

Smooth backfitting in additive inverse regression

Nicolai Bissantz · Holger Dette · Thimo Hildebrandt · Kathrin Bissantz

Received: 10 December 2013 / Revised: 5 December 2014 / Published online: 26 March 2015 © The Institute of Statistical Mathematics, Tokyo 2015

Abstract We consider the problem of estimating an additive regression function in an inverse regression model with a convolution type operator. A smooth backfitting procedure is developed and asymptotic normality of the resulting estimator is established. Compared to other methods for the estimation in additive models the new approach neither requires observations on a regular grid nor the estimation of the joint density of the predictor. It is also demonstrated by means of a simulation study that the backfitting estimator outperforms the marginal integration method at least by a factor of two with respect to the integrated mean squared error criterion. The methodology is illustrated by a problem of live cell imaging in fluorescence microscopy.

Keywords Inverse regression \cdot Additive models \cdot Curse of dimensionality \cdot Smooth backfitting

1 Introduction

In this paper we consider the regression model

$$Y_k = g(\mathbf{X}_k) + \varepsilon_k \qquad k \in \{1, \dots, N\},\tag{1}$$

where $\varepsilon_1, \ldots, \varepsilon_N$ are independent identically distributed random variables and $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independent identically distributed *d*-dimensional predictors with

N. Bissantz \cdot H. Dette (\boxtimes) \cdot T. Hildebrandt

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany e-mail: holger.dette@rub.de

components $\mathbf{X}_k = (X_{k,1}, \dots, X_{k,d})^T$ $(k = 1, \dots, N)$. We assume that the function g is related to a signal θ by a convolution type operator, that is

$$g(\mathbf{z}) = \int_{\mathbb{R}^d} \psi(\mathbf{z} - \mathbf{t})\theta(\mathbf{t}) d\mathbf{t},$$
(2)

where $\psi : \mathbb{R}^d \to \mathbb{R}$ is a known function with $\int_{\mathbb{R}^d} \psi(\mathbf{t}) d\mathbf{t} = 1$. The interest of the experiment is the nonparametric estimation of the signal θ . Models of the type (1) and (2), which are closely related to deconvolution (e.g. see Stefanski and Carroll 1990; Fan 1991), belong to the class of inverse regression models and have important applications in the recovery of images from astronomical telescopes or fluorescence microscopes in biology. Deterministic inverse regression models have been considered for a long time in the literature (Engl et al. 1996; Saitoh 1997). However, in the past decade statistical inference in ill-posed problems has become a very active field of research (see Bertero et al. 2009; Kaipio and Somersalo 2010 for a Bayesian approach and Mair and Ruymgaart 1996; Cavalier 2008; Bissantz et al. 2007 for nonparametric methods).

While most of these methods have been developed for models with a onedimensional predictor, nonparametric estimation in the multivariate setting is of practical importance because in many applications one has to deal with an at least twodimensional predictor. A typical example is image reconstruction since a picture is a two-dimensional object. Also in addition to the spatial dimensions, the data might depend on the time thus introducing a third component. For a multivariate predictor the estimation of the signal θ in the inverse regression model (1) is a much harder problem due to the curse of dimensionality. In direct regression usually qualitative assumptions regarding the signal such as additivity or multiplicativity are made, which allow the estimation of the regression function at reasonable rates (see Linton and Nielsen 1995; Mammen et al. 1999; Carroll et al. 2002; Hengartner and Sperlich 2005; Nielsen and Sperlich 2005). In the present paper we investigate the problem of estimating the signal θ in the inverse regression model with a convolution type operator under the additional assumption of additivity, that is

$$\theta(\mathbf{x}) = \theta_0 + \theta_1(x_1) + \dots + \theta_d(x_d), \tag{3}$$

where $\mathbf{x} = (x_1, \dots, x_d)^T$. In a recent paper Hildebrandt et al. (2014) proposed an estimator of the signal θ if observations are available on a regular grid in \mathbb{R}^d . They also considered the case of a random design and investigated the statistical properties of a marginal integration type estimate with known density of the predictor. The asymptotic analysis of both estimates is based on these rather restrictive assumptions regarding the predictor \mathbf{X} . A regular grid or explicit knowledge of the density of the predictor \mathbf{X} might not be available in all applications. Moreover, estimation of this density in the marginal integration method cannot be performed at one-dimensional rates (see Hildebrandt et al. 2014). In particular, it changes the asymptotic properties of additive estimates such that the signal cannot be reconstructed with one-dimensional nonparametric rates. In the present paper we consider the construction of an estimate in the inverse additive regression model (3) with random design, which is applicable under less restrictive assumptions in particular without knowledge of the density of the

predictor. For this purpose we combine in Sect. 2 smooth backfitting (see Mammen et al. 1999) with Fourier estimation methods in inverse regression models (see Diggle and Hall 1993; Mair and Ruymgaart 1996). Besides several advantages of the smooth backfitting approach observed in the literature in direct regression models (see Nielsen and Sperlich 2005), the backfitting methodology only requires the estimation of the marginal densities of the predictor. As a consequence, the resulting estimate does not suffer from the curse of dimensionality. Section 3 is devoted to the investigation of the asymptotic properties of the new estimator, while we study the finite sample properties by means of a simulation study in Sect. 4. In particular we demonstrate that the smooth backfitting approach results in estimates with an at least two times smaller integrated mean squared error than the marginal integration method. We also illustrate the methodology analyzing data from live cell imaging in fluorescence microscopy. Finally, all proofs and technical arguments are presented in Appendix.

2 Smooth backfitting in inverse regression

Note that the linearity of the convolution operator and assumption (3) imply that the function g is also additive, and consequently the model (1) can be rewritten as

$$Y_k = g_0 + g_1(X_{k,1}) + \dots + g_d(X_{k,d}) + \varepsilon_k,$$
(4)

where $\mathbf{X}_k = (X_{k,1}, \dots, X_{k,d})^T$ and the functions g_0, g_1, \dots, g_d in model (4) are related to the components $\theta_0, \theta_1, \dots, \theta_d$ of the signal θ in model (3) by $g_0 = \theta_0$,

$$g_j(z_j) = \int_{\mathbb{R}} \psi_j(z_j - t)\theta_j(t) \mathrm{d}t \qquad j = 1, \dots, d.$$
(5)

Here ψ_j is the marginal of the convolution function ψ , that is

$$\psi_j(t_j) = \int_{\mathbb{R}^{d-1}} \psi(\mathbf{t}) \mathrm{d}\mathbf{t}_{-\mathbf{j}}$$
(6)

and $\mathbf{t} = (t_1, \ldots, t_d)^T \in \mathbb{R}^d$, $\mathbf{t}_{-\mathbf{j}} = (t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_d)^T \in \mathbb{R}^{d-1}$. The estimation of the additive signal is now performed in several steps and combines Fourier transform estimation methods for inverse regression models (see Diggle and Hall 1993; Mair and Ruymgaart 1996) with the smooth backfitting technique developed for direct nonparametric regression models (see Mammen et al. 1999).

(1) We assume for a moment that the design density is known and denote by f_j and F_j the density and cumulative distribution function of the *j*th marginal distribution of the random variable **X**. In a first step all explanatory variables are transformed to the unit cube using the probability transformation in each component, that is

$$Z_{k,j} = F_j(X_{k,j}) \quad j = 1, \dots, d; \ k = 1, \dots, N.$$
 (7)

This transformation is necessary because of two reasons. On the one hand, the asymptotic analysis of methods based on Fourier estimation requires (with positive

probability) observations at points \mathbf{X}_k with a norm $\| \mathbf{X}_k \|$ converging to infinity, because one has to estimate the Fourier transform of the function g_j on the real axis. On the other hand, the asymptotic analysis of the smooth backfitting method requires a distribution of the explanatory variables with a compact support.

In practice the unknown marginal distributions of the predictor are estimated by standard methods and this estimation does not change the asymptotic properties of the statistic. We refer to Remark 1 for more details.

(2) The transformation in Step (1) yields the representation

$$Y_k = g_0 + g_1^*(Z_{k,1}) + \dots + g_d^*(Z_{k,d}) + \varepsilon_k; \quad k = 1, \dots, N,$$
(8)

where the functions g_j^* are defined by $g_j^* = g_j \circ F_j^{-1}$ (j = 1, ..., d). We now use the smooth backfitting algorithm (see Mammen et al. 1999) to estimate each function g_j^* in (8) from the data $(Z_{1,1}, ..., Z_{1,d}, Y_1), ..., (Z_{N,1}, ..., Z_{N,d}, Y_N)$. This algorithm determines estimates of the components $g_0, g_1^*, ..., g_d^*$ recursively, where $\hat{g}_0 = \overline{Y} = \frac{1}{N} \sum_{k=1}^N Y_k$. For starting values $\hat{g}_1^{*(0)}, ..., \hat{g}_d^{*(0)}$ we calculate for r = 1, 2, ... the estimators $\hat{g}_1^{*(r)}, ..., \hat{g}_d^{*(r)}$ by the recursive relation

$$\hat{g}_{j}^{*(r)}(z_{j}) = \hat{g}_{j}^{*}(z_{j}) - \sum_{k < j} \int \hat{g}_{k}^{*(r)}(z_{k}) \left[\frac{\hat{p}_{jk}(z_{j}, z_{k})}{\hat{p}_{j}(z_{j})} - \hat{p}_{k,[j+]}(z_{k}) \right] dz_{k} - \sum_{k > j} \int g_{k}^{*(r-1)}(z_{k}) \left[\frac{\hat{p}_{jk}(z_{j}, z_{k})}{\hat{p}_{j}(z_{j})} - \hat{p}_{k,[j+]}(z_{k}) \right] dz_{k} - g_{0,j}^{*}.$$
(9)

Here

$$\hat{g}_{j}^{*}(z_{j}) = \frac{\sum_{k=1}^{N} L\left(\frac{Z_{k,j}-z_{j}}{h_{B}}\right)Y_{k}}{\sum_{k=1}^{N} L\left(\frac{Z_{k,j}-z_{j}}{h_{B}}\right)}$$
(10)

denotes the one-dimensional Nadaraya–Watson estimator of the *j*th component (with kernel *L* and bandwidth h_B), \hat{p}_{jk} and \hat{p}_j are the (j, k)th and *j*th marginals of the common kernel density estimator \hat{p} for the density *p* of the predictor $(Z_1, \ldots, Z_d)^T$, and we use the notation

$$\hat{p}_{k,[j+]}(z_k) = \int \hat{p}_{jk}(z_j, z_k) dz_j \left[\int \hat{p}_j(z_j) dz_j \right]^{-1} g_{0,j}^* = \frac{\int \hat{g}_j^*(z_j) \hat{p}_j(z_j) dz_j}{\int \hat{p}_j(z_j) dz_j}.$$
(11)

(3) Estimators of the functions g_j in (4) are now easily obtained by the transformation

$$\hat{g}_j = \hat{g}_j^{*(r_0)} \circ F_j, \tag{12}$$

where $\hat{g}_j^{*(r_0)}$ denotes the estimator obtained after terminating the recursive relation (9) at step r_0 (j = 1, ..., d). In order to recover the signal θ_j from \hat{g}_j we now introduce the random variables

$$U_{k,j} = Y_k - \sum_{\substack{i=1\\i\neq j}}^{d} \hat{g}_i(X_{k,i}) - \hat{g}_0$$
(13)

and use the data $(X_{1,j}, U_{1,j}), \ldots, (X_{N,j}, U_{N,j})$ to estimate the *j*th component θ_j of the signal θ by Fourier transform estimation methods (see Diggle and Hall 1993 for example). For this purpose we note that the relation (5) implies for the Fourier transforms Φ_{g_i} and Φ_{θ_j} of the functions g_j and θ_j the relation

$$\Phi_{\theta_j} = \frac{\Phi_{g_j}}{\Phi_{\psi_j}},$$

where

$$\Phi_{\psi_j}(w) = \int_{\mathbb{R}} \psi_j(x) e^{iwx} \mathrm{d}x$$

is the Fourier transform of the *j*th marginal of the convolution function. Now the Fourier transform $\Phi_{g_j}(w)$ of the function g_j is estimated by its empirical counterpart

$$\hat{\Phi}_{g_j}(w) = \frac{1}{N} \sum_{k=1}^{N} e^{iwX_{k,j}} \frac{U_{k,j}}{\max\{f_j(X_{k,j}), f_j(\frac{1}{a_N})\}},\tag{14}$$

where f_j is the density of the *j*th marginal distribution and a_N is a real-valued sequence converging to 0 as $N \to \infty$. The estimator of $\hat{\theta}_j$ is now obtained from a "smoothed" inversion of the Fourier transform, that is

$$\hat{\theta}_j(x_j) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iwx_j} \Phi_K(hw) \frac{\hat{\Phi}_{g_j}(w)}{\Phi_{\psi_j}(w)} \mathrm{d}w, \tag{15}$$

where Φ_K is the Fourier transform of a kernel *K* and *h* is a bandwidth converging to 0 with increasing sample size.

(4) Finally, the additive estimate of the signal θ is given by

$$\hat{\theta}(\mathbf{x}) = \hat{\theta}_0 + \hat{\theta}_1(x_1) + \dots + \hat{\theta}_d(x_d), \tag{16}$$

where $\hat{\theta}_0 = \hat{g}_0 = \overline{Y}$. and $\hat{\theta}_j$ is defined in (15) for j = 1, ..., d.

Remark 1 (a) Note that we use the term $\max\{f_j(X_{k,j}), f_j(\frac{1}{a_N})\}$ in the denominator of the estimate (14) instead of the more intuitive term $f_j(X_{k,j})$. This "truncation"

avoids situations where the denominator becomes too small, which would yield unstable estimates with a too large variance.

(b) In practical applications, knowledge of the marginal distributions might not be available and in this case the transformation (7) can be achieved by

$$\hat{Z}_{k,j} = \hat{\mathbb{F}}_j(X_{k,j}); \quad j = 1, \dots, d; \ k = 1, \dots, N,$$
 (17)

where for $j = 1, \ldots, d$

$$\hat{\mathbb{F}}_{j}(x) = \frac{1}{N+1} \sum_{k=1}^{N} \mathbb{I}\{X_{k,j} \le x\}$$
(18)

denotes the empirical distribution function of the *j*th components $X_{1,j}, \ldots, X_{N,j}$. Similarly, the density f_j in (14) can be estimated by kernel density methods, that is

$$\hat{f}_{j}(x_{j}) = \frac{1}{Nh_{d,j}} \sum_{k=1}^{N} M\left(\frac{X_{k,j} - x_{j}}{h_{d,j}}\right); \quad j = 1, \dots, d,$$
(19)

where *M* denotes a kernel and $h_{d,j}$ is a bandwidth proportional to $N^{-1/5}$. We note that the estimators $\hat{\mathbb{F}}_j$ and \hat{f}_j converge uniformly to F_j and f_j at rates $(\frac{\log \log N}{N})^{1/2}$ and $(\frac{\log N}{Nh_{d,j}})^{1/2}$, respectively (see van der Vaart 1998; Giné and Guillou 2002). The rates of convergence in inverse deconvolution problems are slower and consequently the asymptotic properties of the estimates $\hat{\theta}_j$ do not change if f_j and F_j are replaced by their empirical counterparts \hat{f}_j and $\hat{\mathbb{F}}_j$ defined in (17) and (19), respectively.

3 Asymptotic properties

In this section we investigate the asymptotic properties of the estimators defined in Sect. 2. In particular we establish weak convergence. For this purpose we require the following assumptions:

- (A1) The kernel *L* in the Nadaraya–Watson estimator \hat{g}_j^* in the backfitting recursion (9) is symmetric, Lipschitz continuous and has compact support, say [-1, 1]. The bandwidth h_B of this estimator is proportional to $N^{-1/5}$.
- (A2) $\mathbb{E}[|Y_j|^{\alpha}] < \infty$ for some $\alpha > \frac{5}{2}$.
- (A3) The functions g_1, \ldots, g_d in model (4) are bounded and twice differentiable with Lipschitz continuous second-order derivatives.
- (A4) The Fourier transforms Φ_{ψ_j} of the marginals ψ_j of the convolution function ψ satisfy

$$\begin{split} &\int_{\mathbb{R}} \frac{|\Phi_{K}(w)|}{|\Phi_{\psi_{j}}(\frac{w}{h})|} \mathrm{d}w \leq C_{1} h^{-\beta_{j}}, \quad \int_{\mathbb{R}} \frac{|\Phi_{K}(w)|^{2}}{|\Phi_{\psi_{j}}(\frac{w}{h})|^{2}} \mathrm{d}w \sim C_{2} h^{-2\beta_{j}}, \\ &\left|\frac{1}{h} \int \int e^{-iw(x-x_{j})/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} \mathrm{d}w \frac{f_{j}(x)}{\max\{f_{j}(x), f_{j}(\frac{1}{a_{N}}\}} \mathrm{d}x\right| = o\left(h^{-2\beta-1}\right) \end{split}$$

uniformly with respect to x_j for some constants $\beta_j > 0$ (j = 1, ..., d) and constants $C_1, C_2, C_3 > 0$, where the constant C_3 does not depend on x_j .

- (A5) The Fourier transform Φ_K of the kernel *K* is symmetric and supported on the interval [-1, 1]. Additionally, there exists a constant $b \in (0, 1]$ such that $\Phi_K(w) = 1$ for all $w \in [-b, b], b > 0$, and $|\Phi_K(w)| \le 1$ for all $w \in \mathbb{R}$.
- (A6) The Fourier transforms $\Phi_{\theta_1}, \ldots, \Phi_{\theta_d}$ of the functions $\theta_1, \ldots, \theta_d$ in the additive model (3) satisfy

$$\int_{\mathbb{R}} |\Phi_{\theta_j}(w)| |w|^{s-1} \mathrm{d}w < \infty \quad \text{for some} \quad s > 1 \quad \text{and} \quad j = 1, \dots, d.$$

(A7) The functions g_1, \ldots, g_d defined in model (5) satisfy

$$\int_{\mathbb{R}} |g_j(z)| |z|^r \mathrm{d} z < \infty \quad \text{for} \quad j = 1, \dots, d$$

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

(A8) For each $N \in \mathbb{N}$ let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ denote independent identically distributed *d*dimensional random variables with marginal densities f_1, \ldots, f_d (which may depend on *N*) such that $f_j(x) \neq 0$ for all $x \in [-\frac{1}{a_N}, \frac{1}{a_N}]$. We also assume that F_j^{-1} exists, where F_j is the distribution function of $X_{1,j}$. Furthermore we assume, that for sufficiently large $N \in \mathbb{N}$

$$f_j(x) \ge f_j\left(\frac{1}{a_N}\right)$$
 whenever $x \in \left[-\frac{1}{a_N}, \frac{1}{a_N}\right]$,

for all j = 1, ..., d.

(A9) If $f_{ijk}(t_i, t_j | t_k)$ and $f_{ij}(t_i | t_j)$ denote the densities of the conditional distribution $\mathbb{P}^{X_i, X_j | X_k}$ and $\mathbb{P}^{X_i | X_j}$, respectively, we assume that there exist integrable functions (with respect to the Lebesgue measure), say $U_{ijk} : \mathbb{R}^2 \to \mathbb{R}$, $\eta_{ij} : \mathbb{R} \to \mathbb{R}$, such that the inequalities

$$f_{ijk}(t_i, t_j | t_k) \le U_{ijk}(t_i, t_j); \ f_{ij}(t_i | t_j) \le \eta_{ij}(t_i)$$

are satisfied for all $t_i, t_j, t_k \in \mathbb{R}$.

Remark 2 Assumption (A1)–(A3) are required for the asymptotic analysis of the backfitting estimator, while (A4)–(A8) are used to analyse the Fourier estimation methods used in the second step of the procedure. In order to demonstrate that these assumptions are satisfied in several cases of practical importance we consider exemplarily Assumption (A4) and (A6). (a) To illustrate Assumption (A4) the convolution function ψ and the kernel *K* are chosen as

$$\psi_j(x) = \frac{\lambda}{2} e^{-\lambda |x|}; \quad K(x) = \frac{\sin(x)}{\pi x},$$

respectively. Furthermore, we choose f_j as density of a uniform distribution on the interval $\left[-\frac{1}{a_N}, \frac{1}{a_N}\right]$ and consider exemplarily the point $x_j = 0$. Note that $\Phi_K(w) = \mathbb{I}_{[-1,1]}(w)$. The integrals in (A4) are obtained by straightforward calculation, that is

$$\begin{split} \int_{\mathbb{R}} \frac{|\Phi_{K}(w)|}{|\Phi_{\psi_{j}}(\frac{w}{h})|} dw &= \int_{[-1,1]} \left(1 + \frac{w^{2}}{h^{2}}\right) dw = \frac{2}{3h^{2}} + 2\\ \int_{\mathbb{R}} \frac{|\Phi_{K}(w)|^{2}}{|\Phi_{\psi_{j}}(\frac{w}{h})|^{2}} dw &= \int_{[-1,1]} \left(1 + \frac{w^{2}}{h^{2}}\right)^{2} dw = \frac{2}{5h^{4}} + \frac{4}{3h^{2}} + 2\\ \frac{1}{h} \int_{[-1/a_{N}, 1/a_{N}]} \int_{[-1,1]} e^{-iw(x - x_{j}^{*})/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} dw \frac{f_{j}(x)}{\max\{f_{j}(x), f_{j}(\frac{1}{a_{N}})\}} dx\\ &= \frac{2}{h} \int_{[-1/a_{N}, 1/a_{N}]} \frac{((h^{2}(x^{2} - 2) + x^{2})\sin(\frac{x}{h}) + 2hx\cos(\frac{x}{h}))}{hx^{3}} dx\\ &= \frac{-2a_{N}\cos(\frac{1}{a_{N}h}) + 2a_{N}^{2}h\sin(\frac{1}{a_{N}h}) + 2hSi(\frac{1}{a_{N}h})}{h} \end{split}$$

and Si(x) denotes the sine-integral $\int_0^x \frac{\sin(y)}{y} dy$. This shows that condition (A4) is satisfied.

(b) In order to illustrate Assumption (A6) let W^m(ℝ) denote the Sobolev space of order m ∈ N; then the assumption θ_j ∈ W^s(ℝ) with s ∈ N\{1} implies condition (A6). Conversely, if (A6) holds with s ∈ N\{1}, then θ_j is (s − 1) times continuously differentiable (see Folland 1984). In other words, (A6) is an assumption regarding the smoothness of the components of the signal θ_j (j = 1,..., d).

Our main result, which is proved in the Appendix, establishes the weak convergence of the estimator $\hat{\theta}_j$ for the *j*th component of the additive signal in model (3). Throughout this paper the symbol \Rightarrow denotes weak convergence.

Theorem 1 *Consider the additive inverse regression model defined by* (1)–(3). *If Assumptions* (A1)–(A8) *are satisfied and additionally the conditions*

$$N^{1/2}h^{\beta_j+1/2}f_j\left(\frac{1}{a_N}\right)^{1/2} \to \infty$$
(20)

$$N^{1/2}h^{3/2}f_j\left(\frac{1}{a_N}\right)^3 \to \infty, \quad N^{1/5}h^{s+\beta_j}f_j\left(\frac{1}{a_N}\right) \to \infty$$
 (21)

🖉 Springer

are fulfilled, then a standardized version of the estimator $\hat{\theta}_j$ defined in (15) converges weakly, that is

$$V_{N,j}^{-1/2}\left(\hat{\theta}_j(x_j) - \mathbb{E}[\hat{\theta}_j(x_j)]\right) \Rightarrow \mathcal{N}(0,1),$$

where

$$\mathbb{E}[\hat{\theta}_j(x_j)] = \theta_j(x_j) + o(h^{s-1}),$$

and the normalizing sequence is given by

$$V_{N,j} = \frac{1}{Nh^2(2\pi)^2} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{-iw(x_j - y)/h} \frac{\Phi_K(w)}{\Phi_{\psi_j}(\frac{w}{h})} dw \right|^2 \frac{(g_j^2(y) + \sigma^2) f_j(y)}{\max\{f_j(y), f_j(\frac{1}{a_N})\}^2} dy$$
(22)

and satisfies

$$N^{1/2}h^{\beta_j+1/2}f_j\left(\frac{1}{a_N}\right)^{1/2} \le V_{N,j}^{-1/2} \le N^{1/2}h^{\beta_j+1/2}.$$
(23)

As a consequence of Theorem 1 we obtain the weak convergence of the additive estimate $\hat{\theta}$ of the signal θ .

Remark 3 If all components except one would be known, it follows from Theorem 3.1 in Hildebrandt et al. (2014) that this component can be estimated at a rate R_N satisfying

$$\frac{c_1}{N^{1/2}h^{1/2+\beta_j}} \le R_n \le \frac{c_2}{N^{1/2}h^{1/2+\beta_j}f_j(a_N^{-1})}$$

(with appropriate constants c_1 and c_2). Consequently, it follows from Theorem 1 that the smooth backfitting operator $\hat{\theta}_j$ defined in (15) has an oracle property and estimates the *j*th component at the one-dimensional rate.

Corollary 1 Consider the inverse regression model defined by (1)–(3) and assume that the assumptions of Theorem 1 are satisfied for all j = 1, ..., d. Then a standardized version of the additive estimator $\hat{\theta}$ defined in (16) converges weakly, that is

$$V_N^{-1/2}\left(\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)]\right) \Rightarrow \mathcal{N}(0, 1).$$

Here

$$\mathbb{E}[\hat{\theta}(x)] = \theta(x) + o(h^{s-1}),$$

and the normalizing factor is given by $V_N = \sum_{j=1}^d V_{N,j} + \sum_{1 \le k \ne l \le d} V_{N,k,l}$, where $V_{N,j}$ is defined in (22),

$$V_{N,k,l} = \frac{1}{Nh^{2}(2\pi)^{2}} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-iw(x_{k}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{k}}(\frac{w}{h})} dw \overline{\int_{\mathbb{R}} e^{-iw(x_{l}-z)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{l}}(\frac{w}{h})}} dw}$$
$$\times \frac{(\sigma^{2} + g_{k}(y)g_{l}(z))f_{k,l}(y,z)}{\max\{f_{k}(y), f_{k}(\frac{1}{a_{N}})\}\max\{f_{l}(y), f_{l}(\frac{1}{a_{N}})\}} d(y,z),$$

and $f_{k,l}$ denotes the joint density of the pair $(X_{k,1}, X_{l,1})$. Moreover, V_N satisfies

$$N^{1/2}h^{\beta_{j^*}+1/2}f_{j^*}\left(\frac{1}{a_N}\right)^{1/2} \le V_N^{-1/2} \le N^{1/2}h^{\beta_{j^*}+1/2}.$$

where $j^* = \operatorname{argmin}_j h^{\beta_j} f_j(1/a_N)$.

Remark 4 It is worthwhile to mention that the results can be extended to dependent data using additional technical arguments. For example, Hildebrandt et al. (2014) investigated marginal integration estimators in an additive inverse regression model, where the error process is m-dependent and similar results as presented in Theorem 1 and Corollary 1 of this paper can derived for the smooth backfitting estimator under this assumption. The details are omitted for the sake of brevity.

Remark 5 It was pointed out by a referee that structured nonparametric (direct) regression techniques have recently found applications in estimating outstanding claims in non-life insurance (the so-called reserving problem). More precisely, Martinez-Miranda et al. (2013) considered the "single chain ladder" model and proposed a continuous approach, which—roughly speaking—consists in an approximation by a two-dimensional multiplicative density model and the corresponding estimation problem (see section 3.3 in this reference). Although this continuous approximation does not contain a convolution, there exist other discrete non-life insurance models, where convolution type structure may occur. For example, Martinez-Miranda et al. (2012) investigated the "double chain ladder" model with a convolution type structure, which is similar to the one considered in this paper. A very interesting problem for future research is the development of a continuous approach to the "double chain ladder" (which might also include covariates) and to use the new techniques of this paper to develop modern statistical estimation tools.

4 Finite sample properties

4.1 Simulation study

In this section we briefly investigate the finite sample properties of the new backfitting estimators by means of a small simulation study. We also compare the two estimators obtained by the marginal integration method with the backfitting estimator proposed in this paper, where the empirical distribution function $\hat{\mathbb{F}}_j$ in (18) and the statistic (19)

are used for the estimation of the unknown marginal distribution F_j and its density f_j , respectively. All results are based on 500 simulation runs. For the sake of brevity we concentrate on three models with a two-dimensional predictor and two distributions for the predictor. To be precise, we consider the models

$$\theta(x_1, x_2) = \theta_1(x_1) + \theta_2(x_2) = e^{-(x_1 - 0.4)^2} + e^{-(x_2 - 0.1)^2},$$
(24)

$$\theta(x_1, x_2) = \theta_1(x_1) + \theta_2(x_2) = x_1 e^{-|x_1|} + (1 + x_2^2)^{-1},$$
(25)

$$\theta(x_1, x_2) = \theta_1(x_1) + \theta_2(x_2) = e^{-|x_1|} + (1 + x_2^2)^{-1},$$
(26)

and assume that the convolution function is given by

$$\psi(x_1, x_2) = \frac{9}{4}e^{-3(|x_1| + |x_2|)}.$$
(27)

Note that the signals in (24) and (25) satisfy the assumptions posed in Sect. 3, while this is not the case for the first component of the signal (26). For the distribution of the explanatory variable we consider an independent and correlated case, that is

a uniform distribution on the square $[1/a_N, 1/a_N]^2$ (28) a two-dimensional normal distribution with mean **0** and variance $\Sigma = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}$.

The variance of the error distribution is given by $\sigma^2 = 0.25$ and for the sequence a_N we used 0.5. In the simulation the bandwidths are chosen in several (nested) steps. At first the bandwidths $h_{d,i}$ in (19) are calculated minimizing the mean integrated squared error of the density estimate. These bandwidths are used in the calculation of the mean integrated squared error of the estimate \hat{g}_i in (12), which is then minimized with respect to the choice of h_B . The final step consists of a calculation of the bandwidth h minimizing the mean integrated squared error of the resulting inverse Fourier transform (15). In practice, this procedure of the mean squared error requires knowledge of the quantities f_i , g_i and for a concrete application we recommend least squares leave-one-out cross-validation for each of these steps. We emphasize that this procedure is computationally expensive because the estimation procedure consists of three nested estimators. Therefore, it cannot be implemented in a simulation study, where the estimate has to be calculated 500 times. However, in a concrete application the estimator has only to be calculated once and works in reasonable time if $N \leq 501$ and $d \leq 5$. For larger sample sizes and larger dimensions a non-exhaustive cross validation should be performed to save computational time.

In Fig. 1 we illustrate the effect of the sample size N = 51, 701 and N = 1001 on the quality of the estimates. For this purpose we choose the model (24) and the design (28). The left and right panels correspond to the first and second components of the additive function. The figures also contain the (pointwise) estimated 5- and 95 %-quantile curves to illustrate the variation of the estimators. We observe that even for the small sample size N = 51 the estimates already reflect the structure

(29)



Fig. 1 Simulated mean, 5 and 95 % quantile of the backfitting estimate on the basis of 500 simulation runs, where model is given by (24) and the design is given by (28). Upper panel N = 51. Middle panel N = 701 Lower panel N = 1001. Left part θ_1 ; right part θ_2 .

of the additive components. With increasing sample size the quality of the estimates improves substantially. For the sake of brevity we restrict ourselves to the sample size N = 701 in the following discussion and an investigation of the impact of the design on the quality of the estimates. In Figs. 2, 3 and 4 we present the estimated mean curves for both components corresponding to model (24)–(26), respectively. Upper parts of the tables show the results for independent components of the predictor, where the case of correlated explanatory variables is displayed in the lower panels. We observe that in models (24) and (25) both components are estimated with reasonable precision (see Figs. 2, 3). The estimators are slightly more accurate under the assumption of an

-0.4

0.0

0.2

Left part θ_1 ; right part θ_2

0.4

0.6

0.8



Fig. 2 Simulated mean, 5 and 95 % quantile of the backfitting estimate on the basis of 500 simulation runs, where model is given by (24) and the design is given by (28) (*upper panel*) and (29) (*lower panel*).

1.0

-0.4

0.0

0.2

0.4

0.6

0.8

1.0

independent design where the differences are more substantial for the estimators of the second component. We note that in the lower-right plot of the Fig. 3, the true function is

outside the limits of the simulation intervals. A possible explanation of this observation is that correlation in the predictors produces a larger systematic bias in the estimates. Problems of this type could be addressed by choosing slightly smaller bandwidths h_B and h in the estimate \hat{g}_j and the inverse Fourier transform (15), respectively.

Model (26) has been included in the study to illustrate the properties of the estimates in models which do not satisfy the assumptions made in Sect. 3. In this case the results are displayed in Fig. 4 and the differences between the uncorrelated and correlated case are even more visible. Here we observe that the first component is not estimated accurately in a neighbourhood of the origin. Consequently, the resulting estimates of the first component are biased in a neighbourhood of the origin. On the other hand, the second component satisfies these assumptions and the right panels of Fig. 4 shows that the second component can be estimated with similar precision as in model (24) and (25).



Fig. 3 Simulated mean, 5- and 95 % quantile of the backfitting estimate on the basis of 500 simulation runs, where model is given by (25) and the design is given by (28) (*upper panel*) and (29) (*lower panel*). *Left part* θ_1 ; *right part* θ_2

In order to compare the new method with the marginal integration method proposed in Hildebrandt et al. (2014), we finally display in Table 1 the simulated integrated mean squared error of both estimators for the models (24) and (25) (model (26) is not included in this comparison because it does not satisfy the assumptions made in Sect. 3). We observe in the case of independent predictors that the backfitting approach yields an improvement of 50 % with respect to the integrated mean squared error criterion. Moreover, in the situation of dependent predictors as considered in (29) the improvement is even more substantial and varies between a factor 3 and 4. We expect that the advantages of the backfitting methodology are even larger with an increasing dimension of the predictor **X**.

4.2 Data example

In this section we illustrate the application of the smooth backfitting estimator for an inverse regression in a problem of live cell imaging in fluorescence microscopy.



Fig. 4 Simulated mean, 5 and 95 % quantile of the backfitting estimate on the basis of 500 simulation runs, where model is given by (26) and the design is given by (28) (*upper panel*) and (29) (*lower panel*). *Left part* θ_1 ; *right part* θ_2

Table 1 Simulated mean integrated squared error of the smooth backfitting estimator $\hat{\theta}_j$ (j = 1, 2) proposed in this paper and of the marginal estimator $\hat{\theta}_j^{MI}$ proposed by Hildebrandt et al. (2014)	Design model	(28)		(29)	
		(24)	(25)	(24)	(25)
	$\hat{ heta}_1$	0.00179	0.00189	0.00500	0.00353
	$\hat{ heta}_2$	0.00154	0.00258	0.00488	0.00345
	$\hat{\theta}_1^{MI}$	0.00347	0.00365	0.02219	0.00934
	$\hat{\theta}_2^{MI}$	0.00311	0.00354	0.01917	0.01092
	-2				

It is well known that images from optical microscopy do not show the true shape of the object, but a (at least slightly) blurred image which can be described as the convolution of the true image with the point spread function (PSF) of the microscope. The PSF describes the imaging properties of the optical device, in particular due to the diffraction of the beam at optical surfaces. Moreover, the image at our disposal is acquired with a CCD (charge-coupled device).



Fig. 5 First, middle and last timestep of the sequence of nine images taken over ≈ 1 min. Intracellular tubular shaped membrane compartments were labelled with the *red* lipophilic fluorescent dye DiI

Figure 5 shows a time stack of living HeLa cells which are stained with a fluorescent dye. The dye DiI labels the intracellular lipid membranes in living cells and is, e.g. used to monitor intracellular transport processes by fluorescence microscopy. For live-cell imaging the HeLa-Cells (ATCC) cultured on coverslips were incubated with the lipophilic carbocyan dye DiI (Molecular Probes). The living cells were then subjected to fluorescence microscopy using a confocal laser scanning microscope (Leica TCS) equipped with a HeNe-Laser. The samples were excited at 543 nm. Images were processed using the supplied Leica Confocal Software. Finally, the number of pixels per image is $512 \times 512 = 262,144$. The resulting image series consists of nine images which were taken over ≈ 1 min and show the trafficking tubular membrane-coated compartments in the cell which are part of the intracellular transport system. The first, middle and last time step of the image sequence are shown in Fig. 5.

At our disposal is a sequence of 9 images of cells imaged by fluorescence microscopy measured of 512×512 pixels each showing the same true object. In addition to a small variability of the images due to ongoing processes in the living cells, there is (may) be time-dependent contributions to the images from a general decay in the overall brightness of the image to a decrease of luminosity of the fluorophors on the one hand or temperature changes in the sample which cools down to ambient temperature on the other hand, resulting in changing background noise levels during the imaging.

We are interested in understanding such a general time-dependent contribution to the image because a profound characterization of such general effects may enhance the quality of the biological analysis substantially. If such general time-dependent background effects are quantified by the presented method it might in a subsequent step be possible to subtract these artificial effects from the image series in a way that, e.g. enables the detection of small-scale events in the images that would otherwise remain hidden.

For the biological analysis of the sample it is essential that the image is taken under constant conditions throughout the complete time period, i.e. changes in the background fluorescence may impair later analysis of the image, especially with regard to time-dependent processes. For this purpose we model the data in an additive framework with an additive time-dependent component for the contribution to the general image structure, that is



Fig. 6 Left estimated signal \hat{f}_{image} , right estimated time-dependent contribution $\hat{f}_{image}^{\dagger}$

 $f(x_1, x_2, t) = f_{\text{image}}(x_1, x_2) + f_{\text{time}}(t).$

Here f_{image} is the true shape of the image, which is only at very small scales changing with time in shape (cf. Fig. 5 for a comparison of the first, middle and last time step of the image sequence), and the function f_{image} is modelling a time-dependent contribution to the image.

We apply the smooth backfitting methodology with a kernel density estimator with a Gaussian kernel *L* (with bandwidth 0.2) and design points $X_{ij} = (i, j, k)$ ($1 \le i, j \le 512, 1 \le k \le 9$). The PSF is estimated from the shape of a point-like test object (bead, size = 200 nm) with shape

$$\psi(x_1, x_2) = \frac{\lambda}{2} \exp(-\lambda(|x_1| + |x_2|)).$$

Figure 6 shows the resulting estimates \hat{f}_{image} and \hat{f}_{image} , respectively. Moreover, $\hat{g}_0 = 76.0$. From this figure (r.h.s.) we observe that the time-dependent component is monotonically decreasing which can be probably attributed to an overall decrease in fluorescence. To observe, e.g. intracellular transport processes in living cells the transport compartments inside mammalian are labelled with a fluorescent dye and then subjected to fluorescence microscopy to take a time series of images over several minutes. In order to analyse the resulting image series it is important that the background conditions remain constant during the imaging period and that it is possible to differ between biologically relevant changes over time and artificial changes due to the experimental setup or the imaging device.

Appendix: Proof of Theorem 1

Let *p* denote the density of the transformed predictor $(Z_{1,1}, \ldots, Z_{1,d})^T$. It is shown in Mammen et al. (1999) that the smooth backfitting algorithm (9) produces a sequence of estimates $(\hat{g}_1^{*(r)}, \ldots, \hat{g}_d^{*(r)})_{r=0,1,\ldots}$ converging in $L^2(p)$ with geometric rate to a vector $(\overline{g}_1, \ldots, \overline{g}_d)$ which satisfies the system of equations

$$\overline{g}_{j}(z_{j}) = \hat{g}_{j}^{*}(z_{j}) - \sum_{k \neq j} \int \overline{g}_{k}(z_{k}) \left[\frac{\hat{p}_{jk}(z_{j}, z_{k})}{\hat{p}_{j}(z_{j})} - \hat{p}_{k,[j+]}(z_{k}) \right] dz_{k} - g_{0,j}^{*} \quad j = 1, \dots, d,$$
(30)

where $g_{0,j}^*$ is defined in (11). Therefore, the asymptotic properties of the smooth backfitting operator can be investigated replacing in (14) the random variables $U_{k,j}$ defined in (13) by their theoretical counterparts

$$\tilde{U}_{k,j} = Y_k - \sum_{\substack{i=1\\i\neq j}}^d \tilde{g}_i(X_{k,i}) - \hat{g}_0,$$

where $\tilde{g}_i(X_{k,i}) = \bar{g}_i(Z_{k,i})$ (i = 1, ..., d; k = 1, ..., N) and $\tilde{g}_i = \bar{g}_i \circ F$ (i = 1, ..., d). This yields the representation

$$\tilde{U}_{k,j} = g_j(X_{k,j}) + \varepsilon_k + \sum_{\substack{i=1\\i\neq j}}^d (g_i(X_{k,i}) - \tilde{g}_i(X_{k,i})) = g_j(X_{k,j}) + \varepsilon_k + B_{j,k,N}, \quad (31)$$

where the last equality defines the random variables $B_{j,k,N}$ in an obvious manner. The results of Mammen et al. (1999) imply

$$B_{j,k,N} = O_p(N^{-1/5})$$
(32)

uniformly with respect to $j \in \{1, ..., d\}$ and $k \in \{1, ..., N\}$.

The assertion of Theorem 1 is now proved in four steps establishing the following statements:

$$b_{\hat{\theta}_j}(x_j) = \mathbb{E}[\hat{\theta}_j(x_j)] - \theta_j(x_j) = o(h^{s-1})$$
(33)

$$Var(\hat{\theta}_{i}(x_{i})) = V_{N,i}(1+o(1))$$
(34)

$$V_{n,j}$$
 satisfies (23) (35)

$$|\operatorname{cum}_{l}(V_{N,j}^{-1/2}\hat{\theta}_{j}(x_{j}))| = o(1) \text{ for all } l \ge 3,$$
(36)

where $V_{N,j}$ is the normalizing factor defined in (22) and cum_l denotes the *l*th cumulant (see Brillinger 2001).

Proof of (33): We first determine the expectation of the estimator $\hat{\theta}_j$ observing that the estimator $\hat{\theta}_j$ is linear, i.e.

$$\hat{\theta}_j(x_j) = \sum_{k=1}^N w_{j,N}(x_j, X_{k,j}) \tilde{U}_{k,j},$$
(37)

🖄 Springer

where the weights $w_{i,N}(x_i, X_{k,i})$ are defined by

$$w_{j,N}(x_j, X_{k,j}) = \frac{1}{2\pi Nh} \int_{\mathbb{R}} e^{-iw(x_j - X_{k,j})/h} \frac{\Phi_K(w)}{\Phi_{\psi_j}(\frac{w}{h})} dw \frac{1}{\max\{f_j(X_{k,j}), f_j(\frac{1}{a_n})\}},$$
(38)

and we have replaced the quantities $U_{k,j}$ by $\tilde{U}_{k,j}$ as described at the beginning of the proof. This representation gives

$$\mathbb{E}[\hat{\theta}_i(x_i)] = E_1 + E_2, \tag{39}$$

where the terms E_1 and E_2 are defined by

$$E_{1} = \mathbb{E}\left[\sum_{k=1}^{N} g_{j}(X_{k,j})w_{j,N}(x_{j}, X_{k,j})\right], \quad E_{2} = \mathbb{E}\left[\sum_{k=1}^{N} B_{j,k,N}w_{j,N}(x_{j}, X_{k,j})\right].$$
(40)

Using the definition of $B_{j,k,N}$ and (32) the term E_2 can be estimated as follows:

$$|E_{2}| \leq \mathbb{E}\left[\sum_{k=1}^{N} \sum_{\substack{i=1\\i\neq j}}^{d} |g_{i}(X_{k,i}) - \tilde{g}_{i}(X_{k,i})| \max_{k} |w_{j,N}(x_{j}, X_{k,j})|\right] \\ \leq \frac{C}{h^{\beta_{j}+1} f_{j}(\frac{1}{a_{N}})} \mathbb{E}\left[\sum_{\substack{i=1\\i\neq j}}^{d} |g_{i}(X_{k,i}) - \tilde{g}_{i}(X_{k,i})|\right] \leq \frac{C}{N^{1/5} h^{\beta_{j}+1} f_{j}(\frac{1}{a_{N}})} = o(h^{s-1}),$$

$$(41)$$

where we used the representation (38) and Assumption (A4). The second inequality in (41) follows from the fact that

$$\mathbb{E}[|g_i(X_{k,i}) - \tilde{g}_i(X_{k,i})|] = O(N^{-1/5}).$$
(42)

In order to establish this statement note that $g_i(X_{k,i}) - \tilde{g}_i(X_{k,i}) = O_P(N^{-1/5})$ (uniformly with respect to k = 1, ..., N). The proof of the L^1 -convergence follows along the lines of the proof of the stochastic convergence in Mammen et al. (1999). Here one additionally shows in each step of the backfitting iteration stochastic convergence and L^1 - convergence (see Hildebrandt 2013 for details).

Similarly, we obtain from the definition of the weights $w_{j,N}(x_j, X_{k,j})$ in (38) the representation

$$E_{1} = \frac{1}{2\pi h} \int_{\mathbb{R}} g_{j}(y) \int_{\mathbb{R}} e^{-iw(x_{j}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} dw \frac{f_{j}(y)}{\max\{f_{j}(y), f_{j}(\frac{1}{a_{N}})\}} dy$$

$$= \frac{1}{2\pi h} \int_{\mathbb{R}} \Phi_{g_{j}}\left(\frac{w}{h}\right) e^{-iwx_{j}/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} dw$$

$$-\frac{1}{2\pi h} \int_{\mathbb{R}} g_{j}(y) \int_{\mathbb{R}} e^{-iw(x_{j}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} dw \left(1 - \frac{f_{j}(y)}{\max\{f_{j}(y), f_{j}(\frac{1}{a_{N}})\}}\right) dy,$$

$$= \theta_{j}(x_{j}) - F_{1} - F_{2},$$
(43)

where the terms F_1 and F_2 are defined by

$$F_1 = \frac{1}{2\pi h} \int_{\mathbb{R}} \Phi_{\theta_j}\left(\frac{w}{h}\right) e^{-iwx_j/h} \left(1 - \Phi_K(w)\right) \mathrm{d}w,$$

$$F_2 = \frac{1}{2\pi h} \int_{\mathbb{R}} g_j(y) \int_{\mathbb{R}} e^{-iw(x_j - y)/h} \frac{\Phi_K(w)}{\Phi_{\psi_j}(\frac{w}{h})} \mathrm{d}w \left(1 - \frac{f_j(y)}{\max\{f_j(y), f_j(\frac{1}{a_N})\}}\right) \mathrm{d}y,$$

respectively. The term F_1 can be estimated using Assumption (A6), that is

$$\begin{split} |F_{1}| &\leq \frac{1}{2\pi h} \int_{\mathbb{R}} |\Phi_{\theta_{j}}\left(\frac{w}{h}\right) |1 - \Phi_{K}(w)| \mathrm{d}w \leq \frac{1}{\pi h} \int_{[-b,b]^{c}} |\Phi_{\theta_{j}}\left(\frac{w}{h}\right)| \mathrm{d}w \\ &\leq \frac{1}{\pi} \int_{[-b/h,b/h]^{c}} \frac{1}{|y|^{s-1}} |y|^{s-1} |\Phi_{\theta_{j}}(y)| \mathrm{d}y \\ &\leq \frac{h^{s-1}}{b^{s-1}\pi} \int_{[-b/h,b/h]^{c}} |y|^{s-1} |\Phi_{\theta_{j}}(y)| \mathrm{d}y = o(h^{s-1}), \end{split}$$

while the term F_2 is estimated similarly, using Assumption (A4), (A7) and (A8) that is

$$\begin{split} |F_2| &\leq \frac{1}{2\pi h} \int_{\mathbb{R}} |g_j(y)| \int_{\mathbb{R}} \frac{|\Phi_K(w)|}{|\Phi_{\psi_j}(\frac{w}{h})|} \mathrm{d}w \Big| 1 - \frac{f_j(y)}{\max\{f_j(y), f_j(\frac{1}{a_N})\}} \Big| \mathrm{d}y \\ &\leq \frac{1}{2\pi h} \int_{([-1/a_N, 1/a_N])^c} |g_j(y)| \mathrm{d}y \int_{\mathbb{R}} \frac{|\Phi_K(w)|}{|\Phi_{\psi_j}(\frac{w}{h})|} \mathrm{d}w = O\left(\frac{a_N^r}{h^{1+\beta_j}}\right) = o(h^{s-1}). \end{split}$$

From these estimates and (43) we obtain $E_1 = \theta_j(x_j) + o(h^{s-1})$, and the assertion (33) now follows from the decomposition (39) and (41).

Proof of (34): using standard results for cumulants (see Brillinger 2001) the variance of the estimate $\hat{\theta}_i$ can be calculated as

$$\operatorname{Var}(\hat{\theta}_j(x_j)) = S_1 + S_2 + S_3 + 2S_4 + 2S_5 + 2S_6, \tag{44}$$

where

$$S_{1} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(\varepsilon_{k}w_{j,N}(x_{j}, X_{k,j}), \varepsilon_{l}\overline{w_{j,N}(x_{j}, X_{l,j})})$$

$$S_{2} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(g_{j}(X_{k,j})w_{j,N}(x_{j}, X_{k,j}), g_{j}(X_{l,j})\overline{w_{j,N}(x_{j}, X_{l,j})})$$

$$S_{3} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(B_{j,k,N}w_{j,N}(x_{j}, X_{k,j}), B_{j,l,N}\overline{w_{j,N}(x_{j}, X_{l,j})})$$

$$S_{4} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(\varepsilon_{k}w_{j,N}(x_{j}, X_{k,j}), g_{j}(X_{l,j})\overline{w_{j,N}(x_{j}, X_{l,j})})$$

$$S_{5} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(\varepsilon_{k}w_{j,N}(x_{j}, X_{k,j}), B_{j,l,N}\overline{w_{j,N}(x_{j}, X_{l,j})})$$

$$S_{6} = \sum_{k=1}^{N} \sum_{l=1}^{N} \operatorname{cum}(g_{j}(X_{k,j})w_{j,N}(x_{j}, X_{k,j}), B_{j,l,N}\overline{w_{j,N}(x_{j}, X_{l,j})}).$$

It is easy to see that $S_4 = 0$ because of $\mathbb{E}[\varepsilon_k] = 0$ and the independence of ε_k and \mathbf{X}_k . We will show that the first two terms S_1 and S_2 determine the variance and that the terms S_3 , S_5 and S_6 are of smaller order. For a proof of the latter result we concentrate on the sixth term because the results for the terms S_3 and S_5 can be treated analogously.

As ε_k , ε_l , $X_{k,j}$ and $X_{l,j}$ are independent for $k \neq l$ the term S_1 can be written as

$$N\operatorname{cum}(\varepsilon_{k}w_{j,N}(x_{j}, X_{k,j}), \varepsilon_{k}\overline{w_{j,N}(x_{j}, X_{k,j})})$$

= $N\operatorname{cum}(\varepsilon_{k}, \varepsilon_{k})\operatorname{cum}(w_{j,N}(x_{j}, X_{k,j}), \overline{w_{j,N}(x_{j}, X_{k,j})})$
+ $N\operatorname{cum}(\varepsilon_{k}, \varepsilon_{k})\operatorname{cum}(w_{j,N}(x_{j}, X_{k,j}))\operatorname{cum}(\overline{w_{j,N}(x_{j}, X_{k,j})}),$

where we used the product theorem for cumulants and $\mathbb{E}[\varepsilon_k] = 0$. Now a straightforward calculation gives

$$S_{1} = \frac{\sigma^{2}}{Nh^{2}(2\pi)^{2}} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{-iw(x_{j}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} dw \right|^{2} \frac{f_{j}(y)}{\max\{f_{j}(y), f_{j}(\frac{1}{a_{N}})\}^{2}} dy$$

× (1 + o(1)).

The second summand in (45) can be calculated in the same way and we obtain

$$S_{2} = \frac{1}{Nh^{2}(2\pi)^{2}} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{-iw(x_{j}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} \mathrm{d}w \right|^{2} \frac{g_{j}^{2}(y)f_{j}(y)}{\max\{f_{j}(y), f_{j}(\frac{1}{a_{N}})\}^{2}} \mathrm{d}y$$

× (1 + o(1)).

Deringer

In a last step we investigate the sixth summand of (45) (the other terms S_3 and S_5 are treated in the same way). By the product theorem and the definition of the cumulants we obtain for this term

$$S_{6} = -\sum_{k \neq l} \sum_{\substack{i=1 \ i \neq j}}^{d} \operatorname{Cov} \left(g_{j}(X_{k,j}) w_{j,N}(x_{j}, X_{k,j}), \overline{g}_{i}(F(X_{l,i})) w_{j,N}(x_{j}, X_{l,j}) \right) \times (1 + o(1)),$$

where we used the definitions of $B_{j,l,N} = \sum_{i \neq j} (g_i(X_{l,i}) - \tilde{g}_i(X_{l,i}))$ and $\overline{g}_i = \tilde{g}_i \circ F_i^{-1}$. We introduce the weights

$$q_{mj}(X_{l,i}) = \frac{L\left(\frac{F_j(X_{m,i}) - F_j(X_{l,i})}{h_B}\right)}{\sum_{s=1}^N L\left(\frac{F_j(X_{s,i}) - F_j(X_{l,i})}{h_B}\right)} \quad l, m = 1, \dots, N; \quad i = 1, \dots, d,$$

denoted by

$$\hat{g}_i^*(F_i(X_{l,i})) = \sum_{m=1}^N q_{mi}(X_{l,i})Y_m \quad l = 1, \dots, N; \quad i = 1, \dots, d$$
(45)

the one-dimensional Nadaraya–Watson estimator from the data $F_i(X_{1,i}), \ldots, F_i(X_{N,i})$ evaluated at the point $F_i(X_{l,i})$ and define

$$v_{mi}(X_{l,i}, z_m) = \frac{\hat{p}_{im}(F_i(X_{l,i}), z_m)}{\hat{p}_i(F_i(X_{l,i}))} - \hat{p}_{m,[i+]}(z_m) \quad i, m = 1, \dots, d; \quad l = 1, \dots, N$$

as the integrand in equation (30). This yields for the term S_6 the decomposition

$$S_6 = (B - A)(1 + o(1)), \tag{46}$$

where the terms A and B are defined by

$$A = \sum_{k \neq l} \sum_{\substack{i=1 \ i \neq j}}^{d} \operatorname{Cov}\left(g_j(X_{k,j}) w_{j,N}(x_j, X_{k,j}), w_{j,N}(x_j, X_{l,j}) \sum_{m=1}^{N} q_{mi}(X_{l,i}) Y_m\right)$$

and

$$B = \sum_{k \neq l} \sum_{\substack{i=1 \ i \neq j}}^{d} \sum_{\substack{m=1 \ m \neq i}}^{d} \operatorname{Cov} \Big(g_j(X_{k,j}) w_{j,N}(x_j, X_{k,j}), \\ \times \int \tilde{g}_m(z_m) v_{mi}(X_{l,i}, z_m) dz_m + g_{0,i}^* \Big) w_{j,N}(x_j, X_{l,j}) \Big),$$

Deringer

respectively. We start with the estimation of the term A calculating each covariance separately, that is

$$\left| \operatorname{Cov} \left(g_j(X_{k,j}) w_{j,N}(x_j, X_{k,j}), w_{j,N}(x_j, X_{l,j}) \sum_{m=1}^N q_{mi}(X_{l,i}) Y_m \right) \right| \le \left(H_1 + H_2 \right) (1 + o(1)),$$
(47)

where the terms H_1 and H_2 are defined by

$$H_{1} = \frac{1}{Nh_{B}} \left| \sum_{r=1}^{d} \mathbb{E} \left[g_{j}(X_{k,j}) w_{j,N}(x_{j}, X_{k,j}) L \left(\frac{F_{i}(X_{k,i}) - F_{i}(X_{l,i})}{h_{B}} \right) \right]$$
$$g_{r}(X_{k,r}) \overline{w_{j,N}(x_{j}, X_{l,j})} \right]$$
$$H_{2} = \frac{1}{Nh_{B}} \left| \mathbb{E} \left[g_{j}(X_{k,j}) w_{j,N}(x_{j}, X_{k,j}) \right] \mathbb{E} \left[L \left(\frac{F_{i}(X_{k,i}) - F_{i}(X_{l,i})}{h_{B}} \right) \right]$$
$$\times \sum_{r=1}^{d} g_{r}(X_{k,r}) \overline{w_{j,N}(x_{j}, X_{l,j})} \right]$$

and we used the fact that the kernel density estimate

$$\frac{1}{Nh_B}\sum_m L\left(\frac{F_i(X_{m,i}) - F_i(X_{l,i})}{h_B}\right) = \hat{p}_i(F_i(X_{l,i}))$$

in the denominator of the Nadaraya–Watson estimate (45) converges uniformly to 1 as $F_i(X_{l,i})$ is uniformly distributed on the interval [0, 1] (see Giné and Guillou 2002). We first investigate the term H_1 and obtain by a tedious calculation using assumption (A4) and (A9)

$$\begin{aligned} H_{1} &\leq \frac{(1+o(1))}{N^{3}h^{2}} \Big| \int_{\mathbb{R}^{2}} \left(\int_{\mathbb{R}} g_{j}(t_{j}) \int_{\mathbb{R}} e^{-iw(x_{j}-t_{j})/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}\left(\frac{w}{h}\right)} \mathrm{d}w \right. \\ &\times \frac{f_{j}(t_{j}) f_{irj}(t_{i}, t_{r}|t_{j})}{\max\{f_{j}(t_{j}), f_{j}(1/a_{N})\}} dt_{j} \Big) \\ &\times \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-iw(x_{j}-s_{j})/h} \frac{\Phi_{K}(w)}{\Phi(w/h)} \mathrm{d}w \frac{f_{j}(s_{j}) f_{ij}(t_{i}|s_{j})}{\max\{f_{j}(t_{j}), f_{j}(1/a_{N})\}} ds_{j} \Big) g_{j}(t_{r}) dt_{i} dt_{r} \Big| \\ &\leq \frac{C}{N^{3}h^{2}} \int_{\mathbb{R}^{2}} \left| \int_{\mathbb{R}} e^{-iw(x_{j}-t_{j})/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(w/h)} \mathrm{d}w \frac{f_{j}(t_{j})}{\max\{f_{j}(t_{j}), f_{j}(1/a_{N})\}} dt_{j} \right| \end{aligned}$$

Deringer

 \square

$$\times \left| \int_{\mathbb{R}} e^{-iw(x_j - s_j)/h} \frac{\Phi_K(w)}{\Phi_{\psi_j}(w/h)} dw \frac{f_j(s_j)}{\max\{f_j(t_j), f_j(1/a_N)\}} ds_j \right|$$
$$U_{irj}(t_i, t_r)\eta_{ij}(t_i)dt_i dt_r$$
$$= o\left(\frac{1}{N^3 h^{2\beta + 1}}\right)$$

uniformly with respect to k, l. A similar calculation yields

$$H_2 \leq \frac{1}{Nh_B} \left| \frac{E_1}{N} \mathbb{E} \left[L \left(\frac{F_i(X_{k,i}) - F_i(X_{l,i})}{h_B} \right) \right] \right|$$
$$\sum_{r=1}^d g_r(X_{k,r}) \overline{w_{j,N}(x_j, X_{l,j})} \right] = o \left(\frac{1}{N^3 h^{2\beta + 1}} \right)$$

(uniformly with respect to k, l) where we use the estimate (40) in the first step. Consequently, the term A in (47) can be bounded by $A = o(1/Nh^{2\beta+1})$ A tedious calculation using similar arguments yields for the term $B = O(1/Nh^{2\beta+1})$ and by (46) the sum S_6 is of the same order. Moreover, it will be shown in the proof of (35) below that this order is smaller than the order of the first two summands S_1 and S_2 in (45) which gives

$$S_6 = O\left(\frac{1}{Nh^{2\beta+1}}\right) = o(S_j) \quad j = 1, 2.$$

A similar calculation for the terms S_3 and S_5 finally yields

$$\begin{aligned} \operatorname{Var}(\hat{\theta}_{j}(x_{j})) &= \frac{1}{Nh^{2}(2\pi)^{2}} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{-iw(x_{j}-y)/h} \frac{\Phi_{K}(w)}{\Phi_{\psi_{j}}(\frac{w}{h})} \mathrm{d}w \right|^{2} \\ &\times \frac{(g_{j}^{2}(y) + \sigma^{2}) f_{j}(y)}{\max\{f_{j}(y), f_{j}(\frac{1}{a_{N}})\}^{2}} \mathrm{d}y \times (1 + o(1)), \\ &= V_{N,j} (1 + o(1)), \end{aligned}$$

which proves (34).

Proof of (35): As g_j is bounded for all j = 1, ..., d and $\max\{f_j(y), f_j(\frac{1}{a_N})\}^2 \ge f_j(y)f_j(\frac{1}{a_N})$ the term $V_{N,j}$ defined in (22) can be estimated as follows:

$$\begin{aligned} |V_{N,j}| &\leq \frac{C}{Nh(2\pi)^2 f_j(\frac{1}{a_N})} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{-iw(x_j/h-y)} \frac{\Phi_K(w)}{\Phi_{\psi_j}(\frac{w}{h})} \mathrm{d}w \right|^2 \mathrm{d}y \\ &= \frac{C}{Nh(2\pi)^2 f_j(\frac{1}{a_N})} \int_{\mathbb{R}} \frac{|\Phi_K(w)|^2}{|\Phi_{\psi_j}(\frac{w}{h})|^2} \mathrm{d}w \;, \end{aligned}$$

where C is a constant and we used Parseval's equality for the last identity (see Kammler 2007). Now assumption (A4) yields the upper bound, that is $|V_{N,j}| \leq$

 $C/Nh^{1+2\beta_j} f_j(\frac{1}{a_N})$. From the assumption $f_j(x)^{-1} \ge C$ and again Parsevals equality we also get the lower bound $|V_{N,j}| \ge C/Nh^{1+2\beta_j}$, which completes the proof of the estimate (23).

Proof of (36): Observing the representation (37) the *l*th cumulant of the estimate $\hat{\theta}_j$ can be estimated as follows:

$$|\operatorname{cum}_{l}(\hat{\theta}_{j}(x_{j}))| = \left| \sum_{k_{1},\dots,k_{l}=1}^{N} \operatorname{cum}\left(\tilde{U}_{k_{1},j}w_{j,N}(x_{j},X_{k_{1},j}),\dots,\tilde{U}_{k_{l},j}w_{j,N}(x_{j},X_{k_{l},j})\right) \right| \le G_{1} + G_{2},$$

where the terms G_1 and G_2 are defined by

$$G_{1} = \Big| \sum_{k_{1},\dots,k_{l}=1}^{N} \operatorname{cum} \Big(A_{k_{1},j} w_{j,N}(x_{j}, X_{k_{1},j}), \dots, A_{k_{l},j} w_{j,N}(x_{j}, X_{k_{l},j}) \Big) \Big|$$

$$G_{2} = \Big| \sum_{k_{1},\dots,k_{l}=1}^{N} \sum_{s=1}^{l} \binom{l}{s} \operatorname{cum} \Big(B_{j,k_{1},N} w_{j,N}(x_{j}, X_{k_{1},j}), \dots, B_{j,k_{s},N} w_{j,N}(x_{j}, X_{k_{s},j}), A_{k_{s+1},j} w_{j,N}(x_{j}, X_{k_{s+1},j}), \dots, A_{k_{l},j} w_{j,N}(x_{j}, X_{k_{l},j}) \Big) \Big|,$$

and we introduce the notation $A_{k_i,j} = g_j(X_{k_i,j}) + \varepsilon_{k_i}$. Exemplarily, we investigate the first term of this decomposition; the term G_2 is treated similarly. As the random variables $A_{k_1,j}w_{j,N}(x_j, X_{k_1,j})$ and $A_{k_2,j}w_{j,N}(x_j, X_{k_2,j})$ are independent for $k_1 \neq k_2$ and identically distributed for $k_1 = k_2$, it follows that

$$G_{1} = N \left| \operatorname{cum}_{l}(A_{k,j} w_{j,N}(x_{j}, X_{k,j})) \right| \leq N \sum_{s=0}^{l} {\binom{l}{s}} \sum_{\substack{\mathbf{j} \in \{0,1\}^{l} \\ j_{1} + \dots + j_{l} = s}} \\ \times \sum_{\nu} \prod_{k=1}^{p} \operatorname{cum}(A_{ij}, ij \in \nu_{k}) \right|,$$

where we used the product theorem for cumulants (see Brillinger 2001) and the third sum extends over all indecomposable partitions of the table



with $A_{i1} = \varepsilon_1$ $(1 \le i \le s)$, $A_{i2} = w_{j,N}(x_j, X_{1,j})$ $(1 \le i \le s)$ and $A_{ij} = g_j(X_{1,j})w_{j,N}(x_j, X_{1,j})$ $(s + 1 \le i \le l)$. In order to illustrate how to estimate this expression we consider exemplarily the case l = 3, where G_1 reduces to

$$G_{1} = N \sum_{s=0}^{3} {\binom{3}{s}} \sum_{\substack{\mathbf{j} \in \{0,1\}^{3} \\ j_{1} + \dots + j_{3} = s}} \left| \sum_{\nu} \prod_{k=1}^{p} \operatorname{cum}(A_{ij}, ij \in \nu_{k}) \right|.$$

As ε is independent of X_1 and has mean 0 the partitions in G_1 with s = 1 vanish. The terms corresponding to s = 0, 2, 3 contain only quantities of the form

$$\begin{aligned} & \operatorname{cum}_{3}(g_{j}(X_{1,j})w_{j,N}(x_{j}, X_{1,j})), \\ & \sigma^{2}\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}), w_{j,N}(x_{j}, X_{1,j}), g_{j}(X_{1,j})w_{j,N}(x_{j}, X_{1,j})), \\ & \sigma^{2}\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}))\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}), g_{j}(X_{1,j})w_{j,N}(x_{j}, X_{1,j}))) \\ & \kappa_{3}\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}), w_{j,N}(x_{j}, X_{1,j}), w_{j,N}(x_{j}, X_{1,j})), \\ & \kappa_{3}\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}), w_{j,N}(x_{j}, X_{1,j}))\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j})), \\ & \kappa_{3}\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j}))\operatorname{cum}(w_{j,N}(x_{j}, X_{1,j})), \\ \end{aligned}$$

where κ_3 denotes the third cumulant of ε_1 . As the inequality

$$\mathbb{E}\Big[|g_j(X_{1,j})w_{j,N}(x_j,X_{1,j})|^{b_r}|w_{j,N}(x_j,X_{1,j})|^{a_r-b_r}\Big] \le \frac{C}{N^{a_r}h^{a_r(\beta_j+1)}f_j(\frac{1}{a_n})^{a_r}}$$

holds for $0 \le b_r \le a_r$ all terms can be bounded by $C/(N^3 h^{3(\beta_j+1)} f_j(\frac{1}{a_n})^3)$. This yields

$$N^{3/2}h^{3\beta_j+3/2}G_1 \le CN^{3/2+1}h^{3\beta_j+3/2}\frac{1}{N^3h^{3(\beta_j+1)}f_j(\frac{1}{a_n})^3} = o(1)$$

where we used the conditions on the bandwidth in the last step. Similar calculations for the general case show

$$N^{l/2}h^{l\beta_j+l/2}G_1 = O\left(\left(N^{l/2-1}h^{l/2}f_j\left(\frac{1}{a_N}\right)^l\right)^{-1}\right) = o(1),$$

whenever $l \ge 3$. The term G_2 can be calculated in the same way, where for example, one additionally has to use the estimate $\text{Cov}(B_{j,k,N}, \varepsilon_l) = O(1/N)$ uniformly with respect to all j = 1, ..., d, and k, l = 1, ..., N, which follows from the definition of the backfitting estimator.

Acknowledgments The authors thank Martina Stein and Alina Dette, who typed parts of this manuscript with considerable technical expertise. 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). The authors are also grateful to two unknown referees. Their constructive comments on the first version of the paper led to a substantial improvement of the manuscript.

References

- Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G. (2009). Image deblurring with Poisson data: From cells to galaxies. *Inverse Problems*, 25(12), 123006.
- Bissantz, N., Hohage, T., Munk, A., Ruymgaart, F. (2007). Convergence rates of general regularization methods for statistical inverse problems. SIAM Journal on Numerical Analysis, 45, 2610–2636.
- Brillinger, D. R. (2001). Time series data analysis and theory. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Carroll, R. J., Härdle, W., Mammen, E. (2002). Estimation in an additive model when the parameters are linked parametrically. *Econometric Theory*, 18(4), 886–912.
- Cavalier, L. (2008). Nonparametric statistical inverse problems. Inverse Problems, 24(3), 034004.
- Diggle, P. J., Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society Series B*, 55, 523–531.
- Engl, H. W., Hanke, M., Neubauer, A. (1996). Regularization of inverse problems, volume 375 of mathematics and its applications. Dordrecht: Kluwer Academic Publishers Group.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. Annals of Statistics, 19, 1257–1272.
- Folland, G. B. (1984). Real Analysis-Modern Techniques and their Applications. New York: Wiley.
- Giné, E., Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, 38(6), 907–921.
- Hengartner, N. W., Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, 95(2), 246–272.
- Hildebrandt, T. (2013). Additive Modelle im inversen Regressionsproblem mit Faltungsoperator. PhD thesis, Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.
- Hildebrandt, T., Bissantz, N., Dette, H. (2014). Additive inverse regression models with convolution-type operators. *Electronic Journal of Statistics*, 8, 1–40.
- Kaipio, J., Somersalo, E. (2010). Statistical and computational inverse problems. Berlin: Springer.
- Kammler, D. W. (2007). A first course in fourier analysis. Cambridge: Cambridge University Press.
- Linton, O. B., Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1), 93–100.
- Mair, B. A., Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. SIAM Journal on Applied Mathematic, 56, 1424–1444.
- Mammen, E., Linton, O. B., Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, 27(5), 1443–1490.
- Martinez-Miranda, M. D., Nielsen, J. P., Sperlich, S., Verrall, R. (2013). Continuous chain ladder: Reformulating and generalising a classical insurance problem. *Expert Systems with Applications*, 40(14), 5588–5603.
- Martinez-Miranda, M. D., Nielsen, J. P., Verrall, R. (2012). Double chain ladder. Astin Bulletin, 42(1), 59–76.
- Nielsen, J. P., Sperlich, S. (2005). Smooth backfitting in practice. Journal of the Royal Statistical Society Serial B, 67(1), 43–61.
- Saitoh, S. (1997). Integral transforms, reproducing kernels and their applications. Harlow: Longman.
- Stefanski, L., Carroll, R. J. (1990). Deconvoluting kernel density estimators. Statistics, 21, 169–184.
- van der Vaart, A. W. (1998). Asymptotic statistics, Cambridge series in statistical and probabilistic mathematics. Cambridge: Cambridge University Press.