HAL
open science

Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding

Eric Gautier, Erwan Le Pennec

To cite this version:

Eric Gautier, Erwan Le Pennec. Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding. 2016. inria-00601274v2

HAL Id: inria-00601274

https://inria.hal.science/inria-00601274v2

Preprint submitted on 15 Sep 2016 (v2), last revised 28 Nov 2017 (v4)

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

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

ADAPTIVE ESTIMATION IN THE NONPARAMETRIC RANDOM COEFFICIENTS BINARY CHOICE MODEL BY NEEDLET THRESHOLDING

ERIC GAUTIER AND ERWAN LE PENNEC

Abstract. In the random coefficients binary choice model, a binary variable equals 1 iff an index $X^\top\beta$ is positive. The vectors X and β are independent and belong to the sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$. We prove lower bounds on the minimax risk for estimation of the density f_β over Besov bodies where the loss is a power of the $\mathrm{L}^p(\mathbb{S}^{d-1})$ norm for $1 \leq p \leq \infty$. We show that a hard thresholding estimator based on a needlet expansion with data-driven thresholds achieves these lower bounds up to logarithmic factors.

1. Introduction

Discrete choice models (see, *e.g.*, [18]) have applications in many areas ranging from planning of public transportation, economics of industrial organizations, evaluation of public policies, among others. We consider the binary choice model where individuals choose between two exclusive alternatives that we label $\{1, -1\}$, for example buying a good or not. In a random utility framework, an individual chooses the alternative that yields the highest utility. Assume that the utility that an individual i gets from alternative -1 has the form $u_{-1,i} = z_{-1,i}^\top \gamma_i + \epsilon_{-1,i}$ while the utility from alternative 1 has the form $u_{1,i} = z_{1,i}^\top \gamma_i + \epsilon_{1,i}$ where $z_{-1,i}$ (resp. $z_{1,i}$) is a vector of $d-1$ characteristics of alternative -1 (resp. 1) for individual i where $d \geq 2$, γ_i are preferences of individual i for the characteristics, and $\epsilon_{-1,i}$ and $\epsilon_{1,i}$ absorb both the usual error terms and constants. In this model, the preferences are allowed to vary across individuals, namely they are heterogeneous. The characteristics of the alternatives are indexed by the individuals, for example they can be characteristics of two goods that a consumer has to choose upon interacted with individual characteristics like age or distance. Alternatively it is common to assume that one observes variation of the product characteristics through multiple markets (*e.g.*, cell phone plans in different countries) in which case a third index

Date: This version: September 15, 2016, first version arXiv:1106.3503.
Keywords: Discrete choice models, random coefficients, inverse problems, minimax rate optimality, adaptation, needlets, data-driven thresholding.
AMS 2010 Subject Classifications: Primary 62P20 ; secondary 42C15, 62C20, 62G07, 62G08, 62G20.
Eric Gautier acknowledges financial support from the grants ERC POEMH and ANR-13-BSH1-0004.
The authors are grateful to Andrii Babii (TSE) for research assistance on the simulations.

for the market is introduced. Individual i strictly prefers 1 ($y_i = 1$) if and only if the net utility for 1 is positive, *i.e.*

$$u_{1,i} - u_{-1,i} = \epsilon_{1,i} - \epsilon_{-1,i} + (z_{1,i} - z_{-1,i})^\top \gamma_i > 0, \tag{1}$$

he is indifferent when $u_{1,i} - u_{-1,i} = 0$, and he strictly prefers -1 ($y_i = -1$) when $u_{1,i} - u_{-1,i} < 0$. Denote now $x_i = (1, (z_{1,i} - z_{-1,i})^\top)^\top / \left|(1, (z_{1,i} - z_{-1,i})^\top)^\top\right|$ and $\beta_i = (\epsilon_{1,i} - \epsilon_{-1,i}, \gamma_i^\top)^\top / \left|(\epsilon_{1,i} - \epsilon_{-1,i}, \gamma_i^\top)^\top\right|$ where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$. The denominator in the definition of β_i is non zero because it is assumed that $u_{1,i} - u_{-1,i} = 0$ never occurs since there are only two alternatives. We assume that the researcher observes $(y_i, x_i^\top)$ for individuals $i = 1, \ldots, n$ and that $(\beta_i^\top, x_i^\top)$ are i.i.d. draws from a continuous distribution. Denoting the population quantities as Y, Z_1, Z_{-1}, X, ϵ_1, ϵ_{-1}, γ and β, with X and β random vectors in the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$, we obtain the model

$$Y = \text{sign}\left(X^\top \beta\right) \tag{2}$$

where, for a real number a, sign(a) is 1 if a is positive, is -1 if a is negative, and it is 0 if $a = 0$. This model extends the Logit, Probit or Mixed-Logit models. Like in [3, 4, 8, 11] among others, we consider a nonparametric specification of the joint distribution of β.

We maintain the following restrictions on the joint distribution of $(\beta^\top, X^\top)$ and relax (A2.3) at the end of section 5.

Assumption 1. *(A1.1) X and β are independent,*
(A1.2) X and β have densities f_X and f_β with respect to the spherical measure σ.

Assumption 2. *(A2.1) $f_\beta(x) f_\beta(-x) = 0$ for a.e. x in $\mathbb{S}^{d-1}$,*
(A2.2) The support of X, denoted by $\text{supp}(f_X)$, is $H^+ = \{x \in \mathbb{S}^{d-1} : \ x_1 \geq 0\}$,
(A2.3) f_X is known and we have $A_X \stackrel{\text{def}}{=} \|f_X\|_{\mathrm{L}^\infty(H^+)} < \infty$ and $B_X \stackrel{\text{def}}{=} \|1/f_X\|_{\mathrm{L}^\infty(H^+)} < \infty$.

Under assumption 1, f_β is solution of the ill-posed inverse problem: for a.e. $x \in H^+$

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{S}^{d-1}} \text{sign}\left(x^\top y\right) f_\beta(y) d\sigma(y) \stackrel{\text{def}}{=} \mathcal{K} f_\beta(x). \tag{3}$$

The operator $\mathcal{K}$ in (3) is a convolution on $\mathbb{S}^{d-1}$.

The identification problem in this model stems from the fact that: (1) the distribution of the observed data only characterizes $\mathcal{K} f_\beta$ on $\text{supp}(f_X)$ which is a proper subset of $\mathbb{S}^{d-1}$ and (2) $\mathcal{K}$ has an infinite dimensional null space. The support of X can only be as large as H^+ because the first coordinate of X is positive (*i.e.* we allow the term $\epsilon_{1,i} - \epsilon_{-1,i}$ in (1)). Condition (A2.2) requires that the support of $Z_1 - Z_{-1}$ is $\mathbb{R}^d$. This is demanding but we do not restrict dependence in the vector $(\epsilon_1 - \epsilon_{-1}, \gamma^\top)$ or its tails. It is important to avoid restricting the dependence between the coordinates of $(\epsilon_1 - \epsilon_{-1}, \gamma^\top)$ since they can be functions of a deep heterogeneity parameter (*e.g.*, the type of a consumer). Tail assumptions on unobservables (see, *e.g.*, [4]) are problematic since they are impossible to

test. In [8, 11], it is assumed that there exists n (unknown) in $\mathbb{S}^{d-1}$ such that $\mathbb{P}(n^\top\beta > 0) = 1$. It implies that for some difference of the characteristics, or taking a limit of these, everyone chooses the same alternative. In contrast, (A2.1) does not restrict the support of β and does not imply "unselected samples".

Estimation of f_β in (3) is related to statistical deconvolution on $\mathbb{S}^{d-1}$ (see, *e.g.*, [10, 13, 16]). However, the left-hand side of (3) is not a density but a regression function where the regressors are random and supported on a subset of $\mathbb{S}^{d-1}$ and the convolution is not injective. [8] provides a straightforward to compute estimator, rates of convergence for the L^p-losses for $1 \leq p \leq \infty$, and confidence sets. This paper extends [8] by providing lower bounds on the minimax risk where the degree of integrability in the loss - specified by the statistician - can differ from the degree of integrability of the smoothness ellipsoid containing the unknown f_β, giving rise to sparse and dense regimes. We show that the estimator in [8] is a plug-in of a linear needlet estimator. Needlets have been successfully used in statistics in [2, 14, 15]. We show that working in high-dimension and replacing the linear estimator by a nonlinear estimator based on hard thresholding achieves the bounds, up to a logarithmic factor. We also contribute to the literature on estimation using needlets by providing data-driven thresholds (similar in spirit to [5] for density estimation using the Dantzig selector). Proofs are given in the appendix.

2. Preliminaries

We use the notation $x \wedge y$ and $x \vee y$ for the minimum and the maximum between x and y. We write $x \lesssim y$ when there exists c such that $x \leq cy$, $x \gtrsim y$ when there exists c such that $x \geq cy$, and $x \simeq y$ when $x \lesssim y$ and $x \gtrsim y$. We denote by $|A|$ and $\mathbf{1}_A$ the cardinal and indicator of the set A, by $\mathbb{N}$ the nonnegative integers, by $\mathbb{N}^*$ the positive integers, by a.e. almost every, and by a.s. almost surely. We denote for $1 \leq p \leq \infty$ by $\|\cdot\|_{\ell^p}$ the ℓ^p-norm of a vector, by $\|\cdot\|_p$ the usual norm on the space $\mathrm{L}^p(\mathbb{S}^{d-1})$ of p integrable real-valued functions with respect to the spherical measure σ. We write $\mathrm{L}^p_{\mathrm{odd}}(\mathbb{S}^{d-1})$ (resp. $\mathrm{L}^p_{\mathrm{even}}(\mathbb{S}^{d-1})$) the closure in $\mathrm{L}^p(\mathbb{S}^{d-1})$ of continuous functions on $\mathbb{S}^{d-1}$ which are odd (*i.e.* for every $x \in \mathbb{S}^{d-1}$, $f(-x) = -f(x)$) (resp. even). Every $f \in \mathrm{L}^p(\mathbb{S}^{d-1})$ can be uniquely decomposed as the sum of an odd and even function f^- and f^+ in $\mathrm{L}^p(\mathbb{S}^{d-1})$. The space $\mathrm{L}^2(\mathbb{S}^{d-1})$ is a Hilbert space with the scalar product $\langle\ ,\ \rangle$ derived from the norm, there f^- and f^+ are orthogonal. $\mathcal{D}$ is the set of densities and $\nu(d) \stackrel{\mathrm{def}}{=} d/2$ is the degree of ill-posedness of the inverse problem.

2.1. Harmonic analysis.

The basic element is the orthogonal decomposition $\mathrm{L}^2(\mathbb{S}^{d-1}) = \bigoplus_{k\in\mathbb{N}} H^{k,d}$ where $H^{k,d}$ are the eigenspaces of the Laplacian Δ on $\mathbb{S}^{d-1}$, corresponding to the eigenvalues $-\zeta_{k,d}$ where $\zeta_{k,d} \stackrel{\mathrm{def}}{=} k(k+d-2)$, of dimension $L(k,d) \stackrel{\mathrm{def}}{=} (2k+d-2)(k+d-2)!/(k!(d-2)!(k+d-2))$. The space $H^{k,d}$ is spanned by an orthonormal basis $(h_{k,l})_{l=1}^{L(k,d)}$ and $H^{0,d}$ by 1. We also have $\mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1}) = \bigoplus_{p\in\mathbb{N}} H^{2p+1,d}$ and $\mathrm{L}^2_{\mathrm{even}}(\mathbb{S}^{d-1}) = \bigoplus_{p\in\mathbb{N}} H^{2p,d}$.

The projector $L_{k,d}$ onto $H^{k,d}$ is the operator with kernel

$$L_{k,d}(x,y) = \sum_{l=1}^{L(k,d)} h_{k,l}(x)h_{k,l}(y) = \frac{L(k,d)}{\sigma(\mathbb{S}^{d-1})P_k^{\mu(d)}(1)} P_k^{\mu(d)}\left(x^\top y\right), \tag{4}$$

where $\mu(d) = (d-1)/2$, the surface of $\mathbb{S}^{d-1}$ is $\sigma(\mathbb{S}^{d-1}) = 2\pi^{d/2}/\Gamma(d/2)$, and C_k^μ are the Gegenbauer polynomials. The Gegenbauer polynomials, defined for $\mu > -1/2$, are orthogonal in the space of square integrable functions on $[-1,1]$ with measure $(1-t^2)^{\mu-1/2}dt$. We have $P_0^\mu(t) = 1$, $P_1^\mu(t) = 2\mu t$ for $\mu \neq 0$, $P_1^0(t) = 2t$, and for every $k \in \mathbb{N}$

$$(k+2)P_{k+2}^\mu(t) = 2(\mu+k+1)tP_{k+1}^\mu(t) - (2\mu+k)P_k^\mu(t). \tag{5}$$

Clearly, for $f \in \mathrm{L}^2(\mathbb{S}^{d-1})$, we have $f = \sum_{k=0}^\infty L_{k,d}f$ and, due to (4),

$$\forall x \in \mathbb{S}^{d-1},\ \|L_{k,d}(x,\cdot)\|_2^2 = \sum_{l=1}^{L(k,d)} |h_{k,l}(x)|^2 = \frac{L(k,d)}{\sigma(\mathbb{S}^{d-1})}. \tag{6}$$

Powers $(-\Delta)^s f$ for $s \in \mathbb{R}$ and f in a Banach space E_1 are defined in a Banach space E_2 when $L_{k,d}f$ is defined in E_2 and $(-\Delta)^s f \stackrel{\text{def}}{=} \sum_{k=0}^\infty \zeta_{k,d}^s L_{k,d}f$ converges in E_2. The best approximation in $\mathrm{L}^r(\mathbb{S}^{d-1})$ of a function f by harmonics of degree less or equal to m is

$$E_m(f)_r = \inf_{P \in \bigoplus_{k=0}^m H^{k,d}} \|f - P\|_r .$$

Definition 3. *For $s > 0$ and $1 \le r \le \infty$, f belongs to the Sobolev space $\mathrm{W}_r^s(\mathbb{S}^{d-1})$ if*

$$\|f\|_{r,s} = \|f\|_r + \left\|(-\Delta)^{s/2} f\right\|_r < \infty.$$

We denote by $\mathrm{W}_{r\ \mathrm{odd}}^s(\mathbb{S}^{d-1})$ the restriction of $\mathrm{W}_r^s(\mathbb{S}^{d-1})$ to odd functions.

Definition 4. *For $s > 0$, $1 \le r \le \infty$, and $0 < q \le \infty$, f belongs to the Besov space $B_{r,q}^s(\mathbb{S}^{d-1})$ if*

$$\|f\|_{B_{r,q}^s}^A = \|f\|_r + \left\|\left(2^{js}E_{2^j}(f)_r\right)_{j\in\mathbb{N}}\right\|_{\ell^q} < \infty.$$

2.2. **The operator.**

Proposition 5. *The operator $\mathcal{K}$ satisfies the following properties:*
(P1.1) For every $f \in \mathrm{L}^1(\mathbb{S}^{d-1})$, $\mathcal{K}f = \mathcal{K}(f^-)$,
(P1.2) If $\mathcal{K}f = \mathcal{K}g$ with $f, g \in \mathrm{L}_{\mathrm{odd}}^1(\mathbb{S}^{d-1})$ then $g = f$,
(P1.3) For every $1 \le r \le \infty$,

$$\mathrm{W}_{r\ \mathrm{odd}}^{\nu(d)-|1/r-1/2|(d-2)}(\mathbb{S}^{d-1}) \subseteq \mathcal{K}(\mathrm{L}_{\mathrm{odd}}^r(\mathbb{S}^{d-1})) \subseteq \mathrm{W}_{r\ \mathrm{odd}}^{\nu(d)+|1/r-1/2|(d-2)}(\mathbb{S}^{d-1})$$

where the exponents $\nu(d) \pm |1/r - 1/2|(d-2)$ cannot be improved,

(P1.4) For every $1 \leq r \leq \infty$, there exists $B(d,r)$ such that

$$\forall K \in \mathbb{N},\ \forall P \in \bigoplus_{\substack{k=0 \\ k \text{ odd}}}^{K} H^{k,d},\ \|\mathcal{K}^{-1}P\|_r \leq B(d,r)K^{\nu(d)}\|P\|_r. \tag{7}$$

Moreover, $\mathcal{K}$ is a self-adjoint and compact operator on $\mathrm{L}^2(\mathbb{S}^{d-1})$ with null space $\mathrm{L}^2_{\text{even}}(\mathbb{S}^{d-1})$, nonzero eigenvalues $(\lambda_{2p+1,d})_{p\in\mathbb{N}}$ corresponding to the eigenspaces $H^{2p+1,d}$ for $p \in \mathbb{N}$

$$\lambda_{1,d} = \frac{2|\mathbb{S}^{d-2}|}{d-1},\ \forall p \in \mathbb{N}^*\ \lambda_{2p+1,d} = \frac{2(-1)^p|\mathbb{S}^{d-2}|1\cdot 3\cdots(2p-1)}{(d-1)(d+1)\cdots(d+2p-1)}.$$

For every $d \in \mathbb{N}\setminus\{1\}$, for every $p \in \mathbb{N}$, there exists $c_\lambda(d), C_\lambda(d) > 0$ such that

$$c_\lambda^{-1}(d)p^{-\nu(d)} \leq |\lambda_{2p+1,d}| \leq C_\lambda(d)p^{-\nu(d)}. \tag{8}$$

$\mathcal{K}$ is a homeomorphism between $\mathrm{L}^2_{\text{odd}}(\mathbb{S}^{d-1})$ and $\mathrm{W}^{\nu(d)}_{2\ \text{odd}}(\mathbb{S}^{d-1})$.

Proposition 5 implies that every $R \in \mathrm{W}^{\nu(d)}_{2\ \text{odd}}(\mathbb{S}^{d-1})$ has a unique inverse given by

$$\mathcal{K}^{-1}(R) = \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} L_{k,d}(R) = \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} \sum_{l=1}^{L(k,d)} \langle R, h_{k,l}\rangle h_{k,l}. \tag{9}$$

2.3. **Needlets.** Smoothed projection operators (see [8]) have good approximation properties in all $\mathrm{L}^p(\mathbb{S}^{d-1})$ spaces and are uniformly bounded from $\mathrm{L}^p(\mathbb{S}^{d-1})$ to $\mathrm{L}^p(\mathbb{S}^{d-1})$. One such operator, the delayed means, is the integral operator with kernel

$$K^{a,J}(x,y) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} a\left(\frac{k}{2^J}\right) L_{k,d}(x,y) \tag{10}$$

where J is an integer, a is a C^∞ and decreasing function on $[0,\infty)$ supported on $[0,2]$ such that, for every $0 \leq t \leq 2$, $0 \leq a(t) \leq 1$ and, for every $0 \leq t \leq 1$, $a(t) = 1$. The delayed means operator exhibits nearly exponential localization (see theorem 2.2 in [21]) and is a building block for the construction of needlets.

Define b such that $b^2(t) = a(t) - a(2t)$ for $t \geq 0$. It is nonzero only when $1/2 \leq t \leq 2$, satisfies $b^2(t) + b^2(2t) = 1$ for $1/2 \leq t \leq 1$ and thus for every $t \geq 1$, $\sum_{j=0}^{\infty} b^2\left(\frac{t}{2^j}\right) = 1$, also $b^2(t) = a(t)$ for $1 \leq t \leq 2$. Take a such that b is bounded away from 0 on $3/5 \leq t \leq 5/3$.

The second ingredient for the construction of needlets is a quadrature formula (corollary 2.9 of [21]) with positive weights $(\omega(j,\xi)^2)_{\xi\in\Xi_j}$ and nodes $\xi \in \Xi_j$ which integrates functions in $\bigoplus_{k=0}^{2^j} H^{k,d}$ and satisfy, for a constant C_Ξ which depends on d,

$$\begin{aligned} \forall j \in \mathbb{N},\ \forall \xi \in \Xi_j,\quad & C_\Xi^{-1}2^{j(d-1)} \leq |\Xi_j| \leq C_\Xi 2^{j(d-1)} \\ & C_\Xi^{-1}2^{-j(d-1)/2} \leq \omega(j,\xi) \leq C_\Xi 2^{-j(d-1)/2}. \end{aligned}$$

Needlets are defined as

$$\psi_{j,\xi}(x) \stackrel{\text{def}}{=} \omega(j,\xi) \sum_{k=0}^{\infty} b\left(\frac{k}{2^{j-1}}\right) L_{k,d}(\xi,x) \quad \text{if } j \in \mathbb{N},\ \xi \in \Xi_j, \tag{11}$$

$$\psi_{0,\xi}(x) \stackrel{\text{def}}{=} L_{0,d}(\xi,x). \tag{12}$$

For $j=0$, $\psi_{0,\xi}(x)$ is constant and Ξ_0 is a singleton.

The L^p-norms of the needlets satisfy, for a constant C_p that can depend on d,

$$\forall j \in \mathbb{N},\ \forall \xi \in \Xi_j,\ C_p^{-1} 2^{j(d-1)(1/2-1/p)} \leq \|\psi_{j,\xi}\|_p \leq C_p 2^{j(d-1)(1/2-1/p)}. \tag{13}$$

If $f \in \mathrm{L}^p(\mathbb{S}^{d-1})$ for $1 \leq p \leq \infty$, then $f = \sum_{j=0}^{\infty} \sum_{\xi \in \Xi_j} \langle f, \psi_{j,\xi} \rangle \psi_{j,\xi}$. The needlets form a tight frame, this means that for $f \in \mathrm{L}^2(\mathbb{S}^{d-1})$

$$\|f\|_2^2 = \sum_{j=0}^{\infty} \sum_{\xi \in \Xi_j} |\langle f, \psi_{j,\xi} \rangle|^2 .$$

Needlets do not form a basis and there is redundancy. Lemma 6 (see [2]) relates $\mathrm{L}^p(\mathbb{S}^{d-1})$ norms at level j to ℓ^p norms of needlet coefficients. Constants may depend on d.

Lemma 6. *(i) For every $1 \leq p \leq \infty$, there exists a constant C'_p such that for every $j \in \mathbb{N}$ and $(\beta_\xi)_{\xi \in \Xi_j} \in \mathbb{R}^{\Xi_j}$*

$$\left\| \sum_{\xi \in \Xi_j} \beta_\xi \psi_{j,\xi} \right\|_p \leq C'_p 2^{j(d-1)(1/2-1/p)} \left\| (\beta_\xi)_{\xi \in \Xi_j} \right\|_{\ell^p}, \tag{14}$$

(ii) There exists a constant c_A such that for every $j \in \mathbb{N}$ there exists $A_j \subset \Xi_j$ with $|A_j| \geq c_A 2^{j(d-1)}$ such that for every $1 \leq p \leq \infty$, there exists a constant $c_{p,A}$ such that for every $j \in \mathbb{N}$ and $(\beta_\xi)_{\xi \in A_j} \in \mathbb{R}^{A_j}$

$$\left\| \sum_{\xi \in A_j} \beta_\xi \psi_{j,\xi} \right\|_p \geq c_{p,A} 2^{j(d-1)(1/2-1/p)} \left\| (\beta_\xi)_{\xi \in A_j} \right\|_{\ell^p}, \tag{15}$$

(iii) For every $1 \leq p \leq \infty$, there exists a constant C''_p such that for every $j \in \mathbb{N}$

$$\left(\sum_{\xi \in \Xi_j} |\langle f, \psi_{j,\xi} \rangle|^p \right)^{1/p} 2^{j(d-1)(1/2-1/p)} \leq C''_p \|f\|_p. \tag{16}$$

Needlets are such that (see [21]), for all function a in the definition of the smoothed projection operators, the norm $\|\cdot\|^A_{B^s_{r,q}}$ defining the Besov spaces is equivalent to

$$\|f\|_{B^s_{r,q}} = \left\| \left(2^{j(s+(d-1)(1/2-1/r))} \left\| (\langle f, \psi_{j,\xi} \rangle)_{\xi \in \Xi_j} \right\|_{\ell^r} \right)_{j \in \mathbb{N}} \right\|_{\ell^q}.$$

The ball of radius M for this norm is denoted by $B^s_{r,q}(M)$.

Recall the following consequence of the proof of the continuous embeddings in [2].

Lemma 7. *(i) If $p \leq r \leq \infty$, then we have $B^s_{r,q}(M) \subseteq B^s_{p,q}(C_\Xi^{1/p-1/r}M)$,*
(ii) If $s > (d-1)(1/r-1/p)$ and $r \leq p \leq \infty$, then we have $B^s_{r,q}(M) \subseteq B^{s-(d-1)(1/r-1/p)}_{p,q}(M)$,
(iii) If $f \in B^s_{r,q}(M)$ and $(\beta_{j,\xi})_{\xi\in\Xi_j, j\in\mathbb{N}}$ are its needlet coefficients, then there exists $(D_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ such that $\|(D_j)_{j\in\mathbb{N}}\|_{\ell^q} \leq M$ and

$$\forall z \geq 1,\ \forall j \in \mathbb{N},\ \sum_{\xi\in\Xi_j} |\beta_{j,\xi}|^z \leq C_\Xi^{1-(z\wedge r)/r} D_j^z 2^{-jz(s+(d-1)(1/2-1/(z\wedge r)))}. \tag{17}$$

Finally recall that, when $f \in B^s_{r,q}$ with $s > (d-1)/r$, then f is continuous.

3. Identification of f_β

Let us present the arguments for the identification of f_β. Proposition 5 (P1.1) implies that $\mathcal{K}f_\beta = \mathcal{K}f_\beta^-$ is odd. Thus under (A2.2) we can define the odd function R as

$$R(x) = \begin{cases} \mathbb{E}[Y|X=x] & \text{for a.e. } x \in H^+ \\ -\mathbb{E}[Y|X=-x] & \text{for a.e. } x \in -H^+ \end{cases} \tag{18}$$

and we have, for a.e. $x \in \mathbb{S}^{d-1}$, $R(x) = \mathcal{K}f_\beta^-(x)$. Uniqueness of f_β^- follows from (P1.2). Using that, for a.e. $x \in \mathbb{S}^{d-1}$, $f_\beta(x) \geq 0$, we conclude that $f_\beta^-(x) = (f_\beta(x) - f_\beta(-x))/2$, and condition (A2.1), that, for a.e. $x \in \mathbb{S}^{d-1}$, we have

$$f_\beta(x) = 2f_\beta^-(x)\mathbf{1}_{f_\beta^-(x)>0}. \tag{19}$$

In this paper we normalize the vectors of random coefficients and covariates to have unit norm. Indeed, since only the sign of the net utility (1) matters for choosing between 1 and -1 and the index is linear, a scale normalization of $(\epsilon_1 - \epsilon_{-1}, \gamma^\top)$ is in order. Let us compare with the normalization in [7]. It is based on the following assumption, which is stronger than the condition in [11] that the support of β is a subset of some (unknown) hemisphere, which itself is stronger than (A2.1).

(H): a.s. there exists $j \in \{1, \ldots, d\}$, the coordinate γ_j of γ has a sign (excluding 0).

assumption (H) is likely to hold when Z_{1j} and Z_{-1j} are cost factors, since consumers dislike an increase in cost. If (H) holds we can identify for which index j γ_j has a sign since it amounts to finding for which coordinate z_j of z $z_j \to \mathbb{E}[Y|Z_1 - Z_{-1} = z]$ is (globally) monotone. We can identify the sign of the coefficient by assessing whether the function is increasing (positive) or decreasing (negative). If $\gamma_j > 0$ then we normalize the vector of coefficients by dividing by γ_j. If $\gamma_j < 0$ we change the sign of $Z_{1j} - Z_{-1j}$ to make it positive. A potential issue with this normalization is that if β_j can take small values then estimators could differ in finite samples depending on which coefficient is used for normalization. Also, monotonicity in one regressor of the conditional mean function implies a type of weak monotonicity (in the sense used to identify treatment

effects, see, *e.g.*, [7]) at the individual level as we now explain. Assuming that $\gamma_j > 0$, $z_{1i} - z_{-1i} = z$ for all $i = 1, \ldots, n$, and that we change z_j to $z_j' > z_j$ while leaving unchanged $(\epsilon_{1i} - \epsilon_{-1i}, \gamma_i^\top)$ (the characteristics of the individuals) and the other components of z, then some people do not change their decision and some chose alternative 1 while originally they had chosen alternative -1, but no one changes from alternative 1 to alternative -1. Monotonicity of the conditional mean function implies monotonicity for every individual. This is sometimes not a realistic model of individuals making choices. Clearly (A2.1) allows both individuals to switch from 1 to -1 and individuals to switch from -1 to 1 after similar changes in z (or x). On the other hand, if (H) holds then (A2.2) can be relaxed and we can consider an index which is nonlinear in X (*c.f.* [7]).

4. Lower bounds

We take $1 \le p, r \le \infty$, $0 \le q \le \infty$, $z \ge 1$, and $s > 0$, and consider the minimax risk

$$\mathcal{R}_n^* \stackrel{\text{def}}{=} \inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f_\beta} - f_\beta \right\|_p^z$$

where the infimum is over all estimators based on the i.i.d. sample of size n. The degree of integrability r in the smoothness class $B^s_{r,q}(M)$ is allowed to differ from the degree of integrability p in the loss function. We distinguish two zones for s, r, q, d, and p:
(1) the *dense zone* where $s \ge p\left(\nu(d) + (d-1)/2\right)(1/r - 1/p)$ with the restriction $q \le r$ if $s = p\left(\nu(d) + (d-1)/2\right)(1/r - 1/p)$, where the rate involves

$$\mu_{\text{dense}}(d, p, r, s) \stackrel{\text{def}}{=} s/(s + \nu(d) + (d-1)/2),$$

(2) the *sparse zone* where $(d-1)/r < s < p\left(\nu(d) + (d-1)/2\right)(1/r - 1/p)$, where the rate involves

$$\mu_{\text{sparse}}(d, p, r, s) \stackrel{\text{def}}{=} (s - (d-1)(1/r - 1/p))/(s + \nu(d) - (d-1)(1/r - 1/2)).$$

Theorem 8. *(i) In the dense zone we have*

$$\mathcal{R}_n^* \ge c_{\text{dense}} \left(\frac{1}{\sqrt{nA_X}} \right)^{\mu_{\text{dense}}(d,p,r,s)z}, \tag{20}$$

(ii) In the sparse zone we have

$$\mathcal{R}_n^* \ge c_{\text{sparse}} \left(\sqrt{\frac{\ln(nA_X)}{nA_X}} \right)^{\mu_{\text{sparse}}(d,p,r,s)z} \tag{21}$$

where the constants c_{dense} and c_{sparse} depend on d, M, p, r, s and z.

The values of μ_{dense} and μ_{sparse} depend on d through the dimension of $\mathbb{S}^{d-1}$. This is the usual curse of dimensionality in nonparametric regression or density estimation. They also depend on d through the degree of ill-posedness $\nu(d) = d/2$ of the inverse problem.

5. Adaptive estimation by needlet thresholding

Consider the estimator $\widehat{f_\beta} = 2\widehat{f_\beta^-}\mathbf{1}_{\widehat{f_\beta^-}>0}$ where $\widehat{f_\beta^-}$ is an estimator of f_β^-.

5.1. **Smoothed projections and linear needlet estimators.** A smoothed projection estimator of f_β^- with kernel (10), window a, and $J \in \mathbb{N}$, is given for $x \in \mathbb{S}^{d-1}$ by

$$\widehat{f_\beta^-}^{a,J}(x) = \sum_{k \text{ odd}} \frac{a\left(\frac{k}{2^J}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(x)$$

with the unbiased estimator of $L_{k,d}R(x)$ (see lemma 10): $\widehat{L_{k,d}R}(x) = 0$ if k is even, else

$$\widehat{L_{k,d}R}(x) = \frac{2}{n}\sum_{i=1}^{n} \frac{y_i L_{k,d}(x_i, x)}{f_X(x_i)}.$$

Alternatively we can estimate f_β^- using the needlet frame with smoothing window a. The coefficients $\beta^a_{j,\xi} = \langle f_\beta^-, \psi_{j,\xi}\rangle$ are such that

$$\begin{aligned}
\beta^a_{j,\xi} &= \omega(j,\xi) \sum_{k \text{ odd}} b\left(\frac{k}{2^{j-1}}\right) \langle f_\beta^-, L_{k,d}(\xi,\cdot)\rangle \\
&= \omega(j,\xi) \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \langle L_{k,d}R, L_{k,d}(\xi,\cdot)\rangle \\
&= \omega(j,\xi) \sum_{\substack{k \text{ odd} \\ 2^{j-2}<k<2^j}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} L_{k,d}R(\xi).
\end{aligned}$$

Using that $a\left(\frac{k}{2^j}\right) = 1$ for $k = 0, \dots, 2^j$ and denoting by $f_\beta^{-a,J} = \mathbb{E}\left[\widehat{f_\beta^-}^{a,J}\right]$, we obtain that, for $1 \le j \le J$, $\beta^a_{j,\xi} = \left\langle f_\beta^{-a,J}, \psi_{j,\xi}\right\rangle$, which can be estimated without bias by

$$\widehat{\beta}^a_{j,\xi} = \omega(j,\xi) \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(\xi) \overset{(\triangle_1)}{=} \left\langle \widehat{f_\beta^-}^{a,J}, \psi_{j,\xi}\right\rangle.$$

Moreover, for $x \in \mathbb{S}^{d-1}$,

$$\widehat{\beta}^a_{j,\xi}\psi_{j,\xi}(x) = \omega(j,\xi)^2 \left(\sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(\xi)\right)\left(\sum_k b\left(\frac{k}{2^{j-1}}\right) L_{k,d}(\xi, x)\right)$$

belongs to $\bigoplus_{k=0}^{2^j} H^{k,d}$, thus by the quadrature formula

$$\sum_{\xi \in \Xi_j} \widehat{\beta}^a_{j,\xi}\psi_{j,\xi}(x) = \sum_{k \text{ odd}} \frac{b^2\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(x).$$

This yields $\sum_{j=0}^{J}\sum_{\xi\in\Xi_j}\widehat{\beta}^a_{j,\xi}\psi_{j,\xi}=\widehat{f^-_\beta}^{a,J-1}$, indeed

$$\sum_{j=0}^{J}\sum_{\xi\in\Xi_j}\widehat{\beta}^a_{j,\xi}\psi_{j,\xi}=\sum_{j=1}^{J}\sum_{\xi\in\Xi_j}\widehat{\beta}^a_{j,\xi}\psi_{j,\xi}\quad\text{(due to }(\triangle_1)\text{ and because }\widehat{f^-_\beta}^{a,J}\text{ is odd)}$$

$$\overset{(\triangle_2)}{=}\sum_{\substack{1\le k<2^{J-1}\\ k\text{ odd}}}\frac{1}{\lambda_{k,d}}\widehat{L_{k,d}R}+\sum_{\substack{2^{J-1}\le k\le 2^J\\ k\text{ odd}}}\frac{b^2\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}}\widehat{L_{k,d}R}$$

$$\overset{(\triangle_3)}{=}\sum_{\substack{1\le k<2^{J-1}\\ k\text{ odd}}}\frac{1}{\lambda_{k,d}}\widehat{L_{k,d}R}+\sum_{\substack{2^{J-1}\le k\le 2^J\\ k\text{ odd}}}\frac{a\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}}\widehat{L_{k,d}R}$$

where $(\triangle_2)$ uses that for $1/2\le t\le 1$, $b^2(t)+b^2(2t)=1$, while $(\triangle_3)$ that $b^2(t)=a(t)$ for $1\le t\le 2$. Thus, the smoothed projection and needlet estimators coincide.

5.2. **Nonlinear estimator with data-driven thresholds.** Consider instead, for $\gamma\ge 1$ and $\rho_{T_{j,\xi,\gamma}}(x)=x\mathbf{1}_{|x|>T_{j,\xi,\gamma}}$, the nonlinear estimator of f^-_β:

$$\widehat{f^-_\beta}^{a,\rho}=\sum_{j=0}^{J}\sum_{\xi\in\Xi_j}\rho_{T_{j,\xi,\gamma}}\left(\widehat{\beta}^a_{j,\xi}\right)\psi_{j,\xi}.$$

It is classical that the optimal choice of J for linear estimators depends on the parameters of the smoothness ellipsoid. In contrast, using a thresholded estimator allows to take J large and independent of the parameters. Thresholding induces additional bias compared to linear estimators which allows to reduce the variance incurred by taking J large.

The level of thresholding should depend on the size of the coefficients relative to their variance. This variance is proportional to $1/\sqrt{n}$ so that the level of the threshold do not have to depend on the smoothness of the unknown function. Instead of using a conservative upper bound on their variance, as is usually the case in estimation using wavelets, we use data-driven levels of thresholding. These provide better estimators in small samples. Lemma 14 gives a theoretical guarantee that the performance is almost as good as that of an oracle which would known the variance of the estimators of the coefficients. The data-driven thresholding rule uses that $\widehat{\beta}^a_{j,\xi}=\frac{1}{n}\sum_{i=1}^{n}G_{j,\xi}(x_i,y_i)$ with

$$G_{j,\xi}(x_i,y_i)\overset{\text{def}}{=}\frac{2}{n}\sum_{i=1}^{n}\omega(j,\xi)\frac{y_i}{f_X(x_i)}\sum_{k\text{ odd}}\frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}}L_{k,d}(x_i,\xi).\tag{22}$$

Define the estimator of the variance by

$$\hat{\sigma}_{j,\xi}\overset{\text{def}}{=}\sqrt{\frac{1}{n(n-1)}\sum_{i=2}^{n}\sum_{k=1}^{i-1}\left(G_{j,\xi}(x_i,y_i)-G_{j,\xi}(x_k,y_k)\right)^2},\tag{23}$$

$t_n = \sqrt{\log n/n}$, and the data-driven thresholds

$$T_{j,\xi,\gamma} \stackrel{\text{def}}{=} 2\sqrt{2\gamma} t_n \widehat{\sigma}_{j,\xi} + \frac{28}{3} M_{j,\xi} \frac{\gamma \log n}{n-1}$$

where $M_{j,\xi}$ is an upper bound on the sup-norm over $H^+ \times \{\pm 1\}$ of $G_{j,\xi}(x,y) - \mathbb{E}\left[G_{j,\xi}(X,Y)\right] = G_{j,\xi}(x,y) - \beta^a_{j,\xi}$ (*e.g.*, $2\|G_{j,\xi}\|_\infty$). For example, using (13) and proposition 5, we get

$$2\|G_{j,\xi}\|_\infty \leq 2 \left\|\mathcal{K}^{-1}\left(\psi^-_{j,\xi}\right)\right\|_\infty B_X \leq 2C_\infty B(d,\infty) 2^{j(\nu(d)+(d-1)/2)} B_X \stackrel{\text{def}}{=} M_j. \tag{24}$$

The second term in $T_{j,\xi,\gamma}$ controls the error in estimating the threshold.

Theorem 9. *For J such that $2^{J(\nu(d)+(d-1)/2)} B_X^{1/2} \simeq t_n^{-1}$, $M > 0$, and $s > (d-1)/r$,*

(i) If $z > 1$ and $\gamma > z/2 + 1$, we have

$$\sup_{f_\beta \in B^s_{r,q}(M) \cap \mathcal{D}} \mathbb{E} \left\|\widehat{f_\beta}^{a,\rho} - f_\beta\right\|_\infty^z \leq \tilde{c}(d,\infty,r,s,\gamma)(\log n)^{z-1} M^r \left(B_X t_n\right)^{\mu_{\text{sparse}}(d,\infty,r,s)z} \tag{25}$$

(ii) If $p < \infty$ and $\gamma > p/2$, we have

$$\sup_{f_\beta \in B^s_{r,q}(M) \cap \mathcal{D}} \mathbb{E} \left\|\widehat{f_\beta}^{a,\rho} - f_\beta\right\|_p^p \leq \tilde{c}(d,p,r,s,\gamma)(\log n)^{p-1} M^\varpi \left(B_X t_n\right)^{\mu(d,p,r,s)p} \tag{26}$$

where $\mu(d,p,r,s) = \mu_{\text{dense}}(d,p,r,s)$, $\varpi = r$ in the dense zone while $\mu(d,p,r,s) = \mu_{\text{sparse}}(d,p,r,s)$, $\varpi > p\frac{\nu(d)+(d-1)(1/2-1/(p\vee z))}{s+\nu(d)-(d-1)(1/r-1/2)}$ is arbitrary in the sparse zone, and $\tilde{c}(d,p,r,s,\gamma)$ is a constant which depends on d,p,r,s, and γ.

Consider now the case where f_X is unknown but we have at our disposal an estimator $\widehat{f_X}$ obtained with an auxiliary data set. Theorem 9 holds on the event

$$\left\|\frac{f_X}{\widehat{f_X}} - 1\right\|_{\mathrm{L}^\pi(H^+)} \leq c\left(B_X t_n\right)^{\frac{s+\nu(d-1)(1/\pi-1/r)_+}{\nu+(d-1)/2}}$$

if the expectations are conditional on the auxiliary data, $B_X = \|1/\widehat{f_X}\|_{\mathrm{L}^\infty(H^+)}$, f_X is replaced by $\widehat{f_X}$ in (22) and (23), $\pi \geq 1$ and $c \in \mathbb{R}$ are arbitrary. The probability of this event is close to 1 for large sets of f_X where f_X is small on a small subset of the sphere (possibly unbounded from below) and is smooth enough. A more general result is proved in the first version of this paper on arXiv (see also [8]).

6. Simulation study

We study the performance of the estimator when $d = 3$, $n = 3000, 5000, 10000$, and X is uniform on H^+. We use of the Von Mises-Fisher distribution $\mathrm{vMF}(\mu,\kappa)$ with density

$$f(\beta;\mu,\kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp\left(\kappa \mu^\top \beta\right)$$

with respect to σ. We take $\beta = (\tilde{\beta}_1, \tilde{\beta}_2, |\tilde{\beta}_3|)$ in the cases:

- $\tilde{\beta}$ follows a $\mathrm{vMF}(\mu, \kappa)$ distribution where $\mu = (0\ 0\ 1)^\top$ and $\kappa = 10$.
- $\tilde{\beta}$ follows a mixture $\lambda\mathrm{vMF}(\mu_1, \kappa) + (1-\lambda)\mathrm{vMF}(\mu_2, \kappa)$, where $\mu_1 = (2^{-1/2}\ 0\ 2^{-1/2})^\top$, $\mu_2 = (-2^{-1/2}\ 0\ 2^{-1/2})^\top$, $\kappa = 10$ and $\lambda = 0.3$.

We use the cubature defined in spherical coordinates as a product of the Gauss-Legendre quadrature with m nodes and trapezoid rule with $2m$ subdivisions (see [1]). The resulting cubature has $2m^2$ nodes and integrates exactly all polynomials on the sphere up to degree $2m-1$. We take the same function a as in [2]. Figure 1 displays the distribution of estimates based on a Monte-Carlo experiments with 1000 replications and $n = 10000$. We plot the the Lambert equal-area projection on the disk which is defined (see [19])

$$(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^\top \mapsto 2\sin\left(\frac{\theta}{2}\right)(\cos\phi, \sin\phi)^\top.$$

Table 1 displays the risk, approximated using Monte-Carlo and 1000 replications.

	Unimodal			Mixture		
Risk \ n	3000	5000	10000	3000	5000	10000
$\mathbb{E}\left[\left\|\widehat{f}_\beta - f_\beta\right\|_1\right]$	0.521	0.429	0.330	0.592	0.478	0.364
$\mathbb{E}\left[\left\|\widehat{f}_\beta - f_\beta\right\|_2^2\right]$	0.091	0.061	0.035	0.101	0.064	0.036
$\mathbb{E}\left[\left\|\widehat{f}_\beta - f_\beta\right\|_4^4\right]$	0.006	0.002	0.001	0.006	0.002	0.001
$\mathbb{E}\left[\left\|\widehat{f}_\beta - f_\beta\right\|_\infty\right]$	0.394	0.326	0.266	0.387	0.305	0.234

TABLE 1. Risk.

7. Appendix

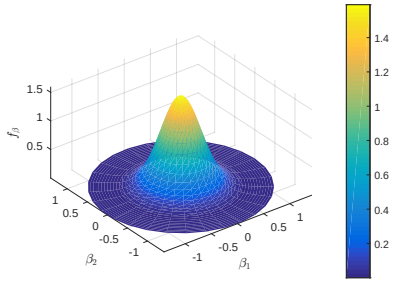
7.1. A preliminary lemma.

Lemma 10. *The following equality holds for every $g \in \mathrm{L}^2(\mathbb{S}^{d-1})$.*

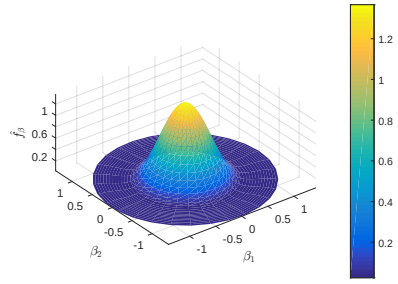
$$\langle R, g\rangle = 2\mathbb{E}\left[\frac{Y g^-(X)}{f_X(X)}\right].$$

Proof. The result is based on the following

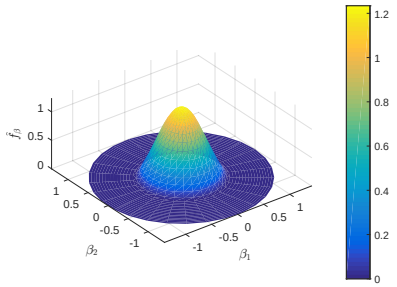
$$\begin{aligned}\langle R, g\rangle &= \langle R, g^-\rangle \quad \text{(because } R \text{ is odd)}\\ &= 2\int_{H^+} \frac{R(x)g^-(x)}{f_X(x)} f_X(x)d\sigma(x)\\ &= 2\mathbb{E}\left[\frac{R(X)g^-(X)}{f_X(X)}\right]\end{aligned}$$



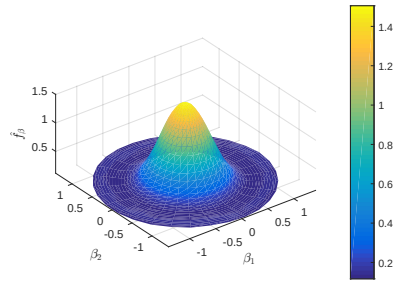
(A) True density



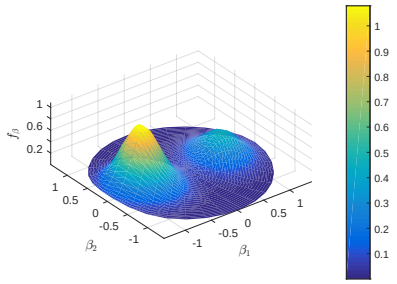
(B) Mean of estimates



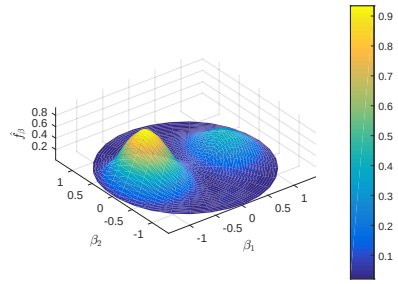
(C) 5% quantile of estimates



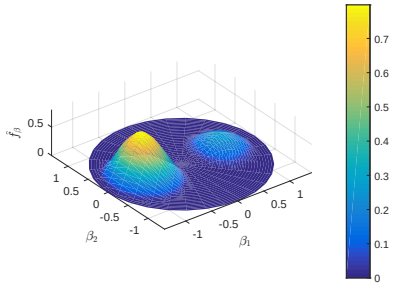
(D) 95% quantile of estimates



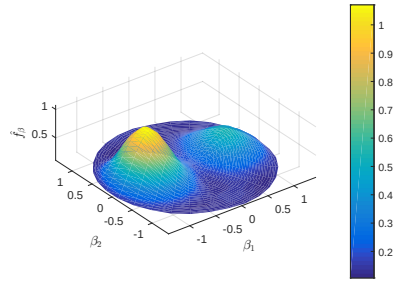
(E) True density



(F) Mean of estimates



(G) 5% quantile of estimates



(H) 95% quantile of estimates

FIGURE 1. True density and distribution of the estimates.

$$= 2\mathbb{E}\left[\frac{\mathbb{E}[Y|X]g^-(X)}{f_X(X)}\right] \qquad \square$$

7.2. **Proof of proposition 5.** The operator $\mathcal{K}$ is related to the Hemispherical transform (see [8, 22]) defined for $f \in \mathrm{L}^1(\mathbb{S}^{d-1})$ and a.e. $x \in \mathbb{S}^{d-1}$ by

$$\mathcal{H}(f)(x) \stackrel{\text{def}}{=} \int_{\mathbb{S}^{d-1}} \mathbf{1}_{x^\top y>0} f(y) d\sigma(y),$$

through

$$Kf = 2\mathcal{H}(f) - \int_{\mathbb{S}^{d-1}} f(y) d\sigma(y).$$

(P1.1) is a consequence of the fact that $y \to x^\top y \in \mathrm{L}^\infty_{\text{odd}}(\mathbb{S}^{d-1})$. (P1.2) follows from theorem 2 (ii), and (P1.3) follows from theorem C in [22]. The second part of the proposition together with (P1.4) are consequences of the properties of $\mathcal{H}$ detailed in [8]. The inequalities (8) correspond to lemma A.2. Note however that there is a typo in the proof and we should read $1.3\ldots(2p-1) \asymp p^{-1/2} 2.4\ldots(2p)$ but the result still holds.

7.3. **Proof of theorem 8.** Start by noting that for every $j \in \mathbb{N}$ and $\xi \in \Xi_j$,

$$\int_{\mathbb{S}^{d-1}} \psi_{j,\xi}(x) dx = \omega(j,\xi) b(0) = \omega(j,\xi)(a(0) - a(0)) = 0.$$

This implies that the functions f_m that we introduce below integrate to 1.

7.3.1. *Proof of the lower bound in the dense zone.* Consider the family $(P_m)_{m=0}^{\mathcal{M}}$, where $\mathcal{M} \in \mathbb{N}^*$, of distributions of an i.i.d. sample of (Y, X) of size n when $f_\beta = f_m$ and the density of X is f_X. These probabilities are absolutely continuous with respect to the product of $\delta_1 + \delta_{-1}$ where δ_y denotes the Dirac mass at y and σ. Take $j \in \mathbb{N}$, $f_0 = 1/\sigma(\mathbb{S}^{d-1})$, and consider the set A_j from lemma 6 (ii). By the Varshamov-Guilbert bound (lemma 2.9 in [23]) there exists $\Omega \subseteq \{0,1\}^{A_j}$ containing $(0,\ldots,0)$ such that $|\Omega| = 2^{|A_j|/8}$ and $\forall(\omega_1,\omega_2) \in \Omega^2$, $\|\omega_1 - \omega_2\|_{\ell^1} \geq |A_j|/8$. Enumerate the elements of Ω from 0 (corresponding to the zero vector) to $\mathcal{M} \stackrel{\text{def}}{=} |\Omega| - 1$ and define

$$f_m \stackrel{\text{def}}{=} f_0 + \gamma \sum_{\xi \in A_j} \omega_\xi \psi_{j,\xi}$$

when $(\omega_\xi)_{\xi \in A_j}$ is the mth element of Ω and $\gamma = cC_\Xi^{-1/r} M 2^{-j(s+(d-1)/2)}$ for $0 < c < 1$ such that all f_m are nonnegative. We now use the following result (see theorem 2.5 in [23]).

Lemma 11. *If for $0 < \alpha < 1/8$ we have:*

(i) $f_m \in B^s_{r,q}(M) \cap \mathcal{D}$ *for* $m = 0,\ldots,\mathcal{M}$,
(ii) $\forall\ 0 \leq m < l \leq \mathcal{M},\ \|f_m - f_l\|_p \geq 2h > 0$,
(iii) $\frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} K(P_m, P_0) \leq \alpha \ln(\mathcal{M})$,

then for every $z \geq 1$

$$\inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)\cap\mathcal{D}} \mathbb{E}\left\|\widehat{f_\beta} - f_\beta\right\|_p^z \geq h^z \frac{\sqrt{\mathcal{M}}}{1+\sqrt{\mathcal{M}}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\ln(\mathcal{M})}}\right). \tag{27}$$

Start by checking (i) in lemma 11. It is enough to show that $f_m \in B^s_{r,q}(M)$. Indeed, for $r \geq 1$ and $\omega \in \Omega$, we have $\left\|(\omega_\xi)_{\xi\in A_j}\right\|_{\ell^r} \leq \left\|(\omega_\xi)_{\xi\in A_j}\right\|_{\ell^1}^{1/r} \leq C_\Xi^{1/r} 2^{j(d-1)/r}$, we obtain

$$\gamma 2^{j(s+(d-1)(1/2-1/r))}\left\|(\omega_\xi)_{\xi\in A_j}\right\|_{\ell^r} \leq \gamma C_\Xi^{1/r} 2^{j(s+(d-1)/2)} \leq M.$$

Lemma 6 (ii) now yields that for every $1 \leq p \leq \infty$ and $0 \leq m < l \leq \mathcal{M}$

$$\|f_m - f_l\|_p \geq \gamma c_{p,A} 2^{j(d-1)(1/2-1/p)}\left(\frac{c_A}{8}2^{j(d-1)}\right)^{1/p} = 2h.$$

Thus (ii) in lemma 11 with $h = c_{p,A}\left(\frac{c_A}{8}\right)^{1/p} c C_\Xi^{-1/r} M 2^{-js-1}$.
By independence, the Kullback-Leibler divergence between P_m and P_0 is given by

$$K(P_m, P_0) = n\mathbb{E}\left[\mathcal{H}(f_m)(X)\ln\left(\frac{\mathcal{H}(f_m)(X)}{\mathcal{H}(f_0)(X)}\right) + (1-\mathcal{H}(f_m)(X))\ln\left(\frac{1-\mathcal{H}(f_m)(X)}{1-\mathcal{H}(f_0)(X)}\right)\right].$$

Using that, for $x > 0$, $\ln(x) \leq x - 1$, we obtain

$$K(P_m, P_0) \leq n\mathbb{E}\left[\frac{\mathcal{H}(f_m - f_0)(X)^2}{\mathcal{H}(f_0)(X)\,(1-\mathcal{H}(f_0)(X))}\right]$$

and thus

$$\begin{aligned} K(P_m, P_0) &\leq 4nA_X \left\|\mathcal{H}(f_m - f_0)\right\|_2^2 \\ &\leq 4nA_X \lambda_{2^j+1,d}^2 \left\|f_m - f_0\right\|_2^2 \end{aligned}$$

where the last display comes from the fact that $f_m - f_0 \in \bigoplus_{2^j+1\leq k\leq 2^{j+2}-1} H^{k,d}$. From (8) we get

$$K(P_m, P_0) \leq 4C_\lambda(d)^2 nA_X 2^{-2j\nu(d)}\left\|f_m - f_0\right\|_2^2$$

which yields using lemma 6 (i)

$$\begin{aligned} K(P_m, P_0) &\leq \left(2C_\lambda(d)C_2'\gamma\right)^2 nA_X 2^{-2j\nu(d)}\left\|(\omega_\xi)_{\xi\in A_j}\right\|_{\ell^2}^2 \\ &\leq \left(2C_\lambda(d)C_2'\gamma\right)^2 nA_X 2^{-2j\nu(d)}\left\|(\omega_\xi)_{\xi\in A_j}\right\|_{\ell^1} \\ &\leq \left(2C_\lambda(d)C_2'\gamma\right)^2 C_\Xi nA_X 2^{j(d-1-2\nu(d))} \\ &\leq \left(2C_\lambda(d)C_2'cM\right)^2 C_\Xi^{1-2/r} nA_X 2^{-2j(s+\nu(d))}. \end{aligned}$$

Condition (iii) of lemma 11 is satisfied once

$$\frac{2^5\left(C_\lambda(d)C_2'cM\right)^2}{\ln(2)}C_\Xi^{-2/r}nA_X2^{-2j(s+\nu(d)+(d-1)/2)}\leq\alpha<\frac{1}{8}. \tag{28}$$

For $\alpha<1/8$, the lower bound (27) yields that

$$\begin{aligned}\inf_{\widehat{f_\beta}}\sup_{f_\beta\in B^s_{r,q}(M)}\mathbb{E}\left\|\widehat{f_\beta}-f_\beta\right\|_p^z &\geq\left(c_{p,A}\left(\frac{c_A}{8}\right)^{1/p}cC_\Xi^{-1/r}M2^{-js-1}\right)^z\left(\frac{3}{4}-\frac{1}{2\sqrt{\ln(\mathcal{M})}}\right)\\ &\geq\frac{1}{2}\left(c_{p,A}\left(\frac{c_A}{8}\right)^{1/p}cC_\Xi^{-1/r}\frac{M}{2}\right)^z2^{-jsz}\end{aligned}$$

where the inequality leading to the second display holds when $\ln(\mathcal{M})\geq 4$, for example for $j(d-1)\geq\ln(5/c_A\ln(2))/\ln(2)$. Now (28) is satisfied for

$$j\geq j_0\overset{\text{def}}{=}1+\frac{\ln\left(2^8\left(C_\lambda(d)C_2'cM\right)^2C_\Xi^{-2/r}nA_X/\ln(2)\right)}{2\ln(2)(s+\nu(d)+(d-1)/2)}$$

which implies the lower bound

$$\begin{aligned}&\inf_{\widehat{f_\beta}}\sup_{f_\beta\in B^s_{r,q}(M)}\mathbb{E}\left\|\widehat{f_\beta}-f_\beta\right\|_p^z\\ &\geq\frac{1}{2}\left(c_{p,A}\left(\frac{c_A}{8}\right)^{1/p}cC_\Xi^{-1/r}M2^{-s-1}\right)^z\left(\frac{2^8\left(C_\lambda(d)C_2'cM\right)^2C_\Xi^{-2/r}nA_X}{\ln(2)}\right)^{-\mu_{\text{dense}}(d,p,r,s)z/2}.\end{aligned}$$

7.3.2. *Proof of the lower bound in the sparse zone.* In this proof we consider asymptotic orders for simplicity. The various constants can be obtained like in Section 7.3.1. Consider the hypotheses

$$f_m=\frac{1}{\sigma(\mathbb{S}^{d-1})}+\gamma\psi_{j,\xi_m}$$

where $\xi_m\in A_j$ and $|\gamma|\lesssim 2^{-j(d-1)/2}$ to ensure the functions are positive. The constant is adjusted so that for one of the f_m that we denote f_0, $\forall x\in H^+$, $\left|\mathcal{H}(f_0^-)(x)\right|\leq c_b$ with $c_b\in(0,\frac{1}{2})$. The function f_m also integrate to 1. We denote by $\mathcal{M}$ the cardinality of A_j ($\mathcal{M}\simeq 2^{j(d-1)}$), P_m the distributions of an i.i.d. sample of (Y,X) of size n when $f_\beta=f_m$ and for a given f_X, and $\Lambda(P_m,P_0)$ the likelihood ratio. Recall that $K(P_m,P_0)=\mathbb{E}_{P_m}\left[\Lambda(P_m,P_0)\right]$. We make use of the following lemma from [17].

Lemma 12. *If for $\pi_0>0$ and $\mathcal{M}\in\mathbb{N}^*$ the following three condition hold*

(i) $f_m\in B^s_{r,q}(M)\cap\mathcal{D}$ *for* $m=1,\dots,\mathcal{M}$,
(ii) $\forall m\neq l,\ \|f_m-f_l\|_p\geq 2h>0$,

(iii) $\forall m = 1, \dots, \mathcal{M}$, $\Lambda(P_0, P_m) = \exp(z_n^m - v_n^m)$, *where* z_n^m *are random variables and* v_n^m *constants such that* $\mathbb{P}(z_n^m > 0) \geq \pi_0$ *and* $\exp\left(\sup_{m=1,\dots,\mathcal{M}} v_n^m\right) \leq \mathcal{M}$,

then

$$\forall z \geq 1,\ \inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)\cap\mathcal{D}} \mathbb{E}\left\|\widehat{f_\beta} - f_\beta\right\|_p^z \geq \frac{h^{-z}\pi_0}{2}.$$

Item (i) is satisfied when $|\gamma| \leq M2^{-j(s-(d-1)(1/r-1/2)}$. This is more restrictive than the condition to ensure positivity because we assume that $s \geq (d-1)/r$. Thus, now we take $\gamma = 2cM2^{-j(s-(d-1)(1/r-1/2)}$ for a well chosen constant c.
The constant h in (ii) is obtained as follows, if $m \neq m'$,

$$\begin{aligned}\|f_m - f_{m'}\|_p &= \gamma\|\psi_{j,\xi_m} - \psi_{j,\xi_{m'}}\|_p \\ &\geq \gamma c_{p,A} 2^{j(d-1)(1/2-1/p)} \\ &\geq 2cM2^{-j(s-(d-1)(1/r-1/p))}.\end{aligned}$$

Let us now consider item (iii), we obtain

$$\begin{aligned}P_m\left(\log\left(\Lambda(P_0,P_m)\right) \geq -j(d-1)\log 2\right) &\geq 1 - P_m\left(\left|\log\left(\Lambda(P_0,P_m)\right)\right| \geq j(d-1)\log 2\right) \\ &\geq 1 - \frac{\mathbb{E}_{P_m}\left[\left|\log\left(\Lambda(P_0,P_m)\right)\right|\right]}{j(d-1)\log 2}.\end{aligned}$$

Thus, condition (iii) is satisfied when

$$\mathbb{E}_{P_m}\left[\left|\log\left(\Lambda(P_0,P_m)\right)\right|\right] \leq \alpha j(d-1)\log 2$$

for $\alpha \in (0,1)$. The same computations as in the beginning of Section 5.1 yield that we need to impose $n2^{-2j\nu(d)}\gamma^2 \lesssim j$, thus

$$A_X n 2^{-2j(s+\nu(d)-(d-1)(1/r-1/2))} \lesssim j.$$

The desired rate is obtained by taking

$$2^j \simeq \left(\frac{nA_X}{\log\left(nA_X\right)}\right)^{\frac{1}{2(s+\nu(d)-(d-1)(1/r-1/2))}}.$$

7.4. **Comparison between Besov ellipsoids of a function and its odd part.**

Lemma 13. *For* $0 < s, q \leq \infty$ *and* $1 \leq r \leq \infty$, *there exists a constant* c_{eq} *that can depend on* d *such that, for every* $f \in B^s_{r,q}$, $\|f^-\|_{B^s_{r,q}} \leq c_{\text{eq}}\|f\|_{B^s_{r,q}}$

Proof. In definition 4 every $f \in B^s_{r,q}(\mathbb{S}^{d-1})$ has same norm as $x \to f(-x)$, thus by the triangle inequality $\|f^-\|^A_{B^s_{r,q}} \leq \|f\|^A_{B^s_{r,q}}$. We conclude by equivalence of the norms $\square$

7.5. **A general inequality.** We make use of the constants $c_{1,z}$ and $c_{2,z}$ such that

$$\int_{\mathbb{R}^+} z\tau^{z-1}e^{-\beta\tau}d\tau \leq c_{1,z}\beta^{-z} \tag{29}$$

$$\int_{\mathbb{R}^+} z\tau^{z-1}e^{-\alpha\tau^2}d\tau \leq c_{2,z}\alpha^{-z/2}. \tag{30}$$

Lemma 14. *For every $\tau,\ \gamma,\ z>1$ and*

$$T^{s,++}_{j,\xi,\gamma} \geq 3\sqrt{2\gamma}t_n\widehat{\sigma}_{j,\xi} + 26M_{j,\xi}\frac{\gamma\log n}{n-1} \stackrel{\text{def}}{=} T^{s,+}_{j,\xi,\gamma},$$

the two following inequalities hold:
when $p=\infty$,

$$\begin{aligned}
&\frac{1}{2^{z-1}}\mathbb{E}\left[\left\|\widehat{f_\beta}^{a,\rho} - f_\beta\right\|_\infty^z\right] \\
&\leq \left\|f_\beta^{-a,J} - f_\beta^-\right\|_\infty^z + (J+1)^{z-1}C'^z_\infty\Bigg\{ \\
&\quad a_{n,\infty,z,J}\sum_{j=0}^{J}2^{j(d-1)z/2}\left(\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{|\beta^a_{j,\xi}|\leq T^{s,++}_{j,\xi,\gamma}} + \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{|\beta^a_{j,\xi}|>T^{s,++}_{j,\xi,\gamma}}\right]\right) \\
&\quad + \frac{4C_\Xi}{n^\gamma}\sum_{j=0}^{J}2^{j(d-1)(z/2+1)}\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \\
&\quad + \left(\frac{C_\Xi 4}{n^\gamma}\right)^{1-1/\tau}\left(\frac{1}{\sqrt{n}}B_X^{1/2}2^{Jz(\nu(d)+(d-1)/2)}\right)^z 2^{J(d-1)(1-1/\tau)}b_{n,\infty,z,J,\tau}\Bigg\}
\end{aligned}$$

where

$$\begin{aligned}
a_{n,\infty,z,J} &= 1 + \left(\frac{2}{\sqrt{\gamma\log n}}\right)^z\left(2+\left(\log\left(C_\Xi 2^{J(d-1)}c_{2,z}\right)\right)^{z/2}\right) \\
&\quad + \left(\frac{4}{\gamma\log n}\right)^z\left(2+\left(\log\left(C_\Xi 2^{J(d-1)}c_{1,z}\right)\right)^{z}\right) \\
b_{n,\infty,z,J,\tau} &= \frac{\left(2\sqrt{2}C_2B(d,2)\right)^z\left(2^{1/\tau}+\left(\log\left(C_\Xi 2^{J(d-1)}c_{2,z\tau}\right)\right)^{z/2}\right)}{1-2^{-(z\nu(d)+(d-1)(z/2+1-1/\tau))}} \\
&\quad + \frac{(8C_\infty B(d,\infty)/3)^z\left(2^{1/\tau}+\left(\log\left(C_\Xi 2^{J(d-1)}c_{1,z}\right)\right)^{z}\right)}{1-2^{-(z\nu(d)+(d-1)(z+1-1/\tau))}}\left(\frac{2^{J(d-1)}}{n}B_X\right)^{z/2};
\end{aligned}$$

while when $1\leq p\leq\infty$,

$$\frac{1}{2^{z-1}}\mathbb{E}\left[\left\|\widehat{f_\beta}^{a,\rho} - f_\beta\right\|_p^z\right]$$

$$\leq \left\| f_\beta^{-a,J} - f_\beta^- \right\|_p^z + (J+1)^{z-1} C_p'^z C_\Xi^{z/(p\wedge z)-1} \Big\{$$

$$a_{n,p,z,J} \sum_{j=0}^{J} 2^{j(d-1)(z/2-z/(p\vee z))} \sum_{\xi\in\Xi_j} \left(\left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{s,++}} \right)$$

$$+ \frac{4}{n^\gamma} \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))} \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z$$

$$+ \frac{2^{2-1/\tau}}{n^{\gamma(1-\frac{1}{\tau})}} C_\Xi \left(\frac{1}{\sqrt{n}} B_X^{1/2} 2^{J(\nu(d)+(d-1)/2)} \right)^z 2^{J(d-1)(1-z/(p\vee z))} b_{n,p,z,J,\tau} \Big\}$$

where

$$a_{n,p,z,J} = 1 + 2\left(\left(\frac{\sqrt{2} c_{2,z}^{1/z}}{\sqrt{\gamma \log n}} \right)^z + \left(\frac{2 c_{1,z}^{1/z}}{\gamma \log n} \right)^z \right)$$

$$b_{n,p,z,J,\tau} = \frac{\left(2 c_{2,z\tau}^{1/(z\tau)} C_2 B(d,2)\right)^z}{1 - 2^{-(z\nu(d)+(d-1)(z/2+1-z/(p\vee z)))}} + \frac{\left(\frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d,\infty)\right)^z}{1 - 2^{-(z\nu(d)+(d-1)(z+1-z/(p\vee z)))}} \left(\frac{2^{J(d-1)}}{n} B_X \right)^{z/2}.$$

The inequalities of lemma 14 are similar to oracle inequalities, for a well chosen J depending on n (see theorem 9), where the oracle estimates $\beta_{j,\xi}^a$ if and only if the error made by estimating this coefficient is smaller than the one made by discarding it. This *oracle* strategy would lead to a quantity of the form

$$\left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| \leq \left(\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]\right)^{1/z}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > \left(\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]\right)^{1/z}}.$$

Proving such an oracle inequality would require to lower bound $\left(\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]\right)^{1/z}$. In the inequalities of lemma 14 the ideal quantity $\left(\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]\right)^{1/z}$ is replaced by $T_{j,\xi,\gamma}^{s,++}$, we call this term a *quasi-oracle* term. The remaining terms can be made as small as we want by taking γ large enough. The last term corresponds to the approximation error. Upper bounds of these types, uniform on Besov ellipsoids, yield an approximation error which can be expressed in terms of the regularity of the Besov class and is uniformly small for J large enough and allows to treat the bias/variance trade-off in the quasi-oracle term uniformly over the ellipsoid.

7.6. **Proof of lemma 14.**

7.6.1. *Preliminaries.* Recall from the proof of theorem 4.1 in [8] that for every $1 \leq p \leq \infty$

$$\left\| \widehat{f_\beta}^{a,\rho} - f_\beta \right\|_p \leq 2 \left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^- \right\|_p$$

and that, for $1 \leq z < \infty$, we have

$$\left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^- \right\|_p^z \leq 2^{z-1} \left(\left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J} \right\|_p^z + \left\| f_\beta^{-a,J} - f_\beta^- \right\|_p^z \right). \tag{31}$$

The first term corresponds to the error in the high dimensional space while the second term corresponds to the approximation error. Let us start by studying the first term. Lemma 6 (i) yields

$$\begin{aligned} \left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J} \right\|_p^z &\leq (J+1)^{z-1} \sum_{j=0}^{J} \left\| \sum_{\xi \in \Xi_j} \left(\rho_{T_{j,\xi,\gamma}} \left(\widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right) \psi_{j,\xi} \right\|_p^z \\ &\leq (J+1)^{z-1} \sum_{j=0}^{J} C_p'^z 2^{j(d-1)z(1/2-1/p)} \left\| \rho_{T_{j,\xi,\gamma}} \left(\widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right\|_p^z. \end{aligned}$$

Thus for $p = \infty$ we have

$$\left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J} \right\|_p^z \leq (J+1)^{z-1} \sum_{j=0}^{J} C_\infty'^z 2^{j(d-1)z/2} \sup_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left(\widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z$$

while for $p < \infty$ we have

$$\left\| \widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J} \right\|_p^z \leq (J+1)^{z-1} C_p'^z C_\Xi^{z/(p \wedge z)-1} \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p \vee z))} \sum_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left(\widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z$$

the last inequality is obtained using that, when $p \geq z$, we have

$$\left(\sum_{\xi \in \Xi_j} |b_\xi|^p \right)^{z/p} \leq \sum_{\xi \in \Xi_j} |b_\xi|^z,$$

and by the Hölder inequality, when $p \leq z$, we have

$$\left(\sum_{\xi \in \Xi_j} |b_\xi|^p \right)^{z/p} \leq C_\Xi^{z/p-1} \sum_{\xi \in \Xi_j} |b_\xi|^z.$$

7.6.2. *Coefficientwise analysis.* For the simplicity of the notations we sometimes drop the dependence on γ in the sets of indices.
We first consider the term

$$\delta_{j,\xi,z} \stackrel{\text{def}}{=} \left| \rho_{T_{j,\xi,\gamma}} \left(\widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z.$$

By construction we have

$$\begin{aligned} \delta_{j,\xi,z} &= \left| \beta_{j,\xi}^a \right|^z \mathbf{1}_{\left| \widehat{\beta}_{j,\xi}^a \right| \leq T_{j,\xi,\gamma}} + \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{\left| \widehat{\beta}_{j,\xi}^a \right| > T_{j,\xi,\gamma}} \\ &= \max \left(\left| \beta_{j,\xi}^a \right|^z \mathbf{1}_{\left| \widehat{\beta}_{j,\xi}^a \right| \leq T_{j,\xi,\gamma}}, \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{\left| \widehat{\beta}_{j,\xi}^a \right| > T_{j,\xi,\gamma}} \right). \end{aligned}$$

We introduce two "phantom" random thresholds $T^b_{j,\xi,\gamma} = T_{j,\xi,\gamma} - \Delta_{j,\xi,\gamma}$ and $T^s_{j,\xi,\gamma} = T_{j,\xi,\gamma} + \Delta_{j,\xi,\gamma}$ for some $\Delta_{j,\xi,\gamma}$ to be defined later. They are used to define "big" and "small" original needlets coefficients. We also use $T^{b,-}_{j,\xi,\gamma}$ for a deterministic lower bound on $T^b_{j,\xi,\gamma}$, $T^{s,+}_{j,\xi,\gamma}$ and $\Delta^+_{j,\xi,\gamma}$ for deterministic upper bounds on $T^s_{j,\xi,\gamma}$ and $\Delta_{j,\xi,\gamma}$. These bounds will hold with high probability. We obtain almost surely

$$\begin{aligned}
\delta_{j,\xi,z} = \max\Bigg(& \left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}\right|\le T_{j,\xi,\gamma}}\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}\right|\le T_{j,\xi,\gamma}}\mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^s_{j,\xi,\gamma}}\right), \\
& \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}\right|> T_{j,\xi,\gamma}}\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}\right|> T_{j,\xi,\gamma}}\mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^b_{j,\xi,\gamma}}\right)\Bigg) \\
\le \max\Bigg(& \left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right), \\
& \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^b_{j,\xi,\gamma}}\right)\Bigg) \\
\le \max\Bigg(& \left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}, \mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right), \\
& \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^{b,-}_{j,\xi,\gamma}}, \mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}\right)\Bigg).
\end{aligned}$$

Sorting the terms according to the number of random terms we obtain

$$\begin{aligned}
\delta_{j,\xi,z} \le \max\Bigg(& \left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}, \left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right), \\
& \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^{b,-}_{j,\xi,\gamma}}, \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}\right)\Bigg).
\end{aligned}$$

7.6.3. *Scalewise analysis.* Defining

$$\begin{aligned}
M_{j,z} &\stackrel{\text{def}}{=} \sup_{\xi\in\Xi_j}\left|\rho_{T_{j,\xi,\gamma}}\left(\widehat{\beta}^a_{j,\xi}\right) - \beta^a_{j,\xi}\right|^z = \sup_{\xi\in\Xi_j}\delta_{j,\xi,z} \\
S_{j,z} &\stackrel{\text{def}}{=} \sum_{\xi\in\Xi_j}\left|\rho_{T_{j,\xi,\gamma}}\left(\widehat{\beta}^a_{j,\xi}\right) - \beta^a_{j,\xi}\right|^z = \sum_{\xi\in\Xi_j}\delta_{j,\xi,z},
\end{aligned}$$

we obtain

$$\begin{aligned}
M_{j,z} \le \max\Bigg(& \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}, \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right), \\
& \sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^{b,-}_{j,\xi,\gamma}}, \sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\Bigg) \\
\stackrel{\text{def}}{=} & \max(M^{S0}_{j,z}, M^{S1}_{j,z}, M^{B1}_{j,z}, M^{B2}_{j,z})
\end{aligned}$$

$$
\begin{aligned}
&\leq M_{j,z}^{S0} + M_{j,z}^{S1} + M_{j,z}^{B1} + M_{j,z}^{B2} \\
S_{j,z} &\leq \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| \leq T_{j,\xi,\gamma}^{s,+}} + \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^{s}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right) \\
&\quad + \sum_{\xi\in\Xi_j} \left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}} + \sum_{\xi\in\Xi_j} \left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^{b}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right) \\
&\stackrel{\text{def}}{=} S_{j,z}^{S0} + S_{j,z}^{S1} + S_{j,z}^{B1} + S_{j,z}^{B2}.
\end{aligned}
$$

We bound the expectations of the random terms as follows

$$
\begin{aligned}
\mathbb{E}\left[M_{j,z}^{S1}\right] &\leq \sup_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \mathbb{E}\left[\sup_{\xi\in\Xi_j} \max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^{s}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right)\right] \\
&\leq \sup_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \left(\mathbb{P}\left(\bigcup_{\xi\in\Xi_j} \left\{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^{s}\right\}\right) + \mathbb{P}\left(\bigcup_{\xi\in\Xi_j} \left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\}\right)\right) \\
\mathbb{E}\left[M_{j,z}^{B1}\right] &\leq \mathbb{E}\left[\sup_{\xi\in\Xi_j} \left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}}\right] \\
\mathbb{E}\left[M_{j,z}^{B2}\right] &\leq \mathbb{E}\left[\sup_{\xi\in\Xi_j} \left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^{z\tau}\right]^{1/\tau} \mathbb{E}\left[\sup_{\xi\in\Xi_j} \max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^{b}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right)\right]^{1-1/\tau} \\
&\leq \mathbb{E}\left[\sup_{\xi\in\Xi_j} \left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^{z\tau}\right]^{1/\tau} \left(\mathbb{P}\left(\bigcup_{\xi\in\Xi_j} \left\{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^{b}\right\}\right) + \mathbb{P}\left(\bigcup_{\xi\in\Xi_j} \left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\}\right)\right)^{1-1/\tau} \\
\mathbb{E}\left[S_{j,z}^{S1}\right] &= \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \mathbb{E}\left[\max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^{s}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right)\right] \\
&\leq \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^z \left(\mathbb{P}\left\{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^{s}\right\} + \mathbb{P}\left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\}\right) \\
\mathbb{E}\left[S_{j,z}^{B1}\right] &= \sum_{\xi\in\Xi_j} \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}} \\
\mathbb{E}\left[S_{j,z}^{B2}\right] &= \sum_{\xi\in\Xi_j} \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \max\left(\mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^{b}}, \mathbf{1}_{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}}\right)\right] \\
&\leq \sum_{\xi\in\Xi_j} \left(\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^{z\tau}\right]\right)^{1/\tau} \left(\mathbb{P}\left\{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^{b}\right\} + \mathbb{P}\left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\}\right)^{1-1/\tau}
\end{aligned}
$$

The constant $\tau > 1$ in the Hölder inequality will be specified later.

7.6.4. *Bernstein inequality and the term* $\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z$. Let denote variance of $G_{j,\xi}(X,Y)$

$$\sigma^2_{j,\xi} \stackrel{\text{def}}{=} \mathbb{E}\left[\left(G_{j,\xi}(X,Y) - \beta^a_{j,\xi}\right)^2\right].$$

Lemma 15. *For every* $c_\sigma, c_M > 0$, *we have*

$$\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}}\right)^z\right] \le 2\left(c_{2,z}\left(\frac{2}{\sqrt{n}}\frac{1}{c_\sigma + c_M\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z + c_{1,z}\left(\frac{4}{3n}\frac{1}{c_\sigma\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M}\right)^z\right)$$

Proof. The Bernstein inequality yields

$$\mathbb{P}\left\{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right| \ge u\right\} \le 2e^{-\frac{nu^2}{2\left(\left(\sigma_{j,\xi}\right)^2 + M_{j,\xi}u/3\right)}}.$$

Taking $u = \tau(c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi})$, we obtain

$$\begin{aligned}
\mathbb{P}\left\{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right| \ge \tau(c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi})\right\} &\le 2e^{-\frac{n\tau^2(c_\sigma\sigma_{j,\xi}+c_M M_{j,\xi})^2}{2\left(\left(\sigma_{j,\xi}\right)^2 + M_{j,\xi}\tau(c_\sigma\sigma_{j,\xi}+c_M M_{j,\xi})/3\right)}} \\
&\le 2\left(e^{-\frac{\tau^2 n(c_\sigma\sigma_{j,\xi}+c_M M_{j,\xi})^2}{4\left(\sigma_{j,\xi}\right)^2}} + e^{-\frac{\tau 3n(c_\sigma\sigma_{j,\xi}+c_M M_{j,\xi})}{4M_{j,\xi}}}\right) \\
&\le 2\left(e^{-\frac{1}{4}n\left(c_\sigma + c_M\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2} + e^{-\frac{3}{4}n\left(c_\sigma\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M\right)\tau}\right).
\end{aligned}$$

We use $\mathbb{E}\left[|X|^z\right] = \int_{\mathbb{R}^+} zu^{z-1}\mathbb{P}\{|X| > u\}du$ and the upper bounds (29) and (30) to derive

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}}\right)^z\right] &\le \int_{\mathbb{R}^+} z\tau^{z-1}\mathbb{P}\left\{\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}} \ge \tau\right\}d\tau \\
&\le \int_{\mathbb{R}^+} z\tau^{z-1}2\left(e^{-\frac{1}{4}n\left(c_\sigma + c_M\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2} + e^{-\frac{3}{4}n\left(c_\sigma\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M\right)\tau}\right)d\tau
\end{aligned}$$

this yields the inequality from the lemma □

Lemma 15 allows to obtain the upper bounds (32) and (34) below.
Taking $c_\sigma = 1$ and $c_M = 0$ yields

$$\mathbb{E}\left[\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z\right] \le 2\left(\left(2c_{2,z}^{1/z}\frac{\sigma_{j,\xi}}{\sqrt{n}}\right)^z + \left(\frac{4}{3}c_{1,z}^{1/z}\frac{M_{j,\xi}}{n}\right)^z\right) \tag{32}$$

while taking $c_\sigma = c'_\sigma t_n$ and $c_M = c'_M \log n/(n-1)$ we obtain

$$\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c'_\sigma\sqrt{\log n}\frac{\sigma_{j,\xi}}{\sqrt{n}} + c'_M \log n \frac{M_{j,\xi}}{n-1}}\right)^z\right] \leq 2\left(c_{2,z}\left(2\frac{1}{c'_\sigma\sqrt{\log n} + c'_M\frac{\sqrt{n}\log n}{n-1}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z\right.$$
$$\left. + c_{1,z}\left(\frac{4}{3}\frac{1}{c'_\sigma\sqrt{n}\sqrt{\log n}\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c'_M\log n\frac{n}{n-1}}\right)^z\right)$$
$$\leq 2\left(c_{2,z}\left(2\frac{1}{c'_\sigma\sqrt{\log n}}\right)^z + c_{1,z}\left(\frac{4}{3}\frac{1}{c'_M\log n}\right)^z\right) \tag{33}$$

The following lemma is useful to handle the case $p = \infty$.

Lemma 16. *For any $\Xi'_j \subset \Xi_j$,*

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}}\right)^z\right] \leq \left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{1}{c_\sigma + c_M\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z\left(2 + \left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)\right)^{z/2}\right)$$
$$+ \left(\frac{8}{3n}\frac{1}{c_\sigma\inf_{\xi\in\Xi'_j}\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M}\right)^z\left(2 + \left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right) \tag{34}$$

Proof. A uniform union bound yields

$$\mathbb{P}\left\{\sup_{\xi\in\Xi'_j}\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}} \geq \tau\right\}$$
$$\leq \min\left(1, \left|\Xi'_j\right| 2\left(e^{-\frac{1}{4}n\left(c_\sigma + c_M\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2} + e^{-\frac{3}{4}n\left(c_\sigma\inf_{\xi\in\Xi'_j}\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M\right)\tau}\right)\right)$$
$$\leq \min\left(1, \left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma + c_M\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2}\right) + \min\left(1, \left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma\inf_{\xi\in\Xi'_j}\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M\right)\tau}\right).$$

This yields

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}}\right)^z\right] \leq \int_{\mathbb{R}^+} z\tau^{z-1}\min\left(1, \left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma + c_M\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2}\right)d\tau$$
$$+ \int_{\mathbb{R}^+} z\tau^{z-1}\min\left(1, \left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma\inf_{\xi\in\Xi'_j}\frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M\right)\tau}\right)d\tau$$

and thus, for any $\tau_1 \geq 0$ and $\tau_2 \geq 0$, we get

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma_{j,\xi} + c_M M_{j,\xi}}\right)^z\right] \leq \tau_2^z + \int_{\tau\geq\tau_2} z\tau^{z-1}\left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma + c_M\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2}d\tau$$

$$+\tau_1^z+\int_{\tau\geq\tau_1} z\tau^{z-1}\left|\Xi_j'\right|2e^{-\frac{3}{4}n\left(c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M\right)\tau}d\tau.$$

Take

$$\tau_1=\frac{8}{3n}\frac{\log\left(c_{1,z}\left|\Xi_j'\right|\right)}{c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M}\quad\text{and}\quad\tau_2=\frac{2\sqrt{2}}{\sqrt{n}}\frac{\sqrt{\log\left(c_{2,z}\left|\Xi_j'\right|\right)}}{c_\sigma+c_M\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}$$

by construction

$$\forall\tau\geq\tau_1,\ \left|\Xi_j'\right|2e^{-\frac{3}{4}n\left(c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M\right)\tau}\leq\frac{2}{c_{1,z}}e^{-\frac{3}{8}n\left(c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M\right)\tau}$$

$$\forall\tau\geq\tau_2,\ \left|\Xi_j'\right|2e^{-\frac{1}{4}n\left(c_\sigma+c_M\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2}\leq\frac{2}{c_{2,z}}e^{-\frac{1}{8}n\left(c_\sigma+c_M\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^2\tau^2}.$$

This implies

$$\begin{aligned}&\mathbb{E}\left[\sup_{\xi\in\Xi_j'}\left(\frac{\left|\widehat{\beta}_{j,\xi}^a-\beta_{j,\xi}^a\right|}{c_\sigma\sigma_{j,\xi}+c_M M_{j,\xi}}\right)^z\right]\\&\leq\left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{\sqrt{\log\left(c_{2,z}\left|\Xi_j'\right|\right)}}{c_\sigma+c_M\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z+2\left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{1}{c_\sigma+c_M\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z\\&\quad+\left(\frac{8}{3n}\frac{\log\left(c_{1,z}\left|\Xi_j'\right|\right)}{c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M}\right)^z+2\left(\frac{8}{3n}\frac{1}{c_\sigma\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c_M}\right)^z\end{aligned}$$

which allows to conclude □

Lemma 16 allows to obtain the upper bounds (36) and (37) below.
If $c_\sigma=1$ and $c_M=0$ (34) reduces to

$$\begin{aligned}&\mathbb{E}\left[\sup_{\xi\in\Xi_j'}\left(\frac{\left|\widehat{\beta}_{j,\xi}^a-\beta_{j,\xi}^a\right|}{\sigma_{j,\xi}}\right)^z\right]\\&\leq\left(\frac{2\sqrt{2}}{\sqrt{n}}\right)^z\left(2+\left(\log\left(c_{2,z}\left|\Xi_j'\right|\right)\right)^{z/2}\right)+\left(\frac{8}{3n}\sup_{\xi\in\Xi_j'}\frac{M_{j,\xi}}{\sigma_{j,\xi}}\right)^z\left(2+\left(\log\left(c_{1,z}\left|\Xi_j'\right|\right)\right)^z\right).\end{aligned}$$

For future use, note that we can also use the uniform bounds M_j (see (24)) and

$$\sigma_{j,\xi}\leq C_2B(d,2)B_X^{1/2}2^{j\nu(d)}\stackrel{\text{def}}{=}\sigma_j \tag{35}$$

instead of $M_{j,\xi}$ and $\sigma_{j,\xi}$ and obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\\
&\leq\left(\frac{2\sqrt{2}}{\sqrt{n}}\sigma_j\right)^z\left(2+\left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)\right)^{z/2}\right)+\left(\frac{8}{3n}M_j\right)^z\left(2+\left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right).
\end{aligned}\tag{36}$$

Along the same lines, with $c_\sigma=c'_\sigma t_n$ and $c_M=c'_M\log n/(n-1)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|}{c'_\sigma\sqrt{\log n}\frac{\sigma_{j,\xi}}{\sqrt{n}}+c'_M\log n\frac{M_{j,\xi}}{n-1}}\right)^z\right]\\
&\leq\left(\frac{2\sqrt{2}}{c'_\sigma\sqrt{\log n}+c'_M\log n/\sqrt{n-1}\inf_{\xi\in\Xi'_j}\frac{M_{j,\xi}}{\sigma_{j,\xi}}}\right)^z\left(2+\left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)c_{1,z}\right)^{z/2}\right)\\
&\quad+\left(\frac{8/3}{c'_\sigma\sqrt{n\log n}\inf_{\xi\in\Xi'_j}\frac{\sigma_{j,\xi}}{M_{j,\xi}}+c'_M\log n}\right)^z\left(2+\left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right)\\
&\leq\left(\frac{2\sqrt{2}}{c'_\sigma\sqrt{\log n}}\right)^z\left(2+\left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)\right)^{z/2}\right)+\left(\frac{8}{3c'_M\log n}\right)^z\left(2+\left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right)
\end{aligned}\tag{37}$$

recall that when $\Xi'_j=\Xi_j$, $\left|\Xi'_j\right|\leq C_\Xi 2^{j(d-1)}$.

7.6.5. *Empirical Bernstein and the probabilities.* We take

$$\begin{aligned}
&\Delta_{j,\xi,\gamma}=\sqrt{2\gamma}t_n\widehat{\sigma}_{j,\xi}+\frac{14}{3}M_{j,\xi}\frac{\gamma\log n}{n-1}\\
&T_{j,\xi,\gamma}=2\Delta_{j,\xi,\gamma},\quad T^b_{j,\xi,\gamma}=\Delta_{j,\xi,\gamma},\quad T^s_{j,\xi,\gamma}=3\Delta_{j,\xi,\gamma}\\
&\Delta^+_{j,\xi,\gamma}=\sqrt{2\gamma}t_n\sigma_{j,\xi}+\frac{26}{3}M_{j,\xi}\frac{\gamma\log n}{n-1}\quad\text{and}\quad\Delta^-_{j,\xi,\gamma}=\sqrt{2\gamma}t_n\sigma_{j,\xi}+\frac{2}{3}M_{j,\xi}\frac{\gamma\log n}{n-1}\\
&T^{b,-}_{j,\xi,\gamma}=\Delta^-_{j,\xi,\gamma}\quad\text{and}\quad T^{s,+}_{j,\xi,\gamma}=3\Delta^+_{j,\xi,\gamma}.
\end{aligned}$$

Lemma 17. *The following upper bounds hold*

$$\begin{aligned}
&\mathbb{P}\left\{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}\right\}\leq\frac{1}{n^\gamma}\\
&\mathbb{P}\left\{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}\right\}\leq\frac{1}{n^\gamma}\\
&\mathbb{P}\left\{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\leq\frac{3}{n^\gamma}
\end{aligned}$$

$$\mathbb{P}\left(\bigcup_{\xi\in\Xi_j}\left\{T^{s,+}_{j,\xi,\gamma}<T^{s}_{j,\xi,\gamma}\right\}\right)\leq\sum_{\xi\in\xi_j}\mathbb{P}\left\{\ T^{s,+}_{j,\xi,\gamma}<T^{s}_{j,\xi,\gamma}\right\}\leq C_{\Xi}2^{j(d-1)}\frac{1}{n^{\gamma}}$$

$$\mathbb{P}\left(\bigcup_{\xi\in\Xi_j}\left\{T^{b,-}_{j,\xi,\gamma}>T^{b}_{j,\xi,\gamma}\right\}\right)\leq\sum_{\xi\in\xi_j}\mathbb{P}\left\{T^{b,-}_{j,\xi,\gamma}>T^{b}_{j,\xi,\gamma}\right\}\leq C_{\Xi}2^{j(d-1)}\frac{1}{n^{\gamma}}$$

$$\mathbb{P}\left(\bigcup_{\xi\in\Xi_j}\left\{\left|\widehat{\beta}^{a}_{j,\xi}-\beta^{a}_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\right)\leq\sum_{\xi\in\xi_j}\mathbb{P}\left\{\left|\widehat{\beta}^{a}_{j,\xi}-\beta^{a}_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\leq C_{\Xi}2^{j(d-1)}\frac{3}{n^{\gamma}}.$$

Proof. Using the results of [20] we get

$$\mathbb{P}\left\{\sigma_{j,\xi}>\widehat{\sigma}_{j,\xi}+2\sqrt{2u}\frac{M_{j,\xi}}{\sqrt{n-1}}\right\}\leq e^{-u}$$

$$\mathbb{P}\left\{\sigma_{j,\xi}<\widehat{\sigma}_{j,\xi}-2\sqrt{2u}\frac{M_{j,\xi}}{\sqrt{n-1}}\right\}\leq e^{-u}$$

$$\mathbb{P}\left\{\left|\widehat{\beta}^{a}_{j,\xi}-\beta^{a}_{j,\xi}\right|>\sqrt{2u}\frac{\widehat{\sigma}_{j,\xi}}{\sqrt{n}}+\frac{14}{3}M_{j,\xi}\frac{u}{n-1}\right\}\leq 3e^{-u}$$

which yield the first inequalities. The others follow from the union bound □

7.6.6. *The case $p=\infty$.* Let us consider the various terms one by one.
Error in the high dimensional space.

$$\mathbb{E}\left[M_{j,z}\right]\leq\mathbb{E}\left[M^{S0}_{j,z}\right]+\mathbb{E}\left[M^{S1}_{j,z}\right]+\mathbb{E}\left[M^{B1}_{j,z}\right]+\mathbb{E}\left[M^{B2}_{j,z}\right]$$

with

$$\mathbb{E}\left[M^{S0}_{j,z}\right]=\sup_{\xi\in\Xi_j}\left|\beta^{a}_{j,\xi}\right|^{z}\mathbf{1}_{\left|\beta^{a}_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}}$$

$$\mathbb{E}\left[M^{S1}_{j,z}\right]\leq C_{\Xi}2^{j(d-1)}\frac{4}{n^{\gamma}}\sup_{\xi\in\Xi_j}\left|\beta^{a}_{j,\xi}\right|^{z}$$

$$\mathbb{E}\left[M^{B1}_{j,z}\right]\leq\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{a}_{j,\xi}-\beta^{a}_{j,\xi}\right|^{z}\mathbf{1}_{\left|\beta^{a}_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]$$

$$\mathbb{E}\left[M^{B2}_{j,z}\right]\leq\left(C_{\Xi}2^{j(d-1)}\frac{4}{n^{\gamma}}\right)^{1-1/\tau}\left(\left(\frac{2\sqrt{2}}{\sqrt{n}}\sigma_j\right)^{z}\left(2^{1/\tau}+\left(\sqrt{\log\left(\left|\Xi_j\right|c_{1,z\tau}\right)}\right)^{z}\right)\right.$$
$$\left.+\left(\frac{8}{3n}M_j\right)^{z}\left(2^{1/\tau}+\left(\log\left(\left|\Xi_j\right|c_{1,z\tau}\right)\right)^{z}\right)\right)$$

where we have used $(a+b)^{1/\tau} \leq a^{1/\tau} + b^{1/\tau}$ for $\tau \geq 1$.
This yields

$$
\begin{aligned}
&\frac{\mathbb{E}\left[\left\|\widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J}\right\|_\infty^z\right]}{(J+1)^{z-1}C_\infty'^z} \\
&\leq \sum_{j=0}^{J} 2^{j(d-1)z/2}\left(\sup_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right|\leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{b,-}}\right]\right) \\
&\quad + \frac{4}{n^\gamma}C_\Xi \sum_{j=0}^{J} 2^{j(d-1)(z/2+1)} \sup_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z \\
&\quad + \left(C_\Xi \frac{4}{n^\gamma}\right)^{1-1/\tau} \sum_{j=0}^{J} 2^{j(d-1)(z/2+1-1/\tau)} \\
&\qquad \times \left(\left(\frac{2\sqrt{2}}{\sqrt{n}}\sigma_j\right)^z \left(2^{1/\tau} + \left(\sqrt{\log\left(|\Xi_j|\, c_{1,z\tau}\right)}\right)^z\right) + \left(\frac{8}{3n}M_j\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_j|\, c_{1,z\tau}\right)\right)^z\right)\right) \\
&\stackrel{\text{def}}{=} O_{\infty,z}' + R_{1,\infty,z}' + R_{2,\infty,z}'.
\end{aligned}
$$

The terms $R_{1,\infty,z}'$ and $R_{2,\infty,z}'$. The term $R_{1,\infty,z}'$ is the term which appears in theorem 14 and thus we only need to bound $R_{2,\infty,z}'$. As in the case $p<\infty$, we can use the uniform bounds on $\sigma_{j,\xi}$ and $M_{j,\xi}$, namely (24) and (35), and $|\Xi_j| \leq |\Xi_J|$ to obtain

$$
\begin{aligned}
&R_{2,\infty,z}' \\
&\leq \left(\frac{4C_\Xi}{n^\gamma}\right)^{1-1/\tau} \sum_{j=0}^{J} 2^{j(d-1)(z/2+1-1/\tau)} \\
&\quad \times \left(\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2 B(d,2) 2^{j\nu(d)} B_X^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(c_{2,z\tau}\,|\Xi_J|\right)\right)^{z/2}\right)\right. \\
&\qquad \left. + \left(\frac{8}{3n}C_\infty B(d,\infty) 2^{j(\nu(d)+(d-1)/2)} B_X\right)^z \left(2^{1/\tau} + \left(\log\left(c_{1,z\tau}\,|\Xi_J|\right)\right)^z\right)\right) \\
&\leq \left(\frac{4C_\Xi}{n^\gamma}\right)^{1-1/\tau} \left[\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2 B(d,2) B_X^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_J|\, c_{2,z\tau}\right)\right)^{z/2}\right) \sum_{j=0}^{J} 2^{j(\nu(d)z+(d-1)(z/2+1-1/\tau))}\right. \\
&\qquad \left. + \left(\frac{8}{3n}C_\infty B(d,\infty) B_X\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_J|\, c_{1,z\tau}\right)\right)^z\right) \sum_{j=0}^{J} 2^{j(\nu(d)z+(d-1)(z+1-1/\tau))}\right] \\
&\leq \left(\frac{4C_\Xi}{n^\gamma}\right)^{1-1/\tau} \left[\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2 B(d,2) B_X^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(c_{2,z\tau}\,|\Xi_J|\right)\right)^{z/2}\right) \frac{2^{J(\nu(d)z+(d-1)(z/2+1-1/\tau))}}{1-2^{-(\nu(d)z+(d-1)(z/2+1-1/\tau))}}\right.
\end{aligned}
$$

$$+\left(\frac{8}{3n}C_\infty B(d,\infty)B_X\right)^z\left(2^{1/\tau}+\left(\log\left(c_{1,z\tau}\left|\Xi_J\right|\right)\right)^z\right)\frac{2^{J(\nu(d)z+(d-1)(z+1-1/\tau))}}{1-2^{-(\nu(d)z+(d-1)(z+1-1/\tau))}}\Bigg].$$

The term $O'_{\infty,z}$. Denote by

$$O'_{z,j}=\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]$$

Because $T^{s,++}_{j,\xi,\gamma}\ge T^{s,+}_{j,\xi,\gamma}$, we get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]\\
&=\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}\right]+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{T^{s,++}_{j,\xi,\gamma}\ge\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]\\
&\le\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}\right]\\
&\quad+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|}{T^{b,-}_{j,\xi,\gamma}}\mathbf{1}_{T^{s,++}_{j,\xi,\gamma}\ge\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right)^z\right]\sup_{\xi\in\Xi_j}\left\{\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{T^{s,++}_{j,\xi,\gamma}\ge\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right\},
\end{aligned}$$

thus

$$\begin{aligned}
O'_{z,j}\le&\left(1+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left(\frac{\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|}{T^{b,-}_{j,\xi,\gamma}}\right)^z\right]\right)\sup_{\xi\in\Xi_j}\left\{\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,++}_{j,\xi,\gamma}}\right\}\\
&+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}\right].
\end{aligned}$$

Using now (37) with $c'_\sigma=\sqrt{2\gamma}$ and $c'_M=\frac{2}{3}\gamma$, we get as $|\Xi_j|\le C_\Xi 2^{j(d-1)}$ gives the upper bound in theorem 14.

7.6.7. *The case $p<\infty$.* Let us consider the various terms one by one.
Error in the high dimensional space. We obtain

$$\mathbb{E}\left[S_{j,z}\right]=\mathbb{E}\left[S^{S0}_{j,z}\right]+\mathbb{E}\left[S^{S1}_{j,z}\right]+\mathbb{E}\left[S^{B1}_{j,z}\right]+\mathbb{E}\left[S^{B2}_{j,z}\right].$$

with

$$\begin{aligned}
\mathbb{E}\left[S^{S0}_{j,z}\right]&=\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}\\
\mathbb{E}\left[S^{S1}_{j,z}\right]&\le\frac{4}{n^\gamma}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\\
\mathbb{E}\left[S^{B1}_{j,z}\right]&\le\sum_{\xi\in\Xi_j}\mathbb{E}\left[\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}
\end{aligned}$$

$$\mathbb{E}\left[S_{j,z}^{B2}\right] \leq \frac{4^{1-1/\tau}}{n^{\gamma(1-1/\tau)}} \sum_{\xi\in\Xi_j} 2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_{j,\xi}}{\sqrt{n}}\right)^z + \left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_{j,\xi}}{n}\right)^z\right)$$

where we have used $(a+b)^{1/\tau} \leq \left(a^{1/\tau}+b^{1/\tau}\right)$. This yields

$$\begin{aligned}
&\frac{\mathbb{E}\left[\left\|\widehat{f_\beta^-}^{a,\rho} - f_\beta^{-a,J}\right\|_p^z\right]}{(J+1)^{z-1}C_p'^z C_\Xi^{z/(p\wedge z)-1}} \\
&\leq \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))}\mathbb{E}\left[S_{j,z}\right] \\
&\leq \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))} \sum_{\xi\in\Xi_j}\left(\left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right|\leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]\mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{b,-}}\right) \\
&\quad + \frac{4}{n^\gamma}\sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z \\
&\quad + \frac{2^{2-1/\tau}}{n^{\gamma(1-1/\tau)}}\sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))}\sum_{\xi\in\Xi_j}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_{j,\xi}}{\sqrt{n}}\right)^z + \left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_{j,\xi}}{n}\right)^z\right) \\
&\stackrel{\text{def}}{=} O_{p,z} + R_{1,p,z} + R_{2,p,z}.
\end{aligned}$$

The terms $R_{1,p,z}$ and $R_{2,p,z}$. The term $R_{1,p,z}$ appears as is in lemma 14. To bound the term $R_{2,p,z}$, we rely on (24). We obtain

$$\begin{aligned}
&\sum_{\xi\in\Xi_j} 2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_j}{\sqrt{n}}\right)^z + \left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_j}{n}\right)^z\right) \\
&\leq \sum_{\xi\in\Xi_j} 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)2^{j\nu(d)}B_X^{1/2}\frac{1}{\sqrt{n}}\right)^z \\
&\quad + \sum_{\xi\in\Xi_j} 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)2^{j(\nu(d)+(d-1)/2)}B_X\frac{1}{n}\right)^z \\
&\leq C_\Xi 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z B_X^{z/2}\frac{1}{n^{z/2}}2^{j((d-1)+z\nu(d))} \\
&\quad + C_\Xi 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z B_X^z\frac{1}{n^z}2^{j((d-1)+z(\nu(d)+(d-1)/2))}
\end{aligned}$$

this yields

$$\sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))}\sum_{\xi\in\Xi_j} 2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_j}{\sqrt{n}}\right)^z + \left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_j}{n}\right)^z\right)$$

$$
\begin{aligned}
&\leq \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))} C_{\Xi} 2^{1/\tau} \left(2c_{2,z\tau}^{1/(z\tau)} C_2 B(d,2)\right)^z B_X^{z/2} \frac{1}{n^{z/2}} 2^{j((d-1)+z\nu(d))} \\
&\quad + \sum_{j=0}^{J} 2^{j(d-1)z(1/2-1/(p\vee z))} C_{\Xi} 2^{1/\tau} \left(\frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d,\infty)\right)^z B_X^{z} \frac{1}{n^{z}} 2^{j((d-1)+z(\nu(d)+(d-1)/2))} \\
&\leq C_{\Xi} 2^{1/\tau} \left(2c_{2,z\tau}^{1/(z\tau)} C_2 B(d,2)\right)^z B_X^{z/2} \frac{1}{n^{z/2}} \sum_{j=0}^{J} 2^{jz(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\vee z)))} \\
&\quad + C_{\Xi} 2^{1/\tau} \left(\frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d,\infty)\right)^z B_X^{z} \frac{1}{n^{z}} \sum_{j=0}^{J} 2^{jz(\nu(d)+(d-1)/z+(d-1)(1-1/(p\vee z))} \\
&\leq \frac{C_{\Xi} 2^{1/\tau} \left(2c_{2,z\tau}^{1/(z\tau)} C_2 B(d,2)\right)^z}{1-2^{-z(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\vee z)))}} B_X^{z/2} \frac{1}{n^{z/2}} 2^{Jz(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\vee z)))} \\
&\quad + \frac{C_{\Xi} 2^{1/\tau} \left(\frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d,\infty)\right)^z}{1-2^{-z(\nu(d)+(d-1)/z+(d-1)(1-1/(p\vee z))}} B_X^{z} \frac{1}{n^{z}} 2^{Jz(\nu(d)+(d-1)/z+(d-1)(1-1/(p\vee z))}.
\end{aligned}
$$

The term $O_{p,z}$. Denote by

$$
O_{z,j,\xi} = \left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| \leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}}.
$$

Because $T_{j,\xi,\gamma}^{s,++} \geq T_{j,\xi,\gamma}^{s,+}$, we get

$$
\begin{aligned}
&\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}} \\
&= \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{s,++}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq \left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}} \\
&\leq \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{s,++}} + \frac{\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]}{\left(T_{j,\xi,\gamma}^{b,-}\right)^z} \left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq \left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{b,-}},
\end{aligned}
$$

$$
O_{z,j,\xi} \leq \left(1 + \frac{\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]}{\left(T_{j,\xi,\gamma}^{b,-}\right)^z}\right) \left|\beta_{j,\xi}^a\right|^z \mathbf{1}_{\left|\beta_{j,\xi}^a\right| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right] \mathbf{1}_{\left|\beta_{j,\xi}^a\right| > T_{j,\xi,\gamma}^{s,++}}.
$$

Now using the results of Section 7.6.4 with $T_{j,\xi,\gamma}^{b,-} = \sqrt{2\gamma} t_n \sigma_{j,\xi} + \frac{2}{3}\gamma \frac{\log n}{n-1} M_{j,\xi}$ we obtain

$$
\begin{aligned}
\sup_{j,\xi} \frac{\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right|^z\right]}{\left(T_{j,\xi,\gamma}^{b,-}\right)^z} &\leq 2\left(c_{2,z}\left(2\frac{1}{\sqrt{2\gamma}\sqrt{\log n}}\right)^z + c_{1,z}\left(\frac{4}{3}\frac{1}{(2/3)\gamma \log n}\right)^z\right) \\
&\leq 2\left(\left(\frac{\sqrt{2} c_{2,z}^{1/z}}{\sqrt{\gamma \log n}}\right)^z + \left(\frac{2c_{1,z}^{1/z}}{\gamma \log n}\right)^z\right),
\end{aligned}
$$

this yields

$$O_{p,z} \leq \left(1+2\left(\left(\frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma\log n}}\right)^z+\left(\frac{2c_{1,z}^{1/z}}{\gamma\log n}\right)^z\right)\right)\sum_{j=0}^{J}2^{j(d-1)z(1/2-1/(p\vee z))}$$
$$\sum_{\xi\in\Xi_j}\left(\left|\beta_{j,\xi}^a\right|^z\mathbf{1}_{|\beta_{j,\xi}^a|\leq T_{j,\xi,\gamma}^{s,++}}+\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^a-\beta_{j,\xi}^a\right|^z\right]\mathbf{1}_{|\beta_{j,\xi}^a|>T_{j,\xi,\gamma}^{s,++}}\right).$$

7.7. **Proof of theorem 9.** This proof requires an upper bound on: the approximation error, $R_{1,p,z}$, $R_{1,p,z}$, and $O_{p,z}$. We use that because $f_\beta\in B_{r,q}^s(M)$, we have, by lemma 13, $f_\beta^-\in B_{r,q}^s(c_{\mathrm{eq}}M)$.

7.7.1. *The case* $1\leq p<\infty$. Let us consider the terms one by one.
The approximation error. Start with

$$\left\|f_\beta^{-a,J}-f_\beta^-\right\|_p=\left\|\sum_{j>J}\sum_{\xi\in\Xi_j}\beta_{j,\xi}^a\psi_{j,\xi}\right\|_p$$

From lemma 6 (i) and the definition of the Besov spaces as a sequence space, with $1/q+1/\tilde{q}=1$,

$$\left\|\sum_{j>J}\sum_{\xi\in\Xi_j}\beta_{j,\xi}^a\psi_{j,\xi}\right\|_p\leq C_p'\sum_{j>J}2^{-js}2^{j(s+(d-1)(1/2-1/p))}\left\|\left(\beta_{j,\xi}^a\right)_{\xi\in\Xi_j}\right\|_{\ell^p}$$

which yields

$$\left\|\sum_{j>J}\sum_{\xi\in\Xi_j}\beta_{j,\xi}^a\psi_{j,\xi}\right\|_p\leq C_p'2^{-Js}(2^{s\tilde{q}}-1)^{-1/\tilde{q}}\left\|f_\beta^-\right\|_{B_{p,q}^s}$$
$$\leq\begin{cases}C_p'c_{\mathrm{eq}}MC_\Xi^{1/p-1/r}(2^{s\tilde{q}}-1)^{-1/\tilde{q}}\left\|f_\beta^-\right\|_{B_{p,q}^s}2^{-Js} & \text{if } r\geq p\\ C_p'c_{\mathrm{eq}}M(2^{s\tilde{q}}-1)^{-1/\tilde{q}}\left\|f_\beta^-\right\|_{B_{p,q}^s}2^{-J(s-(d-1)(1/r-1/p))} & \text{if } r\leq p\end{cases}$$

It is enough to consider the worst case where $r\leq p$ and to check that $\frac{s-(d-1)(1/r-1/p)}{\nu(d)+(d-1)/2}\geq\mu$ in the two zones.
In the dense zone, we have

$$s+\nu(d)+\frac{d-1}{2}\geq\left(\nu(d)+\frac{d-1}{2}\right)\frac{p}{r}$$

which yields

$$\frac{s}{s+\nu(d)+\frac{d-1}{2}}\leq\frac{s}{\left(\nu(d)+\frac{d-1}{2}\right)\frac{p}{r}}.$$

Because $s > (d-1)/r$ and $p \geq r$, we have

$$s - \frac{d-1}{r} + \frac{d-1}{p} - \frac{sr}{p} = (d-1)\left(\frac{sr}{d-1} - 1\right)\left(\frac{1}{r} - \frac{1}{p}\right) \geq 0,$$

which yields $s - (d-1)(1/r - 1/p) \geq \frac{sr}{p}$ and gives the result.
In the sparse zone, because $s > (d-1)/r$, we have

$$\frac{s-(d-1)(1/r-1/p)}{\nu(d)+(d-1)/2} \geq \frac{s-(d-1)(1/r-1/p)}{s+\nu(d)-(d-1)(1/r-1/2)}.$$

The terms $R_{1,p,p}$ and $R_{2,p,p}$. Using lemma 7 (iii) we obtain

$$R_{1,p,p} \leq \frac{4}{n^\gamma}(c_{\mathrm{eq}}M)^p C_\Xi^{1-(p\wedge r)/r} \sum_{j=0}^{J} 2^{-jp(s+(d-1)(1/p-1/(p\wedge r)))}$$

where the exponent is nonpositive because $s > (d-1)/r$, thus

$$R_{1,p,p} \leq \frac{4(c_{\mathrm{eq}}M)^p C_\Xi^{1-(p\wedge r)/r}}{n^\gamma\left(1-2^{-p(s+(d-1)(1/p-1/(p\wedge r)))}\right)}.$$

With $\gamma > p/2$, $R_{1,p,p}$ is of lower order than t_n^p.
We also have

$$R_{2,p,p} \leq \frac{2^{2-1/\tau}}{n^{\gamma(1-1/\tau)}} C_\Xi b_{n,p,p,J,\tau}.$$

With the aforementioned choice of J,

$$\frac{1}{\sqrt{n}} 2^{J(\nu(d)+(d-1)/2)} B_X^{1/2} \lesssim 1$$

$$\frac{2^{J(d-1)}}{n} B_X \lesssim 1.$$

Together these yield that $b_{n,p,p,J,\tau}$ is of the order of a constant.
This term is also of lower order than t_n^p for τ large enough such that $\gamma(1-1/\tau) > p/2$.

The term $O_{p,p}$. First note that $a_{n,p,p,J} = 1 + o(1)$.
We take $T_{j,\xi,\gamma}^{s,++}$ uniform in ξ:

$$\begin{aligned} T_{j,\xi,\gamma}^{s,++} &= 3\sqrt{2\gamma} t_n C_2 B(d,2) 2^{j\nu(d)} B_X^{1/2} \\ &\quad + 52 C_\infty B(d,\infty) 2^{j(\nu(d)+(d-1)/2)} B_X \frac{\gamma \log n}{n-1} \\ &\leq 2^{j\nu(d)} \sqrt{\gamma} t_n B_X^{1/2} \left(3\sqrt{2} C_2 B(d,2) + 52 C_\infty B(d,\infty) \frac{n\sqrt{\gamma}}{n-1}\right) \end{aligned}$$

where the last display uses the upper bound on J, this yields, for $n \geq 2$,

$$T^{s,++}_{j,\xi,\gamma} \leq 2^{j\nu(d)}\sqrt{\gamma}t_n B_X^{1/2}\left(3\sqrt{2}C_2B(d,2)+104C_\infty B(d,\infty)\right) \stackrel{\text{def}}{=} T^{s,++}_{j,\gamma}.$$

As a consequence of (32) we get

$$\begin{aligned}
\mathbb{E}\left[\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^p\right] &\leq 2\left(\left(2c_{2,p}^{1/p}\frac{\sigma_j}{\sqrt{n}}\right)^p+\left(\frac{4}{3}c_{1,p}^{1/p}\frac{M_j}{n}\right)^p\right)\\
&\leq 2\left(2c_{2,p}^{1/p}C_2B(d,2)2^{j\nu(d)}B_X^{1/2}\frac{1}{\sqrt{n}}\right)^p\\
&\quad+2\left(\frac{8}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)2^{j(\nu(d)+(d-1)/2)}B_X\frac{1}{n}\right)^p\\
&\leq 2^{jp\nu(d)}\frac{1}{n^{p/2}}B_X^{p/2}2^{p+1}\left(c_{2,p}^{1/p}C_2B(d,2)+\frac{4}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)\right)^p\\
&\leq \frac{\left(T^{s,++}_{j,\gamma}\right)^p}{(\gamma\log n)^{p/2}}2\left(2\frac{c_{2,p}^{1/p}C_2B(d,2)+\frac{4}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)}{3\sqrt{2}C_2B(d,2)+104C_\infty B(d,\infty)\left(\sqrt{\gamma}\right)}\right)^p\\
&\leq \frac{\left(T^{s,++}_{j,\gamma}\right)^p}{(\gamma\log n)^{p/2}}2\left(\frac{\sqrt{2}}{3}c_{2,p}^{1/p}+\frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}}\right)^p.
\end{aligned}$$

Let $\quad C_\gamma = 3\sqrt{2}C_2B(d,2)+104C_\infty B(d,\infty)\sqrt{\gamma}\quad$ and $\quad C_{\sigma,p}=2^{1/p}\left(\frac{\sqrt{2}}{3}c_{2,p}^{1/p}+\frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}}\right).$

For any $0<z<p$,

$$\begin{aligned}
&\sum_{\xi\in\Xi_j}\left(\left|\beta^a_{j,\xi}\right|^p\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,++}_{j,\gamma}}+\mathbb{E}\left[\left|\widehat{\beta}^a_{j,\xi}-\beta^a_{j,\xi}\right|^p\right]\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\gamma}}\right)\\
&\quad\leq\sum_{\xi\in\Xi_j}\left(\left|\beta^a_{j,\xi}\right|^p\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,++}_{j,\gamma}}+\frac{\left(T^{s,++}_{j,\gamma}\right)^p}{(\gamma\log n)^{p/2}}C^p_{\sigma,p}\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}\right)\\
&\quad\leq\left(1+\frac{C^p_{\sigma,p}}{(\gamma\log n)^{p/2}}\right)\left(T^{s,++}_{j,\gamma}\right)^{p-z}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\\
&\quad\leq\left(1+\frac{C^p_{\sigma,p}}{(\gamma\log n)^{p/2}}\right)\left(\sqrt{\gamma}t_nB_X^{1/2}C_\gamma\right)^{p-z}2^{j\nu(d)(p-z)}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z.
\end{aligned}$$

We need to sum over j and take two different values for z, one that we denote z_1 for $j\leq j_0$ and one that we denote z_2 for $j_0<j\leq J$. The values of z_1, z_2, j_0 will be specified later, depending on the value of the parameters r, q, s and p such that we are in the

dense or sparse zone. Up to a multiplying constant, we thus need to control

$$A+B=\left(B_X^{1/2}t_n\right)^{p-z_1}\sum_{j=0}^{j_0}2^{j[\nu(d)(p-z_1)+(d-1)(p/2-1)]}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^{z_1}$$
$$+\left(B_X^{1/2}t_n\right)^{p-z_2}\sum_{j=j_0+1}^{J}2^{j[\nu(d)(p-z_2)+(d-1)(p/2-1)]}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^{z_2}$$

where we choose adequately z_1, z_2 and j_0 in the two zones. Because of lemma 7 (i) we only consider $p\geq r$.
Let us first consider the dense zone. We define

$$\tilde{r}=\frac{p(\nu(d)+(d-1)/2)}{s+\nu(d)+(d-1)/2}.$$

In the dense zone, $\tilde{r}\leq r$, $p>\tilde{r}$ and

$$s=\left(\nu(d)+\frac{d-1}{2}\right)\left(\frac{p}{\tilde{r}}-1\right). \tag{38}$$

With $z_2=r$, we get

$$B\leq\left(B_X^{1/2}t_n\right)^{p-r}\sum_{j=j_0+1}^{J}2^{j[\nu(d)(p-r)+(d-1)(p/2-1)]}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^{r}.$$

Lemma 7 (iii) gives that

$$\sum_{\xi\in\Xi_j}|\beta_{j,\xi}|^r\leq D_j^r2^{-jr(s+(d-1)(1/2-1/r))}$$

where $\forall j\in\mathbb{N},\ D_j\geq 0$, $(D_j)_{j\in\mathbb{N}}\in\ell_q$. Note that

$$s+(d-1)\left(\frac{1}{2}-\frac{1}{r}\right)=\frac{(d-1)p}{2\tilde{r}}-\frac{d-1}{r}+\nu(d)\left(\frac{p}{\tilde{r}}-1\right), \tag{39}$$

thus

$$B\leq\left(B_X^{1/2}t_n\right)^{p-r}\sum_{j=j_0+1}^{J}2^{jp\left(1-\frac{r}{\tilde{r}}\right)\left(\nu(d)+\frac{d-1}{2}\right)}D_j^r$$
$$\lesssim(c_{\mathrm{eq}}M)^r\left(B_X^{1/2}t_n\right)^{p-r}2^{j_0p\left(1-\frac{r}{\tilde{r}}\right)\left(\nu(d)+\frac{d-1}{2}\right)}$$

for $q\geq 1$ if $r>\tilde{r}$ and for $q\leq r$ if $r=\tilde{r}$ (i.e. $s=p\left(\nu(d)+\frac{d-1}{2}\right)\left(\frac{1}{r}-\frac{1}{p}\right)$).
Taking $2^{j_0\frac{p}{\tilde{r}}\left(\nu(d)+\frac{d-1}{2}\right)}\simeq\left(B_X^{1/2}t_n\right)^{-1}$ we get

$$B\lesssim M^r\left(B_X^{1/2}t_n\right)^{p-\tilde{r}}$$

which is the rate that we expect in that zone.
As for A, we take $z_1 = \overline{r} < \tilde{r} \leq r$, this yields, using lemma 7 (iii),

$$\begin{aligned}
A &\leq \left(B_X^{1/2} t_n\right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-\overline{r})+(d-1)(p/2-1)]} \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^{\overline{r}} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-\overline{r})+(d-1)(p/2-1)-\overline{r}(s+(d-1)(1/2-1/\overline{r}))]} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{jp(\nu(d)+(d-1)/2)(1-\overline{r}/\tilde{r})} \quad \text{(using (38))} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\overline{r}} 2^{j_0 p(\nu(d)+(d-1)/2)(1-\overline{r}/\tilde{r})} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\tilde{r}} \quad \text{(from the definition of } j_0\text{)}.
\end{aligned}$$

Let us now consider the sparse zone. We define by

$$\tilde{r} = p\frac{\nu(d)+(d-1)(1/2-1/p)}{s+\nu(d)-(d-1)(1/r-1/2)}$$

in a such a way that

$$\begin{aligned}
p-\tilde{r} &= p\frac{s-(d-1)(1/r-1/p)}{s+\nu(d)-(d-1)(1/r-1/2)} \\
\tilde{r}-r &= \frac{(p-r)((d-1)/2+\nu(d))-rs}{s+\nu(d)-(d-1)(1/r-1/2)} > 0 \\
s+(d-1)\left(\frac{1}{2}-\frac{1}{r}\right) &= \frac{(d-1)p}{2\tilde{r}} - \frac{d-1}{\tilde{r}} + \nu(d)\left(\frac{p}{\tilde{r}}-1\right). \qquad (40)
\end{aligned}$$

For the term A we take $z_1 = r$ and obtain

$$\begin{aligned}
A &\leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-r)+(d-1)(p/2-1)]} \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^{r} \\
&\leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=0}^{j_0} 2^{j\left[\nu(d)+(d-1)(1/2-1/p)\frac{p}{\tilde{r}}(\tilde{r}-r)\right]} D_j^r \quad \text{(using (40))} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-r} 2^{j_0\left[(\nu(d)+(d-1)(1/2-1/p)\frac{p}{\tilde{r}}(\tilde{r}-r)\right]} M^r,
\end{aligned}$$

the last inequality holds because $\nu(d)+(d-1)/2-(d-1)/p > 0$, indeed because we are in the sparse zone $\nu(d)+(d-1)/2 \geq s/(p/r-1) = sr/(p-r) \geq 2/(p-r) \geq (d-1)/p$.

Taking $2^{j_0(\nu(d)+(d-1)(1/2-1/p))\frac{p}{\tilde{r}}} \simeq \left(B_X^{1/2} t_n\right)^{-1}$ yields

$$A \lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\tilde{r}}.$$

For the term B we take $z_2 = \overline{r} > \tilde{r} > r$ and obtain

$$\begin{aligned}
B &\leq \left(B_X^{1/2} t_n\right)^{p-\overline{r}} \sum_{j=j_0+1}^{J} 2^{j[\nu(d)(p-\overline{r})+(d-1)(p/2-1)]} \sum_{\xi\in\Xi_j} \left|\beta_{j,\xi}^a\right|^{\overline{r}} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\overline{r}} \sum_{j=j_0+1}^{J} 2^{j(\nu(d)+(d-1)(1/2-1/p))p(r-\overline{r})/\tilde{r}} D_j^{\overline{r}} \quad \text{(using (40))} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\overline{r}} 2^{j_0(\nu(d)+(d-1)(1/2-1/p))p(r-\overline{r})/\tilde{r}} M^{\overline{r}} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\tilde{r}} M^{\overline{r}}.
\end{aligned}$$

7.7.2. *The case $p = \infty$.* Consider $r = \infty$. The general case follows by lemma 7 (ii).
The approximation error. Because $f_\beta \in B^s_{\infty,q}(M)$, we have by lemma 6 (i)

$$\begin{aligned}
\left\|\sum_{j>J}\sum_{\xi\in\Xi_j} \beta_{j,\xi}^a \psi_{j,\xi}\right\|_\infty &\leq \sum_{j>J} \left\|\sum_{\xi\in\Xi_j} \beta_{j,\xi}^a \psi_{j,\xi}\right\|_\infty \\
&\leq C'_\infty c_{\text{eq}} M \sum_{j>J} 2^{j(d-1)/2} 2^{-j(s+(d-1)/2)} D_j \quad (\text{where } \|(D_j)_{j\in\mathbb{N}}\|_q \leq c_{\text{eq}} M) \\
&\leq C'_\infty c_{\text{eq}} M 2^{-Js} (2^{s\tilde{q}} - 1)^{-1/\tilde{q}}.
\end{aligned}$$

From the choice of J we get

$$\left\|\sum_{j>J}\sum_{\xi\in\Xi_j} \beta_{j,\xi}^a \psi_{j,\xi}\right\|_\infty \lesssim C'_\infty c_{\text{eq}} M (2^{s\tilde{q}} - 1)^{-1/\tilde{q}} \left(t_n B_X^{1/2}\right)^{s/(\nu(d)+(d-1)/2)},$$

this term is negligible because $s/(\nu(d) + (d-1)/2) \geq s/(s\nu(d) + (d-1)/2)$.

The terms $R'_{1,\infty,z}$ and $R'_{2,\infty,z}$. Using the definition of the Besov norm, we obtain

$$\begin{aligned}
R'_{1,\infty,z} &\leq \frac{4}{n^\gamma} (c_{\text{eq}} M)^z C_\Xi \sum_{j=0}^{J} 2^{-jzs} 2^{j(d-1)} \\
&\lesssim \frac{4}{n^\gamma} 2^{J(d-1)} M^z.
\end{aligned}$$

With $\gamma > z/2 + 1$, which holds if $2(\gamma - 1)(1 - 1/\tau) > z$, $R_{1,\infty,z}$ is of lower order than t_n^z. Due to the choice of J the term in bracket in the expression of $R'_{2,\infty,z}$ in theorem 14 is

less than 1. The second term in the expression of $b_{n,\infty,z,J,\tau}$ is of smaller order than the first term. The order of $b_{n,\infty,z,J,\tau}$ is finally $(\log n)^{z/2}$. Thus we have

$$R'_{2,\infty,z} \lesssim \left(n^{-\gamma} 2^{J(d-1)}\right)^{1-1/\tau} (\log n)^{z/2}.$$

This term is also of lower order than t_n^z when τ is such that $2(\gamma-1)(1-1/\tau) > z$.

The term $O'_{\infty,z}$. Note that here $a_{n,\infty,z,J}$ is of the order of a constant. We now proceed like for the term $O_{p,p}$. Using (36), we obtain for arbitrary $\overline{z} \in [0,z]$

$$\begin{aligned}
&\sup_{\xi\in\Xi_j} \left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right| \le T^{s,++}_{j,\xi,\gamma}} + \mathbb{E}\left[\sup_{\xi\in\Xi_j} \left|\widehat{\beta}^a_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right| > T^{s,++}_{j,\xi,\gamma}}\right] \\
&\lesssim \left(\sqrt{\gamma} t_n B_X^{1/2}\right)^{z-\overline{z}} 2^{j\nu(d)(z-\overline{z})} \sup_{\xi\in\Xi_j} \left|\beta^a_{j,\xi}\right|^{\overline{z}}.
\end{aligned}$$

We use an upper bound on $A+B$ where

$$\begin{aligned}
A &= \left(B_X^{1/2} t_n\right)^{z-z_1} \sum_{j=0}^{j_0} 2^{j[\nu(d)(z-z_1)+(d-1)z/2]} \quad \sup_{\xi\in\Xi_j} \left|\beta^a_{j,\xi}\right|^{z_1} \\
B &= \left(B_X^{1/2} t_n\right)^{z-z_2} \sum_{j=j_0+1}^{J} 2^{j[\nu(d)(z-z_2)+(d-1)z/2]} \sup_{\xi\in\Xi_j} \left|\beta^a_{j,\xi}\right|^{z_2}
\end{aligned}$$

for well chosen $0 \le j_0 \le J$, z_1 and z_2. Because $f \in B^s_{\infty,q}(M)$, we have

$$\forall \overline{z} \ge 1,\ \sup_{\xi\in\Xi_j} \left|\beta^a_{j,\xi}\right|^{\overline{z}} \le (c_{\mathrm{eq}} M)^{\overline{z}} 2^{-j(s+(d-1)/2)\overline{z}}.$$

The result follows taking $z_1 = 0$, j_0 such that $2^{j_0} \simeq t_n^{-1/(s+\nu(d)+(d-1)/2)}$, and $z_2 = z$.

References

[1] Atkinson, K. and Han, W. (2012). *Spherical harmonics and approximations on the unit sphere: an introduction*. Springer.

[2] Baldi, P., Kerkyacharian, G., Marinucci, D., and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395.

[3] Beran, R., Feuerverger, A., and Hall, P. (1996). Estimating coefficient distributions in random coefficient regression. *Ann. Statist.* **24** 2569–2592.

[4] Beran, R. and Millar, P. W. (1994). Minimum distance estimation in random coefficient regression models. *Ann. Statist.* **22** 1976–1992.

[5] Bertin, K., Le Pennec, E., and Rivoirard, V. (2011). Adaptive Dantzig density estimation. *Ann. Inst. H. Poincaré Probab. Statist.* **47** 43–74.

[6] Funk, P. (1916). Uber eine geometrische anwendung der abelschen integralgleichung. *Math. Ann.* **77** 129–135.

[7] Gautier, E. and Hoderlein, S. (2011, 2015). A triangular treatment effect model with random coefficients in the selection equation. Preprint arXiv:1109.0362v4.

[8] Gautier, E. and Kitamura, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica* **81** 581–607.
[9] Groemer, H. (1996). *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge University Press, Cambridge, *Encyclopedia of Mathematics and its Applications*.
[10] Healy, D. M., and Kim, P. T. (1996). An empirical bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.* **24** 232–254.
[11] Ichimura, H., and Thompson, T. S. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *J. Econometrics* **86** 269–295.
[12] Kamzolov, A. I. (1983). The best approximation of the class of functions $\mathbf{W}_p^\alpha(\mathbb{S}^n)$ by polynomials in spherical harmonics. *Mat. Zametki* **32** 285–293.
[13] Kim, P. and Koo, J. (2000). Directional mixture models and optimal esimation of the mixing density. *Canad. J. Statist.* **28** 383–398.
[14] Kerkyacharian, G., Kyriazis, G., Le Pennec, E., Petrushev, P., and Picard, D. (2010). Inversion of noisy radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.* **28** 24–45.
[15] Kerkyacharian, G., Petrushev, P., Picard D., and Willer, T. (2007). Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.* **1** 30–76.
[16] Kerkyacharian, G., Phan Ngoc, T. M., and Picard, D. (2009). Localized deconvolution on the sphere. *Ann. Statist.* **39** 1042–1068.
[17] Korostelev, V. and Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction*. Springer.
[18] Mc Fadden, D. (2001). Economic choices - Nobel Lecture, December 2000. *American Economic Review* **91** 351–378.
[19] Mardia, K. V. and Jupp, E. (2000). *Directional Statistics*. Wiley.
[20] Maurer, A. and Pontil, M. (2009). Empirical Bernstein bounds and sample variance penalization. *COLT 2009*.
[21] Narcowich, F., Petrushev, P., and Ward, J. (2006). Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564.
[22] Rubin, B. (1999). Inversion and characterization of the hemispherical transform. *J. Anal. Math.* **77** 105–128.
[23] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer.

Toulouse School of Economics, Université Toulouse Capitole, 21 allée de Brienne, 31015 Toulouse, France.
E-mail address: `Eric.Gautier@tse-fr.eu`

CMAP / Département de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau, France.
E-mail address: `Erwan.Le-Pennec@polytechnique.edu`