

Rate-optimal nonparametric estimation for random coefficient regression models

HAJO HOLZMANN¹ and ALEXANDER MEISTER²

¹*Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, 35037 Marburg, Germany.
E-mail: holzmann@mathematik.uni-marburg.de*

²*Institut für Mathematik, Universität Rostock, 18051 Rostock, Germany.
E-mail: alexander.meister@uni-rostock.de*

Random coefficient regression models are a popular tool for analyzing unobserved heterogeneity, and have seen renewed interest in the recent econometric literature. In this paper, we obtain the optimal pointwise convergence rate for estimating the density in the linear random coefficient model over Hölder smoothness classes, and in particular show how the tail behavior of the design density impacts this rate. In contrast to previous suggestions, the estimator that we propose and that achieves the optimal convergence rate does not require dividing by a nonparametric density estimate. The optimal choice of the tuning parameters in the estimator depends on the tail parameter of the design density and on the smoothness level of the Hölder class, and we also study adaptive estimation with respect to both parameters.

Keywords: adaptive estimation; ill-posed inverse problem; minimax risk; nonparametric estimation

1. Introduction

In this paper, we consider the linear random coefficient regression model, in which i.i.d. (independent and identically distributed) data (X_j, Y_j) , $j = 1, \dots, n$ are observed according to

$$Y_j = A_{0,j} + A_{1,j}X_j. \quad (1.1)$$

Therein $A_j := (A_{0,j}, A_{1,j})$ are unobserved i.i.d. random variables with the bivariate Lebesgue density f_A ; while A_j and X_j are independent. Note that (1.1) represents a randomized extension of the standard linear regression model. We shall derive the optimal convergence rates for estimating f_A over Hölder smoothness classes in case when the X_j have a Lebesgue density f_X with polynomial tail behaviour, as specified in Assumption 1 below.

From a parametric point of view with focus on means and variances of the random coefficients, a multivariate version of model (1.1) is studied by [11]. They assume the coefficients A_j to be mutually independent. The nonparametric analysis of model (1.1) has been initiated by [3] and [4]. [2] use Fourier methods to construct an estimator of f_A . They do not derive the optimal convergence rate, though. Furthermore, their estimator is rather involved as it requires a nonparametric estimator of a conditional characteristic function, which is then plugged into a regularized Fourier inversion.

Extensions of model (1.1) have seen renewed interest in the econometrics literature in recent years. [13] suggest a nonparametric estimator in a multivariate version of model (1.1). They only obtain its convergence rate for very heavy tailed regressors. Moreover, their estimator requires dividing by a nonparametric density estimator for a transformed version of the regressors. This involves an additional smoothing step, and potentially renders the estimator unstable. [5] propose a specification test for model (1.1) against a general nonseparable model as the alternative, while [6] suggest multiscale tests for qualitative hypotheses on f_A . Extensions and modifications of model (1.1) are studied in [1,8–10,17–19] and [12]. Methods of analytic continuation of the coefficients density outside the support of the covariates are considered under more restrictive conditions in [12] and in the recent work of [7].

In this paper, we consider the basic model (1.1) under the following condition.

Assumption 1 (Design density). For some constants $\beta > 0$ and $C_X > c_X > 0$, the density f_X satisfies

$$C_X(1 + |x|)^{-\beta-2} \geq f_X(x) \geq c_X \cdot (1 + |x|)^{-\beta-2}, \quad \forall x \in \mathbb{R}, \tag{1.2}$$

We analyze precisely how the tail parameter β of f_X influences the optimal rate of convergence of f_A at a given point $a \in \mathbb{R}^2$ in a minimax sense in case $\beta > 1$. Note that the heavy tailed setting which is studied in [13] corresponds to $\beta = 0$ in Assumption 1. To our best knowledge a rigorous study of the minimax convergence rate in the more realistic case of $\beta > 1$ has been missing so far. Indeed we fill this gap and derive optimal rates, which are fundamentally new and not known from any other nonparametric estimation problem.

The estimator which we propose is inspired by [12]. It achieves the optimal convergence rate and does not require dividing by a nonparametric density estimator. Instead we exploit the order statistic of the transformed design variables in a Priestley-Chao manner. The optimal choice of the tuning parameters depends both on the two parameters β and on the smoothness parameter of the Hölder class, which is reminiscent of the estimation problem in [14] and in contrast to usual adaptation problems in nonparametric curve estimation, in which the smoothing parameters shall adapt only to an unknown smoothness level. Here we show how to make the estimator adaptive with respect to both of these parameters.

The paper is organized as follows. In Section 2, we introduce our estimation procedure. Section 3 is devoted to upper and lower risk bounds, which yield minimax rate optimality for the pointwise risk. We also derive an upper risk bound for the uniform risk, here, an additional logarithmic factor occurs. In Section 4, we deal with adaptivity. The proofs and technical lemmas are deferred to Section 5.

Let us fix some notation: ψ_A denotes the characteristic function of the A_j , while $\psi_{U|Z}$ is the conditional characteristic function of the random variable U given the random variable Z . Throughout $|\cdot|$ stands for the Euclidean norm of a real or complex vector, and $\mathbb{1}(A)$ denotes the indicator function of the event A . For positive sequences (a_n) and (b_n) we write $a_n \asymp b_n$ if $ca_n \leq b_n \leq Ca_n, n \in \mathbb{N}$ for constants $0 < c < C$.

2. The estimator

In order to construct an estimator for f_A in model (1.1), we transform the data (X_j, Y_j) into (Z_j, U_j) via

$$U_j = Y_j / \sqrt{1 + X_j^2}, \quad (\cos Z_j, \sin Z_j) = (1, X_j) / \sqrt{1 + X_j^2},$$

so that $Z_j \in (-\pi/2, \pi/2)$ almost surely (a.s.), Z_j and A_j are independent, and

$$U_j = A_{0,j} \cos Z_j + A_{1,j} \sin Z_j. \tag{2.1}$$

Then the conditional characteristic function $\psi_{U|Z}(\cdot|z)$ of U_j given $Z_j = z$ equals

$$\psi_{U|Z}(t|z) = \psi_A(t \cos z, t \sin z). \tag{2.2}$$

By Fourier inversion, integral substitution into polar coordinates (with signed radius) and (2.2) we deduce that

$$\begin{aligned} f_A(a) &= \frac{1}{(2\pi)^2} \iint \exp(-ia'b) \psi_A(b) db \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{-\pi/2}^{\pi/2} |t| \exp(-it(a_0 \cos z + a_1 \sin z)) \psi_{U|Z}(t|z) dz dt. \end{aligned} \tag{2.3}$$

The equation (2.3) motivates us to estimate f_A by an empirical version of the conditional characteristic function $\psi_{U|Z}$ which is directly accessible from the data (Z_j, U_j) . For that purpose choose a function w which satisfies the following assumption.

Assumption 2 (Kernel). For a number $\ell \in \mathbb{N}_0$ the function $w : \mathbb{R} \rightarrow \mathbb{R}$ is even, supported on $[-1, 1]$, $(\ell + 1)$ -fold continuously differentiable on the whole real line, satisfies $w(0) = 1$ as well as $w^{(k)}(0) = 0$ for all $k = 1, \dots, \ell$, and $|w|$ is bounded by 1.

Assumption 2 could be relaxed somewhat. In particular, we may assume compact support instead of imposing the support of w to be a subset of $[-1, 1]$ and we may remove the condition that $|w|$ is bounded by 1. Simple boundedness is sufficient, which follows from the other conditions.

Now we consider the regularized version of f_A by kernel smoothing as follows

$$\begin{aligned} \tilde{f}_A(a; h) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{-\pi/2}^{\pi/2} w(th) |t| \exp(-it(a_0 \cos z + a_1 \sin z)) \psi_{U|Z}(t|z) dz dt \\ &= \int_{-\pi/2}^{\pi/2} \int_{\mathbb{R}} K(u - a_0 \cos z - a_1 \sin z; h) f_{U|Z}(u|z) du dz, \end{aligned} \tag{2.4}$$

where

$$K(x; h) := \frac{1}{(2\pi)^2} \int_{\mathbb{R}} w(th)|t| \exp(itx) dt = \frac{2}{(2\pi)^2} \int_0^\infty w(th)t \cos(tx) dt. \tag{2.5}$$

Inspired by (2.4) we introduce a Priestley-Chao type estimator of the density f_A ,

$$\begin{aligned} \hat{f}_A(a; h, \delta) &= \sum_{j=1}^{n-1} K(U_{[j]} - a_0 \cos Z_{(j)} - a_1 \sin Z_{(j)}; h)(Z_{(j+1)} - Z_{(j)}) \\ &\quad \cdot \mathbb{1}(-\pi/2 + \delta \leq Z_{(j)} \leq Z_{(j+1)} \leq \pi/2 - \delta) \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} w(th)|t| \sum_{j=1}^{n-1} \exp(it(U_{[j]} - a_0 \cos Z_{(j)} - a_1 \sin Z_{(j)})) \\ &\quad \cdot (Z_{(j+1)} - Z_{(j)}) \mathbb{1}(-\pi/2 + \delta \leq Z_{(j)} \leq Z_{(j+1)} \leq \pi/2 - \delta) dt, \end{aligned} \tag{2.6}$$

where $(U_{[j]}, Z_{(j)})$, $j = 1, \dots, n$, denotes the sample (U_j, Z_j) , $j = 1, \dots, n$, sorted such that $Z_{(1)} \leq \dots \leq Z_{(n)}$, and where $h = h_n > 0$ is a classical bandwidth parameter and $\delta = \delta_n \geq 0$ is a threshold parameter both of which remain to be selected. By the parameter δ we cut off that subset of the interval $[-\pi/2, \pi/2]$ in which the Z_j are sparse.

In the following, we shall use the symbol

$$\sum_{j,n,\delta} := \sum_{j \in \{1, \dots, n\}, -\pi/2 + \delta \leq Z_{(j)} \leq Z_{(j+1)} \leq \pi/2 - \delta} \tag{2.7}$$

to denote the sum over the random set of indices $1 \leq j \leq n - 1$ for which $-\pi/2 + \delta \leq Z_{(j)} \leq Z_{(j+1)} \leq \pi/2 - \delta$. Thus, we may write the estimator in (2.6) as

$$\hat{f}_A(a; h, \delta) = \sum_{j,n,\delta} K(U_{[j]} - a_0 \cos Z_{(j)} - a_1 \sin Z_{(j)}; h)(Z_{(j+1)} - Z_{(j)}).$$

In this paper, we consider one-dimensional covariates Z_1, \dots, Z_n only. From a methodological point of view, the estimator (2.6) could be extended to the multivariate setting by using Voronoi cells instead of the order statistics. A similar technique is proposed in eq. (36) in [12]. On the other hand, the asymptotic properties of such an estimator might be completely different from the univariate case.

3. Upper and lower risk bounds

We consider the following Hölder smoothness class of densities.

Definition. For a point $a = (a_0, a_1) \in \mathbb{R}^2$, a smoothness index $\alpha > 0$ and constants $c_A, c_B, r_A, c_M > 0$ define the class $\mathcal{F} = \mathcal{F}(a, \alpha, c_A, c_B, r_A, c_M)$ of densities as follows: $f_A \in \mathcal{F}(a, \alpha, c_A, c_B,$

r_A, c_M) is Hölder-smooth of the degree α in the neighborhood $U_{r_A}(a) = \{b \in \mathbb{R}^2 \mid |a - b| < r_A\}$, that is, f_A is $s = \lfloor \alpha \rfloor = \max\{k \in \mathbb{N}_0 \mid k < \alpha\}$ -times continuously differentiable in $U_{r_A}(a)$ and its partial derivatives satisfy

$$\left| \frac{\partial^s f_A}{\partial x^k \partial y^{s-k}}(x, y) - \frac{\partial^s f_A}{\partial x^k \partial y^{s-k}}(a_0, a_1) \right| \leq c_A \cdot |(x, y) - a|^{\alpha-s}, \tag{3.1}$$

for all $k = 0, \dots, s$ and $(x, y) \in U_{r_A}(a)$. Furthermore, assume that the Fourier transform ψ_A of f_A is weakly differentiable and its weak derivative $\nabla \psi_A$ satisfies

$$\int \operatorname{ess\,sup}_{y \in \mathbb{R}} |\nabla \psi_A(x, y)| \, dx \leq c_B, \tag{3.2}$$

and that $f_A(a) \leq c_M$ for all $a \in \mathbb{R}^2$.

For the proof of the first theorem, the global partial tail and smoothness condition (3.2) of the order 1 is required in addition to the local smoothness assumption (3.1) of the order α . The theorem provides an upper bound on the convergence rate for the estimator in (2.6).

Theorem 3.1. *Consider model (1.1) and assume that f_X satisfies (1.2) for some $\beta > 1$. If w satisfies Assumption 2 for $l \geq 2\lfloor \alpha \rfloor$, and if $\delta = \delta_n$ and $h = h_n$ are chosen such that*

$$\delta \asymp n^{-\frac{1}{\beta+1}} \quad \text{and} \quad h \asymp n^{-\frac{1}{(\alpha+2)(\beta+1)}},$$

then the estimator (2.6) attains the following asymptotic risk upper bound over the function class $\mathcal{F} = \mathcal{F}(a, \alpha, c_A, c_B, r_A, c_M)$,

$$\sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} [|\hat{f}_A(a; h, \delta) - f_A(a)|^2] = \mathcal{O}\left(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}}\right).$$

The following theorem yields that the convergence rates which our estimator (2.6) achieves according to Theorem 3.1 are optimal for the pointwise risk in the minimax sense.

Theorem 3.2. *Fix $a = 0$ and the constants c_A, c_B sufficiently large for any $\alpha > 0$ and $\beta > 1$. Let $(\hat{f}_n)_n$ be an arbitrary sequence of estimators of f_A , where \hat{f}_n is based on the data (X_j, Y_j) , $j = 1, \dots, n$, for each n . Assume that f_X satisfies (1.2). Then*

$$\liminf_{n \rightarrow \infty} n^{\frac{2\alpha}{(\alpha+2)(\beta+1)}} \sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} [|\hat{f}_n(0) - f_A(0)|^2] > 0.$$

The convergence rates from Theorems 3.1 and 3.2 differ significantly from standard rates in nonparametric estimation. While they become faster as α increases, they become slower as β gets larger. It is remarkable that they do not approach the (squared) parametric rate n^{-1} but the slower rate $n^{-2/(\beta+1)}$ for large α .

The case $\beta \leq 1$. An analysis of the proof of Theorem 3.1 shows that in case $\beta < 1$, choosing $\delta \asymp n^{-\frac{1}{\beta+1}}$ and $h \asymp n^{-\frac{1}{2\alpha+4}}$ gives the rate

$$\sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} [|\hat{f}_A(a; h, \delta) - f_A(a)|^2] = \mathcal{O}(n^{-\frac{2\alpha}{2\alpha+4}});$$

in case $\beta = 1$, an additional logarithmic factor occurs. The upper bound no longer depends on β in this regime. For $\beta = 0$, [13] obtain the faster rate $\mathcal{O}(n^{-\frac{2\alpha}{2\alpha+3}})$; their rate is in \mathcal{L}_2 but could be transferred to a pointwise rate. However, they additionally impose the assumption that the density f_A is uniformly bounded with a bounded support. This implies that $f_{U|Z}$ is also uniformly bounded. Under this additional assumption, instead of (5.4) in our analysis, we have the sharper bound

$$\text{Var}_{f_A}(\hat{f}_A(a; h, \delta) | \sigma_Z) \leq \text{const.} \cdot h^{-3} \cdot \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2$$

since $\int_{\mathbb{R}} K^2(u; h) du \leq \text{const.} \cdot h^{-3}$. Then one can show that our estimator also achieves the rate $\mathcal{O}(n^{-\frac{2\alpha}{2\alpha+3}})$ for $\beta = 0$, even with the choice $\delta = 0$.

Finally, we consider the uniform rate of convergence, again in the case $\beta > 1$.

Theorem 3.3. Consider model (1.1) and assume that f_X satisfies (1.2) for some $\beta > 1$. Suppose that w satisfies Assumption 2 for $l \geq 2\lfloor \alpha \rfloor$, and that $\delta = \delta_n$ and $h = h_n$ are chosen such that

$$\delta \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{\beta+1}} \quad \text{and} \quad h \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{(\alpha+2)(\beta+1)}}.$$

For a compact rectangle $K \subseteq \mathbb{R}^2$ let $\mathcal{F}(K, \alpha, c_A, c_B, r_A, c_M)$ denote the class of densities on \mathbb{R}^2 such that $f \in \mathcal{F}(a, \alpha, c_A, c_B, r_A, c_M)$ for each $a \in K$. Then the estimator (2.6) attains the following uniform asymptotic risk upper bound over the function class $\mathcal{F} = \mathcal{F}(K, \alpha, c_A, c_B, r_A, c_M)$,

$$\sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} \left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - f_A(a)|^2 \right] = \mathcal{O} \left(\left(\frac{\log n}{n} \right)^{\frac{2\alpha}{(\alpha+2)(\beta+1)}} \right).$$

4. Adaptation

4.1. Adaptation with respect to β for given smoothness

Assume that (1.2) holds with unknown $\beta > 1$. If there are at least two observations Z_j in the interval $[-\pi/2 + \delta, \pi/2 - \delta]$ so that $\sum_{j,n,\delta}$ is not the sum over the empty set, we set

$$L_n(\delta) = \min\{Z_j \mid Z_j \geq -\pi/2 + \delta\}, \quad R_n(\delta) = \max\{Z_j \mid Z_j \leq \pi/2 - \delta\}, \quad (4.1)$$

otherwise we put $L_n(\delta) = -\pi/2$ and $R_n(\delta) = \pi/2$. To define a selection rule for δ , define the function

$$C_n(\delta) := \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2 + \delta^{-1} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3 + (L_n(\delta) + \pi/2)^2 + (\pi/2 - R_n(\delta))^2 + \delta^2,$$

which is continuous except at the sites $\pi/2$, $Z_j + \pi/2$ and $\pi/2 - Z_j$ for $j = 1, \dots, n$. Now choose $\delta = \hat{\delta}_n$ in the interval $[n^{-1/2}, \pi/4]$ such that

$$C_n(\hat{\delta}_n) \leq \exp(-n) + \inf_{\delta \in [n^{-1/2}, \pi/4]} C_n(\delta). \tag{4.2}$$

The next proposition shows that the convergence rate from Theorem 3.1 does not deteriorate if only β is unknown but α is known.

Proposition 4.1. *Consider model (1.1) and assume that f_X satisfies (1.2) for some unknown $\beta > 1$. Choose w satisfying the Assumption 2 for $2[\alpha] \leq l$ for given $\alpha > 0$. If $\hat{\delta}_n$ is chosen in (4.2) and*

$$\hat{h}_n = (C_n(\hat{\delta}_n))^{\frac{1}{2(\alpha+2)}},$$

then for the estimator $\hat{f}_A(a; \hat{h}_n, \hat{\delta}_n)$ we have that

$$\sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} [|\hat{f}_A(a; \hat{h}_n, \hat{\delta}_n) - f_A(a)|^2] = \mathcal{O}(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}}),$$

where $\mathcal{F} = \mathcal{F}(a, \alpha, c_A, c_B, r_A, c_M)$.

4.2. Adaptation by the Lepski method

Finally we consider adaptivity with respect to both parameters β and α based on a combination of Lepski’s method, see [16] and [15], and the choice (4.2). Consider the grid of bandwidths

$$h_k = \hat{\delta}_n^{1/2} q^k, \quad k \in \mathcal{K}_n = \{0, \dots, K\},$$

where $q > 1$, $K = K_n = \lfloor \log_q n \rfloor$ and $\hat{\delta}_n$ is defined in (4.2). Fix $a \in \mathbb{R}^2$ and denote

$$\hat{f}_k = \hat{f}_A(a; h_k, \hat{\delta}_n).$$

For $C_{Lep} > 0$ sufficiently large to be chosen we let

$$\hat{k} = \max\{k \in \mathcal{K}_n \mid |\hat{f}_k - \hat{f}_l|^2 \leq C_{Lep} \sigma(l, n) \forall l \leq k, l \in \mathcal{K}_n\},$$

where

$$\sigma(k, n) = h_k^{-4} C_n(\hat{\delta}_n) \log n, \quad k \in \mathcal{K}_n.$$

Theorem 4.1. Consider model (1.1) and assume that f_X satisfies (1.2) for some unknown $\beta > 1$. Choose w according to Assumption 2 for some $l \in \mathbb{N}_0$. Then for sufficiently large $C_{\text{Lep}} > 0$ (e.g. $C_{\text{Lep}} = 20^2$ suffices), we have, for every $\alpha > 0$ with $2\lfloor \alpha \rfloor \leq l$, that

$$\sup_{f_A \in \mathcal{F}} \mathbb{E}_{f_A} [|\hat{f}_A(a; h_{\hat{k}}, \hat{\delta}_n) - f_A(a)|^2] = \mathcal{O}\left(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}} (\log n)^{\frac{\alpha}{\alpha+2}}\right),$$

where $\mathcal{F} := \mathcal{F}(a, \alpha, c_A, c_B, r_A, c_M)$.

Thus for adaptivity an additional logarithmic factor occurs in the pointwise rate under Hölder smoothness constraints.

5. Proofs

In the proofs we drop $f_A \in \mathcal{F}$ in $\mathbb{E} = \mathbb{E}_{f_A}$ and in $\mathbb{P} = \mathbb{P}_{f_A}$ from the notation.

5.1. Proofs for Section 3

Proof of Theorem 3.1. By passing to Cartesian coordinates in (2.4), we can write

$$\begin{aligned} \tilde{f}_A(a; h) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-ia'b) \psi_A(b) w(h\|b\|) db = (f_A * \tilde{w}(\cdot/h)/h^2)(a), \\ \tilde{w}(a) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-ia'b) w(\|b\|) db. \end{aligned} \tag{5.1}$$

Assumption 2 guarantees that \tilde{w} is a kernel of order ℓ . Then, using Taylor approximation as usual in kernel regularization, see pages 37–38 in [20] for the argument in case of non-compactly supported kernels, the following asymptotic rate of the regularization bias term occurs

$$\begin{aligned} |f_A(a) - \tilde{f}_A(a; h)| &= \left| f_A(a) - \int \tilde{w}(z) f_A(a - hz) dz \right| \\ &\leq C_{\text{Bias}}(\alpha, w, c_A, c_M) \cdot h^\alpha, \end{aligned} \tag{5.2}$$

where the constant factor $C_{\text{Bias}}(\alpha, w, c_A, c_M)$ only depends on c_A, c_M, w and α .

Now let σ_Z denote the σ -field generated by Z_1, \dots, Z_n , and consider the conditional bias-variance decomposition

$$\begin{aligned} \mathbb{E}[|\hat{f}_A(a; h, \delta) - \tilde{f}_A(a; h)|^2] &= \mathbb{E}[\text{Var}(\hat{f}_A(a; h, \delta) | \sigma_Z)] \\ &\quad + \mathbb{E}[|\mathbb{E}[\hat{f}_A(a; h, \delta) | \sigma_Z] - \tilde{f}_A(a; h)|^2]. \end{aligned} \tag{5.3}$$

Since $U_{[1]}, \dots, U_{[n]}$ are independent given σ_Z , observing from (2.5) that $\|K(\cdot; h)\|_\infty = \mathcal{O}(h^{-2})$, we may bound

$$\begin{aligned} \text{Var}(\hat{f}_A(a; h, \delta) | \sigma_Z) &\leq \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2 \\ &\quad \cdot \int_{\mathbb{R}} K^2(u - a_0 \cos Z_{(j)} - a_1 \sin Z_{(j)}; h) f_{U|Z}(u | Z_{(j)}) \, du \\ &\leq \text{const.} \cdot h^{-4} \cdot \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2, \end{aligned} \tag{5.4}$$

where the constant factor only depends on w . Therein we use the notation (2.7). For the conditional expectation, we obtain that

$$\mathbb{E}[\hat{f}_A(a; h, \delta) | \sigma_Z] = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} w(th) |t| \int_{-\pi/2}^{\pi/2} \tilde{\psi}(t, z) \, dz \, dt,$$

where we set

$$\tilde{\psi}(t, z) = \sum_{j,n,\delta} \psi_{U|Z}(t | Z_{(j)}) \exp(-it a_0 \cos Z_{(j)} - it a_1 \sin Z_{(j)}) \mathbb{1}(Z_{(j)} \leq z \leq Z_{(j+1)}).$$

We deduce that

$$|\mathbb{E}[\hat{f}_A(a; h, \delta) | \sigma_Z] - \tilde{f}_A(a; h)|^2 \leq I_1 + I_2 + I_3, \tag{5.5}$$

where

$$\begin{aligned} I_1 &:= \frac{3}{(2\pi)^4} \left| \int_{L_n(\delta)}^{R_n(\delta)} \int_{\mathbb{R}} w(th) |t| (\tilde{\psi}(t, z) - \exp(-it(a_0 \cos z + a_1 \sin z))) \right. \\ &\quad \left. \cdot \psi_{U|Z}(t | z) \, dt \, dz \right|^2 \\ I_2 &:= \frac{3}{(2\pi)^4} \left| \int_{-\pi/2}^{L_n(\delta)} \int_{\mathbb{R}} w(th) |t| \exp(-it(a_0 \cos z + a_1 \sin z)) \right. \\ &\quad \left. \cdot \psi_{U|Z}(t | z) \, dz \, dt \right|^2 \\ I_3 &:= \frac{3}{(2\pi)^4} \left| \int_{R_n(\delta)}^{\pi/2} \int_{\mathbb{R}} w(th) |t| \exp(-it(a_0 \cos z + a_1 \sin z)) \right. \\ &\quad \left. \cdot \psi_{U|Z}(t | z) \, dz \, dt \right|^2, \end{aligned}$$

where $L_n(\delta)$ and $R_n(\delta)$ are defined in (4.1). If there are no two consecutive Z_j in the interval $[-\pi/2 + \delta, \pi/2 - \delta]$, then $\tilde{\psi}(t, z) = 0$ (indeed $\hat{f}_A(a; h, \delta) = 0$). In this case, by our convention we have $L_n(\delta) = -\pi/2$ and $R_n(\delta) = \pi/2$ so that $I_2 = I_3 = 0$ and I_1 is the integral from $-\pi/2$ to $\pi/2$, as required for the estimate (5.5) to remain true in this case.

First, consider the term I_3 . Using the Cauchy–Schwarz inequality, it holds that

$$\begin{aligned}
 I_3 &\leq \frac{3}{(2\pi)^4} \int_{-1/h}^{1/h} t^2 dt \int_{-1/h}^{1/h} \left| \int_{R_n(\delta)}^{\pi/2} \exp(-it(a_0 \cos z + a_1 \sin z)) \right. \\
 &\quad \left. \cdot \psi_{U|Z}(t|z) dz \right|^2 dt \\
 &\leq \frac{4}{(2\pi)^4} \cdot h^{-4} \cdot (\pi/2 - R_n(\delta))^2.
 \end{aligned}
 \tag{5.6}$$

Analogously we establish that

$$I_2 \leq \frac{4}{(2\pi)^4} \cdot h^{-4} \cdot (L_n(\delta) - \pi/2)^2.$$

Finally, consider the term I_1 . In case when there are two consecutive Z_j in the interval $[-\pi/2 + \delta, \pi/2 - \delta]$ so that the sum in (2.7) is not empty, it holds that

$$\begin{aligned}
 I_1 &\leq \frac{3}{(2\pi)^4} h^{-2} \cdot \left\{ \sum_{j,n,\delta} \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{|t| \leq 1/h} |\tilde{\psi}(t, z) - \exp(-it(a_0 \cos z + a_1 \sin z)) \right. \\
 &\quad \left. \cdot \psi_{U|Z}(t|z) dt dz \right\}^2.
 \end{aligned}$$

Now, for $z \in [Z_{(j)}, Z_{(j+1)})$, we get that

$$\begin{aligned}
 &|\tilde{\psi}(t, z) - \exp(-it(a_0 \cos z + a_1 \sin z)) \psi_{U|Z}(t|z)| \\
 &= |\psi_{U|Z}(t|Z_{(j)}) \exp(-ita_0 \cos Z_{(j)} - ita_1 \sin Z_{(j)}) - \psi_{U|Z}(t|z) \\
 &\quad \cdot \exp(-it(a_0 \cos z + a_1 \sin z))| \\
 &\leq |\psi_{U|Z}(t|Z_{(j)}) - \psi_{U|Z}(t|z)| + |t| \cdot |a| \cdot (Z_{(j+1)} - Z_{(j)}) \\
 &= |\psi_A(t \cos Z_{(j)}, t \sin Z_{(j)}) - \psi_A(t \cos z, t \sin z)| + |t| \cdot |a| \cdot (Z_{(j+1)} - Z_{(j)}),
 \end{aligned}$$

according to (2.2). Hence, we may bound

$$\begin{aligned}
 I_1 &\leq \text{const.} \cdot h^{-2} \left(\left\{ \int_{-\pi/2}^{\pi/2} \int_{|t| \leq h^{-1}} \sum_{j,n,\delta} \mathbb{1}(z \in [Z_{(j)}, Z_{(j+1)}]) |\psi_A(t \cos Z_{(j)}, t \sin Z_{(j)}) \right. \right. \\
 &\quad \left. \left. - \psi_A(t \cos z, t \sin z) \right| dt dz \right\}^2
 \end{aligned}$$

$$\begin{aligned}
 & + \left\{ \int_{-\pi/2}^{\pi/2} \int_{|t| \leq h^{-1}} \sum_{j,n,\delta} \mathbb{1}(z \in [Z_{(j)}, Z_{(j+1)}]) |t| \cdot |a| \cdot (Z_{(j+1)} - Z_{(j)}) \, dt \, dz \right\}^2 \\
 & = \text{const.} \cdot h^{-2} (I_{1,1} + I_{1,2}).
 \end{aligned}$$

Applying the Cauchy–Schwarz inequality gives for $I_{1,2}$

$$I_{1,2} \leq \text{const.} \cdot h^{-4} |a|^4 \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3.$$

For $I_{1,1}$ interchanging sum and integrals we obtain

$$\begin{aligned}
 & \int_{-\pi/2}^{\pi/2} \int_{|t| \leq 1/h} \mathbb{1}(z \in [Z_{(j)}, Z_{(j+1)}]) |\psi_A(t \cos Z_{(j)}, t \sin Z_{(j)}) - \psi_A(t \cos z, t \sin z)| \, dt \, dz \\
 & = \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{|t| \leq 1/h} |\psi_A(t \cos Z_{(j)}, t \sin Z_{(j)}) - \psi_A(t \cos z, t \sin z)| \, dt \, dz \\
 & \leq \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{|t| \leq 1/h} |t| \left| \int_{Z_{(j)}}^z \langle \nabla \psi_A(t \cos u, t \sin u), (-\sin u, \cos u) \rangle \, du \right| \, dt \, dz \\
 & \leq \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{|t| \leq 1/h} |t| \int_{Z_{(j)}}^z \sup_{y \in \mathbb{R}} |\nabla \psi_A(t \cos u, y)| \, du \, dt \, dz \\
 & \leq 2h^{-1} \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{Z_{(j)}}^z \sup_{t \in \mathbb{R}} |\nabla \psi_A(t \cos u, y)| \, dt \, du \, dz \\
 & \leq 2c_B h^{-1} \int_{Z_{(j)}}^{Z_{(j+1)}} \int_{Z_{(j)}}^z \frac{1}{\cos u} \, du \, dz \\
 & = 2c_B h^{-1} (Z_{(j+1)} - Z_{(j)}) \int_{Z_{(j)}}^{Z_{(j+1)}} \frac{1}{\cos u} \, du.
 \end{aligned}$$

Using the Cauchy–Schwarz inequality twice yields

$$\begin{aligned}
 I_{1,1} & \leq \text{const.} \cdot h^{-2} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)}) \int_{Z_{(j)}}^{Z_{(j+1)}} \frac{1}{\cos u} \, du \\
 & \leq \text{const.} \cdot h^{-2} \left(\int_{-\pi/2+\delta}^{\pi/2-\delta} \sum_{j,n,\delta} \mathbb{1}(z \in [Z_{(j)}, Z_{(j+1)}]) (Z_{(j+1)} - Z_{(j)}) \frac{1}{\cos z} \, dz \right)^2 \\
 & \leq \text{const.} \cdot h^{-2} \int_{-\pi/2+\delta}^{\pi/2-\delta} \frac{1}{\cos^2 z} \, dz \int_{-\pi/2+\delta}^{\pi/2-\delta} \sum_{j,n,\delta} \mathbb{1}(z \in [Z_{(j)}, Z_{(j+1)}]) (Z_{(j+1)} - Z_{(j)})^2 \, dz
 \end{aligned}$$

$$\begin{aligned} &\leq \text{const.} \cdot h^{-2} \int_{-\pi/2+\delta}^{\pi/2-\delta} \frac{1}{\cos^2 z} dz \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3 \\ &\leq \text{const.} \cdot \delta^{-1} h^{-2} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3. \end{aligned}$$

Hence, the term I_1 obeys the upper bound

$$I_1 \leq \text{const.} \cdot (|a|^2 \cdot h^{-6} + \delta^{-1} h^{-2}) \cdot \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3.$$

Finally, if there are no two consecutive Z_j in the interval $[-\pi/2 + \delta, \pi/2 - \delta]$, we simply have $I_1 \leq |\tilde{f}_A(a; h)|^2 \leq f_A(a)^2 + \text{const.} \cdot h^{2\alpha} \leq \text{const.}$ Collecting the terms that bound (5.5) and using (5.4), from (5.3) we obtain that

$$\begin{aligned} &\mathbb{E}[|\hat{f}_A(a; h, \delta) - \tilde{f}_A(a; h)|^2 | \sigma_Z] \\ &\leq \text{const.} \cdot h^{-4} \left\{ (\pi/2 - R_n(\delta))^2 + (L_n(\delta) + \pi/2)^2 \right. \\ &\quad \left. + \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2 + \delta^{-1} \cdot \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3 \right\} \\ &\quad + \text{const.} \cdot \left\{ |a|^2 h^{-6} \cdot \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3 \right. \\ &\quad \left. + \mathbb{1}(Z_{(j)} < -\pi/2 + \delta \text{ or } Z_{(j+1)} > \pi/2 - \delta \forall j = 1, \dots, n-1) \right\}. \end{aligned} \tag{5.7}$$

Here, the last term takes care of the event in which the sum $\sum_{j,n,\delta}$ is empty and the estimator actually is zero. In order to bound the terms in (5.7) involving the order statistics, we note that since $\beta > 1$,

$$\int_{\delta}^{\pi/2} u^{-\beta} du \asymp \delta^{1-\beta}, \quad \int_{\delta}^{\pi/2} u^{-2\beta} du \asymp \delta^{1-2\beta}.$$

From (5.2) and (5.7) and Lemma 5.1, we obtain for $\delta \leq \pi/4$ that

$$\begin{aligned} &\mathbb{E}[|\hat{f}_A(a; h, \delta) - f_A(a)|^2] \\ &\leq \text{const.} \cdot \left\{ h^{2\alpha} + h^{-4} \left(\delta + \frac{1}{c_Z n \delta^\beta} \right)^2 + h^{-4} n^{-1} \delta^{1-\beta} + h^{-4} \delta^{-1} n^{-2} \delta^{1-2\beta} \right. \\ &\quad \left. + h^{-6} n^{-2} \delta^{1-2\beta} + n \exp(-c_Z(n-1)(\pi/4)^\beta) \right\}. \end{aligned}$$

Upon inserting the rates for δ and h we obtain the result. □

Proof of Theorem 3.2. We introduce the functions

$$f_{A,\theta}(a_0, a_1) := \alpha_n \beta_n f_0(\alpha_n a_0, \beta_n a_1) + c_L \cdot \theta \cdot \cos(2\beta_n a_1) \cdot \alpha_n \beta_n \phi(\alpha_n a_0, \beta_n a_1),$$

for $\theta \in \{0, 1\}$, some constant $c_L > 0$ and some sequences $(\alpha_n)_n \downarrow 0$ and $(\beta_n)_n \uparrow \infty$ which remain to be selected; moreover we specify

$$f_0(a_0, a_1) := \frac{1}{\pi^2(1 + a_0^2)(1 + a_1^2)},$$

and

$$\phi(a_0, a_1) := \varphi(a_0)\varphi(a_1),$$

where

$$\varphi(x) := \frac{1 - \cos(x)}{\pi x^2}.$$

We verify that $f_{A,0}$ is a probability density as f_0 and φ are probability densities. The Fourier transform of $f_{A,\theta}$ equals

$$\begin{aligned} f_{A,\theta}^{ft}(x, y) &= f_0^{ft}(x/\alpha_n, y/\beta_n) + \frac{1}{2}c_L \cdot \theta \cdot \phi^{ft}(x/\alpha_n, (y + 2\beta_n)/\beta_n) \\ &\quad + \frac{1}{2}c_L \cdot \theta \cdot \phi^{ft}(x/\alpha_n, (y - 2\beta_n)/\beta_n), \end{aligned}$$

so that

$$\iint f_{A,\theta}(a_0, a_1) da_0 da_1 = f_{A,\theta}^{ft}(0) = f_{A,0}^{ft}(0) = 1,$$

since φ^{ft} is supported on the interval $[-1, 1]$. Choosing the constant $c_L > 0$ sufficiently small we can guarantee that $f_{A,1}$ is a non-negative function and satisfies the inequality

$$f_{A,1}(a_0, a_1) \geq c_L^* \alpha_n \beta_n f_0(\alpha_n a_0, \beta_n a_1) \geq 0, \quad \forall a_0, a_1 \in \mathbb{R}, \tag{5.8}$$

for some constant $c_L^* \in (0, 1)$. Thus, $f_{A,1}$ is a probability density as well. Furthermore, we verify that $f_{A,\theta} \in \mathcal{F}$ for both $\theta \in \{0, 1\}$ under the constraint

$$\alpha_n \asymp \beta_n^{-\alpha-1}, \tag{5.9}$$

as c_A and c_B may be viewed as sufficiently large. Therein note that (3.2) is satisfied as $\psi_{A,\theta}$ can be written as the sum of two functions $(x, y) \mapsto \psi_0(x/\alpha_n) \cdot \psi_1(y/\beta_n)$ where $\psi_j, j = 0, 1$ are bounded, weakly differentiable, integrable functions whose weak derivatives are essentially bounded and integrable as well.

The squared pointwise distance between $f_{A,0}$ and $f_{A,1}$ at 0 equals

$$|f_{A,0}(0) - f_{A,1}(0)|^2 = c_L^2 \alpha_n^2 \beta_n^2 / (4\pi^2) \asymp \beta_n^{-2\alpha}. \tag{5.10}$$

Using (5.8), the conditional density of Y_j given X_j under the parameter θ equals

$$\begin{aligned} f_{Y_j|X_j,\theta}(y) &= \int f_{A,\theta}(y - a_1 X_j, a_1) da_1 \geq c_L^* \alpha_n \beta_n \int f_0(\alpha_n(y - a_1 X_j), \beta_n a_1) da_1 \\ &\geq \frac{c_L^* \alpha_n \beta_n}{\pi^2} \int \frac{1}{1 + 2\alpha_n^2 y^2 + 2\alpha_n^2 a_1^2 X_j^2} \cdot \frac{1}{1 + \beta_n^2 a_1^2} da_1 \\ &\geq \frac{c_L^* \alpha_n \beta_n}{2\pi^2} \int_0^{1/\beta_n} \frac{1}{1 + 2\alpha_n^2 y^2 + 2\alpha_n^2 a_1^2 X_j^2} da_1 \\ &\geq \frac{c_L^*}{2\pi^2} \cdot \frac{\alpha_n}{1 + 2\alpha_n^2 y^2 + 2X_j^2 \alpha_n^2 / \beta_n^2}, \end{aligned}$$

for all $y \in \mathbb{R}$. Moreover, we have that

$$\begin{aligned} f_{Y_j|X_j,1}(y) - f_{Y_j|X_j,0}(y) &= c_L \alpha_n \beta_n \int \cos(2\beta_n a_1) \cdot \phi(\alpha_n(y - a_1 X_j), \beta_n a_1) da_1 \\ &= c_L \alpha_n \int \cos(2a_1 \beta_n / \beta_n) \cdot \phi(\alpha_n(y - a_1 X_j / \beta_n), a_1) da_1, \end{aligned}$$

where the Fourier transform equals

$$\begin{aligned} f_{Y_j|X_j,1}^{ft}(t) - f_{Y_j|X_j,0}^{ft}(t) &= \frac{1}{2} c_L \phi^{ft}(t/\alpha_n, (tX_j + 2\beta_n)/\beta_n) \\ &\quad + \frac{1}{2} c_L \phi^{ft}(t/\alpha_n, (tX_j - 2\beta_n)/\beta_n). \end{aligned}$$

Therefore the χ^2 -distance between the competing observation densities is bounded from above as follows,

$$\begin{aligned} &c_L^* \cdot \chi^2(f_{Y_j|X_j,\theta=0}, f_{Y_j|X_j,\theta=1}) \\ &\leq (1/\alpha_n + 2X_j^2 \alpha_n / \beta_n^2) c_L^2 \int |\phi^{ft}(t/\alpha_n, (tX_j + 2\beta_n)/\beta_n)|^2 dt \\ &\quad + (1/\alpha_n + 2X_j^2 \alpha_n / \beta_n^2) c_L^2 \int |\phi^{ft}(t/\alpha_n, (tX_j - 2\beta_n)/\beta_n)|^2 dt \\ &\quad + 2c_L^2 \alpha_n \int \left| \frac{d}{dt} \phi^{ft}(t/\alpha_n, (tX_j + 2\beta_n)/\beta_n) \right|^2 dt \\ &\quad + 2c_L^2 \alpha_n \int \left| \frac{d}{dt} \phi^{ft}(t/\alpha_n, (tX_j - 2\beta_n)/\beta_n) \right|^2 dt, \end{aligned} \tag{5.11}$$

where

$$\begin{aligned} \frac{d}{dt} \phi^{ft}(t/\alpha_n, (tX_j \pm 2\beta_n)/\beta_n) &= \alpha_n^{-1} \{\varphi^{ft}\}'(t/\alpha_n) \cdot \{\varphi^{ft}\}((tX_j \pm 2\beta_n)/\beta_n) \\ &\quad + \frac{X_j}{\beta_n} \cdot \{\varphi^{ft}\}(t/\alpha_n) \cdot \{\varphi^{ft}\}'((tX_j \pm 2\beta_n)/\beta_n). \end{aligned}$$

Moreover, this choice also guarantees that $f_{C,\theta}$ integrates to 1 and, hence, is a probability density. Then the integrals in (5.11) range over a subset of

$$[-\alpha_n, \alpha_n] \setminus (-\beta_n/|X_j|, \beta_n/|X_j|)$$

if H_0^{ft} and its (weak) derivative are supported on $[-1, 1]$. Also these functions are uniformly bounded by 1. Thus the integrals vanish whenever $|X_j| < \beta_n/\alpha_n$. It follows that

$$\chi^2(f_{Y_j|X_j, \theta=0}, f_{Y_j|X_j, \theta=1}) \leq (6 + 8X_j^2\alpha_n^2/\beta_n^2)c_L^2/c_L^*,$$

if $|X_j| \geq \beta_n/\alpha_n$; and $\chi^2(f_{Y_j|X_j, \theta=0}, f_{Y_j|X_j, \theta=1}) = 0$ otherwise. According to standard arguments from decision theory, (5.10) represents a lower bound on the attainable rate if the Hellinger distance between the competing data distributions $f_{X,Y;\theta}^{(n)}$ (for $\theta = 0$ and θ_1 , respectively) obeys an upper bound which is smaller than 1 – uniformly with respect to n , see, for example, [21]. Writing \mathcal{H} for the Hellinger distance, it holds that

$$\begin{aligned} \mathcal{H}^2(f_{X,Y;\theta=0}^{(n)}, f_{X,Y;\theta=1}^{(n)}) &\leq \sum_{j=1}^n \mathcal{H}^2(f_{X_j} f_{Y_j|X_j,0}, f_{X_j} f_{Y_j|X_j,1}) \\ &= \sum_{j=1}^n \int f_{X_j}(x) \int (f_{Y_j|X_j,0}^{1/2}(y|x) - f_{Y_j|X_j,1}^{1/2}(y|x))^2 dy dx \\ &\leq \mathbb{E} \left[\sum_{j=1}^n \chi^2(f_{Y_j|X_j,0}, f_{Y_j|X_j,1}) \right] \end{aligned} \tag{5.12}$$

as the distribution of the X_j is identical for $\theta = 0$ and $\theta = 1$. Then, the term (5.12) is bounded from above by

$$\begin{aligned} &\mathbb{E} \left[\sum_{j=1}^n 1_{[\beta_n/\alpha_n, \infty)}(|X_j|) \cdot (6 + 8X_j^2\alpha_n^2/\beta_n^2)c_L^2 \right] \\ &= 6nc_L^2 \int_{|x| \geq \beta_n/\alpha_n} f_X(x) dx + 8nc_L^2\alpha_n^2\beta_n^{-2} \int_{|x| \geq \beta_n/\alpha_n} x^2 f_X(x) dx \\ &= \mathcal{O}(n(\beta_n/\alpha_n)^{-\beta-1}) = \mathcal{O}(n \cdot \beta_n^{-(\alpha+2)(\beta+1)}), \end{aligned} \tag{5.13}$$

as $\beta > 1$. We choose $\beta_n \asymp n^{1/[(2+\alpha)(1+\beta)]}$ so that the χ^2 -distance between the joint densities of the observations under $\theta = 0$ and $\theta = 1$ in (5.13) is bounded from above as n tends to infinity. By elementary decision theoretic arguments and by (5.10), a lower bound on the attainable convergence rate is given by

$$\beta_n^{-2\alpha} \asymp n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}},$$

which completes the proof of the theorem. \square

Proof of Theorem 3.3. We estimate

$$\begin{aligned} \mathbb{E}\left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - f_A(a)|^2\right] &\leq 2\mathbb{E}\left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - \tilde{f}_A(a; h)|^2\right] \\ &\quad + 2 \sup_{a \in K} |\tilde{f}_A(a; h) - f_A(a)|^2, \end{aligned}$$

where $\tilde{f}_A(a; h)$ is defined in (5.1). The second term – the regularization bias – is bounded in (5.2), and that bound is uniform in $a \in K$ from the assumptions on the function class $\mathcal{F}(K, \alpha, c_A, c_B, r_A, c_M)$. For the first term we have, similarly to (5.3), that

$$\begin{aligned} \mathbb{E}\left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - \tilde{f}_A(a; h, \delta)|^2\right] &\leq 2\mathbb{E}\left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]|^2\right] \\ &\quad + 2\mathbb{E}\left[\sup_{a \in K} |\mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z] - \tilde{f}_A(a; h)|^2\right]. \end{aligned} \quad (5.14)$$

The second term in (5.14) is bounded by

$$\sup_{a \in K} I_1(a) + \sup_{a \in K} I_2(a) + \sup_{a \in K} I_3(a),$$

where $I_j(a)$ are defined as in (5.5), and the dependence on a is stressed in the notation. The bounds on the $I_j(a)$ derived after (5.5) are uniform in a over a bounded set K . Thus, it remains to bound the first term in (5.14).

Given $\epsilon > 0$ let I_ϵ be a subset of K for which the ϵ -balls with centers at points in I_ϵ cover K . It is possible to choose such a set with a cardinality of order $\text{card}(I_\epsilon) \leq C_K \epsilon^{-2}$, where $c_K > 0$ depends on K but not on ϵ . Then

$$\begin{aligned} &\sup_{a \in K} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]|^2 \\ &\leq 2 \sup_{a \in I_\epsilon} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]|^2 \\ &\quad + 2 \sup_{a \in K} \inf_{a' \in I_\epsilon} |\hat{f}_A(a; h, \delta) - E[\hat{f}_A(a; h, \delta)|\sigma_Z] \\ &\quad \quad - (\hat{f}_A(a'; h, \delta) - E[\hat{f}_A(a'; h, \delta)|\sigma_Z])|^2. \end{aligned}$$

Since $\|\partial_x K(x; h)\|_\infty \leq h^{-3}$, see the formula (2.5) for $K(\cdot; h)$ and the Assumption 2 in w , by Lipschitz-continuity the second term is $\leq 8\epsilon^2 h^{-6}$. From the Hoeffding inequality, since $\|K(\cdot; h)\|_\infty \leq h^{-2}$ we obtain for $t > 0$ that

$$\begin{aligned} & \mathbb{P}(|\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]| \\ & \geq t|\sigma_Z|) \leq 2 \exp\left(-\frac{t^2}{2h^{-4} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2}\right). \end{aligned} \tag{5.15}$$

Set

$$r_n = (\log n) \cdot h^{-4} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^2.$$

Then, for $\kappa > 0$ we estimate

$$\begin{aligned} & \mathbb{E}\left[r_n^{-1} \sup_{a \in I_\epsilon} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]|^2 |\sigma_Z|\right] \\ & \leq \kappa^2 + 2 \int_\kappa^\infty t \mathbb{P}\left(\sup_{a \in I_\epsilon} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]| \geq r_n^{1/2} t |\sigma_Z|\right) dt \\ & \leq \kappa^2 + 4 \text{card}(I_\epsilon) \int_\kappa^\infty t \exp\left(-\frac{t^2 \log n}{2}\right) dt \quad (\text{from (5.15) and union bound}) \\ & \leq \kappa^2 + 4C_K \epsilon^{-2} \exp\left(-\frac{\kappa^2 \log n}{2}\right) (\log n)^{-1}. \end{aligned}$$

Choose $\epsilon = n^{-2}$ and $\kappa = 10^{1/2}$. Then if $h^{-1} = \mathcal{O}(n^{1/2})$ we obtain from Lemma 5.1 that

$$\mathbb{E}\left[\sup_{a \in I_\epsilon} |\hat{f}_A(a; h, \delta) - \mathbb{E}[\hat{f}_A(a; h, \delta)|\sigma_Z]|^2\right] = \mathcal{O}(\mathbb{E}[r_n]) = h^{-4} \delta^{1-\beta} \log(n/n),$$

and overall

$$\begin{aligned} & \mathbb{E}\left[\sup_{a \in K} |\hat{f}_A(a; h, \delta) - f_A(a)|^2\right] \\ & \leq \text{const.} \cdot \left\{ h^{2\alpha} + h^{-4} \delta^{1-\beta} \log(n/n) + n^{-1} + h^{-4} \left(\delta + \frac{1}{c_Z n \delta^\beta}\right)^2 + h^{-4} \delta^{-1} n^{-2} \delta^{1-2\beta} \right. \\ & \quad \left. + h^{-6} n^{-2} \delta^{1-2\beta} + n \exp(-c_Z(n-1)(\pi/4)^\beta) \right\}. \end{aligned}$$

Plugging in the choices of δ and h gives the result. □

5.2. Proofs for Section 4

Proof of Proposition 4.1. From (5.7) and (5.2), we estimate

$$\begin{aligned} & \mathbb{E}[|\hat{f}_A(a; \hat{h}_n, \hat{\delta}_n) - f_A(a)|^2 | \sigma_Z] \\ & \leq \text{const.} \cdot \{ \hat{h}_n^{2\alpha} + \hat{h}_n^{-4} \mathcal{C}_n(\hat{\delta}_n) \} + \text{const.} \cdot \left\{ |a|^2 \hat{h}_n^{-6} \cdot \sum_{j,n,\hat{\delta}_n} (Z_{(j+1)} - Z_{(j)})^3 \right. \\ & \quad \left. + \mathbb{1}(Z_{(j)} < -\pi/2 + \hat{\delta}_n \text{ or } Z_{(j+1)} > \pi/2 - \hat{\delta}_n \forall j = 1, \dots, n-1) \right\}. \end{aligned} \tag{5.16}$$

Observe that from the term δ^2 in the definition of $\mathcal{C}_n(\delta)$,

$$\hat{h}_n^2 = (\mathcal{C}_n(\hat{\delta}_n))^{\frac{1}{\alpha+2}} \geq \hat{\delta}_n^{\frac{2}{\alpha+2}} \geq \hat{\delta}_n.$$

Since $\hat{\delta}_n \leq \pi/4 \leq 1$, and since $\mathcal{C}_n(\delta)$ contains the term $\delta^{-1} \sum_{j,n,\delta} (Z_{(j+1)} - Z_{(j)})^3$, from (5.16) and the choice of \hat{h}_n we obtain the bound

$$\begin{aligned} & \mathbb{E}[|\hat{f}_A(a; \hat{h}_n, \hat{\delta}_n) - f_A(a)|^2 | \sigma_Z] \\ & \leq \text{const.} \cdot \{ [\mathcal{C}_n(\hat{\delta}_n)]^{\frac{\alpha}{\alpha+2}} + \mathbb{1}(Z_{(j)} < -\pi/4 \text{ or } Z_{(j+1)} > \pi/4 \forall j = 1, \dots, n-1) \}. \end{aligned} \tag{5.17}$$

By definition of $\hat{\delta}_n$,

$$\mathcal{C}_n(\hat{\delta}_n) \leq \exp(-n) + \inf_{\delta \in [n^{-1/2}, \pi/4]} \mathcal{C}_n(\delta) \leq \exp(-n) + \mathcal{C}_n(\delta_n)$$

for the deterministic choice $\delta_n = n^{-1/(\beta+1)}$, which is contained in $[n^{-1/2}, \pi/4]$ for sufficiently large n since $\beta > 1$. Further, by Jensen’s inequality, Lemma 5.1 and the choice of δ_n ,

$$\mathbb{E}[(\mathcal{C}_n(\delta_n))^{\frac{\alpha}{\alpha+2}}] \leq (\mathbb{E}[\mathcal{C}_n(\delta_n)])^{\frac{\alpha}{\alpha+2}} = \mathcal{O}(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}}).$$

Substituting these estimates into (5.17), and using (5.28) finally gives

$$\begin{aligned} \mathbb{E}[|\hat{f}_A(a; \hat{h}_n, \hat{\delta}_n) - f_A(a)|^2] & \leq \mathcal{O}(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}}) + \text{const.} \{ [\exp(-n)]^{\frac{\alpha}{\alpha+2}} \\ & \quad + \mathbb{P}(Z_{(j)} < -\pi/4 \text{ or } Z_{(j+1)} > \pi/4 \forall j = 1, \dots, n-1) \} \\ & = \mathcal{O}(n^{-\frac{2\alpha}{(\alpha+2)(\beta+1)}}). \end{aligned} \quad \square$$

Proof of Theorem 4.1. Fix $0 < \alpha$ with $2[\alpha] \leq l$ and $f_A \in \mathcal{F}(a, c_A, c_B, r_A, \alpha, c_M)$, and set

$$b(k, \alpha) = C_{\text{Bias}}^2(\alpha, w, c_A, c_M) h_k^{2\alpha}, \quad k \in \mathcal{K}_n,$$

see the bound for the regularization bias in (5.2). We shall abbreviate $f_A(a) = f$.

On the event

$$\{Z_{(j)} < -\pi/2 + \hat{\delta}_n \text{ or } Z_{(j+1)} > \pi/2 - \hat{\delta}_n \forall j = 1, \dots, n - 1\},$$

where $\hat{f}_{\hat{k}} = 0$, we may estimate

$$\mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 | \sigma_Z] \leq \text{const.} \cdot \mathbb{1}(Z_{(j)} < -\pi/4 \text{ or } Z_{(j+1)} > \pi/4 \forall j = 1, \dots, n - 1)$$

since $\hat{\delta}_n \leq \pi/4$. In the following, suppose that there are two design points Z_j in the interval $[-\pi/2 + \hat{\delta}_n, \pi/2 - \hat{\delta}_n]$. Since $h_k \geq \hat{\delta}_n^{1/2}$ for each $k \in \mathcal{K}_n$, as in the proof of Proposition 4.1 the term involving h_k^{-6} in (5.16) is negligible as compared to that with the factor $\hat{\delta}_n^{-1} h_k^{-4}$. Hence using (5.7) and (5.2), we estimate

$$\mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 | \sigma_Z] \leq \text{const.} \cdot \{b(k, \alpha) + \sigma(k, n)\}. \tag{5.18}$$

Define the ‘oracle index’ k^* by

$$k^* = k_n^*(\alpha) = \max\{k \in \mathcal{K}_n \mid b(k, \alpha) \leq C_{\text{Lep}} \sigma(k, n)/16\}.$$

Note that $b(0, \alpha) = C_{\text{Bias}}^2(\alpha, w, c_A, c_M) \hat{\delta}_n^\alpha \leq \text{const.}$ since $\hat{\delta}_n^\alpha \leq 1$, while $\sigma(0, n) = \delta_n^{-2} C_n(\hat{\delta}_n) \times \log n \geq \log n$ since $C_n(\hat{\delta}_n) \hat{\delta}_n^{-2} \geq 1$ from the definition of $C_n(\delta)$. Further, since by the choice of K we have that $q^K \geq n/q$ we estimate

$$b(K, \alpha) \geq C_{\text{Bias}}^2(\alpha, w, c_A, c_M) \hat{\delta}_n^\alpha (n/q)^{2\alpha} \geq \text{const.} n^{3\alpha/2}$$

since $\hat{\delta}_n^\alpha \geq n^{-\alpha/2}$ by the choice of $\hat{\delta}_n$. Finally,

$$\sigma(K, n) \leq \delta_n^{-2} (q/n)^4 C_n(\hat{\delta}_n) (\log n) \leq \text{const.} n^{-3/2} \log(n),$$

since $C_n(\hat{\delta}_n) \hat{\delta}_n^{-2} \leq \text{const.} \cdot n^{5/2}$ since from the definition of $C_n(\delta)$ and since $\hat{\delta}_n \geq n^{-1/2}$.

Since $b(k, \alpha)$ increase by factors $q^{2\alpha}$ in k , and $\sigma(k, n)$ decrease by factors q^{-4} in k , it follows from the above estimates that $k^* \rightarrow \infty$ and $K - k^* \rightarrow \infty$, and that there are constants $0 < \tilde{c}_1 < \tilde{c}_2$ such that $\tilde{c}_1 \leq \sigma(k^*, n)/b(k^*, \alpha) \leq \tilde{c}_2$. Rearranging yields

$$c_1 (C_n(\hat{\delta}_n) \log n)^{\frac{1}{2(\alpha+2)}} \leq h_{k^*} \leq c_2 (C_n(\hat{\delta}_n) \log n)^{\frac{1}{2(\alpha+2)}} \tag{5.19}$$

for constants $c_2 > c_1 > 0$. We obtain from (5.18) that

$$\mathbb{E}[|\hat{f}_{k^*} - f|^2 | \sigma_Z] \leq \text{const.} \cdot [C_n(\hat{\delta}_n) \log n]^{\frac{\alpha}{\alpha+2}}. \tag{5.20}$$

Now, for $\hat{f}_{\hat{k}}$ we estimate

$$\begin{aligned} \mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 | \sigma_Z] &\leq 2\mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} \leq k^* - 1) | \sigma_Z] \\ &\quad + 2\mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} \geq k^*) | \sigma_Z]. \end{aligned} \tag{5.21}$$

For the second term, we have that

$$\begin{aligned} \mathbb{E}[|f_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} \geq k^*) | \sigma_Z] &\leq 2\mathbb{E}[|\hat{f}_{\hat{k}} - \hat{f}_{k^*}|^2 \mathbb{1}(\hat{k} \geq k^*) | \sigma_Z] \\ &\quad + 2\mathbb{E}[|f_{\hat{k}^*} - f|^2 \mathbb{1}(\hat{k} \geq k^*) | \sigma_Z]. \end{aligned} \tag{5.22}$$

The second term in (5.22) is bounded by (5.20) after a trivial estimate of the indicator. Further, from the definition of \hat{k} and (5.19) we have the bound

$$|\hat{f}_{\hat{k}} - \hat{f}_{k^*}|^2 \mathbb{1}_{\hat{k} \geq k^*} \leq C_{\text{Lep}} \sigma(k^*, n) \leq \text{const.} \cdot [C_n(\delta_n) \log n]^{\frac{\alpha}{\alpha+2}},$$

which also holds in conditional expectation given σ_Z .

For the first term in (5.21), we estimate

$$\begin{aligned} \mathbb{E}[|f_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} \leq k^* - 1) | \sigma_Z] &= \sum_{k=0}^{k^*-1} \mathbb{E}[|f_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} = k) | \sigma_Z] \\ &\leq \sum_{k=0}^{k^*-1} (\mathbb{E}[|f_{\hat{k}} - f|^4 | \sigma_Z])^{1/2} [\mathbb{P}(\hat{k} = k | \sigma_Z)]^{1/2}. \end{aligned} \tag{5.23}$$

Then

$$\{\hat{k} = k\} \subseteq \bigcup_{l=0}^k \{|f_{k+1} - \hat{f}_l|^2 > C_{\text{Lep}} \sigma(l, n)\}, \quad k = 0, \dots, K - 1.$$

Now let

$$p_{l,k} = \mathbb{P}(|f_{\hat{k}} - \hat{f}_l| > C_{\text{Lep}}^{1/2}(\sigma(l, n))^{1/2} | \sigma_Z), \quad 0 \leq l < k \leq k^*.$$

By choice of k^* , for $0 \leq l < k \leq k^*$ we have that

$$b(l, \alpha) \leq b(k, \alpha) \leq C_{\text{Lep}} \sigma(k, n) / 16 \leq C_{\text{Lep}} \sigma(l, n) / 16.$$

Hence, setting $\tilde{f}_k = \tilde{f}_A(a; h_k)$ we may estimate

$$\begin{aligned} |\hat{f}_k - \hat{f}_l| &\leq |\hat{f}_k - \tilde{f}_k| + |\hat{f}_l - \tilde{f}_l| + |\tilde{f}_k - f| + |\tilde{f}_l - f| \\ &\leq |\hat{f}_k - \tilde{f}_k| + |\hat{f}_l - \tilde{f}_l| + b(k, \alpha)^{1/2} + b(l, \alpha)^{1/2} \\ &\leq |\hat{f}_k - \tilde{f}_k| + |\hat{f}_l - \tilde{f}_l| + C_{\text{Lep}}^{1/2} \sigma(l, n)^{1/2} / 2. \end{aligned}$$

Therefore, for $0 \leq l < k \leq k^*$,

$$\begin{aligned} p_{l,k} &\leq \mathbb{P}(|\hat{f}_k - \tilde{f}_k| > C_{\text{Lep}}^{1/2} \sigma(l, n)^{1/2} / 4 | \sigma_Z) \\ &\quad + \mathbb{P}_{f_A}(|\hat{f}_l - \tilde{f}_l| > C_{\text{Lep}}^{1/2} \sigma(l, n)^{1/2} / 4 | \sigma_Z). \end{aligned}$$

Since $\sigma(l, n) > \sigma(k, n)$, $l < k$, it suffices to bound

$$\mathbb{P}(|\hat{f}_l - \tilde{f}_l| > C_{\text{Lep}}^{1/2} \sigma(l, n)^{1/2} / 4 | \sigma_Z), \quad 0 \leq l \leq k^*.$$

By choice of the grid \mathcal{K}_n , $h_l^2 \geq h_0^2 = \hat{\delta}_n$, therefore

$$|\mathbb{E}[\hat{f}_l | \sigma_Z] - \tilde{f}_l| \leq \text{const.} \cdot [h_l^{-4} C_n(\hat{\delta}_n)]^{1/2} \leq \sigma(l, n)^{1/2}$$

for n sufficiently large. Hence,,

$$\mathbb{P}(|\hat{f}_l - \tilde{f}_l| > C_{\text{Lep}}^{1/2} \sigma(l, n)^{1/2} / 4 | \sigma_Z) \leq \mathbb{P}(|\hat{f}_l - \mathbb{E}[\hat{f}_l | \sigma_Z]| > \tilde{C} \sigma(l, n)^{1/2} | \sigma_Z),$$

where $\tilde{C} = (C_{\text{Lep}}^{1/2} / 4 - 1)$. Using the bound $\|K(\cdot; h)\|_\infty \leq h^{-2}$, see the formula (2.5) for $K(\cdot; h)$ and the Assumption 2 in w , we use the conditional Hoeffding inequality in order to estimate

$$\begin{aligned} &\mathbb{P}(|\hat{f}_l - \mathbb{E}[\hat{f}_l | \sigma_Z]| > \tilde{C} \sigma(l, n)^{1/2} | \sigma_Z) \\ &\leq 2 \exp\left(-\frac{\tilde{C}^2 \sigma(l, n)}{2h_l^{-4} \sum_{j,n,\hat{\delta}_n} (Z_{(j+1)} - Z_{(j)})^2}\right) \\ &\leq 2 \exp(-\tilde{C} \log n), \end{aligned}$$

see (5.4), where

$$\tilde{C} = \tilde{C}^2 / 2 = (C_{\text{Lep}}^{1/2} / 4 - 1)^2 / 2 = 8$$

for the choice $C_{\text{Lep}} = 20^2$. Note that in this step, the logarithmic factor is essential.

Hence

$$\mathbb{P}(\hat{k} = k | \sigma_Z) \leq 2Kn^{-8}, \quad k = 0, \dots, k^*,$$

and in (5.23) we obtain the bound

$$\mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 \mathbb{1}(\hat{k} \leq k^* - 1) | \sigma_Z] \leq 2K^{1/2} n^{-8/2} \sum_{k=0}^{k^*-1} (\mathbb{E}[|\hat{f}_k - f|^4 | \sigma_Z])^{1/2}. \quad (5.24)$$

The crude bound

$$\begin{aligned} \mathbb{E}[|\hat{f}_k - f|^4 | \sigma_Z] &\leq \mathbb{E}[|\hat{f}_k|^4 | \sigma_Z] + \text{const.} \leq \text{const.} \cdot h_k^{-8} \\ &\leq \text{const.} \cdot \hat{\delta}_n^{-4} \leq \text{const.} \cdot n^2, \quad k \in \mathcal{K}_n, \end{aligned}$$

now suffices to conclude that for sufficiently large choice of the constant C_{Lep} ,

$$\begin{aligned} \mathbb{E}[|\hat{f}_{\hat{k}} - f|^2 | \sigma_Z] &\leq \mathcal{O}([C_n(\hat{\delta}_n) \log n]^{\frac{\alpha}{\alpha+2}}) + \mathcal{O}(n^{-1}) \\ &\quad + \text{const.} \cdot \mathbb{1}(Z_{(j)} < -\pi/4 \text{ or } Z_{(j+1)} > \pi/4, j = 1, \dots, n-1). \end{aligned}$$

The remainder of the proof is as that of Proposition 4.1. □

5.3. Spacings

As $Z_j = \arctan X_j$ the density of Z_j equals

$$f_Z(z) = f_X(\tan(z)) / \cos^2 z, \quad \forall z \in (-\pi/2, \pi/2),$$

so that (1.2) implies

$$C_Z ||z| - \pi/2|^\beta \geq f_Z(z) \geq c_Z ||z| - \pi/2|^\beta, \quad \forall z \in (-\pi/2, \pi/2), \tag{5.25}$$

for some constants $C_Z, c_Z > 0$.

Lemma 5.1. *If f_X satisfies (1.2) and hence f_Z fulfills (5.25), then for $\kappa > 1$ we have that*

$$\begin{aligned} & \mathbb{E} \left[\sum_{j=1}^{n-1} (Z_{(j+1)} - Z_{(j)})^\kappa \cdot \mathbb{1}(\delta - \pi/2 \leq Z_{(j)}, Z_{(j+1)} \leq \pi/2 - \delta) \right] \\ & \leq 2\kappa C_Z c_Z^{-\kappa} \Gamma(\kappa) n(n-1)^{-\kappa} \int_\delta^{\pi/2} u^{-\beta(\kappa-1)} du. \end{aligned} \tag{5.26}$$

Furthermore,

$$\begin{aligned} & \max(\mathbb{E}[(L_n(\delta) + \pi/2)^2], \mathbb{E}[(R_n(\delta) - \pi/2)^2]) \\ & \leq 2 \left(\delta + \frac{1}{c_Z n \delta^\beta} \right)^2 + \pi^2 \cdot \exp(-c_Z n (\pi/2 - \delta) \delta^\beta) \\ & \leq 32 \left(\delta + \frac{1}{c_Z n \delta^\beta} \right)^2, \quad \delta \leq \pi/4, \end{aligned} \tag{5.27}$$

and for $\delta \leq \pi/4$ that

$$\begin{aligned} & \mathbb{P}(Z_{(j)} < -\pi/2 + \delta \text{ or } Z_{(j+1)} > \pi/2 - \delta, j = 1, \dots, n-1) \\ & \leq n \exp(-c_Z (n-1) (\pi/4)^\beta). \end{aligned} \tag{5.28}$$

Proof of Lemma 5.1. Setting

$$Z_j^* := \begin{cases} Z_j, & \text{if } Z_j \geq Z_k, \forall k = 1, \dots, n, \\ \min\{Z_k : Z_k > Z_j\}, & \text{otherwise,} \end{cases}$$

we deduce under (5.25) that

$$\begin{aligned}
 & \mathbb{E} \left[\sum_{j=1}^n (Z_{(j+1)} - Z_{(j)})^\kappa \right] \\
 &= \mathbb{E} \left[\sum_{j=1}^n (Z_j^* - Z_j)^\kappa \mathbb{1}(\delta - \pi/2 \leq Z_j, Z_j^* \leq \pi/2 - \delta) \right] \\
 &\leq n \mathbb{E} \left[\mathbb{E}[(Z_1^* - Z_1)^\kappa \mid Z_1] \mathbb{1}(\delta - \pi/2 \leq Z_1 \leq \pi/2 - \delta) \right] \\
 &= n \mathbb{E} \left[\int_0^{(\pi/2 - Z_1)^\kappa} \mathbb{P}(Z_1^* > Z_1 + t^{1/\kappa} \mid Z_1) dt \mathbb{1}(\delta - \pi/2 \leq Z_1 \leq \pi/2 - \delta) \right] \\
 &\leq n \mathbb{E} \left[\int_0^{(\pi/2 - Z_1)^\kappa} \mathbb{P}(Z_k \notin (Z_1, Z_1 + t^{1/\kappa}), \forall k \neq 1 \mid Z_1) dt \mathbb{1}(\delta - \pi/2 \leq Z_1 \leq \pi/2 - \delta) \right] \\
 &= n \int_{\delta - \pi/2}^{\pi/2 - \delta} \int_0^{(\pi/2 - z)^\kappa} \left(1 - \int_z^{z+t^{1/\kappa}} f_Z(x) dx \right)^{n-1} dt f_Z(z) dz \\
 &= n \int_{\delta - \pi/2}^{\pi/2 - \delta} \int_0^{\pi/2 - z} \left(1 - \int_z^{z+s} f_Z(x) dx \right)^{n-1} \kappa s^{\kappa-1} ds f_Z(z) dz \\
 &\leq C_Z n \int_{\delta - \pi/2}^{\pi/2 - \delta} \int_0^\infty \exp(-(n-1)c_Z ||z| - \pi/2|^\beta s) \kappa s^{\kappa-1} ds ||z| - \pi/2|^\beta dz \\
 &= \kappa C_Z c_Z^{-\kappa} n(n-1)^{-\kappa} \int_{\delta - \pi/2}^{\pi/2 - \delta} ||z| - \pi/2|^{-\beta(\kappa-1)} dz \int_0^\infty \exp(-s) s^{\kappa-1} ds \\
 &= 2\kappa C_Z c_Z^{-\kappa} \Gamma(\kappa) n(n-1)^{-\kappa} \int_\delta^{\pi/2} u^{-\beta(\kappa-1)} du,
 \end{aligned}$$

that is, (5.26). Moreover, we write $Z_j^* := Z_j + \pi/2$ and $L_n^*(\delta) := L_n(\delta) + \pi/2$ so that

$$\begin{aligned}
 \mathbb{E}[L_n^*(\delta)^2] &= 2 \int_0^\pi z \mathbb{P}(L_n^*(\delta) > z) dz \\
 &\leq 2 \int_0^\delta z dz + 2 \int_\delta^\pi z \mathbb{P}(\min\{Z_j^* : Z_j \geq \delta\} \geq z) dz \\
 &= \delta^2 + 2 \int_\delta^\pi z \left(1 - \int_\delta^z f_Z(x - \pi/2) dx \right)^n dz \\
 &\leq \delta^2 + 2 \int_\delta^\pi z \exp\left(-n \int_\delta^z f_Z(x - \pi/2) dx\right) dz \\
 &\leq \delta^2 + 2 \int_\delta^{\pi/2} z \exp(-nc_Z(z - \delta)\delta^\beta) dz + \pi^2 \cdot \exp(-nc_Z(\pi/2 - \delta)\delta^\beta)
 \end{aligned}$$

$$\begin{aligned}
 &= \delta^2 + 2 \int_0^{\pi/2-\delta} (z + \delta) \exp(-nz\delta^\beta) \, dz + \pi^2 \cdot \exp(-nc_Z(\pi/2 - \delta)\delta^\beta) \\
 &\leq \delta^2 + 2 \frac{\delta}{c_Z n \delta^\beta} + 2 \frac{1}{(c_Z n \delta^\beta)^2} + \pi^2 \cdot \exp(-nc_Z(\pi/2 - \delta)\delta^\beta) \\
 &\leq 2 \left(\delta + \frac{1}{c_Z n \delta^\beta} \right)^2 + \pi^2 \cdot \exp(-nc_Z(\pi/2 - \delta)\delta^\beta),
 \end{aligned}$$

as $\delta \downarrow 0$. The term $\mathbb{E}[(R_n(\delta) - \pi/2)^2]$ can be bounded analogously.

Concerning (5.28), we bound the probability that there is at most one observation in $[-\pi/2 + \delta, \pi/2 - \delta]$ for $\delta \leq \pi/4$ by

$$\begin{aligned}
 &\mathbb{P}(Z_{(j)} < -\pi/4 \text{ or } Z_{(j+1)} > \pi/4, j = 1, \dots, n - 1) \\
 &\leq n \mathbb{P}(Z_j \in [-\pi/2, -\pi/4) \cup (\pi/4, \pi/2], j = 2, \dots, n) \\
 &\leq n \left(1 - \int_{-\pi/4}^{\pi/4} f_Z(z) \, dz \right)^{n-1} \\
 &\leq n \exp(-c_Z(n - 1)(\pi/2)(\pi/4)^\beta),
 \end{aligned}$$

which implies the result. □

Acknowledgements

The authors are grateful to the editors and a referee for their thorough review and very helpful and constructive comments. H. Holzmann gratefully acknowledges financial support of the DFG, grant Ho 3260/5-1.

References

- [1] Arellano, M. and Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *Rev. Econ. Stud.* **79** 987–1020. MR2986389 <https://doi.org/10.1093/restud/rdr045>
- [2] Beran, R., Feuerverger, A. and Hall, P. (1996). On nonparametric estimation of intercept and slope distributions in random coefficient regression. *Ann. Statist.* **24** 2569–2592. MR1425969 <https://doi.org/10.1214/aos/1032181170>
- [3] Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Ann. Statist.* **20** 1970–1984. MR1193321 <https://doi.org/10.1214/aos/1176348898>
- [4] Beran, R. and Millar, P.W. (1994). Minimum distance estimation in random coefficient regression models. *Ann. Statist.* **22** 1976–1992. MR1329178 <https://doi.org/10.1214/aos/1176325767>
- [5] Breunig, C. and Hoderlein, S. (2018). Specification testing in random coefficient models. *Quant. Econ.* **9** 1371–1417. MR3891570 <https://doi.org/10.3982/QE757>

- [6] Dunker, F., Eckle, K., Proksch, K. and Schmidt-Hieber, J. (2019). Tests for qualitative features in the random coefficients model. *Electron. J. Stat.* **13** 2257–2306. MR3980958 <https://doi.org/10.1214/19-EJS1570>
- [7] Gaillac, C. and Gautier, E. (2019). Adaptive estimation in the linear random coefficients model when regressors have limited variation. Preprint. Available at [arXiv:1905.06584](https://arxiv.org/abs/1905.06584).
- [8] Gautier, E. and Hoderlein, S. (2011). A triangular treatment effect model with random coefficients in the selection equation. Preprint. Available at [arXiv:1109.0362](https://arxiv.org/abs/1109.0362).
- [9] Gautier, E. and Kitamura, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica* **81** 581–607. MR3043343 <https://doi.org/10.3982/ECTA8675>
- [10] Gautier, E. and Le Pennec, E. (2018). Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding. *Electron. J. Stat.* **12** 277–320. MR3763073 <https://doi.org/10.1214/17-EJS1383>
- [11] Hildreth, C. and Houck, J.P. (1968). Some estimators for a linear model with random coefficients. *J. Amer. Statist. Assoc.* **63** 584–595. MR0232502
- [12] Hoderlein, S., Holzmann, H. and Meister, A. (2017). The triangular model with random coefficients. *J. Econometrics* **201** 144–169. MR3704681 <https://doi.org/10.1016/j.jeconom.2017.05.018>
- [13] Hoderlein, S., Klemelä, J. and Mammen, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory* **26** 804–837. MR2646480 <https://doi.org/10.1017/S0266466609990119>
- [14] Jirak, M., Meister, A. and Reiss, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. MR3262474 <https://doi.org/10.1214/14-AOS1248>
- [15] Lepski, O.V. and Spokoiny, V.G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. MR1604408 <https://doi.org/10.1214/aos/1030741083>
- [16] Lepskiĭ, O.V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167 <https://doi.org/10.1137/1136085>
- [17] Lewbel, A. and Pendakur, K. (2017). Unobserved preference heterogeneity in demand using generalized random coefficients. *J. Polit. Econ.* **125** 1100–1148.
- [18] Masten, M.A. (2018). Random coefficients on endogenous variables in simultaneous equations models. *Rev. Econ. Stud.* **85** 1193–1250. MR3782604 <https://doi.org/10.1093/restud/rdx047>
- [19] Masten, M.A. and Torgovitsky, A. (2013). Instrumental variables estimation of a generalized correlated random coefficients model. Preprint. Available at [arXiv:1310.6643](https://arxiv.org/abs/1310.6643).
- [20] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics. Lecture Notes in Statistics* **193**. Berlin: Springer. MR2768576 <https://doi.org/10.1007/978-3-540-87557-4>
- [21] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. New York: Springer. MR2724359 <https://doi.org/10.1007/b13794>

Received March 2019 and revised January 2020