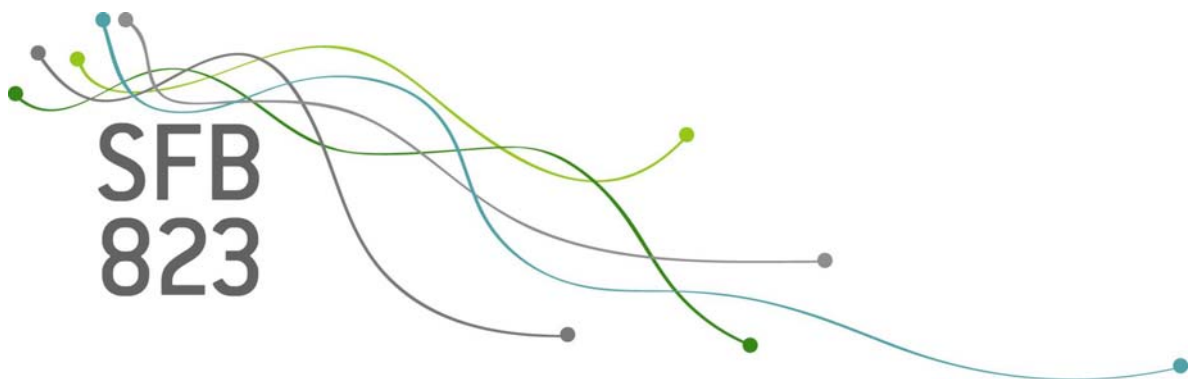# The empirical process of residuals from an inverse regression

Tim Kutta, Nicolai Bissantz,
Justin Chown, Holger Dette

THE EMPIRICAL PROCESS OF RESIDUALS FROM AN INVERSE REGRESSION

TIM KUTTA, NICOLAI BISSANTZ, JUSTIN CHOWN AND HOLGER DETTE

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 cut-off technique. Further, we investigate the empirical process of residuals from this regression, and show that it satsifies a functional central limit theorem.

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 a statistical analysis of data gathered from a single slice. For this purpose we introduce the inverse regression model

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

where $(\varepsilon_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}}$ are independent and 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 $0 \leqslant s \leqslant 1$ as the distance from the origin and $0 \leqslant \phi \leqslant 2\pi$ 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 $0 \leqslant s \leqslant 1$ and $0 \leqslant \phi \leqslant 2\pi$,

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

Details on the underlying physics and applications of CT can be found in Buzug (2008).

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 of Natterer (1986), Engl et al. (1996) and Helgason (2011). Other examples of linear inverse problems are the heat equation and convolution transforms (see Mair and Ruymgaart (1996), Saitoh (1997), and Cavalier (2008), among others). Additional statistical inverse problems include errors-in-variables models and the Berkson error model (see, for example, Bissantz et al. (2007), Carroll et al. (2007), Koul and Song (2008, 2009), Bertero et al. (2009), Kaipio and Somersalo (2010), Delaigle et al. (2014), and Kato and Sasaki (2017)). The Radon transform is usually discussed in the contexts of positron emission tomograpy (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 Johnstone and Silverman (1990), Korostelev and Tsybakov (1993), and Cavalier (2000), among others). On the other hand, CT leads to the inverse regression (1.1) (see, for example, Cavalier (1999) and Kerkyacharian et al. (2010, 2012)).

We contribute to this discussion by deriving the rate of uniform, strong consistency for a nonparametric estimator $\hat{g}$ of the unknown function g based on the popular spectral cutoff method. Further, we derive a functional central limit theorem for the empirical process of the resulting model residuals $\hat{\varepsilon}_{\mathbf{k}}$, i.e. we investigate the estimator

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

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: Bissantz et al. (2018), who study an inverse regression model characterized by a convolution transformation.

In direct regression problems, residual-based empirical processes arising from non- and semi-parametric regression estimators have been considered by numerous authors (see Akritas and van Keilegom (2001), Neumeyer (2009), Müller et al. (2012), Colling and Van Keilegom (2016), and Zhang et al. (2018), among others). Dette et al. (2007) consider tests 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. Neumeyer and Keilegom (2010) work with a similar approach as the previous authors to propose tests for verifying convenient forms of the regression function. Khmaladze and Koul (2009) introduce a popular distribution free approach to addressing goodness-of-fit problems for the errors from a nonparametric regression, where these authors introduce 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 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 Johnstone and Silverman (1990) let

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

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

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

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

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

with corresponding probability measure

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

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 usally 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 below required smoothness of g and $\mathcal{R}g$ entail 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 Johnstone and Silverman (1990) and Born and Wolf (1970). Let

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

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

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

where

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

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

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

where U_m denotes the $m$ths 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,

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

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

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

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 modelling of refraction errors, c.f. Zernike (1934) and more recently Lakshminarayanan and Fleck (2011). We refer to Deans (1983) 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

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

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. In order to specify the required regularity we define

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

the smoothness class. We assume throughout this paper that the regression function g in model (1.1) is an element of $\mathfrak{O}(v)$ (for some $v \geqslant 1$). 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 Mair and Ruymgaart (1996), Cavalier and Tsybakov (2002), Bissantz and Holzmann (2013) or Blanchard and Mücke (2018)).

Proposition 2.1. *Suppose that $g \in \mathfrak{O}(v)$ with $v \geqslant 1$, then the following four identities hold everywhere:*

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

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

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

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

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 Nawab et al. (1996) 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 $v \geqslant 2k + 1$ it also entails periodicity of g and its continuous derivatives in the angular component 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 common in computed tomography we will assume a parallel scanning procedure, corresponding to a grid of design points on the detector space. Adopting our results to fan beam geometry, which underlies most modern scanners, is then mathematically simple.

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

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

and decompose the detector space in rectangular boxes of the form

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

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

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

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 $q \to \infty$ and that $p = p(q) \to \infty$ depends on q as follows:

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

Denoting the number of rows and columns in the grid of design points by q and p respectively 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 Natterer and Wübbelling (2001), p. 74). Sometimes we will use the notation $n \to \infty$, 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 centred 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

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

Here

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

is an estimator of the inner product

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

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. In the literature it is common 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 Cavalier (2008)).

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 of Vogel (2002), where multiple techniques are gathered. More closely related to our case is the risk hull method by Cavalier and Golubev (2006) in the white noise model and the smooth bootstrap examined by Bissantz et al. (2018) in a different context.

Remark 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 $\Lambda : \mathbb{R} \to [0,1]$ denote a function with compact support and define

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

as an alternative estimator of g. Note that the estimate $\hat{g}$ in (2.15) is obtained for $\Lambda(x) = \mathbb{1}_{[0,1]}(x)$. All results presented in this paper remain valid for the estimator (2.18). However, for 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 residual process

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

where $\mathbb{F}$ denotes the residual distribution function and

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

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 $g \in \mathfrak{O}(v)$ for some $v \geq 5$. Then the estimator $\hat{g}$ in (2.15) satisfies*

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

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

Next we derive a uniform bound for the random error of the estimator $\hat{g}$. (1.1).

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*

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

Balancing the two upper bounds for the deterministic and random part yields an optimal choice of the bandwidth. More precisely for $v \geqslant 5$ the choice

$$(3.2) \qquad\qquad t_n := \Theta\left(\left(\log(n)^{-1} n \right)^{\frac{1}{2(v+3)}} \right)$$

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 $g \in \mathfrak{O}(v)$ for some $v \geqslant 5$ and that $\mathbb{E}|\varepsilon|^\kappa < \infty$ for some $\kappa > 3$. Additionally let t_n be chosen as in (3.2). Then*

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

and for all $\tau \leqslant v$

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

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 $v \geqslant 2k + 1$ for some $k \in \mathbb{N}_0$. Additionally let $\alpha, \beta \in \mathbb{N}_0$, such that $\alpha + \beta = k$. Then*

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

In order to prove the weak convergence of the process $\sqrt{n}(\hat{\mathbb{F}}_n - \mathbb{F})$ we consider the *bracketing metric entropy* of the subclass

$$(3.6) \quad \mathfrak{O}(\tau, 1, 1) := \left\{ g : \mathcal{B} \to \mathbb{R} \middle| g \ continuous, \|g\|_\infty \leqslant 1, \sum_{m=0}^{\infty} \sum_{l=-m}^{m} |R(l,m)| \, (m+1)^\tau \leqslant 1, \right\},$$

for some $\tau > 0$. Theorem 3.3 implies that for all $\tau < v$ the difference $\hat{g} - g$ eventually lies in $\mathfrak{O}(\tau, 1, 1)$. As we know from Proposition 2.1 the condition

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

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 these considerations to derive the following result about the complexity of the class $\mathfrak{O}(\tau, 1, 1)$, which is of independent interest and is proven in Appendix B (see section B.4).

Proposition 3.5. *Let $\tau > 3$, then for any $t \in (3, \tau)$ and sufficiently small $\epsilon > 0$*

$$(3.7) \qquad \log\left(N_{[]}(\epsilon, \mathfrak{O}(\tau, 1, 1), \|\cdot\|_\infty)\right) \leqslant C \left(\frac{1}{\epsilon}\right)^{\frac{2}{\tau - t}}.$$

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

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 3.6. *Assume that $g \in \mathfrak{O}(v)$ for some $v > 5$, $\mathbb{E}|\varepsilon|^\kappa < \infty$ 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*

$$(3.8) \qquad \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \left[\mathbb{1}\{\hat\varepsilon_{\mathbf{k}} \leqslant t\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t\} - \varepsilon_{\mathbf{k}} f_\varepsilon(t) \right] \right| = o_P\left(n^{-1/2}\right).$$

Corollary 3.7. *Under the assumptions of Theorem 3.6, the process*

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

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

$$\Sigma(t, \tilde{t}) \;:=\; \frac{8\pi^2}{3} \Big(\mathbb{F}(\min(t, \tilde{t})) - \mathbb{F}(t)\mathbb{F}(\tilde{t}) + f_\varepsilon(t)\mathbb{E}\left[\varepsilon \mathbb{1}\{\varepsilon \leqslant \tilde{t}\} \right]$$

$$+\; f_\varepsilon(\tilde{t})\mathbb{E}\left[\varepsilon \mathbb{1}\{\varepsilon \leqslant t\} \right] + \sigma^2 f_\varepsilon(t) f_\varepsilon(\tilde{t}) \Big), \;\; t, \tilde{t} \in \mathbb{R}.$$

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 $g \in \mathfrak{O}(v)$ for $v \geqslant 5$ and that Assumption 2.2 holds. Then for all $(l, m) \in \mathcal{N}$ it follows that*

$$(A.1) \qquad |\mathbb{E}\hat{R}(l, m) - R(l, m)| \leqslant Cm^5 n^{-1}$$

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 accounts for the average approximation error of Radon coefficients with index m smaller than t_n and the other 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

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

where the terms A_1 and A_2 are given by

$$A_1 := \sum_{m=0}^{t_n} \sum_{l=-m}^{m} \sqrt{m+1} \, \|\varphi_{(l,m)}\|_\infty \, \left| \mathbb{E}\hat{R}(l, m) - R(l, m) \right|$$

$$A_2 := \sum_{m>t_n} \sum_{l=-m}^{m} \sqrt{m+1} \, \|\varphi_{(l,m)}\|_\infty \, |R(l, m)| \, .$$

For the term A_1 it follows that

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

where we have used that Proposition B.2 in Appendix B implies the estimate

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

in the first and the approximation result from Lemma A.1 in the second inequality. Similarly we have

$$A_2 \quad \leqslant \quad \sum_{m>t_n} \sum_{l=-m}^{m} (m+1)|R(l, m)| \leqslant \sum_{m>t_n} \sum_{l=-m}^{m} (m+1)^v t_n^{1-v}|R(l, m)|$$

$$\leqslant \quad t_n^{1-v} \sum_{m=0}^{\infty} \sum_{l=-m}^{m} (m+1)^v |R(l, m)| = \mathcal{O}\left(t_n^{1-v} \right).$$

In the last step we have used that g complies to the smoothness condition of $\mathfrak{O}(v)$ (see (2.9)) and thus the series converges. $\qquad\square$

A.2. **Proof of Lemma 3.2.** We first rewrite $\hat{g} - \mathbb{E}\hat{g}$ employing (2.16) and (A.2)

$$
\begin{aligned}
\left\| \hat{g}(z) - \mathbb{E}\hat{g}(z) \right\|_\infty &= \left\| \sum_{m=0}^{t_n} \sum_{l=-m}^{m} \sqrt{m+1}\varphi_{(l,m)} \left(\hat{R}(l,m) - \mathbb{E}\hat{R}(l,m) \right) \right\|_\infty \\
&= \left\| \sum_{m=0}^{t_n} \sum_{l=-m}^{m} \sqrt{m+1}\varphi_{(l,m)} \left(\sum_{\mathbf{k}\in\mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) w_{\mathbf{k}}\varepsilon_{\mathbf{k}} \right) \right\|_\infty \\
&\leqslant \sum_{m=0}^{t_n} \sum_{l=-m}^{m} (m+1) \left| \sum_{\mathbf{k}\in\mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) w_{\mathbf{k}}\varepsilon_{\mathbf{k}} \right| \\
&= \sum_{m=0}^{t_n} \sum_{l=-m}^{m} (m+1)^2 \left| \sum_{\mathbf{k}\in\mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}}\varepsilon_{\mathbf{k}} \right| \\
&\leqslant \sum_{m=0}^{t_n} 2(m+1)^3 \max_{\substack{(l,m)\in\mathcal{N} \\ m\leqslant t_n}} \left| \sum_{\mathbf{k}\in\mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}}\varepsilon_{\mathbf{k}} \right| \\
&\leqslant C t_n^4 \max_{\substack{(l,m)\in\mathcal{N} \\ m\leqslant t_n}} \left| \sum_{\mathbf{k}\in\mathbf{K}} \overline{\psi_{(l,m)}}(z_{\mathbf{k}}) (m+1)^{-1} w_{\mathbf{k}}\varepsilon_{\mathbf{k}} \right|
\end{aligned}
$$

We proceed 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

$$
\text{(A.3)} \qquad\qquad \varepsilon_{\mathbf{k}}^{d_n} := \mathbb{1}\left\{ |\varepsilon_{\mathbf{k}}| \leqslant d_n \right\} \varepsilon_{\mathbf{k}}.
$$

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

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

and therewith it follows that

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

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

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

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

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

where B_1 and B_2 are defined by

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

$$B_2 := \sum_{\mathbf{k}\in\mathbf{K}} \left|\overline{\psi_{(l,m)}}(z_{\mathbf{k}})\right|\,(m+1)^{-1}w_{\mathbf{k}}\left|\mathbb{E}\varepsilon_{\mathbf{k}}^{d_n}\right|.$$

Using the inequality

$$(A.5) \qquad \|\psi_{(l,m)}\|_\infty \leqslant m+1,$$

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

$$B_2 \leqslant \left|\mathbb{E}\varepsilon^{d_n}\right| \sum_{\mathbf{k}\in\mathbf{K}} w_{\mathbf{k}} = \mathcal{O}(n^{-1/2}),$$

wherewe exploit the decay rate $\left|\mathbb{E}\varepsilon^{d_n}\right| = \mathcal{O}(n^{-1/2})$ 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

$$\left|\mathbb{E}\varepsilon^{d_n}\right| = \left|\mathbb{E}\left[\varepsilon\mathbb{1}\{|\varepsilon| > d_n\}\right]\right| \leqslant \int_{d_n}^\infty \mathbb{P}\left(|\varepsilon| > s\right)ds \leqslant \mathbb{E}\left[|\varepsilon|^\kappa\right](\kappa-1)^{-1}d_n^{1-\kappa} = \mathcal{O}\left(n^{-1/2}\right).$$

For the term B_1 we note that for a fixed constant $C^\star$

$$(A.6) \qquad \mathbb{P}\left(|B_1| > \log(n)^{1/2}n^{-1/2}C^\star\right)$$

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

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

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

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

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

which is summable for sufficiently large $C^\star$. The Borel-Cantelli Lemma therefore implies that

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

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. $\qquad\square$

A.3. **Proof of Theorem 3.3.** Combining Lemma 3.1 and Lemma 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

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

which gives for the left hand side of (3.4)

$$(A.7) \quad \sum_{m=0}^{\infty}\sum_{l=-m}^{m} m^{\tau}\left|\left\langle \mathcal{R}\left[g - \hat{g}\right], \psi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{D},\lambda)}\right| = \sum_{m=0}^{\infty}\sum_{l=-m}^{m} m^{\tau}\left|R(l,m) - \hat{R}(l,m)\mathbb{1}\{m \leqslant t_n\}\right|$$

$$\leqslant \quad D_1 + D_2 + D_3.$$

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

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

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

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

By Proposition A.1 we receive the upper bound

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

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

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

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

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

$$\max_{\substack{(l,m)\in\mathcal{N} \\ m\leqslant t_n}}\left\{(m+1)^{-1}\left|\hat{R}(l,m) - \mathbb{E}\hat{R}(l,m)\right|\right\}$$

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

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

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

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

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

and the definition of t_n in (3.2) yield the desired result. $\square$

A.4. **Proof of Theorem 3.6 .** Proposition 3.5 is used to verify an equicontinuity argument, which is the central building block in the proof of Theorem 3.6. For this purpose we define the L_n^2-bracketing number as follows:

Definition A.2 Let $Z_{1,n},...,Z_{n,n}$ be stochastic processes, indexed in $\mathcal{F}$, and $\epsilon > 0$. The L_2^n-bracketing number of $\mathcal{F}$, denoted by $N_{[]}^{L_2^n}(\epsilon,\mathcal{F})$, is the minimal number N_ϵ of sets $\mathcal{F}_{\epsilon,j}^n$ in a partition of $\mathcal{F} = \bigcup_{j=1}^{N_\epsilon} \mathcal{F}_{\epsilon,j}^n$ such that for each j

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

Lemma A.3. *Define*

$$\text{(A.9)} \qquad M_n(t) := \left| \sum_{\mathbf{k}\in\mathbf{K}} w_\mathbf{k}\left\{ \mathbb{1}\{\hat{\varepsilon}_\mathbf{k}\leqslant t\} - \mathbb{1}\{\varepsilon_\mathbf{k}\leqslant t\} + \mathbb{F}(t) - \mathbb{P}\left(\hat{\varepsilon}_\mathbf{k}\leqslant t\right)\right\}\right|.$$

Then, under the assumptions of Theorem 3.6 it follows that $\sup_{t\in\mathbb{R}}|M_N(t)| = o_P(n^{-1/2})$.

Proof of Lemma A.1. Using the definition of the estimated residuals in (3.1) we have

$$\mathbb{1}\{\hat{\varepsilon}_\mathbf{k}\leqslant t\} - \mathbb{1}\{\varepsilon_\mathbf{k}\leqslant t\} + \mathbb{F}(t) - \mathbb{P}\left(\hat{\varepsilon}_\mathbf{k}\leqslant t\right)$$

$$= \mathbb{1}\left\{\varepsilon_\mathbf{k}\leqslant t + \mathcal{R}[\hat{g}-g](z_\mathbf{k})\right\} - \mathbb{1}\{\varepsilon_\mathbf{k}\leqslant t\} + \mathbb{P}\left(\varepsilon\leqslant t\right) - \mathbb{P}\left(\varepsilon_\mathbf{k}\leqslant t + \mathcal{R}[\hat{g}-g](z_\mathbf{k})\right).$$

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

$$Z_{n,\mathbf{k}}(t,d) := n^{1/2}w_\mathbf{k}\left(\mathbb{1}\left\{\varepsilon_\mathbf{k}\leqslant t + d(z_\mathbf{k})\right\} - \mathbb{1}\{\varepsilon_\mathbf{k}\leqslant t\}\right),$$

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

$$(A.10) \qquad \rho((t, d), (\tilde{t}, \tilde{d})) := \max \left\{ \sup_{x \in [-1, 1]} \left\{ |\mathbb{F}(t + x) - \mathbb{F}(\tilde{t} + x)|, \|d - \tilde{d}\|_\infty \right\} \right\}.$$

Notice that for ρ to be a semimetric 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 $\epsilon > 0$

$$(A.11) \quad \mathbb{P}\left(\sup_{\substack{(t, d), (\tilde{t}, \tilde{d}) \in \mathcal{F} \\ \rho((t, d), (\tilde{t}, \tilde{d})) < \delta_n}} \left| \sum_{\mathbf{k} \in \mathbf{K}} \left(Z_{n, \mathbf{k}}(t, d) - \mathbb{E}Z_{n, \mathbf{k}}(t, d) - Z_{n, \mathbf{k}}(\tilde{t}, \tilde{d}) + \mathbb{E}Z_{n, \mathbf{k}}(\tilde{t}, \tilde{d}) \right) \right| > \epsilon \right) \to 0.$$

If (A.11) holds, then the assertion of Lemma A.3 can be shown as follows: Firstly note that we can derive a lower bound for the probability on the left hand side of (A.11) by

$$(A.12) \quad \mathbb{P}\left(\left\{ \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} Z_{n, \mathbf{k}}(t, d_n) - \mathbb{E}Z_{n, \mathbf{k}}(t, d_n) - Z_{n, \mathbf{k}}(t, 0) + \mathbb{E}Z_{n, \mathbf{k}}(t, 0) \right| > \epsilon \right\} \right.$$

$$\bigcap \left\{ d_n \in \mathcal{R}(\mathfrak{O}(\tau, 1, 1)) \right\} \bigcap \left\{ \sup_{t \in \mathbb{R}} \rho\left((t, 0), (t, d_n) \right) < \delta_n \right\} \right)$$

$$= \quad \mathbb{P}\left(\left\{ \sup_{t \in \mathbb{R}} |n^{1/2} M_n(t)| > \epsilon \right\} \bigcap \left\{ d_n \in \mathcal{R}(\mathfrak{O}(\tau, 1, 1)) \right\} \bigcap \left\{ \sup_{t \in \mathbb{R}} \rho((t, 0), (t, d_n)) < \delta_n \right\} \right).$$

By the second part of Theorem 3.3 we know that

$$\mathbb{P}\left(d_n \in \mathcal{R}(\mathfrak{O}(\tau, 1, 1)) \right) \to 1,$$

for $\tau < v$. Furthermore we notice that $\|\mathcal{R}d\|_\infty \leqslant \|d\|_\infty$ for all 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

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

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

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

Combining these considerations with the right side of (A.12) yields that $n^{1/2} M_n(t) = o_P(1)$ uniformly in t, proving the Lemma provided that (A.11) holds.

This statement is a consequence of Lemma A.19 in Neumeyer (2006) , 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:

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

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

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

$$\text{(A.13)} \qquad \sup_{\rho((t,d),(\tilde{t},\tilde{d})) < \delta_n} \Big| \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E}\Big[\big(Z_{n,\mathbf{k}}(t,d) - Z_{n,\mathbf{k}}(\tilde{t},\tilde{d}) \big)^2 \Big] \Big| \to 0.$$

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

$$\mathbb{E}\left[\big(Z_{n,\mathbf{k}}(t,d) - Z_{n,\mathbf{k}}(\tilde{t},\tilde{d}) \big)^2 \right]$$

$$\leqslant \ C n^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + |\mathbb{F}(t) - \mathbb{F}(\tilde{t})| \right]$$

$$\leqslant \ C n^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + \delta_n \right]$$

$$\leqslant \ C n^{-1} \left[|\mathbb{F}(t + d(z_{\mathbf{k}})) - \mathbb{F}(t + \tilde{d}(z_{\mathbf{k}}))| + |\mathbb{F}(t + \tilde{d}(z_{\mathbf{k}})) - \mathbb{F}(\tilde{t} + \tilde{d}(z_{\mathbf{k}}))| + \delta_n \right].$$

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:

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

The middle term is bounded by δ_n by definition of our semimetric ρ in (A.10), when we consider that $\tilde{d} \in \mathcal{R}(\mathfrak{O}(\tau, 1, 1))$ and therefore $\|\tilde{d}\|_\infty \leqslant 1$. 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, by $N_{[]}^{L_2^n}(\epsilon, \mathcal{F})$, the condition we have to check next is, that for every sequence $\delta_n \downarrow 0$:

$$\text{(A.14)} \qquad \int_0^{\delta_n} \sqrt{\log\big(N_{[]}^{L_2^n}(\epsilon, \mathcal{F}) \big)} \, d\epsilon \to 0.$$

For the construction of an adequate partition of $\mathcal{F}$ satisfying (A.8), consider the ϵ^2-brackets $[g_j^L, g_j^U]$, $j = 1, ..., J = \mathcal{O}(\exp(\epsilon^{4/(\tau-t)}))$ of $\mathfrak{O}(\tau, 1, 1)$, where $t > 3$ such that $\tau - t > 2$ (note that $\tau > 5$ by assumption). 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 receive ϵ^2-brackets $[\mathcal{R}g_j^L, \mathcal{R}g_j^U]$ of the whole class $\mathcal{R}(\mathfrak{O}(\tau, 1, 1))$.

Additionally choose $y_i^L < y_i^U$ with $i = 1, ..., I = \mathcal{O}(\epsilon^{-2})$, 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 $\leqslant \epsilon^2$. Then the sets

(A.15)
$$\mathcal{F}_{i,j,\epsilon}^n := [y_i^L, y_i^U] \times [\mathcal{R}g_j^L, \mathcal{R}g_j^U]$$

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 such 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 $\mathcal{F}_{i,j,\epsilon}^n$

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

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

$$Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E}\left[\sup_{(t,d)(\tilde{t},\tilde{d}) \in \mathcal{F}_{i,j,\epsilon}^n} \left(\left|\mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t + d(z_{\mathbf{k}})\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant \tilde{t} + \tilde{d}(z_{\mathbf{k}})\}\right|\right.\right.$$

$$\left.\left. + \left|\mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant \tilde{t}\}\right|\right)^2\right]$$

$$\leqslant Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \mathbb{E}\left[\left|\mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}})\}\right|\right.$$

$$\left. + \left|\mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant y_i^U\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant y_i^L\}\right|\right]$$

$$\leqslant Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \left[\left|\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))\right|\right.$$

$$\left. + C\left|\mathbb{F}\left(y_i^U\right) - \mathbb{F}\left(y_i^L\right)\right|\right]$$

$$\leqslant Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \left|\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))\right| + C\epsilon^2$$

$$\leqslant Cn^{-1} \sum_{\mathbf{k} \in \mathbf{K}} \{\left|\mathbb{F}(y_i^U + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^U(z_{\mathbf{k}}))\right|$$

$$+ \left|\mathbb{F}(y_i^L + \mathcal{R}g_j^U(z_{\mathbf{k}})) - \mathbb{F}(y_i^L + \mathcal{R}g_j^L(z_{\mathbf{k}}))\right|\} + C\epsilon^2$$

$$\leqslant C\left(\epsilon^2 + \|f_\varepsilon(y)\|_\infty \|\mathcal{R}g_j^U - \mathcal{R}g_j^L\|_\infty + \epsilon^2\right) \leqslant C\epsilon^2.$$

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.14) converges since $\tau - t > 2$.

(4) Finally we have to prove that $(\mathcal{F}, \rho)$ is totally bounded. By definition ρ is a maximum semimetric defined on the product space $\mathbb{R} \times \mathcal{R}(\mathfrak{O}(\tau, 1, 1))$. Hence it suffices to show that each of the spaces $(\mathbb{R}, \rho_1)$, $(\mathcal{R}(\mathfrak{O}(\tau, 1, 1)), \rho_2)$ is totally bounded, where we define for $t, \tilde{t} \in \mathbb{R}$ and $d_1, d_2 \in \mathcal{R}(\mathfrak{O}(\tau, 1, 1))$

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

$$\rho_2(d_1, d_2) \quad := \quad \|d_1 - d_2\|_\infty.$$

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

(A.16)
$$\sup_{x \in [-1,1]} |\mathbb{F}(t + x) - \mathbb{F}(t_j + x)| \leqslant \epsilon.$$

Let $M := \max_{t \in \mathbb{R}} |f_\varepsilon(t)|$ 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 $j = 1, ..., J$ 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:

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

For ρ_2 we recall that by our above observations for every $\epsilon > 0$ the bracketing number of $\mathcal{R}(\mathfrak{O}(\tau, 1, 1))$ 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 (2006) equicontinuity follows which completes the proof of Lemma A.3.

$\square$

Besides Lemma A.3 we require some additional approximation results for a proof of Theorem 3.6.

Proposition A.4. *Under the assumptions of Theorem 3.6 we have*

(A.17)
$$\sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \mathbb{P}(\hat{\varepsilon}_{\mathbf{k}} \leqslant t) - \int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) \, d\lambda(z) = o_P(n^{-1/2})$$

(A.18)
$$\int_{\mathcal{D}} \mathbb{F}(t + \mathcal{R}[\hat{g} - g](z)) d\lambda(z) - \mathbb{F}(t) - f_\varepsilon(t) \int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) = o_P(n^{-1/2})$$

(A.19)
$$\int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) - \sum_{\mathbf{k} \in \mathbf{K}} \varepsilon_{\mathbf{k}} w_{\mathbf{k}} = o_P(n^{-1/2})$$

Proof of Proposition A.4. Recalling the definitions of the estimated residuals $\hat{\varepsilon}_{\mathbf{k}} := \varepsilon_{\mathbf{k}} - \mathcal{R}[\hat{g} - g](z_{\mathbf{k}})$ and the weights $w_{\mathbf{k}} := \lambda(B_{\mathbf{k}})$, we begin by rewriting the left side of (A.17)

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

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

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

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

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

By Assumption 2.2 $\|z - z_{\mathbf{k}}\|_2 = \mathcal{O}(1/\sqrt{n})$. Moreover by Corollary 3.7 the gradient $\nabla \mathcal{R}[\hat{g} - g]$ converges uniformly to 0. Thus we get the desired result. The estimate (A.18) follows by similar arguments, while (A.19) is based on two observations. Firstly, since $\psi_{(0,0)} = 1$ we can rewrite the integral

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

Secondly as the errors are centered

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

Combining these results yields the representation $\mathbb{E}[\hat{R}(0,0)] - R(0,0)$ for the left side of (A.19). By Proposition A.1 this difference is of order $\mathcal{O}(n^{-1})$.

$\square$

Equipped with our observations in Lemma A.3 and Proposition A.4 Theorem 3.6 is easily deduced:

Proof of Theorem 3.6 We apply the triangular inequality to arrive at the following decomposition

$$\sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \left[\mathbb{1}\{\hat{\varepsilon}_{\mathbf{k}} \leqslant t\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t\} - \varepsilon_{\mathbf{k}} f_{\varepsilon}(t) \right] \right|$$

$$\leqslant \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \left[\mathbb{1}\{\hat{\varepsilon}_{\mathbf{k}} \leqslant t\} - \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t\} + \mathbb{F}(t) - \mathbb{P}\left(\hat{\varepsilon}_{\mathbf{k}} \leqslant t\right) \right] \right|$$

$$+ \quad \sup_{t \in \mathbb{R}} \left| \sum_{\mathbf{k} \in \mathbf{K}} w_{\mathbf{k}} \mathbb{P}\left(\hat{\varepsilon}_{\mathbf{k}} \leqslant t\right) - \int_{\mathcal{D}} \mathbb{F}\left(t + \mathcal{R}[\hat{g} - g](z)\right) d\lambda(z) \right|$$

$$+ \quad \sup_{t \in \mathbb{R}} \left| \int_{\mathcal{D}} \mathbb{F}\left(t + \mathcal{R}[\hat{g} - g](z)\right) d\lambda(z) - \mathbb{F}(t) - f_{\varepsilon}(t) \int_{\mathcal{D}} \mathcal{R}[\hat{g} - g](z) d\lambda(z) \right|$$

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

Each of the terms on the right side is of order $o_P\left(n^{-1/2}\right)$, the first one by Lemma A.3 and the other ones by Proposition A.4. $\qquad\square$

A.5. **Proof of Corollary 3.7.** This is a consequence of Theorem 3.6, as we can represent the process $\sqrt{n}(\mathbb{F} - \hat{\mathbb{F}}_n)$ as sum of independent stochastic processes and a negligible term:

$$\sqrt{n}(\mathbb{F} - \hat{\mathbb{F}}_n) \quad = \quad \sum_{\mathbf{k} \in \mathbf{K}} \sqrt{n} w_{\mathbf{k}} \left\{ \mathbb{1}\{\hat{\varepsilon}_{\mathbf{k}} \leqslant t\} - \mathbb{F}(t) \right\}$$

$$= \quad \sum_{\mathbf{k} \in \mathbf{K}} \sqrt{n} w_{\mathbf{k}} \left\{ \mathbb{1}\{\varepsilon_{\mathbf{k}} \leqslant t\} - \mathbb{F}(t) + \varepsilon_{\mathbf{k}} f_{\varepsilon}(t) \right\} + o_P(1).$$

The sum on the right side converges to a Gaussian process, by application of a functional CLT for triangular arrays found in Neumeyer (2006).

APPENDIX B. AUXILIARY RESULTS

B.1. **Uniform bounds.** We begin stating some frequently used properties of the radial polynomials which are taken from Born and Wolf (1970) and Janssen (2014).

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

(B.1)
$$\sup_{0 \leqslant r \leqslant 1} |R_m^{|l|}(r)| = 1.$$

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

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

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

(B.3)
$$\frac{d}{dr} R_m^{|l|}(r) \quad = \quad \sum_{j=0}^{\frac{1}{2}(m-1-|l-1|)} (m - 2j) R_{m-1-2j}^{|l-1|}(r)$$

$$+ \quad \sum_{j=0}^{\frac{1}{2}(m-1-|l+1|)} (m - 2j) R_{m-1-2j}^{|l+1|}(r).$$

Next we provide upper bounds on the $\|\cdot\|_\infty$-norm of the derivatives of the Chebychev and radial polynomials. The bounds on the radial polynomials follow by the above Proposition and the bounds for the Chebychevs by identities from Mason and Handscomb (2002).

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

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

and

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

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

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

By virtue of (B.3) we have

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

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

The case of the Chebychev polynomials is similar. In order to prove the second identity in Proposition B.2 we cite a few well known facts about Chebychev polynomials from Mason and Handscomb (2002)

(1) For all $m \in \mathbb{N}$ U_m is uniformly bounded by $m+1$.
(2) Let T_m denote the Chebychev 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 an induction, analogous to that of the first part. $\qquad\square$

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 the identities (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

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

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 the continuity of $\tilde{g}$ (recall that g is continuous by assumption). Continuity of $\tilde{g}$ is implied by the uniform convergence of the sequence of continuous functions

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

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

$$\left\| \tilde{g} - \sum_{m=0}^{N} \sum_{l=-m}^{m} \varphi_{(l,m)} \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right\|_{\infty} = \left\| \sum_{m=N}^{\infty} \sum_{l=-m}^{m} \varphi_{(l,m)} \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right\|_{\infty}$$

$$\leqslant \sum_{m=N}^{\infty} \sum_{l=-m}^{m} \left\| \varphi_{(l,m)} \right\|_{\infty} \left| \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right| \leqslant \sum_{m=N}^{\infty} \sum_{l=-m}^{m} \sqrt{m+1} \left| \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right|$$

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

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

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:

$$\left| R(l, m) - \mathbb{E}\hat{R}(l, m) \right| \leqslant \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \overline{\psi_{(l,m)}}(z)\mathcal{R}g(z) - \overline{\psi_{(l,m)}}(z_{\mathbf{k}})\mathcal{R}g(z_{\mathbf{k}})d\lambda(z) \right|.$$

$$\leqslant \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \mathfrak{Re}(\psi_{(l,m)}(z))\mathcal{R}g(z) - \mathfrak{Re}(\psi_{(l,m)}(z_{\mathbf{k}}))\mathcal{R}g(z_{\mathbf{k}})d\lambda(z) \right|$$

$$+ \sum_{\mathbf{k} \in \mathbf{K}} \left| \int_{B_{\mathbf{k}}} \mathfrak{Im}(\psi_{(l,m)}(z))\mathcal{R}g(z) - \mathfrak{Im}(\psi_{(l,m)}(z_{\mathbf{k}}))\mathcal{R}g(z_{\mathbf{k}})d\lambda(z) \right|.$$

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

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

and the imaginary part

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

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 $\mathbf{k} = (k_1, k_2)$

$$R(\mathbf{k}) \quad := \quad \int_{B_{\mathbf{k}}} \mathfrak{Re}(\psi_{(l,m)}(z))\mathcal{R}g(z) - \mathfrak{Re}(\psi_{(l,m)}(z_{\mathbf{k}}))\mathcal{R}g(z_{\mathbf{k}})d\lambda(z)$$

$$= \quad \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (s - z_{k_1})\frac{d}{ds}\left[\mathfrak{Re}(\psi_{(l,m)})\mathcal{R}g\right](z_{\mathbf{k}})2\pi^{-1}\sqrt{1 - s^2}d\phi ds$$

$$+ \quad \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (\phi - z_{k_2})\frac{d}{d\phi}\left[\mathfrak{Re}(\psi_{(l,m)})\mathcal{R}g\right](z_{\mathbf{k}})2\pi^{-1}\sqrt{1 - s^2}d\phi ds$$

$$+ \quad \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} (s - z_{k_1})(\phi - z_{k_2})\frac{d}{d\phi}\frac{d}{ds}\left[\mathfrak{Re}(\psi_{(l,m)})\mathcal{R}g\right](\xi_1)2\pi^{-1}\sqrt{1 - s^2}d\phi ds$$

$$+ \quad \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} 2^{-1}(s - z_{k_1})^2\frac{d^2}{ds^2}\left[\mathfrak{Re}(\psi_{(l,m)})\mathcal{R}g\right](\xi_2)2\pi^{-1}\sqrt{1 - s^2}d\phi ds$$

$$+ \int_{k_1/d}^{(k_1+1)/d} \int_{2\pi k_2/n}^{2\pi(k_2+1)/n} 2^{-1}(\phi - z_{k_2})^2 \frac{d^2}{d\phi^2} \left[\mathfrak{Re}(\psi_{(l,m)}) \mathcal{R}g \right] (\xi_3) 2\pi^{-1}\sqrt{1-s^2} d\phi ds.$$

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 $\mathfrak{Re}(\psi_{(l,m)})$ 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. $\qquad\square$

B.4. **Proof of Proposition 3.5.** We begin by rewriting the series condition (3.6) as

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

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 van der Vaart and Wellner (1996). If we can find suitable functions $g_1, ..., g_L$ with finite $\|\cdot\|_\infty$-norm, such that the class $\mathfrak{O}(\tau, 1, 1)$ is included in the union of the $\|\cdot\|_\infty$-balls with radius ϵ, i.e.

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

then the $\|\cdot\|_\infty$-bracketing number of $\mathfrak{O}(\tau, 1, 1)$ for 2ϵ is upper bounded by L. The corresponding brackets are then simply given by $[g_l - \epsilon, g_l + \epsilon]$ for all $l \in \{1, ..., L\}$. We will thus confine ourselves to showing that the covering number of $\mathfrak{O}(\tau, 1, 1)$ for some arbitrary but fixed $\epsilon > 0$ is upper bounded by $L = L(\epsilon) \leqslant \exp(C\epsilon^{-2/(\tilde\tau - \tilde t)})$, where $\tilde t := t - 1/2$.

The rest of the proof consists of the construction of such a class of functions, breaking up $\mathfrak{O}(\tau, 1, 1)$ 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 $g \in \mathfrak{O}(\tau, 1, 1)$ is identical to its $\mathcal{L}^2$-expansion

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

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

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

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

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

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

$$\|g - \tilde{g}\|_{\infty} \leqslant \sum_{m=0}^{\infty} \sum_{l=-m}^{m} \left| \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - \left\langle \tilde{g}, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right| \|\varphi_{(l,m)}\|_{\infty}$$

$$\leqslant \sum_{m=0}^{\infty} \sum_{l=-m}^{m} \sqrt{m+1} \left| \left\langle g, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} - \left\langle \tilde{g}, \varphi_{(l,m)} \right\rangle_{\mathcal{L}^2(\mathcal{B},\mu)} \right|$$

$$\leqslant \frac{\epsilon}{C} + \sum_{m=1}^{\infty} (m+1)^{3/2} \frac{\epsilon}{C(m+1)^{\tilde{t}}} \leqslant \frac{\epsilon}{C}\left(1 + \sum_{m=1}^{\infty} (m+1)^{3/2-\tilde{t}}\right) < \epsilon.$$

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. It is notable that the estimate (B.11) already implies

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

for all $m \geqslant (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})}$ i.e. substantially different coefficients can only occur for smaller m.

Now let us consider those coefficients with $m \leqslant \lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil$. In order to construct the desired functions for a covering of $\mathfrak{O}(\tau, 1, 1)$ as in (B.10), we decompose the domains of possible Radon coefficients in the following way: For each $(l,m) \in \mathcal{N}$, the estimate (B.11) implies that

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

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 that for each function g in $\mathfrak{O}(\tau, 1, 1)$ we can find a vector of coefficients

$$\mathbf{a} := \left(a_{(l,m)}\right) \in \underset{m=0}{\overset{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil}{\times}} \underset{l=-m}{\overset{\bullet m}{\times}} G_{(l,m)},$$

($\bullet$ denotes multiplications with those indices only where $m - l$ is even) such that the corresponding function

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

satisfies (B.12) and hence has maximal distance ϵ to g. Here the coefficients $a_{(l,m)}$ for $m-l$ odd are simply assumed to be 0. The covering number will hence be bounded by the total number

of such coefficients, which can be calculated as follows:

$$(B.13) \qquad \left| \times_{m=0}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \times_{l=-m}^{\bullet m} G_{(l,m)} \right| = \left| G_{(0,0)} \right| \prod_{m=1}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \prod_{l=-m}^{\bullet m} \left| G_{(l,m)} \right|$$

$$\leqslant \left\lceil \frac{4C}{\epsilon} \right\rceil^2 \prod_{m=1}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \left\lceil \left(\frac{4Cm^{\tilde{t}-\tilde{\tau}}}{\epsilon} \right) \right\rceil^{2(m+1)} \leqslant \left\lceil \frac{4C}{\epsilon} \right\rceil^2 \prod_{m=1}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil} \left\lceil \left(\frac{4C}{\epsilon} \right) \right\rceil^{2(m+1)}$$

$$= \left\lceil \frac{C}{\epsilon} \right\rceil^{2 \sum_{m=1}^{\lceil (C/\epsilon)^{1/(\tilde{\tau}-\tilde{t})} \rceil+1} m} \leqslant \left\lceil \frac{C}{\epsilon} \right\rceil^{8(C/\epsilon)^{2/(\tilde{\tau}-\tilde{t})}} \leqslant \exp \left\{ \log \left(\frac{C}{\epsilon} \right) \left(\frac{C}{\epsilon} \right)^{2/(\tilde{\tau}-\tilde{t})} \right\}.$$

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 $\epsilon > 0$ it follows that

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

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

References

Akritas, M. G. and I. van Keilegom (2001). Nonparametric estimation of the residual distribution. *Scandinavian journal of Statistics 28*, 549–567.

Bertero, M., P. Boccacci, G. Desiderà, and G. Vicidomini (2009). Image deblurring with Poisson data: From cells to galaxies. *Inverse Problems 25*(12), 123006, 26.

Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics 1*, 1071–1095.

Bissantz, N., J. Chown, and H. Dette (2018). Regularization parameter selection in inidirect regression by residual based bootstrap. *to appear in Statist. Sinica*.

Bissantz, N., T. Hohage, A. Munk, and F. Ruymgaart (2007). Convergence rates of general regularization methods for statistical inverse problems. *SIAM Journal on Numerical Analysis 45*, 2610–2636.

Bissantz, N. and H. Holzmann (2013). Asymptotics for spectral regularization estimators in statistical inverse problems. *Computational Statistics 28*, 435–453.

Blanchard, G. and N. Mücke (2018, Aug). Optimal rates for regularization of statistical inverse learning problems. *Foundations of Computational Mathematics 18*(4), 971–1013.

Born, M. and E. Wolf (1970). *Principles of Optics*. Oxford: Pergamon Press.

Buzug, T. (2008). *Computed Tomography. From Photon Statistics to Modern Cone- Beam CT*. Berlin Heidelberg: Springer-Verlag.

Carroll, R. J., A. Delaigle, and P. Hall (2007). Non-parametric regression estimation from data contaminated by a mixture of berkson and classical errors. *J. R. Stat. Soc. Ser. B Stat. Methodol. 69*, 859–878.

Cavalier, L. (1999). Asymptotically efficient estimation in a problem related to tomography. *Math. Methods Statist. 7*, 445–456.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Ann. Statist. 28*, 630–647.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems 24*(3), 034004, 19.

Cavalier, L. and Y. Golubev (2006, 08). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist. 34*(4), 1653–1677.

Cavalier, L. and A. Tsybakov (2002, Jul). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields 123*(3), 323–354.

Colling, B. and I. Van Keilegom (2016, Jun). Goodness-of-fit tests in semiparametric transformation models. *TEST 25*(2), 291–308.

Deans, S. R. (1983). *The Radon Transform and some of its Applications.* New York: Wiley,.

Delaigle, A. and P. Hall (2011). Estimation of observation-error variance in errors-in-variables regression. *Statist. Sinica 21*, 103–1063.

Delaigle, A., P. Hall, and F. Jamshidi (2014, 04). Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*.

Dette, H., N. Neumeyer, and I. V. Keilegom (2007). A new test for the parametric form of the variance function in non-parametric regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 69*(5), 903–917.

Engl, H. W., M. Hanke, and A. Neubauer (1996). *Regularization of inverse problems*, Volume 375 of *Mathematics and its Applications.* Dordrecht: Kluwer Academic Publishers Group.

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

Hotz, T., P. Marnitz, R. Stichtenoth, L. Davies, Z. Kabluchko, and A. Munk (2012). Locally adaptive image denoising by a statistical multiresolution criterion. *Comput. Stat. Data An. 56*, 543 –558.

Janssen, A. J. E. M. (2014, Jul). Zernike expansion of derivatives and laplacians of the zernike circle polynomials. *J. Opt. Soc. Am. A 31*(7), 1604–1613.

Johnstone, I. M. and B. W. Silverman (1990). Speed of estimation in positron emission tomography and related inverse problems. *The Annals of Statistics 18*(1), 251–280.

Kaipio, J. and E. Somersalo (2010). *Statistical and Computational Inverse Problems.* Berlin: Springer.

Kato, K. and Y. Sasaki (2017). Uniform confidence bands for nonparametric errors-in-variables regression. arXiv:1702.03377v3.

Kerkyacharian, G., L. P. E., and D. Picard (2012). Radon needlet thresholding. *Bernoulli 18*(2), 391–433.

Kerkyacharian, G., G. Kyriazis, E. Le Pennec, P. Petrushev, and D. Picard (2010). Inversion of noisy radon transform by svd based needlets. *Applied and Computational Harmonic Analysis 28*(1), 24 – 45.

Khmaladze, E. V. and H. L. Koul (2009, 12). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist. 37*(6A), 3165–3185.

Korostelev, A. P. and A. B. Tsybakov (1993). *Minimax theory of image reconstruction*, Volume 82 of *Lecture Notes in Statistics*. Springer-Verlag.

Koul, H. L. and W. Song (2008). Regression model checking with berkson measurement errors. *J. Statist. Plann. Inference 138*, 1615–1628.

Koul, H. L. and W. Song (2009). Minimum distance regression model checking with berkson measurement errors. *Ann. Statist. 37*, 132–156.

Lakshminarayanan, V. and A. Fleck (2011). Zernike polynomials: a guide. *Journal of Modern Optics 58*(7), 545–561.

Mair, B. A. and F. H. Ruymgaart (1996). Statistical inverse estimation in hilbert scales. *SIAM Journal on Applied Mathematics 56*(5), 1424–1444.

Mason, J. C. and D. Handscomb (2002). *Chebyshev polynomials.* London: A CRC Press Company.

Müller, U., A. Schick, and W. Wefelmeyer (2012, 02). Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference 142*, 552–566.

Natterer, F. (1983a). Computerized tomography with unknown sources. *SIAM Journal on Applied Mathematics 43*(5), 1201–1212.

Natterer, F. (1983b). Computerized tomography with unknown sources. *SIAM Journal on Applied Mathematics 43*(5), 1201–1212.

Natterer, F. (1986). *The mathematics of computerized tomography.* B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester.

Natterer, F. and F. Wübbelling (2001). *Mathematical Methods in Image Reconstruction.* Philadelphia: Society for Industrial and Applied Mathematics.

Nawab, S. H., A. S. Willsky, and A. V. Oppenheim (1996). *Signals and Systems.* New Jersey: Prentice Hall.

Neumeyer, N. (2006). *Bootstrap procedures for empirical processes of nonparametric residuals.* Ruhr-Universitt Bochum: habilitation thesis https://www.math.uni-hamburg.de/home/neumeyer/habil.ps.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics 36*(2), 204–228.

Neumeyer, N. and I. V. Keilegom (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis 101*(5), 1067 – 1078.

Saitoh, S. (1997). *Integral Transforms, Reproducing Kernels and their Applications.* Harlow: Longman.

Silverman, B. W. (1978, 01). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist. 6*(1), 177–184.

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes. With applications to statistics.* New York: Springer Series in Statistics.

Vogel, C. R. (2002). *Computational Methods for Inverse Problems.* Bozeman: SIAM.

Zernike, F. (1934). Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica 1(8)*, 689–704.

Zhang, J., Z. Feng, and X. Wang (2018, Oct). A constructive hypothesis test for the single-index models with two groups. *Annals of the Institute of Statistical Mathematics 70*(5), 1077–1114.