Electronic Journal of Statistics
Vol. 8 (2014) 1–40
ISSN: 1935-7524
DOI: 10.1214/13-EJS874

Additive inverse regression models with convolution-type operators

Thimo Hildebrandt, Nicolai Bissantz and Holger Dette

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany
e-mail: thimo.hildebrandt@rub.de; nicolai.bissantz@rub.de; holger.dette@rub.de

Abstract: In a recent paper Birke and Bissantz (2009) considered the problem of nonparametric estimation in inverse regression models with convolution-type operators. For multivariate predictors nonparametric methods suffer from the curse of dimensionality and we consider inverse regression models with the additional qualitative assumption of additivity. In these models several additive estimators are studied. In particular, we propose a new estimation method for observations on regular spaced grid and investigate estimators under the random design assumption which are applicable when observations are not available on a grid. Finally, we compare these estimators with the marginal integration and the non-additive estimator by means of a simulation study. It is demonstrated that the new method yields a substantial improvement of the currently available procedures.

MSC 2010 subject classifications: Primary 62G08; secondary 62G15, 62G20.
Keywords and phrases: Inverse regression, additive models, convolution-type operators.

Received March 2013.

1. Introduction

Inverse models have numerous applications in such important fields as biology, astronomy, economy or physics, where they have been intensively studied in a deterministic framework [Engl et al. (1996), Saitoh (1997)]. Recently inverse problems have also found considerable interest in the statistical literature. These investigations reflect the demand in applications to quantify the uncertainty of estimates or to validate the model assumptions by the construction of statistical confidence regions or hypotheses tests, respectively [see Mair and Ruymgaart (1996), Kaipio and Somersalo (2010), Bissantz et al. (2007b), Cavalier (2008), Bertero et al. (2009), Bertero et al. (2009) or Birke et al. (2010) among others]. In this paper we are interested in the convolution type inverse regression model

$$Y \quad = \quad g(\mathbf{z}) + \varepsilon = \int_{\mathbb{R}^d} \psi(\mathbf{z}-\mathbf{t})\theta(\mathbf{t})d(\mathbf{t}) + \varepsilon \tag{1.1}$$

with a known function $\psi : \mathbb{R}^d \to \mathbb{R}$ [e.g. Adorf (1995)] and a centered noise term ε. The goal of the experiment is to recover the signal $\theta : \mathbb{R}^d \to \mathbb{R}$ from data $(\mathbf{z_1}, Y_\mathbf{1}), \ldots, (\mathbf{z_n}, Y_\mathbf{n})$ which is closely related to deconvolution [e.g. Stefanski and Carroll (1990) and Fan (1991)]. Models of the type (1.1) have important

applications in the recovery of images from astronomical telescopes or fluorescence microscopes in biology. Therefore statistical inference for the problem of estimating the signal θ in model (1.1) has become an important field of research in recent years, where the main focus is on a one dimensional predictor. Bayesian methods have been investigated in Bertero et al. (2009) and Kaipio and Somersalo (2010) and nonparametric methods have been proposed by Mair and Ruymgaart (1996), Cavalier (2008) and Bissantz et al. (2007b) among others.

In the present paper we investigate convergence properties of Fourier-based estimators for the function θ with the following purposes. Firstly, our research is motivated by the fact that deconvolution problems often arise with a multivariate predictor such as location and time. For this situation Birke and Bissantz (2009) proposed a nonparametric estimate of the signal θ and derived its asymptotic properties under rather strong assumptions. We will discuss the nonparametric estimation problem for the signal θ under substantially weaker assumptions. Secondly, because nonparametric estimation usually suffers from the curse of dimensionality improved estimators incorporating qualitative assumptions such as additivity or multiplicity are investigated under the fixed and the random design assumption. While additive estimation has been intensively discussed for direct problems from different perspectives [see Linton and Nielsen (1995), Mammen et al. (1999), Carroll et al. (2002), Hengartner and Sperlich (2005), Nielsen and Sperlich (2005), Doksum and Koo (2000), Horowitz and Lee (2005), Lee et al. (2010), Dette and Scheder (2011)] – to our best knowledge – only one additive estimator is available for indirect inverse regression models so far where it is assumed that the observations are available on a grid [see Birke et al. (2012)]. In this paper we are particularly interested in two alternative additive estimators. The first one is applicable if observations are available on a grid but has a substantially simpler structure than the method proposed by the last-named authors, which makes it very attractive for practitioners. Moreover, it also yields much more precise estimates than the method of Birke et al. (2012). The second estimator is additionally applicable in the case of random predictors.

Thirdly, we will also investigate the case of correlated errors in the inverse regression model (1.1), which has – to our best knowledge – not been considered so far although it appears frequently in applications. Finally, we do not assume that the kernel ψ is periodic, which is a common assertion in inverse regression models with convolution operator [see e.g. Cavalier and Tsybakov (2002)]. Note that for many problems such as the reconstruction of astronomical and biological images from telescopic and microscopic imaging devices this assumption is unrealistic.

The remaining part of this paper is organized as follows. In Section 2 we consider inverse additive regression models with a random design and derive the weak convergence of an unrestricted estimator. We introduce the marginal integration estimator and demonstrate that this method yields improved estimators under the assumption of additivity. Section 3 is devoted to the inverse regression model, where observations are available on a regular grid. We introduce a new additive estimate, which is similar in spirit with the backfitting

methodology. Asymptotic properties of these estimators are investigated and asymptotic normality of (appropriately standardized) statistics will be established. In Section 4 we explain how the results are changing for dependent data while Section 5 presents a small simulation study of the finite sample properties of the proposed methods. In particular we compare the new additive estimator with the currently available methods and demonstrate its superiority by a factor 6–8 with respect to the mean squared error criterion. Finally all technical details regarding the proofs of our asymptotic results can be found in Section 6.

2. Additive inverse regression with random design

2.1. Unrestricted estimation for random design

Recall the definition of model (1.1) where we assume that the moments $E[\varepsilon^k]$ exist for all $k \in \mathbb{N}$ such that $E[\varepsilon] = 0$ and $\sigma^2 = E[\varepsilon^2] > 0$. For the sake of transparency we also assume at this point that the errors corresponding to different predictors are independent – for the more general case of an error process with an MA(q)-structure, see Section 4. In this section we concentrate on the random design case, i.e. the explanatory variables $\mathbf{z}_k$ are realizations of random variables.

(RD) Under the *random design* assumption we assume that the explanatory variables are realizations of independent, identically distributed random variables $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$ with a density f_n. We will not reflect the triangular structure in the notation and use $Y_k, \mathbf{X}_k, \varepsilon_k$ and f instead of $Y_{k,n}, \mathbf{X}_{k,n}, \varepsilon_{k,n}$ and f_n, respectively, that is

$$Y_k = g(\mathbf{X}_k) + \varepsilon_k = \int_{\mathbb{R}^d} \psi(\mathbf{X}_k - \mathbf{t})\theta(\mathbf{t})d\mathbf{t} + \varepsilon_k; \quad k \in \{1, \ldots, N\}, \quad (2.1)$$

where $\varepsilon_1, \ldots, \varepsilon_N$ are independent identically distributed random variables.

Fourier-based estimators have been considered by numerous authors in the univariate case (e.g. Diggle and Hall (1993), Mair and Ruymgaart (1996), Cavalier and Tsybakov (2002) and Bissantz et al. (2007a)) and its generalization to the multivariate case considered in model (2.1) is straightforward. A Fourier-based estimator is given by

$$\hat{\theta}^{RD}(\mathbf{x}^*) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w}, \mathbf{x}^* \rangle} \Phi_K(h\mathbf{w}) \frac{\hat{\Phi}^{RD}(\mathbf{w})}{\Phi_\psi(\mathbf{w})} d\mathbf{w}, \quad (2.2)$$

where $\langle \mathbf{v}, \mathbf{w} \rangle$ denotes the standard inner product of the vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$ and Φ_K and Φ_ψ denote the Fourier transform of a kernel function K and the convolution function ψ (which is assumed to be known), respectively. Moreover, in (2.2) the quantity h is a bandwidth converging to 0 with increasing sample size, the empirical Fourier transform $\hat{\Phi}^{RD}(w)$ is defined by

$$\hat{\Phi}^{RD}(\mathbf{w}) \quad = \quad \frac{1}{n} \sum_{k=1}^{n} e^{i\langle \mathbf{w}, \mathbf{X}_k \rangle} \frac{Y_k}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}}, \quad (2.3)$$

$1/\mathbf{a_N} = (1/a_N, \ldots, 1/a_N) \in \mathbb{R}^d$ and a_N is a sequence converging to 0 with increasing sample size. The resulting estimator will be denoted by $\hat{\theta}^{RD}(\mathbf{x}^*)$. In (2.3) f denotes the density of $\mathbf{X}_1$ and we take the maximum of $f(\mathbf{X}_k)$ and $f(\frac{1}{\mathbf{a_N}})$ to ensure that the variance of $\hat{\theta}^{RD}(\mathbf{x}^*)$ is bounded. We also note that the estimator $\hat{\theta}^{RD}$ admits the representation

$$\hat{\theta}^{RD}(\mathbf{x}^*) \quad = \quad \sum_{k=1}^{n} Y_k w_N(\mathbf{x}^*, \mathbf{X}_k), \tag{2.4}$$

where the weights are given by

$$w_N(\mathbf{x}^*, \mathbf{X}_k) = \frac{1}{N \max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w}, \mathbf{x}^* - \mathbf{X}_k \rangle} \frac{\Phi_K(h\mathbf{w})}{\Phi_\psi(\mathbf{w})} d\mathbf{w}. \tag{2.5}$$

Remark 2.1. Note that we use the same bandwidth for all components of the predictor. This assumption is made for the sake of a transparent presentation of the results. In applications the components of the vector $\mathbf{x}$ represent different physical quantities such that different bandwidths have to be used. All results presented in this paper can be modified to this case with an additional amount of notation.

2.2. Estimation of additive inverse regression models

It is well known that in practical applications nonparametric methods as introduced in Section 2.1 suffer from the curse of dimensionality and therefore do not yield precise estimates of the signal θ with a multivariate predictor. A common approach in nonparametric statistics to deal with this problem is to postulate an additive structure of the signal θ. In the case of direct regression models several estimation techniques such as marginal integration [see Linton and Nielsen (1995), Carroll et al. (2002), Hengartner and Sperlich (2005)] and backfitting [Mammen et al. (1999), Nielsen and Sperlich (2005)] have been proposed in the literature for estimating an additive regression function. Recently the estimation problem of an additive (direct) regression model has also found considerable interest in the context of quantile regression [see Doksum and Koo (2000), De Gooijer and Zerom (2003), Horowitz and Lee (2005), Lee et al. (2010), Dette and Scheder (2011) among others] but – to our best knowledge – only one estimator has been proposed for additive inverse regression models under the assumption that observations are available on a grid [see Birke et al. (2012)]. For this situation we will propose an alternative estimator in the following section, which yields an improvement by a factor 6–10 with respect to mean squared error (see our numerical results in Section 5). Before we introduce the marginal integration method in the context of inverse models, we illustrate it for a direct additive regression model of the form

$$y = g_1(x) + g_2(z) + \varepsilon \tag{2.6}$$

where y is a one-dimensional response and x and z are d_1 and d_2-dimensional predictors, respectively. Following Linton and Nielsen (1995) we denote by Q_1 a measure with density q_1 satisfying $\int q_1(z)dz = 1$ and consider the contrast

$$\alpha_{Q_1}(x) = \int (g_1(x) + g_2(z))q_1(z)dz = g_1(x) + c_1$$

where $c_1 = \int g_2(z)q_1(z)dz$. Estimators of the component $g_1 + c_1$ can now be easily obtained by integrating a nonparametric estimator of the regression curve, say $\hat{g}(x, z)$, in the same way, that is

$$\hat{\alpha}_{Q_1}(x) = \int \hat{g}(x, z)q_1(z)dz.$$

Using a similar argument for the second component we obtain

$$\hat{\alpha}_{Q_2}(z) = \int \hat{g}(x, z)q_2(x)dx$$

as an estimator of $g_2 + c_2$, where $c_2 = \int g_1(x)q_2(x)dx$. The resulting estimator of the regression function is finally constructed as

$$\hat{g}_{add}(x, y) = \hat{\alpha}_{Q_1}(x) + \hat{\alpha}_Q(z) - \hat{c}$$

where $\hat{c} = \int \hat{g}(x, z)q_1(z)q_2(x)dzdz$ is an estimate of $c_1 + c_2$. In the following discussion we will combine the marginal integration methodology with Fourier-estimation techniques [see Diggle and Hall (1993)] for the construction of estimators of an additive signal in the inverse regression model (2.1). To be precise we assume an additive structure for the signal θ, that is

$$\theta(\mathbf{x}^*) = \theta^{add}(\mathbf{x}^*) := \theta_0^{add} + \sum_{j=1}^{m} \theta_{I_j}^{add}(\mathbf{x}_{I_j}^*) \tag{2.7}$$

[see Hastie and Tibishirani (2008)]. Here $\{I_1, \ldots, I_m\}$ denotes a partition of the set $\{1, \ldots, d\}$ with cardinalities $|I_j| = d_j$ and $\mathbf{x}_{I_j}^*$ is the vector which includes all components of the vector $\mathbf{x}^*$ with corresponding indices $i \in I_j$. Furthermore θ_0^{add} is a constant and $\theta_{I_j}^{add} : \mathbb{R}^{d_j} \to \mathbb{R}$ denote functions normalized such that

$$\int \theta_{I_j}^{add}(\mathbf{x})d(\mathbf{x}) = 0 \quad (j = 1, \ldots, m).$$

Note that the completely additive case is obtained for the choice $m = d$, that is $d_1 = \cdots = d_d = 1$.

To construct an estimator in the additive inverse regression model (2.7) with random design we apply the marginal integration method introduced in Linton and Nielsen (1995) to the statistic defined in (2.4). To this end we consider weighting functions $Q_{I_1}, \ldots, Q_{I_m}, Q_{I_j} : \mathbb{R}^{d_j} \to \mathbb{R}$ and define

$$\begin{aligned} Q(\mathbf{x}^*) &= Q_{I_1}(\mathbf{x}_{I_1}^*) \cdots Q_{I_m}(\mathbf{x}_{I_m}^*) \\ Q_{I_j^c}(\mathbf{x}_{I_j^c}^*) &= Q_{I_1}(\mathbf{x}_{I_1}^*) \cdots Q_{I_{j-1}}(\mathbf{x}_{I_{j-1}}^*) Q_{I_{j+1}}(\mathbf{x}_{I_{j+1}}^*) \cdots Q_{I_m}(\mathbf{x}_{I_m}^*), \end{aligned} \tag{2.8}$$

where $I_j^c = \{1, \dots, d\} \setminus I_j$. With this notation we introduce the quantities

$$\alpha_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j}) = \int_{\mathbb{R}^{d-d_j}} \theta(\mathbf{x}^*) dQ_{I_j^c}(\mathbf{x}^*_{I_j^c}), \quad j = 1, \dots, m, \tag{2.9}$$

$$c = \int_{\mathbb{R}^d} \theta(\mathbf{x}^*) dQ(\mathbf{x}^*). \tag{2.10}$$

Now let $\hat{\theta}^{RD}$ denote the unrestricted estimator introduced in Section 2.1 for the random design model, then the additive estimator for the signal θ is finally defined by

$$\hat{\theta}^{add,RD}(\mathbf{x}^*) = \hat{\alpha}_{1,Q_{I_1^c}}(\mathbf{x}^*_{I_1}) + \cdots + \hat{\alpha}_{m,Q_{I_m^c}}(\mathbf{x}^*_{I_m}) - (m-1)\hat{c} \tag{2.11}$$

where $\hat{c}$ and $\hat{\alpha}_{j,Q_{I_j^c}}$ denote estimates for the quantities c and $\alpha_{j,Q_{I_j^c}}$ which are obtained by replacing in (2.9) and (2.10) the signal θ by its estimator $\hat{\theta}^{RD}$, respectively. Recalling the definition of the unrestricted estimator in (2.2) and (2.3), we obtain from (2.9) the representation

$$\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j}) = \sum_{k=1}^n Y_k w_N^{add}(\mathbf{x}^*_{I_j}, \mathbf{X}_k), \tag{2.12}$$

where the weights are given by

$$w_N^{add}(\mathbf{x}^*_{I_j}, \mathbf{X}_k) = \frac{1}{Nh^d(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{X}_k\rangle/h} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\left(\frac{\mathbf{w}_{I_j^c}}{h}\right) \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \frac{1}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a}_N})\}}.$$

Here and throughout this paper we define for $\mathbf{y} \in \mathbb{R}^{d-d_j}$

$$L_{I_j^c}(\mathbf{y}) = \int_{\mathbb{R}^{d-d_j}} e^{-i\langle \mathbf{y}, \mathbf{x}_{I_j^c}\rangle} dQ_{I_j^c}(\mathbf{x}_{I_j^c}), \tag{2.13}$$

where $Q_{I_j^c}(\mathbf{x}_{I_j^c})$ is defined in (2.8).

2.3. Technical Assumptions

In this Section we will derive important asymptotic properties of the proposed estimators in the additive inverse regression model with a random design. For this purpose the following assumptions are required. Throughout this paper $\|\cdot\|$ denotes the Euclidean norm and the symbol $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = c$ for some positive constant c.

Assumption (A.1). The Fourier transform Φ_ψ of the function ψ satisfies (as $h \to 0$)

$$\int_{\mathbb{R}^d} \frac{|\Phi_K(\mathbf{w})|}{|\Phi_\psi(\frac{\mathbf{w}}{h})|} d\mathbf{w} \leq C_1 h^{-\beta}, \quad \int_{\mathbb{R}^d} \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} \sim C_2 h^{-2\beta}$$

for some $\beta > 0$ and constants $C_1, C_2 > 0$.

Assumption (A.2). The Fourier transform Φ_K of the kernel K in (2.2) is symmetric, supported on the cube $[-1,1]^d$ and there exists a constant $b \in (0,1]$ such that $\Phi_K(\mathbf{w}) = 1$ for $\mathbf{w} \in [-b,b]^d, b > 0$, and $|\Phi_K(\mathbf{w})| \leq 1$ for all $\mathbf{w} \in [-1,1]^d$.

Assumption (A.3).

(A) The Fourier transform Φ_θ of the signal θ in model (1.1) exists and satisfies

$$\int_{\mathbb{R}^d} |\Phi_\theta(\mathbf{w})| \parallel \mathbf{w} \parallel^{s-1} d\mathbf{w} < \infty \quad \text{for some} \quad s > 1.$$

(B) The function g in model (1.1) satisfies

$$\int_{\mathbb{R}^d} |g(\mathbf{z})| \parallel \mathbf{z} \parallel^{r} d\mathbf{z} < \infty$$

for some $r > 0$ such that $a_n^r = O(h^{\beta+d+s-1})$.

(C) The Fourier transforms $\Phi_{\theta^{add}_{I_1}}, \ldots, \Phi_{\theta^{add}_{I_m}}$ of the functions $\theta^{add}_{I_1}, \ldots, \theta^{add}_{I_m}$ in the additive model (2.7) satisfy

$$\int_{\mathbb{R}^d} |\Phi_{\theta^{add}_{I_j}}(\mathbf{w})| \parallel \mathbf{w} \parallel^{s-1} d\mathbf{w} < \infty \quad \text{for some} \quad s > 1 \text{and} j = 1, \ldots, m.$$

(D) The functions $g_{I_1}, \ldots, g_{I_m}$ defined in (3.3) satisfy

$$\int_{\mathbb{R}^{d_j}} |g_{I_j}(\mathbf{z})| \parallel \mathbf{z} \parallel^{r} d\mathbf{z} < \infty \quad \text{for} \quad j = 1, \ldots, m$$

for some $r > 0$ such that $a_N^{r-d_j} = O(h^{\beta_j+s+d_j-1})$.

Assumption (A.4). For each $N \in \mathbb{N}$ let $\mathbf{X}_1, \ldots, \mathbf{X_N}$ denote independent identically distributed d-dimensional random variables with density f (which may depend on N) such that $f(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in [-\frac{1}{a_N}, \frac{1}{a_N}]^d$. Furthermore we assume, that for sufficiently large $N \in \mathbb{N}$

$$f(\mathbf{x}) \geq f(\frac{1}{\mathbf{a_N}}) \quad \text{for} \quad \mathbf{x} \in \left[-\frac{1}{a_N}, \frac{1}{a_N}\right]^d.$$

The final assumption is required for the marginal integration estimator and is an extension of Assumption (A.1).

Assumption (A.5). There exist positive constants $\gamma_1, \ldots, \gamma_m$ such that the Fourier transform Φ_ψ of the convolution function ψ satisfies

(A) $\int_{\mathbb{R}^d} \left| L_{I_j^c}\left(\frac{\mathbf{w}_{I_j^c}}{h}\right)\right|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} \sim C_3 h^{-2\beta+\gamma_j}$ $(j = 1, \ldots, m)$

(B) $\int_{\mathbb{R}^d} \left| \sum_{j=1}^m e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\left(\frac{\mathbf{w}_{I_j^c}}{h}\right)\right|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} \sim C_4 h^{-2\beta+\gamma_{min}}$, where $\gamma_{min} = \min_{j=1}^m \gamma_j$

(C) $\int_{\mathbb{R}^d} \left(\prod_{j=1}^m \left| L_{I_j^c}\left(\frac{\mathbf{w}_{I_j^c}}{h}\right)\right|^2 \right) \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} = o\left(h^{-2\beta+\gamma_{min}}\right).$

Remark 2.2.

1. Assumptions (A.3)(B) and (A.3)(D) are needed for the computation of the bias, where we have to ensure that $g(\mathbf{x})$ converges sufficiently fast to zero as $x \to \infty$.
2. The results of this and the following Section can be extended to multiplicative signals of the form

$$\theta(\mathbf{x}^*) \quad = \quad \prod_{j=1}^m \theta_{I_j}(\mathbf{x}^*_{I_j}). \tag{2.14}$$

The details are omitted for the sake of brevity.

Example 2.3. In order to demonstrate that the assumptions are satisfied in many cases of practical importance we consider exemplarily Assumptions (A.1) and (A.5) and a two dimensional additive signal that is $\mathbf{x} = (x_1, x_2)$,

$$\theta(x_1, x_2) = \theta_1(x_1) + \theta_2(x_2),$$

($I_1 = I_2^c = \{1\}, I_2 = I_1^c = \{2\}$). For the convolution function in (1.1) and the weight (2.8) we choose

$$\begin{aligned} \psi(\mathbf{x}) &= \frac{\lambda^2}{4} e^{-\lambda(|x_1|+|x_2|)} \\ Q(\mathbf{x}) &= 1_{[-1,1]^2}(\mathbf{x}), \end{aligned}$$

respectively, and the kernel K is defined by

$$K(\mathbf{x}) = \frac{\sin(x_1)\sin(x_2)}{\pi^2 x_1 x_2}.$$

Note that the Fourier transform of the kernel is given by $\Phi_K(\mathbf{w}) = I_{[0,1]^2}(\mathbf{w})$. The integrals in Assumptions (A.1) and (A.5) are obtained by a straightforward calculation

$$\int_{\mathbb{R}^2} \frac{|\Phi_K(\mathbf{w})|}{|\Phi_\psi(\frac{\mathbf{w}}{h})|} d\mathbf{w} = \int_{[-1,1]^2} \left(1 + \frac{w_1^2}{h^2}\right)\left(1 + \frac{w_2^2}{h^2}\right) d\mathbf{w} = \left(\frac{2}{3h^2} + 2\right)^2$$

$$\int_{\mathbb{R}^2} \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} = \int_{[-1,1]^2} \left(1 + \frac{w_1^2}{h^2}\right)^2\left(1 + \frac{w_2^2}{h^2}\right)^2 d\mathbf{w} = \left(\frac{2}{5h^4} + \frac{4}{3h^2} + 2\right)^2$$

$$\begin{aligned} \int_{\mathbb{R}^2} \left| L_1\left(\frac{w_1}{h}\right)\right|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} &= \int_{[-1,1]^2} \frac{4h^2 |\sin\left(\frac{w_1}{h}\right)|^2 \left(1 + \frac{w_1^2}{h^2}\right)^2 \left(1 + \frac{w_2^2}{h^2}\right)^2}{w_1^2} d\mathbf{w} \\ &= \frac{8}{15h^6} + o(h^{-6}). \end{aligned}$$

Similarly we obtain for the remaining integrals in (A.5)

$$\int_{\mathbb{R}^2} \left| \sum_{j=1}^{2} e^{-i\langle w_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\left(\frac{w_{I_j^c}}{h}\right)\right|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w}$$
$$= h^2 \int_{[-1,1]^2} \left| e^{-iw_1x_1/h} \frac{\sin\left(\frac{w_2}{h}\right)}{w_2} + e^{-iw_2x_2/h} \frac{\sin\left(\frac{w_1}{h}\right)}{w_1}\right|^2$$
$$\times \left(1+\frac{w_1^2}{h^2}\right)^2 \left(1+\frac{w_2^2}{h^2}\right)^2 d\mathbf{w} = \frac{16}{15h^6} + o\left(h^{-6}\right)$$

$$\int_{\mathbb{R}^2} \prod_{j=1}^{2} \left| L_{I_j}\left(\frac{w_{I_j}}{h}\right)\right|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} = \left(\int_{[-1,1]} \frac{4h^2 |\sin\left(\frac{w_1}{h}\right)|^2 \left(1+\frac{w_1^2}{h^2}\right)^2}{w_1^2} dw_1 \right)^2$$
$$= \frac{16}{9h^4} + o(h^{-4})$$

2.4. Asymptotic properties

We begin with a discussion of the weak convergence of the unrestricted estimator $\hat{\theta}^{RD}$ for the signal θ in the inverse regression model (2.1) corresponding to the random design assumption. Here the estimator is given by

$$\hat{\theta}^{RD}(\mathbf{x}^*) = \frac{1}{Nh^d(2\pi)^d} \sum_{k=1}^{n} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w}, \mathbf{x}^* - \mathbf{X}_k\rangle/h} \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \frac{Y_k}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}} \tag{2.15}$$

and its asymptotic properties are described in our first main result which is proved in the Appendix. Throughout this paper the symbol $\Rightarrow$ denotes weak convergence.

Theorem 2.4. *Consider the inverse regression model (2.1) under the random design assumption (RD). Let Assumptions (A.1), (A.2), (A.3)(A), (A.3)(B), (A.4) and (A.5) be fulfilled and assume that $h \to 0$ and $a_N \to 0$ as $N \to \infty$ such that*

$$N^{1/2} h^{\beta + d/2} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{1/2} \to \infty \qquad \text{and} \qquad N^{1/2} h^{3d/2} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{3/2} \to \infty.$$

Furthermore, assume that the errors in model (2.1) are independent, identically distributed with mean zero and variance σ^2. Then

$$V_1^{-1/2}\left(\hat{\theta}^{RD}(\boldsymbol{x}^*) - E[\hat{\theta}^{RD}(\boldsymbol{x}^*)]\right) \Rightarrow \mathcal{N}(0,1), \tag{2.16}$$

where $E[\hat{\theta}^{RD}(\boldsymbol{x}^*)] = \theta(\boldsymbol{x}^*) + O(h^{s-1})$ and the normalizing sequence

$$V_1 = \frac{1}{N(2\pi)^{2d}} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{s}, (\boldsymbol{x}^* - \boldsymbol{y})\rangle} \frac{\Phi_K(h\boldsymbol{s})}{\Phi_\psi(\boldsymbol{s})} d\boldsymbol{s} \right|^2 \frac{(\sigma^2 + g^2(\boldsymbol{y}))f(\boldsymbol{y})}{\max\left\{f(\boldsymbol{y}), f(\frac{1}{\boldsymbol{a_N}})\right\}^2} d\boldsymbol{y} \tag{2.17}$$

is bounded by

$$C_l N^{1/2} h^{d/2+\beta} f(\boldsymbol{a_N^{-1}})^{1/2} \leq V_1^{-1/2} \leq C_u N^{1/2} h^{d/2+\beta}. \tag{2.18}$$

Remark 2.5. Note that the rate of convergence in Theorem 2.4 depends sensitively on the design density. We demonstrate this by providing two examples, one for the fastest and one for the slowest possible rate. First, assume that the predictors are uniformly distributed on the cube $[-\frac{1}{a_N}, \frac{1}{a_N}]^d$ and that the convolution function is the d-dimensional Laplace density function. This yields $\beta = 2d$ in Assumption (A.1) and we get a rate of convergence of order $N^{1/2}h^{5d/2}a_N^{d/2}$, which is exactly the lower bound in Theorem 2.4. However, a rate of order $N^{1/2}h^{5d/2}$ is obtained for the design density

$$f(x_1, \ldots, x_d) \quad = \quad \prod_{k=1}^{d} g_{a,b}(x_k),$$

where the function $g_{a,b} : \mathbb{R} \to \mathbb{R}$ is defined by

$$g_{a,b}(x) = \begin{cases} a, & \text{if } x \in [-1, 1] \\ \frac{a}{|x|^b}, & \text{else,} \end{cases}$$

and the parameters a and b are given by $b > 1, a = (2 + \frac{2}{b-1})^{-1}$. In this case we have

$$V_1^{-1/2} \sim N^{-1/2}h^{-5d/2} + N^{-1/2}h^{-2d}a_N^{(-b+1)/2}.$$

For the choice $h = o(a_N^{b-1})$ we therefore obtain $V_1^{-1/2} \sim N^{-1/2}h^{-5d/2}$.

We now consider the marginal integration estimator $\hat{\theta}^{add,RD}$ defined in (2.11) under the random design assumption. Lemma 2.6 below gives the asymptotic behaviour of the j-th component $\hat{\alpha}_{j,Q_{I_j^c}}$ and Theorem 2.8 the asymptotic distribution of $\hat{\theta}^{add,RD}$. The proofs are complicated and therefore also deferred to Section 6.

Lemma 2.6. *If Assumptions (A.1), (A.2), (A.3)(C), (A.3)(D), (A.4) and (A.5) are satisfied and*

$$N^{1/2}h^{\beta+d/2-\gamma_j/2} f(\boldsymbol{a_N^{-1}})^{1/2} \to \infty \qquad \text{and} \qquad N^{1/2}h^{3/2(d-\gamma_j)} f(\boldsymbol{a_N^{-1}}) \to \infty$$

as $N \to \infty$. Then the appropriately standardized estimator $\hat{\alpha}_{j,Q_{I_j^c}}(\boldsymbol{x}^*_{I_j})$ defined in (2.12) converges weakly to a standard normal distribution, that is

$$V_2^{-1/2}\big(\hat{\alpha}_{j,Q_{I_j^c}}(\boldsymbol{x}^*_{I_j}) - E[\hat{\alpha}_{j,Q_{I_j^c}}(\boldsymbol{x}^*_{I_j})]\big) \Rightarrow \mathcal{N}(0,1) \tag{2.19}$$

for $j = 1, \ldots, m$, where $E[\hat{\alpha}_{j,Q_{I_j^c}}(\boldsymbol{x}^*_{I_j})] = \alpha_{j,Q_{I_j^c}}(\boldsymbol{x}^*_{I_j}) + O(h^{s-1})$ and the standardizing factor

$$\begin{aligned} V_2 \quad = \quad & \frac{1}{N(2\pi)^d} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{w}, \boldsymbol{x}\rangle} e^{i\langle \boldsymbol{w}_{I_j}, \boldsymbol{x}_{I_j}\rangle} L_{I_j^c}\big(\boldsymbol{w}_{I_j^c}\big) \frac{\Phi_K(h\boldsymbol{w})}{\Phi_\psi(\boldsymbol{w})} d\boldsymbol{w} \right|^2 \\ & \times \frac{(\sigma^2 + g(\boldsymbol{x})^2) f(\boldsymbol{x})}{\max\{f(\boldsymbol{x}), f(\frac{1}{\boldsymbol{a_N}})\}^2} d\boldsymbol{x}. \end{aligned}$$

satisfies

$$C_l N^{1/2} h^{d/2+\beta-\gamma_j} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{1/2} \le V_2^{-1/2} \le C_u N^{1/2} h^{d/2+\beta-\gamma_j}. \tag{2.20}$$

Remark 2.7.

(a) Similar to the unrestricted case, the rate of convergence depends on the design density f. Note that under the given assumptions the rate of convergence of the estimator $\hat{\alpha}_{j,Q_{I_j^c}}$ is by the factor h^{γ_j} faster than the rate of the unrestricted estimator.

(b) In the cases of direct regression the problem of estimating an additive function has been studied extensively in the literature [see Linton and Nielsen (1995), Mammen et al. (1999) Hengartner and Sperlich (2005) among many others]. In the direct additive regression model

$$Y_k = g(\mathbf{X_k}) + \varepsilon_k = g_0 + g_1(X_{1,k}) + \cdots + g_d(X_{d,k}) + \varepsilon_k$$

the rate of the convergence of the marginal integration estimators of the component g_j is given by $N^{1/2}h^{1/2}$ (assuming appropriate smoothness conditions). On the other hand it follows from Lemma 2.6 that in the inverse regression model the corresponding rate of convergence is bounded by $N^{1/2}h^{1/2+\beta-\gamma_j} f^{1/2}(1/\mathbf{a_N})$ and $N^{1/2}h^{1/2+\beta-\gamma_j}$. The additional factor $f^{1/2}(\mathbf{1}/\mathbf{a_N})h^{\beta-\gamma_j}$ can be interpreted as the price which has to be paid for the ill-posed problem.

(c) Assume that $m = 1$ (i.e. $d_j = 1; j = 1, \ldots, d$). If all components except one in model (2.6) would be known, this component could be estimated at a rate varying between $N^{1/2}h^{\beta+1/2}$ and $N^{1/2}h^{\beta+1/2}f(\mathbf{1}/\mathbf{a_N})^{1/2}$. Comparing these bound with (2.20) we conclude that in the case of the random design the marginal integration estimate does not have an oracle property.

Theorem 2.8. *If Assumptions* (A.1), (A.2), (A.3)(C), (A.3)(D), (A.4) *and* (A.5) *are satisfied and*

$$N^{1/2}h^{\beta+(3d+\gamma_{min})/2} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{1/2} \to \infty, \quad N^{1/2}h^{\beta+(d-\gamma_{min})/2} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{1/2} \to \infty$$
$$N^{1/2}h^{3/2(d-\gamma_j)} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{3} \to \infty \quad (j = 1, \ldots, m)$$

as $N \to \infty$, then the appropriately standardized additive estimator $\hat{\theta}^{add,RD}$ converges weakly to a standard normal distribution, that is

$$V_3^{-1/2}\big(\hat{\theta}^{add,RD}(\boldsymbol{x}^*) - E[\hat{\theta}^{add,RD}(\boldsymbol{x}^*)]\big) \Rightarrow \mathcal{N}(0,1), \tag{2.21}$$

where $E[\hat{\theta}^{add,RD}(\boldsymbol{x}^*)] = \theta^{add}(\boldsymbol{x}^*) + O(h^{s-1})$ and the standardizing factor

$$V_3 = \frac{1}{N(2\pi)^{2d}} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{w},\boldsymbol{s}\rangle} \left(\sum_{j=1}^{m} e^{-i\langle \boldsymbol{w}_{I_j}, \boldsymbol{x}^*_{I_j}\rangle} L_{I_j^c}(\boldsymbol{w}_{I_j^c}) \right) \frac{\Phi_K(h\boldsymbol{w})}{\Phi_\psi(\boldsymbol{w})} d\boldsymbol{w} \right|^2$$
$$\times \frac{(\sigma^2 + g(\boldsymbol{s})^2) f(\boldsymbol{s})}{\max\{f(\boldsymbol{s}), f(\frac{1}{\boldsymbol{a}_N})\}^2} d\boldsymbol{s}.$$

satisfies

$$C_l N^{1/2} h^{d/2+\beta-\gamma_{min}} f(\boldsymbol{a}_{\boldsymbol{N}}^{-1})^{1/2} \le V_3^{-1/2} \le C_u N^{1/2} h^{d/2+\beta-\gamma_{min}}.$$

3. Additive inverse regression with fixed design

An important application of inverse regression is image reconstruction. In these cases observations are often available on a regular grid and this situation will be studied in this section. We assume again the existence of the moments $E[\varepsilon^k]$ of the errors in model (1.1) for all $k \in \mathbb{N}$ such that $E[\varepsilon] = 0$ and $\sigma^2 = E[\varepsilon^2] > 0$. Regarding the design we make the following assumption.

(FD) Under the *fixed design* assumption we assume that observations are available on a grid of increasing size. More precisely we consider a sequence $a_n \to 0$ as $n \to \infty$ and assume that at each location $\mathbf{z_k} = \frac{\mathbf{k}}{na_n} \in \mathbb{R}^d$ with $\mathbf{k} = (k_1, \dots, k_d) \in \{-n, \dots, n\}^d$ a pair of observations $(\mathbf{z_k}, Y_\mathbf{k})$ is available in the model

$$Y_\mathbf{k} = g(\mathbf{z_k}) + \varepsilon_\mathbf{k} = \int_{\mathbb{R}^d} \psi(\mathbf{z_k} - \mathbf{t})\theta(\mathbf{t})d\mathbf{t} + \varepsilon_\mathbf{k}, \tag{3.1}$$

where $\{\varepsilon_\mathbf{k} \,|\, \mathbf{k} \in \{-n, \dots, n\}^d\}$ are independent and identically distributed random variables. Under this assumption the sample size is $N = (2n+1)^d$. Note that formally the random variables $\{Y_\mathbf{k} \,|\, \mathbf{k} \in \{-n, \dots, n\}^d\}$ form a triangular array, but we do not reflect this dependence in the notation. In other words we will use the notation $Y_\mathbf{k}, \mathbf{z_k}, \varepsilon_\mathbf{k}$ instead of $Y_{\mathbf{k},n}, \mathbf{z}_{\mathbf{k},n}, \varepsilon_{\mathbf{k},n}$ throughout this paper.

3.1. An alternative additive estimator for a fixed design

In principle the marginal integration estimator could also be used under the fixed design assumption (FD) and its asymptotic properties have been studied by Birke et al. (2012). This estimator is defined as in (2.2), where the empirical Fourier transform is replaced by

$$\hat{\Phi}^{FD}(\mathbf{w}) \quad = \quad \frac{1}{n^d a_n^d} \sum_{\mathbf{k} \in \{-n, \dots, n\}^d} Y_\mathbf{k} e^{i\langle \mathbf{w}, \mathbf{z}_k \rangle}.$$

However, it turns out that for observations on a grid a simpler and more efficient estimator can be defined. This idea is closely related to the backfitting approach. To be precise we note that the assumption of additivity for the signal θ implies additivity of the observable signal g due to the linearity of the convolution operator. Hence, model (3.1) is equivalent to

$$Y_\mathbf{k} \quad = \quad g_0 + g_{I_1}(\mathbf{z}_{k_{I_1}}) + \cdots + g_{I_m}(\mathbf{z}_{k_{I_m}}) + \varepsilon_\mathbf{k}, \tag{3.2}$$

where $g_0 = \int_{\mathbb{R}^d} \psi(\mathbf{z} - \mathbf{t})\theta_0 d\mathbf{t}$,

$$g_{I_j}(\mathbf{z}_{\mathbf{k}_{I_j}}) = \int_{\mathbb{R}^{d_j}} \psi_{I_j}(\mathbf{z}_{\mathbf{k}_{I_j}} - \mathbf{t}_{I_j})\theta_{I_j}^{add}(\mathbf{t}_{I_j})d\mathbf{t}_{I_j} \qquad (j = 1, \dots, m) \tag{3.3}$$

and $\psi_{I_1}, \ldots, \psi_{I_d}$ are the marginals of ψ, that is

$$\psi_{I_j}(\mathbf{t}_{I_j}) \;=\; \int_{\mathbb{R}^{d-d_j}} \psi(\mathbf{t}) d\mathbf{t}_{I_j^c}.$$

Recall the definition of $\mathbf{k}_{I_j}$ and $\mathbf{k}_{I_j^c}$ as the d_j and $(d-d_j)$-dimensional vector corresponding to the components $(k_l \mid l \in I_j)$ and $(k_l \mid l \in I_j^c)$ of the vector $\mathbf{k} = (k_1, \ldots, k_d)$, respectively. In order to define estimators of these terms we consider the empirical Fourier transforms in dimension d_j

$$\hat{\Psi}_{I_j}(\mathbf{w}) \;=\; \frac{1}{(na_n)^{d_j}} \sum_{\mathbf{k}_{I_j} \in \{-n,\ldots,n\}^{d_j}} Z_{\mathbf{k}_{I_j}} e^{i\langle \mathbf{w}, \mathbf{z}_{\mathbf{k}_{I_j}} \rangle} \quad (j = 1, \ldots, m),$$

where the random variables $Z_{\mathbf{k}_{I_j}}$ are given by

$$Z_{\mathbf{k}_{I_j}} = \frac{1}{(2n+1)^{d-d_j}} \sum_{\mathbf{k}_{I_j^c} \in \{-n,\ldots,n\}^{d-d_j}} Y_{\mathbf{k}}. \tag{3.4}$$

The additive estimator is now defined by

$$\hat{\theta}^{add,FD}(\mathbf{x}^*) \;=\; \hat{\theta}_0 + \hat{\theta}^{FD}_{I_1}(\mathbf{x}^*_{I_1}) + \cdots + \hat{\theta}^{FD}_{I_m}(\mathbf{x}^*_{I_m}), \tag{3.5}$$

where

$$\hat{\theta}_0 = \frac{1}{N} \sum_{\mathbf{k} \in \{-n,\ldots,n\}^d} Y_{\mathbf{k}},$$

$$\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}) \;=\; \frac{1}{(2\pi)^{d_j}} \int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w}, \mathbf{x}^*_{I_j} \rangle} \Phi_{K_j}(h\mathbf{w}) \frac{\hat{\Psi}_{g_{I_j}}(\mathbf{w})}{\Phi_{\psi_{I_j}}(\mathbf{w})} d\mathbf{w} \tag{3.6}$$

and K_j is a d_j-dimensional kernel $(j = 1, \ldots, m)$. Note that by the lattice structure the statistic $Z_{\mathbf{k}_{I_j}}$ in (3.4) is a $\sqrt{n^{d-d_j}}$-consistent estimator of $g_{I_j}(\mathbf{z}_{\mathbf{k}_{I_j}})$. Therefore the deconvolution problem for the j-th component is reduced to a problem in dimension d_j and the estimator $\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j})$ can be rewritten as

$$\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}) \;=\; \sum_{\mathbf{k}_{I_j} \in \{-n,\ldots,n\}^{d_j}} Z_{\mathbf{k}_{I_j}} w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j}), \tag{3.7}$$

where the weights $w_{\mathbf{k}_{I_j},n}$ are defined by

$$w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j}) = \frac{1}{(nha_n 2\pi)^{d_j}} \int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w}, (\mathbf{x}^*_{I_j} - \mathbf{z}_{\mathbf{k}_{I_j}}) \rangle / h} \frac{\Phi_{K_j}(\mathbf{w})}{\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})} d\mathbf{w}. \tag{3.8}$$

Remark 3.1. It is worthwhile to mention that under the fixed design assumption the estimator $\hat{\theta}^{FD}_{I_j}$ can be defined explicitly (see equation (3.6)), because observations are available on a regular grid. On the other hand in the random design case the situation is substantially more complicated because – even in

the direct case – the estimator can only be defined implicity as a solution of the backfitting equations, which have to be solved iteratively. A detailed analysis including a proof of the existence and uniqueness of a solution and the geometric convergence of the backfitting equations in direct regression models can be found in Mammen et al. (1999). It is an open and very challenging problem for future research to investigate if these techniques can be combined with Fourier based estimation methods to construct backfitting estimators in inverse additive regression models.

3.2. Technical Assumptions

In order to derive the asymptotic properties of the estimator (3.6) we require the assumptions (A.3) from Section 2.3, which refer to the signal θ, the convolution function ψ and the kernel K. Additionally the following assumptions corresponding to the fixed design are required.

Assumption (B.1). The Fourier transforms $\Phi_{\psi_{I_j}}$ of the marginals ψ_{I_j} of ψ satisfy

$$\int_{\mathbb{R}^{d_j}} \frac{|\Phi_{K_j}(\mathbf{w})|}{|\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})|} d\mathbf{w} \quad \leq \quad C_1 h^{-\beta_j}, \quad \int_{\mathbb{R}^{d_j}} \frac{|\Phi_{K_j}(\mathbf{w})|^2}{|\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} \sim C_2 h^{-2\beta_j}$$

for some $\beta_j > 0$ $(j = 1, \ldots, m)$ and constants $C_1, C_2 > 0$.

Assumption (B.2). Under the assumption of additivity the Fourier transform Φ_{K_j} of the kernel K_j is symmetric and supported on $[-1, 1]^{d_j}$ and there exists a constant $b \in (0, 1]$ such that $\Phi_{K_j}(\mathbf{w}) = 1$ for $\mathbf{w} \in [-b, b]^{d_j}, b > 0$, and $|\Phi_{K_j}(\mathbf{w})| \leq 1$ for all $\mathbf{w} \in [-1, 1]^{d_j}$ for all $j = 1, \ldots, m$.

Remark 3.2.

1. The common assumption on the convolution function ψ in the fixed design case is

$$\Phi_{\psi}(\mathbf{w}) \parallel \mathbf{w} \parallel^{\beta} \to C \quad \mathbf{w} \to \infty, \tag{3.9}$$

   [see Birke and Bissantz (2009)]. Assumption (B.1) is substantially weaker because we do not assume the Fourier transform Φ_{ψ} to be asymptotically radial-symmetric. It is satisfied for many commonly used convolution functions such as the multivariate Laplace density and the density of several Gamma distributions for which (3.9) does not hold.
2. Note that the asymptotic theory for the new estimator in the completely additive case $m = d$ $(d_1 = \cdots = d_m = 1)$ does not require the additive functions to have compact support as it is assumed in Birke et al. (2012).

3.3. Asymptotic properties

In the case of a fixed design on a grid (assumption (FD)) the asymptotic properties of the unrestricted estimator have been studied in Birke and Bissantz (2009)

and in this section we investigate the asymptotic properties of the backfitting estimator introduced in the previous section. Our first result, Lemma 3.3, gives the weak convergence of $\hat{\theta}_{I_j}^{FD}$, whereas Theorem 3.4 contains the asymptotic distribution of the estimator $\hat{\theta}^{add,FD}$ defined in (3.5). The proofs are again deferred to Section 6.

Lemma 3.3. *Consider the inverse regression model under the fixed design assumption (FD). Let Assumptions (B.1), (B.2), (A.3)(C) and (A.3)(D) be fulfilled for some $j \in \{1, \dots, m\}$, $h \to 0$ and $a_n \to 0$ as $n \to \infty$ such that*

$$n^d h^{d_j+2\beta_j} a_n^{d_j} \to \infty \quad \text{and} \quad n^2 h^{2+d_j+\beta_j} a_n^3 \to \infty,$$

then

$$U_{nj}(\boldsymbol{x}_{I_j}^*)^{-1/2}(\hat{\theta}_{I_j}^{FD}(\boldsymbol{x}_{I_j}^*) - E[\hat{\theta}_{I_j}^{FD}(\boldsymbol{x}_{I_j}^*)]) \Rightarrow \mathcal{N}(0,1), \tag{3.10}$$

where the normalizing sequence is defined by

$$U_{nj}(\boldsymbol{x}_{I_j}^*) \quad = \quad \frac{\sigma^2}{(2n+1)^{d-d_j}} \sum_{k_{I_j} \in \{-n,\dots,n\}^{d_j}} w_{k_{I_j},n}(\boldsymbol{x}_{I_j}^*)^2,$$

the weights $w_{k_{I_j},n}$ are defined in (3.8) and

$$E[\hat{\theta}_{I_j}^{FD}(\boldsymbol{x}_{I_j}^*)] = \theta_{I_j}^{add}(\boldsymbol{x}_{I_j}^*) + O(h^{s-1}) + O(n^{-2}h^{-d_j-\beta_j-2}a_n^{-3}).$$

The result of Theorem 3.4 below follows immediately from Lemma 3.3. The bias is of the same order as the bias in Lemma 3.3 and we define $j^* = \operatorname{argmax}_j (d_j + 2\beta_j)$.

Theorem 3.4. *Consider the inverse regression model under the fixed design assumption (FD). Let Assumptions (B.1), (B.2), (A.3)(C) and (A.3)(D) be fulfilled, $h \to 0$ and $a_n \to 0$ as $n \to \infty$ such that*

$$n^d h^{d_{j^*}+2\beta_{j^*}} a_n^{d_{j^*}} \to \infty \quad \text{and} \quad n^2 h^{2+d_{j^*}+\beta_{j^*}} a_n^3 \to \infty.$$

Then

$$U_n(\boldsymbol{x}^*)^{-1/2}(\hat{\theta}^{add,FD}(\boldsymbol{x}^*) - E[\hat{\theta}^{add,FD}(\boldsymbol{x}^*)]) \Rightarrow \mathcal{N}(0,1), \tag{3.11}$$

where the normalizing sequence is defined by

$$U_n(\boldsymbol{x}^*) \quad = \quad \sigma^2 \sum_{\boldsymbol{k} \in \{-n,\dots,n\}^d} \left(\sum_{j=1}^m \frac{1}{(2n+1)^{d-d_j}} w_{k_{I_j},n}(\boldsymbol{x}_{I_j}^*) \right)^2,$$

the weights $w_{k_{I_j},n}$ are defined in (3.8) and

$$E[\hat{\theta}^{add,FD}(\boldsymbol{x}^*)] = \theta^{add}(\boldsymbol{x}^*) + O(h^{s-1}) + O\left(\frac{1}{n^2 h^{2+\max_j d_j + \max_j \beta_j} a_n^3} \right).$$

Remark 3.5.

(1) The normalizing sequence $U_n(\mathbf{x}^*)$ in (3.11) is of order $n^d h^{d_{j^*}+2\beta_{j^*}} a_n^{d_{j^*}}$.
(2) The bias of the additive estimator in the fixed design case is only vanishing if the subsets I_j in the decomposition (2.7) satisfy $d_j \leq 3$ for all $j = 1, \ldots, m$.
(3) Theorem 3.4 can easily be extended to multiplicative models of the form (1.1) with

$$\theta(\mathbf{x}^*) \quad = \quad \prod_{j=1}^{m} \theta_{I_j}(\mathbf{x}^*_{I_j})$$

if the convolution function ψ is also multiplicative. Otherwise the estimator is not consistent and other techniques such as the marginal integration method have to be used.

Remark 3.6. Note that for a one-dimensional predictor the estimator $\hat{\theta}^{FD}$ in (2.2) for the signal θ in the inverse regression model converges with a rate $n^{-d/2} h^{-1/2-\beta} a_n^{-1/2}$ under the fixed design assumption. Consequently if all components in the inverse additive regression model with a d-dimensional predictor were known except for one remaining component, this component could be estimated at rate $N^{-1/2} h^{-1/2-\beta} a_n^{-1/2}$. It follows from Lemma 3.3 (with $d_j = 1$) and its proof in the Appendix that the backfitting estimator $\hat{\theta}^{FD}_{I_j}$ defined in (3.7) converges with the same rate. In other words under the fixed design assumption the backfitting estimator (3.7) attains this oracle rate for each component.

Remark 3.7. The choice of the sequence a_n is a delicate problem. Note that the estimators (2.2) and (2.4) are based on the Fourier inversion formula. This formula is applied to the empirical Fourier-transform, which estimates the Fourier transform

$$\Phi(\mathbf{w}) = \int_{\mathbb{R}^d} g(\mathbf{z}) e^{i\langle \mathbf{w}, \mathbf{z} \rangle} d\mathbf{z}.$$

Thus in order to obtain good estimates it is necessary to observe data, where the signal deviates significantly from 0. If a_N is chosen too large, there appear serious bias problems, while a too small choice will increase the variability of the estimate. Consequently in the cases of a fixed design the cube $[1/a_n, 1/a_n]^d$ should be as small as possible such that it contains a reasonable part of the region, where the function g deviates from 0, say 90% or 95%. In applications it is recommended to investigate d two-dimensional scatterplots of the data and choose the sequence a_n for each component of the predictor appropriately.

In the case of a random design this procedure has to be slightly modified. Here we choose each component of the vector $\mathbf{a_N} = (a_{N1}, \ldots, a_{Nd})^T$ from the empirical quantiles of the corresponding marginal distribution. More precisely, we propose

$$a_{Nj} = \left(\frac{1}{2} (\hat{\mathbb{F}}_j^{-1}(0.5) + \hat{\mathbb{F}}_j(0.95)) \right)^{-1}$$

for the jth coordinate, where $\hat{\mathbb{F}}_j(x)$ denotes the empirical distribution of $X_{j,1},\dots,X_{j,N}$ and $\mathbf{X_k} = (X_{1,k},\dots,X_{d,k})^T$.

4. Dependent data

In this Section we briefly discuss the case of dependent data. To be precise we assume that the errors in the inverse regression models have an MA(q) structure. Under the random design assumption this structure is given by

$$\varepsilon_t = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \tag{4.1}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ denotes a white noise process with variance σ^2. A careful inspection of the proof of Theorem 2.4, which is based on the investigation of the asymptotic properties of cumulants, shows that the result of Theorem 2.4 remains valid under this assumption.

Theorem 4.1.
(1) Consider the inverse regression model (2.1) under the random design assumption (RD). If the Assumptions of Theorem 2.4 are satisfied, then

$$V_1^{-1/2}\big(\hat\theta^{RD}(\boldsymbol{x}^*) - E[\hat\theta^{RD}(\boldsymbol{x}^*)]\big) \Rightarrow \mathcal{N}(0,1), \tag{4.2}$$

where the normalizing sequence is given by

$$\begin{aligned} V_1 &= \frac{1}{Nh^d(2\pi)^{2d}}\int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{s},(\boldsymbol{x}^*/h-\boldsymbol{y})\rangle}\frac{\Phi_K(\boldsymbol{s})}{\Phi_\psi(\frac{\boldsymbol{s}}{h})}d\boldsymbol{s}\Big|^2 \\ &\qquad\times \frac{(\sigma^2\sum_{k,l=0}^q \beta_k\beta_l + g^2(h\boldsymbol{y}))f(h\boldsymbol{y})}{\max\{f(h\boldsymbol{y}), f(\frac{1}{\boldsymbol{a}_N})\}^2}d\boldsymbol{y}, \end{aligned}$$

$\beta_0 = 1$ *and* $E[\hat\theta^{RD}(\boldsymbol{x}^*)] = \theta(\boldsymbol{x}^*) + O(h^{s-1})$.
(2) If the assumptions of Theorem 2.8 are satisfied, then the appropriately standardized additive estimator $\hat\theta^{add,RD}$ *converges weakly to a standard normal distribution, that is*

$$V_3^{-1/2}\big(\hat\theta^{add,RD}(\boldsymbol{x}^*) - E[\hat\theta^{add,RD}(\boldsymbol{x}^*)]\big) \Rightarrow \mathcal{N}(0,1), \tag{4.3}$$

where the standardizing factor is given by

$$\begin{aligned} V_3 &= \frac{1}{N(2\pi)^{2d}}\int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} e^{i\langle \boldsymbol{w},\boldsymbol{s}\rangle}\Big(\sum_{j=1}^m e^{-i\langle \boldsymbol{w}_{I_j},\boldsymbol{x}^*_{I_j}\rangle}L_{I_j^c}(\boldsymbol{w}_{I_j^c})\Big)\frac{\Phi_K(h\boldsymbol{w})}{\Phi_\psi(\boldsymbol{w})}d\boldsymbol{w}\Big|^2 \\ &\qquad\times \frac{(\sigma^2\sum_{k,l=0}^q \beta_k\beta_l + g(\boldsymbol{s})^2)f(\boldsymbol{s})}{\max\{f(\boldsymbol{s}), f(\frac{1}{\boldsymbol{a}_N})\}^2}d\boldsymbol{s}. \end{aligned}$$

and $E[\hat\theta^{add,RD}(\boldsymbol{x}^*)] = \theta^{add}(\boldsymbol{x}^*) + O(h^{s-1})$.

Under the assumption of a fixed design on a grid we consider an error process with an MA(q) structure defined by

$$\varepsilon_{\mathbf{k}} = \sum_{\mathbf{r}\in\{-q,\ldots,q\}^d} \beta_{\mathbf{r}} Z_{\mathbf{k}-\mathbf{r}}, \tag{4.4}$$

where $\{Z_{\mathbf{j}}\}_{j\in\mathbb{Z}^d}$ are i.i.d. random variables with mean zero and variance σ^2. This means, that the noise terms are influenced by all shocks, which have a distance on the lattice lower or equal q regarding the ∞-norm. The following result can be obtained by similar arguments as used for the proof of Theorem 3.4.

Theorem 4.2. *Consider the inverse regression model (3.1) under the fixed design assumption with an MA(q) dependent error process of the form (4.4). If the assumptions of Theorem 3.3 are satisfied we have*

$$V_{MA}^{-1/2}(\boldsymbol{x}^*)\left(\hat{\theta}^{add,FD}(\boldsymbol{x}^*) - E[\hat{\theta}^{add,FD}(\boldsymbol{x}^*)]\right) \Rightarrow \mathcal{N}(0,1),$$

where the normalizing sequence is given by

$$V_{MA}(\boldsymbol{x}^*) = \sigma^2 \sum_{\substack{\boldsymbol{l}\in\mathbb{Z}^d \\ \|\boldsymbol{l}\|_\infty \le 2q}} \sum_{\boldsymbol{r}_1\in\{-q,\ldots,q\}^d} \beta_{\boldsymbol{r}_1}\beta_{\boldsymbol{l}+\boldsymbol{r}_1} \times \sum_{\boldsymbol{k}\in\{-n,\ldots,n\}^d} \left|\sum_{j=1}^m \frac{1}{(2n+1)^{d-d_j}} w_{\boldsymbol{k}_{I_j},n}(\boldsymbol{x}^*_{I_j})\right|^2.$$

and $E[\hat{\theta}^{add,FD}(\boldsymbol{x}^*)] = \theta^{add}(\boldsymbol{x}^*) + O(h^{s-1}) + O(\frac{1}{n^2 h^{2+\max_j d_j + \max_j \beta_j} a_n^3})$.

Remark 4.3. If $\varepsilon_{\mathbf{t}}$ has an MA(∞) representation Theorems 4.1 and 4.2 will not hold in general, because without additional assumptions the l-th cumulant of the normalized statistic does not converge to zero for all $l \ge 3$.

5. Finite sample properties

In this Section we investigate the finite sample properties of the new estimators and also provide a comparison with competing methods. We first investigate the case of a fixed design in model (1.1) with the convolution function

$$\psi(x_1,x_2) = \frac{9}{4} e^{-3(|x_1|+|x_2|)},$$

and two additive signals

$$\theta^{(1)}(x_1,x_2) = e^{-(x_1-0.1)^2} + e^{-(x_2-0.4)^2} \tag{5.1}$$

$$\theta^{(2)}(x_1,x_2) = e^{-|x_1-0.4|} + 2e^{-2x_2^2} \tag{5.2}$$

For the kernel K in the Fourier transform Φ_K we use the kernel $K(\mathbf{x}) = \frac{\sin(x_1)\sin(x_2)}{\pi^2 x_1 x_2}$. We consider a fixed design on the grid $\{(\frac{k_1}{na_n}, \frac{k_2}{na_n} | k_1,k_2 \in \{-n,\ldots,n\}\}$ with $N=(2n+1)^2$ points where $n \in \{30,50\}$. In both cases we choose the design parameter as $a_n = 0.25$, such that the cube $[\frac{-1}{a_n}, \frac{1}{a_n}]^2$ covers most

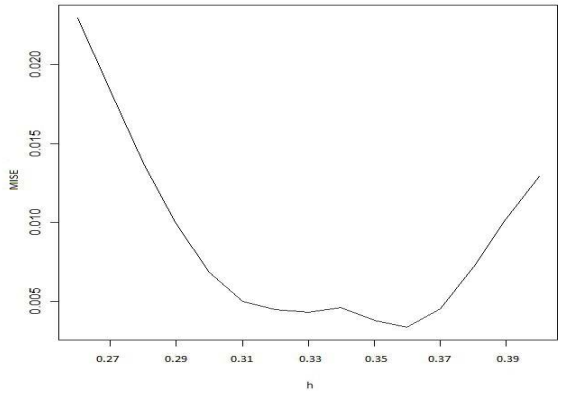

FIG 1. *Mean integrated squared error of the estimator* $\hat{\theta}^{add,FD}$ *for different bandwidths in model* (5.1)*, where* $\sigma = 0.5$.

of the region where the functions $\theta^{(1)}$ and $\theta^{(2)}$ deviate significantly from 0. In all simulations we use (independent) noise terms, which are normal distributed with mean 0 and variance 0.25.

The bandwidth h in the estimator (3.6) is chosen such that the mean integrated squared error (MISE)

$$\mathbb{E}\Big[\int_{\mathbb{R}^2} (\hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}))^2 d\mathbf{x}\Big]$$

is minimized. Figure 1 shows a typical example of the MISE as a function of the bandwidth h. Figure 2 shows the contour plot of the function $\theta^{(1)}$ defined in (5.1) and contour plots of three typical additive estimates where $n = 50$ and the bandwidths are chosen as $h = 0.32, 0.36, 0.4$ (the bandwidth $h = 0.36$ minimizes the MISE). We observe that the shapes in all figures are very similar. The bandwidths $h = 0.32$ and $h = 0.4$ yield stronger deviations from the true function especially at the boundary, but the main structure is even for these choices still recovered. Because other simulations showed a similar picture we conclude that small changes in the bandwidth do not effect the general structure of the estimator significantly.

In order to investigate the finite sample properties of the new estimate $\hat{\theta}^{add,FD}$ defined in (3.5) we performed 1000 iterations with the signal $\theta^{(2)}$ (the results for the signal $\theta^{(1)}$ are similar and are not depicted for the sake of brevity). The simulated mean, variance and mean squared error (MSE) of $\hat{\theta}^{add,FD}$ are given in Table 1 for different choices of $\mathbf{x} = (x_1, x_2)$ where the sample size is $N = 10201$ and the variance of the errors is 0.25. We observe that in most cases the mean squared error is dominated by the bias.

In the second part of this section we compare three different estimates for the signal in the inverse regression model (1.1). The first estimate for θ is the

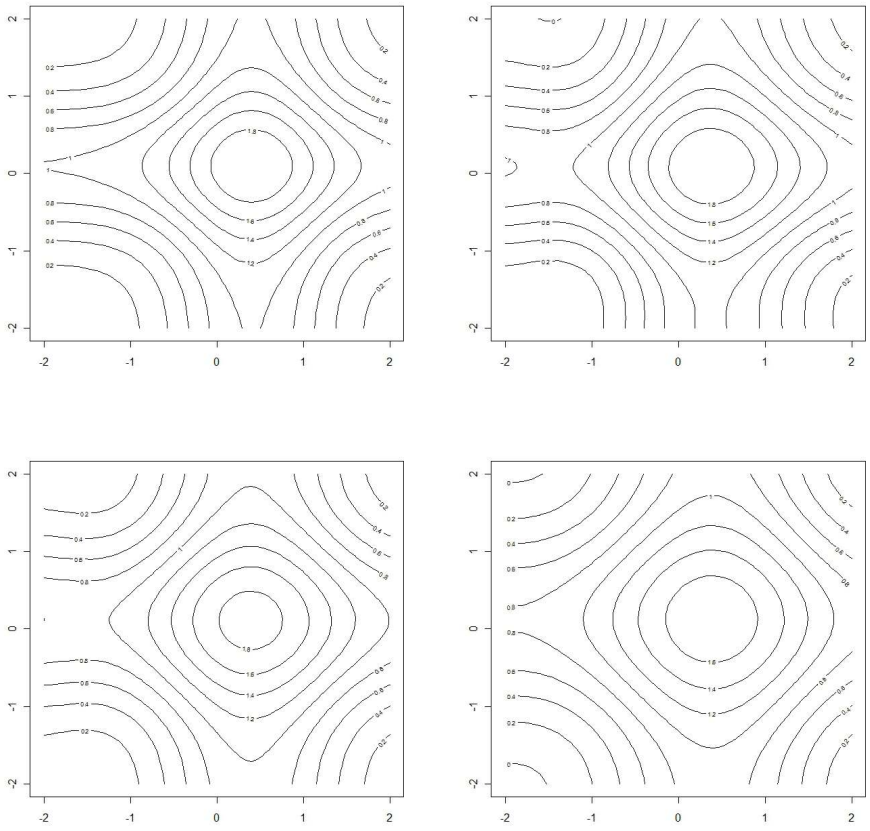

FIG 2. *Contour plot of the function* $\theta^{(1)}$ *defined in* (5.1) *(left upper panel) and its estimates* $\hat{\theta}^{add,FD}$ *defined in* (3.5) *with different bandwidths. Upper right panel:* $h = 0.32$*; Lower left panel:* $h = 0.36$ *(which minimizes the MISE); Lower right panel:* $h = 0.4$.

statistic $\hat{\theta}^{add,FD}$ proposed in this paper [see formula (3.5)]. The second method is the marginal integration estimator suggested by Birke et al. (2012) and the third method is the non additive estimate of Birke and Bissantz (2009). The results are shown in Table 2 for the sample size $N = 3721$ and selected values of the predictor. We observe that the additive estimate of Birke et al. (2012) improves the unrestricted estimate with respect to mean squared error by 20–50%. However, the new additive estimate $\hat{\theta}^{add,FD}$ yields a much larger improvement. The MSE is about 14 and 7–10 times smaller than the MSE obtained by the unrestricted estimator or the estimator proposed by Birke et al. (2012). Further simulations for the signal $\theta^{(2)}$ in (5.2) show similar results and are not depicted for the sake of brevity.

TABLE 1
Mean, variance and mean squared error of the new additive estimator $\hat{\theta} = \hat{\theta}^{add,FD}$ in the case of a fixed design. The model is given by (5.2) with variance $\sigma^2 = 0.25$

N	x_1	x_2	$\theta^{(2)}(\mathbf{x})$	$E[\hat{\theta}(\mathbf{x})]$	$\mathrm{Var}(\hat{\theta}(\mathbf{x}))$	$\mathrm{MSE}(\hat{\theta}(\mathbf{x}))$
10201	−1.6	−1.6	0.1473	0.2522	0.0017	0.0127
		−0.8	0.3131	0.3805	0.0017	0.0063
		0	0.6823	0.8296	0.0017	0.0234
		0.8	0.6823	0.8159	0.0017	0.0195
		1.6	0.3131	0.3827	0.0017	0.0065
10201	−0.8	−1.6	0.6914	0.8216	0.0017	0.0187
		−0.8	0.8573	0.9446	0.0018	0.0094
		0	1.2264	1.3977	0.0017	0.0310
		0.8	1.2264	1.3864	0.0017	0.0273
		1.6	0.8573	0.9496	0.0018	0.0103
10201	0	−1.6	2.1353	2.1887	0.0018	0.0046
		−0.8	2.3012	2.3123	0.0017	0.0018
		0	2.6703	2.7640	0.0018	0.0106
		0.8	2.6703	2.7548	0.0016	0.0087
		1.6	2.3012	2.3178	0.0018	0.0020
10201	0.8	−1.6	0.6914	0.8181	0.0017	0.0178
		−0.8	0.8573	0.9445	0.0018	0.0094
		0	1.2264	1.3967	0.0017	0.0307
		0.8	1.2264	1.3864	0.0017	0.0273
		1.6	0.8573	0.9496	0.0018	0.0103
10201	1.6	−1.6	0.1473	0.2532	0.0016	0.0128
		−0.8	0.3131	0.3785	0.0017	0.0060
		0	0.6823	0.8290	0.0018	0.0233
		0.8	0.6823	0.8168	0.0019	0.0200
		1.6	0.3131	0.3855	0.0017	0.0069

For the sake of comparison, the first two rows of Table 2 contain results of the estimators $\hat{\theta}^{RD}$ and $\hat{\theta}^{add,RD}$, where the explanatory variables follow a uniform distribution on the same cube $[\frac{1}{a_n}, \frac{1}{a_n}]^2$ as used for the fixed design. We observe a similar behaviour of the unrestricted estimators under the fixed and random design assumption. This corresponds to the asymptotic theory, which shows that in the case of a uniform distribution the unrestricted estimators converge with the same rate of convergence (see Remark 2.5). On the other hand, the additive estimator $\hat{\theta}^{add,RD}$ produces a substantially larger mean squared error compared to the estimator $\hat{\theta}^{add,FD}$, which is of similar size as the mean squared error of the estimator proposed by Birke et al. (2012).

Because the performance of the estimators depends on the correct specification of the convolution function ψ we next investigate the performance of the estimators under misspecification of the function ψ. In Figure 3 we display the contour plots of the estimates $\hat{\theta}^{add,FD}$, where in every panel the convolution function is misspecificated as Laplace distribution $Lap(\alpha, \beta)$ with parameters $\alpha = 0$ and $\beta = \frac{1}{3}$. In the upper left and upper right panel the parameter β of the Laplace distribution $Lap(\alpha, \beta)$ is misspecificated, whereas in the lower left panel the true convolution function is the density of a standard normal distribution and in the lower right panel it is a gamma distribution. We observe, that a miss-specification of the shape of the convolution function (as it occurs if a Laplace density is used instead of the density of a Gamma(3,2) distribution)

TABLE 2
Mean, variance and mean squared error of the unrestricted estimator $\hat{\theta}^{FD}$ proposed in Birke and Bissantz (2009), the estimator $\hat{\theta}^{BBH}$ proposed by Birke et al. (2012) and the new estimators $\hat{\theta}^{RD}$, $\hat{\theta}^{add,RD}$ and $\hat{\theta}^{add,FD}$ proposed in this paper. The model is given by (5.1), where $\sigma^2 = 0.25$

	N	x_1	x_2	$\theta^{(1)}(\mathbf{x})$	$\mathbb{E}\hat{\theta}(\mathbf{x})$	Var $\hat{\theta}(\mathbf{x})$	MSE $\hat{\theta}(\mathbf{x})$
$\hat{\theta}^{RD}$	3721	0	0	1.8422	1.9667	0.0516	0.0671
	3721	0	1	1.6877	1.6983	0.0458	0.0459
	3721	1	1	1.1425	1.1909	0.0329	0.0352
	3721	1	1.8	0.5857	0.6624	0.0189	0.0248
$\hat{\theta}^{add,RD}$	3721	0	0	1.8422	1.8680	0.0440	0.0301
	3721	0	1	1.6877	1.6405	0.0195	0.0217
	3721	1	1	1.1425	1.3371	0.0232	0.0610
	3721	1	1.8	0.5857	0.8184	0.0199	0.0740
$\hat{\theta}^{FD}$	3721	0	0	1.8422	1.8123	0.0426	0.0435
	3721	0	1	1.6877	1.7305	0.0425	0.0443
	3721	1	1	1.1425	1.2143	0.0418	0.0470
	3721	1	1.8	0.5857	0.4774	0.0416	0.0533
$\hat{\theta}^{add,FD}$	3721	0	0	1.8422	1.8234	0.0027	0.0031
	3721	0	1	1.6877	1.6589	0.0024	0.0032
	3721	1	1	1.1425	1.1097	0.0025	0.0036
	3721	1	1.8	0.5857	0.5494	0.0023	0.0036
$\hat{\theta}^{BBH}$	3721	0	0	1.8422	1.8874	0.0194	0.0214
	3721	0	1	1.6877	1.7316	0.0191	0.0210
	3721	1	1	1.1425	1.1833	0.0201	0.0218
	3721	1	1.8	0.5857	0.4438	0.0207	0.0408

yields to an estimator with a different structure as the true signal (see the lower right panel in Figure 3). All other panels show the same structure as the upper left panel Figure 2 which gives the contour plot of the true signal $\theta^{(1)}$. This indicates that the structure of the signal can be reconstructed, as long as the chosen convolution kernel exhibits similar modal properties as the "true kernel". However, we also observe from Figure 3 that the levels of the contour differ from those of the true signal.

It was pointed out by a referee that it might also be of interest to investigate the properties of the estimates under misspecification of the density f and corresponding results are shown in Figure 4. All estimates are calculated under the assumption of a $\mathcal{N}(0, I_2)$ distribution for density f. The upper left panel shows the corresponding contour plots of the estimate for the first component in model (5.1) if this density is in fact the correct density. Results for a misspecification in the mean and a factor for the covariance matrix are shown in the upper right panel and lower left panel of Figure 4, where f is the density of a $\mathcal{N}(\mu, I_2)$ distribution with $\mu = (0.1, 0.1)^T$ and a $\mathcal{N}(0, 0.9I_2)$ distribution, respectively. We observe that the structure of the estimates is still preserved. On the other hand, if the true density is a $\mathcal{N}(0, \Sigma)$ distribution with

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix} \tag{5.3}$$

the estimate of $\theta^{(1)}$ does not reflect the main features of the signal.

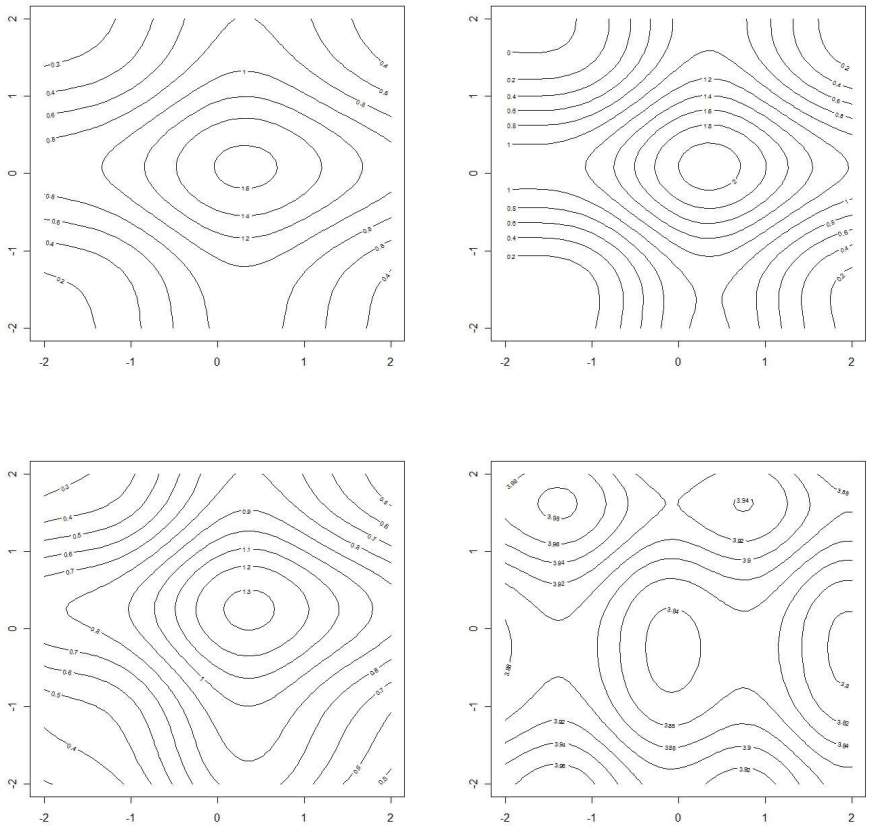

FIG 3. *Contour plot of the estimate* $\hat{\theta}^{add,FD}$ *of* $\theta^{(1)}$ *with misspecificated convolution function. Upper left panel:* ψ *misspecificated as Lap(0,*$\frac{1}{3}$*), where the true convolution function is Lap(0,1); Upper right panel:* ψ *misspecificated as Lap(0,*$\frac{1}{3}$*), where the true convolution function is Lap(0,*$\frac{1}{5}$*); Lower left panel:* ψ *misspecificated as Lap(0,*$\frac{1}{3}$*), where the true convolution function is* $\mathcal{N}(0,1)$*; Lower right panel:* ψ *misspecificated as Lap(0,*$\frac{1}{3}$*), where the true convolution function is Gamma(3,2). The model is given by* (5.1)*, where* $\sigma^2 = 0.25$.

We conclude this section with a brief discussion of the performance of the unrestricted estimator $\hat{\theta}^{RD}$ under the assumption (RD) of a non-uniform random design. In Table 3 we display the simulated mean, variance and mean squared error for various distributions of the predictor $\mathbf{X}$, where the components are independent and identically distributed. In most cases we observe similar results for the bias, independently of the distribution of $\mathbf{X}$ and the choice of the sequence a_n. On the other hand the mean squared error is dominated by the variance, which depends sensitively on the choice of the parameter a_n. This observation corresponds with the representation of the asymptotic variance of

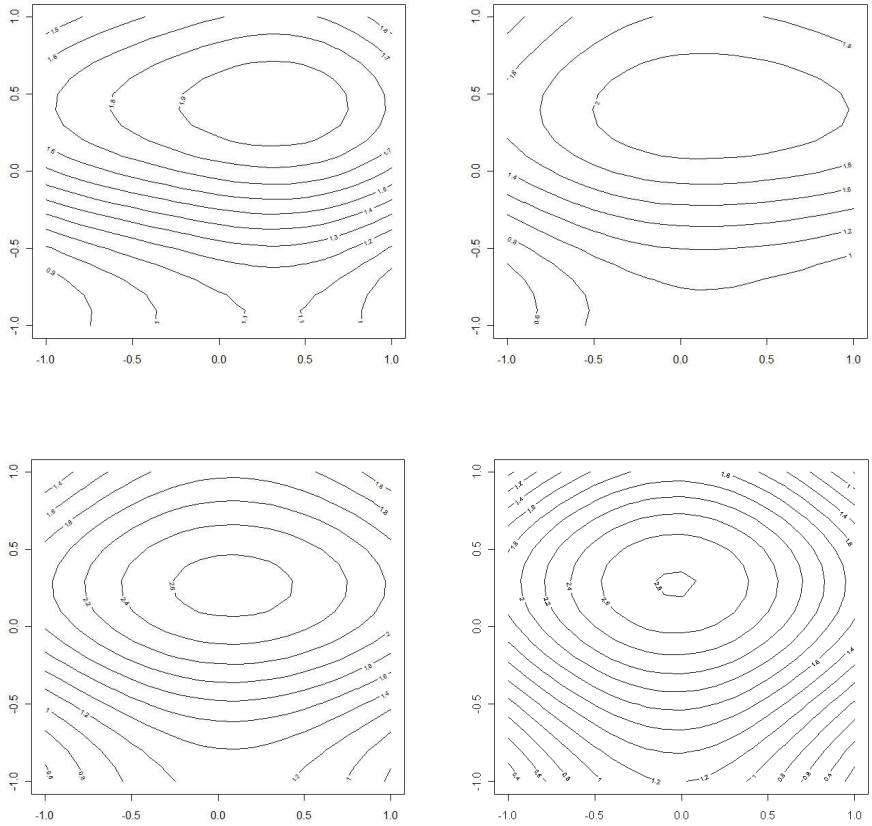

FIG 4. *Contour plot of the estimate* $\hat{\theta}^{add,FD}$ *of* $\theta^{(1)}$ *with misspecified density function. Upper left panel: density* f *correctly specified as* $\mathcal{N}(0, I_2)$*; Upper right panel:* f *misspecified as* $\mathcal{N}(0, I_2)$*, where the true convolution function is* $\mathcal{N}(\mu, I_2)$ *and* $\mu = (0.1, 0.1)^T$*; Lower left panel:* f *misspecified as* $\mathcal{N}(0, I_2)$*, where the true convolution function is* $\mathcal{N}(0, 0.9I_2)$*; Lower right panel:* f *misspecified as* $\mathcal{N}(0, I_2)$*, where the true convolution function is* $\mathcal{N}(0, \Sigma)$ *and* Σ *is defined in* (5.3)*. The model is given by* (5.1)*, where* $\sigma^2 = 0.25$*.*

$\hat{\theta}^{RD}$ in formula (2.17) of Theorem 2.4. We also observe that the impact of the distribution of the explanatory variable on the variance of the estimate $\hat{\theta}^{RD}$ is much smaller.

Acknowledgements

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C1, C4) of the German Research Foundation (DFG). We would like to thank

TABLE 3
Mean, variance and mean squared error of the unrestricted estimator $\hat{\theta}^{RD}$ proposed in this paper for different distributions of the explanatory variables $\mathbf{X}$ and different choices for the parameter a_n. The model is given by (5.1) and the variance is $\sigma^2 = 0.25$

$\mathbf{X}$	N	a_n	x_1	x_2	$\theta^{(1)}(\mathbf{x})$	$\mathbb{E}\hat{\theta}(\mathbf{x})$	Var $\hat{\theta}(\mathbf{x})$	MSE $\hat{\theta}(\mathbf{x})$
$U[\frac{-1}{a_n}, \frac{1}{a_n}]$	10201	0.25	0	0	1.8422	1.7421	0.0297	0.0397
	10201	0.25	0	1	1.6877	1.7163	0.0272	0.0283
	10201	0.25	1	1	1.1425	1.2858	0.0194	0.0399
	10201	0.25	1	1.8	0.5857	0.6105	0.0117	0.0123
$U[\frac{-1}{a_n}, \frac{1}{a_n}]$	10201	0.5	0	0	1.8422	1.4957	0.0076	0.1277
	10201	0.5	0	1	1.6877	1.8123	0.0070	0.0225
	10201	0.5	1	1	1.1425	1.5438	0.0044	0.1654
	10201	0.5	1	1.8	0.5857	0.5695	0.0023	0.0026
$N(0,1)$	10201	0.25	0	0	1.8422	1.8512	0.3271	0.3271
	10201	0.25	0	1	1.6877	1.7019	0.7098	0.7100
	10201	0.25	1	1	1.1425	1.2038	0.7077	0.7115
	10201	0.25	1	1.8	0.5857	0.5983	0.4477	0.4479
$N(0,1)$	10201	0.5	0	0	1.8422	1.8229	0.0079	0.0083
	10201	0.5	0	1	1.6877	1.7466	0.0107	0.0143
	10201	0.5	1	1	1.1425	1.2531	0.0114	0.0236
	10201	0.5	1	1.8	0.5857	0.6366	0.0135	0.0161
$t(2)$	10201	0.25	0	0	1.8422	1.8758	0.0174	0.0185
	10201	0.25	0	1	1.6877	1.7129	0.0255	0.0261
	10201	0.25	1	1	1.1425	1.1786	0.0271	0.0284
	10201	0.25	1	1.8	0.5857	0.6138	0.0324	0.0332
$t(2)$	10201	0.5	0	0	1.8422	1.8590	0.0115	0.0118
	10201	0.5	0	1	1.6877	1.7260	0.0158	0.0173
	10201	0.5	1	1	1.1425	1.2069	0.0182	0.0223
	10201	0.5	1	1.8	0.5857	0.6275	0.0174	0.0191

Alina Dette and Martina Stein who wrote parts of this paper with considerable technical expertise. The authors are also grateful to three unknown referees and the associate editor for their constructive comments on an earlier version of this paper.

6. Appendix

For the proofs we make frequent use of the cumulant method, which is a common tool in time series analysis. Following Brillinger (2001) the r-th order joint cumulant $cum(Y_1, \ldots, Y_r)$ of a r-dimensional complex valued random vector $(Y_1, \ldots, Y_r)$ is given by

$$cum(Y_1, \ldots, Y_r) = \sum (-1)^{p-1}(p-1)! \Big(E \prod_{j \in \nu_1} Y_j \Big) \ldots \Big(E \prod_{j \in \nu_p} Y_j \Big), \qquad (6.1)$$

where we assume the existence of moments of order r, i.e. $E(|Y_j^r|) < \infty$ $(j = 1, \ldots, r)$ and the summation extends over all partitions $(\nu_1, \ldots, \nu_p), p = 1, \ldots, r$ of $(1, \ldots, r)$. If we choose $Y_j = Y, j = 1, \ldots, r$ we denote with $cum_r(Y) = cum(Y, \ldots, Y)$ the r-th order cumulant of a univariate random variable. The following properties of the cumulant will be used frequently in our proofs [see e.g. Brillinger (2001)].

(B1) $cum(a_1Y_1, \ldots, a_rY_r) = a_1 \ldots a_r cum(Y_1, \ldots, Y_r)$ for constants $a_1, \ldots, a_r \in \mathbb{C}$
(B2) if any group of the Y's is independent of the remaining Y's, then $cum(Y_1, \ldots, Y_r) = 0$
(B3) for the random variable $(Z_1, Y_1, \ldots, Y_r)$ we have

$$cum(Z_1 + Y_1, Y_2, \ldots, Y_r) = cum(Z_1, Y_2, \ldots, Y_r) + cum(Y_1, Y_2, \ldots, Y_r)$$

(B4) if the random variables $(Y_1, \ldots, Y_r)$ and $(Z_1, \ldots, Z_r)$ are independent, then

$$cum(Y_1 + Z_1, \ldots, Y_r + Z_r) = cum(Y_1, \ldots, Y_r) + cum(Z_1, \ldots, Z_r)$$

(B5) $cum(Y_j) = E(Y_j)$ for $j = 1, \ldots, r$
(B6) $cum(Y_j, \overline{Y}_j) = Var(Y_j)$ for $j = 1, \ldots, r$

We finally state a result which can easily be proved by using the definition (6.1) and the properties of the mean.

Theorem 6.1. *Let $\boldsymbol{Y} = (Y_1, \ldots, Y_r)$ be a random variable, b_n a sequence and $C > 0$ a constant with*

$$E\Big[\prod_{j=1}^{l} |Y_{i_j}|\Big] \leq C^l b_n^l \qquad \text{for all } 1 \leq l \leq r,$$

then $|cum(Y_{i_1}, \ldots, Y_{i_m})| \leq (m-1)! C^m b_n^m \sum_{j=1}^{m} S_{m,j}$, where $S_{m,j}$ denotes the Sterling number of the second kind.

We will also make use of the fact that the normal distribution with mean μ and variance σ^2 is characterized by its cumulants, where the first two cumulants are equal to μ and σ^2 respectively and all cumulants of larger order are zero. To show asymptotic normality in our proofs we have to calculate the first two cumulants which give the asymptotic mean and variance and show in a second step that all cumulants of order $l \geq 3$ are vanishing asymptotically. In the following discussion all constants which do not depend on the sample size (but may differ in different steps of the proofs) will be denoted by C.

6.1. Proof of Theorem 2.4

For the sake of brevity we write $\hat{\theta}$ instead of $\hat{\theta}^{RD}$ throughout this proof. By the discussion of the previous paragraph we have to calculate the mean and the variance of $\hat{\theta}(\mathbf{x}^*)$ and all cumulants of order $l \geq 3$. We start with the mean conditional on $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, which can be calculated as

$$E[\hat{\theta}(\mathbf{x}^*)|\mathbf{X}] = \sum_{k=1}^{N} g(\mathbf{X}_k) w_N(\mathbf{x}^*, \mathbf{X}_k),$$

where the weights w_N are defined in (2.5). By iterative expectation we get

$$E[\hat{\theta}(\mathbf{x}^*)] = \frac{1}{h^d(2\pi)^d}\int_{\mathbb{R}^d} g(\mathbf{x})\int_{\mathbb{R}^d} e^{-i\langle \mathbf{s},(\mathbf{x}^*-\mathbf{x})\rangle/h}\frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} \times \frac{f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}}d\mathbf{s}d\mathbf{x},$$

which yields a bias of the form $bias_{\hat{\theta}} = E[\hat{\theta}(\mathbf{x}^*)] - \theta(\mathbf{x}^*) = A_1 + A_2$, where (note that $\Phi_g = \Phi_\psi \cdot \Phi_\theta$)

$$\begin{aligned} A_1 &= \frac{1}{h^d(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle \mathbf{s},\mathbf{x}^*\rangle/h}\Phi_K(\mathbf{s})\Phi_\theta\left(\frac{\mathbf{s}}{h}\right)d\mathbf{s} - \theta(\mathbf{x}^*), \\ A_2 &= \frac{1}{h^d(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle \mathbf{s},\mathbf{x}^*\rangle/h}\frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})}\int_{\mathbb{R}^d} g(\mathbf{x})e^{i\langle s,\mathbf{x}\rangle/h} \\ &\quad \times \Big(\frac{f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} - 1\Big)d\mathbf{x}d\mathbf{s}. \end{aligned}$$

For the summand A_1 we can use exactly the same calculation as in Birke and Bissantz (2009) to obtain $A_1 = O(h^{s-1})$. For the second term A_2 we have

$$\begin{aligned} A_2 &\leq \frac{1}{h^d(2\pi)^d}\int_{\mathbb{R}^d}\frac{|\Phi_K(\mathbf{s})|}{|\Phi_\psi(\frac{\mathbf{s}}{h})|}\int_{\mathbb{R}^d}|g(\mathbf{x})|\Big|\frac{f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} - 1\Big|d\mathbf{x}d\mathbf{s} \\ &\leq \frac{C}{h^{d+\beta}(2\pi)^d}\int_{([-\frac{1}{a_N},\frac{1}{a_N}]^d)^c}|g(\mathbf{x})|\Big|\frac{f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} - 1\Big|d\mathbf{x}, \end{aligned}$$

where we used Assumptions (A.1)(A) and (A.4) in the last inequality. In the next step we will use the fact that $0 \leq \frac{f(\mathbf{x})}{\max\left\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\right\}} \leq 1$ $(\mathbf{x} \in \mathbb{R}^d)$ and Assumption (A.3)(B) to obtain

$$A_2 \leq \frac{C}{h^{d+\beta}(2\pi)^d}\int_{([-\frac{1}{a_N},\frac{1}{a_N}]^d)^c}|g(\mathbf{x})|\parallel \mathbf{x}\parallel^r\frac{1}{\parallel \mathbf{x}\parallel^r}d\mathbf{x} = O\Big(\frac{a_N^r}{h^{d+\beta}}\Big) = O(h^{s-1}).$$

This shows that the bias of $\hat{\theta}(\mathbf{x}^*)$ is of order $O(h^{s-1})$. By the definition of $\hat{\theta}(\mathbf{x}^*)$ and (2.5) it follows

$$V(\hat{\theta}(\mathbf{x}^*)|\mathbf{X}) = \frac{\sigma^2}{N^2h^{2d}(2\pi)^{2d}}\sum_{k=1}^n\Big|\int_{\mathbb{R}^d} e^{-i\langle \mathbf{s},(\mathbf{x}^*-\mathbf{X}_k)\rangle/h}\frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})}d\mathbf{s}\Big|^2 \times \frac{1}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}^2}$$

which yields

$$E[V(\hat{\theta}(\mathbf{x}^*)|\mathbf{X})] = \frac{\sigma^2}{Nh^d(2\pi)^{2d}}\int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} e^{-i\langle \mathbf{s},(\mathbf{x}^*/h-y)\rangle}\frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})}d\mathbf{s}\Big|^2 \times \frac{f(h\mathbf{y})}{\max\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\}^2}d\mathbf{y}.$$

The variance of the conditional expectation is given by (observe again the definition of the weight w_N in (2.5))

$$\begin{aligned}
&V(E[\hat{\theta}(\mathbf{x}^*)|\mathbf{X}]) \\
&\quad = V\left(\sum_{k=1}^{N} g(\mathbf{X}_k) w_n(\mathbf{x}^*, \mathbf{X}_k)\right) \\
&\quad = \frac{1}{Nh^d(2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^*/h - \mathbf{y})\rangle} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} d\mathbf{s} \Big|^2 \\
&\qquad\qquad \times \frac{g^2(h\mathbf{y}) f(h\mathbf{y})}{\max\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{y} \\
&\quad - \frac{1}{N(2\pi)^{2d}} \Big| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^*/h - \mathbf{y})\rangle} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} d\mathbf{s} \frac{g(h\mathbf{y}) f(h\mathbf{y})}{\max\left\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\right\}} d\mathbf{y} \Big|^2,
\end{aligned}$$

where the second summand is of order $O(N^{-1})$. Thus the variance can be written as

$$\begin{aligned}
&V(\hat{\theta}(\mathbf{x}^*)) \\
&\quad = E[V(\hat{\theta}(\mathbf{x}^*)|\mathbf{X})] + V(E[\hat{\theta}(\mathbf{x}^*)|\mathbf{X}]) \\
&\quad = \frac{1}{Nh^d(2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^*/h - y)\rangle} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} d\mathbf{s} \Big|^2 \frac{(\sigma^2 + g^2(h\mathbf{y})) f(h\mathbf{y})}{\max\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{y} \\
&\quad + O(N^{-1})
\end{aligned} \tag{6.2}$$

and the rate of convergence has a lower bound given by

$$V(\hat{\theta}(\mathbf{x}^*))^{-1/2} = \Omega\big(N^{1/2} h^{\beta + d/2} f(\mathbf{a_N}^{-1})^{1/2}\big),$$

where the symbol $b_N = \Omega(c_N)$ means that there exists a constant C and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ we have $|b_N| \geq C|c_N|$. The variance has a lower bound

$$\begin{aligned}
V(\hat{\theta}(\mathbf{x}^*)) &\geq \frac{1}{Nh^d(2\pi)^{2d}} \int_{([\frac{-1}{ha_N}, \frac{1}{ha_N}]^d)} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^*/h - y)\rangle} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} d\mathbf{s} \Big|^2 \\
&\qquad\qquad \times \frac{(\sigma^2 + g^2(h\mathbf{y})) f(h\mathbf{y})}{f(h\mathbf{y})^2} d\mathbf{y} \\
&\geq \frac{C}{Nh^d(2\pi)^{2d}} \int_{([\frac{-1}{ha_N}, \frac{1}{ha_N}]^d)} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^*/h - y)\rangle} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} d\mathbf{s} \Big|^2 d\mathbf{y} \\
&= C(Nh^{d+2\beta})^{-1}(1 + o(1)),
\end{aligned}$$

where we used Assumption (A.4) and Parsevals equality. This yields to the upper bound

$$V(\hat{\theta}(\mathbf{x}^*))^{-1/2} \quad = \quad O\left(N^{1/2} h^{\beta + d/2}\right) \tag{6.3}$$

For the proof of asymptotic normality we now show that the l-th cumulant of

$$G_l = \big| cum_l\big(V(\hat{\theta}(x^*))^{-1/2} \hat{\theta}(\mathbf{x}^*)\big)\big|$$

is vanishing asymptotically, whenever $l \geq 3$. For this purpose we recall the definition of the weights w_N in (2.5) and obtain from (6.3) the estimate

$$\begin{aligned} G_l &\leq CN^{l/2}h^{l\beta+dl/2} \sum_{k_1,\ldots,k_l=1}^{N} |cum\big(Y_{k_1}w_N(\mathbf{x}^*,\mathbf{X}_{k_1}),\ldots,Y_{k_l}w_N(\mathbf{x}^*,\mathbf{X}_{k_l})\big)| \\ &= CN^{l/2}h^{l\beta+dl/2} \sum_{k=1}^{N} |cum_l\big(Y_k w_N(\mathbf{x}^*,\mathbf{X}_k)\big)| \\ &= CN^{l/2+1}h^{l\beta+dl/2} \sum_{\mathbf{j}\in\{0,1\}^l} |(cum(U^{j_1}w_N(\mathbf{x}^*,\mathbf{X}_1),\ldots,U^{j_l}w_N(\mathbf{x}^*,\mathbf{X}_1))|, \end{aligned} \tag{6.4}$$

where we used (B2) and the notation $U^0 = g(\mathbf{X}_1)$ and $U^1 = \varepsilon$. This term can be written as

$$CN^{l/2+1}h^{l\beta+dl/2} \sum_{s=0}^{l} \binom{l}{s} \sum_{\substack{\mathbf{j}\in\{0,1\}^l \\ j_1+\cdots+j_l=s}} \big|cum(U^{j_1}w_N(\mathbf{x}^*,\mathbf{X}_1),\ldots,U^{j_l}w_N(\mathbf{x}^*,\mathbf{X}_1))\big|.$$

By using the product theorem for cumulants [see e.g. Brillinger (2001)], we obtain

$$CN^{l/2+1}h^{l\beta+dl/2} \sum_{s=0}^{l} \binom{l}{s} \sum_{\substack{\mathbf{j}\in\{0,1\}^l \\ j_1+\cdots+j_l=s}} \Big|\sum_{\nu}\prod_{k=1}^{p} cum(A_{ij}, ij \in \nu_k)\Big|, \tag{6.5}$$

where the third sum is calculated over all indecomposable partitions $\nu = (\nu_1,\ldots,\nu_p)$ of the table

$$\begin{array}{ccc} A_{i1} & & A_{i2} \\ \vdots & & \vdots \\ A_{i1} & & A_{i2} \\ & A_{ij} & \\ & \vdots & \\ & A_{ij} & \end{array}$$

(here the first s rows have two and the last $l-s$ rows have one column) and

$$\begin{aligned} A_{i1} &= \varepsilon && 1 \leq i \leq s \\ A_{i2} &= w_N(\mathbf{x}^*,\mathbf{X}_1)) && 1 \leq i \leq s \\ A_{ij} &= g(\mathbf{X}_1)w_N(\mathbf{x}^*,\mathbf{X}_1)) && s+1 \leq i \leq l. \end{aligned}$$

As ε is independent of $\mathbf{X}$ only those indecomposable partitions yield a non zero cumulant, which separate all ε's from the other terms. This means that for a partition ν there are $m(\nu)$ sets $\nu_1,\ldots,\nu_{m(\nu)}$ which include only $\varepsilon's$ while

$\nu_{m(\nu)+1}, \ldots, \nu_p$ contain only $w_N(\mathbf{x}^*, \mathbf{X})$'s and $g(\mathbf{X})w_N(\mathbf{x}^*, \mathbf{X})$'s. Thus the term (6.5) can be written as

$$CN^{l/2+1}h^{l\beta+dl/2}\sum_{s=0}^{l}\binom{l}{s}\sum_{\substack{\mathbf{j}\in\{0,1\}^l\\ j_1+\cdots+j_l=s}}\Big|\sum_{\nu}\prod_{k=1}^{m(\nu)}cum_{s_k}(\varepsilon) \times\prod_{k=m(\nu)+1}^{p}cum(A_{ij}, ij\in\nu_k)\Big| \tag{6.6}$$

with

$$\begin{aligned} A_{ij} &= w_N(\mathbf{x}^*-\mathbf{X}_1)) && 1\le i\le s\\ A_{ij} &= g(\mathbf{X})w_N(\mathbf{x}^*-\mathbf{X}_1)) && s+1\le i\le l. \end{aligned}$$

and $s_1+\cdots+s_{m(\nu)}=s$. Furthermore we have $s_i\ge 2$, because the noise terms ε have mean zero, and each set $\nu_{m(\nu)+1}, \ldots, \nu_p$ includes at least one A_{ij} with $1\le i\le s$ because otherwise the partition would not be indecomposable. Let $a_r=|\nu_r|$ denote the number of elements in the set ν_r ($r=m(\nu)+1, \ldots, p$), then we get $a_{m+1}+\cdots+a_p=l$. Furthermore for $r\in\{m+1, \ldots, p\}$ the cumulant $cum(A_{ij}, ij\in\nu_r)$ equals

$$cum(g(\mathbf{X}_1)w_N(\mathbf{x}^*,\mathbf{X}_1)), \ldots, g(\mathbf{X}_1)w_N(\mathbf{x}^*,\mathbf{X}_1)), w_N(\mathbf{x}^*,\mathbf{X}_1)), \ldots, w_N(\mathbf{x}^*,\mathbf{X}_1))) \tag{6.7}$$

because of the symmetry of the arguments in the cumulant. In the next step we denote by b_r the number of components of the form $g(\mathbf{X}_1)w_N(\mathbf{x}^*,\mathbf{X}_1)$ and show the estimate

$$E\Big[\prod_{i=1}^{b_r}|g(\mathbf{X}_1)w_N(\mathbf{x}^*,\mathbf{X}_1))|\prod_{j=1}^{a_r-b_r}|w_N(\mathbf{x}^*,\mathbf{X}_1))|\Big]\le\frac{C^{a_r}}{N^{a_r}h^{a_r(\beta+d)}f(\frac{1}{\mathbf{a_N}})^{a_r}} \tag{6.8}$$

(which does not depend on b_r). From Theorem 6.1 we then obtain that the term in (6.7) is of order $O(N^{-a_r}h^{-a_r(\beta+d)}f(1/\mathbf{a_N})^{-a_r})$. Equations (6.4), (6.6) and (6.7) yield for the cumulants of order $l\ge 3$

$$\begin{aligned} G_l &\le CN^{l/2+1}h^{l\beta+dl/2}\sum_{s=0}^{l}\binom{l}{s}\sum_{\substack{\mathbf{j}\in\{0,1\}^l\\ j_1+\cdots+j_l=s}}\Big|\sum_{\nu}\prod_{k=1}^{m(\nu)}cum_{s_k}(\varepsilon)\\ &\qquad\times\prod_{r=m(\nu)+1}^{p}\frac{C^{a_r}}{N^{a_r}h^{a_r(d+\beta)}f(\frac{1}{\mathbf{a_N}})^{a_r}}\Big|\\ &= O\big((N^{l/2-1}h^{ld/2}f(\mathbf{a_N}^{-1})^l)^{-1}\big)=o(1), \end{aligned}$$

which shows the asymptotic normality.

In order to prove the remaining estimate (6.8) we use the definition of $w_N(\mathbf{x}^*, \mathbf{X}_1)$ and obtain for the term on the left hand side of (6.8)

$$\begin{aligned} L_N &= \int_{\mathbb{R}^d} |g(\mathbf{x})|^{b_r} \Big(\frac{1}{Nh^d(2\pi)^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{s}, (\mathbf{x}^* - \mathbf{x})\rangle/h} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} \\ &\qquad \times \frac{1}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} d\mathbf{s} \Big| \Big)^{a_r} f(\mathbf{x}) d\mathbf{x} \\ &\leq \frac{C}{N^{a_r} h^{a_r d}} \int_{\mathbb{R}^d} |g(\mathbf{x})|^{b_r} \Big(\int_{\mathbb{R}^d} \frac{|\Phi_K(\mathbf{s})|}{|\Phi_\psi(\frac{\mathbf{s}}{h})|} \frac{1}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} d\mathbf{s} \Big)^{a_r} f(\mathbf{x}) d\mathbf{x} \\ &\leq \frac{C}{N^{a_r} h^{a_r d} f(\frac{1}{\mathbf{a_N}})^{a_r}} \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \frac{|\Phi_K(\mathbf{s})|}{|\Phi_\psi(\frac{\mathbf{s}}{h})|} d\mathbf{s} \Big)^{a_r} f(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

where we used the fact that the function g is bounded. Using this inequality and Assumption (A.1)(A) it follows that $L_n \leq C/N^{a_r} h^{a_r(d+\beta)} f(\frac{1}{\mathbf{a_N}})^{a_r}$, which proves (6.8).

6.2. Proof of Lemma 2.6

Similar to the proof of Theorem 2.4, we have to calculate the cumulants of the estimators $\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})$. We start with the first order cumulant

$$\begin{aligned} E[\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})] &= \frac{1}{h^d(2\pi)^d} \int_{\mathbb{R}^{d-d_j}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(\mathbf{x}) \int_{\mathbb{R}^d} \frac{\Phi_K(\mathbf{s})}{\Phi_\psi(\frac{\mathbf{s}}{h})} \\ &\qquad \times \frac{e^{-i\langle \mathbf{s}, (\mathbf{x}^* - \mathbf{x})\rangle/h} f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} d\mathbf{s} d\mathbf{x} dQ_{I_j^c}(\mathbf{x}^*_{I_j^c}) \end{aligned}$$

and with the same arguments as in the proof of Theorem 2.4, we obtain a bias of order $O(h^{s-1})$. For the calculation of the variance of $\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})$ we investigate its conditional variance. Recalling the definitions (2.5) and (2.13) it follows by a straightforward argument

$$\begin{aligned} V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})|\mathbf{X}) &= \frac{\sigma^2}{N^2 h^{2d}(2\pi)^{2d}} \sum_{k=1}^n \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w}, \mathbf{X}_k\rangle/h} e^{i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} \\ &\qquad \times L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 \frac{1}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}^2}, \end{aligned}$$

which gives

$$\begin{aligned} E\big[V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})|\mathbf{X})\big] &= \frac{\sigma^2}{Nh^{2d}(2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w}, \mathbf{x}\rangle/h} e^{i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} \\ &\qquad \times L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 \frac{f(\mathbf{x})}{\max\{f(\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{x}. \end{aligned}$$

The variance of the conditional expectation can be calculated as

$$\begin{aligned} V\big(E[\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})|\mathbf{X}]\big) &= \frac{1}{Nh^d(2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w},\mathbf{x}\rangle} e^{i\langle \mathbf{w}_{I_j},\mathbf{x}^*_{I_j}\rangle/h} \\ &\quad \times L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w}\Big|^2 \frac{g(h\mathbf{x})^2 f(h\mathbf{x})}{\max\{f(h\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{x} \\ &- \frac{1}{N(2\pi)^{2d}} \Big| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w},\mathbf{x}\rangle} e^{i\langle \mathbf{w}_{I_j},\mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big) \\ &\quad \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \frac{g(h\mathbf{x}) f(h\mathbf{x})}{\max\{f(h\mathbf{x}), f(\frac{1}{\mathbf{a_N}})\}} d\mathbf{x}\Big|^2, \end{aligned}$$

where the second summand is of order $O(N^{-1})$. Therefore it follows

$$\begin{aligned} V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})) &= \frac{1}{Nh^d(2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{-i\langle \mathbf{w},\mathbf{y}\rangle} e^{i\langle \mathbf{w}_{I_j},\mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big) \\ &\quad \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w}\Big|^2 \frac{(\sigma^2 + g(h\mathbf{y})^2) f(h\mathbf{y})}{\max\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{y} + O(n^{-1}). \end{aligned}$$

The upper bound for this term is obtained from Assumption (A.4) which gives

$$\frac{(\sigma^2 + g(h\mathbf{y})^2) f(h\mathbf{y})}{\max\{f(h\mathbf{y}), f(\frac{1}{\mathbf{a_N}})\}^2} = O(f(\mathbf{a_N}^{-1})). \tag{6.9}$$

Therefore an application of Parseval's equality and Assumption (A.5)(C) yields

$$V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})) \leq \frac{C}{Nh^{d+2\beta-\gamma_j} f(\frac{1}{\mathbf{a_N}})}. \tag{6.10}$$

A similar argument as in the proof of Theorem 2.4 gives the lower bound $V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})) \geq C/Nh^{d+2\beta-\gamma_j}$. Finally the statement that the l-th cumulant of $V(\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})^{-1/2})\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})$ is of order $o(1)$ can be shown by similar arguments as in the proof of Theorem 2.4.

6.3. *Proof of Theorem 2.8*

The proof follows by similar arguments as given in the previous Sections. For the sake of brevity we restrict ourselves for the calculation of the first and second order cumulants. For this purpose we show, that the estimate $\hat{c}$ has a faster rate of convergence than $\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})$ for at least one $j \in \{1,\dots,m\}$. If this statement is correct the asymptotic variance of the statistic

$$\hat{\theta}^{add,RD}(\mathbf{x}^*) = \sum_{j=1}^{m} \hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j}) - (m-1)\hat{c}$$

is determined by its first term. Recalling the notation (2.12) this term has the representation

$$\hat{D}_N = \sum_{j=1}^{m} \hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j}) \quad = \quad \sum_{j=1}^{m}\sum_{k=1}^{N} Y_k w_N^{add,RD}(\mathbf{x}^*_{I_j}, \mathbf{X}_k) \tag{6.11}$$

and can be treated in the same way as before. The resulting bias of $\hat{D}_N$ is the sum of the biases of the individual term and therefore also of order $O(h^{s-1})$. The conditional variance is given by

$$\begin{aligned} V(\hat{D}_N|\mathbf{X}) &= \sigma^2 \sum_{k=1}^{N} \Big| \sum_{j=1}^{m} w_N^{add,RD}(\mathbf{x}^*_{I_j}, \mathbf{X}_k) \Big|^2 \\ &= \frac{\sigma^2}{N^2 h^{2d} (2\pi)^{2d}} \sum_{k=1}^{n} \Big| \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{X}_k\rangle/h} \Big(\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \\ &\quad \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \frac{1}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}} \Big|^2 . \end{aligned}$$

This yields for expectation of the conditional variance

$$\begin{aligned} &E\big[V(\hat{D}_N|\mathbf{X})\big] \\ &= \frac{\sigma^2}{N h^{2d} (2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{s}\rangle/h} \Big(\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 \\ &\quad \times \frac{f(\mathbf{s})}{\max\{f(\mathbf{s}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{s} \end{aligned}$$

and the variance of the conditional expectation is obtained as

$$\begin{aligned} &V\big(E[\hat{D}_N|\mathbf{X}]\big) \\ &= \frac{1}{N h^{d} (2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{s}\rangle} \Big(\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 \\ &\quad \times \frac{g(h\mathbf{s})^2 f(h\mathbf{s})}{\max\{f(h\mathbf{s}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{s} \\ &- \frac{1}{N(2\pi)^{2d}} \Big| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{s}\rangle} \Big(\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \\ &\quad \times \frac{g(h\mathbf{s}) f(h\mathbf{s})}{\max\{f(h\mathbf{s}), f(\frac{1}{\mathbf{a_N}})\}} d\mathbf{s} \Big|^2 , \end{aligned}$$

where the second summand is of order $O(N^{-1})$. This gives for the variance

$$\begin{aligned} V(\hat{D}_N) &= \frac{1}{N h^{2d} (2\pi)^{2d}} \int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{i\langle \mathbf{w}, \mathbf{s}\rangle/h} \Big(\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j}, \mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \\ &\quad \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 \frac{(\sigma^2 + g(\mathbf{s})^2) f(\mathbf{s})}{\max\{f(\mathbf{s}), f(\frac{1}{\mathbf{a_N}})\}^2} d\mathbf{s} + O(N^{-1}). \end{aligned}$$

In order to obtain bounds for the rate of the variance, we use the lower bound for $\max\{f(h\mathbf{s}), f(\frac{1}{\mathbf{a_N}})\}$ mentioned in (6.9) and Parseval's equality which yields

$$\Big(\frac{1}{Nh^d f(\frac{1}{\mathbf{a_N}})}\int_{\mathbb{R}^d}\Big|\sum_{j=1}^{m} e^{-i\langle \mathbf{w}_{I_j},\mathbf{x}^*_{I_j}\rangle/h} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big|^2 \frac{|\Phi_K(\mathbf{w})|^2}{|\Phi_\psi(\frac{\mathbf{w}}{h})|^2}d\mathbf{w}\Big)^{1/2} = O\left((nh^{d+2\beta-\gamma_{min}} f(\mathbf{a_N}^{-\mathbf{1}}))^{-\mathbf{1}}\right)$$

as an upper bound, where the last estimate follows from Assumption (A.5). The lower bound is of order $\Omega((Nh^{d+2\beta-\gamma_{min}})^{-1})$, where we use Assumption (A.4) and the same calculations as in the previous Section. These are in fact the same bounds as for $\hat{\alpha}_{j^*,Q_{I^c_{j^*}}}(\mathbf{x}^*_{I_{j^*}})$ with $j^* = argmin_j\, \gamma_j$. This means that

$$\hat{D}_N - E[\hat{D}_N] = O_P(N^{-1/2}h^{-d/2-\beta-\gamma_{min}/2} f(\mathbf{a_N}^{-1})^{-1/2})$$

In the last step we show that the estimate $\hat{c}$ has a faster rate of convergence. For this purpose we write $\hat{c}$ as weighted sum of independent random variables that is

$$\hat{c} = \int_{\mathbb{R}^d}\hat{\theta}(\mathbf{x}^*)dQ(\mathbf{x}^*) = \frac{1}{Nh^d(2\pi)^d}\sum_{k=1}^{n}\int_{\mathbb{R}^d} e^{i\langle \mathbf{w},\mathbf{X}_k\rangle/h}\Big(\prod_{j=1}^{m} L_{I_j^c}\Big(\frac{\mathbf{w}_{I_j^c}}{h}\Big)\Big) \times \frac{\Phi_K(\mathbf{w})}{\Phi_\psi(\frac{\mathbf{w}}{h})}d\mathbf{w}\frac{Y_k}{\max\{f(\mathbf{X}_k), f(\frac{1}{\mathbf{a_N}})\}}$$

It now follows by similar calculations as given in the previous paragraph and Assumption (A.5)(C) that

$$V(\hat{c}) = o\Big(V\Big(\sum_{j=1}^{m}\hat{\alpha}_{j,Q_{I_j^c}}(\mathbf{x}^*_{I_j})\Big)\Big)$$

and thus we can ignore the term $\hat{c}$ for the calculation of the asymptotic variance of the statistic $\hat{\theta}^{add,RD}$.

6.4. *Proof of Lemma 3.3*

Observing the representation (3.4) and (3.7) we decompose the estimator into its deterministic and stochastic part, that is

$$\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}) = \hat{E}_{1n} + \hat{E}_{2n} \tag{6.12}$$

where

$$\hat{E}_{1n} = \frac{1}{(2n+1)^{d-d_j}}\sum_{\mathbf{k}\in\{-n,\ldots,n\}^d}(g_{I_1}(\mathbf{z}_{\mathbf{k}_{I_1}}) + \cdots + g_{I_m}(\mathbf{z}_{\mathbf{k}_{I_m}}))w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j})$$

$$\hat{E}_{2n} \quad = \quad \frac{1}{(2n+1)^{d-d_j}} \sum_{\mathbf{k}\in\{-n,\ldots,n\}^d}^{n} \varepsilon_{\mathbf{k}} w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j})$$

and $w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j})$ are defined in (3.8). In a first step we show, that the bias of $\hat{\theta}^{FD}_{I_j}$ is of order $O(\frac{1}{n^2h^{2+d_j+\beta_j}a_n^3})$. For this purpose we rewrite the deterministic part as

$$\hat{E}_{1n} \quad = \quad \hat{E}^{(1)}_{1n} + \hat{E}^{(2)}_{1n}$$

where

$$\begin{aligned}
\hat{E}^{(1)}_{1n} \quad &= \quad \sum_{\mathbf{k}_{I_j}\in\{-n,\ldots,n\}^{d_j}} g_{I_j}(\mathbf{z}_{\mathbf{k}_{I_j}}) w_{\mathbf{k}_{I_j},n}(\mathbf{x}_{I_j}) \\
\hat{E}^{(2)}_{1n} \quad &= \quad \frac{1}{(2n+1)^{d-d_j}} \sum_{\mathbf{k}_{I_j^c}\in\{-n,\ldots,n\}^{d-d_j}} \Big(g_{I_1}(\mathbf{z}_{\mathbf{k}_{I_1}}) + \cdots + g_{I_{j-1}}(\mathbf{z}_{\mathbf{k}_{I_{j-1}}}) \\
&\qquad + g_{I_{j+1}}(\mathbf{z}_{\mathbf{k}_{I_{j+1}}}) + \cdots + g_{I_m}(\mathbf{z}_{\mathbf{k}_{I_m}}) \Big) \sum_{\mathbf{k}_{I_j}\in\{-n,\ldots,n\}^{d_j}} w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j}),
\end{aligned}$$

where the second summand is of order $\hat{E}^{(2)}_{1n} = o(\frac{a_n^{r-d_j}}{h^{\beta_j+d_j}}) = O(h^{s-1})$, which follows from Assumption (A.3)(D). For the difference of the first summand and $\theta^{add}_{I_1}(\mathbf{x}^*_{I_j})$ we use the same calculation as in Birke and Bissantz (2009) and obtain

$$\hat{E}^{(1)}_{1n} - \theta^{add}_{I_j}(\mathbf{x}_{I_j}) \quad = \quad O(h^{s-1}) + O(\frac{1}{n^2a_n^2h^{d_j+2+\beta_j}}).$$

Note that the Riemann approximation does not provide an error of order

$$O((na_n)^{-d}),$$

but we can show that the lattice structure yields an error term of order

$$O((n^2h^2a_n^3)^{-1}).$$

In the next step we derive the variance of the estimator $\hat{\theta}^{FD}_{I_j}$. We can neglect the deterministic part $\hat{E}_{2n}$ in (6.12) and obtain from Parseval's equality and Assumption (A.1)(B)

$$\begin{aligned}
V(\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j})) \quad &= \quad \frac{\sigma^2}{(2n+1)^{d-d_j}} \sum_{\mathbf{k}_{I_j}\in\{-n,\ldots,n\}^{d_j}} |w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j})|^2 \\
&= \quad \frac{\sigma^2}{(2n+1)^{d-d_j}n^{2d_j}h^{2d_j}a_n^{2d_j}(2\pi)^{2d_j}} \\
&\quad \times \sum_{\mathbf{k}_{I_j}\in\{-n,\ldots,n\}^{d_j}} \Big| \int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w},(\mathbf{x}^*_{I_j}-\mathbf{z}_{k_{I_j}})\rangle/h} \frac{\Phi_{K_j}(\mathbf{w})}{\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{(2n+1)^{d-d_j} n^{d_j} h^{d_j} a_n^{d_j} (2\pi)^{2d_j}} \\
&\quad \Big(\int_{[-1/(ha_n),1/(ha_n)]^{d_j}} \Big| \int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w}, (\mathbf{x}^*_{I_j}/h - \mathbf{s})\rangle} \frac{\Phi_{K_j}(\mathbf{w})}{\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 ds \\
&\qquad + O((na_n)^{-1}) \Big) \\
&\sim \frac{\sigma^2}{(2n+1)^{d-d_j} n^{d_j} h^{d_j} a_n^{d_j} (2\pi)^{2d_j}} \int_{\mathbb{R}^{d_j}} \Big| \int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w}, (\mathbf{x}^*_{I_j}/h - \mathbf{s})\rangle} \\
&\qquad \times \frac{\Phi_{K_j}(\mathbf{w})}{\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big|^2 ds (1 + o(1)) \\
&= \frac{\sigma^2}{(2n+1)^{d-d_j} n^{d_j} h^{d_j} a_n^{d_j} (2\pi)^{2d_j}} \int_{\mathbb{R}^{d_j}} \frac{|\Phi_{K_j}(\mathbf{w})|^2}{|\Phi_{\psi_{I_j}}(\frac{\mathbf{w}}{h})|^2} d\mathbf{w} (1 + o(1)) \\
&= \frac{\sigma^2 C}{(2n+1)^{d-d_j} n^{d_j} h^{d_j + 2\beta_j} a_n^{d_j} (2\pi)^{2d_j}} \sim \frac{C}{n^d h^{d_j + 2\beta_j} a_n^{d_j}}.
\end{aligned}
$$

For the proof of the asymptotic normality, we finally show that the l-th cumulant of

$$V(\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}))^{-1/2} \hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j})$$

converges to zero for $l \geq 3$, which completes the proof of Lemma 3.3. For this purpose we note that

$$
\begin{aligned}
&|cum_l(V(\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}))^{-1/2} \hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}))| \leq |C n^{ld/2} h^{ld_j/2 + l\beta_j} a_n^{ld_j/2} cum_l(\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j}))| \\
&\leq \Big| \frac{C}{(n^{ld/2} h^{ld_j/2} a_n^{ld_j/2}} \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_l \in \{-n, \ldots, n\}^d} \prod_{m=1}^{l} \Big(\int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w}, (\mathbf{x}^*_{I_j} - z_{k_{m,I_j}})\rangle/h} \\
&\qquad \times \frac{\Phi_{K_j}(\mathbf{w})}{\Phi_{\psi}(\frac{\mathbf{w}}{h})} d\mathbf{w} \Big) cum(\varepsilon_{\mathbf{k}_1}, \ldots, \varepsilon_{\mathbf{k}_l}) \Big| \\
&\leq \frac{C}{n^{ld/2} h^{ld_j/2} a_n^{ld_j/2}} \sum_{\mathbf{k}_1 \in \{-n, \ldots, n\}^d} \prod_{m=1}^{l} \left(\int_{\mathbb{R}^{d_j}} \frac{|\Phi_{K_j}(\mathbf{w})|}{|\Phi_{\psi}(\frac{\mathbf{w}}{h})|} d\mathbf{w} \right),
\end{aligned}
$$

where κ_l denotes the l-th cumulant of ε. From Assumption 1 it follows that this term is bounded by

$$\frac{C}{n^{ld/2} h^{ld_j/2} a_n^{ld_j}/2} (2n+1)^d h^{-l\beta_j} = C n^{-ld/2+1} h^{-ld_j/2} a_n^{ld_j/2},$$

which converges to zero for $l \geq 3$.

6.5. *Proof of Theorem 3.4*

In the following discussion we ignore the constant term $g_0 = \theta_0$ because the mean

$$\hat{\theta}_0 = \frac{1}{n^d} \sum_{\mathbf{k} \in \{-n, \ldots, n\}^d} Y_{\mathbf{k}}$$

is a $\sqrt{n^d}$-consistent estimator for this constant and the nonparametric components in (3.2) can only be estimated at slower rates. Note that

$$\hat{\theta}^{add,FD}(\mathbf{x}^*) = \sum_{\mathbf{k}\in\{-n,\ldots,n\}^d} Y_{\mathbf{k}} \sum_{j=1}^{n} \frac{1}{(2n+1)^{d-d_j}} w_{\mathbf{k}_{I_j},n}(\mathbf{x}^*_{I_j})$$

and obtain the asymptotic distribution with the same arguments as in the proof of Lemma 3.3.

6.6. Proof of Theorem 4.2

Under the assumption of an MA(q)-dependency structure (4.4) there are no changes in the calculation of the mean of the estimator $\hat{\theta}^{FD}_{I_j}$ and we only have to calculate the cumulants of order $l \geq 2$ in order to establish the asymptotic normality. We start with the variance, which is given by

$$\begin{aligned}
V(\hat{\theta}^{FD}_{I_j}(\mathbf{x}^*_{I_j})) &= \frac{1}{(2n+1)^{2(d-d_j)}} \sum_{\mathbf{k}_1,\mathbf{k}_2\in\{-n,\ldots,n\}^d} w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\overline{w}_{\mathbf{k}_{2,I_j},n}(\mathbf{x}^*_{I_j}) \\
&\qquad \times cum(\varepsilon_{\mathbf{k}_1},\varepsilon_{\mathbf{k}_2}) \\
&= \frac{1}{(2n+1)^{2(d-d_j)}} \sum_{\mathbf{k}_1\in\{-n,\ldots,n\}^d} \sum_{\mathbf{k}_2:\|\mathbf{k}_2-\mathbf{k}_1\|_\infty\leq 2q} \sum_{\mathbf{r}_1,\mathbf{r}_2\in\{-q,\ldots,q\}^d} \\
&\qquad \times w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\overline{w}_{\mathbf{k}_{2,I_j},n}(\mathbf{x}^*_{I_j}) cum(\beta_{\mathbf{r}_1}Z_{\mathbf{k}_1-\mathbf{r}_1},\beta_{\mathbf{r}_2}Z_{\mathbf{k}_2-\mathbf{r}_2}) \\
&= \frac{1}{(2n+1)^{2(d-d_j)}} \sum_{\mathbf{k}_1\in\{-n,\ldots,n\}^d} \sum_{\mathbf{k}_2:\|\mathbf{k}_2-\mathbf{k}_1\|_\infty\leq 2q} \sum_{\mathbf{r}_1\in\{-q,\ldots,q\}^d} \\
&\times \quad w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\overline{w}_{\mathbf{k}_{2,I_j},n}(\mathbf{x}^*_{I_j}) cum(\beta_{\mathbf{r}_1}Z_{\mathbf{k}_1-\mathbf{r}_1},\beta_{\mathbf{k}_2-\mathbf{k}_1+\mathbf{r}_1}Z_{\mathbf{k}_1-\mathbf{r}_1}) \\
&= \frac{\sigma^2}{(2n+1)^{2(d-d_j)}} \sum_{\mathbf{k}_1\in\{-n,\ldots,n\}^d} \sum_{\mathbf{k}_2:\|\mathbf{k}_2-\mathbf{k}_1\|_\infty\leq 2q} \sum_{\mathbf{r}_1\in\{-q,\ldots,q\}^d} \\
&\qquad \times w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\overline{w}_{\mathbf{k}_{2,I_j},n}(\mathbf{x}^*_{I_j})\beta_{\mathbf{r}_1}\beta_{\mathbf{k}_2-\mathbf{k}_1+\mathbf{r}_1} \\
&= \frac{\sigma^2}{(2n+1)^{2(d-d_j)}} \sum_{\mathbf{k}_1\in\{-n,\ldots,n\}^d} \sum_{\substack{\mathbf{l}\in\mathbb{Z}^d \\ \|\mathbf{l}\|_\infty\leq 2q}} \sum_{\mathbf{r}_1\in\{-q,\ldots,q\}^d} \\
&\qquad \times w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\overline{w}_{\mathbf{l}_{I_j}+\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})\beta_{\mathbf{r}_1}\beta_{\mathbf{l}+\mathbf{r}_1} \\
&= \frac{\sigma^2(1+o(1))}{(2n+1)^{(d-d_j)}} \sum_{\mathbf{k}_{1,I_j}\in\{-n,\ldots,n\}^{d_j}} \sum_{\substack{\mathbf{l}\in\mathbb{Z}^d \\ \|\mathbf{l}\|_\infty\leq 2q}} \sum_{\mathbf{r}_1\in\{-q,\ldots,q\}^d} \\
&\qquad |w_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})|^2\beta_{\mathbf{r}_1}\beta_{\mathbf{l}+\mathbf{r}_1},
\end{aligned}$$

where we used a Taylor-approximation for the weights

$$\overline{w}_{\mathbf{l}_{I_j}+\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j}) = \overline{w}_{\mathbf{k}_{1,I_j},n}(\mathbf{x}^*_{I_j})(1+o(1))$$

in the last step. This gives the expression for the variance in Lemma 4.2. For the calculation of the cumulants of $V^{-1/2}\hat{\theta}^{add,FD}_{I_j}$ we first note that the order of the variance $V = V(\hat{\theta}^{add,FD}_{I_j}(\mathbf{x}_{I_j}))$ can be calculated in the same way as in the proof of Lemma 3.3, which gives $V = O(n^{-d}h^{-d_j-2\beta_j}a_n^{-d_j})$. Therefore we have to show

$$|cum_l(n^{d/2}h^{d_j/2+\beta}a_n^{d_j/2}\hat{\theta}^{add,FD}_{I_j})| = n^{ld/2}h^{l(d_j/2+\beta}a_n^{ld_j/2}|cum_l(\hat{\theta}^{add,FD}_{I_j})| \to 0$$

for $l \geq 3$. By a straightforward calculation it follows that

$$\begin{aligned}
&|cum_l(\hat{\theta}^{FD}_{I_j})(\mathbf{x}^*_{I_j})| \\
&= \Big|\frac{1}{(2n+1)^{l(d-d_j)}n^{ld_j}h^{ld_j}a_n^{ld_j}} \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_l\in\{-n,\ldots,n\}^d} \prod_{m=1}^{l}\Big(\int_{\mathbb{R}^{d_j}} e^{-i\langle \mathbf{w},(\mathbf{x}^*_{I_j}-z_{k_{m,I_j}})\rangle/h} \\
&\qquad\qquad \times \frac{\Phi_{K_j}(\mathbf{w})}{\phi_\psi(\frac{\mathbf{w}}{h})}d\mathbf{w}\Big) cum(\varepsilon_{\mathbf{k}_1},\ldots,\varepsilon_{\mathbf{k}_l})\Big| \\
&\leq \frac{C}{(2n+1)^{l(d-d_j)}n^{ld_j}h^{ld_j}a_n^{ld_j}} \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_l\in\{-n,\ldots,n\}^d} \prod_{m=1}^{l}\Big(\int_{\mathbb{R}^{d_j}} \frac{|\Phi_{K_j}(\mathbf{w})|}{|\phi_\psi(\frac{\mathbf{w}}{h})|}d\mathbf{w}\Big) \\
&\qquad\qquad \times |cum(\varepsilon_{\mathbf{k}_1},\ldots,\varepsilon_{\mathbf{k}_l})| \\
&= \frac{C}{(2n+1)^{l(d-d_j)}n^{ld_j}h^{ld_j}a_n^{ld_j}}\frac{1}{h^{l\beta}} \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_l\in\{-n,\ldots,n\}^d} |cum(\varepsilon_{\mathbf{k}_1},\ldots,\varepsilon_{\mathbf{k}_l})| \\
&= \frac{C}{(2n+1)^{l(d-d_j)}n^{ld_j}h^{ld_j}a_n^{ld_j}}\frac{1}{h^{l\beta}}(2n+1)^d,
\end{aligned}$$

because by (4.4) $\mathbf{k}_1$ can be chosen arbitrarily and $\mathbf{k}_2,\ldots,\mathbf{k}_l$ have only $(4q+1)^d$ possibilities to be chosen and their bound is independent of n. Thus the l-th cumulant is of order $n^{-ld/2+1}h^{-ld_j/2}a_n^{-ld_j/2}$, which converges to zero for $l \geq 3$. The result for $\hat{\theta}^{add,FD}$ follows immediately from the results of $\hat{\theta}^{FD}_{I_j}$.

References

Adorf, H. M. (1995). Hubble Space Telescope image restoration in its fourth year. *Inverse Problems*, 11:639–653.

Bertero, M., Boccacci, P., Desiderà, G., and Vicidomini, G. (2009). Image deblurring with Poisson data: From cells to galaxies. *Inverse Problems*, 25(12):123006, 26. MR2565572

Birke, M. and Bissantz, N. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *Journal of Multivariate Statistics*, 100(10):2364–2375. MR2560377

Birke, M., Bissantz, N., and Hildebrandt, T. (2012). Asymptotic normality and confidence intervals for time-dependent inverse regression models with convolution-type operators. *Submitted for publication.*

Birke, M., Bissantz, N., and Holzmann, H. (2010). Confidance bands for inverse regression models. *Inverse Problems*, 26:115020. MR2732912

Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007a). Nonparametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society, Series B*, 69:483–506. MR2323764

Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. (2007b). Convergence rates of general regularization methods for statistical inverse problems. *SIAM J. Num. Anal.*, 45:2610–2636. MR2361904

Brillinger, D. R. (2001). *Time Series Data Analysis and Theory*. SIAM. MR1853554

Carroll, R. J., Härdle, W., and Mammen, E. (2002). Estimation in an additive model when the parameters are linked parametrically. *Econometric Theory*, 18(4):886–912. MR1918327

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 19. MR2421941

Cavalier, L. and Tsybakov, A. (2002). Sharp adaption for inverse problems with random noise. *Prob. Theory Related Fields*, 123:323–354. MR1918537

De Gooijer, J. G. and Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association*, 98(461):135–146. MR1965680

Dette, H. and Scheder, R. (2011). Estimation of additive quantile regression. *Annals of the Institute of Statistical Mathematics*, 63(2):245–265. MR2780485

Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society, Series B*, 55:523–531. MR1224414

Doksum, K. and Koo, J. Y. (2000). On spline estimators and prediction intervals in nonparametric regression. *Computational Statistics and Data Analysis*, 35:67–82. MR1815574

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers Group, Dordrecht. MR1408680

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19:1257–1272. MR1126324

Hastie, T. J. and Tibishirani, R. J. (2008). *Generalized Additive Models*. Chapman & Hall.

Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, 95(2):246–272. MR2170397

Horowitz, J. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100(472):1238–1249. MR2236438

Kaipio, J. and Somersalo, E. (2010). *Statistical and Computational Inverse Problems*. Springer, Berlin. MR2102218

Lee, Y. K., Mammen, E., and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Annals of Statistics*, 38(5):2857–2883. MR2722458

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1):93–100. MR1332841

Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56:1424–1444. MR1409127

Mammen, E., Linton, O. B., and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, 27(5):1443–1490. MR1742496

Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society, Ser. B*, 67(1):43–61. MR2136638

Saitoh, S. (1997). *Integral Transforms, Reproducing Kernels and Their Applications*. Longman, Harlow. MR1478165

Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184. MR1054861