NONPARAMETRIC ESTIMATION IN RANDOM COEFFICIENTS BINARY CHOICE MODELS

ERIC GAUTIER AND YUICHI KITAMURA

Preliminary Draft

Abstract. This paper considers random coefficients binary choice models. The main goal is to estimate the density of the random coefficients nonparametrically. This is an ill-posed inverse problem characterized by an integral transform. A new density estimator for the random coefficients is developed, utilizing Fourier series expansions on spheres. This approach offers a clear insight on the identification problem. More importantly, it leads to a closed form estimator formula. This allows a simple plug-in procedure that requires no numerical optimization. The new estimator, therefore, is easy to implement in empirical applications, while being flexible about the treatment of unobserved heterogeneity. Extensions including treatments of non-random coefficients and models with endogeneity are discussed.

1. Introduction

Consider a binary choice model

$$Y = \mathbb{I}\left\{X'\beta \geq 0\right\} \tag{1.1}$$

where $\mathbb{I}$ denotes the indicator function and X is a d-vector of covariates. We assume that the first element of X is 1, the vector X is thus of the form $X = (1, \tilde{X}')'$. The vector β is random. The random vector $(Y, \tilde{X}, \beta)$ is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $(y_i, \tilde{x}_i, \beta_i), i = 1, ..., N$ denote its realizations. The econometrician observes $(y_i, \tilde{x}_i), i = 1, ..., N$, but $\beta_i, i = 1, ..., N$ remain unobserved. Therefore $\tilde{X}$ and β correspond to observed and unobserved heterogeneity across agents, respectively. Note that the first element of β in this formulation absorbs the usual scalar stochastic shock term as well as a constant in standard binary choice with non-random coefficients. This formulation is used

Date: This Version: November 5, 2007.

Keywords: Inverse problems, Discrete Choice Models.

JEL Classification Number: C14.

We thank seminar participants at Harvard/MIT, Simon Fraser, Tilburg University and the Tinbergen Institute for helpful comments. Kitamura acknowledges financial support from the National Science Foundation via grants SES-0241770 and SES-0551271.

in Ichimura and Thompson (1993), and is convenient for the subsequent development in the paper. Throughout the article we assume

Assumption 1.1. *β is independent of $\tilde{X}$.*

The choice probability is given by

$$r(x) = \mathbb{P}(Y = 1|X = x) = \mathbb{E}_\beta[\mathbb{I}(x'\beta > 0)]. \tag{1.2}$$

Discrete choice models with random coefficients variation are useful in applied research since it is often crucial to incorporate unobserved heterogeneity in the choice behavior of individuals. There is a vast and active literature on this topic. Recent contributions include Briesch, Chintagunta and Matzkin (1996), Brownstone and Train (1999), Chesher and Santos Silva (2002), Hess, Bolduc and Polak (2005), Harding and Hausman (2006), Athey and Imbens (2007) Bajari, fox and Ryan (2007) and Train (2007). A common approach in estimating random coefficient discrete choice models is to assume parametric specifications. A leading example is the mixed Logit model, which is discussed in details by Train (2003). If one does not impose a parametric distributional assumption, the distribution of β itself is the structural parameter of interest. The goal for the econometrician is then to back out the distribution of β from the information about $r(x)$ obtained from the data.

Nonparametric treatments for unobserved heterogeneity distributions is an important issue in econometrics. Heckman and Singer (1984) study the issue of unobserved heterogeneity distributions in duration models and propose a treatment by NPMLE. Elbers and Ridder (1982) also develop some identification results in such models. Beran and Hall (1992) and Hoderlein, Klemela and Mammen (2007) discuss nonparametric estimation of random coefficients linear regression models. Despite the tremendous importance of random coefficient discrete choice models, as exemplified in the above references, nonparametrics in this area is relatively underdeveloped. An important paper by Ichimura and Thompson (1998) proposes a nonparametric maximum likelihood estimator (NPMLE) for the CDF of β. They present sufficient conditions for identification and prove the consistency of the NPMLE. The NPMLE requires high dimensional numerical maximization and can be computationally intensive even for a moderate sample size.

Here we develop a different approach that shares many similarities with standard deconvolution methods in the Euclidean space. This allows us to revisit the identification issue. Moreover, once sufficient constraints are imposed on the parameter, we are able to estimate the density with a simple closed form formula. This is a simple plug-in procedure that requires no numerical optimization. The new estimator, therefore, is easy to implement in empirical applications, while being flexible about the treatment of unobserved heterogeneity.

Since the scale of β is not identified in the binary choice model, we normalize the scale so that β is a vector of Euclidean norm 1 in $\mathbb{R}^d$. Then β belongs to the $d-1$ dimensional sphere $\mathbb{S}^{d-1}$. This is not a restriction as long as the probability that β is equal to 0 is 0. Also, since only the angle between X and β matters, we replace X by $X/\|X\|$ and assume X is on the sphere. We aim to recover the joint probability density function f_β of β with respect to the spherical measure $d\sigma$ over $\mathbb{S}^{d-1}$ from the N observations $(y_1, x_1), \ldots, (y_N, x_N)$ of (Y, X).

The problem considered here is a linear ill-posed inverse problem. Let f_X denote the density of the covariates x, again with respect $d\sigma$. We can write

$$r(x) = \int_{b\in\mathbb{S}^{d-1}} \mathbb{I}\left\{x'b \geq 0\right\} f_\beta(b)d\sigma(b) = \int_{H(x)} f_\beta(b)d\sigma(b) := \mathcal{T}\left(f_\beta\right)(x) \tag{1.3}$$

where the set $H(x)$ is the hemisphere $\{b : x'b \geq 0\}$. The mapping $\mathcal{T}$ is called the hemispherical transformation and some of its properties are discussed in Groemer (1996). As we shall see, $\mathcal{T}$ is not injective without further restrictions. Thus conditions need to be imposed to ensure identification. Even under an additional condition which guarantees identification, however, the inverse of $\mathcal{T}$ is not a continuous mapping, making the problem ill-posed. To see this, suppose we restrict f_β to be in $\mathrm{L}^2(\mathbb{S}^{d-1})$. Since the kernel is square integrable by compactness of the sphere, the operator is Hilbert-Schmidt and thus compact. Therefore if the inverse of $\mathcal{T}$ were continuous, $\mathcal{T}^{-1}\mathcal{T}$ would map the closed unit ball in $\mathrm{L}^2(\mathbb{S}^{d-1})$ to a compact set. But the Riesz theorem states that the unit ball is relatively compact if and only if the vector space has finite dimension. The fact that $\mathrm{L}^2(\mathbb{S}^{d-1})$ is an infinite dimensional space contradicts this. Therefore the inverse of $\mathcal{T}$ cannot be continuous. In order to overcome this problem, we use a one parameter family of regularized inverses that are continuous and converge to the inverse when the parameter goes to infinity. This is a common approach to ill-posed inverse problems in statistics (see, e.g. Carrasco et al., 2007).

Due to the particular form of the kernel of the operator $\mathcal{T}$ involving the scalar product $x'b$, we are able to show that the operator is an analogue of convolution in $\mathbb{R}^d$. This analogy provides a clear insight into the identification issue. It is also useful in deriving an estimator based on a series expansion on the Fourier basis or its extension to higher dimensional spheres. These bases are defined via the Laplacian on the sphere, and they diagonalize the operator $\mathcal{T}$ on $\mathrm{L}^2\left(\mathbb{S}^{d-1}\right)$. Such techniques are used in Healy and Kim (1996) for empirical Bayes estimation in the case of the sphere $\mathbb{S}^2$. The kernel of our integral operator $\mathcal{T}$, however, does not satisfy the assumptions made by Healy and Kim. Moreover, our approach utilizes the so-called "condensed" expressions. The approach replaces a full expansion on a Fourier basis by an expansion in terms of the projections on the finite dimensional eigenspaces of the Laplacian on the sphere. This is useful since an explicit expression of the projector is available. It allows us to work in any dimension and does not require a parametrization by hyperspherical coordinates nor the actual knowledge of an orthonormal basis. This approach, to the best of our knowledge, appears to be new in the econometrics literature.

The paper is organized as follows. In Section 2 we introduce a toy model and the tools from harmonic analysis that are used for the development of our estimation procedure and its asymptotic analysis. We also present a new estimator for densities on a hypersphere. This, in turn, is potentially useful for implementing our procedure for random coefficient binary choice models. Section3 presents the main results of this article. It deals with both the identification and the estimation of the density of the random coefficient. Extensions such as estimation of marginals, models with non-random coefficients, treatment of endogeneity and multinomial discrete choice models are presented in Section 4.

2. Preliminaries

In this section we introduce some tools that are used to relate the estimation of the density of β to a deconvolution problem. Useful geometric concepts are postponed to the appendix. We first study the case where X is of dimension 2 to gain basic insights. We parameterize the vector $b = (b_1, b_2)'$ of $\mathbb{S}^1$ by the angle $\phi = \arccos(b_1)$ in $[0, 2\pi)$. As it is often the case when standard Fourier series techniques are used, we consider spaces of complex valued functions. Let $\mathrm{L}^p(\mathbb{S}^1)$ denote the Banach spaces of Lebesgue p-integrable functions and its norm by $\|\cdot\|_p$. In the case of $\mathrm{L}^2(\mathbb{S}^1)$, the norm is derived from the hermitian product $\int_0^{2\pi} f(\theta)\overline{g(\theta)}d\theta$. With the parametrization by angles we obtain

$$\mathcal{T}(f_\beta)(\theta) = \int_0^{2\pi} \mathbb{I}\{|\theta - \phi| < \pi/2\} f_\beta(\phi)d\phi. \tag{2.1}$$

This expression suggests that the hemispherical transformation is a usual convolution of functions on $\mathbb{R}/(2\pi\mathbb{Z})$. Rewrite (2.1) as

$$\frac{\mathcal{T}(f_\beta)}{\pi}(\theta) = \int_0^{2\pi} \left(\frac{1}{\pi}\mathbb{I}\{|\theta - \phi| < \pi/2\}\right) f_\beta(\phi)d\phi. \tag{2.2}$$

It is then possible to link estimation of f_β with statistical deconvolution problems. $\mathcal{T}(f_\beta)/\pi$ is then interpreted as the density of θ, which is generated by adding (on $\mathbb{R}/(2\pi\mathbb{Z})$) a "noise" drawn from the uniform density $\frac{1}{\pi}\mathbb{I}\{|x|<\pi/2\}$ to the "signal" ϕ drawn from f_β. Differentiating the right hand-side of expression (2.1) we obtain $f_\beta(\theta+\pi/2)-f_\beta(\theta-\pi/2)$ where f_β is defined on the line by periodicity. Under an assumption such that f_β is supported on a hemisphere, this assumption is discussed further in Section 3.1, we obtain either $f_\beta(\theta+\pi/2)$ or $-f_\beta(\theta-\pi/2)$. When the model is identified properly the inverse is a differential operator and as such unbounded. It is typically the case that the inverse of kernel operator is a differential operator but, in order to generalize the inversion to any dimension, we prefer to use an approach based on Fourier series and their generalizations to higher dimensional spheres.

Fourier series is a useful tool for deconvolution problems on the circle. Either $(\exp(-int)/(2\pi))_{n\in\mathbb{Z}}$ or the real valued functions $(\cos(nt)/(\sqrt{\pi}),\sin(nt)/(\sqrt{\pi}))_{n\in\mathbb{N}}$ can be used as the basis of $\mathrm{L}^2(\mathbb{S}^1)$ for Fourier series. Denoting by $c_n(f)=\int_0^{2\pi}f(t)\exp(-int)dt/(2\pi)$ the Fourier coefficients of $f\in\mathrm{L}^2(\mathbb{S}^1)$

$$f_\beta(\theta)=\sum_{n\in\mathbb{Z}}c_n(f_\beta)\exp(in\theta) \tag{2.3}$$

in the $\mathrm{L}^2(\mathbb{S}^1)$ sense. Recall also that for f and g in $\mathrm{L}^1(\mathbb{S}^1)$,

$$c_n(f*g)=2\pi c_n(f)c_n(g). \tag{2.4}$$

Using equation (2.4) we obtain the following proposition.

Proposition 2.1. *$c_0(\mathcal{T}(f_\beta))=\pi c_0(f_\beta)$ and for $n\in\mathbb{Z}\setminus\{0\}$, $c_n(\mathcal{T}(f_\beta))=c_n(f_\beta)2\sin(n\pi/2)/n$.*

As in classical deconvolution problems on the real line, our aim is to obtain f_β using equation (2.3) and Proposition 2.1. Notice that among the Fourier coefficients $c_n(f_\beta), n=1,2,...$ it is only possible to recover the first coefficient $c_0(f_\beta)$ (which is easily seen to be $1/2\pi$, by integrating both sides of (2.1) and noting that f_β is a probability density function) and the odd coefficients. Indeed, Proposition 2.1 shows that $c_{2p}(\mathcal{T}(f_\beta))=0$ holds for all non-zero p's, regardless of the value of $c_{2p}(f_\beta)$. In other words, any f_β with the same coefficient $c_0(f_\beta)$ and odd coefficients gives rise to the same hemispherical transformation. Variations in r do not allow to identify the coefficients $c_{2p}(f_\beta)$ for a non zero p. The same phenomenon occurs in higher dimensions, as shown below.

Remark 2.1. If we make the stronger assumption that f_β belongs to $\mathrm{L}^2(\mathbb{S}^1)$, we may interpret this result in terms of operators. This view is useful for the analysis of higher dimensional cases, which is our main concern. For $n\neq 0$ the vector spaces $H^{n,2}=\mathrm{span}\{\exp(int)/(2\pi),\exp(-int)/(2\pi)\}=\mathrm{span}\left\{\cos(nt)/(\sqrt{2}\pi),\sin(nt)/(\sqrt{2}\pi)\right\}$ are eigenspaces of the compact self-adjoint operator $\mathcal{T}(f_\beta)$.

These eigenspaces are associated with the eigenvalues $2\sin(n\pi/2)/n$. Also, $\overline{\bigoplus_{p\in\mathbb{Z}} H^{2p,2}}$ is the null space ker $\mathcal{T}$ of $\mathcal{T}$. The eigenvalue π is simple and associated with the eigenvector $1/2\pi$.

Before turning to the general case where $d \geq 2$, let us introduce some concepts. We consider functions defined on the sphere $\mathbb{S}^{d-1}$, which is a $d-1$ dimensional smooth submanifold of $\mathbb{R}^d$. The canonical measure on $\mathbb{S}^{d-1}$ (or spherical measure) is denoted by $d\sigma$ and is such that $\int_{\mathbb{S}^{d-1}} d\sigma = |\mathbb{S}^{d-1}|$ which is the area of the sphere; see, for example, Gallot et al (2004). It is given for $d \geq 1$ by $\left|\mathbb{S}^{d-1}\right| = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ where Γ is the usual Gamma function, with $\left|\mathbb{S}^{0}\right|$ being 2. When $d \geq 2$ we consider spaces of real valued functions defined on the sphere. The norm of $\mathrm{L}^2(\mathbb{S}^{d-1})$ is derived from the scalar product $(f,g)_{\mathrm{L}^2(\mathbb{S}^{d-1})} = \int_{\mathbb{S}^{d-1}} f(x)g(x)d\sigma(x)$ for $f,g \in \mathrm{L}^2(\mathbb{S}^{d-1})$. Finally, the Laplacian Δ^S on the sphere proves to be important as well. It allows to extend the Fourier basis to any dimension. We can indeed check that the functions $\sin(nt)/(\sqrt{2}\pi)$ and $\cos(nt)/(\sqrt{2}\pi)$ are eigenfunctions of $-\frac{\mathrm{d}}{\mathrm{d}t^2}$ associated with the eigenvalue n^2. Let Δ denote the usual Laplacian in $\mathbb{R}^d$. Let the notation $\check{f}$ signify the radial extension of f, that is, $\check{f}(x) = f(x/\|x\|)$. Similarly, let $\hat{f}$ denote the restriction of the function f to $\mathbb{S}^{d-1}$. We define Δ^S according to:

$$\Delta^S f = (\Delta \check{f})\hat{}. \tag{2.5}$$

Similarly the gradient on the sphere ∇^S is related to the gradient in $\mathbb{R}^d$ through the formula

$$\nabla^S f = (\nabla \check{f})\hat{}. \tag{2.6}$$

These simple definitions match the classical definitions that hold for any Riemannian manifold. See the appendix for a brief exposition of some notions from Riemannian geometry that are useful for our purpose.

Definition 2.1. A surface harmonic of degree n is the restriction to $\mathbb{S}^{d-1}$ of a homogeneous harmonic polynomial of degree n in $\mathbb{R}^d$.

The reader is referred to Groemer (1996) for clear and detailed expositions on these concepts and important results concerning spherical harmonics used in this paper. Erdélyi et al. (1953, vol. 2 chapter 9) provide detailed accounts focusing on special functions. The proofs and results below can be found in the above references.

Lemma 2.1. *The following properties hold:*

(*i*) $-\Delta^S$ *is a positive self-adjoint unbounded operator on* $\mathrm{L}^2(\mathbb{S}^{d-1})$*, thus it has orthogonal eigenspaces and a basis of eigenfunctions;*

(*ii*) *Surface harmonics of degree* n *are eigenfunctions of* $-\Delta^S$ *associated with the eigenvalue* $n(n+d-2)$*;*

(*iii*) *The dimension of the vector space* $H^{n,d}$ *of spherical harmonics of degree* n *is*

$$\dim H^{n,d} = \frac{(2n+d-2)(n+d-2)!}{n!(d-2)!(n+d-2)};$$

(*iv*) *A system formed of the collection of orthonormal bases of* $H^{n,d}$ *for each degree* $n = 0, \ldots, \infty$ *is complete.*

Notation. We let $h(n,d)$ denote $\dim H^{n,d}$. □

Lemma 2.1 (i) and (iv) give the decomposition

$$L^2(\mathbb{S}^{d-1}) = \bigoplus_{n \in \mathbb{N}} H^{n,d}$$

with orthogonal $H^{n,d}$'s. The $H^{n,d}$ are the eigenspaces of Δ^S. The space of surface harmonics of degree 0 is the one dimensional space spanned by 1. The real Fourier basis on the line is an orthonormal basis of surface harmonics and the vector spaces of surface harmonics of degree n are of dimension 2. A series expansion on an orthonormal basis of surface harmonics is thus also called a Fourier series expansion.

Orthonormal basis of surface harmonics could be obtained via the Schmidt orthonormalization procedure. These usually involve parametrization by angles, such as the spherical coordinates when $d = 3$ as used by Healy and Kim (1996) or hyperspherical coordinates in higher dimensions. In contrast, here we work with the decomposition of a function on the spaces $H^{n,d}$. This leads to a simple method both in terms of theoretical developments and practical implementations. Note that the projector $Q_{n,d}$ on $H^{n,d}$ in $L^2(\mathbb{S}^{d-1})$ can be expressed as an integral operator with kernel

$$q_{n,d}(x,y) = \sum_{l=1}^{h(n,d)} S_n^l(x) S_n^l(y), \tag{2.7}$$

where $(S_n^l)_{l=1}^{h(n,d)}$ is any basis of $H^{n,d}$. This kernel has another simple and useful expression as we shall see below.

Definition 2.2. The condensed harmonic expansion of a function f in $L^2(\mathbb{S}^{d-1})$ is the series expansion $\sum_{n=0}^{\infty} Q_{n,d} f$.

The odd and even part of a function defined on the sphere are important notions in the development of our analysis.

Definition 2.3. We difine the odd part and the even part of a function f by:

$$f^{-}(b) = (f(b) - f(-b))/2$$

and

$$f^{+}(b) = (f(b) + f(-b))/2,$$

for every b in $\mathbb{S}^{d-1}$

If the function f is in $\mathrm{L}^2(\mathbb{S}^{d-1})$ then using equations (2.10) and (6.10) we can check that for p nonnegative $Q_{2p,d}f(x) = Q_{2p,d}f(-x)$ and $Q_{2p+1,d}f(x) = -Q_{2p+1,d}f(-x)$. Thus the sum of the odd terms in the condensed harmonic expansion corresponds to f^- and the sum of the even terms corresponds to f^+. If a positive function f has its support included in some hemisphere then

$$\begin{aligned} f(x) &= 2f^{-}(x)\mathbb{I}\{x \in \mathrm{supp}f\} \\ &= 2f^{-}(x)\mathbb{I}\left\{f^{-}(x) > 0\right\} \end{aligned} \tag{2.8}$$

where we denote by $\mathrm{supp}f$ the support of f. This follows from the fact that $f^-(x) = f^+(x) \geq 0$ on $\mathrm{supp}f$ while $f^-(x) = -f^+(x) \leq 0$ on $-\mathrm{supp}f$ and both f^- and f^+ are 0 on $\mathbb{S}^{d-1} \setminus (\mathrm{supp}f \bigcup -\mathrm{supp}f)$. If f is a probability density function, the coefficient of degree 0 in the expansion of f on surface harmonics is $1/|\mathbb{S}^{d-1}|$.

Remark 2.2. Reciprocally, any harmonic polynomial or series such that the degree 0 coefficient is $1/|\mathbb{S}^{d-1}|$ integrates to one. Thus, truncation used below as a regularization procedure, preserves the probability mass. On the other hand, non-negativity on the whole sphere of a truncated expansion is not guaranteed. When $d = 2$, non-negativity is equivalent to positive definitiveness of the sequence of coefficients of the Fourier series expansion by the Herglotz theorem, see Katznelson (2004, p. 41). This is clearly not preserved by truncation.

Lemma 2.1 (ii) allows to define the Sobolev spaces based on $\mathrm{L}^2(\mathbb{S}^{d-1})$. A relation with the usual definition in terms of derivatives rather than Fourier series is given in the appendix.

Definition 2.4. The Sobolev space $\mathrm{H}^s(\mathbb{S}^{d-1})$ for s positive is the space of functions f in $\mathrm{L}^2(\mathbb{S}^{d-1})$ such that $\sum_{n=0}^{\infty}(n(n+d-2))^s\|Q_{n,d}f\|_2^2 < \infty$ where the $Q_{n,d}f$ are the factors of the condensed harmonic expansion of f.

Equipped with the norm

$$\|f\|_{2,s}^2 = \sum_{n=0}^{\infty} (1 + n(n+d-2))^s \|Q_{n,d} f\|_2^2,$$

$\mathrm{H}^s(\mathbb{S}^{d-1})$ is complete. It is also a Hilbert space for the scalar product related to the norm by polarization. We use these spaces to make smoothness assumptions. Working in the Sobolev spaces instead of spaces like $\mathrm{L}^p(\mathbb{S}^{d-1})$ or the space of continuous functions is convenient since Hilbert spaces techniques and Sobolev embeddings are available. The next result establishes a continuous embedding that is used in this article.

Proposition 2.2. *When $s > (d-1)/2$, $\mathrm{H}^s(\mathbb{S}^{d-1})$ is continuously embedded in the space of bounded continuous functions.*

A proof of the proposition is provided in the appendix. It means that $\mathrm{H}^s(\mathbb{S}^{d-1})$ is a subset of the space of bounded continuous functions for such s and that the injection in the bigger space, which is a linear mapping, is continuous. In other words there exists a constant C such that for every $f \in \mathrm{H}^s(\mathbb{S}^{d-1})$, $\|f\|_\infty \leq C\|f\|_{2,s}$. A similar inequality holds for the more familiar Euclidean space $\mathbb{R}^d$, with the same condition on the exponent s. A consequence of Proposition 2.2 is that consistency in $\mathrm{H}^s(\mathbb{S}^{d-1})$ for $s > (d-1)/2$ implies consistency in sup norm $\|\cdot\|_\infty$.

The next theorem gives an explicit formula for the kernels $q_{n,d}$ in terms of the Gegenbauer polynomials C_n^ν, see Erdélyi (1953, vol. 1 p. 175-179) and the appendix for the properties used in this article. These polynomials are defined for $\nu > 1/2$ and are orthogonal with respect to the weight function $(1-t^2)^{\nu-1/2}dt$ on $[-1,1]$. They correspond to the well known Legendre polynomials when $d = 3$. Note that $C_0^\nu(t) = 1$ and $C_1^\nu(t) = 2\nu t$ for $\nu \neq 0$ while $C_1^0(t) = 2t$. Moreover, they satisfy the following recursion relation

$$(n+2)C_{n+2}^\nu(t) = 2(\nu+n+1)tC_{n+1}^\nu(t) - (2\nu+n)C_n^\nu(t). \tag{2.9}$$

In our approach the Gegenbauer polynomials will be evaluated at N points for a series of successive values of the degree n. The recursion relation (2.9) is therefore a powerful tool.

Theorem 2.1 (Addition Formula)**.** *The following identity holds*

$$q_{n,d}(x,y) = \frac{h(n,d)C_n^{\nu(d)}(x'y)}{|\mathbb{S}^{d-1}|C_n^{\nu(d)}(1)} \tag{2.10}$$

where

$$\nu(d) = (d-2)/2.$$

Notation. We denote for every integer n,

$$c(n,d) = \frac{h(n,d)}{C_n^{\nu(d)}(1)} \tag{2.11}$$

it is such that $c(n,2) = n$ and $c(n,d) = \frac{2n+d-2}{d-2}$ for $d \geq 3$. □

It is interesting to note that the infinite sum of functions $\sum_{n=0}^{T} q_{n,d}(x,y)$ converges in the sense of distributions as T goes to infinity to the Dirac measure δ_x. The integral operator

$$\varphi \in \mathrm{L}^2(\mathbb{S}^{d-1}) \mapsto \int_{\mathbb{S}^{d-1}} \sum_{n=0}^{T} q_{n,d}(x,y)\varphi(y)d\sigma(y)$$

is the analogue of a kernel operator. Theorem 2.1 is later used to obtain closed form estimates in any dimensions in the nonparametric estimation procedures.

The next theorem is an important result which shows that Fourier series on spheres is a very natural tool for deconvolution purposes.

Theorem 2.2 (Funk-Hecke Theorem). *If g belongs to $H^{n,d}$ for some n and F is such that*

$$\int_{-1}^{1} |F(t)|^2(1-t^2)^{(d-3)/2}dt < \infty,$$

then

$$\int_{\mathbb{S}^{d-1}} F(x'y)g(y)d\sigma(y) = \lambda_n(F)g(x) \tag{2.12}$$

where

$$\lambda_n(F) = \frac{|\mathbb{S}^{d-2}|}{C_n^{\nu(d)}(1)} \int_{-1}^{1} F(t)C_n^{\nu(d)}(t)(1-t^2)^{(d-3)/2}.$$

In other words, the kernel operator K defined by

$$f \in \mathrm{L}^2(\mathbb{S}^{d-1}) \mapsto \left(x \mapsto \int_{\mathbb{S}^{d-1}} F(x'y)f(y)d\sigma(y)\right) \in \mathrm{L}^2(\mathbb{S}^{d-1})$$

is, when restricted to a subspace $H^{n,d}$, the multiplication by $\lambda_n(F)$. Thus a basis of surface harmonics diagonalizes any appropriate integral operator defined through a kernel function of the scalar product $x'y$.

Remark 2.3. Healy and Kim (1996) use an approach related to ours to analyze a deconvolution problem on the sphere in dimension $d = 3$. As we shall see below, the Addition Formula along with condensed harmonic expansions provide a general treatment that works for cases with arbitrary dimension.

Notation. We define $\lambda(n,d) = \lambda_n\left(\mathbb{I}\left\{t \in [0,1]\right\}\right)$ for $d \geq 3$ and $\lambda(n,2) = \frac{2\sin(n\pi/2)}{n}$ of Proposition 2.1. □

Proposition 2.3. *When $d \geq 2$, the coefficients $\lambda(n,d)$ have the following expression*

(*i*) $\lambda(0,d) = \frac{2}{|\mathbb{S}^{d-1}|}$

(*ii*) $\forall p > 0,\ \lambda(2p,d) = 0$

(*iii*) $\forall p \geq 0,\ \lambda(2p+1,d) = \frac{(-1)^p|\mathbb{S}^{d-2}|1\cdot 3\cdots(2p-1)}{(d-1)(d+1)\cdots(d+2p-1)}$.

The proof is given in the appendix.

Remark 2.4. These eigenvalues of the hemispherical transformation take alternatively positive and negative signs: the operator is not a positive operator. $\mathcal{T}^*\mathcal{T} = \mathcal{T}^2$ is however compact, self-adjoint and positive and the Spectral Theorem applies. We can easily verify that the sequence of eigenvalues converges to zero, which is also implied by the compactness of $\mathcal{T}^2$.

The following corollary corresponds to an observation made in Remark 2.1 for the $d = 2$ case.

Corollary 2.1. *The null space of $\mathcal{T}$ seen as an operator on $\mathrm{L}^2(\mathbb{S}^{d-1})$ is*

$$\ker\mathcal{T} = \overline{\bigoplus_{p=1}^{\infty} H^{2p,d}}.$$

The spaces $H^{0,d}$ and $H^{2p+1,d}$ for p non negative are the eigenspaces associated with non zero eigenvalues.

Remark 2.5. Because of the lack of continuity of the inverse of the operator $\mathcal{T}$, we use a truncation of the series expansion of the inverse. Our approach amounts to the spectral cut-off method used in the statistics of inverse problems.

We use the following notation.

Notation. For two sequences of positive numbers $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we write $a_n \asymp b_n$ when there exists M positive such that $M^{-1}b_n \leq a_n \leq Mb_n$ for every n positive. □

Before closing this section on preliminary materials, let us present some new results on density estimation on the sphere. This provides some theoretical results which are useful for the estimation of the random coefficients density f_β. Also, we propose new estimators for densities on the sphere which in turn can be used to construct a feasible estimator for f_β in the next section.

Estimation of densities on compact manifolds have been studied by several authors. Some authors have considered Fourier series type estimates, see for example Devroye and Gyorfi (1985) for the circle and Hendriks (1990) for general compact Riemannian manifolds. Hendriks (2003) considers fast spherical Fourier transform on the sphere $\mathbb{S}^2$ using the algorithm of Driscoll and Healy (1994). Kernel estimates have also been considered, see Devroye and Gyorfi (1985) for the case of the circle, Hall, Watson and Cabrera (1987) for higher dimensional spheres and Pelletier (2005) for general Riemannian manifolds. We present a new estimate of f_X. It is in the spirit of Hendricks (1990) but, using condensed harmonic expansions together with the Addition Formula, it has a simple closed form in any dimension. Again neither a basis of spherical harmonics nor a hyperspherical parametrization are required. See the appendix for the proofs of the following theorems.

Theorem 2.3. *The density f_X is given by*

$$f_X(x) = \frac{1}{|\mathbb{S}^{d-1}|}\left(1+\sum_{n=1}^{\infty} c(n,d)\mathbb{E}\left[C_n^{\nu(d)}(X'x)\right]\right). \tag{2.13}$$

If the support of X is contained in some hemisphere H, the density f_X is also given by

$$f_X(x) = 2f_X^-(x)\mathbb{I}\{x\in H\}. \tag{2.14}$$

where

$$f_X^-(x) = \frac{1}{|\mathbb{S}^{d-1}|}\sum_{p=0}^{\infty} c(2p+1,d)\mathbb{E}\left[C_{2p+1}^{\nu(d)}(X'x)\right].$$

The specification of our discrete choice model (1.1) implies that $X = (1,\tilde{X})'/\|(1,\tilde{X})'\|$, thus the support of X is automatically a subset of the closed upper hemisphere, denoted by H^+.

We consider the following estimate

$$\hat{f}_X^{N,T}(x) = \frac{1}{|\mathbb{S}^{d-1}|}\left(1+\frac{1}{N}\sum_{i=1}^{N}\sum_{n=1}^{T} c(n,d)C_n^{\nu(d)}(x_i'x)\right) \tag{2.15}$$

as a sample analogue of (2.13). On the other hand, (2.14) suggests the following estimator

$$\hat{f}_X^{N,T}(x) = 2\hat{f}_X^{-N,T}(x)\mathbb{I}\{x\in H\} = \left(\frac{2}{N|\mathbb{S}^{d-1}|}\sum_{i=1}^{N}\sum_{p=0}^{T} c(2p+1,d)C_{2p+1}^{\nu(d)}(x_i'x)\right)\mathbb{I}\left\{x\in H\right\} \tag{2.16}$$

when the support of X is a subset of H. If the support of X is known to be a subset of a hemisphere yet a priori knowledge of the hemisphere is not available, it is still possible to obtain an appropriate estimator using the identity (2.8). See Section 3 for an application of this approach for the estimation of f_β. Note that when X is supported in some hemisphere we could also write f_X can be represented by even terms. The estimate (2.16) seems more convenient from a practical point of view, however, since it involves terms we are later use to estimate f_β.

To study consistency, we consider the mean square error where the loss function is the norm of some Sobolev space $\mathrm{H}^v(\mathbb{S}^{d-1})$. The case where $v=0$ corresponds to the integrated MSE. Using Sobolev embeddings, the result can be used to consider loss functions expressed in $\mathrm{L}^p(\mathbb{S}^{d-1})$ norms and $\mathrm{L}^p(\mathbb{S}^{d-1})$ norms of derivatives for some $2\le p\le p^*(v)$ and p^* a function of the exponent v. In particular, Sobolev embedding is convenient in showing the L_∞ convergence result of our estimator for f_X. This, in turn, is useful for the asymptotic analysis of our estimator for f_β, which requires a preliminary estimator for f_X.

Theorem 2.4. *When f_X belongs to $\mathrm{H}^\sigma(\mathbb{S}^{d-1})$, $T_N \asymp N^{1/(2\sigma+d-1)}$ and v in $[0,\sigma]$, then the estimator (2.15) satisfies*

$$\mathbb{E}\left[\left\|\hat{f}_X^{N,T_N}-f_X\right\|_{2,v}^2\right]=O\left(N^{-\frac{\sigma-v}{\sigma+(d-1)/2}}\right). \tag{2.17}$$

If the support of X is contained in some hemisphere, (2.17) holds for the estimator (2.16). Also, both estimators satisfy

$$\left\|\hat{f}_X^{N,T_N}-f_X\right\|_\infty=o_p\left(N^{-\frac{\sigma-(d-1)/2-\epsilon}{2\sigma+d-1}}\right) \tag{2.18}$$

when $\sigma>(d-1)/2$, for all ϵ in $(0,\sigma-(d-1)/2)$.

These rates are the same as the ones obtained by Hendriks (1990) and the other references mentioned above for estimation of densities on manifolds. It also matches the convergence rate for nonparametric estimation in the conventional space $\mathbb{R}^d$ if d is replaced by the dimension of the manifold, here $d-1$.

Asymptotic normality is obtained under either of the following two scenarios:

(i) f_X is bounded from below by $1/C_X$ on the whole sphere; the estimator (2.15) is used.

(ii) f_X is supported on a hemisphere H and is bounded from below by on H; the estimator (2.16) is used.

Theorem 2.5. *For either* (i) *or* (ii)*, if* $\sigma > (d-1)/2$*,* x *in* $\text{supp} f_X$*, and* $T_N \asymp N^{\alpha}$*,* $\frac{1}{2\sigma} \le \alpha < \frac{1}{d-1}$*, then*

$$N^{\frac{1}{2}} s_N^{-1} \left(\hat{f}_X^{N,T_N}(x) - f_X(x) \right) \xrightarrow{d} N(0,1) \tag{2.19}$$

holds with $s_N^2 := 4\text{var}(Z_{ni})$*,* $Z_{ni} = \sum_{n=0}^{T_N} q_{n,d}(X_i, x)$ *for (i), and* $Z_{ni} = \sum_{p=0}^{T_N} q_{2p+1,d}(X_i, x)$ *for (ii).*

Note that the condition on the rate α amounts to under-smoothing. The variance in the denominator can be estimated by a sample counterpart.

3. Main Results

3.1. Identification in the Random Coefficient Model.

This section analyzes the identifiability of f_β and discusses sufficient conditions for identification. We make the following assumption which also appears in Ichimura and Thompson (1998). It is used to extend the choice probability $r(x)$ to a function on the whole sphere and as a result to identify f_β.

Assumption 3.1. *The support of* f_X *is the whole hemisphere* H^+*.*

This assumption demands that $\tilde{X}$ is supported on the whole space $\mathbb{R}^{d-1}$. It rules out discrete or bounded $\tilde{X}$. (See Section 4 for a potential approach to deal with such regressors as dummy variables.) For simplicity we make the following stronger assumption on f_X.

Assumption 3.2. f_X *is bounded from below on* H^+ *by a constant* $1/C_X$*.*

This assumption can be relaxed using thresholding for X drawn near the boundary of H^+ as in Hoderlein, Klemelä and Mammen (2007) in the case of the linear random coefficients model. We make the following assumption on the smoothness of f_X and f_β.

Assumption 3.3. f_X^- *and* f_β^- *belong to* H^σ *and* H^s*, respectively for* $\sigma > (d-1)/2$ *and* $s > (d-1)/2$*.*

We now consider choice probabilities $r(x)$ given by (1.2) which are invariant by dilatation

$$\forall x \in \mathbb{R}^d, \ \mathbb{P}(Y=1|X=x) = \mathbb{P}(Y=1|X=x/\|x\|).$$

As such they can be studied as function on the sphere. The invariance by dilatation is satisfied in the case of the random coefficient model (1.2). They are not strictly speaking functions on the sphere as they are only defined on the support of X. Under Assumption 3.1 it is possible to extend such functions $r(x)$ to a *bona fide* function on the whole sphere. If we again think that the choice probability is such that model (1.2) holds then, as f_β is a probability density function, we obtain for x in H^+

$$\mathcal{T}(f_\beta)(-x) = \int_{H(-x)} f_\beta(b)d\sigma(b) = 1 - r(x) = 1 - \mathcal{T}(f_\beta)(x). \tag{3.1}$$

We thus consider the extension R such that

$$\forall x \in H^+, R(x) = r(x), \text{ and } \forall x \in -H^+, R(x) = 1 - r(-x) = 1 - R(-x). \tag{3.2}$$

This extension is one possibility among infinitely many alternatives, though it is the only one consistent with the random coefficient model. R is now a function defined on the whole sphere and, provided it is square integrable, it has a condensed harmonic expansion which enables us to obtain the expression in the next theorem. R, and not r, is the quantity we want to invert in order to obtain the density of the random coefficients. Note that R is entirely determined by its odd part. Indeed, we can check that

$$\begin{aligned} R(x) &= R^+(x) + R^-(x) \\ &= \frac{1}{2}\left[R(x) + R(-x)\right] + R^-(x) \\ &= \frac{1}{2}\left[R(x) + (1 - R(x))\right] + R^-(x) \quad \text{by (3.2)} \\ &= \frac{1}{2} + R^-(x). \end{aligned} \tag{3.3}$$

This proves to be important for the identification of the model (1.2).

Theorem 3.1. *When* $\text{infess} f_X(X) > 0$ *in* $\mathrm{L}^\infty(\Omega)$,

$$R(x) = \frac{1}{2} + \frac{1}{|\mathbb{S}^{d-1}|}\sum_{p=0}^{\infty} c(2p+1,d)\mathbb{E}\left[\frac{(2Y-1)C_{2p+1}^{\nu(d)}(X'x)}{f_X(X)}\right] \tag{3.4}$$

.

The assumption on f_X is satisfied under the stronger Assumption 3.2. The following result is a consequence of the Funk-Hecke theorem.

Lemma 3.1. *For all* n *in* $\mathbb{N}$ *and* f *in* $\mathrm{L}^2(\mathbb{S}^{d-1})$, $Q_{n,d}\mathcal{T}(f) = \mathcal{T}(Q_{n,d}f) = \lambda_n\left(\mathbb{I}\left\{t \in [0,1]\right\}\right) Q_{n,d}f$.

Also, we have seen in Theorem 3.1 that

$$R^-(x) = \sum_{p=0}^{\infty} \lambda(2p+1,d)\mathbb{E}\left[\frac{(2Y-1)q_{2p+1,d}(X,x)}{f_X(X)\lambda(2p+1,d)}\right].$$

Therefore, provided that
$$\sum_{p=0}^{\infty}\left\|\mathbb{E}\left[\frac{(2Y-1)q_{2p+1,d}(X,x)}{f_X(X)\lambda(2p+1,d)}\right]\right\|_2^2<\infty,$$
the relation
$$\begin{aligned}R^-(x)&=\sum_{p=0}^{\infty}Q_{2p+1,d}R^-\\&=\sum_{p=0}^{\infty}Q_{2p+1,d}\mathcal{T}(f_\beta^-)\end{aligned}$$
holds with
$$\begin{aligned}f_\beta^-(b)&=\sum_{p=0}^{\infty}\mathbb{E}\left[\frac{(2Y-1)q_{2p+1,d}(X,b)}{f_X(X)\lambda(2p+1,d)}\right]\\&=\frac{1}{|\mathbb{S}^{d-1}|}\sum_{p=0}^{\infty}\frac{c(2p+1,d)}{\lambda(2p+1,d)}\mathbb{E}\left[\frac{(2Y-1)C_{2p+1}^{\nu(d)}(X'b)}{f_X(X)}\right].\end{aligned}\tag{3.5}$$
Therefore $R^-=\mathcal{T}(f_\beta^-)$. Moreover, since $R-1/2$ is odd and $\mathcal{T}\left(\frac{1}{|\mathbb{S}^{d-1}|}\right)=\frac{1}{2}$, we can check that whatever the even function g having $\frac{1}{|\mathbb{S}^{d-1}|}$ as coefficient of degree 0,
$$\mathcal{T}(g)=\frac{1}{2}$$
and thus
$$R=R^-+\frac{1}{2}=\mathcal{T}\left(g+f_\beta^-\right).$$
That is, such $g+f_\beta^-$ solves the inverse problem. Expressions for f_β^- other than can be obtained using the explicit expression of the Gegenbauer polynomials in the appendix. Formulae simpler than (3.5) are available for the case with $d=2$ or 4; in the other cases one can use relation (2.9) to obtain Gegenbauer polynomials.

A consequence of the above result is that only the odd part f_β^- of the density of the random coefficient and the coefficient of degree 0 are identified. An infinite number of functions f_β are compatible with the choice probability $r(x)$ since $g-\frac{1}{|\mathbb{S}^{d-1}|}$ could be arbitrary. A more general result along this line for the hemispherical transform applied to signed measures is given in Groemer (1996, Proposition 3.4.11).

Ichimura and Thompson (1998, Theorem 1) give a sufficient condition for identification of the model (1.1). Their assumption postulates that there exists c on $\mathbb{S}^{d-1}$ such that $\mathbb{P}(c'\beta>0)=1$. This, in our terminology, means that:

Assumption 3.4. *The support of β is a subset of some hemisphere.*

As noted by Ichimura and Thompson (1998) Assumption 3.4 does not seem to be too stringent in economic applications. It is often reasonable to assume that one of the random coefficients, such as a price coefficient, has a known sign. Assumption 3.4 implies the following mapping from f_β^- to f_β developed in (2.8):

$$f_\beta(b) = 2f_\beta^-(b)\mathbb{I}\left\{f_\beta^-(b) > 0\right\}. \tag{3.6}$$

This relation is useful because (i) it shows that Assumption 3.4 guarantees identification if f_β^- is identified, (ii) it enables us to derive a key formula that leads to a simple and practical estimation algorithm and (iii) it guaranties that f_β is nonnegative.

Remark 3.1. Assumption 3.4 is testable since it yields implications in terms of f_β^- which is identified under weak conditions. For example, we can compare the positivity of f_β^- with its negativity on the corresponding points on the opposite side of the sphere. Or, it implies that f_β^- integrates to $1/(2|\mathbb{S}^{d-1}|)$ on H and $-1/(2|\mathbb{S}^{d-1}|)$ on $-H$. An estimator for f_β^- and its asymptotic properties are presented in the next section.

3.2. **Nonparametric Estimator for f_β and Its Asymptotic Properties.** As noted earlier the relation (3.6) holds under Assumption 3.4. This suggests the following form as an estimator for f_β:

$$\hat{f}_\beta(b) = 2\hat{f}_\beta(b)\mathbb{I}\left\{\hat{f}_\beta(b) > 0\right\} \tag{3.7}$$

with $\hat{f}_\beta$ being an estimator for f_β^-. The expression (3.5) offers a simple way to estimate f_β^-, though some consideration is in order. Note that the expectations in (3.5) need to be estimated empirically, which are divided by $\lambda(2p+1,d), p = 0,1,....$ The factor $\lambda(2p+1,d)$ tends to zero as p increases, as a manifestation of the ill-posed nature of our estimation problem. We need a regularization procedure to deal with this. Our approach is to truncate the infinite sums to achieve appropriate convergence properties. The unknown f_X can be replaced by a nonparametric estimator $\hat{f}_X$ of the density of f_X. Those considerations suggest

$$\hat{f}_\beta^{-,N,T}(b) = \frac{1}{N|\mathbb{S}^{d-1}|}\sum_{i=1}^{N}\frac{2y_i - 1}{\hat{f}_X(x_i)}\sum_{p=0}^{T}\frac{c(2p+1,d)}{\lambda(2p+1,d)}C_{2p+1}^{\nu(d)}(x_i'b) \tag{3.8}$$

as the estimator to be used in (3.7). Here we propose to use the estimator $\hat{f}_X^{N,T}$ developed in Section 2 as $\hat{f}_X$. Our nonparametric estimator for f_β is therefore:

$$\hat{f}_\beta^{N,T}(b) = 2\hat{f}_\beta^{-,N,T}(b)\mathbb{I}\left\{\hat{f}_\beta^{-,N,T}(b) > 0\right\}. \tag{3.9}$$

The proof of the following result is given in the appendix.

Theorem 3.2. *Under Assumptions 1.1, 3.2 and 3.3, if $s > (d-1)/2$, $\sigma \geq s + \frac{d}{2}$ and T_N satisfies $T_N \asymp N^{1/(2s+2d-1)}$,*

$$\left\| \hat{f}_{\beta}^{N,T_N} - f_{\beta} \right\|_2 = O_p\left(N^{-\frac{s}{2s+2d-1}} \right).$$

Our rate is in accordance with the rate in Healy and Kim (1996), who study deconvolution on $\mathbb{S}^2$ for non degenerate kernels. Kim and Koo (2000) prove that the rate in Healy and Kim (1996) is optimal in a minimax sense. Hoderlein, Klemelä and Mammen (2007) study estimation of densities in a linear model with random coefficients.

Let us now consider asymptotic normality.

Theorem 3.3. *Under the assumptions of Theorem 3.2, if $T_N \asymp N^{\alpha}$, $\left(\frac{1}{2s+d} \vee \frac{2}{2\sigma+d-1}\right) < \alpha < \frac{1}{2d-1}$, then*

$$N^{\frac{1}{2}} s_N^{-1} \left(\hat{f}_{\beta}^{N,T_N}(b) - f_{\beta}(b) \right) \xrightarrow{d} N(0,1) \tag{3.10}$$

holds for $f_{\beta}(b) \neq 0$, where $s_N^2 := 4\text{var}(Z_{ni})$, $Z_{ni} = \frac{2Y_i - 1}{f_X(X_i)} \sum_{p=0}^{T_N} \frac{q_{2p+1,d}(X_i, b)}{\lambda(2p+1,d)}$.

As in the case of the estimation of f_X, the condition imposed for α corresponds to undersmoothing, therefore no bias term is present in the above result.

4. Discussion

4.1. **Estimation of Marginals.** In Section 3 we have provided an expression for the estimate of the full joint density of β, from which an estimator for a marginal density can be obtained. Let $d\sigma_k$ denote the surface measure and $d\underline{\sigma}_k = d\sigma_k / |\mathbb{S}^k|$ the uniform measure on $\mathbb{S}^k$. We write $\beta = \left(\overline{\beta}', \overline{\overline{\beta}}' \right)'$ and wish to obtain the density of the marginal of $\overline{\beta}$ which is a vector of dimension $d-k$. We also define $\overline{P}$ and $\overline{\overline{P}}$ the projectors such that $\overline{\beta} = \overline{P}\beta$ and $\overline{\overline{\beta}} = \overline{\overline{P}}\beta$ and denote by $d\overline{P}_* \underline{\sigma}_{d-1}$ and $d\overline{\overline{P}}_* \underline{\sigma}_{d-1}$ the direct image probability measures. One possibility is to define the marginal law of $\overline{\overline{\beta}}$ as the measure $\overline{\overline{P}}_* f_{\beta} d\sigma$. This may not be convenient, however, since then a uniform distribution would have U-shaped marginals. The U-shape becomes more pronounced as the dimension of β increases. In order to obtain a flat density for the marginals of the uniform joint distribution on the sphere it is enough to consider densities with respect to the dominating measure $d\overline{\overline{P}}_* \underline{\sigma}_{d-1}$. Notice that sampling U uniformly on $\mathbb{S}^{d-1}$ is equivalent to sampling $\overline{\overline{U}}$ according to $\overline{\overline{P}}_* \underline{\sigma}_{d-1}$ and then given $\overline{\overline{U}}$ forming $\rho\left(\overline{\overline{U}}\right) V$ where V is a draw from the uniform distribution $\underline{\sigma}_{d-1-k}$ on $\mathbb{S}^{d-1-k}$ and $\rho\left(\overline{\overline{U}}\right) = \sqrt{1 - \left\|\overline{\overline{U}}\right\|^2}$. Indeed given $\overline{\overline{U}}$,

$\overline{U}/\rho\left(\overline{\overline{U}}\right)$ is uniformly distributed on $\mathbb{S}^{d-1-k}$. Thus, when g is an element of $\mathrm{L}^1(\mathbb{S}^{d-1})$ we can write for k in $\{1,\dots,d-1\}$,

$$\int_{\mathbb{S}^{d-1}} g(b)d\underline{\sigma}_{d-1}(b) = \int_{\mathbb{B}^k}\left[\int_{\mathbb{S}^{d-1-k}} g\left(\rho\left(\overline{\overline{b}}\right)u,\overline{\overline{b}}\right)d\underline{\sigma}_{d-1-k}(u)\right]d\overline{\overline{P}}_*\underline{\sigma}_{d-1}\left(\overline{\overline{b}}\right) \tag{4.1}$$

where $\mathbb{B}^k$ is the k dimensional ball of radius 1. Setting $g=|\mathbb{S}^{d-1}|f_\beta(b)\mathbb{I}\left\{\overline{\overline{b}}\in A\right\}$ for A Borel set of $\mathbb{B}^k$ shows that the marginal density of $\overline{\overline{\beta}}$ with respect to the dominating measure $d\overline{\overline{P}}_*\underline{\sigma}_{d-1}$ is given by

$$f_{\overline{\overline{\beta}}}\left(\overline{\overline{b}}\right) = |\mathbb{S}^{d-1}|\int_{\mathbb{S}^{d-1-k}} f_\beta\left(\rho\left(\overline{\overline{b}}\right)u,\overline{\overline{b}}\right)d\underline{\sigma}_{d-1-k}(u). \tag{4.2}$$

In the particular case where $k=d-1$, *i.e.* we are interested in the marginal of $\tilde{\beta}$, we use $d\underline{\sigma}_0 = (\delta_1+\delta_{-1})/2$ where δ denotes the Dirac mass.

When the dimension of the variables in the integral is small we can use hyperspherical parametrization (polar coordinates when $k=d-2$ and spherical coordinates when $k=d-3$) and deterministic numerical integration methods. When it is not, one may use Monte-Carlo methods, by forming

$$\hat{f}_{\overline{\overline{\beta}}}^{N,T,M}\left(\overline{\overline{b}}\right) = \frac{1}{M}\sum_{j=1}^{M}\hat{f}_\beta^{N,T}\left(\rho\left(\overline{\overline{b}}\right)u_j,\overline{\overline{b}}\right) \tag{4.3}$$

where u_j are draws from independent uniform random variables on $\mathbb{S}^{d-1-k}$. Draws u_j could be obtained by computing $u_j = v_j/\|v_j\|$ where v_j are draws from a standard Gaussian random vector of dimension $d-1-k$. When $\overline{\overline{\beta}}$ is of dimension 2 we could draw contour plots on the disk, that is, level sets of the density. When β is of dimension 3 it is possible to draw contour plots on $\mathbb{S}^2$.

4.2. **Treatment of non-random coefficients.** It may be useful to develop an extension of the method described in the previous sections to models that have non-random coefficients, at least for two reasons. First, the convergence rate of our estimator of the joint density of β slows down as the dimension d of β grows, which is a manifestation of the curse of dimensionality. Treating some coefficient as fixed parameters alleviates this problem. Second, our identification assumption in Section 3.1 precludes covariates with discrete or bounded support. This may not be desirable as many random coefficient discrete choice models in economics involve dummy variables as covariates. The following identification/estimation strategy allows such covariates as far as their coefficients are non-random. Note that Hoderlein, Klemelä and Mammen (2007) suggest a method to deal with non-random coefficients in their treatment of random coefficient linear regression models. Identification in random coefficient binary choice models with covariates with limited support is somewhat tricky. As we shall see shortly, identification is possible in a model where the coefficients on covariates with limited support are non-random, provided that at least one of the covariates with "large support" has a non-random coefficient as well. More precisely, consider the model:

$$y_i = \mathbb{I}\{\beta_{1i} + \beta_{2i}' x_{2i} + \alpha_1 z_{1i} + \alpha_2' z_{2i} \geq 0\} \tag{4.4}$$

where $\beta_1 \in \mathbb{R}$ and $\beta_2 \in \mathbb{R}^{d_X - 1}$ are random coefficients, whereas the coefficients $\alpha_1 \in \mathbb{R}$ and $\alpha_2 \in \mathbb{R}^{d_Z - 1}$ are nonrandom. The covariate vector $(z_1, z_2')'$ is in $\mathbb{R}^{d_Z}$, though the $(d_Z - 1)$-subvector z_2 might have limited support: for example, it can be a vector of dummies. The covariate vector $(x_2', z_1)'$ is assumed to be, among other things, continuously distributed. Normalizing the coefficients vector and the vector of covariates to be elements of the unit sphere works well for the development of our procedure, as we have seen in the prevous sections. The model (4.4), however, is presented "in the original scale" to avoid confusion.

Define $\beta_1^*(z_2) := \beta_1 + \alpha_2' z_2$, $\tau(z_2) = [\beta_1^*(z_2), \alpha_1, \beta_2]'$ and $w = [1, z_1, x_2']'$. We also use the notation

$$\tau(z_2) := \frac{[\beta_1^*(z_2), \alpha_1, \beta_2]'}{\|[\beta_1^*(z_2), \alpha_1, \beta_2]'\|} \in \mathbb{S}^{d_X+1}, w := \frac{[1, z_1, x_2']'}{\|[1, z_1, x_2']'\|} \in \mathbb{S}^{d_X+1}.$$

Then (4.4) is equivalent to:

$$\begin{aligned} y_i &= \mathbb{I}\{\beta_1^*(z_2) + [\alpha_1, \beta_2][z_1, x_2']' \geq 0\} \\ &= \mathbb{I}\{[\beta_1^*(z_2), \alpha_1, \beta_2][1, z_1, x_2']' \geq 0\} \\ &= \mathbb{I}\left\{\frac{[\beta_1^*(z_2), \alpha_1, \beta_2]}{\|[\beta_1^*(z_2), \alpha_1, \beta_2]'\|} \frac{[1, z_1, x_2']'}{\|[1, z_1, x_2']'\|} \geq 0\right\} \\ &= \mathbb{I}\left\{\tau(z_2)' w \geq 0\right\}. \end{aligned}$$

This has the same form as our original model if we condition on $Z_2 = z_2$. We can then apply previous results for identification and estimation under the following assumptions. First, suppose $(\beta_1, \beta_2')'$ and w are independent, instead of Assumption 1.1. Second, we impose some condition on $f_{W|Z_2=z_2}$, the conditional density of W given $Z_2 = z_2$. More specifically, suppose there exists a set $\mathcal{Z}_2 \in \mathbb{R}^{d_Z-1}$, such that Assumptions 3.1, 3.2 and 3.3 hold if we replace f_X and d with $f_{W|Z_2=z_2}$ and $d_X + 1$ for all $z_2 \in \mathcal{Z}_2$. If Z_2 is a vector of dummies, for example, $\mathcal{Z}_2$ would be a set of discrete points. By (3.5) we obtain

$$f^-_{\tau(z_2)|Z_2=z_2}(t) = \frac{1}{|\mathbb{S}^{d_X}|} \sum_{p=0}^{\infty} \frac{h(2p+1, d_X+1)}{\lambda(2p+1, d_X+1)} \mathbb{E}\left[\left. \frac{(2Y-1) C_{2p+1}^{\nu(d_X+1)}(W' t)}{f_{W|Z_2=z_2}(W) C_{2p+1}^{\nu(d_X+1)}(1)} \right| Z_2 = z_2\right] \tag{4.5}$$

for all $z_2 \in \mathcal{Z}_2$, where the right hand side consists of observables. This determines $f_{\tau(z_2)|Z_2=z_2}$. That is, the conditional density

$$f\left(\frac{[\beta_1^*(z_2), \alpha_1, \beta_2]}{\|[\beta_1^*(z_2), \alpha_1, \beta_2]'\|}\middle| Z_2 = z_2\right)$$

is identified for all $z_2 \in \mathcal{Z}_2$. (Here and henceforth we use the notation $f(\cdot|\cdot)$ to denote conditional densities with appropriate arguments when adding subscripts is too cumbersome.) This obviously identifies

$$f\left(\frac{[\beta_1^*(z_2), \alpha_1, \beta_2]}{\|\beta_2\|}\middle| Z_2 = z_2\right) \tag{4.6}$$

for all $z_2 \in \mathcal{Z}_2$ as well. If we are only interested in the joint distribution of β_2 under a suitable normalization, we can stop here. The presence of the term $\alpha_1 z_1$ in (4.4) is unimportant so far.

Some more work is necessary, however, if one is interested in the joint distribution of the coefficients on all the regressors. Notice that the distribution (4.6) gives

$$f\left(\frac{\beta_1^*(z_2)}{\|\beta_2\|}\middle| Z_2 = z_2\right) = f\left(\frac{\beta_1 + \alpha_2' z_2}{\|\beta_2\|}\middle| Z_2 = z_2\right),$$

from which we can, for example, get

$$E\left(\frac{\beta_1^*(z_2)}{\|\beta_2\|}\middle| Z_2 = z_2\right) = E\left(\frac{\beta_1}{\|\beta_2\|}\right) + E\left(\frac{1}{\|\beta_2\|}\right)\alpha_2' z_2 \quad \text{for all } z_2 \in \mathcal{Z}_2.$$

Define a constant

$$c := E\left(\frac{1}{\|\beta_2\|}\right)$$

then we can identify $c\alpha_2$ as far as $z_2 \in \mathcal{Z}_2$ has enough variation. From the second marginal of the conditional joint density (4.6)

$$E\left(\frac{\alpha_1}{\|\beta_2\|}\right) = E\left[E\left(\frac{\alpha_1}{\|\beta_2\|}\middle| Z_2\right)\middle| Z_2 \in \mathcal{Z}_2\right] = c\alpha_1$$

is identified as well.[1] Let

$$f\left(\frac{[\beta_{2i}', \alpha_1, \alpha_2']'}{\|\beta_{2i}\|}\right) \tag{4.7}$$

denote the joint density of all the coefficient (except for β_1, which corresponds to the conventional disturbance term in the original model (4.4), normalized by the length of β_{2i}. Then

$$f\left(\frac{[\beta_{2i}', \alpha_1, \alpha_2']'}{\|\beta_{2i}\|}\right) = f\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \frac{c\alpha_2}{c\alpha_1} \end{bmatrix} \begin{bmatrix} \frac{\beta_{2i}}{\|\beta_{2i}\|} \\ \frac{\alpha_1}{\|\beta_{2i}|} \end{bmatrix}\right).$$

[1]This can be done by just using the conditional distribution at one value of z_2 as well.

In the expression on the right hand side, $f([\beta_{2i}', \alpha_1]'/\|\beta_{2i}\|)$ is already available from (4.6), and $c\alpha_1$ and $c\alpha_2$ are identified already, therefore the desired joint density (4.7) is identified. Obviously (4.7) also determines the joint density of $[\beta_{2i}', \alpha_1, \alpha_2']'$ under other suitable normalizations as well.

The density (4.5) is estimable: when Z_2 is discrete, one can truncate the infinite sum by a parameter T_N, then compute its sample counterpart by applying the formula (3.8) to each subsample corresponding to each value of Z_2. If Z_2 continuous we can estimate $f_{W|z_2}$ and the conditional expectation by nonparametric smoothing. An estimate for the density (4.6) can be then obtained with some numerical techniques.

4.3. **Endogenous Regressors.** Assumption 1.1 is violated if some of the regressors are endogenous in the sense that the random coefficients and the covariates are not independent. This problem can be solved if an appropriate vector of instruments is available. To be more specific, suppose we observe (Y, X, Z) generated from the following model

$$Y_i = \mathbb{I}\{\alpha_i + \beta_{2i}' X_i \geq 0\} \tag{4.8}$$

with

$$X_i = \Gamma Z_i + V_i \tag{4.9}$$

where V is a vector of reduced form residuals and Z is independent of (β, V). The equations (4.8) and (4.9) yield

$$Y_i = \mathbb{I}\{\left(\alpha_i + V_i'\beta_i\right) + Z_i'\Gamma'\beta_i\}.$$

Suppose the distribution of ΓZ_i satisfy the support conditions assumed for the basic model. It is then possible to estimate the density of $\overline{\tau} = \tau/\|\tau\|$ where $\tau = \left(\beta_1 + V'\tilde{\beta}, \tilde{\beta}\right)'$ by replacing Γ with a consistent estimator, which is easy to obtain under the maintained assumptions. This yields an estimate for the joint density of $\beta_{2i}/\|\beta_{2i}\|$, the random coefficients on the covariates under scale normalization.

5. Conclusion

To be added.

6. Appendix

Let us start this appendix by recalling useful notions of Riemannian geometry. The tangent space $T_x\mathbb{S}^{d-1}$ at a point x on the sphere is the vector space of tangents $\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\big|_{t=0}$ of curves $\gamma : (-\epsilon,\epsilon) \to U$ where $\epsilon > 0$ and U is a neighborhood of x in $\mathbb{R}^d$, drawn on $\mathbb{S}^{d-1}$. We can easily check that it is the orthogonal in $\mathbb{R}^d$ of x. Given a map f from $\mathbb{S}^{d-1}$ to $\mathbb{R}$, its differential d_xf at x in $\mathbb{R}^d$ is a linear form acting on $T_x\mathbb{S}^{d-1}$. It is such that for h in $T_x\mathbb{S}^{d-1}$ corresponding to a curve γ, $d_xf.h$ is defined as $\frac{\mathrm{d}}{\mathrm{d}t}[f(\gamma)]\big|_{t=0}$. A useful example in the case of derivatives of choice probabilities is the height function, see do Carmo (1976) p.86, defined for z in $\mathbb{S}^{d-1}$ as $x \in \mathbb{S}^{d-1} \mapsto z'x$. Its differential is the mapping

$$h \in T_x\mathbb{S}^{d-1} \mapsto z'h. \tag{6.1}$$

As in the Euclidian plane, the gradient on the sphere is related to the above defined differential using the scalar product. The gradient of f at x is denoted by $\nabla_x^S f$ and defined as the vector of $T_x\mathbb{S}^{d-1}$ such that for h in $T_x\mathbb{S}^{d-1}$, $\nabla_x^S f'h = d_xf.h$. The scalar product on the tangent spaces is the restriction of the scalar product in $\mathbb{R}^d$. This is a general construction of a gradient on smooth submanifolds of $\mathbb{R}^d$. It matches in the particular case of the sphere the definition provided by identity (2.6). The Laplace operator on a smooth submanifolds of $\mathbb{R}^d$ is defined through the generalization of the formula div∇. The generalization of the divergence is defined as follows. A vector field X is a map which to x in $\mathbb{S}^{d-1}$ assigns a vector $X(x)$ of $T_x\mathbb{S}^{d-1}$. It is differentiable if given a local parametrization of $\mathbb{S}^{d-1}$, for example using the stereographic projection, consisting of to maps φ from an open set U in $\mathbb{R}^{d-1}$ to $V \cap \mathbb{S}^{d-1}$ where V is an open set of $\mathbb{R}^d$, $X(\varphi)$ is differentiable. The linear mapping which to v in $T_x\mathbb{S}^{d-1}$ corresponding to some curve $\gamma(-\epsilon,\epsilon) \to U$ and X a vector field, assigns the orthogonal projection of $\frac{\mathrm{d}}{\mathrm{d}t}X(\gamma)\big|_{t=0}$ on $T_x\mathbb{S}^{d-1}$ is denoted by D. Then Δ^S is defined as $\mathrm{tr}D\nabla^S$. Also, see for example Gallot et al (2004) p.209, we have

$$-\int_{\mathbb{S}^{d-1}} f(x)\Delta^S f(x)d\sigma(x) = \int_{\mathbb{S}^{d-1}} |||df_x|||^2 d\sigma(x) = \int_{\mathbb{S}^{d-1}} \nabla_x^S f' \nabla_x^S f d\sigma(x) \tag{6.2}$$

where $|||\cdot|||$ denotes the operator norm. We can check using the condensed harmonic expansion, Lemma 2.1 (ii) and relation (6.2) that

$$\|f\|_{2,1}^2 = \|f\|_2^2 + \left\|\|\nabla^S f\|\right\|_2^2$$

where the last term denotes the right hand-side of (6.2) and is the $\mathrm{L}^2(\mathbb{S}^{d-1})$ norm of the Euclidian norm of the gradient. For higher order integer Sobolev norm, we can check with iterations of identity

(6.2) that the norm introduced is equivalent to the usual square root of the some of the squares of the $\mathrm{L}^2(\mathbb{S}^{d-1})$ norms of the derivatives.

We now give some results on the Gegenbauer polynomials. These results can be found in Erdélyi et al. (1953) and Groemer (1996). The Gegenbauer polynomials have the following explicit representation

$$C_n^{\nu}(t) = \sum_{l=0}^{[n/2]} \frac{(-1)^l (\nu)_{n-l}}{l!(n-2l)!}(2t)^{n-2l} \tag{6.3}$$

where $(a)_0 = 1$ and for n in $\mathbb{N}\setminus\{0\}$, $(a)_n = a(a+1)\cdots(a+n-1) = \Gamma(a+n)/\Gamma(a)$. When $\nu = 0$, case $d = 2$, it is related to the Chebychev polynomials of the first kind as follows

$$\forall n \in \mathbb{N}\setminus\{0\},\ C_n^0(t) = \frac{2}{n}T_n(t)$$

and

$$C_0^0(t) = T_0(t) = 1$$

where

$$\forall n \in \mathbb{N},\ T_n(t) = \cos\left(n \arccos(t)\right).$$

When $\nu = 1$, case $d = 4$, $C_n^1(t)$ coincides with the Chebychev polynomial of the second kind $U_n(t)$ which is such that

$$\forall n \in \mathbb{N},\ U_n(t) = \frac{\sin[(n+1)\arccos(t)]}{\sin[\arccos(t)]}.$$

The Gegenbauer polynomials are stable by differentiation, they satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}C_n^{\nu}(t) = 2\nu C_{n-1}^{\nu+1}(t) \tag{6.4}$$

for $\nu > 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}C_n^{0}(t) = 2C_{n-1}^{1}(t). \tag{6.5}$$

For $\nu \neq 0$, the Rodrigues formula

$$C_n^{\nu}(t) = (-2)^{-n}(1-t^2)^{-\nu+1/2}\frac{(2\nu)_n}{(\nu+1/2)_n n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t^2)^{n+\nu-1/2}, \tag{6.6}$$

can be used. The following results are also used in the paper

$$\sup_{t\in[-1,1]}\left|\frac{C_n^{\nu}(t)}{C_n^{\nu}(1)}\right| \leq 1, \tag{6.7}$$

for ν positive,

$$\forall n \in \mathbb{N},\ C_n^{\nu}(1) = \begin{pmatrix} n+2\nu-1 \\ n \end{pmatrix} \tag{6.8}$$

while for $\nu = 0$

$$C_0^0(1) = 1 \text{ and } \forall n \in \mathbb{N}\setminus\{0\},\ C_n^0(1) = \frac{2}{n}, \tag{6.9}$$

the following parity relation holds

$$C_n^{\nu}(-t) = (-1)^n C_n^{\nu}(t). \tag{6.10}$$

The normalization is such that

$$\int_{-1}^{1} (C_n^{\nu(d)}(t))^2(1-t^2)^{(d-3)/2}dt = \frac{|\mathbb{S}^{d-1}|(C_n^{\nu(d)}(1))^2}{|\mathbb{S}^{d-2}|h(n,d)}. \tag{6.11}$$

Let us present some useful inequalities. We herein use the following notation

Notation. $\zeta_n = 1 + n(n+d-2)$. □

Lemma 6.1. *For constants depending only on the dimension, the following holds:*

$$h(n,d) \asymp n^{d-2}, \tag{6.12}$$

$$\sum_{n=1}^{T} h(n,d)\zeta_n^v \asymp T^{2v+d-1}, \tag{6.13}$$

$$|\lambda(2p+1,d)| \asymp p^{-d/2}, \tag{6.14}$$

$$\sum_{p=1}^{T} \frac{h(2p+1,d)}{\lambda(2p+1,d)^2}\zeta_{2p+1}^v \asymp T^{2v+2d-1}, \tag{6.15}$$

$$\sum_{p=1}^{T} \frac{h(2p+1,d)}{|\lambda(2p+1,d)|} \asymp T^{3d/2-1}. \tag{6.16}$$

Proof. Estimate (6.12) is clearly satisfied when $d = 2$ and 3 since $h(n,2) = 2$ and $h(n,3) = 2n+1$. When $d \geq 4$ we have

$$h(n,d) = \frac{2}{(d-2)!}(n+(d-2)/2)[(n+1)(n+2)\cdots(n+d-3)],$$

the lower bound is straightforward and the upper bound follows from

$$h(n,d) \leq \frac{2}{(d-2)!}(n+d-3)^{d-2}$$

and $2/((d-2)!)$ by a constant large enough.

Estimate (6.13) follows from (6.12), $\zeta_n \asymp n^2$ and comparing the sum with the integral of $x \mapsto x^{2v+d-2}$.

When d is even and $p \geq d/2$

$$|\lambda(2p+1,d)| = \frac{\kappa_d}{(2p+1)(2p+3)\cdots(2p+d-1)}$$

where

$$\kappa_d = \frac{|\mathbb{S}^{d-2}| 1\cdot 3\cdots(d-1)}{d-1}.$$

The upper bound is straightforward and we can write

$$|\lambda(2p+1,d)| \geq \frac{\kappa_d}{(2p+d-1)^{d/2}}$$

and conclude replacing κ_d by a small enough constant.

Sterling's double inequality, see Feller (1968) p.50-53

$$\sqrt{2\pi} n^{n+1/2} \exp\left(-n + \frac{1}{12n+1}\right) < n! < \sqrt{2\pi} n^{n+1/2} \exp\left(-n + \frac{1}{12n}\right)$$

implies that

$$\frac{(2^p p!)^2}{(2p)!} \asymp \sqrt{p}$$

thus

$$1\cdot 3\cdots(2p-1) \asymp \sqrt{p} 2\cdot 4\cdots(2p).$$

Therefore, for $p \geq d/2$ and d odd we have

$$|\lambda(2p+1,d)| \asymp \frac{\sqrt{p}}{(2p+2)(2p+4)\cdots(2p+d-1)}$$

and (6.14) holds for d even and odd.

(6.15) and (6.16) follow from the above inequality and comparing sums with integrals. □

Proof of Proposition 2.2. Consider $\left(S_n^l\right)_{l=1,\dots,h(n,d);n=1,\dots,\infty}$ an orthonormal basis of spherical harmonics. Given f in $\mathrm{H}^s(\mathbb{S}^{d-1})$ we have for x in $\mathbb{S}^{d-1}$,

$$f(x) = \sum_{n=0}^{\infty} \sum_{l=1}^{h(n,d)} \left(f, S_n^l\right)_{\mathrm{L}^2(\mathbb{S}^{d-1})} S_n^l(x).$$

Using the Hölder inequality we obtain

$$\begin{aligned} |f(x)|^2 &\leq \sum_{n=0}^{\infty} \sum_{l=1}^{h(n,d)} (1+n(n+d-2))^s \left(f, S_n^l\right)^2_{\mathrm{L}^2(\mathbb{S}^{d-1})} \sum_{n=0}^{\infty} \sum_{l=1}^{h(n,d)} (1+n(n+d-2))^{-s} \left(S_n^l(x)\right)^2 \\ &\leq \frac{\|f\|_{2,s}^2}{|\mathbb{S}^{d-1}|} \sum_{n=0}^{\infty} (1+n(n+d-2))^{-s} h(n,d) \end{aligned}$$

indeed using (2.7) and (2.10) $\sum_{l=1}^{h(n,d)}\left(S_n^l(x)\right)^2 = q_{n,d}(x,x) = \frac{h(n,d)}{|\mathbb{S}^{d-1}|}$. Thus using (6.12), for $s > (d-1)/2$ and some constant c depending on the dimension only, we have

$$|f(x)|^2 \leq c\sum_{n=0}^{\infty} n^{-2s+d-2}\|f\|_{2,s}^2.$$

Therefore there exists a constant $\tilde{c}_d$ such that

$$\|f\|_\infty \leq \tilde{c}_d\|f\|_{2,s}$$

thus the injection of $\mathrm{H}^s(\mathbb{S}^{d-1})$ into $\mathrm{L}^\infty(\mathbb{S}^{d-1})$ is continuous. We may check as well that f is continuous.

$\square$

Proof of Proposition 2.3. From the Funk-Hecke theorem we know that the coefficients $\alpha(n,d) = C_n^{\nu(d)}(1)|\mathbb{S}^{d-2}|^{-1}\lambda_n\left(\mathbb{I}\left\{t \in [0,1]\right\}\right)$ are given by

$$\alpha(n,d) = \int_0^1 C_n^{\nu(d)}(t)(1-t^2)^{(d-3)/2}dt$$

using (6.6),

$$\alpha(n,d) = \frac{(-2)^{-n}(d-2)_n}{n!\left((d-1)/2\right)_n}\int_0^1 \frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t^2)^{n+(d-3)/2}dt.$$

Thus for $n \geq 1$ and $d \geq 3$,

$$\alpha(n,d) = -\frac{(-2)^{-n}(d-2)_n}{n!\left((d-1)/2\right)_n}\left.\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}(1-t^2)^{n-1+(d-3)/2}dt\right|_{t=0}$$

since the term on the right hand-side is equal to 0 for $t = 1$. To prove that the coefficients $\alpha(2p,d)$ are equal to zero for p positive it is enough to prove

$$\left.\frac{\mathrm{d}^{2p+1}}{\mathrm{d}t^{2p+1}}(1-t^2)^{2p+1+m}\right|_{t=0} = 0, \quad \forall m \geq 1,\ p \geq 0.$$

The Faá di Bruno formula gives that this quantity is equal to

$$\sum_{k_1+2k_2=2p+1}\left.\frac{(-1)^{2p+1-k_2}(2p+1)!(m+1)\cdots(2p+1+m)}{k_1!k_2!}(1-t^2)^{m+k_2}(2t)^{k_1}\right|_{t=0}.$$

and we conclude since k_1 in the sum cannot be equal to 0.

When $n = 2p+1$ for p nonnegative we obtain, using again the Faá di Bruno formula, that the derivative at $t = 0$ is equal to

$$(-1)^p\frac{(2p)!}{p!}\left[(2p+1+(d-3)/2)(2p+(d-3)/2)\cdots(p+2+(d-3)/2)\right].$$

We obtain the result of Proposition 2.3 using identity (6.8). For the case $d = 2$ we use Proposition 2.1.

$\square$

Let us now prove the theorems concerning the estimation of f_X.

Proof of Theorem 2.3. f_X being a probability density function, it has the following condensed harmonic expansion

$$f_X(x) = \frac{1}{|\mathbb{S}^{d-1}|} + \sum_{n=1}^{\infty}(Q_{n,d}f_X)(x).$$

We then write

$$\begin{aligned}(Q_{n,d}f_X)(x) &= \int_{\mathbb{S}^{d-1}} q_{n,d}(x,y)f_X(y)d\sigma(y)\\ &= \mathbb{E}\left[q_{n,d}(x,X)\right]\\ &= \frac{h(n,d)}{|\mathbb{S}^{d-1}|C_n^{\nu(d)}(1)}\mathbb{E}\left[C_n^{\nu(d)}(x'X)\right],\end{aligned}$$

using (2.10), and use (2.11). The case where f_X is supported in some hemisphere can be treated similarly. □

Lemma 6.2. *For f in H^s and $v < s$,*

$$\left\|\sum_{n=T+1}^{\infty} Q_{n,d}f\right\|_{2,v}^2 \le cT^{-2(s-v)}\left\|f\right\|_{2,s}^2.$$

Proof. The result follows from the next sequence of inequalities

$$\begin{aligned}\left\|\sum_{n=T+1}^{\infty} Q_{n,d}f\right\|_{2,v}^2 &\le \sum_{n=T+1}^{\infty}\zeta^v\|Q_{n,d}f\|_2^2\\ &\le \zeta_{T+1}^{-(s-v)}\sum_{n=T+1}^{\infty}\zeta_n^s\|Q_{n,d}f\|_2^2\\ &\le \zeta_{T+1}^{-(s-v)}\sum_{n=T+1}^{\infty}\zeta_n^s\|Q_{n,d}f\|_2^2\end{aligned}$$

and the fact that $\zeta_{T+1} > T^{-2(s-v)}$. □

Proof of Theorem 2.4. First consider the estimator (2.15). Using Lemma 6.2 and the fact that $f_X(x) - \mathbb{E}\left[\hat{f}_X^{N,T}(x)\right]$ is equivalent in $\mathrm{L}^2(\mathbb{S}^{d-1})$ to $\sum_{n=N+1}^{\infty}\left(Q_{n,d}f_X\right)(x)$, the square of the bias is bounded as follows

$$\left\|f_X - \mathbb{E}\left[\hat{f}_X^{N,T}\right]\right\|_{2,v}^2 \le T^{-2(\sigma-v)}\left\|f_X\right\|_{2,\sigma}^2.$$

Let us now consider the variance and show that

$$\mathbb{E}\left[\left\|\hat{f}_X^{N,T} - \mathbb{E}\left[\hat{f}_X^{N,T}\right]\right\|_{2,v}^2\right] \le \frac{c}{N}T^{2v+d-1}$$

for some constant c depending only on the dimension d.

Using (2.7)

$$\sum_{i=1}^{N} q_{n,d}(x_i, x) = \sum_{l=1}^{h(n,d)} \sum_{i=1}^{N} S_n^l(x_i) S_n^l(x)$$

is a linear combination of elements of $H^{n,d}$ and thus an element of $H^{n,d}$. Also for any orthonormal basis $\left(S_n^l\right)_{l=1}^{h(n,d)}$ of $H^{n,d}$

$$\mathbb{E}\left[q_{n,d}(X, x)\right] = \sum_{l=1}^{h(n,d)} \mathbb{E}\left[S_n^l(X)\right] S_n^l(x),$$

it is therefore also an element of $H^{n,d}$. Thus, we have

$$\left\| \hat{f}_X^{N,T} - \mathbb{E}\left[\hat{f}_X^{N,T}\right] \right\|_{2,v}^2 = \sum_{n=0}^{T} \frac{\zeta_n^v}{N^2} \int_{\mathbb{S}^{d-1}} \left(\sum_{i=1}^{N} q_{n,d}(x_i, x) - \mathbb{E}\left[q_{n,d}(X, x)\right] \right)^2 d\sigma(x).$$

Denoting the expectation of the above term by $\mathbb{V}(N, T)$,

$$\begin{aligned}
\mathbb{V}(N,T) &= \sum_{n=0}^{T} \frac{\zeta_n^v}{N^2} \int_{\mathbb{S}^{d-1}} \mathbb{E}\left(\sum_{i=1}^{N} q_{n,d}(x_i, x) - \mathbb{E}\left[q_{n,d}(X, x)\right] \right)^2 d\sigma(x) \\
&= \sum_{n=0}^{T} \frac{\zeta_n^v}{N} \int_{\mathbb{S}^{d-1}} \mathbb{E}\left(q_{n,d}(X, x) - \mathbb{E}\left[q_{n,d}(X, x)\right]\right)^2 d\sigma(x).
\end{aligned}$$

Thus

$$\mathbb{V}(N,T) \leq \sum_{n=0}^{T} \frac{\zeta_n^v}{N} \mathbb{E} \int_{\mathbb{S}^{d-1}} \left(q_{n,d}(X, x)\right)^2 d\sigma(x).$$

Using successively Theorem 2.1 and relation (6.11) it follows that

$$\begin{aligned}
\mathbb{V}(N,T) &\leq \frac{1}{N|\mathbb{S}^{d-1}|^2} \sum_{n=0}^{T} \frac{\zeta_n^v h(n,d)^2}{\left(C_n^{\nu(d)}(1)\right)^2} \int_{\mathbb{S}^{d-1}} (C_n^{\nu(d)}(X'x))^2 d\sigma(x) \\
&\leq \frac{1}{N|\mathbb{S}^{d-1}|} \sum_{n=0}^{T} \zeta_n^v h(n,d)
\end{aligned}$$

and we conclude using (6.13). Choosing $T = T_N$ of the order indicated in the theorem balances the orders of the squared bias and the variance. (2.18) is a consequence of Proposition 2.2. It is easily seen that (2.17) and (2.18) also hold for $\hat{f}_X^{-,N,T_n}$ and f_X^-. Since $f_X(x) - f_X^{N,T_N}(x) = 2f_X^-(x) - 2f_X^{-,N,T_N}(x)$ for x in H and is equal to 0 on $\mathbb{S}^{d-1} \setminus H$, the same convergence rate result for the estimator (2.16) as well.

$\square$

Proof of Theorem 2.5. The stated lower bound for α guarantees that the bias term due to the truncation by T_N is asymptotically negligible. This can be shown following essentially the same steps as in the argument for the asymptotic negligibility of the term C_3 in the proof of Theorem 3.3. Therefore it suffices to verify that an appropriate CLT holds for the triangular array $\{Z_{N,i}\}_{i=1}^N, N = 1, 2, ...$ Consider the first scenario where Assumption (i) is satisfied. We verify the Lyapounov condition: there exists $\delta > 0$ such that for N going to infinity,

$$\frac{\mathbb{E}\left[|Z_{N,1} - \mathbb{E}[Z_{N,1}]|^{2+\delta}\right]}{N^{\delta/2}\left(\operatorname{var}(Z_{N,1})\right)^{1+\delta/2}} \to 0, \tag{6.17}$$

We need a lower bound on $\operatorname{var}(Z_{N,1})$. Since $\mathbb{E}[Z_{N,1}]$ converges to $f_X(x)$ while the variance blows-up, it is enough to obtain a lower bound on

$$\mathbb{E}[Z_{N,1}^2] = \int_{\mathbb{S}^{d-1}}\left(\sum_{n=0}^{T_N} q_{n,d}(y,x)\right)^2 f_X(y)d\sigma(y). \tag{6.18}$$

Using (i) and the computation in the proof of Theorem 2.4 we obtain

$$\mathbb{E}[Z_{N,1}^2] \geq \frac{1}{C_X|\mathbb{S}^{d-1}|}\sum_{n=0}^{T_N} h(n,d),$$

thus, using (6.13), for some constant c, the denominator of (6.17) is greater than $cN^{\delta/2}N^{\alpha(d-1)(1+\delta/2)}$. As for the denominator, it is enough to obtain an upper bound of $\mathbb{E}\left[|Z_{N,1}|^{2+\delta}\right]$, where

$$\begin{aligned}
\mathbb{E}\left[|Z_{N,1}|^{2+\delta}\right] &= \int_{\mathbb{S}^{d-1}}\left|\sum_{n=0}^{T_N} q_{n,d}(y,x)\right|^{2+\delta} f_X(y)d\sigma(y)\\
&\leq c|\mathbb{S}^{d-1}|\left(\sum_{n=0}^{T_N}\sup_{y,x\in\mathbb{S}^{d-1}}|q_{n,d}(y,x)|\right)^{2+\delta}\\
&\leq c|\mathbb{S}^{d-1}|\left(\sum_{n=0}^{T_N} h(n,d)\right)^{2+\delta}\\
&\leq c|\mathbb{S}^{d-1}|N^{\alpha(d-1)(2+\delta)}
\end{aligned}$$

where c is a constant that take different values and is only a function of the dimension. Note that we have used that, since $\sigma > (d-1)/2$, f_X is bounded. In the last line above we have used equation (2.10) together with inequality (6.7). Using again (6.13) we obtain that the criterion (6.17) is satisfied if

$$\frac{\delta/2}{1+\delta/2} > \alpha(d-1)$$

. Taking δ large enough results in the upper bound for α. The second scenario with Assumption (ii) can be treated using the same argument. We however need to replace (6.18) by is

$$\begin{aligned}
\mathbb{E}[Z_{N,1}^2] &= \int_H \left(\sum_{p=0}^{T_N} q_{2p+1,d}(y,x) \right)^2 f_X(y) d\sigma(y) \\
&\geq \frac{1}{C_X} \int_H \left(\sum_{p=0}^{T_N} q_{2p+1,d}(y,x) \right)^2 d\sigma(y) \\
&= \frac{1}{C_X} \int_{\mathbb{S}^{d-1}} \left(\sum_{p=0}^{T_N} q_{2p+1,d}(y,x) \right)^2 d\sigma(y).
\end{aligned}$$

where we have used (6.10) to obtain the last identity. □

Let us turn to results concerning choice probabilities.

Proof of Theorem 3.1. R has the following condensed harmonic expansion

$$R(x) = \frac{1}{2} + \sum_{p=1}^{\infty} (Q_{2p+1,d}R)(x).$$

We then write using (3.2), changing variables and using (6.10),

$$\begin{aligned}
(Q_{2p+1,d}R)(x) &= \int_{\mathbb{S}^{d-1}} q_{2p+1,d}(x,z) R(z) d\sigma(z) \\
&= \int_{H^+} q_{2p+1,d}(x,z) r(z) d\sigma(z) + \int_{-H^+} q_{2p+1,d}(x,z)(1 - r(-z)) d\sigma(z) \\
&= \int_{H^+} q_{2p+1,d}(x,z) r(z) d\sigma(z) - \int_{H^+} q_{2p+1,d}(x,z)(1 - r(z)) d\sigma(z),
\end{aligned}$$

and using (2.10) and (2.11)

$$\begin{aligned}
(Q_{2p+1,d}R)(x) &= \int_{H^+} q_{2p+1,d}(x,z)(2r(z) - 1) d\sigma(z) \\
&= \int_{H^+} q_{2p+1,d}(x,z) \mathbb{E}\left[\frac{2Y-1}{f_X(z)} \middle| X = z \right] f_X(z) d\sigma(z) \\
&= \mathbb{E}\left[\frac{(2Y-1) q_{2p+1,d}(x,Z)}{f_X(Z)} \right] \\
&= \frac{h(n,d)}{|\mathbb{S}^{d-1}| C_n^{\nu(d)}(1)} \mathbb{E}\left[\frac{(2Y-1) C_n^{\nu(d)}(x'X)}{f_X(X)} \right].
\end{aligned}$$

□

Now we turn to the proofs for Theorems 3.2 and 3.3 for the estimator of f_β. We first provide some notation and auxiliary lemmas used for the proofs. The assumption stated in the theorems are maintained throughout the lemmas.

Notation. In what follows we often drop subscripts/superscripts for the dimension d for notational economy. Thus we write h_n, λ_n, Q_n, and q_n for $h(h,d)$ $\lambda(n,d)$, $Q_{n,d}$ and $q_{n,d}$. Define

$$P_n(x'b) := \frac{C_{2p+1}^{\nu(d)}(x'b)}{C_{2p+1}^{\nu(d)}(1)}.$$

For a density function f on $\mathbb{S}^{d-1}$ let:

$$U_{T,i}(f,b) = \frac{2y_i - 1}{f(x_i)} \sum_{p=0}^{T} \frac{h_{2p+1}}{|\mathbb{S}^{d-1}|\lambda_{2p+1}} P_{2p+1}(x_i'b),$$

$$\hat{f_\beta}^{-}(f,b) = \frac{1}{N}\sum_{i=1}^{N} U_{T,i}(f,b),$$

$$M_T(f,b) = E[U_{T,i}(f,b)]$$

$$J_{1T}(f) = E\left[\int_{\mathbb{S}^{d-1}} U_{T,i}^2(f,b)d\sigma(b)\right] = E[\|U_{T,i}(f,b)\|_2^2],$$

$$J_{2T}(f) = \int_{\mathbb{S}^{d-1}} E[U_{T,i}(f,b)]^2 d\sigma(b) = \int_{\mathbb{S}^{d-1}} M_T^2(f,b)d\sigma(b) = \|M_T(f,b)\|_2^2.$$

Lemma 6.3. *For $g \in L^2(\mathbb{S}^{d-1})$,*

$$\left\|\sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} Q_{2p+1} g\right\|_2 \le \left|\frac{1}{\lambda_{2T+1}}\right| \|g\|_2.$$

Proof of Lemma 6.3.

$$\begin{aligned}
\left\|\sum_{p=0}^{T}\frac{1}{\lambda_{2p+1}}Q_{2p+1}g\right\|_2^2 &= \int_{\mathbb{S}^{d-1}}\left[\sum_{p=0}^{T}\frac{1}{\lambda_{2p+1}}Q_{2p+1}g\right]^2 d\sigma(b)\\
&= \int_{\mathbb{S}^{d-1}}\sum_{p=0}^{T}\frac{1}{\lambda_{2p+1}^2}\left[Q_{2p+1}g\right]^2 d\sigma(b)\\
&\le \int_{\mathbb{S}^{d-1}}\frac{1}{\lambda_{2T+1}^2}\sum_{p=0}^{T}\left[Q_{2p+1}g\right]^2 d\sigma(b)\\
&= \frac{1}{\lambda_{2T+1}^2}\sum_{p=0}^{T}\|Q_{2p+1}g\|_2^2\\
&\le \frac{1}{\lambda_{2T+1}^2}\sum_{p=0}^{\infty}\|Q_{2p+1}g\|_2^2\\
&\le \frac{1}{\lambda_{2T+1}^2}\sum_{n=0}^{\infty}\|Q_{n}g\|_2^2\\
&= \frac{1}{\lambda_{2T+1}^2}\left\|\sum_{n=0}^{\infty}Q_{n}g\right\|_2^2\\
&= \frac{1}{\lambda_{2T+1}^2}\|g\|_2^2.
\end{aligned}$$

□

Following lemmas report asymptotic (stochastic) order results when $T\to\infty$ and $\|f-f_X\|_\infty\downarrow 0$ and/or the sample size N goes to infinity.

Lemma 6.4. *For a density function f on $\mathbb{S}^{d-1}$,*

$$J_{1T}(f)-J_{1T}(f_X)=O(T^{2d-1}\|f-f_X\|_2)O\left(\left\|\frac{1}{f}\left(\frac{1}{f}+\frac{1}{f_X}\right)\right\|_2\right)$$

and

$$J_{2T}(f)-J_{2T}(f_X)=O(T^d)O(\|f-f_X\|_2)O(\|1/f\|_\infty\|(f+f_X)/f\|_2).$$

Proof of Lemma 6.4. Regarding the equation for

$$J_{1T}(f)=E\left[\int_{\mathbb{S}^{d-1}}U_{T,i}^2(f,b)d\sigma(b)\right],$$

note

$$\int_{\mathbb{S}^{d-1}} U_{T,i}^2(f,b)d\sigma(b) = \int_{\mathbb{S}^{d-1}} \frac{1}{f(x_i)^2}\left(\sum_{p=0}^{T}\frac{h_{2p+1}}{\lambda_{2p+1}}\frac{P_{2p+1}(x_i,b)}{|\mathbb{S}^{d-1}|}\right)^2 d\sigma(b)$$
$$= \frac{1}{f^2(x_i)||\mathbb{S}^{d-1}|^2}\sum_{p=0}^{T}\frac{h_{2p+1}^2}{\lambda_{2p+1}^2}\int_{\mathbb{S}^{d-1}} P_{2p+1}^2(x_i'b)d\sigma(b).$$

Now,

$$\int_{\mathbb{S}^{d-1}} P_{2p+1}^2(x_i'b)d\sigma(b) = \frac{1}{(C_{2p+1}^{\nu(d)}(1))^2}\int_{\mathbb{S}^{d-1}}(C_{2p+1}^{\nu(d)}(x_i'b))^2 d\sigma(b)$$
$$= \frac{1}{(C_{2p+1}^{\nu(d)}(1))^2}|\mathbb{S}^{d-2}|\int_{\mathbb{S}^{d-1}}(C_{2p+1}^{\nu(d)})^2(1-t^2)^{\frac{d-3}{2}}dt$$
$$= \frac{1}{(C_{2p+1}^{\nu(d)}(1))^2}|\mathbb{S}^{d-2}|\frac{|\mathbb{S}^{d-1}|(C_{2p+1}^{\nu(d)}(1))^2}{|\mathbb{S}^{d-2}|h_{2p+1}} \quad \text{by (6.11)}$$
$$= |\mathbb{S}^{d-1}|/h_{2p+1}.$$

Therefore

$$\int_{\mathbb{S}^{d-1}} U_{T,i}^2(f,b)d\sigma(b) = \frac{1}{f^2(x_i)||\mathbb{S}^{d-1}|}\sum_{p=0}^{T}\frac{h_{2p+1}}{\lambda_{2p+1}^2},$$

and

$$J_1(f) = \frac{1}{|\mathbb{S}^{d-1}|}\left(\sum_{p=0}^{T}\frac{h_{2p+1}}{\lambda_{2p+1}^2}\right)E\left[\frac{1}{f^2(X)}\right]$$

or

$$J_1(f) - J_1(f_X) = \frac{1}{|\mathbb{S}^{d-1}|}\left(\sum_{p=0}^{T}\frac{h_{2p+1}}{\lambda_{2p+1}^2}\right)E\left[\frac{1}{f^2(X)} - \frac{1}{f_X^2(X)}\right].$$

Finally,

$$E\left[\frac{1}{f^2(X)} - \frac{1}{f_X^2(X)}\right] = \int_{\mathbb{S}^{d-1}}\frac{1}{f}\left(\frac{1}{f}+\frac{1}{f_X}\right)(f_X - f)d\sigma(x)$$
$$\leq \left\|\frac{1}{f}\left(\frac{1}{f}+\frac{1}{f_X}\right)\right\|_2 \|f_X - f\|_2.$$

Using (6.15), the desired result follows. Next, we turn to the equation for J_{2T}. With the notation above,

$$J_{2T}(f) - J_{2T}(f_X) = \int_{\mathbb{S}^{d-1}} M_T^2(f,b)d\sigma(x) - \int_{\mathbb{S}^{d-1}} M_T^2(f_X,b)d\sigma(x) \tag{6.19}$$
$$= \int_{\mathbb{S}^{d-1}} \left(M_T(f,b) + M_T(f_X,b)\right)\left(M_T(f,b) - M_T(f_X,b)\right) d\sigma(x)$$
$$\leq \|(M_T(f,b) + M_T(f_X,b)\|_2 \|(M_T(f,b) - M_T(f_X,b)\|_2.$$

Now,

$$M_T(f,b) + M_T(f_X,b) = E\left[\left(\frac{1}{f(x_i)} + \frac{1}{f_X(x_i)}\right)(2y_i - 1)\sum_{p=0}^{T} \frac{h_{2p+1}}{|\mathbb{S}^{d-1}|\lambda_{2p+1}} P_{2p+1}(x_i'b)\right]$$
$$= E\left[\left(\frac{1}{f(x_i)} + \frac{1}{f_X(x_i)}\right)(2R(x_i) - 1)\sum_{p=0}^{T} \frac{h_{2p+1}}{|\mathbb{S}^{d-1}|\lambda_{2p+1}} P_{2p+1}(x_i'b)\right]$$

by iterated expectations

$$= \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} \int_{\mathbb{S}^{d-1}} \left(\frac{f+f_X}{f}\right)(2R-1)\frac{h_{2p+1}}{|\mathbb{S}^{d-1}|\lambda_{2p+1}} P_{2p+1}(x'b)d\sigma(x)$$
$$= \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} \int_{\mathbb{S}^{d-1}} \left(\frac{f+f_X}{f}\right)(2R-1)q_{2p+1}(x,b)d\sigma(x)$$

by Theorem 2.1 (Addition formula)

$$= \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} Q_{2p+1}\left(\frac{f+f_X}{f}(2R-1)\right).$$

Take the L^2 norm of both sides, then apply Lemma 6.3 and (6.15), to get:

$$\|M_T(f,b) + M_T(f_X,b)\|_2 \leq \left|\frac{1}{\lambda_{2T+1}}\right| \left\|\frac{f+f_X}{f}(2R-1)\right\|_2 \tag{6.20}$$
$$\leq \left|\frac{1}{\lambda_{2T+1}}\right| \left\|\frac{f+f_X}{f}\right\|_2$$
$$= O(T^{\frac{d}{2}})O(\|(f+f_X)/f\|_2).$$

By similar calculations,

$$\|M_T(f,b) - M_T(f_X,b)\|_2 \le \left|\frac{1}{\lambda_{2T+1}}\right| \left\| \left(\frac{f_X}{f} - 1\right)(2R-1)\right\|_2 \tag{6.21}$$

$$\le \left|\frac{1}{\lambda_{2T+1}}\right| \left\| \left(\frac{1}{f}(f - f_X)\right)\right\|_2$$

$$= O(T^{\frac{d}{2}})O(\|f - f_X\|_2)O(\|1/f\|_\infty).$$

The result follows from (6.19), (6.20) and (6.21). □

Now we are ready to establish the rate of convergence of $\hat{f}_\beta^-(\hat{f}_X, \cdot)$ as an estimator for $f_\beta^-(\cdot)$. The first lemma shows that $\hat{f}_\beta^-(\hat{f}_X, \cdot)$ converges to $M_T(\hat{f}_X, \cdot)$ at an appropriate speed, then the second lemma replaces $M_T(\hat{f}_X, b, \cdot)$ with the target, $f_\beta^-(b, \cdot)$.

Lemma 6.5.

$$\|\hat{f}_\beta^-(\hat{f}_X, \cdot) - M_T(\hat{f}_X, \cdot)\|_2 = O_p(N^{-\frac{1}{2}}T^{\frac{2d-1}{2}}).$$

Proof of Lemma 6.5. Let f be an element of $L^2(\mathbb{S}^{d-1})$.

$$\|\hat{f}_\beta^-(f, \cdot) - M_T(f, \cdot)\|_2^2 = \left\| \frac{1}{N}\sum_{i=1}^N (U_{T,i}(f,\cdot) - E[U_{T,i}(f,\cdot)])\right\|_2^2$$

$$= \frac{1}{N^2}\int_{\mathbb{S}^{d-1}} \left[\sum_{i=1}^N (U_{T,i}(f,b) - E[U_{T,i}(f,b)])\right]^2 d\sigma(b),$$

therefore

$$E\|\hat{f}_\beta^-(f, \cdot) - M_T(f, \cdot)\|_2^2 = \frac{1}{N}\int_{\mathbb{S}^{d-1}} E\left[U_{T,i}(f,b) - E[U_{T,i}(f,b)]\right]^2 d\sigma(b)$$

$$= \frac{1}{N}\int_{\mathbb{S}^{d-1}} \left[E\left[U_{T,i}^2(f,b)\right] - E\left[U_{T,i}(f,b)\right]^2\right] d\sigma(b)$$

$$= \frac{1}{N}\left(J_{1T}(f) - J_{2T}(f)\right).$$

In particular, evaluated at the true covariates density $f = f_X$, the above quantity is the integral of the variance of $\hat{f}_\beta^-(f_X, b) =$ with respect to $\sigma(b)$ over $\mathbb{S}^{d-1}$. By regularization, the order of magnitude of this term is given by $O(N^{-1}T^{2d-1})$ (see the proof of Theorem 2.4). That is,

$$\frac{1}{N}(J_{1T}(f_X) - J_{2T}(f_X)) = O(N^{-1}T^{2d-1}).$$

By this and Lemma 6.4,

$$\begin{aligned}
E\|\hat{f}_\beta^-(f,\cdot)-M_T(f,\cdot)\|_2^2 =& \frac{1}{N}\left(J_{1T}(f)-J_{2T}(f)\right)\\
=& \frac{1}{N}\left(J_{1T}(f_X)-J_{2T}(f_X)\right)+\frac{1}{N}\left(J_{1T}(f)-J_{1T}(f_X)\right)-\frac{1}{N}\left(J_{2T}(f)-J_{2T}(f_X)\right)\\
=& O(N^{-1}T^{2d-1})+O(N^{-1}T^{2d-1}\|f-f_X\|_2)O\left(\left\|\frac{1}{f}\left(\frac{1}{f}+\frac{1}{f_X}\right)\right\|_2\right)\\
&+O(N^{-1}T^{d})O(\|f-f_X\|_2)O(\|1/f\|_\infty\|(f+f_X)/f\|_2).
\end{aligned}$$

Therefore the LHS is $O(N^{-1}T^{2d-1})$ if $\|f-f_X\|_2=o(1)$. By Markov and the convergence result for $\hat{f}_X$ (Theorem 2.4),

$$\|\hat{f}_\beta^-(\hat{f}_X,\cdot)-M_T(\hat{f}_X,\cdot)\|_2=O_p(N^{-\frac{1}{2}}T^{\frac{2d-1}{2}}).$$

□

Lemma 6.6. *Suppose* $T\asymp N^{\frac{1}{2s+2d-1}}$*, then*

$$\|\hat{f}_\beta^-(\hat{f}_X,\cdot)-f_\beta^-(\cdot)\|_2=O_p(N^{\frac{-s}{2s+2d-1}}).$$

Proof of Lemma 6.6. Note

$$\begin{aligned}
\|\hat{f}_\beta^-(\hat{f}_X,\cdot)-f_\beta^-(\cdot)\|_2 &= \|\hat{f}_\beta^-(\hat{f}_X,\cdot)-M_T(\hat{f}_X,\cdot)+M_T(\hat{f}_X,\cdot)-M_T(f_X,\cdot)+M_T(f_X,\cdot)-f_\beta^-(\cdot)\|_2\\
&\le \|\hat{f}_\beta^-(\hat{f}_X,\cdot)-M_T(\hat{f}_X^-,\cdot)\|_2+\|M_T(\hat{f}_X^-,\cdot)-M_T(f_X,\cdot)\|_2+\|M_T(f_X,\cdot)-f_\beta^-(\cdot)\|_2\\
&:= \|A_1\|_2+\|A_2\|_2+\|A_3\|_2.
\end{aligned}$$

By Lemma 6.5, $A_1=O_p(N^{-\frac{1}{2}}T^{\frac{2d-1}{2}})$. By (6.21) and Theorem 2.4,

$$\begin{aligned}
A_2 &= O_p(T^{\frac{d}{2}}\|\hat{f}_X-f_X\|_2\|1/\hat{f}_X\|_\infty)\\
&= O_p(T^{\frac{d}{2}}N^{\frac{-\sigma}{2\sigma+d-1}}).
\end{aligned}$$

Lemma 6.2 implies that

$$\|A_3\|_2=\left\|\sum_{p=0}^{T}Q_{2p+1}f_\beta^- -\sum_{p=0}^{\infty}Q_{2p+1}f_\beta^-\right\|_2=O(T^{-s}).$$

If the true density f_X is known, the term A_1 evaluated at f_X corresponds to the standard error and the term A_3 the bias. The optimal choice of T that balances the bias-variance trade-off is $T_N\asymp N^{\frac{1}{2s+2d-1}}$, giving the convergence rate $N^{\frac{-s}{2s+2d-1}}$. The term A_2 reflects the effect of substituting f_X with its estimate. Evaluated at the optimal rate T_N,

$$A_2=O_p(N^{\frac{d}{2(2s+2d-1)}}N^{\frac{-\sigma}{2\sigma+d-1}}).$$

So A_2 becomes negligible if

$$\frac{d}{2(2s+2d-1)} - \frac{\sigma}{2\sigma+d-1} \leq \frac{-s}{2s+2d-1},$$

which holds if $\sigma \geq s + \frac{d}{2}$. □

Lemma 6.6 establishes the L_2 convergence rate of $\hat{f}_\beta^-(\hat{f}_X, b) = \hat{f}_\beta^{-,N,T}(b)$. This is our feasible estimator (3.8) for f_β^-, the odd part of f_β. Now we are ready to show the L_2 convergence result for $\hat{f}_\beta^{N,T}$.

Proof of Theorem 3.2. For notational convenience, in this proof we simply write $\hat{f}_\beta := \hat{f}_\beta^{N,T}$, $\hat{f}_\beta^- := \hat{f}_\beta^{-,N,T}$, $\mathbb{I} := \mathbb{I}\{f_\beta^-(b) > 0\}$ and $\hat{\mathbb{I}} := \mathbb{I}\{\hat{f}_\beta^-(b) > 0\}$. Then $f_\beta = 2f_\beta^- \mathbb{I}$ and $\hat{f}_\beta = 2\hat{f}_\beta^- \hat{\mathbb{I}}$.

$$\begin{aligned}
\|\hat{f}_\beta - f_\beta\|_2^2 &= \int (\hat{f}_\beta(b) - f_\beta(b))^2 d\sigma(b) \\
&= \int_{I(b)=1,\hat{I}(b)=1} (\hat{f}_\beta(b) - f_\beta(b))^2 d\sigma(b) + \int_{I(b)=0,\hat{I}(b)=1} (\hat{f}_\beta(b) - f_\beta(b))^2 d\sigma(b) \\
&\quad + \int_{I(b)=1,\hat{I}(b)=0} (\hat{f}_\beta(b) - f_\beta(b))^2 d\sigma(b) + \int_{I(b)=0,\hat{I}(b)=0} (\hat{f}_\beta(b) - f_\beta(b))^2 d\sigma(b) \\
&:= B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

Obviously

$$B_1 = \int_{I(b)=1,\hat{I}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^2 d\sigma(b)$$

and $B_4 = 0$. Also,

$$B_2 = \int_{I(b)=0,\hat{I}(b)=1} (2\hat{f}_\beta^-(b) - f_\beta(b))^2 d\sigma(b).$$

But given $I(b) = 0$ and $\hat{I}(b) = 1$, $2\hat{f}_\beta^-(b) > 0$, $f_\beta(b) = 0$ and $2f_\beta^-(b) \leq 0$, so replacing f_β with $2f_\beta^-$ in the bracket,

$$B_2 \leq \int_{I(b)=0,\hat{I}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^2 d\sigma(b).$$

Similarly,

$$B_3 = \int_{I(b)=1,\hat{I}(b)=0} (\hat{f}_\beta(b) - 2f_\beta^-(b))^2 d\sigma(b).$$

and given $I(b) = 1$ and $\hat{I}(b) = 0$, $2f_\beta^-(b) > 0$, $\hat{f}_\beta(b) = 0$ and $2\hat{f}_\beta^-(b) \leq 0$, so replacing f_β with $2f_\beta^-$ in the bracket,

$$B_3 \leq \int_{I(b)=0,\hat{I}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^2 d\sigma(b).$$

Overall,

$$\|\hat{f}_\beta - f_\beta\|_2^2 \leq 4\|\hat{f}_\beta^- - f_\beta^-\|_2^2$$

or

$$\|\hat{f}_\beta - f_\beta\|_2 \leq 2\|\hat{f}_\beta^- - f_\beta^-\|_2$$
$$= O_p(N^{\frac{-s}{2s+2d-1}}) \qquad \text{by Lemma 6.6,}$$

which is the desired result. □

Proof of Theorem 3.3. Since $f_\beta^-(b) > 0$ under the assumption of the theorem and the consistency result implied by Theorem 3.2, it is suffices to consider $N^{1/2} s_N^{-1} 2(\hat{f}_\beta^{-,N,T_n}(b) - f_\beta^-(b))$. Now, using the notation introduced in this appendix, write

$$\hat{f}_\beta^{-,N,T_n}(b) - f_\beta^-(b) = \hat{f}_\beta^{-,N,T_n}(b) - M_T(\hat{f}_X, b) + M_T(\hat{f}_X, b) - M_T(f_X, b) + M_T(f_X, b) - f_\beta^-(b)$$
$$:= C_1(b) + C_2(b) + C_3(b).$$

In what follows we show that C_2 and C_3 are asymptotically negligible. Since we are concerned with asymptotic normality at a point b throughout this proof, we sometimes drop the argument "b" from a function symbol when obvious. Note that C_3 is the regularization bias term and $C_3 = \sum_{p=T+1}^{\infty} Q_{2p+1} f_\beta^-$. Therefore under the maintained hypothesis $s > \frac{d-1}{2}$, for a $v < s$

$$\|C_3\|_\infty \leq \left\| \sum_{p=T+1}^{\infty} Q_{2p+1} f_\beta^- \right\|_{2,v}$$
$$\leq c T^{-(s-v)} \|f_\beta^-\|_{2,s} \qquad \text{by Lemma 6.2.}$$

Noting $s_N \asymp T^{d-\frac{1}{2}}$ and letting $v = \frac{d-1}{2} + \epsilon, \epsilon > 0$ in the above result,

$$N^{\frac{1}{2}} s_N^{-1} C_3 = O(N^{\frac{1}{2}} T^{-d+\frac{1}{2}}) O(T^{-(s-\frac{d-1}{2}-\epsilon)})$$
$$= O(N^{\frac{1}{2}+\alpha(s-\frac{d-1}{2}-\epsilon)}).$$

C_3 is negligible if $N^{\frac{1}{2}} s_N^{-1} C_3 = o(1)$, that is,

$$\frac{1}{2} + \alpha(s - \frac{d-1}{2} - \epsilon) < 0$$

or

$$\frac{1}{2s+d-\epsilon} < \alpha.$$

Since $\epsilon > 0$ can be made arbitrarily small, we obtain the condition

$$\frac{1}{2s+d} < \alpha. \tag{6.22}$$

Now we turn to C_2, which can be written as

$$\begin{aligned} C_2 &= \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} Q_{2p+1} \left[\left(\frac{f_X}{\hat{f}_X} - 1 \right) (2R-1) \right] \\ &:= \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} Q_{2p+1} \Delta \end{aligned}$$

Therefore

$$\begin{aligned} \|C_2\|_{2,t}^2 &= \sum_{n=0}^{\infty} \zeta_n^t \|Q_n C_2\|_2^2 \\ &= \sum_{n=0}^{\infty} \zeta_n^t \left\| Q_n \sum_{p=0}^{T} \frac{1}{\lambda_{2p+1}} Q_{2p+1} \Delta \right\|_2^2 \\ &= \sum_{p=0}^{\infty} \zeta_{2p+1}^t \left\| \frac{1}{\lambda_{2p+1}} Q_{2p+1} \Delta \right\|_2^2 \\ &= \sum_{p=0}^{\infty} \frac{\zeta_{2p+1}^t}{\lambda_{2p+1}^2} \|Q_{2p+1}\Delta\|_2^2 \\ &\le \frac{1}{\lambda_{2T+1}^2} \sum_{p=0}^{\infty} \zeta_{2p+1}^t \|Q_{2p+1}\Delta\|_2^2 \\ &\le \frac{1}{\lambda_{2T+1}^2} \sum_{n=0}^{\infty} \zeta_n^t \|Q_n\Delta\|_2^2 \\ &= \frac{\|\Delta\|_{2,t}^2}{\lambda_{2T+1}^2}, \end{aligned}$$

or $\|C_2\|_{2,t} = \frac{\|\Delta\|_{2,t}}{\lambda_{2T+1}}$. Recall that $\lambda_{2T+1} \asymp T^{-d/2}$. By Theorem 2.4 and the results in Tambaca (2001) and Cox and O'Sullivan (1994),

$$\|C_2\|_{2,t} = O(T^{d/2}) O_p(N^{-\frac{\sigma - t - \epsilon}{2\sigma + d - 1}}),$$

$\epsilon > 0$. Let $t = \frac{d-1}{2} + \epsilon$, then

$$\begin{aligned} C_2(b) &\le \|C_2\|_\infty \\ &\le c\|C_2\|_{2,t} \\ &= O(T^{d/2}) O_p(N^{-\frac{\sigma - \frac{d-1}{2} - 2\epsilon}{2\sigma + d - 1}}). \end{aligned}$$

or

$$N^{\frac{1}{2}} s_N^{-1} C_2(b) = O(N^{\frac{1}{2}}) O(T^{-d+\frac{1}{2}}) O(T^{d/2}) O_p(N^{-\frac{\sigma - \frac{d-1}{2} - 2\epsilon}{2\sigma + d - 1}})$$

$$= O(N^{\frac{1}{2}}) O(N^{-\alpha \frac{d-1}{2}}) O_p(N^{-\frac{\sigma - \frac{d-1}{2} - 2\epsilon}{2\sigma + d - 1}}).$$

$C_2(b)$ is negligible if $N^{\frac{1}{2}} s_N^{-1} C_2(b) = o_p(1)$, or

$$\frac{1}{2} - \alpha \frac{d-1}{2} - \frac{\sigma - \frac{d-1}{2} - 2\epsilon}{2\sigma + d - 1} < 0$$

or

$$\frac{2 + \frac{4\epsilon}{d-1}}{2\sigma + d - 1} < \alpha.$$

Since $\epsilon > 0$ can be made arbitrarily small, we obtain the condition

$$\frac{2}{2\sigma + d - 1} < \alpha. \tag{6.23}$$

In sum, $C_2 + C_3$ is asymptotically negligible if α satisfies (6.22) and (6.23), that is,

$$\frac{1}{2s + d} \vee \frac{2}{2\sigma + d - 1} < \alpha.$$

Moreover

$$N^{\frac{1}{2}} s_N^{-1} C_1(b) = N^{1/2} s_N^{-1} (\hat{f}_\beta^-(f_X, b) - M_T(f_X, b)) + o_p(1).$$

holds. Therefore the proof is complete if we can show that $N^{1/2} s_N^{-1} 2(\hat{f}_\beta^-(f_X, b) - M_T(f_X, b))$ converges to the stated distribution in the theorem. Note that

$$N^{1/2} s_N^{-1} \left(\hat{f}_\beta^-(f_X, b) - M_T(f_X, b) \right) = \frac{1}{N^{1/2} s_N} \sum_{i=1}^{N} 2 \left(Z_{Ni} - \mathbb{E}[Z_i] \right).$$

We can show, as in the proof of Theorem 2.5 but using (6.15), that there exists a constant c depending on the dimension such that

$$\operatorname{var}(Z_{N,1}) \geq c N^{\alpha(2d-1)}.$$

Also,

$$\mathbb{E}\left[|Z_{N,1}|^{2+\delta} \right] = \int_{\mathbb{S}^{d-1}} \left| \sum_{p=0}^{T_N} \frac{q_{2p+1,d}(y, x)}{\lambda(2p+1, d)} \right|^{2+\delta} f_X^{-1-\delta}(y) d\sigma(y)$$

$$\leq \left(\frac{C_X}{|\mathbb{S}^{d-1}|} \right)^{1+\delta} \left(\sum_{n=0}^{T_N} \frac{h(2p+1, d)}{|\lambda(2p+1, d)|} \right)^{2+\delta}.$$

Therefore the criterion (6.17) is satisfied if

$$\frac{\delta/2}{1 + \delta/2} > \alpha(2d - 1).$$

Taking δ large enough, we obtain the upper bound for the rate α. An application of the Lyapounov CLT proves the claim. $\square$

References

[1] Athey, S., and Imbens, G. W. (1985): "Discrete Choice Models with Multiple Unobserved Choice Characteristics". Preprint.

[2] Bajari, P., Fox, J. and Ryan, S. (2007): "Linear Regression Estimation of Discrete choice Models with Nonparametric Distribution of Random Coefficients". *American Economic Review, Papers and Proceedings*, 97, 459-463.

[3] Beran, R. and Hall, P. (2007): "Estimating Coefficient Distributions in Random Coefficient Regression". *Annals of Statistics*, 20, 1110-1119.

[4] Briech, R., Chintagunta, P. and Matzkin, R. (1996): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity". Preprint.

[5] Brownstone, D., and Train, K. (1999): "Forecasting New Product Penetration with Flexible Substitution Patterns". *Journal of Econometrics*, 89, 109-129.

[6] Carrasco, M., Florens, J. P., and Renault, E.: "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization", *Handbook of Econometrics,* J.J. Heckman and E.E. Leamer (eds.), vol. 6, North Holland, to appear.

[7] Chesher, A., and Silva Santos, J. M. C.. (2002): "Taste Variation in Discrete Choice Models". *Review of Economic Studies*, 69, 147-168.

[8] Devroye, L., and Gyorfi, L. (1985): *Nonparametric Density Estimation. The L1-View.* Wiley, New York.

[9] do Carmo, M. P. (1976): *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Englewood Cliffs.

[10] Driscoll, J. R., and Healy, D. M. (1994): "Computing Fourier Transforms and Convolutions on the 2-Sphere", *Advances in Applied Mathematics*, 15, 202-250.

[11] Elbers, C. and Ridder, G. (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Models", *Review of Economics Studies*, 49, 403-410.

[12] Erdélyi, A. et al. ed. (1953): *Higher Transcendental Functions*, vol. 1,2 of the *Bateman Manuscript Project.* Mc Graw-Hill, New-York.

[13] Fan, J. (1991): "Asymptotic Normality for Deconvolution Kernel Density Estimators", *Sankhyā*, 53, Series A, 97-110.

[14] Feller, W. (1968): *An Introduction to Probability Theory and Its Applications - Third edition*, vol. 1. New York: Wiley.

[15] Gallot, S., Hulin, D., and Lafontaine, J. (2004): *Riemannian Geometry - Third edition.* Springer, Berlin.

[16] Groemer, H. (1996): *Geometric Applications of Fourier Series and Spherical Harmonics.* Cambridge University Press: Cambridge, *Encyclopedia of Mathematics and its Applications.*

[17] Harding, M. C. and Hausman, J. (2006): "Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Non-linear Least Squares". Preprint.

[18] Heckman, J. J. and Singer, B. (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

[19] Hess, S., Bolduc, D. and and Polak, J. (2005): "Random Covariance Heterogeneity in Discrete Choice Models". Preprint.

[20] Hall, P., Watson, G. S., and Cabrera, J. (1987): "Kernel Density Estimation with Spherical Data", *Biometrika*, 74, 751-62.

[21] Healy, D. M., and Kim, P. T. (1996): "An Empirical Bayes Approach to Directional Data and Efficient Computation on the Sphere", *The Annals of Statistics*, 24, 232-254.

[22] Hendriks, H. (1990): "Nonparametric Estimation of a Probability Density on a Riemannian Manifold Using Fourier Expansions", *The Annals of Statistics*, 18, 832-849.

[23] Hendriks, H. (2003): "Application of Fast Spherical Fourier Transform to Density Estimation", *Journal of Multivariate Analysis*, 84, 209-221.

[24] Hoderlein, S., Klemelä, J., and Mammen, E. (2007) Reconsidering the Random Coefficient Model, *Preprint*.

[25] Ichimura, H., and Thompson, T. S. (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution", *The Journal of Econometrics*, 86, 269-295.

[26] Kim, P. T., and Koo, J. Y. (2000): "Directional Mixture Models and Optimal Esimation of the Mixing Density", *The Canadian Journal of Statistics*, 28, 383-398.

[27] Pelletier, B. (2005): "Kernel Density Estimation on Riemannian Manifolds", *Statistics* & *Probability Letters*, 73, 297-304.

[28] Train, K. E. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520 and CREST, Statistics Laboratory, 3 avenue Pierre Larousse, 92240 Malakoff, France.

E-mail address: eric.gautier@yale.edu

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520.

E-mail address: yuichi.kitamura@yale.edu