

The Empirical Process of Residuals from an Inverse Regression

T. Kutta^{1*}, N. Bissantz^{1**}, J. Chown^{1***}, and H. Dette^{1****}

¹*Ruhr-Universität Bochum, Germany*

Received February 4, 2019; in final form, June 8, 2019

Abstract—In this paper we investigate an indirect regression model characterized by the Radon transformation. This model is useful for recovery of medical images obtained by computed tomography scans. The indirect regression function is estimated using a series estimator motivated by a spectral cutoff technique. Further, we investigate the empirical process of residuals from this regression, and show that it satisfies a functional central limit theorem.

Keywords: indirect regression model, inverse problems, Radon transform, empirical process.

AMS 2010 Subject Classification: 62G08, 62G30, 15A29.

DOI: 10.3103/S1066530719020029

1. INTRODUCTION

Computed tomography (CT) is a noninvasive imaging technique, which is a key method for medical diagnoses. CT is based on measuring the intensity losses of X-rays sent through a body. From these measurements an attenuation profile can be recovered that provides an image of the body's (unobservable) interior. The X-rays are linear and so the scanner rotates to create a two-dimensional slice. Insight into three-dimensional structures is obtained by considering multiple slices. Our investigation is limited to the statistical analysis of data gathered from a single slice. For this purpose we introduce the inverse regression model

$$Y_{\mathbf{k}} = \mathcal{R}g(z_{\mathbf{k}}) + \varepsilon_{\mathbf{k}}, \quad \mathbf{k} \in \mathbf{K}, \quad (1.1)$$

where $(\varepsilon_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}}$ are independent identically distributed random variables with $\mathbb{E}[\varepsilon_{\mathbf{k}}] = 0$. Here \mathbf{K} is a given index set, with each index \mathbf{k} corresponding to an X-ray path and the design point $z_{\mathbf{k}}$ characterizing this path with associated response $Y_{\mathbf{k}}$. Consequently, $z_{\mathbf{k}}$ can be written using coordinates $0 \leq s \leq 1$ as the distance from the origin and $0 \leq \phi \leq 2\pi$ as the angle of inclination. The body's (true) attenuation profile along the slice is represented by g , a function supported on the unit disc. \mathcal{R} is a linear operator acting on g and denotes the normalized *Radon transform*, i.e., for $0 \leq s \leq 1$ and $0 \leq \phi \leq 2\pi$,

$$\mathcal{R}g(s, \phi) := \frac{1}{2}(1 - s^2)^{-\frac{1}{2}} \int_{-\sqrt{1-s^2}}^{\sqrt{1-s^2}} g(s \cos(\phi) - t \sin(\phi), s \sin(\phi) + t \cos(\phi)) dt. \quad (1.2)$$

Details on the underlying physics and applications of CT can be found in [9].

Image reconstruction in CT is a particular case of the broad class of linear inverse problems. An overview of the mathematical aspects of these problems and methods to solving them can be found in the monographs [39], [21], and [22]. Other examples of linear inverse problems are the heat equation and convolution transforms (see [35], [46], and [13], among others). Additional statistical inverse problems include errors-in-variables models and the Berkson error model (see, for example, [5], [10], [32, 33], [2], [26], [19], and [27]). The Radon transform is usually discussed in the contexts of positron emission

* E-mail: Tim.Kutta@ruhr-uni-bochum.de

** E-mail: nicolai.bissantz@web.de

*** E-mail: Justin.Chown@ruhr-uni-bochum.de

**** E-mail: holger.dette@rub.de

tomography (PET) and CT in medical imaging. In the case of PET, lines-of-sight are observed along which emissions have occurred. However, the positions of the emissions on these lines are unknown. Here the aim is to reconstruct the emission density (see [25], [31], and [12], among others). In contrast CT leads to inverse regression models such as considered in [11] and in [28, 29]. In the present paper we contribute to this discussion by considering the model (1.1), which involves a different kind of Radon transform. In particular we do not require the common white noise assumption. We derive uniform strong consistency of a nonparametric estimator \hat{g} of the function g and a functional central limit theorem for the empirical process of the residuals $\hat{\varepsilon}_{\mathbf{k}} = Y_{\mathbf{k}} - \mathcal{R}g(z_{\mathbf{k}})$ resulting from this fit. More precisely, we investigate the empirical process

$$\hat{\mathbb{F}}_n(t) = \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\}, \quad t \in \mathbb{R}, \tag{1.3}$$

where the nonnegative weights $w_{\mathbf{k}}$ sum to 1 (see Section 3). Statistical applications of results of this type include validation of model assumptions. In the context of inverse regression models, to the best of our knowledge only one result is available, namely, [4] studying an inverse regression model characterized by a convolution transformation.

In direct regression problems, residual-based empirical processes arising from non- and semiparametric regression estimators have been considered by numerous authors (see [1], [43], [37], [16], and [51], among others). In [20] tests are considered for a parametric form of the variance function in a heteroscedastic nonparametric regression by comparing the empirical distribution function of standardized residuals calculated under a null model to that of an alternative model. [44] works with a similar approach to propose tests for verifying convenient forms of the regression function. In [30] a popular distribution-free approach to addressing goodness-of-fit problems for the errors from a nonparametric regression is introduced, which consists in providing a transformation of the empirical distribution function of residuals that is useful for forming test statistics with convenient limit distributions. All of these approaches to validating model assumptions crucially rely on a technical asymptotic linearity property of the residual-based empirical distribution function. We show that the estimator (1.3) shares this property as well, and the results of this article can be used immediately in approaches to validating model assumptions in the inverse regression model (1.1) that are in the same spirit as the previously mentioned works.

We have organized the remaining parts of the paper as follows. Model (1.1) is further discussed and we introduce the estimator \hat{g} in Section 2. Our main results are given in Section 3. All of the proofs of our results and additional supporting technical details may be found in the appendices.

2. ESTIMATION IN THE INDIRECT REGRESSION MODEL

In this section we give more details regarding the Radon transform model (1.1) and introduce an estimator of the function g .

2.1. The Radon Transform

Following [25] let

$$\mathcal{B} := \{(r, \theta) : 0 \leq r \leq 1, 0 \leq \theta \leq 2\pi\} \tag{2.1}$$

denote the unit disc, which is the two-dimensional domain of the investigated attenuation profile g and is called *brain space* for historical reasons. It is equipped with the uniform distribution, given in polar coordinates by

$$d\mu(r, \theta) := \pi^{-1} r \, dr d\theta. \tag{2.2}$$

This means that no prior emphasis on any region of the scanned area is given. The *detector space* \mathcal{D} is defined as

$$\mathcal{D} := \{(s, \phi) : 0 \leq s \leq 1, 0 \leq \phi \leq 2\pi\} \tag{2.3}$$

with corresponding probability measure

$$d\lambda(s, \phi) := 2\pi^{-2} \sqrt{1 - s^2} \, ds d\phi. \tag{2.4}$$

The domain of the transformed image $\mathcal{R}g$ is \mathcal{D} , a parametrization of all lines (X-ray paths) crossing the unit disc. It is usually referred to as *detector space*. λ is a probability measure on \mathcal{D} adapted to the length of the line segments inside the disc. For analytic simplicity we allow the angles in \mathcal{B} and \mathcal{D} to be exactly 0 and 2π . This is possible since the smoothness of g and $\mathcal{R}g$ required below entails periodicity with respect to the angular coordinates.

The Radon transform in (1.2) defines a linear operator from $\mathcal{L}^2(\mathcal{B}, \mu)$ to $\mathcal{L}^2(\mathcal{D}, \lambda)$. Identifying corresponding equivalence classes it can be shown that \mathcal{R} is one-to-one, compact, and permits a singular-value decomposition (SVD). The SVD of \mathcal{R} is vital for our subsequent investigations. To state it efficiently we introduce some definitions borrowed from [25] and [8]. Let

$$\mathcal{N} := \{(l, m) : m \in \mathbb{N}_0, l = m, m-2, \dots, -m\}$$

be an index set and define for $(l, m) \in \mathcal{N}$ the function

$$\varphi_{(l,m)}(r, \theta) := \sqrt{m+1} R_m^{|l|}(r) \exp(il\theta), \quad (2.5)$$

where

$$R_m^{|l|}(r) := \sum_{j=0}^{\frac{1}{2}(m-|l|)} (-1)^j \frac{(m-j)!}{j! \left(\frac{m+|l|}{2} - j\right)! \left(\frac{m-|l|}{2} - j\right)!} r^{m-2j}$$

is the so-called *radial polynomial*. Finally, for $(l, m) \in \mathcal{N}$ we define

$$\psi_{(l,m)}(s, \phi) := U_m(s) \exp(il\phi), \quad (2.6)$$

where U_m denotes the m th Chebyshev polynomial of the second kind. For convenience of notation we also define $\varphi_{(l,m)} \equiv 0$ and $\psi_{(l,m)} \equiv 0$ for $(l, m) \notin \mathcal{N}$. Both collections of functions,

$$\{\varphi_{(l,m)} : (l, m) \in \mathcal{N}\} \quad \text{and} \quad \{\psi_{(l,m)} : (l, m) \in \mathcal{N}\}$$

form orthonormal bases of the spaces $\mathcal{L}^2(\mathcal{B}, \mu)$ and $\mathcal{L}^2(\mathcal{D}, \lambda)$ respectively. With these notation the SVD of \mathcal{R} for some $g \in \mathcal{L}^2(\mathcal{B}, \mu)$ is given by

$$\mathcal{R}g(s, \phi) = \sum_{m=0}^{\infty} \sum_{l=-m}^m \frac{1}{\sqrt{m+1}} \psi_{(l,m)}(s, \phi) \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)}. \quad (2.7)$$

In the literature the functions $\varphi_{(l,m)}(r, \theta)(m+1)^{-1/2}$ are commonly referred to as *Zernike polynomials*, which play an important role in the analysis of optical systems, for instance, in the modeling of refraction errors, c.f. [50] and more recently [34]. We refer to [17] for more details on the cited SVD of the normalized Radon transform. Due to injectivity of the operator \mathcal{R} we can immediately access its inverse \mathcal{R}^{-1} pointwise defined for some $\mathcal{R}g \in \mathcal{R}(\mathcal{L}^2(\mathcal{B}, \mu))$, as

$$g = \mathcal{R}^{-1}[\mathcal{R}g](r, \theta) = \sum_{m=0}^{\infty} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)}(r, \theta) \langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}. \quad (2.8)$$

The identities (2.7), (2.8) as well as \mathcal{L}^2 -expansions in the respective spaces apply a priori almost everywhere. However if g is sufficiently smooth they even hold uniformly. The smoothness of the function g in the inverse regression is stated in terms of rapidly decaying \mathcal{L}^2 -coefficients. In analogy to [13], Section 2.2, we define for $v, L \geq 0$ the class

$$\mathfrak{D}(v, L) := \left\{ g \in \mathcal{L}^2(\mathcal{B}, \mu) \mid g \text{ continuous, } \sum_{m=0}^{\infty} \sum_{l=-m}^m |\langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}| (m+1)^v \leq L \right\}. \quad (2.9)$$

We assume throughout this paper that the regression function g in model (1.1) is an element of $\mathfrak{D}(v, L)$ (for some $v \geq 1$ and finite L). Controlling smoothness and thereby the complexity of the class of regression functions by related conditions is common in inverse problems. This is owed to their natural correspondence to singular value decompositions of operators and their suitability to prove minimax optimal rates (see for example [35], [15], [6] or [7]).

Proposition 2.1. *Suppose that $g \in \mathfrak{D}(v, L)$ for some $v \geq 1$ and $L < \infty$, then the following four identities hold everywhere:*

$$g = \sum_{m=0}^{\infty} \sum_{l=-m}^m \varphi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)}, \tag{2.10}$$

$$\mathcal{R}g = \sum_{m=0}^{\infty} \sum_{l=-m}^m \psi_{(l,m)} \langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}, \tag{2.11}$$

$$\mathcal{R}g = \sum_{m=0}^{\infty} \sum_{l=-m}^m \frac{1}{\sqrt{m+1}} \psi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)}, \tag{2.12}$$

$$g = \mathcal{R}^{-1}[\mathcal{R}g] = \sum_{m=0}^{\infty} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)} \langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}. \tag{2.13}$$

Moreover the functions g and $\mathcal{R}g$ are $\lfloor (v - 1)/2 \rfloor$ times continuously differentiable.

The equality of g and its \mathcal{L}^2 -expansion is vital when proving uniform bounds on the distance between g and \hat{g} . In one-dimensional convolution type problems this is usually dealt with by the Dirichlet conditions that directly apply to classes of smooth functions (see [41], pp. 197–198). It should also be noted that the series condition on the function g in (2.9) implies regularity properties beyond mere smoothness. For instance, if $v \geq 2k + 1$, it also entails periodicity in the angular components of g and its continuous derivatives up to the order k . This property follows by periodicity of the basis functions in the angle and is an analogue to periodicity of convergent Fourier series on bounded intervals. Notice that it fits naturally to the scanning regime, since any function transformed from Cartesian into spherical coordinates will comply to periodicity with respect to the angle.

2.2. Design

As is common in computed tomography, we will assume a parallel scanning procedure corresponding to a grid of design points on the detector space. It is then mathematically simple to adopt our results to fan beam geometry, which underlies most modern scanners.

We thus define a grid on the detector space \mathcal{D} , where for given $p, q \in \mathbb{N}$ each of the constituting rectangles has side length $1/q$ in s -direction and $2\pi/p$ in ϕ -direction. More formally, we define an index set

$$\mathbf{K} := \{(k_1, k_2) : 0 \leq k_1 \leq q - 1, 0 \leq k_2 \leq p - 1\}$$

and decompose the detector space in rectangular boxes of the form

$$B_{(k_1, k_2)} := \left\{ (s, \phi) \in \mathcal{D} : \frac{k_1}{q} \leq s \leq \frac{k_1 + 1}{q}, \frac{2\pi k_2}{p} \leq \phi \leq \frac{2\pi(k_2 + 1)}{p} \right\},$$

where $\mathbf{k} = (k_1, k_2) \in \mathbf{K}$. The design points $\{z_{(k_1, k_2)} \mid (k_1, k_2) \in \mathbf{K}\}$ are then defined as follows. The second coordinate of $z_{(k_1, k_2)}$ is given by

$$z_{k_2}^2 := 2\pi \frac{k_2 + \frac{1}{2}}{p}$$

and the first coordinate $z_{k_1}^1$ is determined as the solution of the equation

$$\int_{k_1/q}^{(k_1+1)/q} (s - z_{k_1}^1) \sqrt{1 - s^2} ds = 0. \tag{2.14}$$

Throughout this paper we consider the inverse regression model (1.1) with these $n = pq$ design points. The non-uniform design in radial direction defined by (2.14) is motivated by a midpoint rule to numerically integrate over each box with respect to the measure λ in (2.4). For asymptotic considerations, we assume that $q \rightarrow \infty$ and that $p = p(q) \rightarrow \infty$ depends on q as follows:

Assumption 2.2. *There exist constants $C_1, C_2 > 0$ such that $C_1 q \leq p(q) \leq C_2 q$ for all $q \in \mathbb{N}$.*

Denoting the number of rows and columns in the grid of design points by q and p is common in the literature and numerical programming. Notice that our Assumption 2.2 leaves room for the resolution optimal choice $2\pi q \approx p$ (see [40], p. 74). Sometimes we will use the notation $n \rightarrow \infty$, actually meaning that according to Assumption 2.2 q and thereby p and n diverge. Note also that the index set \mathbf{K} depends on the sample size n in model (1.1). Thus formally we consider a triangular array of independent, identically distributed and centered random variables $(\varepsilon_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}}$, but we do not reflect this dependence on n in our notation.

2.3. The Spectral Cutoff Estimator

Motivated by the representation (2.13) we now define the cutoff estimator \hat{g} for the function g in model (1.1) by

$$\hat{g}(r, \theta) = \sum_{m=0}^{t_n} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)}(r, \theta) \hat{R}(l, m). \quad (2.15)$$

Here

$$\hat{R}(l, m) := \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) Y_{\mathbf{k}} \quad (2.16)$$

is an estimator of the inner product

$$R(l, m) := \langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)} \quad (2.17)$$

and $w_{\mathbf{k}} := \lambda(B_{\mathbf{k}})$ denotes the Lebesgue measure of the cell $B_{\mathbf{k}}$. Comparing (2.13) to our estimator in (2.15), we observe that the inner products have been replaced by the estimates (2.16). Furthermore the series has been truncated at $t_n \in \mathbb{N}$, which represents the application of a regularized inverse. It is common in the literature to refer to either t_n or t_n^{-1} as *bandwidth*, since it is used to balance between bias and variance like the bandwidth in kernel density estimation (see [13]).

The choice of a bandwidth is a non-trivial problem. An optimal bandwidth with respect to some criterion such as the integrated mean squared error will depend on the unknown regression function g . Several data driven selection criteria for the choice of t_n have been proposed and examined in the literature. We refer to the monograph [49], where multiple techniques are gathered. We also note that under the (stronger) assumption of a regression model with Gaussian white noise effective bandwidth selection and adaptive minimax optimal estimation have been discussed in [14] and [28], respectively. The latter authors investigate estimators with respect to various norms, providing insight into different optimal rates. More closely related to our case is the smooth bootstrap examined by [4] in the context of deconvolution.

Remark 2.1. It should be noticed that in practice a smooth dampening of high frequencies usually shows a better performance than the strict spectral cutoff. We can accommodate this by introducing a smooth version of the estimator \hat{g} in (2.15). For this purpose let $\Lambda: \mathbb{R} \rightarrow [0, 1]$ denote a function with compact support and define

$$\hat{g}_{\Lambda}(r, \theta) = \sum_{m=0}^{\infty} \Lambda(mt_n^{-1}) \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)}(r, \theta) \hat{R}(l, m) \quad (2.18)$$

as an alternative estimator of g . Note that the estimate \hat{g} in (2.15) is obtained for $\Lambda(x) = \mathbf{1}_{[0,1]}(x)$. All results presented in this paper remain valid for the estimator (2.18). However, for the sake of brevity and a transparent presentation the subsequent discussion is restricted to the spectral cutoff estimator in (2.15).

3. THE EMPIRICAL PROCESS OF RESIDUALS

In this section we investigate the asymptotic properties of the empirical process of residuals

$$\sqrt{n}(\hat{\mathbb{F}}_n(t) - \mathbb{F}(t)) := \sqrt{n} \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} (\mathbf{1}_{\{\hat{\varepsilon}_{\mathbf{k}} \leq t\}} - \mathbb{F}(t)), \quad t \in \mathbb{R},$$

where \mathbb{F} denotes the distribution function of the error and

$$\hat{\varepsilon}_{\mathbf{k}} := Y_{\mathbf{k}} - \mathcal{R}\hat{g}(z_{\mathbf{k}}), \quad \mathbf{k} \in \mathbf{K}, \tag{3.1}$$

the \mathbf{k} th residual obtained from the estimate \hat{g} . The weights $w_{\mathbf{k}}$ are defined in Section 2.3. We begin by showing a uniform convergence result for \hat{g} . For this purpose we derive uniform approximation rates for bias and variance and subsequently balance these two, to get optimal results. The proofs of the following results are complicated and therefore deferred to the Appendix.

Lemma 3.1. *Suppose that Assumption 2.2 holds and that $g \in \mathfrak{D}(v, L)$ for some $v \geq 5$ and $L < \infty$. Then the estimator \hat{g} in (2.15) satisfies*

$$\|\mathbb{E}\hat{g} - g\|_{\infty} = \mathcal{O}(t_n^{-(v-1)} + t_n^8 n^{-1}),$$

where $\|g\|_{\infty} := \sup_{z \in \mathcal{B}} |g(z)|$.

Lemma 3.2. *Suppose that Assumption 2.2 holds and that $\mathbb{E}|\varepsilon|^{\kappa} < \infty$ for some $\kappa > 3$. Additionally let the sequence $(t_n)_{n \in \mathbb{N}}$ satisfy $t_n n^{-1/2} = \mathcal{O}(1)$. Then the estimator \hat{g} in (2.15) satisfies*

$$\|\hat{g} - \mathbb{E}\hat{g}\|_{\infty} = \mathcal{O}(t_n^4 \log(n)^{1/2} n^{-1/2}) \quad a.s.$$

Remark 3.1. (a) The smoothness assumption $v \geq 5$, i.e. two times continuous differentiability of g can be weakened at the cost of a slower convergence rate of the bias. For instance $v \geq 3$ yields the uniform order $\mathcal{O}(t_n^6 n^{-1/2} + t^{-(v-1)})$. However in this case the bias term is of larger magnitude than the random part (see Lemma 3.2), leading to problems in the residual based analysis. We therefore assume $v \geq 5$ throughout our discussion.

(b) The moment assumption $\kappa > 3$ reflects that our measurement errors are modelled as a triangular array. For arrays of random variables, e.g., the strong law of large numbers requires markedly higher moment assumptions than for sequences ([45], p. 122). For an i.i.d. sequence of errors in model (1.1) one would just require $\kappa > 2$. However this would lead to a rather unnatural formulation of the model.

Balancing the two upper bounds in Lemmas 3.1 and 3.2 yields an optimal choice of the bandwidth. More precisely, for $v \geq 5$ the choice

$$t_n := \mathcal{O}((\log(n))^{-1} n^{\frac{1}{2(v+3)}}) \tag{3.2}$$

balances the upper bound from Lemma 3.2 with the leading term $\mathcal{O}(t_n^{-(v-1)})$ of the bias from Lemma 3.1. Combining these results yields the first part of the following theorem.

Theorem 3.3. *Let Assumption 2.2 hold, suppose that $g \in \mathfrak{D}(v, L)$ for some $v \geq 5$ and $L < \infty$ and that $\mathbb{E}|\varepsilon|^{\kappa} < \infty$ for some $\kappa > 3$. Additionally let t_n be chosen as in (3.2). Then*

$$\|g - \hat{g}\|_{\infty} = \mathcal{O}((\log(n)) n^{-1})^{\frac{v-1}{2(v+3)}} \tag{3.3}$$

and for all $\tau \leq v$

$$\sum_{m=0}^{\infty} \sum_{l=-m}^m m^{\tau} |\langle \mathcal{R}[g - \hat{g}], \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}| = \mathcal{O}(n^{\frac{v-\tau}{2(v+3)}}) \quad a.s. \tag{3.4}$$

Theorem 3.3 is an important tool for our subsequent work. A sufficiently fast convergence rate is crucial for the accurate recovery of the model errors. But beyond that the particular shape of \hat{g} as a truncated series estimator based on an orthonormal system is particularly suited for an asymptotic linearization of the process $\sqrt{n}(\hat{\mathbb{F}}_n - \mathbb{F})$ in Theorem 3.5 (see Proposition A.5 in Appendix A). We also mention that an interesting topic for future research is the investigation of an optimal rate in this context.

By the same techniques uniform bounds can be deduced for the derivatives of our estimators.

Corollary 3.4. *Let the assumptions of Theorem 3.3 hold, let t_n be of order $o(n^{1/4})$ and suppose $v \geq 2k + 1$ for some $k \in \mathbb{N}_0$. Additionally let $\alpha, \beta \in \mathbb{N}_0$ such that $\alpha + \beta = k$. Then*

$$\left\| \frac{\partial^\alpha}{\partial r^\alpha} \frac{\partial^\beta}{\partial \theta^\beta} g - \frac{\partial^\alpha}{\partial r^\alpha} \frac{\partial^\beta}{\partial \theta^\beta} \hat{g} \right\|_\infty = \mathcal{O}((\log(n))^{1/2} n^{-1/2} t_n^{2k+4} + t_n^{v-(2k+1)}) \quad a.s. \quad (3.5)$$

For the next step recall the definition of the estimated residuals $\hat{\varepsilon}_{\mathbf{k}}$ in (3.1), as well as the estimate for the residual distribution function $\hat{\mathbb{F}}_n$ in (1.3). In order to prove a uniform CLT for $\sqrt{n}(\hat{\mathbb{F}}_n - \mathbb{F})$ we disentangle the dependencies of the terms in $\hat{\mathbb{F}}_n$ in the next result.

Theorem 3.5. *Assume that $g \in \mathfrak{D}(v, L)$ for some $v > 5$ and $L < \infty$, $\mathbb{E}|\varepsilon|^\kappa < \infty$ for some $\kappa > 3$, that \mathbb{F} admits a Hölder continuous density f_ε with exponent $\zeta > 4/(v - 1)$ and that Assumption 2.2 holds. If the bandwidth t_n satisfies (3.2), then*

$$\sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} [\mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t\} - \varepsilon_{\mathbf{k}} f_\varepsilon(t)] \right| = o_P(n^{-1/2}). \quad (3.6)$$

Note that the required smoothness of the error density is reciprocal to the smoothness of the function g . This implies that a larger value v yields a larger class of error distributions that can be analyzed.

Corollary 3.6. *Under the assumptions of Theorem 3.5, the process*

$$\left\{ \sum_{\mathbf{k} \in \mathbf{K}} n^{1/2} w_{\mathbf{k}} \{ \mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\} - \mathbb{F}(t) \} \right\}_{t \in \mathbb{R}}$$

converges weakly to a mean zero Gaussian process G with covariance function

$$\begin{aligned} \Sigma(t, \tilde{t}) := & \frac{8\pi^2}{3} \left(\mathbb{F}(\min(t, \tilde{t})) - \mathbb{F}(t)\mathbb{F}(\tilde{t}) + f_\varepsilon(t)\mathbb{E}[\varepsilon \mathbf{1}\{\varepsilon \leq \tilde{t}\}] \right. \\ & \left. + f_\varepsilon(\tilde{t})\mathbb{E}[\varepsilon \mathbf{1}\{\varepsilon \leq t\}] + \sigma^2 f_\varepsilon(t)f_\varepsilon(\tilde{t}) \right), \quad t, \tilde{t} \in \mathbb{R}. \end{aligned}$$

APPENDIX A. PROOFS AND TECHNICAL DETAILS

Throughout our calculations C will denote a positive constant, which may differ from line to line. The dependence of C on other parameters will be highlighted in the specific context.

A.1. Proof of Lemma 3.1

We begin with an auxiliary result which provides an approximation rate for Lemma 3.1 in expectation of $\hat{R}(l, m)$ for $R(l, m)$ and is proven in Appendix B (see Section B.3).

Proposition A.1. *Suppose that $g \in \mathfrak{D}(v, L)$ for $v \geq 5$, $L < \infty$ and that Assumption 2.2 holds. Then for all $(l, m) \in \mathcal{N}$ it follows that*

$$|\mathbb{E}\hat{R}(l, m) - R(l, m)| \leq C m^5 n^{-1}, \quad (A.1)$$

where $C > 0$ is some constant depending on g and C_1 (the constant from Assumption 2.2).

We are now in a position to derive the decay rate of the bias postulated in Lemma 3.1. The decay rate naturally splits up into two parts. One part accounts for the average approximation error of Radon coefficients with index m smaller than t_n and the other one for the error due to frequency limitation of the estimator.

The singular value decomposition of the normalized Radon transform in (2.12) and the definition of our estimator in (2.15) yield

$$\begin{aligned} \|\mathbb{E}\hat{g} - g\|_\infty &= \left\| \sum_{m=0}^{t_n} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)} (\mathbb{E}\hat{R}(l,m) - R(l,m)) - \sum_{m>t_n} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)} R(l,m) \right\|_\infty \\ &\leq A_1 + A_2, \end{aligned}$$

where the terms A_1 and A_2 are given by

$$\begin{aligned} A_1 &:= \sum_{m=0}^{t_n} \sum_{l=-m}^m \sqrt{m+1} \|\varphi_{(l,m)}\|_\infty |\mathbb{E}\hat{R}(l,m) - R(l,m)|, \\ A_2 &:= \sum_{m>t_n} \sum_{l=-m}^m \sqrt{m+1} \|\varphi_{(l,m)}\|_\infty |R(l,m)|. \end{aligned}$$

For the term A_1 it follows that

$$A_1 \leq \sum_{m=0}^{t_n} \sum_{l=-m}^m (m+1) |\mathbb{E}\hat{R}(l,m) - R(l,m)| \leq \sum_{m=0}^{t_n} \sum_{l=-m}^m (m+1) C m^5 n^{-1} = \mathcal{O}(t_n^8 n^{-1}),$$

where we have used Proposition B.2 in Appendix B to see that

$$\|\varphi_{(l,m)}\|_\infty \leq \sqrt{m+1}. \tag{A.2}$$

Similarly we have

$$\begin{aligned} A_2 &\leq \sum_{m>t_n} \sum_{l=-m}^m (m+1) |R(l,m)| \leq \sum_{m>t_n} \sum_{l=-m}^m (m+1)^v t_n^{1-v} |R(l,m)| \\ &\leq t_n^{1-v} \sum_{m=0}^\infty \sum_{l=-m}^m (m+1)^v |R(l,m)| = \mathcal{O}(t_n^{1-v}). \end{aligned}$$

In the last step we have used that g complies to the smoothness condition of $\mathfrak{D}(v, L)$ (see (2.9)). □

A.2. Proof of Lemma 3.2

We first rewrite $\hat{g} - \mathbb{E}\hat{g}$ employing (2.16) and (A.2)

$$\begin{aligned} \|\hat{g}(z) - \mathbb{E}\hat{g}(z)\|_\infty &= \left\| \sum_{m=0}^{t_n} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)} (\hat{R}(l,m) - \mathbb{E}\hat{R}(l,m)) \right\|_\infty \\ &= \left\| \sum_{m=0}^{t_n} \sum_{l=-m}^m \sqrt{m+1} \varphi_{(l,m)} \left(\sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right) \right\|_\infty \\ &\leq \sum_{m=0}^{t_n} \sum_{l=-m}^m (m+1) \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right| \\ &= \sum_{m=0}^{t_n} \sum_{l=-m}^m (m+1)^2 \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right| \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{m=0}^{t_n} 2(m+1)^3 \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right| \\
&\leq C t_n^4 \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right|.
\end{aligned}$$

We proceed with deriving an upper bound for the maximum. For this purpose we introduce a truncation parameter $d_n := n^{1/2} \log(n)^{-1/2}$ and define the truncated error

$$\varepsilon_{\mathbf{k}}^{d_n} := \mathbf{1}\{|\varepsilon_{\mathbf{k}}| \leq d_n\} \varepsilon_{\mathbf{k}}. \quad (\text{A.3})$$

We will now show that all of the errors $\varepsilon_{\mathbf{k}}$ with $\mathbf{k} \in \mathbf{K}$ eventually equal their truncated versions $\varepsilon_{\mathbf{k}}^{d_n}$ almost surely. Via Markov's inequality we conclude that

$$\mathbb{P}(|\varepsilon_{\mathbf{k}}| > d_n) \leq \mathbb{E}[|\varepsilon|^{\kappa}] d_n^{-\kappa}$$

and therewith it follows that

$$\sum_{n=pq} \mathbb{P}(\exists \mathbf{k} \in \mathbf{K}: \varepsilon_{\mathbf{k}}^{d_n} \neq \varepsilon_{\mathbf{k}}) = \sum_{n=pq} \mathbb{P}(\exists \mathbf{k} \in \mathbf{K}: |\varepsilon_{\mathbf{k}}| > d_n) \leq \sum_{n=pq} n d_n^{-\kappa} \mathbb{E}[|\varepsilon|^{\kappa}].$$

Recalling that $n = pq$ and that there exists some $C_2 > 0$ such that $p \leq C_2 q$ by Assumption 2.2, we derive

$$C \sum_{n=pq} n d_n^{-\kappa} = C \sum_{n=pq} n^{1-\kappa/2} \log(n)^{\kappa/2} \leq C \sum_{q \geq 1} q^{2-\kappa} \log(C_2 q^2)^{\kappa/2} < \infty.$$

Summability is entailed by $2 - \kappa < -1$. The Borel–Cantelli Lemma implies that almost surely eventually all measurement errors and their truncated versions are equal. Thus we can confine ourselves to the maximum

$$\max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} \varepsilon_{\mathbf{k}}^{d_n} \right| \leq B_1 + B_2, \quad (\text{A.4})$$

where B_1 and B_2 are defined by

$$\begin{aligned}
B_1 &:= \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} [\varepsilon_{\mathbf{k}}^{d_n} - \mathbb{E} \varepsilon_{\mathbf{k}}^{d_n}] \right|, \\
B_2 &:= \sum_{\mathbf{k} \in \mathbf{K}} |\overline{\psi_{(l,m)}}(z_{\mathbf{k}})| (m+1)^{-1} w_{\mathbf{k}} |\mathbb{E} \varepsilon_{\mathbf{k}}^{d_n}|.
\end{aligned}$$

Using the inequality

$$\|\psi_{(l,m)}\|_{\infty} \leq m+1, \quad (\text{A.5})$$

which is a consequence of Proposition B.2, it follows that

$$B_2 \leq |\mathbb{E} \varepsilon^{d_n}| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} = o(n^{-1/2}),$$

where we exploit the decay rate $|\mathbb{E} \varepsilon^{d_n}| = o(n^{-1/2})$ in the last estimate. For the proof of this fact we recall the notation (A.3) and note that the condition $\mathbb{E} \varepsilon = 0$ implies

$$|\mathbb{E} \varepsilon^{d_n}| = |\mathbb{E}[\varepsilon \mathbf{1}\{|\varepsilon| > d_n\}]| \leq \int_{d_n}^{\infty} \mathbb{P}(|\varepsilon| > s) ds \leq \mathbb{E}[|\varepsilon|^{\kappa}] (\kappa - 1)^{-1} d_n^{1-\kappa} = o(n^{-1/2}).$$

For the term B_1 we note that for a fixed constant C^*

$$\mathbb{P}(|B_1| > \log(n)^{1/2} n^{-1/2} C^*)$$

$$\leq t_n^2 \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \mathbb{P} \left(\left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} [\varepsilon_{\mathbf{k}}^{d_n} - \mathbb{E} \varepsilon_{\mathbf{k}}^{d_n}] \right| > \log(n)^{1/2} n^{-1/2} C^* \right). \quad (\text{A.6})$$

Due to truncation $|\varepsilon_{\mathbf{k}}^{d_n} - \mathbb{E} \varepsilon_{\mathbf{k}}^{d_n}|$ is bounded by $2d_n$ and its variance by σ^2 . Furthermore the weights are uniformly of order $\mathcal{O}(n^{-1})$, since

$$\max_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} = 2\pi^{-2} \max_{\mathbf{k} \in \mathbf{K}} \int_{2\pi k_2/q}^{2\pi(k_2+1)/q} \int_{k_1/p}^{(k_1+1)/p} \sqrt{1-s^2} ds d\phi \leq 4(\pi pq)^{-1} = 4(\pi n)^{-1}.$$

Consequently the Bernstein inequality yields for the right side of (A.6) the upper bound

$$t_n^2 \exp \left(- \frac{CC^* \log(n)/n}{n^{-1} + \log(n)^{1/2} d_n/n^{3/2}} \right) \leq t_n^2 \exp(-CC^* \log(n)) \leq t_n^2 n^{-CC^*},$$

which is summable for sufficiently large C^* . The Borel–Cantelli Lemma therefore implies that

$$\max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}} [\varepsilon_{\mathbf{k}}^{d_n} - \mathbb{E} \varepsilon_{\mathbf{k}}^{d_n}] \right| = \mathcal{O}(\log(n)^{1/2} n^{-1/2}) \quad a.s.$$

Combining these estimates we see that the left side of (A.4) is almost surely of order $\mathcal{O}(\log(n)^{1/2} n^{-1/2})$. Consequently the right side of (A.3) is of order $\mathcal{O}(t_n^4 \log(n)^{1/2} n^{-1/2})$ almost surely, which proves the assertion. \square

A.3. Proof of Theorem 3.3

Combining Lemmas 3.1 and 3.2 yields the first part of Theorem 3.3 when the truncation parameter is chosen as in (3.2). For the proof of the second property we note the identity

$$\langle \mathcal{R}[g - \hat{g}], \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)} = R(l, m) - \hat{R}(l, m) \mathbf{1}\{m \leq t_n\},$$

which gives for the left-hand side of (3.4)

$$\begin{aligned} \sum_{m=0}^{\infty} \sum_{l=-m}^m m^\tau |\langle \mathcal{R}[g - \hat{g}], \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}| &= \sum_{m=0}^{\infty} \sum_{l=-m}^m m^\tau |R(l, m) - \hat{R}(l, m) \mathbf{1}\{m \leq t_n\}| \\ &\leq D_1 + D_2 + D_3. \end{aligned} \quad (\text{A.7})$$

The terms D_1 , D_2 , and D_3 are defined as follows:

$$D_1 := \sum_{m=0}^{t_n} \sum_{l=-m}^m m^\tau |R(l, m) - \mathbb{E} \hat{R}(l, m)|,$$

$$D_2 := \sum_{m=0}^{t_n} \sum_{l=-m}^m m^\tau |\hat{R}(l, m) - \mathbb{E} \hat{R}(l, m)|,$$

$$D_3 := \sum_{m > t_n} \sum_{l=-m}^m m^\tau |R(l, m)|.$$

By Proposition A.1 we obtain the upper bound

$$D_1 \leq \sum_{m=0}^{t_n} \sum_{l=-m}^m C m^{\tau+5} n^{-1} = \mathcal{O}(t_n^{\tau+7} n^{-1}).$$

For the second sum on right of (A.7) we use the bound

$$D_2 = \sum_{m=0}^{t_n} \sum_{l=-m}^m m^\tau (m+1) [(m+1)^{-1} |\hat{R}(l, m) - \mathbb{E} \hat{R}(l, m)|]$$

$$\leq C t_n^{\tau+3} \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \{ (m+1)^{-1} |\hat{R}(l, m) - \mathbb{E}\hat{R}(l, m)| \} = \mathcal{O}(n^{-1/2} \log(n)^{1/2} t_n^{\tau+3}) \text{ a.s.}$$

In the last equality we have used the following bound established in the proof of Lemma 3.2:

$$\begin{aligned} & \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \{ (m+1)^{-1} |\hat{R}(l, m) - \mathbb{E}\hat{R}(l, m)| \} \\ &= \max_{\substack{(l,m) \in \mathcal{N} \\ m \leq t_n}} \left| (m+1)^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) w_{\mathbf{k}} \varepsilon_{\mathbf{k}} \right| = \mathcal{O}(n^{-1/2} \log(n)^{1/2}) \text{ a.s.} \end{aligned}$$

The third term in (A.7) can be bounded by

$$D_3 \leq \sum_{m > t_n} \sum_{l=-m}^m (m t_n)^{v-\tau} m^\tau |R(l, m)| = t_n^{v-\tau} \sum_{m > t_n} \sum_{l=-m}^m m^v |R(l, m)|.$$

Due to the smoothness condition in (2.9) the double sum is finite. Since $t_n \rightarrow \infty$ it follows that the series converges to 0 for $n \rightarrow \infty$. Consequently

$$t_n^{v-\tau} \sum_{m > t_n} \sum_{l=-m}^m m^v |R(l, m)| = o(t_n^{v-\tau})$$

and the definition of t_n in (3.2) yield the desired result. □

A.4. Proof of Theorem 3.5

The proof of Theorem 3.5 relies on an equicontinuity argument, which requires an investigation of the smoothness class $\mathfrak{D}(v, L)$. To access its complexity we introduce two measures of entropy:

Definition A.2. (i) Let $\mathcal{F} := \{f: \Theta \rightarrow \mathbb{R}\}$ be a class of continuous functions defined on $\Theta \subset \mathbb{R}^p$. For two continuous functions $g^L, g^U: \Theta \rightarrow \mathbb{R}$ with $\|g^L - g^U\|_\infty \leq \varepsilon$, we call

$$[g^L, g^U] := \{f \in \mathcal{F}: g^L \leq f \leq g^U\}$$

an ε -bracket. The *maximum bracketing number* $J = N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\infty)$ of \mathcal{F} is the minimum number of ε -brackets $[g_1^L, g_1^U], \dots, [g_J^L, g_J^U]$ such that

$$\mathcal{F} \subset \bigcup_{j=1}^J [g_j^L, g_j^U].$$

Notice that the functions g_j^L, g_j^U do not have to be elements of \mathcal{F} .

(ii) Let $Z_{1,n}, \dots, Z_{n,n}$ be stochastic processes indexed in \mathcal{F} and $\varepsilon > 0$. The L_2^n -bracketing number of \mathcal{F} , denoted by $N_{[]}^{L_2^n}(\varepsilon, \mathcal{F})$, is the minimal number N_ε of sets $\mathcal{F}_{\varepsilon,j}^n$ in a partition of $\mathcal{F} = \bigcup_{j=1}^{N_\varepsilon} \mathcal{F}_{\varepsilon,j}^n$ such that for each j

$$\sum_{i=1}^n \mathbb{E} \left[\sup_{f, \tilde{f} \in \mathcal{F}_{\varepsilon,j}^n} (Z_{i,n}(f) - Z_{i,n}(\tilde{f}))^2 \right] \leq \varepsilon^2. \tag{A.8}$$

As we know from Proposition 2.1, the series condition

$$\sum_{m=0}^{\infty} \sum_{l=-m}^m (m+1)^\tau |\langle \mathcal{R}h, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)}| \leq 1$$

entails that a function $h \in \mathcal{L}^2(\mathcal{B}, \mu)$ is smooth to a degree determined by τ . This implies that a finite-dimensional representation can be used as an adequate approximation of h , in our case a truncated \mathcal{L}^2 -expansion. We employ this considerations to derive the following result about the complexity of the class $\mathfrak{D}(\tau, 1)$, which is of independent interest and is proven in Appendix B (see Section B.4).

Proposition A.3. *Let $\tau > 3$, then for any $t \in (3, \tau)$ and sufficiently small $\epsilon > 0$*

$$\log(N_{[]}(\epsilon, \mathfrak{D}(\tau, 1), \|\cdot\|_\infty)) \leq C \left(\frac{1}{\epsilon}\right)^{\frac{2}{\tau-t}}. \tag{A.9}$$

Here $N_{[]}(\epsilon, \mathfrak{D}(\tau, 1), \|\cdot\|_\infty)$ denotes the minimal number of ϵ -brackets with respect to $\|\cdot\|_\infty$ needed to cover the smoothness class $\mathfrak{D}(\tau, 1)$.

Lemma A.4. *Define*

$$M_n(t) := \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \{ \mathbf{1}\{\hat{\epsilon}_{\mathbf{k}} \leq t\} - \mathbf{1}\{\epsilon_{\mathbf{k}} \leq t\} + \mathbb{F}(t) - \mathbb{P}(\hat{\epsilon}_{\mathbf{k}} \leq t) \} \right|. \tag{A.10}$$

Then, under the assumptions of Theorem 3.5, it follows that $\sup_{t \in \mathbb{R}} |M_N(t)| = o_P(n^{-1/2})$.

Proof of Lemma A.4. Using the definition of the estimated residuals in (3.1) we have

$$\begin{aligned} & \mathbf{1}\{\hat{\epsilon}_{\mathbf{k}} \leq t\} - \mathbf{1}\{\epsilon_{\mathbf{k}} \leq t\} + \mathbb{F}(t) - \mathbb{P}(\hat{\epsilon}_{\mathbf{k}} \leq t) \\ &= \mathbf{1}\{\epsilon_{\mathbf{k}} \leq t + \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})\} - \mathbf{1}\{\epsilon_{\mathbf{k}} \leq t\} + \mathbb{P}(\epsilon \leq t) - \mathbb{P}(\epsilon_{\mathbf{k}} \leq t + \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})). \end{aligned}$$

As we have seen in Theorem 3.3, the random function $\Delta_n := \mathcal{R}[g - \hat{g}]$ is eventually included in the smoothness class $\mathcal{R}(\mathfrak{D}(\tau, 1))$ for every $\tau < v$. Since $v > 5$ by assumption, we can also choose a $\tau > 5$. Since Δ_n is a complicated object, depending on all residuals, we replace it by general functions in $\mathcal{R}(\mathfrak{D}(\tau, 1))$ and prove a uniform result over $\mathfrak{D}(\tau, 1)$. We thus define the stochastic processes

$$Z_{n,\mathbf{k}}(t, d) := n^{1/2} w_{\mathbf{k}} (\mathbf{1}\{\epsilon_{\mathbf{k}} \leq t + d(z_{\mathbf{k}})\} - \mathbf{1}\{\epsilon_{\mathbf{k}} \leq t\}),$$

indexed in the space $\mathcal{F} := \mathbb{R} \times \mathcal{R}(\mathfrak{D}(\tau, 1))$, equipped with the semi-metric

$$\rho((t, d), (\tilde{t}, \tilde{d})) := \max \left\{ \sup_{x \in [-1, 1]} \{ |\mathbb{F}(t+x) - \mathbb{F}(\tilde{t}+x)|, \|d - \tilde{d}\|_\infty \} \right\}. \tag{A.11}$$

Notice that for ρ to be a semi-metric, the error density f_ϵ must have support \mathbb{R} , which is assumed at this point for the sake of simplicity. Furthermore recall the uniform order of the product $n^{1/2} w_{\mathbf{k}} = \mathcal{O}(n^{-1/2})$. To prove equicontinuity, we have to show that for every sequence $\delta_n \downarrow 0$ and every $\epsilon > 0$

$$\mathbb{P} \left(\sup_{\substack{(t,d),(\tilde{t},\tilde{d}) \in \mathcal{F} \\ \rho((t,d),(\tilde{t},\tilde{d})) < \delta_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} (Z_{n,\mathbf{k}}(t, d) - \mathbb{E}Z_{n,\mathbf{k}}(t, d) - Z_{n,\mathbf{k}}(\tilde{t}, \tilde{d}) + \mathbb{E}Z_{n,\mathbf{k}}(\tilde{t}, \tilde{d})) \right| > \epsilon \right) \rightarrow 0. \tag{A.12}$$

If (A.12) holds, then the assertion of Lemma A.4 can be shown as follows: Firstly note that we can derive a lower bound for the probability on the left-hand side of (A.12) by

$$\begin{aligned} & \mathbb{P} \left(\left\{ \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} Z_{n,\mathbf{k}}(t, \Delta_n) - \mathbb{E}Z_{n,\mathbf{k}}(t, \Delta_n) - Z_{n,\mathbf{k}}(t, 0) + \mathbb{E}Z_{n,\mathbf{k}}(t, 0) \right| > \epsilon \right\} \right. \\ & \quad \left. \cap \{ \Delta_n \in \mathcal{R}(\mathfrak{D}(\tau, 1)) \} \cap \left\{ \sup_{t \in \mathbb{R}} \rho((t, 0), (t, \Delta_n)) < \delta_n \right\} \right) \\ &= \mathbb{P} \left(\left\{ \sup_{t \in \mathbb{R}} n^{1/2} M_n(t) > \epsilon \right\} \cap \{ \Delta_n \in \mathcal{R}(\mathfrak{D}(\tau, 1)) \} \cap \left\{ \sup_{t \in \mathbb{R}} \rho((t, 0), (t, \Delta_n)) < \delta_n \right\} \right). \tag{A.13} \end{aligned}$$

By the second part of Theorem 3.3 we know that

$$\mathbb{P}(\Delta_n \in \mathcal{R}(\mathfrak{D}(\tau, 1))) \rightarrow 1$$

for $\tau < v$. Furthermore we notice that $\|\mathcal{R}d\|_\infty \leq \|d\|_\infty$ for any continuous function d , which follows immediately from the definition of the Radon transform in (1.2). Combining this with the upper bound

$$\|g - \hat{g}\|_\infty = \mathcal{O} \left(\left(\frac{\log(n)}{n} \right)^{\frac{v-1}{2(v+3)}} \right)$$

from the first part of Theorem 3.3 yields

$$\rho((t, 0), (t, \Delta_n)) = \|\mathcal{R}[\hat{g} - g]\|_\infty \leq \|\hat{g} - g\|_\infty = \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\frac{v-1}{2(v+3)}}\right) \text{ a.s.,}$$

so that for a sequence $\delta_n \downarrow 0$, say e.g. $\delta_n = \log(n)^{-1}$,

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} \rho((t, 0), (t, \Delta_n)) < \delta_n\right) \rightarrow 1.$$

Combining these considerations with the right-hand side of (A.13) implies that $n^{1/2}M_n(t) = o_P(1)$ uniformly in t , proving the Lemma provided that (A.12) holds.

This statement is a consequence of Lemma A.19 in [42], which requires four regularity properties of the process under consideration. The rest of the proof consists in verifying these properties.

(1) For all $\eta > 0$ we have to show:

$$\sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E}\left(\sup_{(t,d) \in \mathcal{F}} |Z_{n,\mathbf{k}}(t,d)| \mathbf{1}\{|Z_{n,\mathbf{k}}(t,d)| > \eta\}\right) \rightarrow 0.$$

This is easy to see, since $|Z_{n,\mathbf{k}}| \leq Cn^{-1/2}$ (recall that $\max_{\mathbf{k}} w_{\mathbf{k}} \leq Cn^{-1}$) and so the sum is equal to 0 for all n larger than some n_0 .

(2) For every sequence $\delta_n \downarrow 0$

$$\sup_{\rho((t,d),(\tilde{t},\tilde{d})) < \delta_n} \left| \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E}[(Z_{n,\mathbf{k}}(t,d) - Z_{n,\mathbf{k}}(\tilde{t},\tilde{d}))^2] \right| \rightarrow 0. \tag{A.14}$$

Consider the expectation for some fixed but arbitrary $\mathbf{k} \in \mathbf{K}$ which can be bounded uniformly as follows:

$$\begin{aligned} & \mathbb{E}[(Z_{n,\mathbf{k}}(t,d) - Z_{n,\mathbf{k}}(\tilde{t},\tilde{d}))^2] \\ & \leq Cn^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + |\mathbb{F}(t) - \mathbb{F}(\tilde{t})| \right] \\ & \leq Cn^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + \delta_n \right] \\ & \leq Cn^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(t + \tilde{d}(z_{\mathbf{k}}))| + |\mathbb{F}(t + \tilde{d}(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + \delta_n \right]. \end{aligned}$$

All three terms inside the square brackets are uniformly of order $o(1)$. This can be shown as follows: An application of the mean-value theorem demonstrates that the first term is a null sequence:

$$|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(t + \tilde{d}(z_{\mathbf{k}}))| \leq \|f_\epsilon\|_\infty \|d - \tilde{d}\|_\infty \leq C\delta_n \rightarrow 0.$$

The middle term is bounded by δ_n by the definition of our semi-metric ρ in (A.11), when we consider that $\tilde{d} \in \mathcal{R}(\mathfrak{D}(\tau, 1))$ and therefore $\|\tilde{d}\|_\infty \leq 1$. Consequently it is $o(1)$, as well as the last term by assumption.

(3) Denoting the L_2^n -bracketing number, as given in Definition A.2 (ii), by $N_{\square}^{L_2^n}(\epsilon, \mathcal{F})$, the condition we have to check next is that for every sequence $\delta_n \downarrow 0$

$$\int_0^{\delta_n} \sqrt{\log(N_{\square}^{L_2^n}(\epsilon, \mathcal{F}))} d\epsilon \rightarrow 0. \tag{A.15}$$

For the construction of an adequate partition of \mathcal{F} satisfying (A.8), consider ϵ^2 -brackets $[g_j^L, g_j^U]$, $j = 1, \dots, J = \mathcal{O}(\exp(\epsilon^{4/(\tau-t)}))$, of $\mathfrak{D}(\tau, 1)$ (see Definition A.2 (i)), where $t > 3$ is such that $\tau - t > 2$. Recall that $\tau > 5$ by assumption. Such brackets exist according to Proposition A.3. The images of these brackets under \mathcal{R} are simply $[\mathcal{R}g_j^L, \mathcal{R}g_j^U]$ due to monotonicity of the integral and they are still ϵ^2 -brackets, since \mathcal{R} reduces $\|\cdot\|_\infty$ -distance. As a consequence we obtain ϵ^2 -brackets $[\mathcal{R}g_j^L, \mathcal{R}g_j^U]$ of the whole class $\mathcal{R}(\mathfrak{D}(\tau, 1))$.

Additionally choose $y_i^L < y_i^U$ with $i = 1, \dots, I = \mathcal{O}(\epsilon^{-2})$ such that the intervals $[y_i^L, y_i^U]$ form a partition of the real line (for infinite values we take the intervals to be half open) and such that each interval has probability mass $\leq \epsilon^2$. Then the sets

$$\mathcal{F}_{i,j,\epsilon}^n := [y_i^L, y_i^U] \times [\mathcal{R}g_j^L, \mathcal{R}g_j^U] \tag{A.16}$$

form a partition of \mathcal{F} . Their number is of order $\mathcal{O}(\exp(\epsilon^{4/(\tau-t)}))$, where we might have to slightly shrink t so that still $t > 3$ and $\tau - t > 2$ hold. Now we have to show that (A.8) holds, that is in the present case for an arbitrary $\mathcal{F}_{i,j,\epsilon}^n$

$$\sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E} \left[\sup_{(t,d),(\tilde{t},\tilde{d}) \in \mathcal{F}_{i,j,\epsilon}^n} (Z_{n,\mathbf{k}}(t,d) - Z_{n,\mathbf{k}}(\tilde{t},\tilde{d}))^2 \right] \leq \epsilon^2.$$

In the subsequent calculation we define the expressions $\mathbb{F}(\pm\infty)$ and $f_\epsilon(\pm\infty)$ by taking the respective limits. The left-hand side of the above inequality is bounded by

$$\begin{aligned} & Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E} \left[\sup_{(t,d),(\tilde{t},\tilde{d}) \in \mathcal{F}_{i,j,\epsilon}^n} \left(|\mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t + d(z_{\mathbf{k}})\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq \tilde{t} + \tilde{d}(z_{\mathbf{k}})\}| \right. \right. \\ & \quad \left. \left. + |\mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq \tilde{t}\}| \right)^2 \right] \\ & \leq Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E} \left[|\mathbf{1}\{\varepsilon_{\mathbf{k}} \leq y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}})\}| \right. \\ & \quad \left. + |\mathbf{1}\{\varepsilon_{\mathbf{k}} \leq y_i^U\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq y_i^L\}| \right] \\ & \leq Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \left[|\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))| + C|\mathbb{F}(y_i^U) - \mathbb{F}(y_i^L)| \right] \\ & \leq Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} |\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))| + C\epsilon^2 \\ & \leq Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \{ |\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^U(z_{\mathbf{k}}))| \\ & \quad + |\mathbb{F}(y_i^L + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))| \} + C\epsilon^2 \\ & \leq C(\epsilon^2 + \|f_\epsilon(y)\|_\infty \|\mathcal{R}g_j^U - \mathcal{R}g_j^L\|_\infty + \epsilon^2) \leq C\epsilon^2. \end{aligned}$$

Replacing ϵ by $\epsilon C^{-1/2}$ yields the desired result without changing the rate of the upper bound $\mathcal{O}(\exp(\epsilon^{4/(\tau-t)}))$ of the L_2^n -bracketing number. Thus the integral in (A.15) converges, since $\tau - t > 2$.

(4) Finally we have to prove that (\mathcal{F}, ρ) is totally bounded. By definition, ρ is a maximum semi-metric defined on the product space $\mathbb{R} \times \mathcal{R}(\mathfrak{D}(\tau, 1))$. Hence it suffices to show that each of the spaces (\mathbb{R}, ρ_1) , $(\mathcal{R}(\mathfrak{D}(\tau, 1)), \rho_2)$ is totally bounded, where we define for $t, \tilde{t} \in \mathbb{R}$ and $d_1, d_2 \in \mathcal{R}(\mathfrak{D}(\tau, 1))$

$$\rho_1(t, \tilde{t}) := \sup_{x \in [-1,1]} |\mathbb{F}(t+x) - \mathbb{F}(\tilde{t}+x)|, \quad \rho_2(d_1, d_2) := \|d_1 - d_2\|_\infty.$$

We start with ρ_1 and demonstrate that for every $\epsilon > 0$ we can find a finite number of $t_1, \dots, t_J \in \mathbb{R}$ such that for every $t \in \mathbb{R}$ there exists a t_j such that

$$\sup_{x \in [-1,1]} |\mathbb{F}(t+x) - \mathbb{F}(t_j+x)| \leq \epsilon. \tag{A.17}$$

Let $M := \max_{t \in \mathbb{R}} |f_\epsilon(t)|$ and I be a closed interval with probability mass larger than $1 - \epsilon$. Take an equidistant grid with maximal width ϵ/M of points t_j for $j = 1, \dots, J$ across I (including the boundary

points) and now let, for an arbitrary $t \in \mathbb{R}$ say t_j , be one of the closest points to t of this grid. If $t \notin I$ we choose a boundary point of I and the result is immediate. If $t \in I$, we get by the mean-value theorem:

$$|\mathbb{F}(t+x) - \mathbb{F}(t_j+x)| \leq \epsilon \|f_\epsilon\|_\infty M^{-1} = \epsilon.$$

For ρ_2 we recall that by our above observations for every $\epsilon > 0$ the maximum-bracketing number of $\mathcal{R}(\mathcal{D}, 1)$ with respect to the norm $\|\cdot\|_\infty$ is finite and thus in particular we have total boundedness.

Having established these regularity properties, by Neumeyer’s Lemma A.19 (see [42]) equicontinuity follows, which completes the proof of Lemma A.4. \square

Besides Lemma A.4 we require some additional approximation results for the proof of Theorem 3.5.

Proposition A.5. *Under the assumptions of Theorem 3.5 we have*

$$\sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \mathbb{P}(\hat{\epsilon}_{\mathbf{k}} \leq t) - \int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z) = o_P(n^{-1/2}), \tag{A.18}$$

$$\int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z) - \mathbb{F}(t) - f_\epsilon(t) \int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) = o_P(n^{-1/2}), \tag{A.19}$$

$$\int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) - \sum_{\mathbf{k} \in \mathbf{K}} \epsilon_{\mathbf{k}} w_{\mathbf{k}} = o_P(n^{-1/2}). \tag{A.20}$$

Proof of Proposition A.5. Recalling the definitions of the estimated residuals $\hat{\epsilon}_{\mathbf{k}} := \epsilon_{\mathbf{k}} - \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})$ and the weights $w_{\mathbf{k}} := \lambda(B_{\mathbf{k}})$, we begin by rewriting the left-hand side of (A.18)

$$\sum_{\mathbf{k} \in \mathbf{K}} \int_{B_{\mathbf{k}}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})) - \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z).$$

According to the mean-value theorem the absolute value of this term is bounded by

$$\sum_{\mathbf{k} \in \mathbf{K}} \int_{B_{\mathbf{k}}} |\{\mathcal{R}[\hat{g} - g](z_{\mathbf{k}}) - \mathcal{R}[\hat{g} - g](z)\} f_\epsilon(t_z)| d\lambda(z),$$

where t_z is some suitable point between $t + \mathcal{R}[\hat{g} - g](z)$ and $t + \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})$. Since the density is bounded, it suffices to show that

$$\sup_{z \in B_{\mathbf{k}}} |\mathcal{R}[\hat{g} - g](z_{\mathbf{k}}) - \mathcal{R}[\hat{g} - g](z)| = o_P(n^{-1/2}).$$

An application of Cauchy–Schwarz yields

$$\sup_{z \in B_{\mathbf{k}}} |\mathcal{R}[\hat{g} - g](z_{\mathbf{k}}) - \mathcal{R}[\hat{g} - g](z)| \leq 2 \|\nabla \mathcal{R}[\hat{g} - g]\|_\infty \|z - z_{\mathbf{k}}\|_2.$$

By Assumption 2.2 $\|z - z_{\mathbf{k}}\|_2 = \mathcal{O}(1/\sqrt{n})$. Moreover by Corollary 3.6 the gradient $\nabla \mathcal{R}[\hat{g} - g]$ converges uniformly to 0. Thus we get the desired result. The estimate (A.19) follows by similar arguments, while (A.20) is based on two observations. Firstly, since $\psi_{(0,0)} = 1$, we can rewrite the integral

$$\int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) = \langle \mathcal{R}[\hat{g} - g], \psi_{(0,0)} \rangle_{\mathcal{L}^2(\mathcal{D}, \lambda)} = \hat{R}(0, 0) - R(0, 0).$$

Secondly, as the errors are centered,

$$\sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \epsilon_{\mathbf{k}} = \hat{R}(0, 0) - \mathbb{E}[\hat{R}(0, 0)].$$

Combining these results yields the representation $\mathbb{E}[\hat{R}(0, 0)] - R(0, 0)$ for the left-hand side of (A.20). By Proposition A.1 this difference is of order $\mathcal{O}(n^{-1})$. \square

Equipped with our observations in Lemma A.4 and Proposition A.5 Theorem 3.5 is easily deduced:

Proof of Theorem 3.5. We apply the triangle inequality to arrive at the following decomposition:

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} [\mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t\} - \varepsilon_{\mathbf{k}} f_{\varepsilon}(t)] \right| \\ & \leq \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} [\mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\} - \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t\} + \mathbb{F}(t) - \mathbb{P}(\hat{\varepsilon}_{\mathbf{k}} \leq t)] \right| \\ & \quad + \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \mathbb{P}(\hat{\varepsilon}_{\mathbf{k}} \leq t) - \int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z) \right| \\ & \quad + \sup_{t \in \mathbb{R}} \left| \int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z) - \mathbb{F}(t) - f_{\varepsilon}(t) \int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) \right| \\ & \quad + \sup_{t \in \mathbb{R}} \left| f_{\varepsilon}(t) \int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) - \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \varepsilon_{\mathbf{k}} f_{\varepsilon}(t) \right|. \end{aligned}$$

Each of the terms on the right-hand side is of order $o_P(n^{-1/2})$, the first one by Lemma A.4 and the other ones by Proposition A.5. \square

A.5. Proof of Corollary 3.6

This is a consequence of Theorem 3.5, as we can represent the process $\sqrt{n}(\mathbb{F} - \hat{\mathbb{F}}_n)$ as a sum of independent stochastic processes and a negligible term:

$$\begin{aligned} \sqrt{n}(\mathbb{F} - \hat{\mathbb{F}}_n) &= \sum_{\mathbf{k} \in \mathbf{K}} \sqrt{n} w_{\mathbf{k}} \{ \mathbf{1}\{\hat{\varepsilon}_{\mathbf{k}} \leq t\} - \mathbb{F}(t) \} \\ &= \sum_{\mathbf{k} \in \mathbf{K}} \sqrt{n} w_{\mathbf{k}} \{ \mathbf{1}\{\varepsilon_{\mathbf{k}} \leq t\} - \mathbb{F}(t) + \varepsilon_{\mathbf{k}} f_{\varepsilon}(t) \} + o_P(1). \end{aligned}$$

The sum on the right-hand side converges to a Gaussian process by an application of Lemma A.19 in [42].

APPENDIX B. AUXILIARY RESULTS

B.1. Uniform Bounds

We begin with stating some frequently used properties of the radial polynomials which are taken from [8] and [24].

Proposition B.1. (1) For all $(l, m) \in \mathcal{N}$

$$\sup_{0 \leq r \leq 1} |R_m^{[l]}(r)| = 1. \tag{B.1}$$

(2) For all $(l, m), (l, m') \in \mathcal{N}$

$$\int_0^1 \sqrt{2(m+1)} R_m^{[l]}(r) \sqrt{2(m'+1)} R_{m'}^{[l]}(r) r dr = \delta_{m,m'}. \tag{B.2}$$

(3) For all $(l, m) \in \mathcal{N}$ the derivative of the corresponding radial polynomial has the following structure:

$$\begin{aligned} \frac{d}{dr} R_m^{[l]}(r) &= \sum_{j=0}^{\frac{1}{2}(m-1-|l-1|)} (m-2j) R_{m-1-2j}^{[l-1]}(r) \\ &\quad + \sum_{j=0}^{\frac{1}{2}(m-1-|l+1|)} (m-2j) R_{m-1-2j}^{[l+1]}(r). \end{aligned} \tag{B.3}$$

Next we provide upper bounds on the $\|\cdot\|_\infty$ -norm of the derivatives of the Chebyshev and radial polynomials. The bounds on the radial polynomials follow by the above Proposition and the bounds for Chebyshev's by identities from [36].

Proposition B.2. *Let $k \in \mathbb{N}_0$ and $(l, m) \in \mathcal{N}$, then*

$$\sup_{0 \leq r \leq 1} \left| \frac{d^k}{dr^k} R_m^{|l|}(r) \right| \leq m^{2k}$$

and

$$\sup_{0 \leq s \leq 1} \left| \frac{d^k}{ds^k} U_m(s) \right| \leq (m+1)m^{2k}.$$

Proof of Proposition B.2. In order to show the first statement, we apply the identities (B.1) and (B.3) from Proposition B.1 and use an induction argument. The initial step is given by (B.1) and the induction hypothesis is

$$\sup_{0 \leq r \leq 1} \left| \frac{d^k}{dr^k} R_m^{|l|}(r) \right| \leq m^{2k}.$$

By virtue of (B.3) we have

$$\begin{aligned} \left| \frac{d^{k+1}}{dr^{k+1}} R_m^{|l|}(r) \right| &= \left| \sum_{j=0}^{\frac{1}{2}(m-1-|l-1|)} (m-2j) \frac{d^k}{dr^k} R_{m-1-2j}^{|l-1|}(r) + \sum_{j=0}^{\frac{1}{2}(m-1-|l+1|)} (m-2j) \frac{d^k}{dr^k} R_{m-1-2j}^{|l+1|}(r) \right| \\ &\leq 2 \sum_{j=0}^{\frac{1}{2}(m-1-|l-1|)} (m-2j) m^{2k} \leq m^{2k+2}, \end{aligned}$$

where we have used the induction hypothesis to bound the derivatives of $R_m^{|l|}$.

The case of the Chebyshev polynomials is similar. In order to prove the second identity in Proposition B.2 we cite a few well-known facts about Chebyshev polynomials from [36]:

- (1) For all $m \in \mathbb{N}$, U_m is uniformly bounded by $m+1$.
- (2) Let T_m denote the Chebyshev polynomial of the first kind, which satisfies the differential equation

$$\frac{d}{ds} T_m(s) = U_{m-1}(s)m.$$

For all $m \in \mathbb{N}$, T_m is uniformly bounded by 1.

- (3) For all $m \in \mathbb{N}$ the representation

$$\frac{d}{ds} U_m(s) = \sum_{j=0}^{\bullet m-2} \frac{(m^2 - j^2)m}{m+1} T_j(s)$$

holds, where \bullet indicates that we only sum over such terms where $m-j$ is even.

The proof now follows by induction, analogously to that of the first part. □

B.2. Proof of Proposition 2.1

We employ these bounds to sketch a proof of Proposition 2.1. The techniques are borrowed from the theory of Fourier series. It is well known that a continuous function f on a compact interval with absolutely summable Fourier coefficients is identical to its Fourier series f_∞ . This is most easily proven by observing that f and f_∞ are identical in mean and that by uniform convergence f_∞ is also continuous. We proceed analogously for the proof of (2.10)–(2.13). The differentiability is an immediate consequence of this argument. To avoid redundancy we confine our investigation to equation (2.10).

Firstly we define the function on the right side of (2.10) by \tilde{g} . Obviously

$$\int_{\mathcal{B}} (g - \tilde{g})^2 d\mu = 0. \tag{B.4}$$

As μ is absolutely continuous with respect to the Lebesgue measure the set $\{g = \tilde{g}\}$ has Lebesgue measure 0 and thus (2.10) follows if we can establish continuity of \tilde{g} (recall that g is continuous by assumption). Continuity of \tilde{g} follows from the uniform convergence of the sequence of continuous functions

$$\left(\sum_{m=0}^N \sum_{l=-m}^m \varphi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right)_{N \in \mathbb{N}} \tag{B.5}$$

to \tilde{g} for $N \rightarrow \infty$. To see this we consider the difference

$$\begin{aligned} \left\| \tilde{g} - \sum_{m=0}^N \sum_{l=-m}^m \varphi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right\|_\infty &= \left\| \sum_{m=N}^\infty \sum_{l=-m}^m \varphi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right\|_\infty \\ &\leq \sum_{m=N}^\infty \sum_{l=-m}^m \|\varphi_{(l,m)}\|_\infty |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \leq \sum_{m=N}^\infty \sum_{l=-m}^m \sqrt{m+1} |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}|, \end{aligned}$$

where we used (A.2) in the last step. Plugging the identity (2.13) (recall that we already know it in an \mathcal{L}^2 -sense from equation (2.8)) into the inner products yields

$$\sum_{m=N}^\infty \sum_{l=-m}^m \sqrt{m+1} |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| = \sum_{m=N}^\infty \sum_{l=-m}^m (m+1) |\langle \mathcal{R}g, \psi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{D},\lambda)}|. \tag{B.6}$$

By the series condition in (2.9), the right and thus the left side converge to 0, which proves continuity of \tilde{g} . Consequently, it follows from (B.4) that $g = \tilde{g}$.

To establish differentiability of g and $\mathcal{R}g$ we use their \mathcal{L}^2 -representations (2.10) and (2.11). Differentiability and summation may be interchanged by uniformity arguments, using the bounds from Proposition B.2. Continuity of the derivatives is then derived as in the above argumentation.

B.3. Proof of Proposition A.1

By definition of $\hat{R}(l, m)$ in (2.16) and the weights $w_{\mathbf{k}} = \lambda(B_{\mathbf{k}})$ we obtain:

$$\begin{aligned} |R(l, m) - \mathbb{E}\hat{R}(l, m)| &\leq \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \overline{\psi_{(l,m)}(z)} \mathcal{R}g(z) - \overline{\psi_{(l,m)}(z_{\mathbf{k}})} \mathcal{R}g(z_{\mathbf{k}}) d\lambda(z) \right| \\ &\leq \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \Re(\psi_{(l,m)}(z)) \mathcal{R}g(z) - \Re(\psi_{(l,m)}(z_{\mathbf{k}})) \mathcal{R}g(z_{\mathbf{k}}) d\lambda(z) \right| \\ &\quad + \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \Im(\psi_{(l,m)}(z)) \mathcal{R}g(z) - \Im(\psi_{(l,m)}(z_{\mathbf{k}})) \mathcal{R}g(z_{\mathbf{k}}) d\lambda(z) \right|. \end{aligned}$$

By Proposition 2.1 the function $\mathcal{R}g$ is twice continuously differentiable. Recalling the definition of $\psi_{(l,m)}$, we observe that the real part

$$\Re(\psi_{(l,m)}(s, \phi)) = U_m(s) \cos(\phi l) \tag{B.7}$$

and the imaginary part

$$\Im(\psi_{(l,m)}(s, \phi)) = U_m(s) \sin(\phi l) \tag{B.8}$$

are infinitely often differentiable. By Proposition B.2 it is now easy to see that all second order derivatives of these functions are uniformly bounded by $2m^5$. We now use a Taylor expansion and obtain for any $\mathbf{k} = (k_1, k_2)$

$$\begin{aligned} R(\mathbf{k}) &:= \int_{B_{\mathbf{k}}} \Re(\psi_{(l,m)}(z)) \mathcal{R}g(z) - \Re(\psi_{(l,m)}(z_{\mathbf{k}})) \mathcal{R}g(z_{\mathbf{k}}) d\lambda(z) \\ &= \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (s - z_{k_1}) \frac{d}{ds} [\Re(\psi_{(l,m)}) \mathcal{R}g](z_{\mathbf{k}}) 2\pi^{-1} \sqrt{1 - s^2} d\phi ds \\ &+ \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (\phi - z_{k_2}) \frac{d}{d\phi} [\Re(\psi_{(l,m)}) \mathcal{R}g](z_{\mathbf{k}}) 2\pi^{-1} \sqrt{1 - s^2} d\phi ds \\ &+ \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (s - z_{k_1})(\phi - z_{k_2}) \frac{d}{d\phi} \frac{d}{ds} [\Re(\psi_{(l,m)}) \mathcal{R}g](\xi_1) 2\pi^{-1} \sqrt{1 - s^2} d\phi ds \\ &+ \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} 2^{-1}(s - z_{k_1})^2 \frac{d^2}{ds^2} [\Re(\psi_{(l,m)}) \mathcal{R}g](\xi_2) 2\pi^{-1} \sqrt{1 - s^2} d\phi ds \\ &+ \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} 2^{-1}(\phi - z_{k_2})^2 \frac{d^2}{d\phi^2} [\Re(\psi_{(l,m)}) \mathcal{R}g](\xi_3) 2\pi^{-1} \sqrt{1 - s^2} d\phi ds. \end{aligned}$$

Here ξ_1, ξ_2, ξ_3 denote points dependent on s, ϕ and $z_{\mathbf{k}}$, which are located inside $B_{\mathbf{k}}$ because of its convexity. The first two integrals vanish because of the choice of our design points. Moreover $|s - z_{k_1}|$ and $|\phi - z_{k_2}|$ are bounded by $Cn^{-1/2}$ by Assumption 1. The second order derivatives of $\mathcal{R}g$ are bounded (because they are continuous) and those of $\Re(\psi_{(l,m)})$ are bounded by $2m^5$, as we have noted above. Thus the term $R(\mathbf{k})$ is of order $\mathcal{O}(m^5 n^{-1})$. Treating the integrals in the sum over the imaginary parts in same fashion yields the result. \square

B.4. Proof of Proposition A.3

We begin by rewriting the series condition (2.9) as

$$1 \geq \sum_{m=0}^{\infty} \sum_{l=-m}^m (m+1)^{\tau} |R(l, m)| = \sum_{m=0}^{\infty} \sum_{l=-m}^m (m+1)^{\tilde{\tau}} |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)}|, \tag{B.9}$$

where we define $\tilde{\tau} := \tau - 1/2$ for convenience of notation. The reason for this modification is that all conditions are now expressed directly by g instead of its Radon transform.

Our proof rests upon an observation found in the monograph [48]. If we can find suitable functions g_1, \dots, g_L with finite $\|\cdot\|_{\infty}$ -norm such that the class $\mathfrak{D}(\tau, 1)$ is included in the union of the $\|\cdot\|_{\infty}$ -balls with radius ϵ , i.e.

$$\mathfrak{D}(\tau, 1) \subset U_{\epsilon}^{\|\cdot\|_{\infty}}(g_1) \cup \dots \cup U_{\epsilon}^{\|\cdot\|_{\infty}}(g_L), \tag{B.10}$$

then the maximum-bracketing number of $\mathfrak{D}(\tau, 1)$ for 2ϵ is upper bounded by L . The corresponding brackets are then simply given by $[g_l - \epsilon, g_l + \epsilon]$ for all $l \in \{1, \dots, L\}$. We will thus confine ourselves

to showing that the covering number of $\mathfrak{D}(\tau, 1)$ for some arbitrary but fixed $\epsilon > 0$ is upper bounded by $L = L(\epsilon) \leq \exp(C\epsilon^{-2/(\tilde{\tau}-\tilde{t})})$, where $\tilde{t} := t - 1/2$.

The rest of the proof consists of the construction of such a class of functions, breaking up $\mathfrak{D}(\tau, 1)$ in ϵ -balls and verifying that their number is bounded in the desired way. We begin by relating closeness of Radon coefficients to closeness in $\|\cdot\|_\infty$ -norm.

Invoking Proposition 2.1, we observe that every function $g \in \mathfrak{D}(\tau, 1)$ is identical to its \mathcal{L}^2 -expansion

$$g = \sum_{m=0}^{\infty} \sum_{l=-m}^m \varphi_{(l,m)} \langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}.$$

Because of (B.9) and $\|g\|_\infty \leq 1$ we get for each $g \in \mathfrak{D}(\tau, 1)$

$$|\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \leq \frac{1}{(m+1)^{\tilde{\tau}}} \quad \forall (l,m) \in \mathcal{N}. \tag{B.11}$$

We will now investigate the distance between two functions g, \tilde{g} in $\mathfrak{D}(\tau, 1)$ which have similar Radon coefficients in the sense that

$$|\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - \langle \tilde{g}, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \leq \frac{\epsilon}{C(m+1)^{\tilde{t}}} \quad \forall (l,m) \in \mathcal{N}, \tag{B.12}$$

for some $\epsilon > 0$. For sufficiently large $C > 0$, depending on \tilde{t} only, the maximal distance between g and \tilde{g} can be bounded via

$$\begin{aligned} \|g - \tilde{g}\|_\infty &\leq \sum_{m=0}^{\infty} \sum_{l=-m}^m |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - \langle \tilde{g}, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \|\varphi_{(l,m)}\|_\infty \\ &\leq \sum_{m=0}^{\infty} \sum_{l=-m}^m \sqrt{m+1} |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - \langle \tilde{g}, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \\ &\leq \frac{\epsilon}{C} + \sum_{m=1}^{\infty} (m+1)^{3/2} \frac{\epsilon}{C(m+1)^{\tilde{t}}} \leq \frac{\epsilon}{C} \left(1 + \sum_{m=1}^{\infty} (m+1)^{3/2-\tilde{t}}\right) < \epsilon. \end{aligned} \tag{B.13}$$

In the second inequality we used (A.2) and in the last step $\tilde{t} > 5/2$ in order to guarantee the convergence of the series.

We now construct a covering of $\mathfrak{D}(\tau, 1)$ as in (B.10). First we note that for any fixed but arbitrary function $g \in \mathfrak{D}(\tau, 1)$ the estimate (B.11) already entails

$$|\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - 0| = |\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)}| \leq \frac{\epsilon}{C(m+1)^{\tilde{t}}}$$

for all $m > (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})}$. For $(l,m) \in \mathcal{N}$ with $m \leq \lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil$, (B.11) implies that

$$\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \in [-(m+1)^{-\tilde{\tau}}, (m+1)^{-\tilde{\tau}}] \times [-i(m+1)^{-\tilde{\tau}}, i(m+1)^{-\tilde{\tau}}].$$

We can introduce $\lceil 4C(m+1)^{\tilde{\tau}-\tilde{t}}/\epsilon \rceil^2$ grid points to this cube such that any two of them have maximal distance $\epsilon/(C(m+1)^{\tilde{t}})$. The set of grid points for each cube will be called $G_{(l,m)}$. It then follows for any $(l,m) \in \mathcal{N}$ with $m \leq \lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil$ that there exists an $a_{(l,m)} \in G_{(l,m)}$ such that

$$|\langle g, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - a_{(l,m)}| \leq \frac{\epsilon}{C(m+1)^{\tilde{t}}}.$$

Setting $a_{l,m} = 0$ for $m - l$ odd, we define the function

$$\tilde{g} = \sum_{m=0}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \sum_{l=-m}^m \varphi_{(l,m)} a_{(l,m)},$$

which by construction fulfills (B.12) and hence by (B.13) complies to the inequality

$$\|g - \tilde{g}\|_\infty \leq \varepsilon.$$

Since $g \in \mathfrak{D}(\tau, 1)$ was arbitrary, the collection of functions \tilde{g} that satisfy

$$\langle \tilde{g}, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)} \in G_{(l,m)}$$

for $m \leq \lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil$ and

$$\langle \tilde{g}, \varphi_{(l,m)} \rangle_{\mathcal{L}^2(\mathcal{B}, \mu)} = 0$$

for $m > \lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil$ fulfills (B.10).

The cardinality of this function class consequently upper bounds the covering number of $\mathfrak{D}(\tau, 1)$. To calculate the number of possible \tilde{g} we count all possible combinations of coefficients $a_{(l,m)}$. Let \bullet in the product denote multiplication only with those factors, where $m - l$ is even. Then

$$\begin{aligned} \left| \prod_{m=0}^{\lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \prod_{l=-m}^{\bullet m} G_{(l,m)} \right| &= |G_{(0,0)}| \prod_{m=1}^{\lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \prod_{l=-m}^{\bullet m} |G_{(l,m)}| \\ &\leq \left[\frac{4C}{\varepsilon} \right]^2 \prod_{m=1}^{\lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \left[\left(\frac{4Cm^{\tilde{t}-\tilde{\tau}}}{\varepsilon} \right) \right]^{2(m+1)} \leq \left[\frac{4C}{\varepsilon} \right]^2 \prod_{m=1}^{\lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \left[\left(\frac{4C}{\varepsilon} \right) \right]^{2(m+1)} \\ &= \left[\frac{C}{\varepsilon} \right]^{2 \sum_{m=1}^{\lceil (C/\varepsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil+1} m} \leq \left[\frac{C}{\varepsilon} \right]^{8(C/\varepsilon)^{2/(\tilde{\tau}-\tilde{t})}} \leq \exp \left\{ \log \left(\frac{C}{\varepsilon} \right) \left(\frac{C}{\varepsilon} \right)^{2/(\tilde{\tau}-\tilde{t})} \right\}. \end{aligned} \quad (\text{B.14})$$

To achieve the desired rate we repeat our above argumentation for a shrunk version of t , say $t - \delta$ which is still larger than 3, i.e. with $\tilde{t} - \delta$ still larger than $5/2$. For sufficiently small $\varepsilon > 0$ it follows that

$$\exp \left\{ \log \left(\frac{C}{\varepsilon} \right) \left(\frac{C}{\varepsilon} \right)^{2/(\tilde{\tau}-\tilde{t}+\delta)} \right\} \leq \exp \left\{ \left(\frac{C}{\varepsilon} \right)^{2/(\tilde{\tau}-\tilde{t})} \right\}.$$

By our auxiliary considerations the bracketing number is thus bounded in the desired way. \square

ACKNOWLEDGMENTS

This work has been supported in part by the Collaborative Research Center ‘‘Statistical modeling of nonlinear dynamic processes’’ (SFB 823, Project C1) of the German Research Foundation (DFG) and the Bundesministerium f ur Bildung und Forschung through the project ‘‘MED4D: Dynamic medical imaging: Modeling and analysis of medical data for improved diagnosis, supervision and drug development’’.

REFERENCES

1. M. G. Akritas and I. van Keilegom, ‘‘Nonparametric Estimation of the Residual Distribution’’, *Scandinavian J. Statist.* **28**, 549–567 (2001).
2. M. Bertero, P. Boccacci, G. Desider , and G. Vicidomini, ‘‘Image Deblurring with Poisson Data: From Cells to Galaxies’’, *Inverse Problems* **25** (12), 123006, 26 (2009).
3. P. J. Bickel and M. Rosenblatt, ‘‘On Some Global Measures of the Deviations of Density Function Estimates’’, *Ann. Statist.* **1**, 1071–1095 (1973).
4. N. Bissantz, J. Chown, and H. Dette, ‘‘Regularization Parameter Selection in Indirect Regression by Residual Based Bootstrap’’, *Statist. Sinica* (2018) (in press).
5. N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart, ‘‘Convergence Rates of General Regularization Methods for Statistical Inverse Problems’’, *SIAM J. Numerical Anal.* **45**, 2610–2636 (2007).
6. N. Bissantz and H. Holzmann, ‘‘Asymptotics for Spectral Regularization Estimators in Statistical Inverse Problems’’, *Comput. Statist.* **28**, 435–453 (2013).
7. G. Blanchard and N. M ucke, ‘‘Optimal Rates for Regularization of Statistical Inverse Learning Problems’’, *Foundations Comput. Math.* **18** (4), 971–1013 (2018).

8. M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Oxford, 1970).
9. T. Buzug, *Computed Tomography. From Photon Statistics to Modern Cone-Beam CT* (Springer-Verlag, Berlin–Heidelberg, 2008).
10. R. J. Carroll, A. Delaigle, and P. Hall, “Non-parametric Regression Estimation from Data Contaminated by a Mixture of Berkson and Classical Errors”, *J. Roy. Statist. Soc.: Ser. B, Statist. Methodol.* **69**, 859–878 (2007).
11. L. Cavalier, “Asymptotically Efficient Estimation in a Problem Related to Tomography”, *Math. Methods Statist.* **7**, 445–456 (1999).
12. L. Cavalier, “Efficient Estimation of a Density in a Problem of Tomography”, *Ann. Statist.* **28**, 630–647 (2000).
13. L. Cavalier, “Nonparametric Statistical Inverse Problems”, *Inverse Problems* **24**, 034004, 19 (2008).
14. L. Cavalier and Yu. Golubev, “Risk Hull Method and Regularization by Projections of Ill-posed Inverse Problems”, *Ann. Statist.* **34**, 1653–1677 (2006).
15. L. Cavalier and A. Tsybakov, “Sharp Adaptation for Inverse Problems with Random Noise”, *Probab. Theory and Related Fields* **123** (3), 323–354 (2002).
16. B. Colling and I. van Keilegom, “Goodness-of-Fit Tests in Semiparametric Transformation Models”, *TEST* **25** (2), 291–308 (2016).
17. S. R. Deans, *The Radon Transform and some of its Applications* (Wiley, New York, 1983).
18. A. Delaigle and P. Hall, “Estimation of Observation-Error Variance in Errors-in-Variables Regression”, *Statist. Sinica* **21**, 103–1063 (2011).
19. A. Delaigle, P. Hall, and F. Jamshidi, “Confidence Bands in Non-parametric Errors-in-Variables Regression”, *J. Roy. Statist. Soc.: Ser. B, Statist. Methodology* **77** (2014).
20. H. Dette, N. Neumeyer, and I. van Keilegom, “A New Test for the Parametric Form of the Variance Function in Non-parametric Regression”, *J. Roy. Statist. Soc.: Ser. B, Statist. Methodology* **69** (5), 903–917 (2007).
21. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, in *Mathematics and its Applications*, (Kluwer Academic Publishers Group, Dordrecht, 1996), Vol. 375.
22. S. Helgason, *Integral Geometry and Radon Transforms* (Springer, New York, 2011).
23. T. Hotz, P. Marnitz, R. Stichtenoth, L. Davies, Z. Kabluchko, and A. Munk, “Locally Adaptive Image Denoising by a Statistical Multiresolution Criterion”, *Comput. Statist. Data Anal.* **56**, 543–558 (2012).
24. A. J. E. M. Janssen, “Zernike Expansion of Derivatives and Laplacians of the Zernike Circle Polynomials”, *J. Opt. Soc. Am. A* **31** (7), 1604–1613 (2014).
25. I. M. Johnstone and B. W. Silverman, “Speed of Estimation in Positron Emission Tomography and Related Inverse Problems”, *Ann. Statist.* **18** (1), 251–280 (1990).
26. J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems* (Springer, Berlin, 2010).
27. K. Kato and Y. Sasaki, *Uniform Confidence Bands for Nonparametric Errors-in-Variables Regression*, arXiv:1702.03377v3 (2017).
28. G. Kerkyacharian, E. Le Pennec, and D. Picard, “Radon Needlet Thresholding”, *Bernoulli* **18** (2), 391–433 (2012).
29. G. Kerkyacharian, G. Kyriazis, E. Le Pennec, P. Petrushev, and D. Picard, “Inversion of Noisy Radon Transform by SVD Based Needlets”, *Appl. Comput. Harmonic Anal.* **28** (1), 24–45 (2010).
30. E. V. Khmaladze and H. L. Koul, “Goodness-of-Fit Problem for Errors in Nonparametric Regression: Distribution Free Approach”, *Ann. Statist.* **37** (6A), 3165–3185 (2009).
31. A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction*, in *Lecture Notes in Statistics* (Springer-Verlag, New York, 1993), Vol. 82.
32. H. L. Koul and W. Song, “Regression Model Checking with Berkson Measurement Errors”, *J. Statist. Plann. Inference* **138**, 1615–1628 (2008).
33. H. L. Koul and W. Song, “Minimum Distance Regression Model Checking with Berkson Measurement Errors”, *Ann. Statist.* **37**, 132–156 (2009).
34. V. Lakshminarayanan and A. Fleck, “Zernike Polynomials: A Guide”, *J. Modern Optics* **58** (7), 545–561 (2011).
35. B. A. Mair and F. H. Ruymgaart, “Statistical Inverse Estimation in Hilbert Scales”, *SIAM J. Appl. Math.* **56** (5), 1424–1444 (1996).
36. J. C. Mason and D. Handscomb, *Chebyshev Polynomials* (A CRC Press Company, London, 2002).
37. U. Müller, A. Schick, and W. Wefelmeyer, “Estimating the Error Distribution Function in Semiparametric Additive Regression Models”, *J. Statist. Plann. Inference* **142**, 552–566 (2012).
38. F. Natterer, “Computerized Tomography with Unknown Sources”, *SIAM J. Appl. Math.* **43** (5), 1201–1212 (1983a).
39. F. Natterer, *The Mathematics of Computerized Tomography* (Teubner, Stuttgart; Wiley, Chichester; 1986).

40. F. Natterer and F. Wübbelling, *Mathematical Methods in Image Reconstruction* (SIAM, Philadelphia, 2001).
41. S. H. Nawab, A. S. Willsky, and A. V. Oppenheim, *Signals and Systems* (Prentice Hall, New Jersey, 1996).
42. N. Neumeyer, *Bootstrap Procedures for Empirical Processes of Nonparametric Residuals* (Ruhr-Universität, Bochum, 2006), habilitation thesis.
<https://www.math.uni-hamburg.de/home/neumeyer/habil.ps>
43. N. Neumeyer, “Smooth Residual Bootstrap for Empirical Processes of Non-parametric Regression Residuals”, *Scandinavian J. Statist.* **36** (2), 204–228 (2009).
44. N. Neumeyer and I. van Keilegom, “Estimating the Error Distribution in Nonparametric Multiple Regression with Applications to Model Testing”, *J. Multivariate Anal.* **101** (5), 1067–1078 (2010).
45. J. P. Romano and A. F. Siegel, *Counterexamples in Probability and Statistics* (Chapman and Hall, New York, 1986).
46. S. Saitoh, “Integral Transforms, Reproducing Kernels and their Applications” (Longman, Harlow, 1997).
47. B. W. Silverman, “Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives”, *Ann. Statist.* **6** (1), 177–184 (1978).
48. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes. With Applications to Statistics*, in *Springer Series in Statistics* (Springer, New York, 1996).
49. C. R. Vogel, *Computational Methods for Inverse Problems* (SIAM, Bozeman, 2002).
50. F. Zernike, “Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode”, *Physica* **1**(8), 689–704 (1934).
51. J. Zhang, Z. Feng, and X. Wang, “A Constructive Hypothesis Test for the Single-Index Models with Two Groups”, *Ann. Inst. Statist. Math.* **70** (5), 1077–1114 (2018).