Courant Research Centre

‘Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis’

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 225

Tests for qualitative features in the random coefficients model

Fabian Dunker, Konstantin Eckle, Katharina Proksch, Johannes Schmidt-Hieber

April 2017

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Tests for qualitative features in the random coefficients model

Fabian Dunker*
University of Goettingen

Konstantin Eckle†
Ruhr-Universitaet Bochum

Katharina Proksch‡
University of Goettingen

Johannes Schmidt-Hieber§
University of Leiden

Abstract

The random coefficients model is an extension of the linear regression model which allows for additional heterogeneity in the population by modeling the regression coefficients as random variables. Given data from this model, the statistical challenge is to recover information about the joint density of the random coefficients which is a multivariate and ill-posed problem. Because of the curse of dimensionality and the ill-posedness, pointwise nonparametric estimation of the joint density is difficult and suffers from slow convergence rates. Larger features, such as an increase of the density along some direction or a well-accentuated mode can, however, be much easier detected from data by means of statistical tests. In this article, we follow this strategy and construct tests and confidence statements for qualitative features of the joint density, such as increases, decreases and modes. We propose a multiple testing approach based on aggregating single tests which are designed to extract shape information on fixed scales and directions. Using recent tools for Gaussian approximations of multivariate empirical processes, we derive expressions for the critical value. We apply our method to simulated and real data.

Keywords: Random coefficients model; Radon transform; ill-posed problems; Gaussian approximation; mode detection; monotonicity; multiscale statistics; shape constraints.

*CRC Poverty, Equity and Growth, University of Goettingen, Humboldtallee 3, 37073 Goettingen, Germany, fdunker@uni-goettingen.de

†Department of Mathematics, Institute of Statistics, 44780 Bochum, Germany, konstantin.eckle@ruhr-uni-bochum.de

‡Institute for Mathematical Stochastics, Georg-August-University of Goettingen, Goldschmidtstrasse 7, 37077 Goettingen, Germany, kproksc@uni-goettingen.de

§Mathematical Institute of the University of Leiden, Niels Bohrweg 1, 2333 CA Leiden, Netherlands, schmidthieberaj@math.leidenuniv.nl

1 Introduction

In the random coefficients model, n i.i.d. random vectors $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$ are observed, with $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d})$ a d-dimensional vector of design variables and

$$Y_i = \beta_{i,1} X_{i,1} + \beta_{i,2} X_{i,2} + \ldots + \beta_{i,d} X_{i,d}, \quad i = 1, \ldots, n. \tag{1.1}$$

The unobserved random coefficients $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,d})$, $i = 1, \ldots, n$, are i.i.d. realizations of an unknown d-dimensional distribution $F_{\boldsymbol{\beta}}$ with Lebesgue density $f_{\boldsymbol{\beta}}$. Design variables and random coefficients are assumed to be independent. The statistical task is to recover properties of the joint density $f_{\boldsymbol{\beta}}$, which is assumed to belong to some nonparametric class. In this work, we derive tests for increases and modes of $f_{\boldsymbol{\beta}}$.

For $d = 1$, the random coefficients model simplifies to nonparametric density estimation. For $d > 1$, recovery of $f_{\boldsymbol{\beta}}$ is an inverse problem with ill-posedness depending on the distribution of the design vectors $\mathbf{X}_i$. If the design is sufficiently regular, the inverse problem is mildly ill-posed. Otherwise, the model can be severely ill-posed or even be non-identifiable. In this work, we study the mildly ill-posed regime and consider in particular the random coefficients model with random intercept

$$Y_i = \beta_{i,1} + \beta_{i,2} X_{i,2} + \ldots + \beta_{i,d} X_{i,d}, \quad i = 1, \ldots, n, \tag{1.2}$$

which can be obtained from (1.1) setting $X_{i,1} = 1$, almost surely.

Random coefficients models appear in econometrics and epidemiology and are used to model heterogeneity in the population. While the standard linear regression model accounts for heterogeneity only by an intercept that varies across the population, the random coefficients model allows in addition that different individuals have different slopes. Applications in epidemiology are considered by Greenland [27]; Gustafson and Greenland [28]. In economics, random coefficients models are frequently used for the evaluation of panel data, cf. Hsiao [33] or Hsiao and Pesaran [34], Chapter 6, for an overview. Modeling and estimating consumer demand in industrial organization and marketing often makes use of random coefficients [7; 40; 39; 8; 14]. In all these works, parametric assumptions on $f_{\boldsymbol{\beta}}$ are imposed. Recently, nonparametric approaches to deal with the presence of random coefficients became popular in microeconometrics [31; 37; 30; 18], frequently combined with binary choice [35; 24; 25; 38; 17; 22, among others].

The random coefficients model also includes quantum homodyne tomography. In this case, we observe Φ_i and

$$Y_i = Q_i \cos(\Phi_i) + P_i \sin(\Phi_i), \quad i = 1, \ldots, n, \tag{1.3}$$

with (Q_i, P_i) i.i.d. random variables which are unobserved and independent of Φ_i. The Φ_i can be chosen by the experimenter and are typically uniform on $[0, \pi]$. The interest is in reconstruction of the Wigner function which takes the role of the joint density of (Q_i, P_i). Because P_i and Q_i are not jointly observable, the Wigner function can take negative values. For more on quantum homodyne tomography and the Wigner function, see Butucea et al. [10].

We propose a nonparametric test for shape information of the joint density $f_{\boldsymbol{\beta}}$ in the random coefficients model. The focus will be on a test for directional derivatives and modes. The nonparametric estimation theory for $f_{\boldsymbol{\beta}}$ has been developed in Beran and Hall [6]; Beran et al. [5]; Feuerverger and Vardi [21]; Hoderlein et al. [31]. Due to the ill-posedness of the problem and the curse of dimensionality induced by d, pointwise estimation rates are slow. The reason is that small perturbations in the signal are indistinguishable from the data. Nevertheless, we can get good detection rates for larger features, such as an accentuated mode or a strong increase in the joint density along some direction. From a practical point of view, the relevant information regarding an unknown density is typically its shape rater than its precise, full reconstruction. As a practitioner, one would like to recover increases/decreases and the modes of a density. If, say, two modes in the joint density of two random quantities are detected, this indicates that two different groups can be identified. Hence, shape information allows to interpret a given dataset.

Larger features of the density will also be discovered by a nonparametric estimator even if it suffers from slow pointwise convergence. There are, however, two important reasons, why a testing approach might be more appropriate. Firstly, with a significance test of level α we can conclude that with probability $1 - \alpha$ a detected feature is not an artifact. Secondly, for an estimator we need to pick one bandwidth or smoothing parameter while detection of different features might require different bandwidth choices depending on the size of the hidden features themselves. Indeed, a short and steep increase will be best detected on a small scale whereas for finding a longer and less strong increase the choice of a larger bandwidth is beneficial. Using multiple testing methods, it is possible to combine a whole range of smoothness parameters into one test and to adapt to different shapes of features.

We construct a so called multiscale test, aggregating single tests on different scales and directions. Multiscale tests can be viewed as a multiple testing procedure specifically designed for non-parametric models. Given a model, the theoretical challenge is to prove that a multiscale statistic can be approximated by a distribution free statistic which is independent of the observations. This allows us then to compute quantiles and to find approximations for the critical values of the multiscale statistic. So far, qualitative feature detection based on multiscale statistics has been studied for various nonparametric models, including the

Gaussian white noise model [15], density estimation [16] and deconvolution [42]. All these results are univariate and based on empirical process theory. In a multivariate and non-Gaussian setting we are facing the problem that the classical KMT construction is unfit for multiscale problems as it imposes rather strong conditions. Instead, very recent results on Gaussian approximations of suprema of multivariate empirical processes developed by Chernozhukov et al. [11] can be used [19; 41, in the context of multivariate deconvolution and multivariate linear inverse problems with additive noise, respectively]. In this work, we extend these techniques. The main difficulties are twofold. First, we need to derive specific properties of the inverse Radon transform for general dimension d. Second, in contrast to the other works on multiscale inference, no distribution free approximation can be obtained and we therefore need to study the approximating process if several unobserved functions are replaced by estimators.

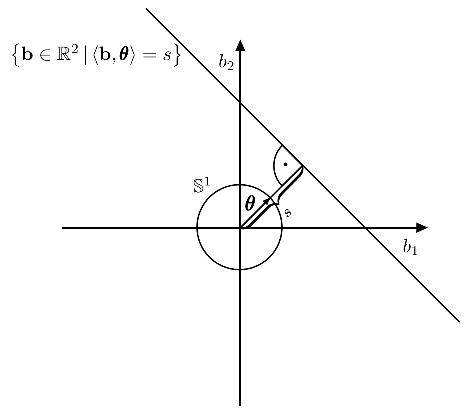
In order to study the power of the multiscale test, a theoretical detection bound and numerical simulations are provided. The theoretical result gives conditions under which a mode will be detected. In a numerical simulation study, we investigate the power of the test for increases/decreases along some direction and mode detection in dependence on the sample size and the design variables. We also analyze real consumer demand data from the British Family Expenditure Survey.

Let us briefly summarize related literature on testing in the random coefficients model. Under a parametric assumption on the density $f_{\boldsymbol{\beta}}$, Beran [4] considers goodness-of-fit testing and Swamy [43]; Andrews [2] test whether some of the random coefficients are deterministic. The only test based on a nonparametric assumption was proposed recently by Breunig and Hoderlein [9]. It allows to assess whether a given set of data follows the random coefficients model.

This paper is organized as follows. In Section 2, we describe the connection between the random coefficients model and the Radon transform. Rewriting the model as an inverse problem in terms of the Radon transform reveals the ill-posed nature of the model. This allows us to construct and to analyze the multiscale test in Section 3. In this part we also derive the asymptotic theory of the estimator and obtain theoretical detection bounds. In Section 4 the test is studied for simulated data. As a real data example, consumer demand is analyzed in Section 5. Proofs and technicalities are deferred to a supplement. The Python source code is be available online.

Notation: Throughout the paper, vectors are displayed by bold letters, e.g. $\mathbf{X}, \boldsymbol{\beta}$. Inequalities between vectors are understood componentwise. The Euclidean norm on $\mathbb{R}^d$ is denoted by $\|\cdot\|$ and the corresponding standard inner product by $\langle\cdot,\cdot\rangle$. We further denote by $\mathbf{e}_1, \ldots, \mathbf{e}_d \in \mathbb{R}^d$ the standard ON-basis of the d-dimensional Euclidean space, $\mathbb{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$ and we write $\mathcal{Z}$ for the cylinder $\mathcal{Z} = \mathbb{R} \times \mathbb{S}^{d-1}$. Furthermore, we write $\mathbf{v}$ for any direction $\mathbf{v} = \sum_{j=1}^{d} \mathrm{v}_j \mathbf{e}_j \in \mathbb{S}^{d-1}$. For two positive sequences $(a_n)_n$, $(b_n)_n$, $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ mean that for some positive constant C, $a_n \leq C b_n$ for all n. As usual, we write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

2 The random coefficients model as an inverse problem



The random coefficients model can be written in terms of the Radon transform [cf. 5]. This allows us then to interpret the model as an inverse problem. In this section, we summarize the main steps and review relevant results on the inversion of the Radon transform. Let H^s denote the L^2-Sobolev space. The Radon transform is the operator $R : H^s(\mathbb{R}^d) \to H^{s+\frac{d-1}{2}}(\mathcal{Z})$, with

$$Rf(s, \boldsymbol{\theta}) = \int_{\langle \mathbf{b}, \boldsymbol{\theta} \rangle = s} f(\mathbf{b}) \, d\mu_{d-1}(\mathbf{b})$$

and μ_{d-1} the surface measure on the $(d-1)$-dimensional hyperplane $\{\mathbf{b} \in \mathbb{R}^d : \langle \mathbf{b}, \boldsymbol{\theta} \rangle = s\}$. The Radon transform maps therefore a function to all its integrals over hyperplanes parametrized by $(s, \boldsymbol{\theta}) \in \mathcal{Z}$. The figure above shows the parametrization in two dimensions.

For the connection between the Radon transform and the random coefficients model (1.1) we consider the normalized observations

$$S_i := \frac{Y_i}{\|\mathbf{X}_i\|}, \quad \boldsymbol{\Theta}_i := \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}, \quad i = 1, \dots, n.$$

The random vectors $\boldsymbol{\Theta}_i$ take values in the $(d-1)$-dimensional sphere $\mathbb{S}^{d-1}$. In the random coefficients model with intercept (1.2), $\boldsymbol{\Theta}_i$ is always in the upper hemisphere, i.e. the first component of $\boldsymbol{\Theta}_i$ is positive. In this case, we extend the distribution of $\boldsymbol{\Theta}_i$ to the whole sphere by randomizing the signs of the design variables. For this purpose, we generate independent random variables ζ_i, $i = 1, \dots, n$, with $\mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i = -1) = 1/2$, which are independent of the data $(\mathbf{X}_i, Y_i)$, $i = 1, \dots, n$, and define $S_i := \zeta_i Y_i / \|\mathbf{X}_i\|$ and $\boldsymbol{\Theta}_i := \zeta_i \mathbf{X}_i / \|\mathbf{X}_i\|$. Independent of the symmetrization, we have $S_i = \langle \boldsymbol{\Theta}_i, \boldsymbol{\beta}_i \rangle$. The conditional distribution of $S_1 | \boldsymbol{\Theta}_1$ is therefore

$$F_{S|\boldsymbol{\Theta}}(x|\boldsymbol{\theta}) = \mathbb{P}(S_1 \leq x | \boldsymbol{\Theta}_1 = \boldsymbol{\theta}) = \mathbb{P}(\langle \boldsymbol{\Theta}_1, \boldsymbol{\beta}_1 \rangle \leq x | \boldsymbol{\Theta}_1 = \boldsymbol{\theta}) = \int_{-\infty}^{x} (Rf_{\boldsymbol{\beta}})(s, \boldsymbol{\theta}) \, ds,$$

and the conditional density becomes

$$f_{S|\boldsymbol{\Theta}}(s|\boldsymbol{\theta}) = (Rf_{\boldsymbol{\beta}})(s, \boldsymbol{\theta}). \tag{2.1}$$

Recall that we have access to an i.i.d. sample $(S_i, \boldsymbol{\Theta}_i)$ of the joint density $f_{S,\boldsymbol{\Theta}}$. This allows for nonparametric estimation of $f_{S|\boldsymbol{\Theta}}$ using the fact that $f_{S|\boldsymbol{\Theta}}(s|\boldsymbol{\theta}) = f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})/f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. Applying the inverse Radon transform to this estimate gives an estimator for the joint density $f_{\boldsymbol{\beta}}$. This inversion scheme suffers from two sources of ill-posedness. Firstly, dividing by $f_{\boldsymbol{\Theta}}$ might result in very unstable reconstructions if $f_{\boldsymbol{\Theta}}$ is small. This can be avoided by imposing regularity on the distribution of the design variables $\mathbf{X}_i$. If this regularity is violated and we systematically miss observations from some directions, the problem becomes unevenly harder and only logarithmic convergence rates can be obtained [see 12; 23; 32]. When the support of $\boldsymbol{\Theta}_i$ does not contain an open ball $f_{\boldsymbol{\beta}}$ might be non-identifiable. Secondly, even with regularity on the distribution of the design, the Radon inversion is known to be an ill-posed problem with degree of ill-posedness $(d-1)/2$. Hence, regularization of the inversion scheme is necessary.

In this work, we study the mildly ill-posed case where the random directions $\boldsymbol{\Theta}_i$, $i = 1, \dots, n$, are sufficiently regularly distributed over the sphere and the ill-posedness is only due to the inversion of the Radon transform. The precise assumptions on the design are stated in Section 3.2.

Our approach makes use of the following explicit inversion formula of the Radon transform. Define the operator Λ via

$$\Lambda f(s, \boldsymbol{\theta}) = \mathcal{H}_d \partial_s^{d-1} Rf(s, \boldsymbol{\theta}), \tag{2.2}$$

where $\mathcal{H}_d$ denotes the identity for d odd and the Hilbert transform with respect to the variable s for d even. Let $c_d^{-1} = (-1)^{(d-1)/2} 2^{-d} \pi^{1-d}$ for d odd and $c_d^{-1} = -(-1)^{d/2} 2^{-d} \pi^{1-d}$ for d even. Let φ be a Schwartz function on $\mathbb{R}^d$. Then we have the inversion formula

$$\varphi = c_d^{-1} R^* \Lambda \varphi, \tag{2.3}$$

cf. Theorem 3.8 in Helgason [29]. Here R^* is the adjoint of the Radon transform with respect to the L^2 scalar product. It is also called the back projection operator. Notice that our constant c_d differs from the constant in [29] as we use the standard definition of the Hilbert transform and defined R^* as the adjoint of the Radon transform (as opposed to the dual transform).

3 Multiscale tests for qualitative features

3.1 Multiscale inference

The goal of this work is to derive confidence statements for qualitative features of the joint density of the random coefficients. In particular, we are interested in the detection of modes (local maxima) of the density. Following the approach of Schmidt-Hieber et al. [42], we express the features in terms of differential operators. To be precise, for a collection of compactly supported, non-negative and sufficiently smooth test functions $\phi_{\mathbf{t},h}$ consider the integral

$$\int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b})\, d\mathbf{b} = \sum_{k=1}^{d} \mathrm{v}_k \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \frac{\partial}{\partial \mathrm{b}_k} f_{\boldsymbol{\beta}}(\mathbf{b})\, d\mathbf{b} \tag{3.1}$$

for some directional vector $\mathbf{v} = (\mathrm{v}_1, \ldots, \mathrm{v_d})^\top$ and $d \geq 2$. Since there should not be any favored direction, we consider in the following radially symmetric test functions,

$$\phi_{\mathbf{t},h}(\cdot) = \frac{1}{h^d \operatorname{Vol}(\mathbb{S}^{d-2})} \phi\Big(\frac{\|\cdot - \mathbf{t}\|}{h}\Big) \tag{3.2}$$

with a non-negative and sufficiently smooth kernel $\phi : [0, \infty) \to [0, \infty)$ with $\int_0^\infty \phi(u)du = 1$ and support on $[0, 1]$. $\operatorname{Vol}(\mathbb{S}^{d-2})$ denotes the volume of the sphere $\mathbb{S}^{d-2} \subset \mathbb{R}^{d-1}$ and $\operatorname{Vol}(\mathbb{S}^0) := 2$. Notice that $\phi_{\mathbf{t},h}$ is supported on the ball $B_h(\mathbf{t})$ with center $\mathbf{t}$ and radius h. The normalization for $\phi_{\mathbf{t},h}$ does not entail that $\phi_{\mathbf{t},h}$ integrates to one but turns out to be convenient.

If the integral (3.1) is positive, there exists a subset of $B_h(\mathbf{t})$ with positive Lebesgue measure on which $\partial_{\mathbf{v}} f_{\boldsymbol{\beta}}$ is positive. On this subset, $f_{\boldsymbol{\beta}}$ is thus strictly increasing in direction $\mathbf{v}$. Similarly, we can recover a decrease if the integral (3.1) is negative. To construct a statistical test for increases and decreases it is therefore natural to use an empirical counterpart of the functional defined in (3.1).

Let $\mathcal{T} = \{(\mathbf{t}, h, \mathbf{v}) : h \in (0, 1], h \leq \mathbf{t} \leq 1 - h, \mathbf{v} \in \mathbb{S}^{d-1}\}$, where the inequalities for the vector $\mathbf{t}$ are understood componentwise. For statistical inference regarding the sign of the directional derivatives of $f_{\boldsymbol{\beta}}$, we fix a subset $\mathcal{T}_n \subset \mathcal{T}$ and test for all $(\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n$ simultaneously the corresponding hypotheses of the form

$$H_{0,+}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} \leq 0 \quad \text{versus} \quad H_{1,+}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} > 0 \tag{3.3}$$

and

$$H_{0,-}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} \geq 0 \quad \text{versus} \quad H_{1,-}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} < 0. \tag{3.4}$$

We would like to point out that all ideas and all asymptotic properties that are presented in the following apply if $h \leq \mathbf{t} \leq 1 - h$ is replaced by $\mathbf{a} + h \leq \mathbf{t} \leq \mathbf{b} - h$ for arbitrary, fixed vectors $\mathbf{a} \leq \mathbf{b} \in \mathbb{R}^d$ in the definition of the set $\mathcal{T}$. For ease of notation, we set $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = \mathbf{1}$ to derive the theory. In the following, we adopt some of the arguments from Eckle et al. [19] to give interesting applications for tests of the form (3.3) and (3.4).

Simultaneous tests for these hypotheses can be used to obtain a graphical representation of the local monotonicity behavior of a bivariate density. To this end, define a subset $\mathcal{T}_n = \{(\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j) : j = 1, \ldots, p\}$ for a fixed scale h_0 of the form $\mathcal{T}_n = \mathcal{T}_{\mathbf{t}} \times \{h_0\} \times \mathcal{T}_{\mathbf{v}}$, where $\mathcal{T}_{\mathbf{t}}$ contains the $p/|\mathcal{T}_{\mathbf{v}}|$ vertices of an equidistant grid of width $2h_0$ and $\mathcal{T}_{\mathbf{v}}$ contains the directions. We consider four equidistant directions on $\mathbb{S}^1$ given by $\mathcal{T}_{\mathbf{v}} = \{\mathbf{v}_1, -\mathbf{v}_1, \mathbf{v}_2, -\mathbf{v}_2\}$. Since $\mathcal{T}_{\mathbf{v}} = -\mathcal{T}_{\mathbf{v}}$ we have symmetry in the hypotheses, i.e. $H_{0,+}^{\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j} = H_{0,-}^{\widetilde{\mathbf{t}}_j, h_0, -\widetilde{\mathbf{v}}_j}$ and we therefore test only $H_{0,+}^{\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j}$ for all triples $(\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j) \in \mathcal{T}_n$. Figure 1 displays an example for the test outcome with the hypotheses in (3.3) and $\mathcal{T}_n$ above. An arrow in a direction $\widetilde{\mathbf{v}}_j$ at a location $\widetilde{\mathbf{t}}_j$ represents a rejection of the corresponding hypothesis $H_{0,+}^{\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j}$ and provides therefore an indication of a positive directional derivative of $f_{\boldsymbol{\beta}}$ in direction $\widetilde{\mathbf{v}}_j$ at the location $\widetilde{\mathbf{t}}_j$. A detailed description of the settings used to generate Figure 1 and an analysis of the results is given in Section 4.1.

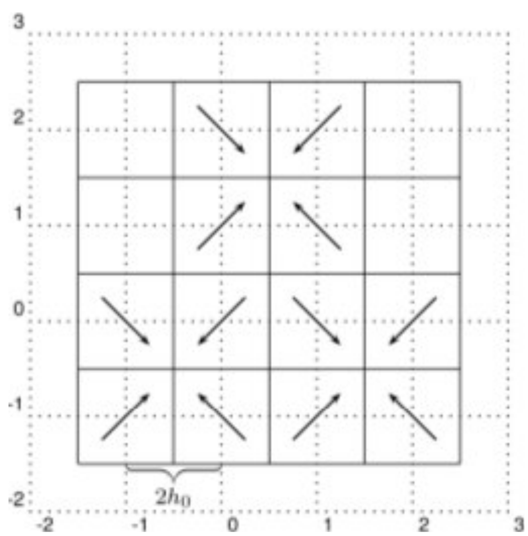


Figure 1: *Example of a global map for monotonicity of a bivariate density.*

A second application of tests for the hypotheses (3.3) and (3.4) is the detection of specific shape constraints such as a mode at a given point $\mathbf{b}_0 \in \mathbb{R}^d$. For this purpose, we fix a bandwidth h and consider pairs $(\mathbf{t}_1, \mathbf{v}_1), \ldots, (\mathbf{t}_p, \mathbf{v}_p)$, where $\mathbf{v}_j$, $j = 1, \ldots, p$, are directional vectors and the test locations $\mathbf{t}_j$ are points on the line $\{\mathbf{b}_0 + r\mathbf{v}_j : r \geq h\}$ $(j = 1, \ldots, p)$ in a neighborhood of $\mathbf{b}_0$. Inference for the presence of a mode at the point $\mathbf{b}_0$ can now be conducted by investigating the hypotheses

$$H_{0,-}^{\mathbf{t}_j,h,\mathbf{v}_j} \quad \text{versus} \quad H_{1,-}^{\mathbf{t}_j,h,\mathbf{v}_j} \quad (j=1,\ldots,p). \tag{3.5}$$

In order to obtain a test procedure which is more flexible in terms of the detection of modes of unknown or varying accentuations we can also include more bandwidths h, defining neighborhoods of $\mathbf{b}_0$ of different sizes. This approach is of particular importance for densities which have a second mode close to the test location $\mathbf{b}_0$. In Section 4.2 we illustrate an example of a bimodal density in which two separate tests for the hypotheses (3.5) with different bandwidths fail to detect the mode, whereas their multiscale aggregation succeeds.

We now derive an empirical counterpart of the functional (3.1) in terms of the Radon transform $Rf_{\boldsymbol{\beta}}$. In order to exploit identity (2.1) we make the following assumptions.

Assumption 1. *Suppose that the function ϕ in (3.2) is $(d+2)$-times continuously differentiable with $\phi'(0)=\phi''(0)=0$.*

Assumption 2. *Suppose that the density $f_{\boldsymbol{\beta}}$ is compactly supported, continuously differentiable and bounded from below in the test region by a constant $c_{\boldsymbol{\beta}}>0$*

$$f_{\boldsymbol{\beta}}(\mathbf{b}) \geq c_{\boldsymbol{\beta}} \quad \textit{for all} \quad \mathbf{b} \in [0,1]^d.$$

Remark 1. *Assumption 2 provides basic and intuitive assumptions on the density $f_{\boldsymbol{\beta}}$. They are however too restrictive for quantum homodyne tomography (model (1.3)), where the density $f_{\boldsymbol{\beta}}$ is given by the Wigner function. The Wigner function can take negative values and is not compactly supported. In the example of quantum homodyne tomography, we therefore replace Assumption 2 by the following conditions. Suppose that $f_{\boldsymbol{\beta}}$ is continuously differentiable and, for some $\gamma,\varepsilon>0$,*

(i) $|f_{\boldsymbol{\beta}}(\mathbf{b})| \lesssim \|\mathbf{b}\|^{-d-\varepsilon}$ *for all* $\mathbf{b}\in\mathbb{R}^d$*;*

(ii) $|f_{\boldsymbol{\beta}}(\mathbf{b}) - f_{\boldsymbol{\beta}}(\mathbf{b}')| \lesssim \dfrac{\|\mathbf{b}-\mathbf{b}'\|^{\gamma}}{(1+\min\{\|\mathbf{b}\|,\|\mathbf{b}'\|\})^{d+\gamma+\varepsilon}}$ *for all* $\mathbf{b},\mathbf{b}'\in\mathbb{R}^d$*;*

(iii) There exist constants $\delta, c_{\boldsymbol{\beta}}>0$ such that for every hyperplane $P\subset\mathbb{R}^d$ with $P\cap[-\delta,1+\delta]^d\neq\emptyset$ it holds that

$$\int_P f_{\boldsymbol{\beta}}(\mathbf{b})d\mu_{d-1}(\mathbf{b}) \geq c_{\boldsymbol{\beta}}.$$

Under the assumptions above the inversion formula (2.3) holds for partial derivatives of the test functions $\partial_{\mathbf{v}}\phi_{\mathbf{t},h}$. This is a direct consequence of Theorem 3.8 in Helgason [29]. The following lemma analyzes the structure of a partial derivative of the test function transformed by the operator Λ introduced in (2.2) and how this transform depends on h.

Lemma 3.1. *Work under Assumption 1 and let*

$$\widetilde{\phi}(z) := \int_0^\infty r^{d-2} \frac{\partial}{\partial z} \phi\left(\sqrt{z^2 + r^2}\right) dr \qquad \textit{with } z \in \mathbb{R}. \tag{3.6}$$

Then

$$\Lambda(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})(s, \boldsymbol{\theta}) = \frac{\langle \boldsymbol{\theta}, \mathbf{v} \rangle}{h^{d+1}} (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta} \rangle}{h} \Big),$$

where Λ *and* $\phi_{\mathbf{t},h}$ *are defined in* (2.2) *and* (3.1)*, respectively. Moreover,*

(i) $\|\Lambda(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})\|_\infty \lesssim h^{-d-1}$*;*

(ii) $\|\Lambda(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})\|^k_{L^k(\mathcal{Z})} \lesssim h^{-dk-k+1}$ *for* $k > 1$*.*

For a given triple $(\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}$ we study the statistic

$$T_{\mathbf{t},h,\mathbf{v}} := \frac{1}{n\sqrt{h}} \sum_{i=1}^n \frac{\langle \boldsymbol{\Theta}_i, \mathbf{v} \rangle}{f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}_i)} (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{S_i - \langle \mathbf{t}, \boldsymbol{\Theta}_i \rangle}{h} \Big).$$

By Lemma 3.1, the expectation of this statistic can be written as

$$\mathbb{E}\big[T_{\mathbf{t},h,\mathbf{v}}\big] = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} h^{d+1/2} \Lambda(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})(s, \boldsymbol{\theta}) f_{S|\boldsymbol{\Theta}}(s|\boldsymbol{\theta}) ds d\boldsymbol{\theta},$$

where $d\boldsymbol{\theta}$ denotes the surface measure on $\mathbb{S}^{d-1}$, i.e. $|S_{\boldsymbol{\Theta}}| = \int_{S_{\boldsymbol{\Theta}}} d\boldsymbol{\theta}$ for any measurable $S_{\boldsymbol{\Theta}} \subseteq \mathbb{S}^{d-1}$. By an application of the inversion formula introduced in (2.3) and Lemma 5.1 in Helgason [29], we obtain

$$\mathbb{E}\big[T_{\mathbf{t},h,\mathbf{v}}\big] = -c_d h^{d+1/2} \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b}. \tag{3.7}$$

Up to rescaling, $T_{\mathbf{t},h,\mathbf{v}}$ is thus an empirical counterpart of the functional defined in (3.1).

The statistic $T_{\mathbf{t},h,\mathbf{v}}$ depends on the density $f_{\boldsymbol{\Theta}}$. In quantum homodyne tomography this density is known. However, for many other applications $f_{\boldsymbol{\Theta}}$ needs to be estimated from the data. In this case, we assume that we have an estimator $\widetilde{f}_{\boldsymbol{\Theta}}$ for $f_{\boldsymbol{\Theta}}$ based on an additional sample $(S_i, \boldsymbol{\Theta}_i)$, $i = n+1, \dots, 2n$ which is independent of $(S_i, \boldsymbol{\Theta}_i)$, $i = 1, \dots, n$. We replace $f_{\boldsymbol{\Theta}}$ by its estimator and consider the test statistic

$$\widehat{T}_{\mathbf{t},h,\mathbf{v}} := \frac{1}{n\sqrt{h}} \sum_{i=1}^n \frac{\langle \boldsymbol{\Theta}_i, \mathbf{v} \rangle}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}_i)} (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{S_i - \langle \mathbf{t}, \boldsymbol{\Theta}_i \rangle}{h} \Big). \tag{3.8}$$

3.2 Assumptions on the design

As mentioned in Section 2, the inverse problem might become severely ill-posed or non-identifiable if the density $f_{\mathbf{\Theta}}$ approaches zero for some directions. This section provides conditions on the design which ensure that $f_{\mathbf{\Theta}}$ has Hölder smoothness and is bounded from below and above. These results are of independent interest.

In the random coefficients model (1.1), the density $f_{\mathbf{\Theta}}$ can be expressed in terms of the density $f_{\mathbf{X}}$ via $f_{\mathbf{\Theta}}(\boldsymbol{\theta}) = \int_0^\infty r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta})dr$. This provides a simple relationship to check whether $f_{\mathbf{\Theta}}$ is bounded from below for a specific design.

The special case of the random coefficients model with intercept (1.2) is more restrictive. In this case, we write $f_{\mathbf{X}}$ as a function of $\boldsymbol{x} = (x_2, \ldots, x_d) \in \mathbb{R}^{d-1}$. This yields the explicit formula

$$f_{\mathbf{\Theta}}(\boldsymbol{\theta}) = \frac{1}{2|\theta_1|^d} f_{\mathbf{X}}\Big(\frac{\theta_2}{\theta_1}, \ldots, \frac{\theta_d}{\theta_1}\Big), \tag{3.9}$$

see (4.1) for a proof. A necessary condition for the assumption that $\inf_{\boldsymbol{\theta}} f_{\mathbf{\Theta}}(\boldsymbol{\theta}) > 0$ is given by $f_{\mathbf{X}}(\boldsymbol{x}) \gtrsim \|\boldsymbol{x}\|^{-d}$ as $\|\boldsymbol{x}\| \to \infty$. This corresponds to Cauchy-type tails of the design variables. Thinner tails will increase the ill-posedness of the problem. To avoid very technical proofs, in the random coefficients model with intercept we consider only the case where $(X_{i,2}, \ldots, X_{i,d})$ follows a multivariate Cauchy distribution, i. e.,

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{\Gamma(d/2)}{\pi^{d/2}|\Sigma|^{1/2}\big(1 + (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)^{d/2}} \quad \text{for} \quad \boldsymbol{x} \in \mathbb{R}^{d-1}, \tag{3.10}$$

with $\boldsymbol{\mu} \in \mathbb{R}^{d-1}$ and $\Sigma \in \mathbb{R}^{(d-1)\times(d-1)}$ a symmetric and positive definite matrix. We can compute $f_{\mathbf{\Theta}}$ explicitly using (3.9)

$$f_{\mathbf{\Theta}}(\boldsymbol{\theta}) = \frac{\Gamma(d/2)}{2\pi^{d/2}|\Sigma|^{1/2}\big(\theta_1^2 + ((\operatorname{sgn}(\theta_1)\theta_j - |\theta_1|\mu_j)_{j=2}^d)^\top \Sigma^{-1}((\operatorname{sgn}(\theta_1)\theta_j - |\theta_1|\mu_j)_{j=2}^d)\big)^{d/2}}, \tag{3.11}$$

where $\operatorname{sgn}(\cdot)$ denotes the signum function. In this case, $f_{\mathbf{\Theta}}$ is bounded from above and below and continuously differentiable on the hemispheres $\mathbb{S}_+^{d-1} := \{\boldsymbol{\theta} \in \mathbb{S}^{d-1} \mid \theta_1 > 0\}$ and $\mathbb{S}_-^{d-1} := \{\boldsymbol{\theta} \in \mathbb{S}^{d-1} \mid \theta_1 < 0\}$. In particular, if $(X_{i,2}, \ldots, X_{i,d})$ is standard Cauchy, then the density $f_{\mathbf{\Theta}}$ is constant. In the general case, we make the following assumption.

Assumption 3. *Suppose that either*

(i) $f_{\mathbf{X}}(\boldsymbol{x}) \lesssim \|\boldsymbol{x}\|^{-d-\varepsilon}$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$ *and some* $\varepsilon > 0$*;*

(ii) $\displaystyle\int_0^\infty r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta})dr \geq c > 0$ *for all* $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$*;*

(iii) $|f_{\mathbf{X}}(\boldsymbol{x}) - f_{\mathbf{X}}(\boldsymbol{x}')| \lesssim \dfrac{\|\boldsymbol{x}-\boldsymbol{x}'\|^\gamma}{1+\|\boldsymbol{x}\|^{d+\gamma+\varepsilon}}$ *for* $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$*,* $\|\boldsymbol{x}\| = \|\boldsymbol{x}'\|$*, and* $\gamma > 0$*;*

or model (1.2) *holds and* $f_{\mathbf{X}}$ *is of the form* (3.10).

3.3 Asymptotic properties

This section presents the main theoretical result of the paper stating that the standardized and properly calibrated test statistic (3.8) can be uniformly approximated by a maximum of a Gaussian process. For that we need the definition of a Gaussian process on the cylinder $\mathcal{Z}$. To this end, let $\mathcal{B}(\mathcal{Z})$ be the Borel σ-algebra on $\mathcal{Z}$. Define the σ-finite measure

$$\nu : \begin{cases} \mathcal{B}(\mathcal{Z}) & \to \mathbb{R}_0^+, \\ E & \mapsto \nu(E) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \mathbb{1}_E(\boldsymbol{\theta}, s)\, ds\, d\boldsymbol{\theta}. \end{cases}$$

Let $\big(\mathcal{B}(\mathcal{Z})\big)_\nu$ denote the collection of all sets of finite ν-measure and let W denote Gaussian ν-noise on $\big(\mathcal{B}(\mathcal{Z}), \nu\big)$. For disjoint sets $E_1, E_2 \in \big(\mathcal{B}(\mathcal{Z})\big)_\nu$ this implies

$$W(E_1) \sim \mathcal{N}(0, \nu(E_1)), \quad W(E_1 \cup E_2) = W(E_1) + W(E_2) \text{ a.s.} \quad \text{and} \quad W(E_1) \perp W(E_2)$$

[1, Chapter 1.4.3]. W is a random, finitely additive, signed measure. Integration w.r.t. W can be defined similarly to Lebesgue-integration, starting with a definition for simple functions and an extension to general $f \in L^2(\nu)$ via approximation by simple functions in the L^2-limit. Integration with respect to W yields

$$\int_E W(dsd\boldsymbol{\theta}) = W(E) \sim \mathcal{N}(0, \nu(E)) \quad \text{for} \quad E \in \big(\mathcal{B}(\mathcal{Z})\big)_\nu,$$

$$W(f) := \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} f(s, \boldsymbol{\theta})\, W(dsd\boldsymbol{\theta}) \sim \mathcal{N}(0, \|f\|_{L^2(\nu)}) \quad \text{for} \quad f \in L^2(\nu),$$

and $\mathrm{Cov}(W(f), W(g)) = \langle W(f), W(g)\rangle_{L^2(\mathbb{P})} = \langle f, g\rangle_{L^2(\nu)}$ for $f, g \in L^2(\nu)$, where $L^k(\mathbb{P})$ denotes the collection of all random variables whose first k absolute moments exist. For more details, cf. Adler and Taylor [1], Chapter 5.2.

Let us provide some heuristic for the Gaussian approximation of $T_{\mathbf{t},h,\mathbf{v}}$. The process $(\mathbf{t}, h, \mathbf{v}) \mapsto \sqrt{n}(T_{\mathbf{t},h,\mathbf{v}} - \mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}])$ has in the important case $\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}] = 0$ the same mean and covariance structure as the Gaussian process

$$X_{\mathbf{t},h,\mathbf{v}} = h^{-1/2} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v}\rangle (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta}\rangle}{h}\Big) \frac{\sqrt{f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})}}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} W(dsd\boldsymbol{\theta}). \tag{3.12}$$

In the proof of Theorem 3.2 below we show that the expectation $\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]$ is asymptotically negligible in the limit process. The test statistic and the Gaussian process depend, however, on the unknown densities $f_{S,\boldsymbol{\Theta}}$ and $f_{\boldsymbol{\Theta}}$ which have to be estimated from the data. For Gaussian ν-noise W that is independent of the data let

$$\widehat{X}_{\mathbf{t},h,\mathbf{v}} := h^{-1/2} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta} \rangle}{h} \Big) \frac{\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})}}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} W(ds d\boldsymbol{\theta})$$

and

$$\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}} := \Big(\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \frac{\widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} ds d\boldsymbol{\theta} \Big)^{1/2}. \tag{3.13}$$

The Gaussian approximation result for the family of test statistics $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$ holds for a finite subset $\mathcal{T}_n \subset \mathcal{T}$. Its cardinality may, however, grow polynomially of arbitrary degree with the sample size. Moreover, the range of bandwidths must be bounded from above and below by $h_{\max}$ and $h_{\min}$, both converging to zero as n goes to infinity. The precise conditions are summarized in the following assumption.

Assumption 4. *Let $h_{\min} := \min\{h : (\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n\}$ and $h_{\max} := \max\{h : (\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n\}$. Suppose that $|\mathcal{T}_n| = p \lesssim n^L$ for some $L > 0$ and $h_{\max} \lesssim \log(n)^{-14\gamma/(d-1)-5} n^{2\gamma/(d-1)-1}$, $h_{\min} \gtrsim n^{-1+\varepsilon}$ for some $\varepsilon > 0$.*

Let $\mathcal{A}_p$ be the set of half-open hyperrectangles in $\mathbb{R}^p$, i.e. every $A \in \mathcal{A}_p$ has the representation $A = \{\boldsymbol{x} \in \mathbb{R}^p : -\infty < \boldsymbol{x} \leq \boldsymbol{a}\}$ for some $\boldsymbol{a} \in \mathbb{R}^p$. For finite sets S_n and two stochastic processes $(X_{s,n})_{s \in S_n}$ and $(\widetilde{X}_{s,n})_{s \in S_n}$, which are defined on the same probability space, we write

$$(X_{s,n})_{s \in S_n} \leftrightarrow (\widetilde{X}_{s,n})_{s \in S_n}$$

if $\lim_n \sup_{A \in \mathcal{A}_{|S_n|}} |\mathbb{P}((X_{s,n})_{s \in S_n} \in A) - \mathbb{P}((\widetilde{X}_{s,n})_{s \in S_n} \in A)| = 0$.

Theorem 3.2. *For the calibration of the standardized statistic define*

$$\alpha_h := \sqrt{(3d-1)\log(1/h)} \quad \text{and} \quad \beta_h := \frac{\sqrt{\log(e/h)}}{\log(\log(e^e/h))}.$$

Then under Assumptions 1-4,

$$\Big(\beta_h \Big(\sqrt{n} \frac{|\widehat{T}_{\mathbf{t},h,\mathbf{v}} - \mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \alpha_h \Big) \Big)_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} \leftrightarrow \Big(\beta_h \Big(\frac{|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \alpha_h \Big) \Big)_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n}.$$

Furthermore, conditionally on the data, the Gaussian approximation is bounded in probability by a constant that is independent of n, almost surely.

3.4 Construction of the multiscale test

With the previous theorem, we can now construct simultaneous statistical tests for the hypotheses (3.3) and (3.4). If the constant c_d is positive then the method consists of rejecting the hypotheses $H_{0,+}^{\mathbf{t},h,\mathbf{v}}$ in (3.3) for small values of $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$ and rejecting $H_{0,-}^{\mathbf{t},h,\mathbf{v}}$ in (3.4) for large values of $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$, and vice versa if c_d is negative. Theorem 3.2 is used to control the multiple level of the tests. Let $\alpha \in (0,1)$ and denote by $\kappa_n(\alpha)$ the smallest number such that

$$\mathbb{P}\left(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h\left(\frac{|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \alpha_h\right) \le \kappa_n(\alpha)\right) \ge 1-\alpha.$$

By Theorem 3.2, $\kappa_n(\alpha)$ is bounded uniformly with respect to n. Define for $(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n$ the quantiles

$$\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha) = \frac{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}}{\sqrt{n}}\left(\beta_h^{-1}\kappa_n(\alpha) + \alpha_h\right) \tag{3.14}$$

and reject the hypothesis (3.3), if

$$\operatorname{sgn}(c_d)\widehat{T}_{\mathbf{t},h,\mathbf{v}} < -\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha). \tag{3.15}$$

Similarly, the hypothesis (3.4) is rejected, whenever

$$\operatorname{sgn}(c_d)\widehat{T}_{\mathbf{t},h,\mathbf{v}} > \kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha). \tag{3.16}$$

Theorem 3.3. *Work under the assumptions of Theorem 3.2 and assume that the tests* (3.15) *and* (3.16) *for the hypotheses* (3.3) *and* (3.4) *are performed simultaneously for all* $(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n$*. The probability of at least one false rejection of any of the tests is asymptotically at most* α*, that is*

$$\mathbb{P}\Big(\exists(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n : |\widehat{T}_{\mathbf{t},h,\mathbf{v}}| > \kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha)\Big) \le \alpha + o(1)$$

for $n \to \infty$*.*

We further introduce a method for the detection and localization of modes which relies on the local tests for modality proposed in (3.5) for a set of candidate modes. We assume that $\mathcal{T}_n$ fulfills the following condition. For any fixed h and $\mathbf{v}$ the set $\{\mathbf{t} : (\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n\}$ is an equidistant grid in $[h,1-h]^d$ with grid width h. Furthermore, for any fixed $\mathbf{t}$ and h the set $\{\mathbf{v} : (\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n\}$ is a grid in S^{d-1} with grid width converging to zero with increasing sample size.

For some $\mathbf{b}_0 \in (0,1)^d$ let $\mathcal{T}_n^{\mathbf{b}_0}$ be the set of all triples $(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n$ such that $ch \ge \|\mathbf{b}_0 - \mathbf{t}\| \ge 2h$ for some sufficiently large $c > 2$. Furthermore, the angle between $\mathbf{t} - \mathbf{b}_0$ and $\mathbf{v}$ must go to 0 when $n \to \infty$. We decide for a mode at the point $\mathbf{b}_0$ if for all triples in $\mathcal{T}_n^{\mathbf{b}_0}$ all local tests (3.16) for the hypotheses (3.5) reject. By choosing the test locations as the vertices of an equidistant grid no prior knowledge about the location of $\mathbf{b}_0$ has to be assumed. Theorem 3.4 below states that the procedure detects all modes of the density with probability converging to one as $n \to \infty$.

Theorem 3.4. *Work under the assumptions of Theorem 3.2 and assume that for any mode $\mathbf{b}_0 \in (0,1)^d$ there are functions $g_{\mathbf{b}_0} : \mathbb{R}^d \to \mathbb{R}$, $\widetilde{f}_{\mathbf{b}_0} : \mathbb{R} \to \mathbb{R}$ such that the density $f_{\boldsymbol{\beta}}$ has a representation of the form*

$$f_{\boldsymbol{\beta}}(\mathbf{b}) \equiv (1 + g_{\mathbf{b}_0}(\mathbf{b}))\widetilde{f}_{\mathbf{b}_0}(\|\mathbf{b} - \mathbf{b}_0\|) \tag{3.17}$$

in an open neighborhood of $\mathbf{b}_0$. Furthermore, let $g_{\mathbf{b}_0}$ be differentiable in an open neighborhood of $\mathbf{b}_0$ with $g_{\mathbf{b}_0}(\mathbf{b}) = o(1)$ and $\langle \nabla g_{\mathbf{b}_0}(\mathbf{b}), \mathbf{e}\rangle = o(\|\mathbf{b} - \mathbf{b}_0\|)$ when $\mathbf{b} \to \mathbf{b}_0$ for all $\mathbf{e} \in \mathbb{R}^d$ with $\|\mathbf{e}\| = 1$. In addition, let $\widetilde{f}_{\mathbf{b}_0}$ be differentiable in an open neighborhood of zero with $\widetilde{f}'_{\mathbf{b}_0}(h) \leq -ch(1 + o(1))$ for $h \to 0$.

If the set

$$\left\{(\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n : h \geq C \log(n)^{1/(2d+3)} n^{-1/(2d+3)}\right\}$$

is nonempty for some sufficiently large constant $C > 0$, then the procedure described in the previous paragraph detects the mode $\mathbf{b}_0$ with probability converging to one as $n \to \infty$.

Note that the rate for the localization of the modes of the density $f_{\boldsymbol{\beta}}$ is given by the grid width and equal to $n^{-1/(2d+3)}$ (up to some logarithmic factor). Recall that we have $h_{\max} \lesssim \log(n)^{-14\gamma/(d-1)-5} n^{2\gamma/(d-1)-1}$ by Assumption 4 such that the consistency of the method requires the condition

$$\gamma > \frac{d^2 - 1}{2d + 3},$$

where the right hand side is smaller than one for $d = 2, 3$.

4 Finite sample properties

In this section we illustrate the finite sample properties of the proposed test in a bivariate and a trivariate setting. In the bivariate setting we illustrate how simultaneous tests for the hypotheses (3.3) and (3.4) can be used to obtain a graphical representation of the local monotonicity properties of the density. In the trivariate setting we investigate the performance of the test for modality at a given point $\mathbf{b}_0$ (see the hypotheses in (3.5)) and the dependence of its power on the distribution of $\mathbf{X}$.

As test function we consider the simplest polynomial which satisfies the conditions of Assumption 1 for $d = 2, 3$, that is,

$$\phi(x) = c(56x^3 + 21x^2 + 6x + 1)(1 - x)^6 \mathbb{1}\{x \leq 1\}, \quad x \in [0, \infty),$$

with c such that $\int \phi = 1$. Figure 2 displays the function $\mathcal{H}_d \widetilde{\phi}^{(d-1)}$ for $d = 2$ (left panel) and $d = 3$ (right panel). Throughout this section the nominal level is fixed as $\alpha = 0.05$.

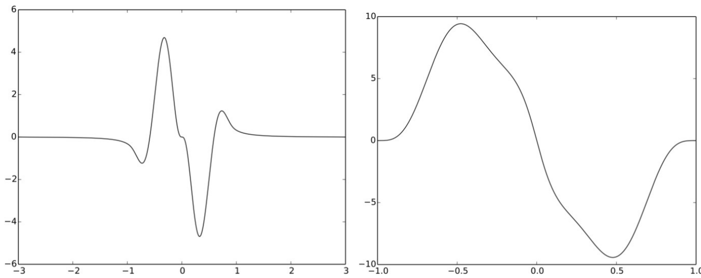


Figure 2: *The function* $\mathcal{H}_d \widetilde{\phi}^{(d-1)}$ *for* $d = 2$ *(left panel) and* $d = 3$ *(right panel).*

4.1 Inference about local monotonicity of a bivariate density

We follow the multiscale approach in Section 3.1 to obtain a graphical representation of the monotonicity behavior for the bivariate density of the random coefficients. To test the hypotheses (3.3) we use (3.15) with $\mathcal{T}_n = \mathcal{T}_{\mathbf{t}} \times \{h_0\} \times \mathcal{T}_{\mathbf{v}}$. Here, $h = h_0 = 0.5$ is fixed and the set of test locations $\mathcal{T}_{\mathbf{t}}$ is defined as the set of vertices on an equidistant grid in the square $[-1, 2]^2$ with width one. Finally, the set of test directions are

$$\mathcal{T}_{\mathbf{v}} = \left\{ \mathbf{v}_1 = -\mathbf{v}_3 = \sqrt{2}^{-1}(1, 1)^\top, \mathbf{v}_2 = -\mathbf{v}_4 = \sqrt{2}^{-1}(-1, 1)^\top \right\}.$$

The data are simulated with $f_{\boldsymbol{\beta}}$ the density of the normal mixture $\frac{1}{3}\mathcal{N}((-0.4, -0.57)^\top, 0.2I) + \frac{1}{3}\mathcal{N}((1.5, -0.52)^\top, 0.2I) + \frac{1}{3}\mathcal{N}((0.45, 1.6)^\top, 0.15I)$. The design is chosen such that $\boldsymbol{\Theta}_i$ is uniformly distributed on the sphere $\mathbb{S}^1$. Figure 1 in Section 3.1 displays the monotonicity behavior of the density $f_{\boldsymbol{\beta}}$ based on sample size $n = 20000$. Each arrow at a location $\mathbf{t}$ in direction $\mathbf{v}$ displays a rejection of a hypothesis (3.3). The map indicates the existence of modes around the points $(-0.5, -0.5)^\top$, $(1.5, -0.5)^\top$, and $(0.5, 1.5)^\top$ and thus reconstructs the true modes fairly well.

4.2 A local test for modality

In this section, we consider the random coefficients model with $d = 3$ and investigate the power of a local test for the existence of a mode at a given location $\mathbf{b}_0$ by testing several hypotheses of the form (3.5) simultaneously. More precisely, six tests for the hypotheses (3.5) are conducted for fixed bandwidth $h = h_0 = 1$ and

$$
\begin{aligned}
\mathbf{t}_1 = \mathbf{v}_1 &= (1,0,0)^\top = -\mathbf{t}_2 = -\mathbf{v}_2, \\
\mathbf{t}_3 = \mathbf{v}_3 &= (0,1,0)^\top = -\mathbf{t}_4 = -\mathbf{v}_4, \\
\mathbf{t}_5 = \mathbf{v}_5 &= (0,0,1)^\top = -\mathbf{t}_6 = -\mathbf{v}_6.
\end{aligned}
$$

The postulated mode is given by the point $\mathbf{b}_0 = (0,0,0)^\top$ and we conclude that $f_{\boldsymbol{\beta}}$ has a local maximum at the point $\mathbf{b}_0$, whenever all hypotheses $H_{0,-}^{\mathbf{t}_j,h_0,\mathbf{v}_j}$, $j = 1,\dots,6$, are rejected, that is

$$
\operatorname{sgn}(c_d)\widehat{T}_{\mathbf{t}_j,h_0,\mathbf{v}_j} > \kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha), \quad \text{for all } j = 1,\dots,6, \tag{4.1}
$$

where $\kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha)$ is defined by (3.14).

Numerical simulations for random coefficients model without intercept: At first, we consider model (1.1) with $\mathbf{X}_i \sim \mathrm{Unif}[-5,5]^3$. In the first two columns of Table 1 level and power of the local test (4.1) are reported for sample size $n \in \{250, 500, 1000\}$ based on 1000 repetitions. For the level of the test, we used $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \propto \mathbf{1}(\boldsymbol{\beta} \in [-5,5]^3)$ and for the power we took $f_{\boldsymbol{\beta}}$ as the density of a trivariate standard normal distribution. By construction, the multiscale method is rather conservative but detects the mode in most cases. We also propose a calibrated version of the test where the quantiles are chosen such that the test keeps its nominal level $\alpha = 0.05$. We point out that this calibration can be conducted without assuming any knowledge about the unknown densities $f_{\boldsymbol{\beta}}$ and $f_{\mathbf{X}}$. Columns four and five in Table 1 contain the level and power of the calibrated tests and should be compared to the second and third column.

n	level	power	level (cal.)	power (cal.)
250	0	12.6	5.2	90.4
500	0	53.0	5.2	98.7
1000	0	93.8	5.4	100

Table 1: *Simulated level and power of the test* (4.1) *for a mode and uniform design* $\mathbf{X}_i \sim \mathrm{Unif}[-5,5]^3$. *Results with theoretical quantiles are in the second and third column. Results where* $\kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha)$ *in* (4.1) *are replaced by calibrated quantiles are in the fourth and fifth column.*

Next we investigate an asymmetric distribution of the directions $\boldsymbol{\Theta}_i$ by sampling $\mathbf{X}_i \sim \mathcal{N}((3,0,0)^\top, 2I)$, with I the 3×3 identity matrix. We consider the same test cases as for Table 1 above. Results are reported in Table 2. Compared to Table 1, we observe a decrease in the power of the test (4.1). The explanation is that the design above is closer to a uniform design of $\mathbf{\Theta}_i$ on the sphere which makes it simpler to recover information about the joint density, see also Section 2.

n	level	power	level (cal.)	power (cal.)
250	0	1.2	5.4	73.3
500	0	8.7	4.8	88.1
1000	0	39.3	5.4	97.5

Table 2: *Same as Table 1 but now for normal design* $\mathbf{X}_i \sim \mathcal{N}((3,0,0)^\top, 2I)$.

Numerical simulations for random coefficients model with intercept: Now, we study model (1.2) with $d = 3$. In a first simulation, we sample the random vectors $(X_{i,2}, X_{i,3})^\top$ from a standard bivariate Cauchy distribution, such that the density $f_{\mathbf{\Theta}}$ is constant. Except for the different design, we consider otherwise the same test settings as above. The simulated level and power of the test (4.1) and of its calibrated version are shown in Table 3.

n	level	power	level (cal.)	power (cal.)
250	0	9.9	4.8	89.3
500	0	48.6	4.5	99.1
1000	0	95.7	4.6	100

Table 3: *Same as Table 1 but now for random coefficients model with intercept and* $(X_{i,2}, X_{i,3})^\top$ *from a standard bivariate Cauchy distribution.*

Finally, we consider two designs which do not satisfy Assumption 3. Table 4 reports the level and power for the same setting as above except that now $(X_{i,2}, X_{i,3})^\top$ is drawn a standard normal distribution. Table 5 is the same for $(X_{i,2}, X_{i,3})^\top \sim \mathrm{Unif}[-5,5]^2$. We observe only a slight decrease in the power of the test for normally distributed design compared to the setting when Assumption 3 holds. Even under uniform design, the test performs fairly well.

n	level	power	level (cal.)	power (cal.)
250	0	6.3	5.2	86.2
500	0	33.1	4.8	98.6
1000	0	84.1	4.6	99.9

Table 4: *Same as Table 1 but now for random coefficients model with intercept and* $(X_{i,2}, X_{i,3})^\top \sim \mathcal{N}((0,0)^\top, I)$.

n	level	power	level (cal.)	power (cal.)
250	0	1.1	4.8	73.8
500	0	11.8	5.0	87.1
1000	0	39.5	5.3	97.1

Table 5: *Same as Table 1 but now for random coefficients model with intercept and* $(X_{i,2}, X_{i,3})^\top \sim \mathrm{Unif}[-5,5]^2$.

4.3 A two-scale test

For multimodal densities which have a second mode close to the test location $\mathbf{b}_0$ testing different bandwidths simultaneously can be advantageous. This is illustrated by the following example, where we consider the random coefficients model without intercept with $d = 2$. The data are simulated with $f_{\boldsymbol{\beta}}$ the density of the normal mixture

$$\frac{1}{2}\mathcal{N}\Big((0,0)^\top, \begin{pmatrix} 0.05 & 0 \\ 0 & 0.4 \end{pmatrix}\Big) + \frac{1}{2}\mathcal{N}((2,0)^\top, 0.1 \cdot I).$$

We consider a design such that $\boldsymbol{\Theta}$ is uniformly distributed on the circle $\mathbb{S}^1$. Sample size is $n = 2000$ and $\alpha = 0.05$. We conducted four tests simultaneously for the hypotheses (3.5) with $\mathbf{b}_0 = (0,0)^\top$, a fixed bandwidth $h = h_0 = 1$, and

$$\mathbf{t}_1 = \mathbf{v}_1 = (1,0)^\top = -\mathbf{t}_2 = -\mathbf{v}_2,$$
$$\mathbf{t}_3 = \mathbf{v}_3 = (0,1)^\top = -\mathbf{t}_4 = -\mathbf{v}_4.$$

This test fails to detect a mode. The same happens (in 94.3 percent of 1000 repetitions) if we take instead $h = h_0 = 0.5$, and test for the directions

$$\mathbf{t}_1 = (0.5,0)^\top = -\mathbf{t}_2,\ \mathbf{v}_1 = (1,0)^\top = -\mathbf{v}_2,$$
$$\mathbf{t}_3 = (0,0.5)^\top = -\mathbf{t}_4,\ \mathbf{v}_3 = (0,1)^\top = -\mathbf{v}_4.$$

The bandwidth $h_0 = 1$ is too large to separate the two modes of $f_{\boldsymbol{\beta}}$. On the contrary, $h_0 = 0.5$ is too small to detect the decrease with small slope corresponding to the eigenvalue 0.4 of the covariance matrix in the first mixture component of $f_{\boldsymbol{\beta}}$. By conducting both tests simultaneously, we are however able to detect the mode at $(0,0)^\top$ in 68.5 percent of 1000 repetitions. Figure 3 provides an illustration of the results of the eight tests for the hypotheses (3.5) conducted simultaneously. Each arrow at a location $\mathbf{t}$ in direction $\mathbf{v}$ displays a rejection of a hypothesis (3.5) and the length of the arrows corresponds to the respective bandwidths.

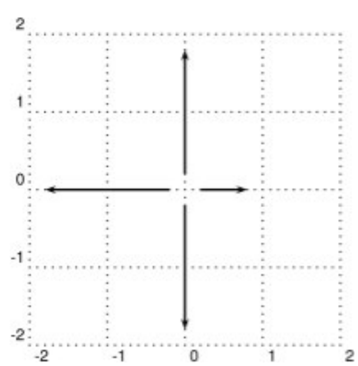


Figure 3: *Rejected hypotheses of the eight tests.*

5 Application to consumer demand data

Heterogeneity of consumers is a major challenge in modeling and estimating consumer demand. In several different demand models random coefficients were proposed to account for the heterogeneity in the population of consumers.

5.1 Model and data

In this section we are interested in the almost ideal demand system (AIDS) which was initially proposed by [13] with fixed coefficients. This model does not explain demand for a product itself but explains the budget share spent on a product by a linear equation. The explanatory variables are log prices and the log of total expenditure divided by a price index. A detailed discussion of the model is contained in [36].

Fixed coefficients in this model mean that all consumers are assumed to react in the same way when the price of a product changes. It is well known that some consumers are very price sensitive and change their behavior significantly with small variations in prices while other consumers are less price sensitive. This type of heterogeneity can be modeled by a random coefficient on log prices which is assumed to vary across the population of consumers. A similar argument suggests a random coefficient on log total expenditure. Recently, applications of the AIDS using a nonparametric random coefficient specification instead of fixed coefficients were presented in [30] and [9].

We apply our test for shape constraints to random coefficients in the AIDS for budget shares for food at home (BSF)

$$BSF_i = \beta_{i,1} + \beta_{i,2} \ln(TotExp_i) + \beta_{i,3} \ln(FoodPrice_i). \tag{4.1}$$

Food expenditure is a large fraction of total expenditure and is roughly about 20%.

We analyze the data of the British Family Expenditure Survey which ran from 1961 to 2001. It reported yearly cross sections for labor income, expenditure and other characteristics of about 7000 households. Following [30] we use data of the years 1994–2000 only. To reduce measurement errors only two person households with no children where at least one is working and the head of the household is a white collar worker is considered. This is quite usual in the demand literature, see [36]. Budget shares of food at home are generated by dividing the expenditure for all food consumed at home by total expenditure. The largest and smallest 2.5% were removed as outliers. Our variables $TotExp$ and $FoodPrice$ are total expenditure and food prices normalized by a general price index to report relative prices.

Assumption 3 and the numerical simulations in Section 4 suggest that our test has more power when the normalized regressors are approximately uniform on the sphere. We achieve this by symmetrizing the design in model (4.1) as follows:

$$BSF_i = \tilde{\beta}_{i,1} + \tilde{\beta}_{i,2}\,(\ln(TotExp_i) - 5) + \tilde{\beta}_{i,3}\,(25\ln(FoodPrice_i) - 0.3)\,. \tag{4.2}$$

The relation of the modified model to the random coefficients in (4.1) is $\beta_{i,1} = \tilde{\beta}_{i,1} - 5\tilde{\beta}_{i,2} - 0.3\tilde{\beta}_{i,3}$, $\beta_{i,2} = \tilde{\beta}_{i,2}$, $\beta_{i,3} = 25\tilde{\beta}_{i,3}$. Observations of the new variable $\ln(TotExp_i) - 5$ lie between -1.7 and 1.9. The observations of $25\ln(FoodPrice_i) - 0.3$ range from -3.2 to 4.

5.2 Results

For a first evaluation of the data we assumed fixed coefficients in model (4.2) and estimated the model with ordinary least squares (OLS).

$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$
0.2302	-0.0743	0.0027
(0.001)	(0.001)	(0.001)

Table 6: *Results of OLS for model* (4.2) *with fixed coefficients.*

In order to find modes of the density we conducted tests on the 5% level of the form (3.5) with a fixed bandwidth $h = h_0 = 1.3$. We were looking for modes on the equidistant grid covering $[-1.5, 1.5]^3$ with grid width 1.5. Hence, the grid had 27 nodes. For every grid point $\mathbf{b} \in \mathbb{R}^3$ tests of the hypotheses (3.5) were conducted for the directions and locations

$$\begin{aligned}
&\mathbf{t}_1 = \mathbf{b} + h_0\mathbf{e}_1, \quad \mathbf{v}_1 = \mathbf{e}_1, && \mathbf{t}_4 = \mathbf{b} - h_0\mathbf{e}_2, \quad \mathbf{v}_4 = -\mathbf{e}_2,\\
&\mathbf{t}_2 = \mathbf{b} - h_0\mathbf{e}_1, \quad \mathbf{v}_2 = -\mathbf{e}_1, && \mathbf{t}_5 = \mathbf{b} + h_0\mathbf{e}_3, \quad \mathbf{v}_5 = \mathbf{e}_3,\\
&\mathbf{t}_3 = \mathbf{b} + h_0\mathbf{e}_2, \quad \mathbf{v}_3 = \mathbf{e}_2, && \mathbf{t}_6 = \mathbf{b} - h_0\mathbf{e}_3, \quad \mathbf{v}_6 = -\mathbf{e}_3.
\end{aligned}$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ denote the unit vectors of $\mathbb{R}^3$. We detected a single mode in the neighborhood of the grid point $(0, 0, 0)^\top$.

In the following we will use nonparametric estimation to motivate hypothesis for our test. It is important that this estimate and the test are independent. We meet the requirement by splitting the sample in two independent equally sized sub-samples. The first sub-sample is used for nonparametric estimation of the random coefficient density in model (4.2) with the estimator in [31]. Figure 4 gives contour plots for the joint densities of $f_{\tilde{\beta}_1,\tilde{\beta}_2}$, $f_{\tilde{\beta}_1,\tilde{\beta}_3}$, and $f_{\tilde{\beta}_2,\tilde{\beta}_3}$. The nonparametric estimate suggest that the random coefficient density of $f_{\tilde{\beta}_1,\tilde{\beta}_2,\tilde{\beta}_3}$ is unimodal and almost symmetric. The estimate suggests that $f_{\tilde{\beta}_1,\tilde{\beta}_2,\tilde{\beta}_3}$ has a mode close to

$$(\beta_1, \beta_2, \beta_3) = (0.25, -0.17, 0.03). \tag{4.3}$$

This is consistent with the results of the test above which found a mode close to $(0, 0, 0)$. Since the marginal densities of $\beta_1, \beta_2, \beta_3$ are nearly symmetric it is also consistent that the mode is close to the OLS estimates given in Table 6. With a significantly skewed or with a multimodal random coefficient density location of modes would differ from OLS.

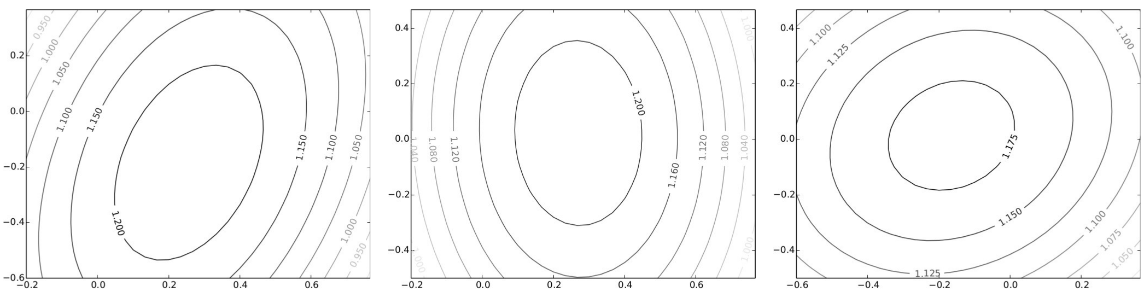


Figure 4: *Nonparametric estimates of the joint densities of* $f_{\tilde{\beta}_1,\tilde{\beta}_2}$ *(left),* $f_{\tilde{\beta}_1,\tilde{\beta}_3}$ *(middle), and* $f_{\tilde{\beta}_2,\tilde{\beta}_3}$ *(right).*

We tested on the 5% level if a mode can be verified for the location given by (4.3) with a smaller bandwidth than in the test above. Six hypotheses (3.5) with a fixed bandwidth $h = h_0 = 0.5$ and the following directions and locations

$$\begin{aligned} &\mathbf{t}_1 = (0.75, -0.17, 0.03)^\top,\ \mathbf{v}_1 = (1, 0, 0)^\top, && \mathbf{t}_4 = (0.25, -0.67, 0.03)^\top,\ \mathbf{v}_4 = -\mathbf{v}_3, \\ &\mathbf{t}_2 = (-0.25, -0.17, 0.03)^\top,\ \mathbf{v}_2 = -\mathbf{v}_1, && \mathbf{t}_5 = (0.25, -0.17, 0.53)^\top,\ \mathbf{v}_5 = (0, 0, 1)^\top, \\ &\mathbf{t}_3 = (0.25, 0.33, 0.03)^\top,\ \mathbf{v}_3 = (0, 1, 0)^\top, && \mathbf{t}_6 = (0.25, -0.17, -0.47)^\top,\ \mathbf{v}_6 = -\mathbf{v}_5 \end{aligned}$$

are tested. However, the bandwidth $h = 0.5$ allows only to conclude that a mode exists within an Euclidean ball with radius 0.5 and center $(0.25, -0.17, 0.03)^\top$. Hence, we cannot be sure that the location of the mode is not $(0, 0, 0)^\top$. Our primary interest is in the coefficients on $\tilde{\beta}_2$ and $\tilde{\beta}_3$ on total expenditure and food prices. In order to see if the mode is indeed in a location where $\tilde{\beta}_2 < 0$ and $\tilde{\beta}_3 > 0$ we add the two hypotheses $H_{0,-}^{\mathbf{t}_7,h_7,\mathbf{v}_7}$ with $h_7 = 0.1$ for

$$\mathbf{t}_7 = (0.25, 0, 0.03)^\top \quad \text{and} \quad \mathbf{v}_7 = (0, 1, 0)^\top$$

as well as $H_{0,+}^{\mathbf{t}_8,h_8,\mathbf{v}_8}$ with $h_8 = 0.03$ for

$$\mathbf{t}_8 = (0.25, -0.17, 0)^\top \quad \text{and} \quad \mathbf{v}_8 = (0, 0, 1)^\top.$$

When testing the eight hypothesis simultaneously the existence of the mode was confirmed by rejection of the first six hypothesis. In addition, $H_{0,-}^{\mathbf{t}_7,h_7,\mathbf{v}_7}$ was rejected, while we failed to reject $H_{0,+}^{\mathbf{t}_8,h_8,\mathbf{v}_8}$. This gives evidence that the mode is in a location where $\tilde{\beta}_2 < 0$ but we cannot decide if $\tilde{\beta}_3 > 0$ or $\tilde{\beta}_3 = 0$ at the mode.

Let us return to the initial model (4.1). The results of our test give evidence that a mode exists close to

$$(\beta_1, \beta_2, \beta_3) = (1.1, -0.17, 0.75)$$

with strong evidence that β_2 is indeed negative. This vector of coefficients describes a representative member of the majority of consumers. Hence, in the majority group food budget shares decrease with increasing log total expenditure. However, the nonparametric estimate in Figure 4 suggests that there is considerable variance among consumers around this representative member.

Acknowledgments

J. Schmidt-Hieber was partially funded by a TOP II grant from the Dutch science foundation. F. Dunker acknowledges support by the Ministry of Education and Cultural Affairs of Lower Saxony in the project Reducing Poverty Risk. K. Proksch acknowledges financial support by the German Research Foundation DFG through subproject A07 of CRC 755. K. Eckle has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project C4) of the German Research Foundation (DFG).

References

[1] Adler, R. J. and Taylor, J. E. (2007). *Random fields and geometry.* Springer Monographs in Mathematics. Springer, New York.

[2] Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.

[3] Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

[4] Beran, R. (1993). Semiparametric random coefficient regression models. *Ann. Inst. Statist. Math.*, 45(4):639–654.

[5] Beran, R., Feuerverger, A., and Hall, P. (1996). On nonparametric estimation of intercept and slope distributions in random coefficient regression. *Ann. Statist.*, 24(6):2569–2592.

[6] Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Ann. Statist.*, 20(4):1970–1984.

[7] Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

[8] Berry, S. and Pakes, A. (2007). The pure characteristics demand model. *Internat. Econom. Rev.*, 48(4):1193–1225.

[9] Breunig, C. and Hoderlein, S. (2016). Nonparametric Specification Testing in Random Parameter Models. Boston College Working Papers in Economics 897, Boston College Department of Economics.

[10] Butucea, C., Gu, M., and Artiles, L. (2007). Minimax and adaptive estimation of the wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.*, 35(2):465–494.

[11] Chernozhukov, V., Chetverikov, D., and Kato, K. (2016). Central limit theorems and bootstrap in high dimensions. *ArXiv Preprint*, arXiv:1412.3661.

[12] Davison, M. E. (1983). The ill-conditioned nature of the limited angle tomography problem. *SIAM J. Appl. Math.*, 43(2):428–448.

[13] Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70:312–326.

[14] Dub, J.-P., Fox, J. T., and Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

[15] Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152.

[16] Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785.

[17] Dunker, F., Hoderlein, S., and Kaido, H. (2013). Random coefficients in static games of complete information. *cemmap Working Papers*, CWP12/13.

[18] Dunker, F., Hoderlein, S., and Kaido, H. (2017). Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models. *cemmap Working Papers*, CWP11/17.

[19] Eckle, K., Bissantz, N., and Dette, H. (2016). Multiscale inference for multivariate deconvolution. *ArXiv Preprint*, arXiv:1611.05201.

[20] Eckle, K., Bissantz, N., Dette, H., Proksch, K., and Einecke, S. (2017). Multiscale inference for a multivariate density with applications to x-ray astronomy. *Ann. Inst. Statist. Math.*, forthcoming.

[21] Feuerverger, A. and Vardi, Y. (2000). Positron emission tomography and random coefficients regression. *Ann. Inst. Statist. Math.*, 52(1):123–138.

[22] Fox, J. T. and Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics*, 47(1):118–139.

[23] Frikel, J. (2013). Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, 34(1):117–141.

[24] Gautier, E. and Hoderlein, S. (2012). A triangular treatment effect model with random coefficients in the selection equation. *cemmap Working Papers*, CWP39/12.

[25] Gautier, E. and Kitamura, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica*, 81(2):581–607.

[26] Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(4):503–522.

[27] Greenland, S. (2000). When should epidemiologic regressions use random coefficients? *Biometrics*, 56(3):915–921.

[28] Gustafson, P. and Greenland, S. (2006). The performance of random coefficient regression in accounting for residual confounding. *Biometrics*, 62(3):760–768.

[29] Helgason, S. (2011). *Integral geometry and Radon transforms.* Springer, New York.

[30] Hoderlein, S., Holzmann, H., and Meister, A. (2015). The triangular model with random coefficients. *cemmap Working Papers*, CWP33/15.

[31] Hoderlein, S., Klemelä, J., and Mammen, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, 26(3):804–837.

[32] Hohmann, D. and Holzmann, H. (2016). Weighted angle radon transform: Convergence rates and efficient estimation. *Statistica Sinica.*, 26:157–175.

[33] Hsiao, C. (2014). *Analysis of Panel Data*. Cambridge University Press. Cambridge Books Online.

[34] Hsiao, C. and Pesaran, M. H. (2004). Random Coefficient Panel Data Models. CESifo Working Paper Series 1233, CESifo Group Munich.

[35] Ichimura, H. and Thompson, T. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics*, 86(2):269 – 295.

[36] Lewbel, A. (1997). Consumer demand systems and household equivalence scales. In Pesaran, M. H. and Schmidt, P., editors, *Handbook of applied econometrics*, volume 2, chapter 4, pages 167–201. Blackwell, Oxford.

[37] Masten, M. (2015). Random coefficients on endogenous variables in simultaneous equations models. *cemmap Working Papers*, CWP25/15.

[38] Masten, M. and Torgovitsky, A. (2014). Instrumental variables estimation of a generalized correlated random coefficients model. *cemmap Working Papers*, CWP02/14.

[39] Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

[40] Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.

[41] Proksch, K., Werner, F., and Munk, A. (2016). Multiscale scanning in inverse problems. *ArXiv Preprint*, arXiv:1611.04537.

[42] Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Ann. Statist.*, 41(3):1299–1328.

[43] Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–323.

SUPPLEMENT

A Nonparametric estimators for the densities $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$

In this section we discuss the estimation of the densities $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$ and of some quantities that rely on the estimates. We use kernel density estimators based on the second half of the observations $(S_i, \boldsymbol{\Theta}_i)$, $i = n+1, \dots, 2n$. The density of the random vector $\mathbf{X}_1$ is denoted by $f_{\mathbf{X}}$. Note that $f_{\mathbf{X}}$ is a d-variate density in the random coefficients model without intercept and a $(d-1)$-variate density in the random coefficients model with intercept. Throughout the following, $K : \mathbb{R} \to \mathbb{R}$ is assumed to be Lipschitz continuous, non-negative and $\int K = 1$.

In the random coefficients model with intercept the symmetrizations $S_i = \zeta_i Y_i / \|\mathbf{X}_i\|$ and $\boldsymbol{\Theta}_i = \zeta_i \mathbf{X}_i / \|\mathbf{X}_i\|$ with Rademacher variables ζ_i correspond to point reflections of the densities at the origin. Thus, the density $f_{\boldsymbol{\Theta}}$ is in general not continuous even when the non-symmetrized density is continuous on the hemisphere $\mathbb{S}_+^{d-1}$ (see also the representation (3.11)). However, continuity or smoothness is necessary to control the bias in nonparametric density estimation. Therefore, we use a two step procedure for the estimation of $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$ in the random coefficients model with intercept. First, we estimate the density of the non-symmetrized samples $Y_i/\|\mathbf{X}_i\|$ and $\mathbf{X}_i/\|\mathbf{X}_i\|$, $i = n+1, \dots, 2n$ on the hemisphere $\mathbb{S}_+^{d-1}$ and on $\mathbb{R} \times \mathbb{S}_+^{d-1}$, respectively. Smoothness assumptions are reasonable in this step. Second, we add to these densities their point reflection at the origin and divide by 2 to get estimates of $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$ for the symmetrized variables.

In the random coefficients model without intercept, we introduce a kernel density estimate

$$\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{C(h_*)}{n h_*^{d-1}} \sum_{i=n+1}^{2n} K\Big(\frac{1 - \langle \boldsymbol{\Theta}_i, \boldsymbol{\theta} \rangle}{h_*^2}\Big), \quad h_* > 0, \tag{4.1}$$

where

$$C(h_*) := h_*^{d-1} \Big(\int_{\mathbb{S}^{d-1}} K\Big(\frac{1 - \langle \boldsymbol{\theta}', \boldsymbol{\theta} \rangle}{h_*^2}\Big) d\boldsymbol{\theta}' \Big)^{-1}.$$

As shown in Bai et al. [3], the integral does not depend on $\boldsymbol{\theta}$ and $C(h_*)$ converges to some positive constant as $h_* \to 0$.

We also propose the following kernel density estimator

$$\widehat{f}_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta}) = \frac{1}{n h_+^d} C(h_+) \sum_{i=n+1}^{2n} K\Big(\frac{1 - \langle \boldsymbol{\Theta}_i, \boldsymbol{\theta} \rangle}{h_+^2}\Big) K\Big(\frac{S_i - s}{h_+}\Big), \quad h_+ > 0, \tag{4.2}$$

to estimate the joint density $f_{S,\boldsymbol{\Theta}}$.

In the random coefficients model with intercept we define estimators for $f_{\mathbf{\Theta}}$ and $f_{S,\mathbf{\Theta}}$ according to (4.1) and (4.2), where $\boldsymbol{\theta} \in \mathbb{S}_+^{d-1}$ and the constant $C(h_*)$ is replaced by a function $\boldsymbol{\theta} \mapsto C(h_*, \boldsymbol{\theta})$, defined by

$$C(h_*, \boldsymbol{\theta}) := h_*^{d-1} \Big(\int_{\mathbb{S}_+^{d-1}} K\Big(\frac{1 - \langle \boldsymbol{\theta}', \boldsymbol{\theta} \rangle}{h_*^2}\Big) d\boldsymbol{\theta}' \Big)^{-1} \quad (\boldsymbol{\theta} \in \mathbb{S}_+^{d-1}).$$

Note that $C(h_*) \leq C(h_*, \boldsymbol{\theta}) \leq 2C(h_*)$ and that $C(h_*, \boldsymbol{\theta})$ is Lipschitz-continuous with respect to $\boldsymbol{\theta}$. Using these properties, the proofs of Lemmas A.1 and A.2 below follow the same lines of proof if an intercept is included and we therefore omit the details.

By (2.1), we have the representation $f_{S,\mathbf{\Theta}}(s, \boldsymbol{\theta}) = f_{\mathbf{\Theta}}(\boldsymbol{\theta}) R f_{\boldsymbol{\beta}}(s, \boldsymbol{\theta})$. Therefore, $f_{S,\mathbf{\Theta}}(s, \boldsymbol{\theta})$ and $s \mapsto \log(|s|)^2 f_{S,\mathbf{\Theta}}(s, \boldsymbol{\theta})$ for $|s| \geq 1$ are uniformly bounded since $f_{\boldsymbol{\beta}}$ is compactly supported by Assumption 2. Moreover, $f_{S,\mathbf{\Theta}}$ is Hölder continuous with Hölder constant γ. This is a straightforward consequence of the Hölder γ-continuity of $f_{\mathbf{\Theta}}$ shown in the proof of Lemma A.1 and the identity

$$R f_{\boldsymbol{\beta}}(s, \boldsymbol{\theta}) = \int_{\mathbb{R}^{d-1}} f_{\boldsymbol{\beta}}(s\boldsymbol{\theta} + x_1 \boldsymbol{\theta}_1 + \ldots + x_{d-1} \boldsymbol{\theta}_{d-1}) d\boldsymbol{x},$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{d-1}$ denote an orthonormal basis of the orthogonal complement of $\operatorname{span}\{\boldsymbol{\theta}\}$, together with the compact support and the Lipschitz-continuity of $f_{\boldsymbol{\beta}}$ (following from Assumption 2). Basic arguments show that the properties of $f_{S,\mathbf{\Theta}}$ discussed above also hold in quantum homodyne tomography under the assumptions presented in Remark 1. We point out that the marginal densities of the Wigner function which are given by the Radon transform are nonnegative.

For the estimation of the test statistic $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$ and the limiting process $\widehat{X}_{\mathbf{t},h,\mathbf{v}}$ the quantities $1/f_{\mathbf{\Theta}}$ and $\sqrt{f_{S,\mathbf{\Theta}}}$ need to be estimated. The functions $(\cdot)^{-1}$ and $\sqrt{\cdot}$ are not smooth in zero and we therefore introduce the cut-off estimators

$$\widetilde{f}_{\mathbf{\Theta}} := \widehat{f}_{\mathbf{\Theta}} \vee \log(n)^{-1} \quad \text{and} \quad \widetilde{f}_{S,\mathbf{\Theta}} := \widehat{f}_{S,\mathbf{\Theta}} \vee \log(n)^{-2}. \tag{4.3}$$

By the boundedness from below of $f_{\mathbf{\Theta}}$ and Lemma A.1 it holds $\widetilde{f}_{\mathbf{\Theta}} = \widehat{f}_{\mathbf{\Theta}}$ almost surely for n sufficiently large.

In the next two lemmas we derive convergence properties for these estimators. If in Lemma A.1 $f_{\mathbf{X}}$ follows a multivariate Cauchy distribution, then $\gamma := 1$. Otherwise, γ comes from Assumption 3 *(iii)*.

Lemma A.1. *Let Assumption 3 hold and consider* $\widehat{f}_{\mathbf{\Theta}}$ *introduced in* (4.1) *with* $h_* = O(\log(n)^{-3/\gamma})$ *and* $h_* \geq \log(n)^{7/(d-1)} n^{-1/(d-1)}$. *Then* *(i)* $\quad \sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \big|\mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big] - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big| = O\big(h_*^{\gamma}\big);$

(ii) $\quad \sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \big|\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big]\big| = O_{\mathbb{P}}\Big(\sqrt{\frac{\log(n)}{nh_*^{d-1}}}\Big);$

(iii) $\quad \sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \big|\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big| = O\big(\log(n)^{-1}\big),$ *for* $n\to\infty$, *almost surely.*

In Lemma A.2 γ comes from Assumption 3 *(iii)* and Remark 1 *(ii)* in the case of quantum homodyne tomography.

Lemma A.2. *Let Assumption 3 hold and consider the estimator* $\widehat{f}_{S,\boldsymbol{\Theta}}$ *in* (4.2) *with bandwidth choice* $h_+ = O(\log(n)^{-3/\gamma})$ *and* $h_+ \geq \log(n)^{3/d} n^{-1/(2d)}$. *Then*

$$\sup_{(s,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{S}^{d-1}} \big|\widehat{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) - f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\big| = O(\log(n)^{-2}) \quad \textit{for } n\to\infty \textit{ almost surely.}$$

Proof of Lemma A.1. We assume that *(i)-(iii)* in Assumption 3 hold. If the design density is multivariate Cauchy, we can derive the properties in a similar way for the hemisphere $\mathbb{S}_+^{d-1}$. In order to prove *(i)*, we start with

$$\begin{aligned}\big|\mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big] - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big| =& \Big|\frac{C(h_*)}{h_*^{d-1}}\int_{\mathbb{S}^{d-1}} K\Big(\frac{1-\langle\boldsymbol{\theta}',\boldsymbol{\theta}\rangle}{h_*^2}\Big) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}')d\boldsymbol{\theta}' - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\Big| \\ \leq& \frac{C(h_*)}{h_*^{d-1}}\int_{\|\boldsymbol{\theta}'-\boldsymbol{\theta}\|\lesssim h_*} K\Big(\frac{1-\langle\boldsymbol{\theta}',\boldsymbol{\theta}\rangle}{h_*^2}\Big)\big|f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}') - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big|d\boldsymbol{\theta}'.\end{aligned}$$

Here, we used the compact support of K and the identity $1-\langle\boldsymbol{\theta}',\boldsymbol{\theta}\rangle = \|\boldsymbol{\theta}'-\boldsymbol{\theta}\|^2/2$. Since $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \int_0^\infty r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta})dr$, we have by Assumption 3 *(iii)* $|f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}') - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})| \leq \int_0^\infty r^{d-1}|f_{\mathbf{X}}(r\boldsymbol{\theta}') - f_{\mathbf{X}}(r\boldsymbol{\theta})|dr \lesssim h_*^\gamma$ for $\|\boldsymbol{\theta}'-\boldsymbol{\theta}\| \lesssim h_*$. By definition of the constant $C(h_*)$, we obtain $\big|\mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big] - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big| \lesssim h_*^\gamma$ and this proves *(i)*.

Next, we bound the stochastic error term *(ii)* using an entropy argument and Bernstein's inequality. Observe that by the Lipschitz-continuity of K, $\widehat{f}_{\boldsymbol{\Theta}} - \mathbb{E}(\widehat{f}_{\boldsymbol{\Theta}})$ is Lipschitz-continuous with Lipschitz constant of order h_*^{-d-1}. For $a_n := \sqrt{\frac{\log(n)}{nh_*^{d-1}}}$, let $\{\boldsymbol{\theta}_j : j = 1,\ldots,M\}$ be defined as the set of smallest cardinality such that $\bigcup_{j=1}^M B_{c'h_*^{d+1}a_n}(\boldsymbol{\theta}_j) \supset \mathbb{S}^{d-1}$ for some constant $c' > 0$. If $c' > 0$ is chosen small enough, then

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \big|\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big]\big| > ca_n\Big) \leq \sum_{j=1}^M \mathbb{P}\Big(\big|\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j) - \mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j)\big]\big| > \frac{c}{2}a_n\Big). \tag{4.4}$$

In order to bound the probability, we apply Bernstein's inequality to $\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j) - \mathbb{E}[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j)] = \sum_{i=n+1}^{2n} Z_i$ with

$$Z_i := \frac{1}{nh_*^{d-1}} C(h_*) K\Big(\frac{1-\langle\boldsymbol{\Theta}_i,\boldsymbol{\theta}\rangle}{h_*^2}\Big) - \frac{1}{n}\mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j)\big].$$

We find $|Z_i| \leq \frac{C_1}{nh_*^{d-1}}$ for some constant $C_1 > 0$, and for some constant $C_2 > 0$,

$$\mathbb{E}\big[Z_i^2\big] \leq \frac{C(h_*)^2}{n^2 h_*^{2d-2}} \int_{\mathbb{S}^{d-1}} \Big(K\Big(\frac{1-\langle \boldsymbol{\theta}', \boldsymbol{\theta}\rangle}{h_*^2}\Big)\Big)^2 f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}')d\boldsymbol{\theta}' \leq \frac{C_2}{n^2 h_*^{d-1}}$$

using the boundedness of K and $f_{\boldsymbol{\Theta}}$ and the definition of $C(h_*)$. Hence, an application of Bernstein's inequality yields with (4.4),

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}} \big|\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big]\big| > ca_n\Big) \leq 2M \exp\Big(\frac{-a_n^2 c^2/8}{C_2 n^{-1} h_*^{-d+1} + cC_1 a_n (6n)^{-1} h_*^{-d+1}}\Big).$$

Since M is a polynomial power of n, the claim follows by choosing the constant c large enough.

For *(iii)* one proceeds similarly with the choice $a_n = (\log(n)\log\log(n))^{-1}$ using the summability of the probabilities. □

Proof of Lemma A.2. We assume that *(i)-(iii)* in Assumption 3 hold. If the design density is multivariate Cauchy, we can derive the properties in a similar way for $\mathbb{R}\times\mathbb{S}_+^{d-1}$. An upper bound of the bias can be derived similarly to Lemma A.1 *(i)*. For the stochastic error

$$\mathbb{P}\Bigg(\sup_{(s,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{S}^{d-1}} \Big|\widehat{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) - \mathbb{E}\big(\widehat{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\big)\Big| > \log\log(n)^{-1}\log(n)^{-2}\Bigg)$$

we make use of a slightly modified version of Proposition 2.2 in Giné and Guillou [26] which we state below as Proposition A.1. We recall the definition of a VC class of functions. Let $\mathcal{F}$ be a uniformly bounded class of measurable functions on a measurable space $(S, \mathcal{S})$, with a measurable and bounded envelope F. We say that $\mathcal{F}$ is a measurable uniformly bounded VC class of functions if $\mathcal{F}$ is measurable and if there are constants $A, v > 0$ such that

$$\sup_Q N(\mathcal{F}, L_2(Q), \varepsilon\|F\|_{L^2(Q)}) \leq \Big(\frac{A}{\varepsilon}\Big)^v$$

for all $0 < \varepsilon < 1$, where $N(T, d, \varepsilon)$ denotes the ε-covering number of the metric space (T, d) and the supremum is taken over all probability measures on $(S, \mathcal{S})$.

In contrast to Proposition 2.2 in Giné and Guillou [26], Proposition A.1 contains the explicit dependence of the right hand side of (4.5) on the constants A and v. This is necessary as we consider the class of functions

$$\mathcal{F}_n := \Big\{(S, \boldsymbol{\Theta}) \mapsto K\Big(\frac{1-\langle \boldsymbol{\Theta}, \boldsymbol{\theta}\rangle}{h_+^2}\Big)K\Big(\frac{S-s}{h_+}\Big),\ (s,\boldsymbol{\theta}) \in \mathbb{R}\times\mathbb{S}^{d-1}\Big\},$$

which depends on n via h_+. By the boundedness of K we find $U, \sigma \lesssim 1$. To show that $\mathcal{F}_n$ is a VC class of functions, we introduce a discretization of $\mathbb{R}\times\mathbb{S}^{d-1}$ as follows: Let

$c > 0$ be a sufficiently small constant only depending on the kernel K. We chose a grid $\{\boldsymbol{\theta}_j : j = 1, \dots, M_1\}$ of $\mathbb{S}^{d-1}$ with grid width at most $c\varepsilon h_+^2$. Obviously, this is possible with $M_1 \lesssim (\varepsilon^{-1} h_+^{-2})^{d-1}$. Moreover, introduce the set of intervals $I_k = [k, k+1)$, $k \in \mathbb{Z}$. For each probability measure Q there are at most $\lceil (c\varepsilon)^{-2} \rceil$ sets $I_{i_j} \times \mathbb{S}^{d-1}$, $j = 1, \dots, \lceil (c\varepsilon)^{-2} \rceil$, such that $Q(I_{i_j} \times \mathbb{S}^{d-1}) \geq (c\varepsilon)^2$. Let $\{s_j : j = 1, \dots, M_2\}$ be an equidistant grid of

$$\widetilde{I}_{h_+} := \left\{ s \in \mathbb{R} : \operatorname{dist}\left(s, \bigcup_{j=1}^{\lceil (c\varepsilon)^{-2} \rceil} I_{i_j} \right) \leq 1 \right\}$$

with grid width $ch_+\varepsilon$ and let s_{M_2+1} denote an arbitrary point in $\widetilde{I}_{h_+}^C$. Basic calculations show $M_2 \lesssim \varepsilon^{-3} h_+^{-1}$. Moreover, the subset of $\mathcal{F}_n$ indexed by

$$\big\{ s_j : j = 1, \dots M_2 + 1 \big\} \times \big\{ \boldsymbol{\theta}_j : j = 1, \dots, M_1 \big\} =: \big\{ (s_j, \boldsymbol{\theta}_j) : j = 1, \dots, M_1(M_2 + 1) \big\}$$

is an ε-covering set of $\mathcal{F}_n$. To see this, fix $(s, \boldsymbol{\theta}) \in \widetilde{I}_{h_+} \times \mathbb{S}^{d-1}$. Then

$$\begin{aligned}
&\int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} \Big| K\Big(\frac{1 - \langle \boldsymbol{\Theta}, \boldsymbol{\theta} \rangle}{h_+^2} \Big) K\Big(\frac{S - s}{h_+} \Big) - K\Big(\frac{1 - \langle \boldsymbol{\Theta}, \boldsymbol{\theta}_j \rangle}{h_+^2} \Big) K\Big(\frac{S - s_j}{h_+} \Big) \Big|^2 dQ(S, \boldsymbol{\Theta}) \\
\lesssim & \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|^2}{h_+^4} + \frac{|s - s_j|^2}{h_+^2}
\end{aligned}$$

by the Lipschitz continuity of K. Hence, by construction of the set $\widetilde{I}_{h_+} \times \mathbb{S}^{d-1}$ there exists $j \in \{1, \dots, M_1(M_2 + 1)\}$ such that

$$\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|^2}{h_+^4} + \frac{|s - s_j|^2}{h_+^2} < \varepsilon^2.$$

For $(s, \boldsymbol{\theta}) \in (\widetilde{I}_{h_+} \times \mathbb{S}^{d-1})^C$ we obtain

$$\int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} \Big(K\Big(\frac{S - s}{h_+} \Big) + K\Big(\frac{S - s_{M_2+1}}{h_+} \Big) \Big)^2 dQ(S, \boldsymbol{\Theta}) < \varepsilon^2$$

since the support of $K(\frac{\cdot - s}{h_+})$ is compact and does not intersect with any of the sets $I_{i_j} \times \mathbb{S}^{d-1}$, $j = 1, \dots, \lceil (c\varepsilon)^{-2} \rceil$ for h_+ sufficiently small. A similar argument applies to $K(h_+^{-1}(\cdot - s_{M_2+1}))$. Hence,

$$N(\mathcal{F}, L_2(Q), \varepsilon) \lesssim \Big(\varepsilon^{-1} h_+^{(-2d+1)/(d+2)} \Big)^{d+2}$$

and $\mathcal{F}_n$ is a VC class of functions with $v = d + 2$ and $A = A_n = h_+^{(-2d+1)/(d+2)}$. An application of Proposition A.1 yields

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{(s, \boldsymbol{\theta}) \in \mathbb{R} \times \mathbb{S}^{d-1}} \big| \widehat{f}_{S, \boldsymbol{\Theta}}(s, \boldsymbol{\theta}) - \mathbb{E}(\widehat{f}_{S, \boldsymbol{\Theta}}(s, \boldsymbol{\theta})) \big| > \log\log(n)^{-1} \log(n)^{-2} \Big) \\
=& \mathbb{P}\Big(\sup_{f \in \mathcal{F}_n} \Big| \sum_{i=n+1}^{2n} \big(f(S_i, \boldsymbol{\Theta}_i) - \mathbb{E}(f(S_i, \boldsymbol{\Theta}_i)) \big) \Big| > \frac{n h_+^d}{C(h_+) \log\log(n) \log(n)^2} \Big) \\
\lesssim & \exp\Bigg(- \frac{1}{4K' \sigma^2 C(h_+)^2} \frac{n h_+^{2d}}{\log\log(n)^2 \log(n)^4} \Bigg)
\end{aligned}$$

for n sufficiently large. We have used that $\log(1+x) = x(1+o(1))$ for $x \to 0$. The last line of the equation converges to zero at a summable rate since $h_+ \geq \log(n)^{3/d} n^{-1/(2d)}$ by assumption which concludes the proof of the uniform almost sure convergence of $\widehat{f}_{S,\boldsymbol{\Theta}}$. $\square$

Proposition A.1. *Let P be any probability measure on $(S, \mathcal{S})$ and let ξ_i, $i = 1, \ldots, n$, be independent with common law P. Let further $\mathcal{F}$ be a measurable uniformly bounded VC class of functions and let σ^2 and U be any numbers such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} \operatorname{Var}_P f$, $U \geq \sup_{f \in \mathcal{F}} \|f\|_\infty$ and $0 < \sigma \leq U$. Then there exist universal constants $C, K', L > 0$ such that the exponential inequality*

$$\begin{aligned} &\mathbb{P}\Big(\sup_{f \in \mathcal{F}} \Big| \sum_{i=1}^{n} \big(f(\xi_i) - \mathbb{E}(f(\xi_i)) \big) \Big| > t \Big) \\ \leq & K' \exp\left(- \frac{1}{K'} \frac{t}{U} \log\left(1 + \frac{tU}{\big(\sqrt{n}\sigma + L\sqrt{v}U\sqrt{\log(AU\sigma^{-1})}\big)^2} \right) \right) \end{aligned} \tag{4.5}$$

is valid for all $t \geq C(vU\log(AU\sigma^{-1}) + \sqrt{vn}\sigma\sqrt{\log(AU\sigma^{-1})})$.

Let us turn to the standard deviation

$$\sigma_{\mathbf{t},h,\mathbf{v}} = \Big(\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \frac{f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} ds d\boldsymbol{\theta} \Big)^{1/2} \tag{4.6}$$

and its estimator defined in (3.13). The following lemma shows that it is uniformly bounded from above and below. The proof is deferred to Appendix B.

Lemma A.3. *Under the Assumptions 1 and 2 there exist universal constants $C_1, C_2, n_0 > 0$ such that for any $n > n_0$,*

$$C_1 \leq \sigma_{\mathbf{t},h,\mathbf{v}} \leq C_2.$$

The consistency of the estimates $\widehat{f}_{\boldsymbol{\Theta}}$ and $\widehat{f}_{S,\boldsymbol{\Theta}}$ shows that $\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}$ is a consistent estimator of the standard deviation $\sigma_{\mathbf{t},h,\mathbf{v}}$.

Lemma A.4. *Under Assumption 3 it holds that*

$$\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} \big| \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t},h,\mathbf{v}} \big| = O\big(\log(n)^{-1} \big) \quad \textit{for } n \to \infty, \textit{ almost surely.}$$

Proof. By Lemma A.3,

$$\begin{aligned} \big| \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t},h,\mathbf{v}} \big| &\leq \frac{\big| \widehat{\sigma}^2_{\mathbf{t},h,\mathbf{v}} - \sigma^2_{\mathbf{t},h,\mathbf{v}} \big|}{\sigma_{\mathbf{t},h,\mathbf{v}}} \\ &\lesssim \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \Big| \frac{f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} - \frac{\widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} \Big| ds d\boldsymbol{\theta}. \end{aligned}$$

By Assumption 3, $f_{\boldsymbol{\Theta}}$ is uniformly bounded from below. Thus, $\widetilde{f}_{\boldsymbol{\Theta}}$ is almost surely uniformly bounded from below for sufficiently large n by Lemma A.1. This shows that

$$\begin{aligned}
&f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})\Big|\frac{1}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} - \frac{1}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2}\Big| + \frac{1}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2}\big|f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta}) - \widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})\big| \\
&= O\big((\log(n)^{-1}\big) \quad \text{almost surely.}
\end{aligned}$$

Here we used the boundedness of $f_{S,\boldsymbol{\Theta}}$ and

$$\begin{aligned}
&\big|f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta}) - \widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})\big| \\
&\leq \big|f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta}) - \widehat{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})\big| + \big|\widehat{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta}) - \widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})\big| \\
&= O\big(\log(n)^{-2}\big) \quad \text{almost surely}
\end{aligned}$$

by Lemma A.2. The claim follows now from the integrability of $\big((\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big)^2$ proved in Lemma 3.1 *(ii)*. □

Lemma A.5. *Under Assumption 3 we have*

$$\sup_{(s,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{S}^{d-1}}\Big|\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})} - \sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}\Big| = O(\log(n)^{-1}) \quad \textit{for } n\to\infty \textit{ almost surely.}$$

Proof. This is a direct consequence of

$$\sup_{(s,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{S}^{d-1}}\Big|\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) - f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\Big| = O(\log(n)^{-2}) \quad \text{for } n\to\infty \text{ almost surely}$$

as shown in the proof of Lemma A.4 and

$$\big|\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})} - \sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}\big| = \frac{\big|\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) - f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\big|}{\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})} + \sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}} \leq \frac{\big|\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) - f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\big|}{\log(n)^{-1}}.$$

□

We discussed in Section 3 that the test statistic $T_{\mathbf{t},h,\mathbf{v}}$ relies on the unknown density $f_{\boldsymbol{\Theta}}$ and therefore we introduced the statistic $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$, where the density $f_{\boldsymbol{\Theta}}$ is replaced by the estimate $\widetilde{f}_{\boldsymbol{\Theta}}$. An important part of the proof of Theorem 3.2 consists of showing that this replacement is asymptotically negligible. To this end, the bias of the estimate $1/\widetilde{f}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ has to be controlled.

Lemma A.6. *Under Assumption 3,*

$$\sup_{\boldsymbol{\theta}\in\mathbb{S}^{d-1}}\Big|\frac{1}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} - \mathbb{E}\Big(\frac{1}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\Big)\Big| = O\big(h_*^{\gamma}\big) \quad \textit{for } n\to\infty.$$

Proof. Recall that $\widetilde{f}_{\mathbf{\Theta}} = \widehat{f}_{\mathbf{\Theta}}$ almost surely for n sufficiently large by the almost sure uniform convergence of $\widehat{f}_{\mathbf{\Theta}}$ proven in Lemma A.1 and the lower bound of $f_{\mathbf{\Theta}}$. Notice that

$$\begin{aligned}\Big|\mathbb{E}\Big(\frac{1}{f_{\mathbf{\Theta}}(\boldsymbol{\theta})} - \frac{1}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})}\Big)\Big| &= \Big|\mathbb{E}\Big(\frac{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) f_{\mathbf{\Theta}}(\boldsymbol{\theta})}\Big)\Big| \\ &\lesssim \Big|\mathbb{E}\Big(\frac{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})}{f_{\mathbf{\Theta}}(\boldsymbol{\theta})^2}\Big)\Big| + \Big|\mathbb{E}\Big(\frac{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) f_{\mathbf{\Theta}}(\boldsymbol{\theta})} \mathbb{1}\big\{\exists \boldsymbol{\theta}' : \widehat{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}') \le f_{\mathbf{\Theta}}(\boldsymbol{\theta}')/2\big\}\Big)\Big| \\ &\lesssim \big|\mathbb{E}\big(\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})\big)\big| + \log(n) h_*^{-d+1} \mathbb{P}\big(\exists \boldsymbol{\theta}' : \widehat{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}') \le f_{\mathbf{\Theta}}(\boldsymbol{\theta}')/2\}\big).\end{aligned}$$

The almost sure uniform convergence of $\widehat{f}_{\mathbf{\Theta}}$ show that

$$\mathbb{E}\big(\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})\big) = \mathbb{E}\big(\widehat{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})\big) \quad \text{and} \quad \mathbb{P}\big(\exists \boldsymbol{\theta}' : \widehat{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}') \le f_{\mathbf{\Theta}}(\boldsymbol{\theta}')/2\}\big) = 0$$

for n sufficiently large. Therefore,

$$\sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \Big|\mathbb{E}\Big(\frac{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - f_{\mathbf{\Theta}}(\boldsymbol{\theta})}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) f_{\mathbf{\Theta}}(\boldsymbol{\theta})}\Big)\Big| \lesssim h_*^{\gamma}$$

since $f_{\mathbf{\Theta}}$ is bounded from below by Lemma A.1. □

B Further proofs

Proof of (3.9)*:* Recall that in the random coefficients model with intercept

$$\mathbf{\Theta}_1 = \zeta_1 \frac{(1, X_{1,2}, X_{1,3}, \ldots, X_{1,d})}{\|(1, X_{1,2}, X_{1,3}, \ldots, X_{1,d})\|},$$

where ζ_1 is a Rademacher variable. Hence, we obtain

$$\begin{aligned}f_{\mathbf{\Theta}}(\boldsymbol{\theta}) &= \frac{1}{2}\int_0^\infty r^{d-1}\delta(r\theta_1 - 1) f_{\mathbf{X}}(r\theta_2, \ldots, r\theta_d) dr \\ &\quad + \frac{1}{2}\int_0^\infty r^{d-1}\delta(r\theta_1 + 1) f_{\mathbf{X}}(-r\theta_2, \ldots, -r\theta_d) dr \\ &= \frac{1}{2}\int_0^\infty \frac{r^{d-1}}{\theta_1^d}\delta(r - 1) f_{\mathbf{X}}\big(\tfrac{r}{\theta_1}\theta_2, \ldots, \tfrac{r}{\theta_1}\theta_d) dr \mathbb{1}\{\theta_1 > 0\} \\ &\quad + \frac{1}{2}\int_0^\infty \frac{r^{d-1}}{\theta_1^{d-1}|\theta_1|}\delta(r + 1) f_{\mathbf{X}}\big(-\tfrac{r}{\theta_1}\theta_2, \ldots, -\tfrac{r}{\theta_1}\theta_d) dr \mathbb{1}\{\theta_1 < 0\} \\ &= \frac{1}{2|\theta_1|^d} f_{\mathbf{X}}\Big(\frac{\theta_2}{\theta_1}, \ldots, \frac{\theta_d}{\theta_1}\Big).\end{aligned} \tag{4.1}$$

□

Proof of Lemma 3.1. By assumption, $\phi_{\mathbf{t},h}$ is radially symmetric and satisfies (3.2). We fix a direction $\mathbf{v} \in \mathbb{S}^{d-1}$ and consider the directional derivative

$$\partial_{\mathbf{v}} \phi_{\mathbf{t},h}(\mathbf{b}) = \frac{1}{h^{d+1}\operatorname{Vol}(\mathbb{S}^{d-2})} \phi' \left(\frac{\|\mathbf{b}-\mathbf{t}\|}{h} \right) \frac{\langle \mathbf{b}-\mathbf{t}, \mathbf{v} \rangle}{\|\mathbf{b}-\mathbf{t}\|},$$

where ϕ' is the usual derivative of ϕ. The Radon transform of this directional derivative is

$$\begin{aligned} R(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})(s,\boldsymbol{\theta}) &= \int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = s} \partial_{\mathbf{v}} \phi_{\mathbf{t},h}(\mathbf{b}) d\mu_{d-1}(\mathbf{b}) \\ &= \frac{1}{h^{d+1}\operatorname{Vol}(\mathbb{S}^{d-2})} \int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = s} \phi' \left(\frac{\|\mathbf{b}-\mathbf{t}\|}{h} \right) \frac{\langle \mathbf{b}-\mathbf{t}, \mathbf{v} \rangle}{\|\mathbf{b}-\mathbf{t}\|} d\mu_{d-1}(\mathbf{b}) \\ &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{d-2})} \int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = h^{-1}(s-\langle \mathbf{t},\boldsymbol{\theta}\rangle)} \phi' \left(\|\mathbf{b}\| \right) \frac{\langle \mathbf{b}, \mathbf{v} \rangle}{\|\mathbf{b}\|} d\mu_{d-1}(\mathbf{b}). \end{aligned}$$

Set $\widetilde{s} = h^{-1}(s - \langle \mathbf{t}, \boldsymbol{\theta} \rangle)$. For $d > 2$, using the definition of $\widetilde{\phi}$ in (3.6),

$$\begin{aligned} R(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})(s,\boldsymbol{\theta}) &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{d-2})} \int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = \widetilde{s}} \frac{\phi'(\|\mathbf{b}\|)}{\|\mathbf{b}\|} \langle \mathbf{b}, \mathbf{v} \rangle d\mu_{d-1}(\mathbf{b}) \\ &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{d-2})} \int_0^\infty \frac{\phi'\big(\sqrt{\widetilde{s}^2+r^2}\big)}{\sqrt{\widetilde{s}^2+r^2}} \int_{\mathbf{w}\perp\boldsymbol{\theta}, \|\mathbf{w}\|=r} \langle \boldsymbol{\theta}\widetilde{s} + \mathbf{w}, \mathbf{v} \rangle d\mathbf{w} dr \\ &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{d-2})} \int_0^\infty \frac{\phi'\big(\sqrt{\widetilde{s}^2+r^2}\big)}{\sqrt{\widetilde{s}^2+r^2}} \int_{\mathbf{w}\perp\boldsymbol{\theta}, \|\mathbf{w}\|=r} \langle \boldsymbol{\theta}\widetilde{s}, \mathbf{v} \rangle d\mathbf{w} dr \\ &= \frac{\langle \boldsymbol{\theta}, \mathbf{v}\rangle}{h^2} \int_0^\infty r^{d-2} \phi'\left(\sqrt{\widetilde{s}^2+r^2}\right) \frac{\widetilde{s}}{\sqrt{\widetilde{s}^2+r^2}} dr \\ &= \frac{\langle \boldsymbol{\theta}, \mathbf{v}\rangle}{h^2} \int_0^\infty r^{d-2} \frac{\partial}{\partial \widetilde{s}} \phi\left(\sqrt{\widetilde{s}^2+r^2}\right) dr \\ &= \frac{\langle \boldsymbol{\theta}, \mathbf{v}\rangle}{h^2} \widetilde{\phi}(\widetilde{s}) \\ &= \frac{\langle \boldsymbol{\theta}, \mathbf{v}\rangle}{h^2} \widetilde{\phi}\left(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta}\rangle}{h}\right). \end{aligned}$$

For $d = 2$ let $\mathbf{w} \perp \boldsymbol{\theta}$ with $\|\mathbf{w}\| = 1$ and write $\mathbf{b} = \boldsymbol{\theta}\widetilde{s} + r\mathbf{w}$ for $r \in \mathbb{R}$. Then

$$\begin{aligned} R(\partial_{\mathbf{v}} \phi_{\mathbf{t},h})(s,\boldsymbol{\theta}) &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{0})} \int_{-\infty}^\infty \frac{\phi'\big(\sqrt{\widetilde{s}^2+r^2}\big)}{\sqrt{\widetilde{s}^2+r^2}} \langle \boldsymbol{\theta}\widetilde{s} + r\mathbf{w}, \mathbf{v} \rangle dr \\ &= \frac{1}{h^{2}\operatorname{Vol}(\mathbb{S}^{0})} \int_{-\infty}^\infty \frac{\phi'\big(\sqrt{\widetilde{s}^2+r^2}\big)}{\sqrt{\widetilde{s}^2+r^2}} \langle \boldsymbol{\theta}\widetilde{s}, \mathbf{v} \rangle dr \\ &= \frac{\langle \boldsymbol{\theta}, \mathbf{v}\rangle}{h^2} \int_0^\infty \phi'\left(\sqrt{\widetilde{s}^2+r^2}\right) \frac{\widetilde{s}}{\sqrt{\widetilde{s}^2+r^2}} dr, \end{aligned}$$

as $r \mapsto \frac{\phi'\left(\sqrt{\widetilde{s}^2+r^2}\right)}{\sqrt{\widetilde{s}^2+r^2}}$ is an even function. Now we can proceed similarly as in the case $d > 2$.

If d is odd, the proof of the representation of $A(\partial_{\mathbf{v}}\phi_{\mathbf{t},h})(s,\boldsymbol{\theta})$ is completed by taking the $(d-1)$-th derivative with respect to the variable s. If d is even, for any function $f:\mathbb{R}\to\mathbb{R}$, for any fixed $z\in\mathbb{R}$, and for $h>0$

$$\begin{aligned}\mathcal{H}_d\Big(s\mapsto f\Big(\frac{s-z}{h}\Big)\Big)(u) &= \frac{1}{\pi}\text{p.v.}\int_{-\infty}^{\infty} f\Big(\frac{s-z}{h}\Big)\frac{1}{u-s}ds\\ &= \frac{1}{\pi}\lim_{\epsilon\to 0^+}\int_{(-\infty,u-\epsilon]\cup[u+\epsilon,\infty)} f\Big(\frac{s-z}{h}\Big)\frac{1}{u-s}ds\\ &= \frac{1}{\pi}\lim_{\epsilon\to 0^+}\int_{(-\infty,(u-z)/h-\epsilon]\cup[(u-z)/h+\epsilon,\infty)} f(s)\frac{1}{(u-z)/h-s}ds\\ &= \big(\mathcal{H}_d f\big)\Big(\frac{u-z}{h}\Big)\qquad\text{for } u\in\mathbb{R},\end{aligned}$$

by substitution. That $\mathcal{H}_d f$ exists is shown below for the choice $f=\widetilde{\phi}^{(d-1)}$. Hence, we obtain $A(\partial_{\mathbf{v}}\phi_{\mathbf{t},h})(s,\boldsymbol{\theta}) = \langle\boldsymbol{\theta},\mathbf{v}\rangle h^{-d-1}(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h})$ for d even.

Next we prove $\|\widetilde{\phi}^{(k)}\|_\infty<\infty$ for $k=0,\dots,d+1$. The case $k=0$ is obvious. We use the chain rule for higher order derivatives given by Faà di Bruno's formula

$$\frac{d^k}{dz^k}f_1(f_2(z)) = \sum_{(m_1,\dots,m_k)\in\mathcal{M}_k}\frac{k!}{m_1!\dots m_k!}f_1^{(m_1+\dots+m_k)}(f_2(z))\prod_{j=1}^{k}\Big(\frac{f_2^{(j)}(z)}{j!}\Big)^{m_j},\tag{4.2}$$

where $\mathcal{M}_k$ is the set of all k-tuples of non-negative integers satisfying $\sum_{j=1}^k jm_j=k$. Since $z\mapsto\phi(\sqrt{z^2+r^2})$ is a.e. $(k+1)$-times continuously differentiable, we can interchange the integral with the k-fold differentiation for the variable z provided that

$$\int_0^\infty\Big|r^{d-2}\frac{\partial^{k+1}}{\partial z^{k+1}}\phi\left(\sqrt{z^2+r^2}\right)\Big|dr$$

exists for all $k=1,\dots,d+1$. Applying (4.2) with $f_1=\phi$ and $f_2=\sqrt{\cdot^2+r^2}$ gives

$$\frac{\partial^{k+1}}{\partial z^{k+1}}\phi\left(\sqrt{z^2+r^2}\right) = \sum_{(m_1,\dots,m_{k+1})\in\mathcal{M}_{k+1}} C_{m_1,\dots,m_{k+1}}\phi^{(M)}\left(\sqrt{z^2+r^2}\right)\prod_{j=1}^{k+1}\big(f_2^{(j)}(z)\big)^{m_j}$$

for suitable constants $C_{m_1,\dots,m_{k+1}}$ and $M=\sum_{j=1}^{k+1}m_j$. Applying the chain rule to $f_2^{(j)}$ yields

$$f_2^{(j)}(z) = \sum_{\{\ell_j,k_j:\ell_j+2k_j=j\}} C_{\ell_j,k_j}z^{\ell_j}(z^2+r^2)^{1/2-\ell_j-k_j}$$

for non-negative integers ℓ_j,k_l and suitable constants C_{ℓ_j,k_j}. As ϕ is compactly supported, it remains to show that each of the functions

$$z\mapsto\int_0^{\sqrt{1-z^2}} r^{d-2}\phi^{(M)}\left(\sqrt{z^2+r^2}\right)|z|^{\sum_{j=1}^{k+1}\ell_jm_j}(z^2+r^2)^{M/2-\sum_{j=1}^{k+1}(\ell_j+k_j)m_j}dr\tag{4.3}$$

for $|z| \leq 1$ is uniformly bounded, where ℓ_j, k_j are arbitrary elements of the set $\{\ell_j, k_j : \ell_j + 2k_j = j\}$, $j = 1, \ldots, k+1$. Notice that

$$M/2 - \sum_{j=1}^{k+1} (\ell_j + k_j) m_j = \sum_{j=1}^{k+1} (\frac{1}{2} - \ell_j - k_j) m_j < 0.$$

A uniform bound for the integral on the right hand side of (4.3) can be found easily when z is bounded away from zero. We can thus assume that $|z| \leq \sqrt{1-z^2}$. Splitting the integral $\int_0^{\sqrt{1-z^2}} = \int_0^{|z|} + \int_{|z|}^{\sqrt{1-z^2}}$ and using that by Taylor expansion and Assumption 1,

$$\phi^{(j)}\left(\sqrt{z^2 + r^2}\right) \lesssim (z^2 + r^2)^{(3-j)/2} \text{ for } j = 1, 2 \text{ and } \phi^{(M)} \lesssim 1 \text{ for } M \leq d+2$$

as well as $\max\{z^2, r^2\} \leq z^2 + r^2 \leq 2\max\{z^2, r^2\}$, we obtain an upper bound (up to some constant) for the integral on the right hand side of (4.3) by

$$\begin{aligned}
&|z|^{\sum_{j=1}^{k+1} \ell_j m_j + M - 2\sum_{j=1}^{k+1} (\ell_j + k_j) m_j + \max\{3-M, 0\}} \int_0^{|z|} r^{d-2} dr \\
&+ |z|^{\sum_{j=1}^{k+1} \ell_j m_j} \int_{|z|}^{\sqrt{1-z^2}} r^{d-2+\max\{3-M,0\}+M-2\sum_{j=1}^{k+1}(\ell_j+k_j)m_j} dr \\
\lesssim &|z|^{\sum_{j=1}^{k+1} \ell_j m_j + M - 2\sum_{j=1}^{k+1} (\ell_j + k_j) m_j + d - 1 + \max\{3-M, 0\}} + 1.
\end{aligned}$$

By the use of $\ell_j + 2k_j = j$, $\sum_{j=1}^{k+1} j m_j = k+1$ and $k \leq d+1$, we find that this is bounded by $z^{-3+M+\max\{3-M,0\}} + 1$ which proves the result.

Next, we prove that $\mathcal{H}_d \widetilde{\phi}^{(d-1)}$ exists. Recall that $\|\widetilde{\phi}^{(d)}\|_\infty < \infty$ and consequently, $\widetilde{\phi}^{(d-1)}$ is Lipschitz continuous. For any Lipschitz continuous function f with compact support,

$$\Big| \int_{-\infty}^{u-1} \frac{f(x)}{u-x} dx \Big| \vee \Big| \int_{u+1}^{\infty} \frac{f(x)}{u-x} dx \Big| \leq \|f\|_\infty \lambda(\operatorname{supp} f),$$

where $\lambda(\operatorname{supp} f)$ denotes the Lebesgue measure of the support of f. Moreover,

$$\lim_{\epsilon \to 0^+} \Big(\int_{u-1}^{u-\varepsilon} \frac{f(x)}{u-x} dx + \int_{u+\varepsilon}^{u+1} \frac{f(x)}{u-x} dx \Big) = \lim_{\epsilon \to 0^+} \int_\varepsilon^1 \frac{f(u-x) - f(u+x)}{x} dx.$$

By the Lipschitz-continuity of f, $|f(u-x) - f(u+x)| \lesssim |x|$ such that the r.h.s. can be bounded by a constant that does not depend on u. The result follows with $f = \widetilde{\phi}^{(d-1)}$. This proves assertion *(i)* in the Lemma.

Finally, we turn to assertion *(ii)*. As shown above, $\widetilde{\phi}^{(d-1)}$ is bounded. For odd dimension d the claim therefore follows from substitution and the compact support of $\widetilde{\phi}^{(d-1)}$. For d even, substitution and the fact that the Hilbert transform $\mathcal{H}_d$ defines a bounded operator $L^k(\mathbb{R}) \to L^k(\mathbb{R})$ for all $1 < k < \infty$ yield the required result.

□

Proof of Lemma A.3. The existence of a uniform upper bound of $\sigma_{\mathbf{t},h,\mathbf{v}}$ follows directly from the boundedness of $f_{S,\boldsymbol{\Theta}}$. A uniform lower bound for $f_{\boldsymbol{\Theta}}$ and the integrability of $\big(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\big)^2$ are shown in the proof of Lemma 3.1 *(ii)*. For the lower bound of $\sigma_{\mathbf{t},h,\mathbf{v}}$ recall that

$$\frac{f_{S,\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} = f_{S|\boldsymbol{\Theta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta}) = \int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = \langle \mathbf{t},\boldsymbol{\theta}\rangle + hs} f_{\boldsymbol{\beta}}(\mathbf{b})d\mu_{d-1}(\mathbf{b}).$$

By Assumption 2, $f_{\boldsymbol{\beta}}(\mathbf{b}) \geq c_{\boldsymbol{\beta}} > 0$ for all $\mathbf{b} \in [0,1]^d$ and $f_{\boldsymbol{\beta}}$ is uniformly continuous. Hence, there exists $\delta > 0$, which does not depend on h, such that $f_{\boldsymbol{\beta}}$ is uniformly bounded from below in the ball $B_\delta(\mathbf{t})$ of radius δ around any $\mathbf{t} \in [0,1]^d$, say, $f_{\boldsymbol{\beta}}(\mathbf{b}) > c_{\boldsymbol{\beta}}/2$ for all $\mathbf{b} \in \bigcup_{\mathbf{t}\in[0,1]^d} B_\delta(\mathbf{t})$. Define for $s^2 < \delta^2/(dh^2)$

$$A_{\delta,\mathbf{t},h} := \big\{\mathbf{b} \in \mathbb{R}^d : \mathbf{b} = \mathbf{t} + hs\boldsymbol{\theta} + \rho_2\boldsymbol{\theta}_2^{\perp} + \ldots + \rho_d\boldsymbol{\theta}_d^{\perp},\, \rho_j^2 < \delta^2/d,\ j = 2,\ldots,d\big\},$$

where $\boldsymbol{\theta}_2^{\perp},\ldots,\boldsymbol{\theta}_d^{\perp}$ form an orthonormal basis of the orthogonal complement of span$\{\boldsymbol{\theta}\}$. Clearly, $\mu_{d-1}(A_{\delta,\mathbf{t},h}) = (2\delta)^{d-1}d^{(1-d)/2} > 0$, and all $\mathbf{b} \in A_{\delta,\mathbf{t},h}$ satisfy

$$\|\mathbf{t} - \mathbf{b}\|^2 = (hs)^2 + \rho_2^2 + \ldots + \rho_d^2 < \frac{\delta^2}{d} + \delta^2\frac{d-1}{d} = \delta^2 \quad \text{and} \quad \langle \mathbf{b},\boldsymbol{\theta}\rangle = \langle \mathbf{t},\boldsymbol{\theta}\rangle + hs.$$

In particular, $A_{\delta,\mathbf{t},h} \subset B_\delta(\mathbf{t})$. Thus,

$$\int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = \langle \mathbf{t},\boldsymbol{\theta}\rangle + hs} f_{\boldsymbol{\beta}}(\mathbf{b})d\mu_{d-1}(\mathbf{b}) \geq \int_{A_{\delta,\mathbf{t},h}} f_{\boldsymbol{\beta}}(\mathbf{b})d\mu_{d-1}(\mathbf{b}) \geq \frac{c_{\boldsymbol{\beta}}}{2}\mu_{d-1}(A_{\delta,\mathbf{t},h}) > 0.$$

Hence, $Rf_{\boldsymbol{\beta}}(\langle \mathbf{t},\boldsymbol{\theta}\rangle + hs,\boldsymbol{\theta})$ is uniformly bounded from below for all $h \leq \mathbf{t} \leq 1-h$, $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ and $|s| < \delta/(\sqrt{d}h)$. Therefore,

$$\sigma_{\mathbf{t},h,\mathbf{v}}^2 \gtrsim \int_{-\delta/(\sqrt{d}h)}^{\delta/(\sqrt{d}h)} \big(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\big)^2 ds \geq \int_{-\delta/(\sqrt{d}h_{\max})}^{\delta/(\sqrt{d}h_{\max})} \big(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\big)^2 ds, \tag{4.4}$$

where the inequality holds uniformly over $\mathcal{T}$.

In quantum homodyne tomography, assumption *(iii)* in Remark 1 yields

$$\int_{\langle \mathbf{b},\boldsymbol{\theta}\rangle = \langle \mathbf{t},\boldsymbol{\theta}\rangle + hs} f_{\boldsymbol{\beta}}(\mathbf{b})d\mu_{d-1}(\mathbf{b}) \geq c_{\boldsymbol{\beta}}$$

for $s^2 < \delta^2/(dh^2)$ if δ is sufficiently small. Hence, (4.4) holds in this case as well. Furthermore, since $\mathcal{H}_d(\widetilde{\phi}^{(d-1)}) \in L^2(\mathbb{R})$, we obtain

$$\sigma_{\mathbf{t},h,\mathbf{v}}^2 \gtrsim \int_{\mathbb{R}} \big(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\big)^2 ds + o(1) \quad \text{for } n \to \infty.$$

If $\|\mathcal{H}_d(\widetilde{\phi}^{(d-1)})\|_2 \neq 0$ there exists $n_0 = n_0(\delta,d,\phi) \in \mathbb{N}$ such that

$$\sigma_{\mathbf{t},h,\mathbf{v}}^2 \gtrsim \frac{1}{2}\int_{\mathbb{R}} \big(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\big)^2 ds = \frac{1}{2}\big\|\widetilde{\phi}^{(d-1)}\big\|_2^2$$

for all $n > n_0$. The equality on the r.h.s. is trivial for odd dimensions d and follows for even dimensions from the anti self-adjointness of the Hilbert transform and $\mathcal{H}_d\mathcal{H}_d f = -f$. $\square$

C Proof of Theorem 3.2

If $\|\widetilde{\phi}^{(d-1)}\|_2 = 0$, Theorem 3.2 obviously holds. In the following we assume $\|\widetilde{\phi}^{(d-1)}\|_2 \neq 0$ and define

$$a_{\mathbf{t},h,\mathbf{v}}(s,\boldsymbol{\theta}) := h^{d+1}\Lambda(\partial_{\mathbf{v}}\phi_{\mathbf{t},h})(s,\boldsymbol{\theta}) = \langle\boldsymbol{\theta},\mathbf{v}\rangle(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\Big(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h}\Big),$$

where the equality follows from Lemma 3.1 *(i)*.

C.1 Controlling the effect of density estimation in the test statistic

Theorem C.1. *Under the assumptions of Theorem 3.2,*

$$\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h\sqrt{n}\frac{\big||\widehat{T}_{\mathbf{t},h,\mathbf{v}}-\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|-|T_{\mathbf{t},h,\mathbf{v}}-\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|\big|}{\sigma_{\mathbf{t},h,\mathbf{v}}} = o_{\mathbb{P}}(1), \ \ \textit{as } n\to\infty.$$

Proof. By the triangle inequality

$$\big||\widehat{T}_{\mathbf{t},h,\mathbf{v}}-\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|-|T_{\mathbf{t},h,\mathbf{v}}-\mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|\big| \leq U_{\mathbf{t},h,\mathbf{v}}+V_{\mathbf{t},h,\mathbf{v}}$$

with $U_{\mathbf{t},h,\mathbf{v}} := |\widehat{T}_{\mathbf{t},h,\mathbf{v}}-T_{\mathbf{t},h,\mathbf{v}}-\mathbb{E}[\widehat{T}_{\mathbf{t},h,\mathbf{v}}-T_{\mathbf{t},h,\mathbf{v}}]|$ and $V_{\mathbf{t},h,\mathbf{v}} := |\mathbb{E}[\widehat{T}_{\mathbf{t},h,\mathbf{v}}-T_{\mathbf{t},h,\mathbf{v}}]|$. We first bound $V_{\mathbf{t},h,\mathbf{v}}$ using

$$V_{\mathbf{t},h,\mathbf{v}} = \Big|\frac{1}{\sqrt{h}}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} a_{\mathbf{t},h,\mathbf{v}}(s,\boldsymbol{\theta})\mathbb{E}\Big[\frac{1}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}-\frac{1}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\Big]f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}\Big|$$

and

$$\int_{\mathbb{R}}\big|a_{\mathbf{t},h,\mathbf{v}}(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\big|ds \lesssim h\int_{\mathbb{R}}\big|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big|f_{S,\boldsymbol{\Theta}}(hs+\langle\mathbf{t},\boldsymbol{\theta}\rangle,\boldsymbol{\theta})ds \lesssim h\log(h)^2. \quad (4.1)$$

The last inequality follows for odd dimension d by the boundedness of $f_{S,\boldsymbol{\Theta}}$ and the integrability of $\widetilde{\phi}^{(d-1)}$. For even dimension, recall that $\mathcal{H}_d\widetilde{\phi}^{(d-1)}$ is bounded as shown in the proof of Lemma 3.1. Notice that

$$\int_2^{4/h^2}\frac{\big|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big|}{\log(s)^2}\log(s)^2 f_{S,\boldsymbol{\Theta}}(hs+\langle\mathbf{t},\boldsymbol{\theta}\rangle,\boldsymbol{\theta})ds \lesssim \log(h)^2$$

by $|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)| \lesssim (1+s^2)^{-1/2}$ (which holds for any function with compact support and bounded Hilbert transform) and the integrability of $(1+s^2)^{-1/2}\log(s)^{-2}$ for $s\geq 2$. For the remainder, we find

$$\int_{4/h^2}^{\infty}\frac{\big|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big|}{\log(s)^2}\frac{\log(s)^2}{\log(hs+\langle\mathbf{t},\boldsymbol{\theta}\rangle)^2}\log(hs+\langle\mathbf{t},\boldsymbol{\theta}\rangle)^2 f_{S,\boldsymbol{\Theta}}(hs+\langle\mathbf{t},\boldsymbol{\theta}\rangle,\boldsymbol{\theta})ds \lesssim 1$$

by the boundedness of $s \mapsto \log(|s|)^2 f_{S,\mathbf{\Theta}}(s,\boldsymbol{\theta})$ for all $|s| \geq 2, \boldsymbol{\theta} \in \mathbb{S}^{d-1}$, and

$$\frac{\log(s)}{\log(hs + \langle \mathbf{t}, \boldsymbol{\theta}\rangle)} \leq \frac{\log(s)}{\log(hs/2)} = \frac{\log(s)}{\log(h/2) + \log(s)} \leq 2,$$

as $\log(h/2) \geq -\log(s)/2$ for $s \geq 4/h^2$. A similar argument can be used to bound the integral $\int_{-\infty}^{-2} |a_{\mathbf{t},h,\mathbf{v}}(s,\boldsymbol{\theta}) f_{S,\mathbf{\Theta}}(s,\boldsymbol{\theta})|ds$. Applying Lemma A.6 with bandwidth $h_* = \log(n)^{7/(d-1)} n^{-1/(d-1)}$ gives

$$\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} V_{\mathbf{t},h,\mathbf{v}} \lesssim \log(h_{\max})^2 \sqrt{h_{\max}} \log(n)^{7\gamma/(d-1)} n^{-\gamma/(d-1)}. \tag{4.2}$$

Next, we prove $\rho_n := \mathbb{P}(\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} U_{\mathbf{t},h,\mathbf{v}} \geq \delta_n) \to 0$ as $n \to \infty$, where $\delta_n := (n \log(n))^{-1/2}$. If for some positive constant c

$$A_n := \Big\{ (S_i, \mathbf{\Theta}_i)_{i=n+1,\ldots,2n} : \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \big| \widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}[\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})] \big| \leq c \sqrt{\tfrac{\log n}{n h_*^{d-1}}} \Big\},$$

then by Lemma A.1

$$\begin{aligned} \rho_n &\leq \mathbb{E}\Big[\mathbb{P}\Big(\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} U_{\mathbf{t},h,\mathbf{v}} \geq \delta_n \bigm| (S_i, \mathbf{\Theta}_i)_{i=n+1,\ldots,2n} \Big) \mathbb{1}(A_n) \Big] + \mathbb{P}(A_n^c) \\ &\leq \sum_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} \mathbb{E}\Big[\mathbb{P}\Big(U_{\mathbf{t},h,\mathbf{v}} \geq \delta_n \bigm| (S_i, \mathbf{\Theta}_i)_{i=n+1,\ldots,2n} \Big) \mathbb{1}(A_n) \Big] + o(1) \end{aligned}$$

for sufficiently large c. Now we apply Bernstein's inequality to

$$U_{\mathbf{t},h,\mathbf{v}} = \Big| \sum_{i=1}^{n} \Big\{ \frac{1}{n\sqrt{h}} a_{\mathbf{t},h,\mathbf{v}}(S_i, \mathbf{\Theta}_i) \Big(\tfrac{1}{\widetilde{f}_{\mathbf{\Theta}}(\mathbf{\Theta}_i)} - \tfrac{1}{f_{\mathbf{\Theta}}(\mathbf{\Theta}_i)} \Big) - \frac{1}{n} \mathbb{E}[\widehat{T}_{\mathbf{t},h,\mathbf{v}} - T_{\mathbf{t},h,\mathbf{v}}] \Big\} \Big|.$$

By Lemma 3.1 *(i)*, $|a_{\mathbf{t},h,\mathbf{v}}(s,\boldsymbol{\theta})|$ can be bounded by a constant uniformly over $(s,\boldsymbol{\theta}) \in \mathbb{R} \times \mathbb{S}^{d-1}$. Moreover,

$$\Big| \frac{1}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})} - \frac{1}{f_{\mathbf{\Theta}}(\boldsymbol{\theta})} \Big| \leq \frac{\big| \widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}[\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})] \big| + \big| \mathbb{E}[\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})] - f_{\mathbf{\Theta}}(\boldsymbol{\theta}) \big|}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta}) f_{\mathbf{\Theta}}(\boldsymbol{\theta})}.$$

The inequality $\widetilde{f}_{\mathbf{\Theta}} \geq \log(n)^{-1}$, the uniform lower bound of $f_{\mathbf{\Theta}}$, $\widetilde{f}_{\mathbf{\Theta}} = \widehat{f}_{\mathbf{\Theta}}$ almost surely for n sufficiently large, Lemma A.1, and the definition of h_* imply that each summand in $U_{\mathbf{t},h,\mathbf{v}}$ is bounded on A_n by

$$\leq C \frac{\log(n)}{n\sqrt{h}} \Big(\sqrt{\frac{\log(n)}{n h_*^{d-1}}} + h_*^{\gamma} \Big) \leq C_1 \frac{1}{n \sqrt{h_{\min}} \log(n)^2}$$

for some constants $C, C_1 > 0$. By a change of variables in the integral for the variable s, the uniform boundedness of $f_{S,\mathbf{\Theta}}$, and the integrability of $a_{\mathbf{t},h,\mathbf{v}}^2$ as shown in Lemma 3.1 *(i)*, we find for the conditional variance with a similar argument as above

$$\operatorname{Var}\big(U_{\mathbf{t},h,\mathbf{v}} \,|\, (S_i, \mathbf{\Theta}_i)_{i=n+1,\ldots,2n} \big) \leq C \frac{1}{n^2} \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \Big(\tfrac{1}{\widetilde{f}_{\mathbf{\Theta}}(\boldsymbol{\theta})} - \tfrac{1}{f_{\mathbf{\Theta}}(\boldsymbol{\theta})} \Big)^2 \leq C_2 n^{-2} \log(n)^{-4}$$

with some constants $C, C_2 > 0$. Bernstein's inequality yields

$$\begin{aligned}\rho_n \lesssim &|\mathcal{T}_n| \exp\Big(- \frac{\delta_n^2/2}{C_2 n^{-1}\log(n)^{-4} + \frac{C_1\delta_n}{3n\sqrt{h_{\min}}\log(n)^2}}\Big) + o(1) \\ =&|\mathcal{T}_n| \exp\Big(- \frac{(n\log(n))^{-1}/2}{C_2 n^{-1}\log(n)^{-4} + \frac{C_1}{3n^{3/2}\sqrt{h_{\min}}\log(n)^{5/2}}}\Big) + o(1) = o(1),\end{aligned}$$

as $h_{\min} \geq n^{-1}$. Finally, the claim follows from $\beta_h \lesssim \frac{\sqrt{\log(n)}}{\log\log(n)}$ and the boundedness from below of $\sigma_{\mathbf{t},h,\mathbf{v}}$ shown in Lemma A.3. □

C.2 Approximation of the limit statistic

Define the process

$$X_{\mathbf{t},h,\mathbf{v}} = h^{-1/2}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v}\rangle (\mathcal{H}_d \widetilde{\phi}^{(d-1)})\left(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta}\rangle}{h}\right) \frac{\sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} W(dsd\boldsymbol{\theta}).$$

Note that $X_{\mathbf{t},h,\mathbf{v}}$ corresponds to the process $\widehat{X}_{\mathbf{t},h,\mathbf{v}}$ where the density estimators have been replaced by the true densities. The proof of Theorem 3.2 relies on a recently obtained Gaussian approximation result which is reproduced here for convenience.

Theorem C.2 ([11], Proposition 2.1)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random vectors in $\mathbb{R}^{2p}$ with $\mathbb{E}(\mathrm{X}_{i,j}) = 0$ and $\mathbb{E}(\mathrm{X}_{i,j}^2) < \infty$ for $i = 1, \ldots, n,\ j = 1, \ldots, 2p$. Moreover, let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be independent random vectors in $\mathbb{R}^{2p}$ with $\mathbf{Y}_i \sim N(\mathbf{0}, \mathbb{E}(\mathbf{X}_i\mathbf{X}_i^\top)),\ i = 1, \ldots, n$. Let $b, q > 0$ be some constants and let $B_n \geq 1$ be a sequence of constants, possibly growing to infinity as $n \to \infty$. Denote further by $\mathcal{A}'_{2p}$ the set of all hyperrectangles in $\mathbb{R}^{2p}$ of the form $A = \{\boldsymbol{x} \in \mathbb{R}^{2p} : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$ for $-\infty \leq \boldsymbol{a} \leq \boldsymbol{b} \leq \infty$. Assume that*

(i) $n^{-1}\sum_{i=1}^n \mathbb{E}(\mathrm{X}_{i,j}^2) \geq b$ *for all* $1 \leq j \leq 2p$;

(ii) $n^{-1}\sum_{i=1}^n \mathbb{E}(|\mathrm{X}_{i,j}|^{2+k}) \leq B_n^k$ *for all* $1 \leq j \leq 2p$ *and* $k = 1, 2$;

(iii) $\mathbb{E}\big(\big(\max_{1\leq j\leq 2p}|\mathrm{X}_{i,j}|/B_n\big)^q\big) \leq 2$ *for all* $i = 1, \ldots, n$

and define

$$D_n^{(1)} := \Big(\frac{B_n^2\log^7(2pn)}{n}\Big)^{\frac{1}{6}}, \quad D_{n,q}^{(2)} := \Big(\frac{B_n^2\log^3(2pn)}{n^{1-2/q}}\Big)^{\frac{1}{3}}.$$

Then there exists a constant C only depending on b and q, such that

$$\sup_{A\in\mathcal{A}'_{2p}}\Big|\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{X}_i \in A\Big) - \mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{Y}_i \in A\Big)\Big| \leq C(D_n^{(1)} + D_{n,q}^{(2)}).$$

Theorem C.3. *Under the assumptions of Theorem 3.2,*

$$\Big(\beta_h\Big(\sqrt{n}\frac{|T_{\mathbf{t},h,\mathbf{v}}-E[T_{\mathbf{t},h,\mathbf{v}}]|}{\sigma_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\Big)\Big)_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\leftrightarrow\Big(\beta_h\Big(\frac{|X_{\mathbf{t},h,\mathbf{v}}|}{\sigma_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\Big)\Big)_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}.$$

Proof. To take absolute values into account, we introduce the set

$$\mathcal{T}_n':=\mathcal{T}_n\cup\big\{(\mathbf{t},h,-\mathbf{v}):(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n\big\}=:\big\{(\mathbf{t}_j,h_j,\mathbf{v}_j):j=1,\dots,2p\big\}. \tag{4.3}$$

Moreover, for $i=1,\dots,n$, let $\mathbf{X}_i:=(\mathrm{X}_{i,1},\dots,\mathrm{X}_{i,2p})^\top$ with

$$\mathrm{X}_{i,j}:=\Upsilon_j(S_i,\mathbf{\Theta}_i)-\mathbb{E}(\Upsilon_j(S_i,\mathbf{\Theta}_i)),\quad\text{and}\quad\Upsilon_j(s,\boldsymbol{\theta}):=\frac{a_{\mathbf{t}_j,h_j,\mathbf{v}_j}(s,\boldsymbol{\theta})}{\sigma_{\mathbf{t}_j,h_j,\mathbf{v}_j}\sqrt{h_j}f_{\mathbf{\Theta}}(\boldsymbol{\theta})},\quad\text{for } j=1,\dots,2p.$$

Notice that $\sum_{i=1}^n X_{i,j}=n\sigma_{\mathbf{t}_j,h_j,\mathbf{v}_j}^{-1}(T_{\mathbf{t}_j,h_j,\mathbf{v}_j}-\mathbb{E}[T_{\mathbf{t}_j,h_j,\mathbf{v}_j}])$. In a first step, we show that for $\mathbf{Z}\sim N(\mathbf{0},\mathbb{E}(\mathbf{X}_1\mathbf{X}_1^\top))$,

$$\sup_{A\in\mathcal{A}_{2p}'}\Big|\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_i\in A\Big)-\mathbb{P}\Big(\mathbf{Z}\in A\Big)\Big|\to 0. \tag{4.4}$$

Observe that by (4.1) and the uniform lower bound of $\sigma_{\mathbf{t}_j,h_j,\mathbf{v}_j}$ established in Lemma A.3,

$$\big|\mathbb{E}(\Upsilon_j(S_1,\mathbf{\Theta}_1))\big|\lesssim\log(h_j)^2\sqrt{h_j}. \tag{4.5}$$

Because of this bound, the expectation $\mathbb{E}(\Upsilon_j(S_1,\mathbf{\Theta}_1))$ in the definition of $X_{i,j}$ will only provide terms of negligible order if we check the conditions of Theorem C.2. In particular, condition *(i)* is a direct consequence of the definition of $\sigma_{\mathbf{t}_j,h_j,\mathbf{v}_j}$ in (4.6). By Lemma 3.1 *(ii)*, the uniform lower bound of $\sigma_{\mathbf{t}_j,h_j,\mathbf{v}_j}$ in Lemma A.3, the lower bound of $f_{\mathbf{\Theta}}$, and the boundedness of $f_{S,\mathbf{\Theta}}$, we find for $k=1,2$, $\max_{j=1,\dots,2p}\mathbb{E}(|\Upsilon_j(S_1,\mathbf{\Theta}_1)|^{2+k})\lesssim h_{\min}^{-k/2}$. This implies condition *(ii)* of Theorem C.2 with $B_n\asymp h_{\min}^{-1/2}$.

Lemma 3.1 *(i)* implies $\max_{j=1,\dots,2p}|\mathrm{X}_{i,j}|\lesssim h_{\min}^{-1/2}$ which proves assertion *(iii)* in the theorem for any $q>0$ and $B_n=ch_{\min}^{-1/2}$, provided that the constant c is chosen sufficiently large. Consequently, Theorem C.2 applies and for $\mathbf{Z}\sim N(\mathbf{0},\mathbb{E}(\mathbf{X}_1\mathbf{X}_1^\top))$

$$\sup_{A\in\mathcal{A}_{2p}'}\Big|\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_i\in A\Big)-\mathbb{P}\Big(\mathbf{Z}\in A\Big)\Big|\lesssim\Big(\frac{h_{\min}^{-1}\log^7(n)}{n}\Big)^{\frac{1}{6}}+\Big(\frac{h_{\min}^{-1}\log^3(n)}{n^{1-2/q}}\Big)^{\frac{1}{3}}\to 0$$

choosing q large enough and using Assumption 4.

In a second step, we show that there exists a version of the Gaussian noise W such that

$$\max_{j=1,\dots,2p}\big|\mathrm{Z}_j-W(\Upsilon_j\sqrt{f_{S,\mathbf{\Theta}}})\big|=O_{\mathbb{P}}\big(|\log(h_{\max})|^3\sqrt{h_{\max}}\big).$$

To this end, we define the Gaussian process $(\widetilde{W}(f))_{f\in L^\infty(\mathcal{Z})}$ indexed by $L^\infty(\mathcal{Z})$ as the centered Gaussian process with covariance function

$$\int_{\mathcal{Z}} f_1(s,\boldsymbol{\theta})f_2(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta} - \int_{\mathcal{Z}} f_1(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}\int_{\mathcal{Z}} f_2(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}.$$

Thus, there exists a version of $\widetilde{W}(f)$ such that $\mathbf{Z} = \big(\widetilde{W}(\Upsilon_1),\dots,\widetilde{W}(\Upsilon_{2p})\big)^\top$. Recall that $(W(f))_{f\in L^2(\nu)}$ defines a Gaussian process whose mean and covariance functions are 0 and $\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} f_1(s,\boldsymbol{\theta})f_2(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}$, respectively. Basic calculations show that there exists a version of W such that

$$\widetilde{W}(f) = W(f\sqrt{f_{S,\boldsymbol{\Theta}}}) - \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} f(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}\, W(\sqrt{f_{S,\boldsymbol{\Theta}}}).$$

Hence,

$$\big|\widetilde{W}(\Upsilon_j) - W(\Upsilon_j\sqrt{f_{S,\boldsymbol{\Theta}}})\big| = \Big|\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} \Upsilon_j(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}\, W(\sqrt{f_{S,\boldsymbol{\Theta}}})\Big|.$$

By (4.5), $|\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} \Upsilon_j(s,\boldsymbol{\theta})f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})dsd\boldsymbol{\theta}| \lesssim \log(h_j)^2\sqrt{h_j}$. Furthermore, $W(\sqrt{f_{S,\boldsymbol{\Theta}}}) \sim N(0,1)$ which implies that $\mathbb{E}(\max_{j=1,\dots,2p}|\widetilde{W}(\Upsilon_j) - W(\Upsilon_j\sqrt{f_{S,\boldsymbol{\Theta}}})|) \lesssim \log(h_{\max})^2\sqrt{h_{\max}}$. An application of Markov's inequality finally proves

$$\max_{j=1,\dots,2p}\big|\widetilde{W}(\Upsilon_j) - W(\Upsilon_j\sqrt{f_{S,\boldsymbol{\Theta}}})\big| = O_{\mathbb{P}}\big(|\log(h_{\max})|^3\sqrt{h_{\max}}\big).$$

The insertion of the bandwidth normalization terms has no influence on the convergence as translation and multiplication preserve the interval structure. $\square$

C.3 Boundedness of the limit statistic

Recall from Lemma A.3 that $\sigma_{\mathbf{t},h,\mathbf{v}}$ is uniformly bounded from below whenever h is sufficiently small, where the upper bound $\overline{h}$ for h only depends on ϕ, d and $f_{\boldsymbol{\beta}}$. We therefore introduce the set

$$\overline{\mathcal{T}} := \big\{(\mathbf{t},h,\mathbf{v})\in\mathcal{T} : h\leq\overline{h}\big\}.$$

Theorem C.4. *Under the assumptions of Theorem 3.2,* $\sup_{(\mathbf{t},h,\mathbf{v})\in\overline{\mathcal{T}}}\beta_h(|X_{\mathbf{t},h,\mathbf{v}}|/\sigma_{\mathbf{t},h,\mathbf{v}} - \alpha_h)$ *is almost surely bounded.*

Proof. We want to apply Theorem 6.1 in Dümbgen and Spokoiny [15] to the non-normalized process

$$Y_{\mathbf{t},h,\mathbf{v}} := \frac{1}{\sigma_{\mathbf{t},h,\mathbf{v}}}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}} \langle\boldsymbol{\theta},\mathbf{v}\rangle(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\Big(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h}\Big)\frac{\sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}W(dsd\boldsymbol{\theta}).$$

Denote by ρ the canonical pseudo-metric on $\mathcal{T}$, induced by $Y_{\mathbf{t},h,\mathbf{v}}$

$$\rho : \begin{cases} \overline{\mathcal{T}} \times \overline{\mathcal{T}} \to \mathbb{R}_0^+ \\ \big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) \mapsto \Big(\mathbb{E}\big|Y_{\mathbf{t},h,\mathbf{v}} - Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2\Big)^{\frac{1}{2}}. \end{cases}$$

In the next step, we prove

$$\rho\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) \lesssim \big(\|\mathbf{v}-\mathbf{v}'\|^2 + \|\mathbf{t}-\mathbf{t}'\| + |h-h'|\big)^{1/2}. \tag{4.6}$$

By the uniform lower and upper bound for $\sigma_{\mathbf{t},h,\mathbf{v}}$,

$$\big|Y_{\mathbf{t},h,\mathbf{v}} - Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2 \lesssim \big|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2 + \big|Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2\big|\sigma_{\mathbf{t}',h',\mathbf{v}'} - \sigma_{\mathbf{t},h,\mathbf{v}}\big|^2. \tag{4.7}$$

In order to bound the expectation of the first term on the right hand side of (4.7), we use the boundedness properties of $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$

$$\begin{aligned}
&\mathbb{E}\big|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2 \\
\lesssim &\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\left|\langle\boldsymbol{\theta},\mathbf{v}-\mathbf{v}'\rangle(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\left(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h}\right)\right|^2 ds\,d\boldsymbol{\theta} \\
&+ \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\left|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\left(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h}\right) - (\mathcal{H}_d\widetilde{\phi}^{(d-1)})\left(\frac{s-\langle\mathbf{t}',\boldsymbol{\theta}\rangle}{h}\right)\right|^2 ds\,d\boldsymbol{\theta} \\
&+ \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\left|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\left(\frac{s-\langle\mathbf{t}',\boldsymbol{\theta}\rangle}{h}\right) - (\mathcal{H}_d\widetilde{\phi}^{(d-1)})\left(\frac{s-\langle\mathbf{t}',\boldsymbol{\theta}\rangle}{h'}\right)\right|^2 ds\,d\boldsymbol{\theta} \\
=:& \,\rho_1+\rho_2+\rho_3.
\end{aligned}$$

We show that the three terms can be bounded by the squared r.h.s. in (4.6). From Lemma 3.1 *(ii)* we obtain $\rho_1 \lesssim h\|\mathbf{v}-\mathbf{v}'\|^2$. For ρ_2, we distinguish between the cases $\|\mathbf{t}-\mathbf{t}'\| > h$ and $\|\mathbf{t}-\mathbf{t}'\| \le h$. In the first case, the triangle inequality and Lemma 3.1 *(ii)* give $\rho_2 \lesssim h < \|\mathbf{t}-\mathbf{t}'\|$. In the second case, the integral w.r.t. the variable s in ρ_2 is equal to

$$2h\int_{\mathbb{R}}\big((\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big)^2 ds - 2h\int_{\mathbb{R}}(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\Big(s+\frac{\langle\mathbf{t},\boldsymbol{\theta}\rangle-\langle\mathbf{t}',\boldsymbol{\theta}\rangle}{h'}\Big)ds.$$

Recall that the Hilbert transform and the differentiation operator commute. Therefore, using the differentiability of $\widetilde{\phi}^{(d-1)}$ which has been shown in the proof of Lemma 3.1, we find that

$$h(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\Big(s+\frac{\langle\mathbf{t},\boldsymbol{\theta}\rangle-\langle\mathbf{t}',\boldsymbol{\theta}\rangle}{h}\Big) = h(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s) + \langle\mathbf{t}-\mathbf{t}',\boldsymbol{\theta}\rangle(\mathcal{H}_d\widetilde{\phi}^{(d)})(\xi)$$

for some ξ between s and $s+\frac{\langle\mathbf{t}-\mathbf{t}',\boldsymbol{\theta}\rangle}{h}$. Hence,

$$\rho_2 \lesssim \|\mathbf{t}-\mathbf{t}'\|\int_{\mathbb{R}}(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)(\mathcal{H}_d\widetilde{\phi}^{(d)})(\xi)ds \lesssim \|\mathbf{t}-\mathbf{t}'\|\int_{\mathbb{R}}\big(1+(s/2)^2\big)^{-1}ds \lesssim \|\mathbf{t}-\mathbf{t}'\|.$$

Here, we used the boundedness of $\mathcal{H}_d\widetilde{\phi}^{(d-1)}$ shown in Lemma 3.1 *(i)*, $|\xi| \geq |s|/2$ for all $|s| \geq 2$ in the case $\|\mathbf{t}-\mathbf{t}'\| \leq h$, and $|(\mathcal{H}_d\widetilde{\phi}^{(d)})(u)| \lesssim (1+u^2)^{-1}$. The latter is obvious for d odd. For d even we find that $\widetilde{\phi}$ is an odd function and therefore $\widetilde{\phi}^{(d)}$ is an odd function. Moreover, for any odd function f such that $\mathcal{H}_d f$ exists, we have, up to some constant,

$$(\mathcal{H}_d f)(u) = \int_{-\infty}^{0} \frac{f(x)}{u-x}dx + \int_0^\infty \frac{f(x)}{u-x}dx = \int_0^\infty f(x)\Big(\frac{-1}{u+x} + \frac{1}{u-x}\Big)dx = \int_0^\infty \frac{2xf(x)}{u^2-x^2}dx.$$

Here all integrals are understood in the principal value sense. Finally, a similarly argument as in the proof of Lemma 3.1 shows that $\mathcal{H}_d\widetilde{\phi}^{(d)}$ exists and that $|(\mathcal{H}_d\widetilde{\phi}^{(d)})(u)| \lesssim (1+u^2)^{-1}$ by the compact support of $\widetilde{\phi}$.

We finally turn to ρ_3. Without loss of generality, we may assume $h \leq h'$. We study the cases $h \leq h'/2$ and $h > h'/2$, separately. In the first case, the triangle inequality and Lemma 3.1 *(ii)* give $\rho_3 \lesssim h + h' \lesssim |h'-h|$. If $h'/2 < h \leq h'$, we argue as for the upper bound of ρ_2 and find

$$\rho_3 \lesssim (h'-h)\int_{\mathbb{R}} \big((\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)\big)^2 ds - 2h\big(-1+\tfrac{h}{h'}\big)\int_{\mathbb{R}} (\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)s(\mathcal{H}_d\widetilde{\phi}^{(d)})(\xi)ds$$

for some ξ between s and $\frac{h}{h'}s$. Recall that $|(\mathcal{H}_d\widetilde{\phi}^{(d)})(u)| \lesssim (1+u^2)^{-1}$ and $|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(u)| \lesssim (1+u^2)^{-1/2}$. Thus,

$$\int_{\mathbb{R}} \big|(\mathcal{H}_d\widetilde{\phi}^{(d-1)})(s)s(\mathcal{H}_d\widetilde{\phi}^{(d)})(\xi)\big|ds \lesssim \int_{\mathbb{R}} \big(1+s^2\big)^{-1/2}|s|\big(1+(s/2)^2\big)^{-1}ds < \infty,$$

where we used that $|\xi| \geq \frac{h}{h'}|s| > |s|/2$. Finally, $|h^2/h' - h| \leq h'-h$ implies $\rho_3 \lesssim |h'-h|$.

For the second term on the right hand side of (4.7) we use that $\mathrm{Var}(Y_{\mathbf{t},h,\mathbf{v}}) = h$. Using again the uniform boundedness from above and below of $\sigma_{\mathbf{t},h,\mathbf{v}}$ and the fact that $|\sqrt{x}-\sqrt{y}|^2 \leq |x-y|$ for all $x, y \geq 0$ gives

$$\begin{aligned}
&\mathbb{E}\big|Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2|\sigma_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}|^2\\
=&h'\Big|\frac{(\mathbb{E}|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}}|^2)^{1/2}}{\sqrt{h}} - \frac{(\mathbb{E}|\sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}|^2)^{1/2}}{\sqrt{h'}}\Big|^2\\
\lesssim&\mathbb{E}\big|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}}\big|^2\Big|\frac{\sqrt{h'}}{\sqrt{h}}-1\Big|^2 + \big|(\mathbb{E}|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}}|^2)^{1/2} - (\mathbb{E}|\sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}|^2)^{1/2}\big|^2\\
\lesssim&|h-h'| + \mathbb{E}\big|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2.
\end{aligned}$$

For the second term in the last line, the bounds above apply which completes the proof for (4.6).

Set $\sigma^2(\mathbf{t},h,\mathbf{v}) := h$ and $\widetilde{\rho}\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) := \big(\|\mathbf{v}-\mathbf{v}'\|^2 + \|\mathbf{t}-\mathbf{t}'\| + |h-h'|\big)^{1/2}$, such that

$$\sigma^2(\mathbf{t},h,\mathbf{v}) - \sigma^2(\mathbf{t}',h',\mathbf{v}') \leq \widetilde{\rho}^2\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) \quad \text{for all } \big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) \in \overline{\mathcal{T}} \times \overline{\mathcal{T}}.$$

For fixed $\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big)\in\overline{\mathcal{T}}\times\overline{\mathcal{T}}$, the random variable $Y_{\mathbf{t},h,\mathbf{v}}-Y_{\mathbf{t}',h',\mathbf{v}'}$ follows a normal distribution with mean zero and variance bounded by a constant multiple of $\widetilde{\rho}^2\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big)$. Thus, there exists a constant $M>0$ such that for any $\eta>0$,

$$\mathbb{P}\big(|Y_{\mathbf{t},h,\mathbf{v}}-Y_{\mathbf{t}',h',\mathbf{v}'}|\geq\widetilde{\rho}((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}'))\eta\big)\lesssim\exp(-\eta^2/M).$$

Furthermore, $\mathbb{P}\big(Y_{\mathbf{t},h,\mathbf{v}}>\sqrt{h}\eta\big)\lesssim\exp(-\eta^2/2)$, as $h^{-1/2}Y_{\mathbf{t},h,\mathbf{v}}$ corresponds to a standard normal distributed random variable. Thus, conditions (i) and (ii) of Theorem 6.1 in Dümbgen and Spokoiny [15] are satisfied. As in [19] one shows that condition (iii) of Theorem 6.1 in [15] holds with $V=(3d-1)/2$ and that the process $Y_{\mathbf{t},h,\mathbf{v}}$ is almost surely continuous on $\overline{\mathcal{T}}$ with respect to ρ. The boundedness of $\sup_{(\mathbf{t},h,\mathbf{v})\in\overline{\mathcal{T}}}\big(\beta_h\frac{|X_{\mathbf{t},h,\mathbf{v}}|}{\sigma_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\beta_h\big)$ follows by an application of Theorem 6.1 and Remark 1 in [15]. □

C.4 Replacing the true densities in the limit process by estimators

Theorem C.5. *Under the assumptions of Theorem 3.2,*

$$\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h\frac{\big||X_{\mathbf{t},h,\mathbf{v}}|-|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|\big|}{\sigma_{\mathbf{t},h,\mathbf{v}}}=o_{\mathbb{P}}(1)\quad\textit{for } n\to\infty.$$

Proof. Recall the definition of the symmetrized set $\mathcal{T}_n'$ in (4.3) and let

$$\widehat{F}(s,\boldsymbol{\theta}):=\frac{\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}-\frac{\sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}.$$

Lemma A.5 and an argument as in the proof of Lemma A.4 show that $\sup_{(s,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{S}^{d-1}}\big|\widehat{F}(s,\boldsymbol{\theta})\big|=O(\log(n)^{-1})$ for $n\to\infty$ almost surely. Define

$$\Delta_{\mathbf{t},h,\mathbf{v}}:=X_{\mathbf{t},h,\mathbf{v}}-\widehat{X}_{\mathbf{t},h,\mathbf{v}}=\frac{1}{\sqrt{h}}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\langle\boldsymbol{\theta},\mathbf{v}\rangle(\mathcal{H}_d\widetilde{\phi}^{(d-1)})\Big(\frac{s-\langle\mathbf{t},\boldsymbol{\theta}\rangle}{h}\Big)\widehat{F}(s,\boldsymbol{\theta})\,dW_{s,\boldsymbol{\theta}}$$

and $\Delta_{\infty,\mathbf{t},h,\mathbf{v}}:=\log(n)^{-1}X_{\mathbf{t},h,\mathbf{v}}$. We write $\widetilde{\mathbb{P}}$ and $\widetilde{\mathbb{E}}$ for the probability and expectation conditionally on $(S_i,\boldsymbol{\Theta}_i)$, $i=n+1,\ldots,2n$. Under $\widetilde{\mathbb{P}}$, the vectors $(\Delta_{\mathbf{t},h,\mathbf{v}})_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'}$ and $(\Delta_{\infty,\mathbf{t},h,\mathbf{v}})_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'}$ are centered and normally distributed with

$$\widetilde{\mathbb{E}}\big(\Delta_{\mathbf{t},h,\mathbf{v}}-\Delta_{\mathbf{t}',h',\mathbf{v}'}\big)^2+\widetilde{\mathbb{E}}\big(\Delta_{\infty,\mathbf{t},h,\mathbf{v}}-\Delta_{\infty,\mathbf{t}',h',\mathbf{v}'}\big)^2\lesssim\log(n)^{-2}\quad\forall(\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\in\mathcal{T}_n'$$

almost surely. Hence, an application of Theorem 2.2.5 in Adler and Taylor [1] gives

$$\Big|\widetilde{\mathbb{E}}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'}\Delta_{\mathbf{t},h,\mathbf{v}}\Big)-\widetilde{\mathbb{E}}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'}\Delta_{\infty,\mathbf{t},h,\mathbf{v}}\Big)\Big|=O\big(\log(n)^{-1/2}\big)\quad\text{almost surely.}$$

Moreover, by the almost sure asymptotic boundedness of $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'} \beta_h(|X_{\mathbf{t},h,\mathbf{v}}|/\sigma_{\mathbf{t},h,\mathbf{v}} - \alpha_h)$ proved in Theorem C.4, we have $|\widetilde{\mathbb{E}}(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'} \Delta_{\infty,\mathbf{t},h,\mathbf{v}})| = O(\log(n)^{-1/2})$ almost surely. Finally, for some constant $C > 0$,

$$\begin{aligned}\widetilde{\mathbb{P}}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \frac{|\Delta_{\mathbf{t},h,\mathbf{v}}|}{\sigma_{\mathbf{t},h,\mathbf{v}}} > \log\log(n)^{-1/2}\Big) &\leq \widetilde{\mathbb{P}}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'} \Delta_{\mathbf{t},h,\mathbf{v}} > C \log\log(n)^{1/2}\log(n)^{-1/2}\Big)\\ &= O\big(\log\log(n)^{-1/2}\big)\end{aligned}$$

for $n \to \infty$ almost surely, by Markov's inequality. The constants introduced above do not depend on the second sample $(S_i, \mathbf{\Theta}_i)$, $i = n+1, \ldots, 2n$ and therefore the claim follows by an application of the law of iterated expectations. □

C.5 Replacement of the standard deviation by an estimator

Theorem C.6. *Under the assumptions of Theorem 3.2,*

(i) $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \sqrt{n}|\widehat{T}_{\mathbf{t},h,\mathbf{v}} - E[T_{\mathbf{t},h,\mathbf{v}}]|\Big|\frac{1}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \frac{1}{\sigma_{\mathbf{t},h,\mathbf{v}}}\Big| = o_{\mathbb{P}}(1)$ *for* $n \to \infty$*;*

(ii) $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h |\widehat{X}_{\mathbf{t},h,\mathbf{v}}|\Big|\frac{1}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \frac{1}{\sigma_{\mathbf{t},h,\mathbf{v}}}\Big| = o_{\mathbb{P}}(1)$ *for* $n \to \infty$*.*

Proof. We only prove *(i)* as *(ii)* follows by a similar argument. By Lemma A.3 and Lemma A.4, $\sigma_{\mathbf{t},h,\mathbf{v}}$ and $\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}$ are almost surely uniformly bounded from below for all sufficiently large n. Thus,

$$\begin{aligned}&\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \sqrt{n}|\widehat{T}_{\mathbf{t},h,\mathbf{v}} - E[T_{\mathbf{t},h,\mathbf{v}}]|\Big|\frac{1}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \frac{1}{\sigma_{\mathbf{t},h,\mathbf{v}}}\Big|\\ \lesssim &\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h\Big(\sqrt{n}\frac{|\widehat{T}_{\mathbf{t},h,\mathbf{v}} - \mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|}{\sigma_{\mathbf{t},h,\mathbf{v}}} - \alpha_h\Big) \sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} |\sigma_{\mathbf{t},h,\mathbf{v}} - \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}|\\ &+ \frac{\log(n)}{\log\log(n)} \sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} |\sigma_{\mathbf{t},h,\mathbf{v}} - \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}|\end{aligned}$$

almost surely. The claim follows from Lemma A.4, Theorems C.1 and C.3 and the almost sure boundedness of $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h(|X_{\mathbf{t},h,\mathbf{v}}|/\sigma_{\mathbf{t},h,\mathbf{v}} - \alpha_h)$ established in Theorem C.4. □

D Proofs of Theorems 3.3 and 3.4

Proof of Theorem 3.3. We have

$$\begin{aligned}\mathbb{P}\Big(\exists(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n:|\widehat{T}_{\mathbf{t},h,\mathbf{v}}|>\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha)\Big) &= 1-\mathbb{P}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h\Big(\sqrt{n}\frac{|\widehat{T}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\Big)\le\kappa_n(\alpha)\Big)\\ &= 1-\mathbb{P}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h\Big(\frac{|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\Big)\le\kappa_n(\alpha)\Big)+o(1)\\ &\le \alpha+o(1)\end{aligned}$$

for $n\to\infty$. Here we used (3.14) for the first equality and Theorem 3.2 for the second. $\square$

Proof of Theorem 3.4. We assume in the following that $c_d>0$, the case $c_d<0$ can be treated similarly. The following statement can be derived similarly as in the proof of Theorem 3.3 in [19]. For a null sequence $0<(\alpha_n)_{n\in\mathbb{N}}<1$ converging sufficiently slowly and for the set $\mathcal{T}_n'\subseteq\mathcal{T}_n$ of all triples for which the inequality

$$\int_{\mathbb{R}^d}\phi_{\mathbf{t},h}(\mathbf{b})\partial_{\mathbf{v}}f_{\boldsymbol{\beta}}(\mathbf{b})\,d\mathbf{b}<-2c_d^{-1}h^{-d-1/2}\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha_n) \tag{4.1}$$

is satisfied it holds that

$$\mathbb{P}\Big(\widehat{T}_{\mathbf{t},h,\mathbf{v}}>\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha_n)\ \text{ for all }(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'\Big)=1-o(1).$$

Hence, the hypotheses (3.5) are rejected simultaneously on the set of scales $\mathcal{T}_n'$ with asymptotic probability one. Moreover, for a mode $\mathbf{b}_0\in(0,1)^d$ of $f_{\boldsymbol{\beta}}$ and any triple $(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n^{\mathbf{b}_0}$, one can prove that $\partial_{\mathbf{v}}f_{\boldsymbol{\beta}}(\mathbf{b})\lesssim -h$ for all $\mathbf{b}\in\operatorname{supp}\phi_{\mathbf{t},h}$ by following the arguments in the proof of Theorem 3.3 in [20]. Consequently, $\int_{\mathbb{R}^d}\phi_{\mathbf{t},h}(\mathbf{b})\partial_{\mathbf{v}}f_{\boldsymbol{\beta}}(\mathbf{b})\,\mathrm{d}\mathbf{b}\lesssim -h$.

As seen in Appendix C.3-C.5

$$\left|\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h\left(\frac{|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\right)-\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h\left(\frac{|X_{\mathbf{t},h,\mathbf{v}}|}{\sigma_{\mathbf{t},h,\mathbf{v}}}-\alpha_h\right)\right|=o_{\mathbb{P}}(1)$$

for $n\to\infty$ and $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n}\beta_h(\frac{|X_{\mathbf{t},h,\mathbf{v}}|}{\sigma_{\mathbf{t},h,\mathbf{v}}}-\alpha_h)$ is finite almost surely for n sufficiently large. Moreover, $\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}$ is almost surely uniformly bounded by Lemmas A.3 and A.4 for n sufficiently large, such that

$$h^{-d-1/2}\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha_n)\lesssim\sqrt{\frac{\log n}{n}}h^{-d-1/2}$$

almost surely. In order to verify (4.1), we need to pick h such that $h^{d+3/2}\gtrsim(\log(n)/n)^{1/2}$. Thus, (4.1) holds for $h\ge C\log(n)^{\frac{1}{2d+3}}n^{-\frac{1}{2d+3}}$ with some sufficiently large constant $C>0$. $\square$