Tests for qualitative features in the random coefficients model

Fabian Dunker[*]

School of Mathematics and Statistics, University of Canterbury, Private Bag 4800,
Christchurch 8140, New Zealand
e-mail: fabian.dunker@canterbury.ac.nz

Konstantin Eckle[*,†]

Mathematical Institute of the University of Leiden, Niels Bohrweg 1, 2333 CA Leiden,
Netherlands
e-mail: konstantin@eckles.de

Katharina Proksch[*,‡]

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

Johannes Schmidt-Hieber[*,§]

University of Twente, 5 Drienerlolaan, 7522 NB Enschede, Netherlands
e-mail: a.j.schmidt-hieber@utwente.nl

Abstract: The random coefficients model is an extension of the linear regression model that allows for unobserved 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, non-parametric 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.

MSC 2010 subject classifications: Primary 62G10, 62G15; secondary 62G20.

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

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

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

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

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

Random coefficients models appear in econometrics and epidemiology and are used to model unobserved heterogeneity in the population. While the standard linear regression model accounts for unobserved heterogeneity only by the additive noise, the random coefficients model allows in addition that different individuals have different slopes. Applications in epidemiology are considered by Greenland (2000); Gustafson and Greenland (2006). In economics, random coefficients models are frequently used to evaluate panel data, cf. Hsiao (2014) or Hsiao and Pesaran (2004), Chapter 6, for an overview. Modeling and estimating consumer demand in industrial organization and marketing often makes use of random coefficients Berry et al. (1995); Petrin (2002); Nevo (2001); Dubé et al. (2012). In all these works, parametric assumptions on $f_{\boldsymbol{\beta}}$ are imposed. Recently, nonparametric approaches for random coefficients became popular in microeconometrics Hoderlein et al. (2010); Masten (2018); Hoderlein et al. (2017), frequently combined with binary choice Ichimura and Thompson (1998); Gautier and Hoderlein (2012); Gautier and Kitamura (2013); Masten and Torgovitsky (2016); Dunker et al. (2013); Fox and Gandhi (2016); Dunker et al.

(2018), among others. Model (1) without intercept has been considered in this context by Berry and Pakes (2007) and Dunker et al. (2017).

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

$$Y_i = Q_i \cos(\Phi_i) + P_i \sin(\Phi_i), \quad i = 1, \ldots, n, \tag{3}$$

with (Q_i, P_i) i.i.d. random variables which are unobserved and independent of Φ_i. The angles 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). In quantum mechanics, it is impossible to measure/observe the random variables P_i and Q_i simultaneously and the Wigner function can take negative values. For more on quantum homodyne tomography and the Wigner function, see Section 2 in Butucea et al. (2007).

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 (1992); Beran et al. (1996); Feuerverger and Vardi (2000); Hoderlein et al. (2010); Hohmann and Holzmann (2016); Holzmann and Meister (2019) Due to the ill-posedness of the problem and the curse of dimensionality induced by d, nonparametric estimation rates are slow. The reason is that small perturbations in the random coefficient density are indistinguishable in 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. It is therefore essential to recover increases/decreases and the modes of a density. If, say, two modes in the joint density 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 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 nonparametric 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 Dümbgen and Spokoiny (2001), density estimation Dümbgen and Walther (2008) and deconvolution Schmidt-Hieber et al. (2013). In multivariate settings the classical KMT approximation suffers from the curse of dimensionality which then leads to very restrictive conditions on the usable scales. Instead, very recent results on Gaussian approximations of suprema of multivariate empirical processes developed by Chernozhukov et al. (2017) can be used (see Eckle et al. (2017); Proksch et al. (2018) for applications to multivariate deconvolution and multivariate linear inverse problems with additive noise). 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 can 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 (1993) considers goodness-of-fit testing and Swamy (1970); Andrews (2001) test whether some of the random coefficients are deterministic. The only test based on a nonparametric assumption was proposed recently by Breunig and Hoderlein (2018). 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. An R package and Python code is available from https://arxiv.org/abs/1704.01066.

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, $a_n \leq Cb_n$ 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, see Beran et al. (1996). 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. For a density $f \in L^1(\mathbb{R}^d)$, let Rf denote its Radon transform, given by

$$Rf(s, \boldsymbol{\theta}) = \int_{\langle \mathbf{b}, \boldsymbol{\theta} \rangle = s} f(\mathbf{b}) \, d\mu_{d-1}(\mathbf{b}),$$
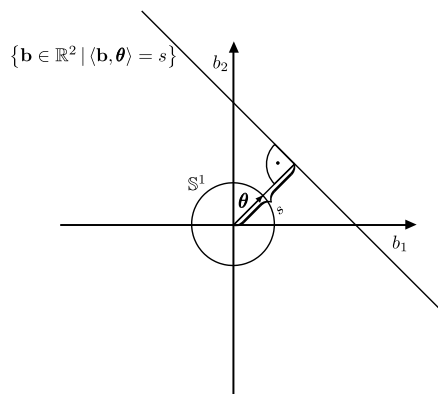
where μ_{d-1} the surface measure on the $(d-1)$-dimensional hyperplane $\{\mathbf{b} \in \mathbb{R}^d : \langle \mathbf{b}, \boldsymbol{\theta} \rangle = s\}$. The Radon transform maps therefore a function to all its integrals over hyperplanes parametrized by $(s, \boldsymbol{\theta}) \in \mathcal{Z}$. The figure above shows the parametrization in two dimensions. For the connection between the Radon transform and the random coefficients model (1) consider the normalized observations

$$S_i := \frac{Y_i}{\|\mathbf{X}_i\|}, \quad \boldsymbol{\Theta}_i := \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}, \quad i = 1, \dots, n. \tag{4}$$

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 (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

$$f_{S|\boldsymbol{\Theta}}(s|\boldsymbol{\theta}) = (Rf_{\boldsymbol{\beta}})(s, \boldsymbol{\theta}). \tag{5}$$

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 nonparametric estimation of $f_{S|\boldsymbol{\Theta}}$ via $f_{S|\boldsymbol{\Theta}}(s|\boldsymbol{\theta}) = f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})/f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$. 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 happens if the normalized design variables $\boldsymbol{\Theta}_i = \mathbf{X}_i/\|\mathbf{X}_i\|$ systematically miss observations from some directions. In this case the problem becomes unevenly harder, see Davison (1983); Frikel (2013); Hohmann and Holzmann (2016). For example, if the support of $\boldsymbol{\Theta}_i$ does contain a hemisphere, only logarithmic convergence rates can be obtained. 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 for $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Define the operator Λ via

$$\Lambda f(s, \boldsymbol{\theta}) = \mathcal{H}_d \partial_s^{d-1} R f(s, \boldsymbol{\theta}), \tag{6}$$

where $\mathcal{H}_d$ denotes the identity for d odd and the Hilbert transform

$$\mathcal{H}_d f(u) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \int_{(-\infty, u-\epsilon] \cup [u+\epsilon, \infty)} \frac{f(t)}{u-t} dt = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{f(t)}{u-t} dt \quad \left(u \in \mathbb{R}\right)$$

taken 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. If φ is a Schwartz function on $\mathbb{R}^d$, then we have the inversion formula

$$\varphi = c_d^{-1} R^* \Lambda \varphi, \tag{7}$$

cf. Theorem 3.8 in Helgason (2011). The so called back projection operator R^* is the adjoint of the Radon transform with respect to the L^2 scalar product. Notice that our constant c_d differs from the constant in Helgason (2011) as we use the standard definition of the Hilbert transform and define 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. (2013), 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

$$\int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) \, d\mathbf{b} = \sum_{k=1}^d \mathrm{v}_k \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \frac{\partial}{\partial \mathrm{b}_k} f_{\boldsymbol{\beta}}(\mathbf{b}) \, d\mathbf{b} \tag{8}$$

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

$$\phi_{\mathbf{t},h}(\cdot) = \frac{1}{h^d \operatorname{Vol}(\mathbb{S}^{d-2})} \phi\left(\frac{\|\cdot - \mathbf{t}\|}{h}\right) \tag{9}$$

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 $[0, 1]$. Moreover, $\operatorname{Vol}(\mathbb{S}^{d-2})$ 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 $\phi_{\mathbf{t},h}$ turns out to be convenient but does not entail that $\phi_{\mathbf{t},h}$ integrates to one.

If the integral (8) 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 (8) 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 (8).

Let $\mathcal{T} = \{(\mathbf{t}, h, \mathbf{v}) : h \in (0, 1], \mathbf{a}_1 + h \leq \mathbf{t} \leq \mathbf{a}_2 - h, \mathbf{v} \in \mathbb{S}^{d-1}\}$ for $\mathbf{a}_1 \leq \mathbf{a}_2 \in \mathbb{R}^d$, where the inequalities for the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{t}$ are understood componentwise and vector plus (minus) scalar means that the scalar is added to (subtracted from) each entry of the vector. 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

$$H_{0,+}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} \leq 0 \quad \text{versus} \quad H_{1,+}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} > 0 \tag{10}$$

and

$$H_{0,-}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} \geq 0 \quad \text{versus} \quad H_{1,-}^{\mathbf{t},h,\mathbf{v}} : \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b} < 0. \tag{11}$$

For constructing global tests, we can now argue as in Eckle et al. (2017). Our main interest are the following three global testing problems (i)-(iii).

(i) Testing for the presence of a mode at a fixed location. Tests for the hypotheses (10) and (11) can be used for 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 use several bandwidths/scales h and for each h 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 studying the testing problem

$$H_{0,-}^{\mathbf{t}_j, h, \mathbf{v}_j} \quad \text{versus} \quad H_{1,-}^{\mathbf{t}_j, h, \mathbf{v}_j}, \tag{12}$$

with h ranging over all chosen scales and $j = 1, \ldots, p$. Level and power of the mode test (12) for different designs are reported in Section 4.2. In Section 4.3, we also show that it is essential to include several bandwidths/scales h in order to separate modes which are close.

(ii) A global testing procedure for all modes. Simultaneous tests for the hypotheses (10) and (11) can be used for a global testing procedure to detect all modes of the density on a domain. Compared to the previous case, we search for evidence over a range of different $\mathbf{b}_0$ which then inflates the number of local tests.

(iii) A graphical representation of the local monotonicity behavior for bivariate densities. Let $d = 2$ and 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 restrict the testing procedure to one fixed scale for an easy to read graphical representation. 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. $H_{0,+}^{\widetilde{\mathbf{t}}_j, h_0, \widetilde{\mathbf{v}}_j} = H_{0,-}^{\widetilde{\mathbf{t}}_j, h_0, -\widetilde{\mathbf{v}}_j}$. Therefore, we 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 (10) and $\mathcal{T}_n$ as 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 an indication of a negative directional derivative of $f_{\boldsymbol{\beta}}$ in direction $\widetilde{\mathbf{v}}_j$ at the location $\widetilde{\mathbf{t}}_j$. Thus, Figure 1 provides strong evidence that the density is trimodal with modes close to the locations $(-0.5, -0.5)^\top$, $(1.5, -0.5)^\top$, and $(0.5, 1.5)^\top$. A detailed description of the settings used to generate Figure 1 and an analysis of the results is given in Section 4.1.

We now derive an empirical counterpart of the functional (8) in terms of the Radon transform $Rf_{\boldsymbol{\beta}}$. We make the following assumptions for the inversion of the Radon transform.

Assumption 1. *Suppose that the function ϕ in (9) 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 $c_{\boldsymbol{\beta}} > 0$*

$$f_{\boldsymbol{\beta}}(\mathbf{b}) \geq c_{\boldsymbol{\beta}} \quad \text{for all} \quad \mathbf{b} \in [\mathbf{a}_1, \mathbf{a}_2].$$
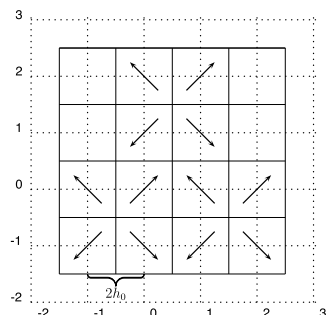
FIG 1. *Example of a global map for monotonicity of a bivariate density. Each arrow indicates a decrease in the respective direction.*

Assumption 2 is too restrictive for quantum homodyne tomography (model (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 this case, we replace Assumption 2 by the following conditions.

Assumption 2′. *Suppose that $f_{\boldsymbol{\beta}}$ is continuously differentiable and, for some $\gamma', \varepsilon > 0$,*

(i) $\mathbf{b} \mapsto \|\mathbf{b}\|^{d+\varepsilon}|f_{\boldsymbol{\beta}}(\mathbf{b})|$ *is bounded;*

(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 $P \cap [\mathbf{a}_1 - \delta, \mathbf{a}_2 + \delta] \neq \emptyset$, $\int_P f_{\boldsymbol{\beta}}(\mathbf{b}) d\mu_{d-1}(\mathbf{b}) \geq c_{\boldsymbol{\beta}}$.

Under the assumptions above the inversion formula (7) 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 (2011). The following lemma analyzes the structure of a partial derivative of the test function transformed by the operator Λ introduced in (6) and how this transform depends on h.

Lemma 3.1. *Work under Assumption 1 and let*

$$\widetilde{\phi}(z) := \int_0^\infty r^{d-2} \frac{\partial}{\partial z} \phi\left(\sqrt{z^2 + r^2}\right) dr \qquad with \ z \in \mathbb{R}. \tag{13}$$

Then

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

where Λ and $\phi_{\mathbf{t},h}$ are defined in (6) and (8), respectively. Moreover,

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

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

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

By an application of the inversion formula introduced in (7) and Lemma 5.1 in Helgason (2011), we obtain

$$\mathbb{E}\big[T_{\mathbf{t},h,\mathbf{v}}\big] = -c_d h^{d+1/2} \int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) d\mathbf{b}. \tag{14}$$

Up to rescaling, $T_{\mathbf{t},h,\mathbf{v}}$ is thus an empirical counterpart of the functional defined in (8).

The statistic $T_{\mathbf{t},h,\mathbf{v}}$ depends on the density $f_{\boldsymbol{\Theta}}$. In quantum homodyne tomography this density is known. For many other applications, however, $f_{\boldsymbol{\Theta}}$ needs to be estimated from the data. In this case, we use a standard cut-off kernel density estimator $\widetilde{f}_{\boldsymbol{\Theta}}$ for $f_{\boldsymbol{\Theta}}$ based on an additional sample $(S_i, \boldsymbol{\Theta}_i)$, $i = n+1, \ldots, 2n$ which is independent of $(S_i, \boldsymbol{\Theta}_i)$, $i = 1, \ldots, n$. See also Appendix A for the definition of the estimator. We replace $f_{\boldsymbol{\Theta}}$ by its estimator and consider the test statistic

$$\widehat{T}_{\mathbf{t},h,\mathbf{v}} := \frac{1}{n\sqrt{h}} \sum_{i=1}^{n} \frac{\langle \boldsymbol{\Theta}_i, \mathbf{v} \rangle}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}_i)} (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{S_i - \langle \mathbf{t}, \boldsymbol{\Theta}_i \rangle}{h} \Big). \tag{15}$$

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 $f_{\boldsymbol{\Theta}}$ approaches zero for some directions. This section provides conditions on the design which ensure that $f_{\boldsymbol{\Theta}}$ has positive Hölder smoothness and is bounded from below and above. These results are of independent interest.

In the random coefficients model (1), the density $f_{\boldsymbol{\Theta}}$ can be expressed in terms of the density $f_{\mathbf{X}}$ via $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \int_0^{\infty} r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta}) dr$. To enforce that $f_{\boldsymbol{\Theta}}$ is bounded from below we restrict ourselves to designs where $\int_0^{\infty} r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta}) dr$ is bounded away from 0. The formula also allows us to relate the smoothness of $f_{\mathbf{X}}$ to the smoothness of $f_{\boldsymbol{\Theta}}$.

Although, the random coefficients model with intercept (2) could be viewed as a special case of the more general model (1) and vice versa, it requires a different set of assumptions. For model (2), we write $f_{\mathbf{X}}$ as a function of $\boldsymbol{x} = (x_2, \ldots, x_d) \in \mathbb{R}^{d-1}$ and obtain

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{1}{2|\theta_1|^d} f_{\mathbf{X}}\Big(\frac{\theta_2}{\theta_1}, \ldots, \frac{\theta_d}{\theta_1} \Big), \tag{16}$$

see Appendix B for a proof. A necessary condition to ensure that $\inf_{\boldsymbol{\theta}} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) > 0$, is given by $f_{\mathbf{X}}(\boldsymbol{x}) \gtrsim 1 \wedge \|\boldsymbol{x}\|^{-d}$ for all $\boldsymbol{x}$. This corresponds to Cauchy-type tails of the design variables. Thinner tails will increase the ill-posedness of the problem. In order to avoid very technical proofs, we consider in the random coefficients model with intercept only the case where $(X_{i,2}, \ldots, X_{i,d})$ follows a multivariate Cauchy distribution, that is,

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{\Gamma(d/2)}{\pi^{d/2}|\Sigma|^{1/2}\big(1 + (\boldsymbol{x} - \boldsymbol{\mu})^{\top}\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\big)^{d/2}} \quad \text{for} \quad \boldsymbol{x} \in \mathbb{R}^{d-1}, \quad (17)$$

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 $f_{\boldsymbol{\Theta}}$ explicitly using (16)

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{\pi^{-d/2}|\Sigma|^{-1/2}\Gamma(d/2)}{2\big(\theta_1^2 + ((\text{sgn}(\theta_1)\theta_j - |\theta_1|\mu_j)_{j=2}^d)^{\top}\Sigma^{-1}((\text{sgn}(\theta_1)\theta_j - |\theta_1|\mu_j)_{j=2}^d)\big)^{d/2}}, \tag{18}$$

where $\text{sgn}(\cdot)$ denotes the signum function. In this case, $f_{\boldsymbol{\Theta}}$ is bounded from above and below and is 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 $f_{\boldsymbol{\Theta}}$ is constant. This leads to the following assumptions on the design.

Assumption 3. *In model* (1), *suppose that*

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

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

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

In model (2), *assume that* $f_{\mathbf{X}}$ *is of the form* (17) *with* Σ *a symmetric and positive definite matrix.*

In quantum homodyne tomography we set a global γ equal to the minimum of γ' from Assumption 2' and γ from Assumption 3.

It is important to notice that statistical testing in the random coefficients model relies on two unrelated sets of assumptions. Firstly, there are assumptions on the density of the random coefficients as introduced in Section 3.1. Restrictions of this kind are common in statistical inference for an unknown density. On the other hand, there are assumptions (see Assumption 3 above) on the design. These assumptions control the ill-posedness of the problem.

Note that Assumption 3 (ii) can be weakened to $\int_{-\infty}^{\infty} r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta}) dr \geq c > 0$ for all $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. For example, if the support of Θ is a hemisphere this condition can hold while Assumption 3 (ii) is violated. This relaxation can be achieved by multiplying independently generated ζ_i to Θ_i and S_i as proposed for model (2) in Section 2.

3.3. Asymptotic properties

This section presents the main theoretical result of the paper stating that the standardized and properly calibrated test statistic (15) 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

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

(Adler and Taylor, 2007, Chapter 1.4.3). W is a random, finitely additive, signed measure. Integration with respect to 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 (2007), 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

$$X_{\mathbf{t},h,\mathbf{v}} = h^{-1/2} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle (\mathcal{H}_d \widetilde{\phi}^{(d-1)}) \Big(\frac{s - \langle \mathbf{t}, \boldsymbol{\theta} \rangle}{h} \Big) \frac{\sqrt{f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})}}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} W(dsd\boldsymbol{\theta}). \tag{19}$$

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 second part of the sample. To this end, we use the

standard cut-off kernel density estimates $\widetilde{f}_{\boldsymbol{\Theta}}$ and $\widetilde{f}_{S,\boldsymbol{\Theta}}$ defined in Appendix A. 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

$$\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}} := \Big(\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \frac{\widetilde{f}_{S,\boldsymbol{\Theta}}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})^2} ds d\boldsymbol{\theta} \Big)^{1/2}. \quad (20)$$

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. *When working in model* (1), *let γ be as defined in Assumption 3. When considering model* (2), *let $\gamma = 1$. Set $h_{\min} := \min\{h : (\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n\}$ 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 $h_{\max} \lesssim \log(n)^{-14\gamma/(d-1)-5} n^{2\gamma/(d-1)-1} \wedge o(1)$, $h_{\min} \gtrsim n^{-1+\varepsilon}$ 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 $A = \{\boldsymbol{x} \in \mathbb{R}^p : -\infty < \boldsymbol{x} \le \boldsymbol{a}\}$ 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*

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

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, almost surely conditionally on the estimators $\widetilde{f}_{S,\boldsymbol{\Theta}}$ and $\widetilde{f}_{\boldsymbol{\Theta}}$, the limit distribution is bounded in probability by a constant that is independent of the sample size n.

3.4. Construction of the multiscale test

With the previous theorem, we can now construct simultaneous statistical tests for the hypotheses (10) and (11). If the constant c_d is positive then the method consists of rejecting the hypotheses $H_{0,+}^{\mathbf{t},h,\mathbf{v}}$ in (10) for small values of $\widehat{T}_{\mathbf{t},h,\mathbf{v}}$ and rejecting $H_{0,-}^{\mathbf{t},h,\mathbf{v}}$ in (11) 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

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

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

$$\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha) = \frac{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}}{\sqrt{n}}\left(\beta_h^{-1}\kappa_n(\alpha) + \alpha_h\right) \tag{21}$$

and reject the hypothesis (10), if

$$\operatorname{sgn}(c_d)\widehat{T}_{\mathbf{t},h,\mathbf{v}} < -\kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha). \tag{22}$$

Similarly, hypothesis (11) is rejected, whenever

$$\operatorname{sgn}(c_d)\widehat{T}_{\mathbf{t},h,\mathbf{v}} > \kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha). \tag{23}$$

Theorem 3.3. *Let the assumptions of Theorem 3.2 hold and assume that the tests* (22) *and* (23) *for the hypotheses* (10) *and* (11) *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, i.e.*

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

for $n \to \infty$.

Based on the previous result, we now propose a method for the detection and localization of modes on a subdomain. For convenience, we only consider the case of a hyperrectangle $[\mathbf{a}_1, \mathbf{a}_2] \subset \mathbb{R}^d$. We study the case that there is a mode $\mathbf{b}_0$ in the interior $(\mathbf{a}_1, \mathbf{a}_2)$. For the multiscale test to have power we need that the set of local tests is rich enough. This can be expressed in terms of conditions on $\mathcal{T}_n$. Let $H(\mathcal{T}_n)$ be the set of all scales/bandwidths such that for every scale h in this set all triples $(\mathbf{t},h,\mathbf{v})$ are in $\mathcal{T}_n$, where $\mathbf{t}$ ranges over all grid points of an equidistant grid with component wise mesh size h in the hyperrectangle $[\mathbf{a}_1 + h, \mathbf{a}_2 - h]$, and $\mathbf{v}$ ranges over all grid points of a grid of S^{d-1} and grid width converging to zero with increasing sample size.

The testing procedure is as follows. For any $\mathbf{b}_0$ in $(\mathbf{a}_1, \mathbf{a}_2)$, let $\mathcal{T}_n^{\mathbf{b}_0}$ be the set of all sequences of triples $(\mathbf{t}_n, h_n, \mathbf{v}_n)_n \in \mathcal{T}_n$ such that $ch_n \geq \|\mathbf{b}_0 - \mathbf{t}_n\| \geq 2h_n$

for some sufficiently large $c > 2$ and $\angle(\mathbf{t}_n - \mathbf{b}_0, \mathbf{v}_n) \to 0$ as $n \to \infty$, where $\angle$ denotes the angle between two vectors. The previous conditions ensure that several such sequences can be found. If for all triples in $\mathcal{T}_n^{\mathbf{b}_0}$ all local tests (23) reject the hypotheses (12), we have evidence for the existence of a mode at the point $\mathbf{b}_0$. 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. *Assume the conditions of Theorem 3.2 and suppose that for any mode* $\mathbf{b}_0$ *in* $(\mathbf{a}_1, \mathbf{a}_2)$ *there are functions* $g_{\mathbf{b}_0} : \mathbb{R}^d \to \mathbb{R}$, $r_{\mathbf{b}_0} : \mathbb{R} \to \mathbb{R}$ *such that the density* $f_{\boldsymbol{\beta}}$ *has a representation of the form*

$$f_{\boldsymbol{\beta}}(\mathbf{b}) \equiv (1 + g_{\mathbf{b}_0}(\mathbf{b})) r_{\mathbf{b}_0}(\|\mathbf{b} - \mathbf{b}_0\|) \tag{24}$$

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 $\min\{h : h \in H(\mathcal{T}_n)\} \geq C \log(n)^{1/(2d+3)} n^{-1/(2d+3)}$ *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* $n \to \infty$.

The rate for the localization of the modes is $n^{-1/(2d+3)}$ (up to some logarithmic factor). The expected optimal rate for mode detection in an inverse problem with ill-posedness of degree r over a 2-Hölder class is $n^{-1/(d+2r+4)}$. This has been rigorously proven in some special cases such as univariate density deconvolution, see Wieczorek (2010). Since the ill-posedness of the Radon transform is $\frac{d-1}{2}$ and $2d + 3 = d + 2\frac{d-1}{2} + 4$, the obtained rate matches the expected optimal rate. Assumption 4 requires $h_{\max} \lesssim \log(n)^{-14\gamma/(d-1)-5} n^{2\gamma/(d-1)-1}$. To be able to include scales of the order $n^{-1/(2d+3)}$ we need $\gamma > (d^2 - 1)/(2d + 3)$. 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 (10) and (11) 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 (12)) 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 $d = 2, 3$. Throughout this section the nominal level is fixed as $\alpha = 0.05$, and all level and power statements are in percent. Except for Table 2, none of the simulations in this section assumes knowledge of the design density f_{Θ} or uses a parametric specification of it.
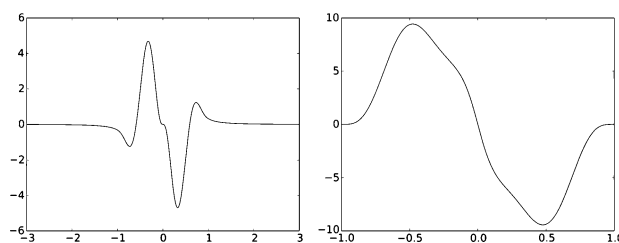

FIG 2. *The function* $\mathcal{H}_d \widetilde{\phi}^{(d-1)}$ *for* $d = 2$ *(left) and* $d = 3$ *(right).*

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 a bivariate density of random coefficients. To test the hypotheses (11) we use (23) 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 is

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

The data are simulated with $f_{\boldsymbol{\beta}}$ the density of the normal mixture

$$\tfrac{1}{3}\mathcal{N}((-0.4, -0.57)^{\top}, 0.2I) + \tfrac{1}{3}\mathcal{N}((1.5, -0.52)^{\top}, 0.2I) + \tfrac{1}{3}\mathcal{N}((0.45, 1.6)^{\top}, 0.15I).$$

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 (11). 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 detects the true modes fairly well.

4.2. Influence of the design on the power

Given the random coefficients model with $d = 3$, we study the power of the test for the existence of a mode at a given location $\mathbf{b}_0$ considering only few local tests. The postulated mode is given by the point $\mathbf{b}_0 = (0, 0, 0)^{\top}$ and we take $\mathcal{T}_n = \{(\mathbf{t}, h, \mathbf{v}) : h = 1, \mathbf{t} = \mathbf{v} \in \{\pm \mathbf{e}_i, i = 1, 2, 3\}\}$ with $\mathbf{e}_i$ the i-th standard

unit vector in $\mathbb{R}^3$. We conclude that $f_{\boldsymbol{\beta}}$ has a local maximum at the point $\mathbf{b}_0$, whenever all hypotheses $H_{0,-}^{\mathbf{t}_j,h_0,\mathbf{v}_j}$, $j = 1, \ldots, 6$, are rejected, that is,

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

with $\kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha)$ as defined in (21). Recall that the quantiles $\kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha)$ are constructed in such a way that the probability of at least one false rejection within the six tests (4.1) is bounded by α. However, the mode test detects the presence of a mode whenever all six tests (4.1) are rejected at the same time. The multiscale method is therefore rather conservative for the specific task of mode detection. In this simulation, we also study a calibrated version of the test where the quantiles are chosen such that the test keeps its nominal level $\alpha = 0.05$ and detects the presence of a non existing mode in about 5 percent of the simulation runs. For the calibration of the test we work under the null hypothesis assuming that $f_{\boldsymbol{\beta}}$ is uniform. Therefore, knowledge about the true unknown density $f_{\boldsymbol{\beta}}$ is not required.

Numerical simulations for random coefficients model without intercept: At first, we consider model (1) with uniform design $\mathbf{X}_i \sim \text{Unif}[-5, 5]^3$. To study the level of the test, we used $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \propto \mathbf{1}(\boldsymbol{\beta} \in [-5, 5]^3)$. For the power we took $f_{\boldsymbol{\beta}}$ as the density of a trivariate standard normal distribution. All results are based on the local tests (4.1) and 1000 repetitions. Level simulations with the theoretical quantiles confirm that the multiscale test keeps its nominal level as the percentage of false rejections of at least one of the six hypotheses in (4.1) for sample size $n \in \{250, 500, 1000\}$ is nearly 5 percent. The results of the mode test are reported in Table 1.

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 only consider the calibrated mode test. Results are reported in columns five and six in Table 1. Compared to the case of uniform design, we observe a decrease in the power of the mode test (4.1). The explanation is that the uniform design on the $\mathbf{X}_i$ induces a more uniform distribution of $\boldsymbol{\Theta}_i$ on the sphere which makes it simpler to recover information about the joint density as discussed in Section 2.

TABLE 1

Simulated level and power of the test (4.1) *for a mode considering uniform design* $\mathbf{X}_i \sim \text{Unif}[-5, 5]^3$ *and normal design* $\mathbf{X}_i \sim \mathcal{N}((3, 0, 0)^\top, 2I)$. *Results with theoretical quantiles are in the second column. Results where* $\kappa_n^{\mathbf{t}_j,h_0,\mathbf{v}_j}(\alpha)$ *in* (4.1) *are replaced by calibrated quantiles are in all other columns.*

		$\mathbf{X}_i \sim \text{Unif}[-5, 5]^3$		$\mathbf{X}_i \sim \mathcal{N}((3, 0, 0)^\top, 2I)$	
n	power	level (cal.)	power (cal.)	level (cal.)	power (cal.)
250	9	4.5	91.0	4.7	66.1
500	15.7	4.6	99.1	4.5	78.8
1000	79.1	5.0	100	4.5	90.6

Numerical simulations for the random coefficients model with intercept: We study model (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_Θ is constant. Except for the different design, we consider otherwise the same test settings as above. The simulated level and power of the calibrated version of the test (4.1) are reported in Table 2. To investigate the influence of the estimation of the design density f_Θ on the power of the test we compare this with simulations for which we assume that the density f_Θ is known to be constant. Compared to the power approximations for unknown f_Θ,

TABLE 2

Same as Table 1 but for random coefficients model with intercept and $(X_{i,2}, X_{i,3})^\top$ from a standard bivariate Cauchy distribution. In the fourth and fifth column we assume that the density f_Θ is known.

	f_Θ unknown		f_Θ known	
n	level (cal.)	power (cal.)	level (cal.)	power (cal.)
250	4.8	91.3	5.0	93.3
500	5.2	99.0	5.2	99.7
1000	5.3	100	5.3	100

we observe only a slight increase in power of the test for known f_Θ.

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

TABLE 3

Same as Table 1 but for random coefficients model with intercept and $(X_{i,2}, X_{i,3})^\top \sim \mathcal{N}((0,0)^\top, I)$ (second and third column) and $(X_{i,2}, X_{i,3})^\top \sim$ Unif$[-5, 5]^2$ (fourth and fifth column).

	$(X_{i,2}, X_{i,3})^\top \sim \mathcal{N}((0,0)^\top, I)$		$(X_{i,2}, X_{i,3})^\top \sim$ Unif$[-5, 5]^2$	
n	level (cal.)	power (cal.)	level (cal.)	power (cal.)
250	5.1	88.3	5.1	64.8
500	5.5	98.3	5.1	78.1
1000	5.1	100	5.5	89.4

4.3. Multiscale mode testing

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

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

We consider a design such that $\boldsymbol{\Theta}$ is uniformly distributed on the circle $\mathbb{S}^1$. The level α is fixed to five percent. We conducted simultaneously twelve tests
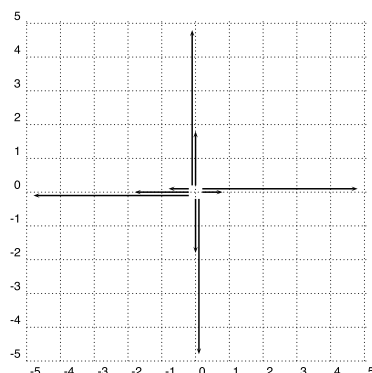
FIG 3. *Rejected hypotheses of the twelve tests. Each arrow indicates a decrease on a scale that corresponds to the length of the arrow.*

with three different scales $h \in \{0.5, 1, 2.5\}$ for the hypotheses (12) with $\mathbf{b}_0 = (0,0)^\top$. The tests are given by $\{(\mathbf{t}_i, h_j, \mathbf{v}_i) : i = 1, \ldots, 4; h_j \in \{0.5, 1, 2.5\}, \mathbf{v}_i \in \{\pm\mathbf{e}_1, \pm\mathbf{e}_2\}, \mathbf{t}_i = h_j\mathbf{v}_i\}$. We analyze the outcome of the twelve tests in two ways. Firstly, we investigate the performance of each of the three tests for modality for the bandwidths $h = 2.5$, $h = 1$ and $h = 0.5$ separately. Secondly, we consider the performance of a combined test for modality using two scales $h = 1$ and $h = 0.5$. The power approximations for sample sizes $n \in \{2000, 5000, 15000\}$ based on 1000 repetitions are displayed in Table 4.

TABLE 4
Power of the multiscale test for modality. The mode tests for the bandwidths $h = 2.5$, $h = 1$ and $h = 0.5$ are in the second, third and fourth column. The combined mode test using $h = 1$ and $h = 0.5$ is in the fifth column.

n	$h = 2.5$	$h = 1$	$h = 0.5$	$h \in \{0.5, 1\}$
2000	100	0	0	25
5000	100	0	1	86.3
15000	100	0	68.7	100

Figure 3 illustrates the results of the twelve tests for the hypotheses (12) conducted simultaneously. Each arrow at a location $\mathbf{t}$ in direction $\mathbf{v}$ displays a rejection of a hypothesis (12) and the length of the arrows corresponds to the respective bandwidths.

The results in Table 4 show that the mode test for bandwidth $h = 2.5$ detects in all cases a mode in a neighborhood of $(0,0)^\top$. However, this bandwidth is too large to distinguish between the underlying modes of $f_{\boldsymbol{\beta}}$ at $(0,0)^\top$ and $(2,0)^\top$. The test for the bandwidth $h = 1$ fails to detect the mode as $h = 1$ is still too large to separate the two modes of $f_{\boldsymbol{\beta}}$. On the contrary, $h = 0.5$ 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 $f_{\boldsymbol{\beta}}$. This effect vanishes for increasing sample size. By conducting the tests for bandwidths $h = 1$ and

$h = 0.5$ simultaneously, we are able to detect the mode at $(0,0)^\top$ in most of the simulations.

4.4. Comparison with the parametric random coefficients model

It is common in the literature to use a parametric specification of the random coefficients. Usually, $\boldsymbol{\beta}$ is assumed to have a multivariate normal distribution. In this section, we compare our nonparametric approach with mode estimation under a parametric specification of the random coefficient distribution. Our findings are similar to what is usually observed when parametric and nonparametric methods are compared. If the parametric specification of the model is consistent with the data generating process, the parametric approach outperforms the nonparametric one in terms of estimation errors, power of tests, and computation time. However, when the parametric specification and the data generating process differ, a model misspecification bias is present in the parametric estimation. This will not vanish, even if the sample size is large. The nonparametric model does not suffer from this problem and can in these cases perform better than the parametric model. We illustrate this effect in the following simulations.

We consider the parametric random coefficients model $Y = \tilde{\boldsymbol{\beta}}^\top \mathbf{X}$ where the density of $\tilde{\boldsymbol{\beta}}$ belongs to some parametric family $f_{\tilde{\boldsymbol{\beta}}}(\mathbf{b}; \eta)$ with parameter η. Note that we can rewrite the equation as a mixed model $Y = \boldsymbol{\gamma}^\top \mathbf{X} + \mathbf{u}^\top \mathbf{X}$. Here $\boldsymbol{\gamma} = \mathbb{E}(\tilde{\boldsymbol{\beta}})$ are fixed effects and $\mathbf{u} = \tilde{\boldsymbol{\beta}} - \mathbb{E}(\tilde{\boldsymbol{\beta}})$ are random effects with expectation $\mathbf{0}$. If $f_{\tilde{\boldsymbol{\beta}}}(\mathbf{b}; \eta)$ is unimodal with a mode at $\mathbb{E}(\tilde{\boldsymbol{\beta}})$, we just need to derive confidence statements for $\boldsymbol{\gamma}$. This is in particular true if $f_{\tilde{\boldsymbol{\beta}}}(\mathbf{b}; \eta)$ is a normal density, which is the most common choice. A simple and efficient way to estimate $\boldsymbol{\gamma}$ is heteroscedasticity robust regression. Obviously, the computational complexity of this method is much smaller than the complexity of our algorithm.

Our test example is the random coefficients model with intercept (2) with $d = 3$ and $(X_{i,2}, X_{i,3})^\top$ sampled from a standard bivariate Cauchy distribution. The distribution of the random coefficients $\boldsymbol{\beta}_i$ in the data generating process is given by *(i)* a standard Gaussian, *(ii)* the normal mixture $0.5\mathcal{N}(\mathbf{0}, 0.1I) + 0.5\mathcal{N}((2,0,0)^\top, 0.1I)$ and *(iii)* an exponential-2 distribution for $\beta_{i,1}$ in addition $(\beta_{i,2}, \beta_{i,3})^\top \sim \mathcal{N}(\mathbf{0}, 0.1I)$ independent of the first component. Table 5 reports estimates for $\boldsymbol{\gamma}$ obtained by transforming Y and $\mathbf{X}$ to S and $\boldsymbol{\Theta}$, and running a heteroscedasticity robust regression as described in Wooldridge (2013), Chapter 8.

In column *(i)* of Table 5 the parametric assumption holds and the procedure detects the true mode $\mathbf{0}$ of the density with high precision compared to the bandwidth choice $h = 1$ in the simulations presented in Table 2 in Section 4.2. In contrast, for the bimodal density in column *(ii)* the coefficient vector $(1,0,0)^\top$ does not describe a representative member of the population because the misspecification bias is too large. For the skewed distribution of column *(iii)* the estimator also fails to detect the mode of the density for the same reason.

By applying our testing procedure in the setting of columns *(ii)* and *(iii)* we can show that the OLS results do not represent modes of the density. To this

TABLE 5
Results of OLS for model $Y = \boldsymbol{\gamma}^\top \mathbf{X} + \mathbf{u}^\top \mathbf{X}$ for $d = 3$, $n = 1000$ and $(X_{i,2}, X_{i,3})^\top$ drawn from a standard bivariate Cauchy distribution. The distribution of the true random coefficients $\boldsymbol{\beta}_i$ is a standard Gaussian in column (i), the normal mixture $0.5\mathcal{N}(\mathbf{0}, 0.1I) + 0.5\mathcal{N}((2,0,0)^\top, 0.1I)$ in column (ii) and an exponential-2 distribution for $\beta_{i,1}$ and $(\beta_{i,2}, \beta_{i,3})^\top \sim \mathcal{N}(\mathbf{0}, 0.1I)$ independent of the first component in column (iii).

	(i)		(ii)		(iii)	
$\widehat{\gamma}_1$	0.00	(0.02)	1.00	(0.02)	2.03	(0.03)
$\widehat{\gamma}_2$	0.02	(0.02)	0.00	(0.01)	-0.02	(0.02)
$\widehat{\gamma}_3$	0.00	(0.02)	0.01	(0.01)	0.02	(0.02)

end, we set in *(ii)*

$$\mathbf{t} = (0.5, 0, 0)^\top, \quad h = 0.5, \quad \mathbf{v} = (1, 0, 0)^\top$$

and in *(iii)*

$$\mathbf{t} = (1, 0, 0)^\top, \quad h = 1, \quad \mathbf{v} = (1, 0, 0)^\top.$$

Both tests reject $H_{0,-}^{\mathbf{t},h,\mathbf{v}}$ (in *(ii)* in 100% and in *(iii)* in 95.5% percent of 1000 repetitions). Therefore, our procedure shows that neither $(1, 0, 0)^\top$ in *(ii)* nor $(2, 0, 0)^\top$ in *(iii)* are modes of the underlying density. Of course, the parametric model would be able to detect the modes in cases *(ii)* and *(iii)* with a different parametric specification. However, this would require considerable a priori knowledge about the data. If we did not interpret the results as estimators for the mode but as estimators for $\mathbb{E}(\boldsymbol{\beta})$, the parametric method would perform well.

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 Deaton and Muellbauer (1980) 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 Lewbel (1997).

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 Hoderlein et al. (2017) and Breunig and Hoderlein (2018).

We apply our multiscale test to detect modes in the random coefficients for budget shares for food at home (BSF)

$$BSF_i = \beta_{i,1} + \beta_{i,2} \ln(TotExp_i) + \beta_{i,3} \ln(FoodPrice_i). \tag{4.2}$$

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 household income, expenditure and other characteristics of roughly 7000 households. We use data of the years 1997–2001 only which gives a sample size of about 33000. Budget shares of food are generated by dividing the expenditure for all food by total expenditure. Food prices are reported as relative prices in comparison to a general prize index. The variable $TotExp$ is normalized to January 2000 real prizes.

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 can achieve this by symmetrizing the design in model (4.2) as follows:

$$BSF_i = \tilde{\beta}_{i,1} + \tilde{\beta}_{i,2} \left(\ln(TotExp_i) - 5\right) + \tilde{\beta}_{i,3} \left(25 \ln(FoodPrice_i) - 0.3\right). \tag{4.3}$$

The relation of the modified model to the random coefficients in (4.2) 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 -5 and 3.7. The observations of $25 \ln(FoodPrice_i) - 0.3$ range from -1 to 1.3.

5.2. Results

For a first evaluation of the data we assumed fixed coefficients in model (4.3) and estimated the model with ordinary least squares (OLS).

TABLE 6
Results of OLS for model (4.3) with fixed coefficients.

$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$
0.1940	-0.0587	0.0005
(0.000)	(0.001)	(0.001)

In order to find modes of the density we conducted simultaneously tests on the 5% level of the form (12) on the two scales $h_1 = 0.75$ and $h_2 = 0.5$. In the simulation settings in Section 4.2 our testing procedure performed well even when the $\mathbf{X}_i$ were not Cauchy distributed. Therefore, to obtain a testing procedure which is more flexible with respect to the design, we use the non-parametric density estimator $\widehat{f}_{\boldsymbol{\Theta}}$ instead of a parametric estimation procedure.

We were testing for modes on the equidistant grid covering $[-1,1]^3$ with grid width 1. Hence, the grid had 27 nodes. For every grid point $\mathbf{b} \in \mathbb{R}^3$ tests of the hypotheses (12) were conducted for the directions and locations

$$
\begin{array}{llll}
\mathbf{t}_1 = \mathbf{b} + h_j \mathbf{e}_1, & \mathbf{v}_1 = \mathbf{e}_1, & \mathbf{t}_4 = \mathbf{b} - h_j \mathbf{e}_2, & \mathbf{v}_4 = -\mathbf{e}_2, \\
\mathbf{t}_2 = \mathbf{b} - h_j \mathbf{e}_1, & \mathbf{v}_2 = -\mathbf{e}_1, & \mathbf{t}_5 = \mathbf{b} + h_j \mathbf{e}_3, & \mathbf{v}_5 = \mathbf{e}_3, \qquad (j=1,2) \\
\mathbf{t}_3 = \mathbf{b} + h_j \mathbf{e}_2, & \mathbf{v}_3 = \mathbf{e}_2, & \mathbf{t}_6 = \mathbf{b} - h_j \mathbf{e}_3, & \mathbf{v}_6 = -\mathbf{e}_3,
\end{array}
$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^3$ denote the standard unit vectors of $\mathbb{R}^3$. We detected a single mode in the neighborhood of the grid point $(0,0,0)^\top$ for the tests with bandwidth $h_1 = 0.75$. The test for the bandwidth $h_2 = 0.5$ did not detect a mode.

In the following we use nonparametric density estimation to motivate hypotheses for the testing procedure. It is important that this estimate and the test are independent, otherwise the testing procedure would be biased and could not guarantee a bound on the error rate. 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 (4.3) with the estimator in Hoderlein et al. (2010). 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 $f_{\tilde{\beta}_2, \tilde{\beta}_3}$ based on about 16500 observations. We chose the smoothing parameters h and g in the estimator in Hoderlein et al. (2010) equal to 0.05 and 0.1, respectively. These values are hand-picked after trying different smoothing parameters. We noticed for smaller parameters an increase of small fluctuations in the estimate which indicated overfitting. Larger smoothing parameters did not change the overall shape of estimate. They just made the mode slightly flatter. We interpreted this as oversmoothing and concluded that 0.05 and 0.1 are a good compromise. Note that these bandwidth choices do not affect the level or the power of the test performed below as the test is completely independent from this estimator. The nonparametric estimate suggests that the random coefficient density of $f_{\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3}$ has one (well-pronounced) mode close to

$$
(\beta_1, \beta_2, \beta_3) = (0.25, -0.07, 0.02). \tag{4.4}
$$

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.

With the second sub-sample we studied whether a mode can be found for the location given by (4.4) on a smaller scale than in the test above. This would then indicate 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 conduct two mode tests (4.4) simultaneously for the bandwidths $h_1 = 0.07$ and $h_2 = 0.02$, that is, we test twelve hypotheses (12) for the following locations and directions
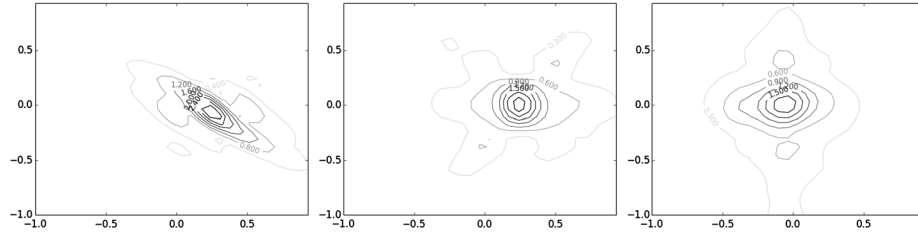
FIG 4. Nonparametric estimates of the joint densities of $f_{\tilde{\beta}_1,\tilde{\beta}_2}$ (left), $f_{\tilde{\beta}_1,\tilde{\beta}_3}$ (middle), and $f_{\tilde{\beta}_2,\tilde{\beta}_3}$ (right).

$(\mathbf{t}_i, \mathbf{v}_i)$. On scale $h_1 = 0.07$, we tested

$$\mathbf{t}_1 = (0.32, -0.07, 0.02)^\top, \ \mathbf{v}_1 = (1, 0, 0)^\top,$$
$$\mathbf{t}_2 = (0.18, -0.07, 0.02)^\top, \ \mathbf{v}_2 = -\mathbf{v}_1,$$
$$\mathbf{t}_3 = (0.25, 0, 0.02)^\top, \ \mathbf{v}_3 = (0, 1, 0)^\top,$$
$$\mathbf{t}_4 = (0.25, -0.14, 0.02)^\top, \ \mathbf{v}_4 = -\mathbf{v}_3,$$
$$\mathbf{t}_5 = (0.25, -0.07, 0.09)^\top, \ \mathbf{v}_5 = (0, 0, 1)^\top,$$
$$\mathbf{t}_6 = (0.25, -0.07, -0.05)^\top, \ \mathbf{v}_6 = -\mathbf{v}_5$$

and on scale $h_2 = 0.02$,

$$\mathbf{t}_7 = (0.27, -0.07, 0.02)^\top, \ \mathbf{v}_7 = \mathbf{v}_1, \quad \mathbf{t}_{10} = (0.25, -0.09, 0.02)^\top, \ \mathbf{v}_{10} = -\mathbf{v}_3,$$
$$\mathbf{t}_8 = (0.23, -0.07, 0.02)^\top, \ \mathbf{v}_8 = -\mathbf{v}_1, \quad \mathbf{t}_{11} = (0.25, -0.07, 0.04)^\top, \ \mathbf{v}_{11} = \mathbf{v}_5,$$
$$\mathbf{t}_9 = (0.25, -0.05, 0.02)^\top, \ \mathbf{v}_9 = \mathbf{v}_3, \quad \mathbf{t}_{12} = (0.25, -0.07, 0)^\top, \ \mathbf{v}_{12} = -\mathbf{v}_5.$$

The test rejected all local hypotheses on scale $h_1 = 0.07$ but not on scale $h_2 = 0.02$. This gives evidence that the mode is in a location where $\tilde{\beta}_2$ is negative but we cannot decide whether $\tilde{\beta}_3$ is positive at the mode.

Let us return to the initial model (4.2). The results of our test give evidence that a mode exists close to

$$(\beta_1, \beta_2, \beta_3) = (0.5, -0.07, 0.5)$$

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

Appendix A: Nonparametric estimators for f_Θ and $f_{S,\Theta}$

In this section we discuss the estimation of the densities f_Θ and $f_{S,\Theta}$ and related quantities. The results are derived by standard arguments. They are necessary

for our analysis of the test. We use kernel density estimators based on the second half of the observations $(S_i, \boldsymbol{\Theta}_i)$, $i = n+1, \ldots, 2n$. The density of the random vector $\mathbf{X}_1$ is denoted by $f_{\mathbf{X}}$. In the random coefficients model without intercept $f_{\mathbf{X}}$ is a d-variate density 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 without intercept, we use the kernel density estimator proposed by Bai et al. (1988)

$$\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{C(h_*)}{nh_*^{d-1}} \sum_{i=n+1}^{2n} K\Big(\frac{1 - \langle \boldsymbol{\Theta}_i, \boldsymbol{\theta} \rangle}{h_*^2}\Big), \quad h_* > 0, \tag{4.5}$$

with normalization constant

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

As shown in Bai et al. (1988), the integral does not depend on $\boldsymbol{\theta}$ and $C(h_*)$ converges to some positive constant as $h_* \to 0$. For the joint density of $(S, \boldsymbol{\Theta})$, we propose the kernel density estimator

$$\widehat{f}_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta}) = \frac{1}{nh_+^d}C(h_+) \sum_{i=n+1}^{2n} K\Big(\frac{1 - \langle \boldsymbol{\Theta}_i, \boldsymbol{\theta} \rangle}{h_+^2}\Big) K\Big(\frac{S_i - s}{h_+}\Big), \quad h_+ > 0. \tag{4.6}$$

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 on the boundary of the hemisphere $\mathbb{S}_+^{d-1}$ (see also (18)). Smoothness is, however, necessary to control the bias. 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 $\mathbf{X}_i / \|\mathbf{X}_i\|$, $i = n+1, \ldots, 2n$ on the hemisphere $\mathbb{S}_+^{d-1}$ and on $\mathbb{R} \times \mathbb{S}_+^{d-1}$, respectively. The estimators for $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$ are then the same as (4.5) and (4.6) except that now $\boldsymbol{\theta} \in \mathbb{S}_+^{d-1}$ and the normalization constant $C(h_*)$ is replaced by a function $\boldsymbol{\theta} \mapsto C(h_*, \boldsymbol{\theta})$, defined by

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

In a second step, we symmetrize the estimators and divide by two to get estimators of the densities $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$ on the whole domain.

In the next lemma we establish convergence of 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. *Suppose Assumption 3 is satisfied for some $\gamma > 0$ and set $\gamma = 1$ in the case of the random coefficients model with intercept. In both models, the estimator $\widehat{f}_{\boldsymbol{\Theta}}$ with bandwidth $h_* = O(\log(n)^{-3/\gamma})$ and $h_* \geq \log(n)^{7/(d-1)}n^{-1/(d-1)}$ satisfies*

(i) $\quad \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \big| \mathbb{E}\big[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big] - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\big| = O(h_*^{\gamma});$

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

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

The proof is deferred to the end of this section. Let us now discuss properties of the density $f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})$. By (5), $f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta}) = f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})Rf_{\boldsymbol{\beta}}(s, \boldsymbol{\theta})$. Under Assumption 2, $f_{\boldsymbol{\beta}}$ is compactly supported and consequently, $f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})$ and $s \mapsto \log(|s|)^2 f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})$ for $|s| \geq 1$ are uniformly bounded. Moreover, $f_{S,\boldsymbol{\Theta}}$ is Hölder continuous with Hölder constant γ. This is a straightforward consequence of the Hölder γ-continuity of $f_{\boldsymbol{\Theta}}$ shown in the proof of Lemma A.1 and the identity

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

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{d-1}$ denote an orthonormal basis of the orthogonal complement of span$\{\boldsymbol{\theta}\}$, together with the compact support and the Lipschitz-continuity of $f_{\boldsymbol{\beta}}$ (following from Assumption 2). Moreover, the properties of $f_{S,\boldsymbol{\Theta}}$ discussed above also hold in quantum homodyne tomography under Assumption 2'. We point out that the marginal densities of the Wigner function, which are given by the Radon transform, are nonnegative. This will be used later to bound the standard deviation $\sigma_{\mathbf{t},h,\mathbf{v}}$ away from zero.

If in Lemma A.2 $f_{\mathbf{X}}$ follows a multivariate Cauchy distribution, then $\gamma := 1$. Otherwise, γ comes from Assumption 3 *(iii)*.

Lemma A.2. *Let Assumptions 2 resp. 2' and 3 hold and consider the estimator $\widehat{f}_{S,\boldsymbol{\Theta}}$ in (4.6) with bandwidth choice $h_+ = O(\log(n)^{-3/\gamma})$ and $h_+ \geq \log(n)^{3/d}n^{-1/(2d)}$. Then*

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

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 $1/f_{\boldsymbol{\Theta}}$ and $\sqrt{f_{S,\boldsymbol{\Theta}}}$ need to be estimated. The functions $(\cdot)^{-1}$ and $\sqrt{\cdot}$ are not smooth in zero and we therefore introduce the cut-off estimators

$$\widetilde{f}_{\boldsymbol{\Theta}} := \widehat{f}_{\boldsymbol{\Theta}} \vee \log(n)^{-1} \quad \text{and} \quad \widetilde{f}_{S,\boldsymbol{\Theta}} := \widehat{f}_{S,\boldsymbol{\Theta}} \vee \log(n)^{-2}. \tag{4.7}$$

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

Proof of Lemma A.1. We only give a detailed proof for the case without intercept, i.e. *(i)-(iii)* in Assumption 3 hold. In the case with intercept, we use $C(h_*) \leq C(h_*, \boldsymbol{\theta}) \leq 2C(h_*)$ and the fact that $C(h_*, \boldsymbol{\theta})$ is Lipschitz-continuous with respect to $\boldsymbol{\theta}$ to arrive at the same conclusion.

To prove *(i)* observe that

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

Here, we used the identity $1 - \langle \boldsymbol{\theta}', \boldsymbol{\theta} \rangle = \|\boldsymbol{\theta}' - \boldsymbol{\theta}\|^2 / 2$ and the compact support of K. 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 $\left| \mathbb{E}\left[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \right] - f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \right| \lesssim h_*^\gamma$ 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

$$
\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \left| \widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}\left[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \right] \right| > ca_n \right) \leq \sum_{j=1}^M \mathbb{P}\left(\left| \widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j) - \mathbb{E}\left[\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_j) \right] \right| > \frac{c}{2} a_n \right).
$$
$$(4.8)$$

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

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

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

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

using the boundedness of K and $f_{\boldsymbol{\Theta}}$ as well as the definition of $C(h_*)$. Hence, an application of Bernstein's inequality yields with (4.8),

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

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. $\qquad\square$

To prove Lemma A.2, we make use of a slightly modified version of Proposition 2.2 in Giné and Guillou (2001) which we state below as Proposition A.1. If $\mathcal{F}$ is a uniformly bounded class of measurable functions on a measurable space $(S, \mathcal{S})$ with a measurable and bounded envelope F, then $\mathcal{F}$ is said to be a measurable uniformly bounded VC class of functions if there are constants $A, v > 0$ such that

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

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 $(S, \mathcal{S})$.

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 $\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathrm{Var}_P f$, $U \geq \sup_{f \in \mathcal{F}} \|f\|_\infty$ and $0 < \sigma \leq U$. Then there exist universal constants $C, K', L > 0$ such that the exponential inequality*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n \left(f(\xi_i) - \mathbb{E}[f(\xi_i)]\right) \right| > t \right)$$

$$\leq K' \exp\left(-\frac{1}{K'} \frac{t}{U} \log\left(1 + \frac{tU}{\left(\sqrt{n}\sigma + L\sqrt{v}U\sqrt{\log(AU\sigma^{-1})}\right)^2} \right) \right) \quad (4.9)$$

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

In contrast to Proposition 2.2 in Giné and Guillou (2001), Proposition A.1 contains the explicit dependence of the right hand side of (4.9) on the constants A and v.

Proof of Lemma A.2. Similarly as in the proof of Lemma A.1, it is enough to consider the random coefficients model without intercept only and to work under Assumption 3, *(i)-(iii)*. 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

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

we apply Proposition A.1 to the function class

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

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 $\{\boldsymbol{\theta}_j : j = 1, \ldots, M_1\}$ 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 $I_{i_j} \times \mathbb{S}^{d-1}$, $j = 1, \ldots, \lceil (c\varepsilon)^{-2} \rceil$, such that $Q(I_{i_j} \times \mathbb{S}^{d-1}) \geq (c\varepsilon)^2$. Let $\{s_j : j = 1, \ldots, M_2\}$ be an equidistant grid of

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

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

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

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

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

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

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

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

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

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

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

$$= \mathbb{P}\left(\sup_{f \in \mathcal{F}_n} \left| \sum_{i=n+1}^{2n} \left(f(S_i, \boldsymbol{\Theta}_i) - \mathbb{E}[f(S_i, \boldsymbol{\Theta}_i)] \right) \right| > \frac{n h_+^d}{C(h_+) \log \log(n) \log(n)^2} \right)$$

$$\lesssim \exp\left(-\frac{1}{4K'\sigma^2 C(h_+)^2} \frac{n h_+^{2d}}{\log \log(n)^2 \log(n)^4} \right)$$

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 $\widehat{f}_{S,\Theta}$. $\qquad\square$

Let us turn to the standard deviation

$$\sigma_{\mathbf{t},h,\mathbf{v}} = \Big(\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \frac{f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{f_\Theta(\boldsymbol{\theta})^2} ds d\boldsymbol{\theta} \Big)^{1/2} \quad (4.10)$$

and its estimator defined in (20). The following lemma shows that it is uniformly bounded from above and below. The proof is deferred to Appendix B.

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

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

The proof is given in Appendix B. The consistency of the estimates $\widehat{f}_\Theta$ and $\widehat{f}_{S,\Theta}$ 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 Assumptions 1- 3,*

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

Proof. By Lemma A.3,

$$\big| \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t},h,\mathbf{v}} \big| \leq \frac{\big| \widehat{\sigma}^2_{\mathbf{t},h,\mathbf{v}} - \sigma^2_{\mathbf{t},h,\mathbf{v}} \big|}{\sigma_{\mathbf{t},h,\mathbf{v}}}$$

$$\lesssim \int_{\mathbb{R} \times \mathbb{S}^{d-1}} \langle \boldsymbol{\theta}, \mathbf{v} \rangle^2 \big((\mathcal{H}_d \widetilde{\phi}^{(d-1)})(s) \big)^2 \Big| \frac{f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{f_\Theta(\boldsymbol{\theta})^2} - \frac{\widetilde{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta})}{\widetilde{f}_\Theta(\boldsymbol{\theta})^2} \Big| d(s, \boldsymbol{\theta}).$$

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

$$f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) \Big| \frac{1}{f_\Theta(\boldsymbol{\theta})^2} - \frac{1}{\widetilde{f}_\Theta(\boldsymbol{\theta})^2} \Big|$$

$$+ \frac{1}{\widetilde{f}_\Theta(\boldsymbol{\theta})^2} \big| f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) - \widetilde{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) \big|$$

$$= O\big((\log(n)^{-1}) \big) \text{ a.s..}$$

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

$$\big| f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) - \widetilde{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) \big|$$

$$\leq \big| f_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) - \widehat{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) \big|$$

$$+ \big| \widehat{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) - \widetilde{f}_{S,\Theta}(\langle \mathbf{t}, \boldsymbol{\theta} \rangle + hs, \boldsymbol{\theta}) \big|$$

$$= O\big(\log(n)^{-2} \big) \quad \text{almost surely}$$

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)*. $\qquad\square$

Lemma A.5. *Under Assumptions 2 resp. 2' and 3 we have*

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

Proof. This is a direct consequence of

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

as shown in the proof of Lemma A.4 and

$$
\begin{aligned}
\left|\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}-\sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}\right| &= \frac{\left|\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})-f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\right|}{\sqrt{\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}+\sqrt{f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})}} \\
&\leq \frac{\left|\widetilde{f}_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})-f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta})\right|}{\log(n)^{-1}}. \qquad\square
\end{aligned}
$$

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,*

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

Proof. Uniformly over $\boldsymbol{\theta}\in\mathbb{S}^{d-1}$,

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

Following the line of arguments in the proof of Lemma A.1, it is easy to see that $\mathbb{P}\big(\exists\boldsymbol{\theta}':\widehat{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}')\leq f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}')/2\big)$ decays at a rate which is faster than polynomial. In particular, $\widetilde{f}_{\boldsymbol{\Theta}}=\widehat{f}_{\boldsymbol{\Theta}}$ except on a set with probability decaying faster than any polynomial, which concludes the proof using Lemma A.1 *(i)*. $\qquad\square$

Appendix B: Proofs related to properties of the Radon transform

Proof of (16)*:* Recall from the discussion subsequent to (4) that in the random coefficients model with intercept

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

where ζ_1 is a Rademacher variable. Recall from Section 3.2 that in the random coefficients model (1), the density $f_{\boldsymbol{\Theta}}$ can be expressed in terms of the density $f_{\mathbf{X}}$ via $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \int_0^\infty r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta}) dr$, which follows since

$$\mathbb{P}\left(\boldsymbol{\Theta} \in S_{\boldsymbol{\Theta}}\right) = \int_{\{\mathbf{x} \,|\, \mathbf{x}/\|\mathbf{x}\| \in S_{\boldsymbol{\Theta}}\}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{S_{\boldsymbol{\Theta}}} \int_0^\infty r^{d-1} f_{\mathbf{X}}(r\boldsymbol{\theta}) dr d\boldsymbol{\theta},$$

for any measurable set $S_{\boldsymbol{\Theta}} \subset \boldsymbol{\theta}$. Let $S_{\boldsymbol{\Theta}} \subset \mathbb{S}^{d-1}$ be measurable. In the random coefficients model with intercept (2), we find for $\zeta_1 = 1$ corresponding to $\theta_1 > 0$

$$
\begin{aligned}
\mathbb{P}\left(\boldsymbol{\Theta} \in S_{\boldsymbol{\Theta}}\right) &= \int_{\{\mathbf{x} \,|\, \mathbf{x}/\|\mathbf{x}\| \in S_{\boldsymbol{\Theta}}\}} \delta(x_1 - 1) f_{\mathbf{X}}(x_2, \ldots, x_d) d\mathbf{x} \\
&= \int_{S_{\boldsymbol{\Theta}}} \int_0^\infty \delta(r\theta_1 - 1) f_{\mathbf{X}}(r\theta_2, \ldots, r\theta_d) r^{d-1} dr d\boldsymbol{\theta} \\
&= \int_{S_{\boldsymbol{\Theta}}} \int_0^\infty \delta(r - 1) f_{\mathbf{X}}\left(\tfrac{r}{\theta_1}\theta_2, \ldots, \tfrac{r}{\theta_1}\theta_d\right) \frac{r^{d-1}}{\theta_1^d} \mathbb{1}\{\theta_1 > 0\} dr d\boldsymbol{\theta} \\
&= \int_{S_{\boldsymbol{\Theta}}} f_{\mathbf{X}}\left(\tfrac{\theta_2}{\theta_1}, \ldots, \tfrac{\theta_d}{\theta_1}\right) \frac{1}{\theta_1^d} \mathbb{1}\{\theta_1 > 0\} d\boldsymbol{\theta},
\end{aligned}
$$

where $f_{\mathbf{X}}$ denotes the density of $(X_{1,2}, \ldots, X_{1,d})$ and with $u \mapsto \delta(u)$ the Dirac delta function. To see why this makes sense, notice that we either have $\mathbb{R}^+ \subset \pi_{\{1\}}(\{rS_{\boldsymbol{\Theta}} \,|\, r > 0\})$ or $\mathbb{R}^+ \cap \pi_{\{1\}}(\{rS_{\boldsymbol{\Theta}} \,|\, r > 0\}) = \emptyset$, where for $j \in \{1, \ldots, d\}$ the function $\pi_{\{j\}} : \mathbb{R}^d \to \mathbb{R}$ denotes the j-th projection map. This means in particular that either the cone $\{rS_{\boldsymbol{\Theta}} \,|\, r > 0\}$ contains a neighborhood of 1 in the first coordinate or there exists a neighborhood of 1, which is disjoint to the projection of the cone. Therefore, given $\zeta_1 = 1$, we find $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = f_{\mathbf{X}}\left(\tfrac{\theta_2}{\theta_1}, \ldots, \tfrac{\theta_d}{\theta_1}\right) \frac{1}{\theta_1^d} \mathbb{1}\{\theta_1 > 0\}$. Analogously, we find $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = f_{\mathbf{X}}\left(\tfrac{\theta_2}{\theta_1}, \ldots, \tfrac{\theta_d}{\theta_1}\right) \frac{1}{|\theta_1|^d} \mathbb{1}\{\theta_1 < 0\}$ for $\zeta_1 = -1$, which yields

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{1}{2|\theta_1|^d} f_{\mathbf{X}}\left(\frac{\theta_2}{\theta_1}, \ldots, \frac{\theta_d}{\theta_1}\right) \mathbb{1}\{\theta_1 \neq 0\},$$

by conditioning on the events $\{\zeta_1 = 1\}$ and $\{\zeta_1 = -1\}$. By Assumption 3 we obtain that $\lim_{\theta_1 \to 0} \frac{1}{2|\theta_1|^d} f_{\mathbf{X}}\left(\frac{\theta_2}{\theta_1}, \ldots, \frac{\theta_d}{\theta_1}\right)$ exists and therefore

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

for the continuous extension. $\qquad\square$

Proof of Lemma 3.1. By assumption, $\phi_{\mathbf{t},h}$ is radially symmetric and satisfies (9). 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

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

Set $\widetilde{s} = h^{-1}(s - \langle\mathbf{t},\boldsymbol{\theta}\rangle)$. For $d > 2$, using the definition of $\widetilde{\phi}$ in (13),

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

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

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

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 $h > 0$

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

by substitution. We show below that $\mathcal{H}_d f$ exists 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 $k = 0,\ldots,d+1$. The case $k = 0$ is obvious. We use the chain rule for higher order derivatives given by Faà di Bruno's formula

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

(4.11)

where $\mathcal{M}_k$ is the set of all k-tuples of non-negative integers which satisfy $k = \sum_{j=1}^{k}jm_j$. 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

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

exists for all $k = 1,\ldots,d+1$. Applying (4.11) with $f_1 = \phi$ and $f_2 = \sqrt{\cdot^2 + r^2}$ gives

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

for suitable constants $C_{m_1,\ldots,m_{k+1}}$ 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

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

(4.12)

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 (4.12) 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 (4.12) by

$$|z|^{\sum_{j=1}^{k+1} \ell_j m_j + M - 2 \sum_{j=1}^{k+1} (\ell_j + k_j) m_j + \max\{3 - M, 0\}} \int_0^{|z|} r^{d-2} dr$$

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

$$\lesssim |z|^{\sum_{j=1}^{k+1} \ell_j m_j + M - 2 \sum_{j=1}^{k+1} (\ell_j + k_j) m_j + d - 1 + \max\{3-M,0\}} + 1.$$

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,

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

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

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

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 prove *(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. $\qquad \square$

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 $f_{S,\boldsymbol{\Theta}}$. The uniform lower bound of $f_{\boldsymbol{\Theta}}$ follows from Assumption 3. The integrability of $\left(\mathcal{H}_d(\widetilde{\phi}^{(d-1)})(s)\right)^2$ is 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 $\mathbf{b} \in [\mathbf{a}_1,\mathbf{a}_2]$ 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 $\mathbf{t} \in [\mathbf{a}_1,\mathbf{a}_2]$, say, $f_{\boldsymbol{\beta}}(\mathbf{b}) > c_{\boldsymbol{\beta}}/2$ for all $\mathbf{b} \in \bigcup_{\mathbf{t}\in[\mathbf{a}_1,\mathbf{a}_2]} B_\delta(\mathbf{t})$. Define for $s^2 < \delta^2/(dh^2)$

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

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

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

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 $\mathbf{a}_1 + h \leq \mathbf{t} \leq \mathbf{a}_2 - h$, $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ and $|s| < \delta/(\sqrt{d}h)$. Therefore,

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

$$(4.13)$$

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

In quantum homodyne tomography, Assumption 2' *(iii)* 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, (4.13) holds in this case as well. Furthermore, since $\mathcal{H}_d(\widetilde{\phi}^{(d-1)}) \in L^2(\mathbb{R})$, we obtain

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

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

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

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 $\mathcal{H}_d\mathcal{H}_d f = -f$. $\square$

Appendix 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,*

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

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

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

and

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

$$(4.14)$$

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 $s \mapsto \log(|s|)^2 f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})$ 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 $\int_{-\infty}^{-2} |a_{\mathbf{t},h,\mathbf{v}}(s, \boldsymbol{\theta}) f_{S,\boldsymbol{\Theta}}(s, \boldsymbol{\theta})| ds$. Applying Lemma A.6 with bandwidth $h_* = \log(n)^{7/(d-1)} n^{-1/(d-1)}$ gives

$$\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} V_{\mathbf{t},h,\mathbf{v}} \lesssim \log(h_{\max})^2 \sqrt{h_{\max}} \log(n)^{7\gamma/(d-1)} n^{-\gamma/(d-1)}. \tag{4.15}$$

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

$$A_n := \left\{ (S_i, \boldsymbol{\Theta}_i)_{i=n+1,\dots,2n} : \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \left| \widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \mathbb{E}[\widetilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})] \right| \leq c \sqrt{\frac{\log n}{nh_*^{d-1}}} \right\},$$

then by Lemma A.1

$$\rho_n \leq \mathbb{E}\left[\mathbb{P}\left(\sup_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} U_{\mathbf{t},h,\mathbf{v}} \geq \delta_n \mid (S_i, \boldsymbol{\Theta}_i)_{i=n+1,\dots,2n} \right) \mathbb{1}(A_n) \right] + \mathbb{P}(A_n^c)$$

$$\leq \sum_{(\mathbf{t},h,\mathbf{v}) \in \mathcal{T}_n} \mathbb{E}\left[\mathbb{P}\left(U_{\mathbf{t},h,\mathbf{v}} \geq \delta_n \mid (S_i, \boldsymbol{\Theta}_i)_{i=n+1,\dots,2n} \right) \mathbb{1}(A_n) \right] + o(1)$$

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

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

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,

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

The inequality $\widetilde{f}_{\boldsymbol{\Theta}} \geq \log(n)^{-1}$, the uniform lower bound of $f_{\boldsymbol{\Theta}}$, $\widetilde{f}_{\boldsymbol{\Theta}} = \widehat{f}_{\boldsymbol{\Theta}}$ 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

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

for some constants $C, C_1 > 0$. By a change of variables in the integral for the variable s, the uniform boundedness of $f_{S,\boldsymbol{\Theta}}$, 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

$$\operatorname{Var}\left(U_{\mathbf{t},h,\mathbf{v}} \mid (S_i, \boldsymbol{\Theta}_i)_{i=n+1,\dots,2n}\right) \le C \frac{1}{n^2} \sup_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}} \left(\frac{1}{\tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} - \frac{1}{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\right)^2$$

$$\le C_2 n^{-2} \log(n)^{-4}$$

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

$$\rho_n \lesssim |\mathcal{T}_n| \exp\left(-\frac{\delta_n^2/2}{C_2 n^{-1} \log(n)^{-4} + \frac{C_1 \delta_n}{3n\sqrt{h_{\min}}\log(n)^2}}\right) + o(1)$$

$$= |\mathcal{T}_n| \exp\left(-\frac{(n\log(n))^{-1}/2}{C_2 n^{-1} \log(n)^{-4} + \frac{C_1}{3n^{3/2}\sqrt{h_{\min}}\log(n)^{5/2}}}\right) + o(1) = o(1),$$

as $h_{\min} \ge n^{-1}$. 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. $\quad\square$

C.2. Approximation of the limit statistic

Define the process

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

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 (Chernozhukov et al. (2017), Proposition 2.1). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ 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 $i = 1, \dots, n$, $j = 1, \dots, 2p$. Moreover, let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be independent random vectors in $\mathbb{R}^{2p}$ with $\mathbf{Y}_i \sim N(\mathbf{0}, \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top])$, $i = 1, \dots, n$. Let $b, q > 0$ be some constants and let $B_n \ge 1$ 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 $A = \{\mathbf{x} \in \mathbb{R}^{2p} : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$ for $-\infty \le \mathbf{a} \le \mathbf{b} \le \infty$. Assume that*

(i) $\quad n^{-1} \sum_{i=1}^n \mathbb{E}[\mathrm{X}_{i,j}^2] \ge b$ *for all* $1 \le j \le 2p$;

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

(iii) $\quad \mathbb{E}\left[\left(\max_{1 \le j \le 2p} |\mathrm{X}_{i,j}|/B_n\right)^q\right] \le 2$ *for all* $i = 1, \dots, n$

and define

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

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

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

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

$$\mathcal{T}'_n := \mathcal{T}_n \cup \big\{(\mathbf{t}, h, -\mathbf{v}) : (\mathbf{t}, h, \mathbf{v}) \in \mathcal{T}_n\big\} =: \big\{(\mathbf{t}_j, h_j, \mathbf{v}_j) : j = 1, \ldots, 2p\big\}. \quad (4.16)$$

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

$$X_{i,j} := \Upsilon_j(S_i, \boldsymbol{\Theta}_i) - \mathbb{E}[\Upsilon_j(S_i, \boldsymbol{\Theta}_i)], \quad \text{and} \quad \Upsilon_j(s, \boldsymbol{\theta}) := \frac{a_{\mathbf{t}_j, h_j, \mathbf{v}_j}(s, \boldsymbol{\theta})}{\sigma_{\mathbf{t}_j, h_j, \mathbf{v}_j} \sqrt{h_j} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}$$

for $j = 1, \ldots, 2p$. 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 $\mathbf{Z} \sim N(\mathbf{0}, \mathbb{E}[\mathbf{X}_1 \mathbf{X}_1^\top])$,

$$\sup_{A \in \mathcal{A}'_{2p}} \left| \mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i \in A\Big) - \mathbb{P}\Big(\mathbf{Z} \in A\Big) \right| \to 0. \quad (4.17)$$

Observe that by (4.14) and the uniform lower bound of $\sigma_{\mathbf{t}_j, h_j, \mathbf{v}_j}$ established in Lemma A.3,

$$\big|\mathbb{E}[\Upsilon_j(S_1, \boldsymbol{\Theta}_1)]\big| \lesssim \log(h_j)^2 \sqrt{h_j}. \quad (4.18)$$

Because of this bound, the expectation $\mathbb{E}[\Upsilon_j(S_1, \boldsymbol{\Theta}_1)]$ 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 (4.10). 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 $f_{\boldsymbol{\Theta}}$, and the boundedness of $f_{S,\boldsymbol{\Theta}}$, we find for $k = 1, 2$, $\max_{j=1,\ldots,2p} \mathbb{E}[|\Upsilon_j(S_1, \boldsymbol{\Theta}_1)|^{2+k}] \lesssim h_{\min}^{-k/2}$. This implies condition *(ii)* of Theorem C.2 with $B_n \asymp h_{\min}^{-1/2}$.

Lemma 3.1 *(i)* implies $\max_{j=1,\ldots,2p} |X_{i,j}| \lesssim h_{\min}^{-1/2}$ 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 $\mathbf{Z} \sim N(\mathbf{0}, \mathbb{E}[\mathbf{X}_1 \mathbf{X}_1^\top])$

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

The right hand side goes to 0 by Assumption 4, when q is chosen large enough.

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

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

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

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

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

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

Hence,

$$\left| \widetilde{W}(\Upsilon_j) - W(\Upsilon_j\sqrt{f_{S,\boldsymbol{\Theta}}}) \right| = \left| \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \Upsilon_j(s,\boldsymbol{\theta}) f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) ds d\boldsymbol{\theta} \, W(\sqrt{f_{S,\boldsymbol{\Theta}}}) \right|$$

and by (4.18), $\left| \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \Upsilon_j(s,\boldsymbol{\theta}) f_{S,\boldsymbol{\Theta}}(s,\boldsymbol{\theta}) ds d\boldsymbol{\theta} \right| \lesssim \log(h_j)^2 \sqrt{h_j}$. Furthermore, $W(\sqrt{f_{S,\boldsymbol{\Theta}}}) \sim N(0,1)$ which implies

$$\mathbb{E}\left[\max_{j=1,\ldots,2p} \left| \widetilde{W}(\Upsilon_j) - W(\Upsilon_j\sqrt{f_{S,\boldsymbol{\Theta}}}) \right| \right] \lesssim \log(h_{\max})^2 \sqrt{h_{\max}}.$$

An application of Markov's inequality finally proves

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

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

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

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 apply Theorem 6.1 in Dümbgen and Spokoiny (2001) to the non-normalized process

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

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

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

In the next step, we prove

$$\rho\big((\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\big) \lesssim \big(\|\mathbf{v}-\mathbf{v}'\|^2 + \|\mathbf{t}-\mathbf{t}'\| + |h-h'|\big)^{1/2}. \tag{4.19}$$

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

$$\begin{aligned} &\big|Y_{\mathbf{t},h,\mathbf{v}} - Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2 \\ &\lesssim \big|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2 + \big|Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2\big|\sigma_{\mathbf{t}',h',\mathbf{v}'} - \sigma_{\mathbf{t},h,\mathbf{v}}\big|^2. \end{aligned} \tag{4.20}$$

In order to bound the expectation of the first term on the right hand side of (4.20), we use the boundedness properties of $f_{\boldsymbol{\Theta}}$ and $f_{S,\boldsymbol{\Theta}}$

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

We show that the three terms can be bounded by the squared r.h.s. in (4.19). 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

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

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 (4.20) 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 $|\sqrt{x} - \sqrt{y}|^2 \le |x - y|$ for all $x, y \ge 0$ gives

$$\mathbb{E}\big|Y_{\mathbf{t}',h',\mathbf{v}'}\big|^2|\sigma_{\mathbf{t},h,\mathbf{v}} - \sigma_{\mathbf{t}',h',\mathbf{v}'}|^2$$

$$=h'\Big|\frac{(\mathbb{E}|\sigma_{\mathbf{t},h,\mathbf{v}}Y_{\mathbf{t},h,\mathbf{v}}|^2)^{1/2}}{\sqrt{h}} - \frac{(\mathbb{E}|\sigma_{\mathbf{t}',h',\mathbf{v}'}Y_{\mathbf{t}',h',\mathbf{v}'}|^2)^{1/2}}{\sqrt{h'}}\Big|^2$$

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

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

For the second term in the last line, the bounds above apply which completes the proof for (4.19).

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

$$\sigma^2(\mathbf{t}, h, \mathbf{v}) - \sigma^2(\mathbf{t}', h', \mathbf{v}') \le \widetilde{\rho}^2\big((\mathbf{t}, h, \mathbf{v}), (\mathbf{t}', h', \mathbf{v}')\big)$$

for all $((\mathbf{t}, h, \mathbf{v}), (\mathbf{t}', h', \mathbf{v}')) \in \overline{\mathcal{T}} \times \overline{\mathcal{T}}$. 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 $\eta > 0$,

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

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 (2001) are satisfied. As in Eckle et al. (2017) one shows that condition (iii) of Theorem 6.1 in Dümbgen and Spokoiny (2001) 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 Dümbgen and Spokoiny (2001). □

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 (4.16) and let

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

Lemma A.5 together with 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

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

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

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

for all $(\mathbf{t},h,\mathbf{v}),(\mathbf{t}',h',\mathbf{v}')\in\mathcal{T}_n'$ almost surely. Hence, an application of Theorem 2.2.5 in Adler and Taylor (2007) gives

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

Moreover, since $\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n'}\beta_h(|X_{\mathbf{t},h,\mathbf{v}}|/\sigma_{\mathbf{t},h,\mathbf{v}} - \alpha_h)$ is almost surely bounded asymptotically as proven 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 $C > 0$,

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

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

C.5. Replacement of the standard deviation by an estimator

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

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

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

$$\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \sqrt{n} |\widehat{T}_{\mathbf{t},h,\mathbf{v}} - E[T_{\mathbf{t},h,\mathbf{v}}]| \Big| \frac{1}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \frac{1}{\sigma_{\mathbf{t},h,\mathbf{v}}} \Big|$$

$$\lesssim \sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \left(\sqrt{n} \frac{|\widehat{T}_{\mathbf{t},h,\mathbf{v}} - \mathbb{E}[T_{\mathbf{t},h,\mathbf{v}}]|}{\sigma_{\mathbf{t},h,\mathbf{v}}} - \alpha_h \right) \sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} |\sigma_{\mathbf{t},h,\mathbf{v}} - \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}|$$

$$+ \frac{\log(n)}{\log\log(n)} \sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} |\sigma_{\mathbf{t},h,\mathbf{v}} - \widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}|$$

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. □

Appendix D: Proofs of Theorems 3.3 and 3.4

Proof of Theorem 3.3. We have

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

$$= 1 - \mathbb{P}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \Big(\sqrt{n} \frac{|\widehat{T}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \alpha_h \Big) \leq \kappa_n(\alpha) \Big)$$

$$= 1 - \mathbb{P}\Big(\sup_{(\mathbf{t},h,\mathbf{v})\in\mathcal{T}_n} \beta_h \Big(\frac{|\widehat{X}_{\mathbf{t},h,\mathbf{v}}|}{\widehat{\sigma}_{\mathbf{t},h,\mathbf{v}}} - \alpha_h \Big) \leq \kappa_n(\alpha) \Big) + o(1)$$

$$\leq \alpha + o(1)$$

for $n \to \infty$. Here we used (21) for the first equality and Theorem 3.2 for the second. □

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 Eckle et al. (2017). 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

$$\int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) \, d\mathbf{b} < -2c_d^{-1} h^{-d-1/2} \kappa_n^{\mathbf{t},h,\mathbf{v}}(\alpha_n) \tag{4.21}$$

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 (12) are rejected simultaneously on the set of scales $\mathcal{T}'_n$ with asymptotic probability one. Moreover, for a mode $\mathbf{b}_0$ in $(\mathbf{a}_1, \mathbf{a}_2)$ 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 $\mathbf{b} \in \text{supp}\phi_{\mathbf{t},h}$ by following the arguments in the proof of Theorem 10 in Eckle et al. (2018). Consequently, $\int_{\mathbb{R}^d} \phi_{\mathbf{t},h}(\mathbf{b}) \partial_{\mathbf{v}} f_{\boldsymbol{\beta}}(\mathbf{b}) \, d\mathbf{b} \lesssim -h$.

As seen in Appendices 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

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

In order to verify (4.21), we need to pick h such that $h^{d+3/2} \gtrsim (\log(n)/n)^{1/2}$. Thus, (4.21) holds for $h \geq C \log(n)^{\frac{1}{2d+3}} n^{-\frac{1}{2d+3}}$ with some sufficiently large constant $C > 0$. $\qquad\square$

Acknowledgments

References

Adler, R. J. and Taylor, J. E. (2007). *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York. MR2319516

Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734. MR1828540

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. MR0971170

Beran, R. (1993). Semiparametric random coefficient regression models. *Ann. Inst. Statist. Math.*, 45(4):639–654. MR1252945

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. MR1425969

Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Ann. Statist.*, 20(4):1970–1984. MR1193321

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

Berry, S. and Pakes, A. (2007). The pure characteristics demand model. *Internat. Econom. Rev.*, 48(4):1193–1225. MR2375623

Breunig, C. and Hoderlein, S. (2018). Specification testing in random coefficient models. *Quantitative Economics*, 9(3):1371–1417. MR3891570

Butucea, C., Guţă, 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. MR2336856

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45(4):2309–2352. MR3693963

Davison, M. E. (1983). The ill-conditioned nature of the limited angle tomography problem. *SIAM J. Appl. Math.*, 43(2):428–448. MR0700347

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

Dubé, 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. MR3013723

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152. MR1833961

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785. MR2435455

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

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.

Dunker, F., Hoderlein, S., Kaido, H., and Sherman, R. (2018). Nonparametric identification of the distribution of random coefficients in binary response static games of complete information. *Journal of Econometrics*, 206(1):83–102. MR3840784

Eckle, K., Bissantz, N., and Dette, H. (2017). Multiscale inference for multivariate deconvolution. *Electron. J. Stat.*, 11(2):4179–4219. MR3716498

Eckle, K., Bissantz, N., Dette, H., Proksch, K., and Einecke, S. (2018). Multiscale inference for a multivariate density with applications to x-ray astronomy. *Annals of the Institute of Statistical Mathematics*, 70(3):647–689. MR3785711

Feuerverger, A. and Vardi, Y. (2000). Positron emission tomography and random coefficients regression. *Ann. Inst. Statist. Math.*, 52(1):123–138. MR1771484

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.

Frikel, J. (2013). Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, 34(1):117–141. MR2981336

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

Gautier, E. and Kitamura, Y. (2013). Nonparametric estimation in random

coefficients binary choice models. *Econometrica*, 81(2):581–607. MR3043343

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. MR1876841

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

Gustafson, P. and Greenland, S. (2006). The performance of random coefficient regression in accounting for residual confounding. *Biometrics*, 62(3):760–768. MR2247204

Helgason, S. (2011). *Integral geometry and Radon transforms.* Springer, New York. MR2743116

Hoderlein, S., Holzmann, H., and Meister, A. (2017). The triangular model with random coefficients. *Journal of Econometrics*, 201(1):144-169.

Hoderlein, S., Klemelä, J., and Mammen, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, 26(3):804–837. MR2646480

Hohmann, D. and Holzmann, H. (2016). Weighted angle radon transform: Convergence rates and efficient estimation. *Statistica Sinica.*, 26(1):157–175. MR3468349

Holzmann, H. and Meister, A. (2019). Rate-optimal nonparametric estimation for random coefficient regression models. *ArXiv:1902.05261.*

Hsiao, C. (2014). *Analysis of Panel Data.* Cambridge University Press. Cambridge Books Online. MR3560738

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

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. MR1649224

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.

Masten, M. A. (2018). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies*, 85(2):1193–1250. MR3782604

Masten, M. A. and Torgovitsky, A. (2016). Identification of instrumental variable correlated random coefficients models. *The Review of Economics and Statistics*, 98(5):1001–1005.

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

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

Proksch, K., Werner, F., and Munk, A. (2018). Multiscale scanning in inverse problems. *Ann. Statist.*, 46(6B):3569–3602. MR3852662

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. MR3113812

Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–323. MR0269032

Wieczorek, B. (2010). On optimal estimation of the mode in nonparametric deconvolution problems. *Journal of Nonparametric Statistics*, 22(1):65–80. MR2598954

Wooldridge, J. (2013). *Introductory Econometrics: A Modern Approach.* Cengage Learning.