Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators

Nicolai Bissantz[1] and Melanie Birke[1]

[1]Fakultät für Mathematik
Ruhr-Universität Bochum, Germany

July 21, 2008

Abstract

We consider inverse regression models with convolution-type operators which mediate convolution on $\mathbb{R}^d$ ($d \geq 1$) and prove a pointwise central limit theorem for spectral regularisation estimators which can be applied to construct pointwise confidence regions. Here, we cope with the unknown bias of such estimators by undersmoothing. Moreover, we prove consistency of the residual bootstrap in this setting and demonstrate the feasibility of the bootstrap confidence bands at moderate sample sizes in a simulation study.

Keywords: Bootstrap, Inverse problems, Model selection, Testing.

1 Introduction

Suppose that we have observations $(\mathbf{z_k}, \mathbf{Y_k})$, $\mathbf{k} = (k_1, \dots, k_d) = \{-n, \dots, n\}^d$, from the model

$$\mathbf{Y_k} = g(\mathbf{z_k}) + \epsilon_{\mathbf{k}}, \tag{1}$$

where $g = \Psi * \theta$ is a one-to-one convolution operator with a function Ψ, the $\mathbf{z_k} = \left(\frac{k_1}{na_n}, \dots, \frac{k_d}{na_n}\right)$ are fixed design points, the $\epsilon_{\mathbf{k}}$'s are i.i.d. errors with $E\epsilon_{\mathbf{k}} = 0$, $E\epsilon_{\mathbf{k}}^2 = \sigma^2$ ($\mathbf{k} = (k_1, \dots, k_d)$), and a_n is a sequence which converges asymptotically to zero. The observable signal g can be represented as the image of the signal θ under the operator

$$(K\theta)(z) = \int_{\mathbb{R}^d} \Psi(z - \mathbf{t})\theta(\mathbf{t})d\mathbf{t}.$$

Recovery of the signal θ from the data $(\mathbf{z_k}, Y_k)$ in model (1) is a statistical inverse problem (e.g. Mair and Ruymgaart, 1996, and Bissantz et al., 2007b) which is closely related to density deconvolution (e.g. Stefanski and Carroll, 1990; Fan, 1991; Delaigle and Gijbels, 2002). It is usually assumed in nonparametric deconvolution regression models (e.g. Cavalier and Tsybakov, 2002) that the function θ is periodic (say on $[0, 1]$), and that A is thus a convolution operator on $[0, 1]$ with periodic Ψ which is, however, often unrealistic in practice. Examples are the deconvolution of astronomical and biological images from telescopic and microscopic imaging devices which involves deconvolution, but where the signal is usually not periodic. In this paper we will discuss the estimation of the signal θ from model (1), which

[1]Address for correspondence: Dr. Nicolai Bissantz, Ruhr-Universität Bochum, Fakultät für Mathematik, NA 3/70, Universitätsstr. 150, D-44780 Bochum, Germany, email: nicolai.bissantz@rub.de, Fon: +49/234/32–23291, Fax: +49/551/32–14559

appears to be more appropriate in this context. An important difficulty in this situation is that the reconstruction of θ from $g = K\theta$ at any location x on the real line requires (at least asymptotically) information on g on the full real line. We therefore use a design which includes an additional sequence $a_n \to 0$ to ensure that the design points $\mathbf{z_k}$ will asymptotically exhaust $\mathbb{R}^d$.

In this paper we discuss pointwise convergence properties of Fourier-based estimators in model (1). The estimator and some useful assumptions are introduced in Section 2. Asymptotic normality and confidence intervals are discussed in Section 3 and a bootstrap version of the confidence intervals in Section 4. Whereas it is known that asymptotic confidence intervals do not perform well for moderate sample sizes (e.g. Hall, 1993, in the direct density estimation context, and Bissantz et al., 2007a, for uniform confidence bands in density deconvolution), we demonstrate a satisfactory performance of the bootstrap confidence intervals in a simulation study in Section 5. Finally, in order to keep the paper more readable, all proofs are deferred to an appendix.

2 Prerequisites. Estimator, notation and assumptions

Notation. In the following, we consider the $\mathbf{j}^{th}$ derivative of a function or estimator $\hat{\theta}_n(\mathbf{x})$, which depends on a d-variate covariable $\mathbf{x}$. By the $\mathbf{j}^{th}$ derivative $\mathbf{j} = (j_1, \ldots, j_d)$ we denote the partial derivative $\partial^j / \partial x_1^{j_1} \cdots \partial x_d^{j_d}$, where $j = j_1 + \cdots + j_d$, and we suppose $j_1, \ldots, j_d$ to be such that $j \le p$, where θ has partial derivatives of order p which are all continuous. Moreover, $\omega^{\mathbf{j}}$, where $\omega \in \mathbb{R}^d$, means $\omega_1^{j_1} \cdots \omega_d^{j_d}$.

The estimator. We consider the Fourier estimator

$$\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) = \hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} e^{-i\langle \omega, \mathbf{x} \rangle} \Phi_k(h\omega) \frac{\widehat{\Phi}_g(\omega)}{\Phi_\Psi(\omega)} d\omega, \quad 0 \le j \le p, \tag{2}$$

Here $h > 0$ is a smoothing parameter called the bandwidth, and $\widehat{\Phi}_g$ is the empirical Fourier transform of g defined by

$$\widehat{\Phi}_g(\omega) = \frac{1}{N a_n^d} \sum_{\mathbf{r} \in \{-n, \ldots, n\}^d} Y_{\mathbf{r}} e^{i\langle \omega, z_{\mathbf{r}} \rangle},$$

where $N = n^d$.

Hence, the estimator $\hat{\theta}_n^{(\mathbf{j})}$ can be written in kernel form as follows:

$$\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) = \frac{1}{N h^{j+d} a_n^d} \sum_{\mathbf{r} \in \{-n, \ldots, n\}^d} Y_{\mathbf{r}} K_n^{(j)} \left(\frac{\mathbf{x} - \mathbf{z_r}}{h} \right),$$

where the kernel

$$K_n^{(\mathbf{j})}(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} e^{-i\langle \omega, \mathbf{x} \rangle} \frac{\Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega, \quad 0 \le \mathbf{j} \le p,$$

depends on n through h and Φ_k, which provides regularisation in the Fourier space, is the Fourier transform of some kernel function k.

Hence, the estimator $\hat{f}_n(\mathbf{x})$ may be written as

$$\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) = \sum_{\mathbf{r} \in \{-n, \ldots, n\}^d} Y_{\mathbf{r}} \frac{1}{N h^{j+d} a_n^d} K_n^{(\mathbf{j})} \left(\frac{\mathbf{x} - \mathbf{z_r}}{h} \right) = \sum_{\mathbf{r} \in \{-n, \ldots, n\}^d} Y_{\mathbf{r}} w_{\mathbf{j},\mathbf{r},n}(\mathbf{x}),$$

with weights

$$w_{\mathbf{j},\mathbf{r},n}(\mathbf{x}) = \frac{1}{Nh^{j+d}a_n^d} K_n^{(\mathbf{j})}\left(\frac{\mathbf{x}-\mathbf{z_r}}{h}\right). \tag{3}$$

Further assumptions. We will make the following common assumptions on Φ_k and Ψ. Our first assumption is that Ψ is ordinary smooth, i.e. we consider mildly ill-posed problems in model (1).

Assumption 1. The Fourier transform Φ_ψ satisfies

$$\Phi_\psi(\omega)||\omega||^\beta \to C_\epsilon, \qquad \omega \to \infty$$

for some $\beta > 0$ and $C_\epsilon \in \mathbb{C} \setminus \{0\}$.

We also make some regularity assumptions on the Fourier transform Φ_k, which in effect causes the regularization of the estimator.

Assumption 2. The Fourier transform Φ_k of k is symmetric and supported on $[-1,1]^d$ with $\Phi_k(\omega) = 1$ for $\omega \in [-b,b]^d$, $b > 0$, and $|\Phi_k(\omega)| \leq 1$ for all $\omega \in [-1,1]^d$.

Finally, we make the following assumptions on the signal θ and its image $g = K\theta$, where the first assumption is about the smoothness of θ. The second assumption is about the tail behaviour of g and will be required in the computation of the bias to determine the impact of the parts of the signal which are not observed at given sample size $2n+1$ due to the fact that the support of the design is then limited to $\left[-\frac{1}{a_n}, \frac{1}{a_n}\right]$.

Assumption 3. A. The Fourier transform Φ_θ of θ satisfies

$$\int_{\mathbb{R}} |\Phi_\theta(\omega)| \; ||\omega||^{s-1} \, d\omega < \infty \qquad \text{for some} \qquad s > p+1.$$

B. The function $g = K\theta$ satisfies

$$\int_{\mathbb{R}} |g(z)| \; ||z||^r \, dz < \infty$$

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

3 Asymptotic normality and asymptotic confidence intervals

3.1 Asymptotic normality

Our purpose in this section is to derive the pointwise asymptotic distribution of the estimator defined above. The result is stated in Theorem 1 and a consequence of the Lindeberg central limit theorem which is also used in Bissantz and Holzmann (2008) to derive the asymptotic distribution of deconvolution estimators of periodic functions on $[0,1]$ in the inverse regression context. If we know the pointwise asymptotic distribution of the estimator we can construct pointwise confidence regions for the unknown function θ which is described in Section 3.2. The following result follows from a general central limit theorem for weighted sums of independent random variables (Eubank, 1999) under the condition

$$\frac{\max_{\mathbf{k}\in\{-n,\ldots,n\}^d} |w_{\mathbf{j},\mathbf{k},n}(\mathbf{x})|}{\left(\sum_{\mathbf{r}\in\{-n,\ldots,n\}^d} w_{\mathbf{j},\mathbf{r},n}^2(\mathbf{x})\right)^{1/2}} \to 0. \tag{4}$$

Theorem 1. *Suppose in model (1) that assumption 1 holds and $h \to 0$ and $a_n \to 0$ as $n \to \infty$ such that $Nh^d a_n^d \to \infty$. Then*

$$\left(\sigma^2 \sum_{\mathbf{r}\in\{-n,\ldots,n\}^d} w_{\mathbf{j},\mathbf{r},n}^2(\mathbf{x})\right)^{-1/2} \left(\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) - E\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

3.2 Asymptotic confidence intervals

With the asymptotic distribution derived in Section 3.1 we are now able to construct pointwise asymptotic confidence intervals for the function θ. In the following subsections we will propose two different methods for this. The first one is the typical way to construct asymptotic confidence intervals where we use the quantiles of the asymptotic distribution of the estimator. Like in all nonparametric regression problems we will see that this method has some drawbacks. As we see in Theorem 1 the asymptotic distribution still depends on the unknown function θ and the standard deviation σ through bias and variance. Furthermore, in general we cannot obtain a closed form of the bias such that we cannot use a plug in estimator for θ in this expression but have to apply undersmoothing. Therefore, we state the results on asymptotic confidence intervals here but also propose a bootstrap approach in the following section.
For the sake of simplicity let in the sequel

$$\begin{aligned} b_{\theta,n}(x) &:= \mathrm{E}[\hat{\theta}_n(x)] - \theta(x) \\ s^2 &:= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \omega^{2\beta}\omega^{2j}|\Phi_k(\omega)|^2 d\omega \end{aligned}$$

denote the bias and part of the asymptotic variance of the estimator $\hat{\theta}_n(x)$. Based on Theorem 1 we can state the following result.

Corollary 2. *Let the assumptions of Theorem 1 be fulfilled. Then, with*

$$\tilde{c}_n(x) := \left(\hat{\theta}_n(x) - b_{\theta,n}(x) - \frac{\sigma^2 s^2 u_{1-\frac{\alpha}{2}}}{\sqrt{Nh^{2\beta+2j+d}a_n^d}}, \hat{\theta}_n(x) - b_{\theta,n}(x) + \frac{\sigma^2 s^2 u_{1-\frac{\alpha}{2}}}{\sqrt{Nh^{2\beta+2j+d}a_n^d}}\right),$$

$$P(\theta(x) \in \tilde{c}_n(x)) \to 1-\alpha$$

for $n \to \infty$,

where $u_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution. The interval $\tilde{c}_n(x)$ in Corollary 2 cannot be used as confidence interval for θ since it still depends on the unknown parameters $b_{\theta,n}(x)$ and σ^2. To construct a confidence interval from $\tilde{c}_n(x)$ we propose to use a variance estimator $\hat{\sigma}^2$ of σ^2 with convergence rate $o\left(\sqrt{Nh^{2\beta+2j+d}a_n^d}\right)$, i.e. faster than the estimator $\hat{\theta}_n(x)$. A possible choice are e.g. difference estimators (Munk, Bissantz, Wagner and Freitag, 2005). A common way to achieve (asymptotic)) negligibility of the bias is to undersmooth the estimator (e.g. Hall, 1993, Bissantz et al., 2007a). Below, in Corollary 4, we will state a result for confidence intervals with undersmoothing bandwidth. To this end we first determine the order of the bias in Lemma 3 under assumptions 1-3.

Lemma 3. *Under the assumptions 1-3, if $Na_n^d h^{\beta+s+d-1} \to \infty$ we have*

$$b_{\theta,n}(x) = o(h^{s-j-1}).$$

We are now in a position to define (asymptotically valid) confidence intervals based on undersmoothing of the estimator. Here, undersmoothing follows immediately from the $o()$-rate of the bias, if $h^{s-j-1} \asymp (Nh^{2\beta+2j+d})^{-1/2}$ due to assumptions 2 and 3.A.

Corollary 4. *Let the assumptions of Theorem 1, assumptions 1-3, $Na_n^d h^{\beta+s+d-1} \to \infty$ and $Nh^{2\beta+2s+d-2}a_n^d = O(1)$ be fulfilled. Then with*

$$c_n(x) \quad := \quad \Big(\hat{\theta}_n(x) - \frac{\hat{\sigma}^2 s^2 u_{1-\frac{\alpha}{2}}}{\sqrt{Nh^{2\beta+2j+d}a_n^d}}, \hat{\theta}_n(x) + \frac{\hat{\sigma}^2 s^2 u_{1-\frac{\alpha}{2}}}{\sqrt{Nh^{2\beta+2j+d}a_n^d}}\Big),$$

$$P(\theta(x) \in c_n(x)) \quad \to \quad 1-\alpha$$

for $n \to \infty$.

Corollary 4 follows from Theorem 1 because $\hat{\sigma}^2$ converges faster than the $\sqrt{Nh^{2\beta+2j+d}a_n^d}$ rate of the estimator $\hat{\theta}^{(j)}(\mathbf{x})$. Now, $c_n(x)$ is a pointwise conficence interval for $\theta(x)$ but usually in nonparametric regression, asymptotic confidence intervals are conservative. Although we can cope with the bias by undersmoothing this may have negative effects on the estimator $\hat{\theta}_n(x)$. Especially the spectral regularisation method used here is very sensitive to undersmoothing (cf. Bissantz et al., 2007a). Therefore, in the next subsection, a bootstrap method is proposed.

4 Bootstrap confidence intervals

Another method to deal with the unknown bias is to use bootstrap to construct confidence intervals. Bootstrap is often used in nonparametric regression to cope with the unknown bias or variance of an asymptotic distribution. Available approaches for fixed design include wild bootstrap (see e.g. Wu, 1986 or Mammen, 1993) and residual bootstrap (see e.g. Efron, 1979 or Härdle and Bowman, 1988). We restrict ourselves to the latter one and prove its consistency in the setting under consideration in this paper. In the first step of the residual bootstrap the distribution of the residuals is estimated from the residuals of (some) estimator $\hat{g}$ of the regression function $g = K\theta$. Here, we have reduced the inverse regression problem to a direct one by writing model (1) as $Y_{\mathbf{k}} = g(\mathbf{z_k}) + \varepsilon_{\mathbf{k}}$. A straightforward choice for $\hat{g}$ is to choose $\hat{g}_{dc,\tilde{h}} = K\theta_{\tilde{h}}$, where $\hat{\theta}_{\tilde{h}}$ is a Fourier estimator determined from the original sample. However, in practice different choices are possible, e.g. to estimate g with a local polynomial estimator $\hat{g}_{lp,\tilde{h}}$ of order $p \geq 0$.
To generate the bootstrap data let ε_k^*, $k = 1, \dots, n$ be drawn with replacement from the distribution of the centered residuals $\hat{\varepsilon}_{\mathbf{k}} - \bar{\hat{\varepsilon}}$, $\mathbf{k} \in \{-n, \dots, n\}^d$ with $\hat{\varepsilon}_k = Y_k - \hat{g}_{i,\tilde{h}}(z_k)$ and $i = dc, lp$, respectively. The data is then generated by

$$Y_{\mathbf{k}}^* = \hat{g}_{i,\tilde{h}}(\mathbf{z_k}) + \varepsilon_{\mathbf{k}}^*, \ i = dc \text{ or } lp$$

where $\tilde{h}$ is a second bandwidth, typically larger than h, and the bootstrap estimator of $\theta^{(j)}$ is defined as

$$\hat{\theta}_n^{(\mathbf{j})\ *}(\mathbf{x}) = \frac{1}{Nh^{j+d}a_n^d} \sum_{\mathbf{r} \in \{-\mathbf{n},\dots,\mathbf{n}\}^{\mathbf{d}}} Y_{\mathbf{r}}^* K_n^{(\mathbf{j})}\Big(\frac{\mathbf{x}-\mathbf{z_r}}{h}\Big).$$

In the following, to simplify notations, we restrict ourselves to the cases $\hat{g}_h = \hat{g}_{dc,h}$ and $d = 1$. To show feasibility of the residual bootstrap in our setting we now prove that the bootstrap estimator $\hat{\theta}_n^{(j)\,*}$ has asymptotically the same distribution as the estimator $\hat{\theta}_n^{(j)}$. To this end we need the following assumptions on the estimator $\hat{g}$.

Assumption 4. A. The Fourier transform $\Phi_{\hat{\theta}_{\tilde{h}}}$ of $\hat{\theta}_{\tilde{h}}$ exists and fulfills

$$\int_{\mathbb{R}} |\Phi_\theta(\omega) - \Phi_{\hat{\theta}_{\tilde{h}}}(\omega)||\omega|^{s-1}\, d\omega = o_P(1).$$

B. The estimator $\hat{g}_{\tilde{h}}$ satisfies

$$\int_{\mathbb{R}} |\hat{g}_{\tilde{h}}(z)||z|^r\, dz < \infty \qquad \text{for some} \qquad r > 0.$$

Assumption 4.A is concerned with the smoothness of the difference function $\theta - \hat{\theta}_{\tilde{h}}$, analogously to assumption 3.A on the smoothness of the function of interest θ, and the second part 4.B of the assumption with the smoothness of the estimated Fourier transform of g.
With this assumption our bootstrap is consistent:

Theorem 5. *With assumptions 1-3 and 4 the bootstrap described above is consistent, that is*

$$\sup_t \left| \mathrm{P}_* \left(\Big(\sigma^2 \sum_{r=-n}^{n} w_{j,r,n}^2(x) \Big)^{-1/2} \Big(\hat{\theta}_n^{(j)\,*}(x) - \mathrm{E}_*[\hat{\theta}_n^{(j)\,*}(x)] \Big) \le t \right) - \Phi(t) \right| = o_P(1)$$

for $t \to \infty$*, where* P_* *and* E_* *denote the probability and expectation conditionally on* $\mathcal{Y} = \{Y_r | r \in \{-n, \dots, n\}\}$.

Hence, we can bootstrap the (pointwise) distribution of the estimator (2) and define bootstrap confidence intervals from the following procedure. We generate B bootstrap samples $(x_k, Y_k^{*,l})$, $k \in \{-n, \dots, n\}^d$, $l = 1, \dots, B$ and bootstrap estimates $\hat{\theta}_n^{(j)\,*,l}(x)$, $l = 1, \dots, B$ and use the $\lfloor B\alpha \rfloor$-th order statistic $\vartheta_{n,\alpha}^*(x) = \hat{\theta}_n^{(\mathbf{j})\,*,(\lfloor \mathbf{B}\alpha \rfloor)}(\mathbf{x})$ as an estimate for the α-quantile of the distribution. This results in the bootstrap confidence interval

$$c_n^*(x) = \left(2\hat{\theta}(x) - \vartheta_{n,1-\alpha/2}^*(x)\, ,\, 2\hat{\theta}(x) - \vartheta_{n,\alpha/2}^*(x)\right)$$

of level $(1 - \alpha)$.

5 Simulations

In this section we demonstrate the performance of the bootstrap confidence intervals with the results of a small simulation study. First, in Section 5.1 we introduce the simulation framework and discuss the problem of bandwidth choice. Then, in Section 5.2, we present the results of our simulation study of the bootstrap confidence intervals for the Fourier estimator on $\mathbb{R}^1$.

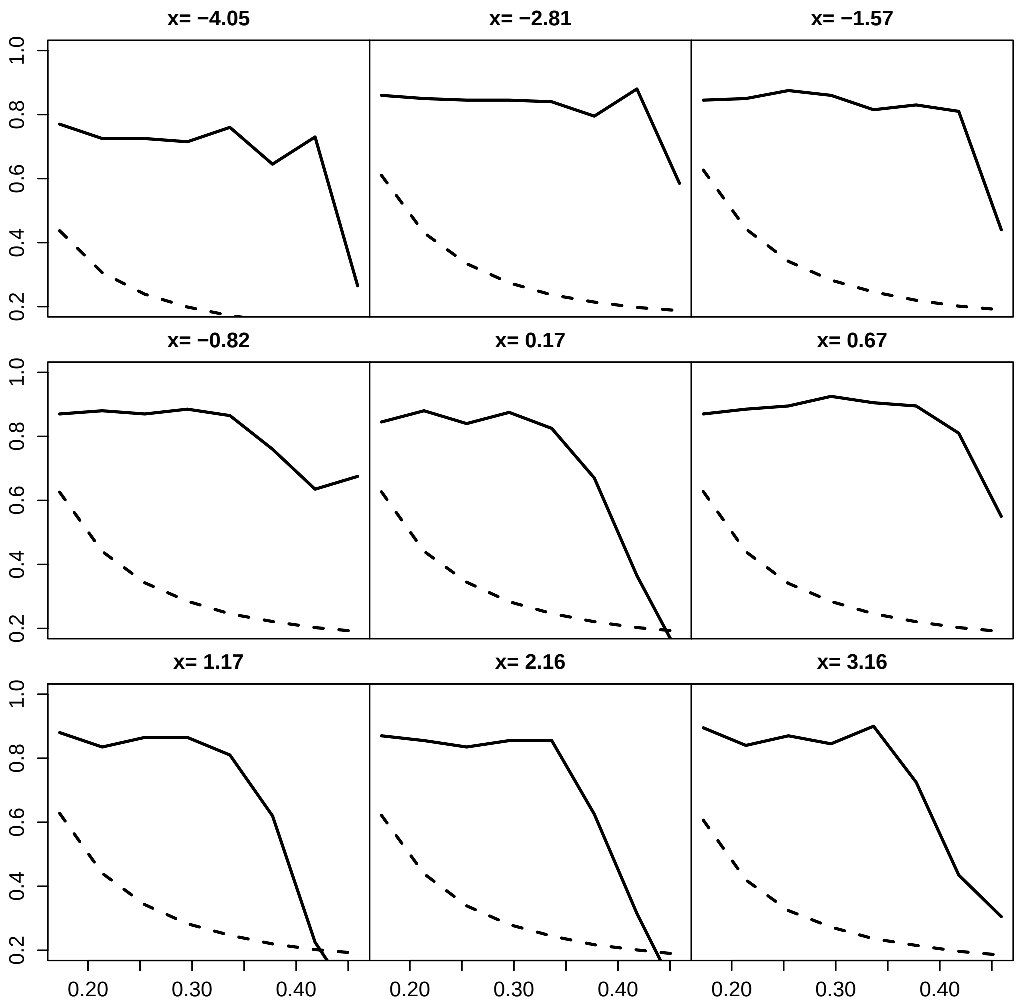

Figure 1: Simulated effective coverage probabilities and interval length for (residual) bootstrap confidence intervals with 90% nominal coverage probability at a number of locations along the x-axis, and in dependence of the bandwidth h. The true function is the Gaussian $\theta_1(x)$, and the simulation parameters are $2n+1=201$, $\sigma=0.1$ and $a_n=0.25$. Solid lines indicate the effective coverage probability, and dashed lines the effective confidence interval length (multiplied by a factor of 3).

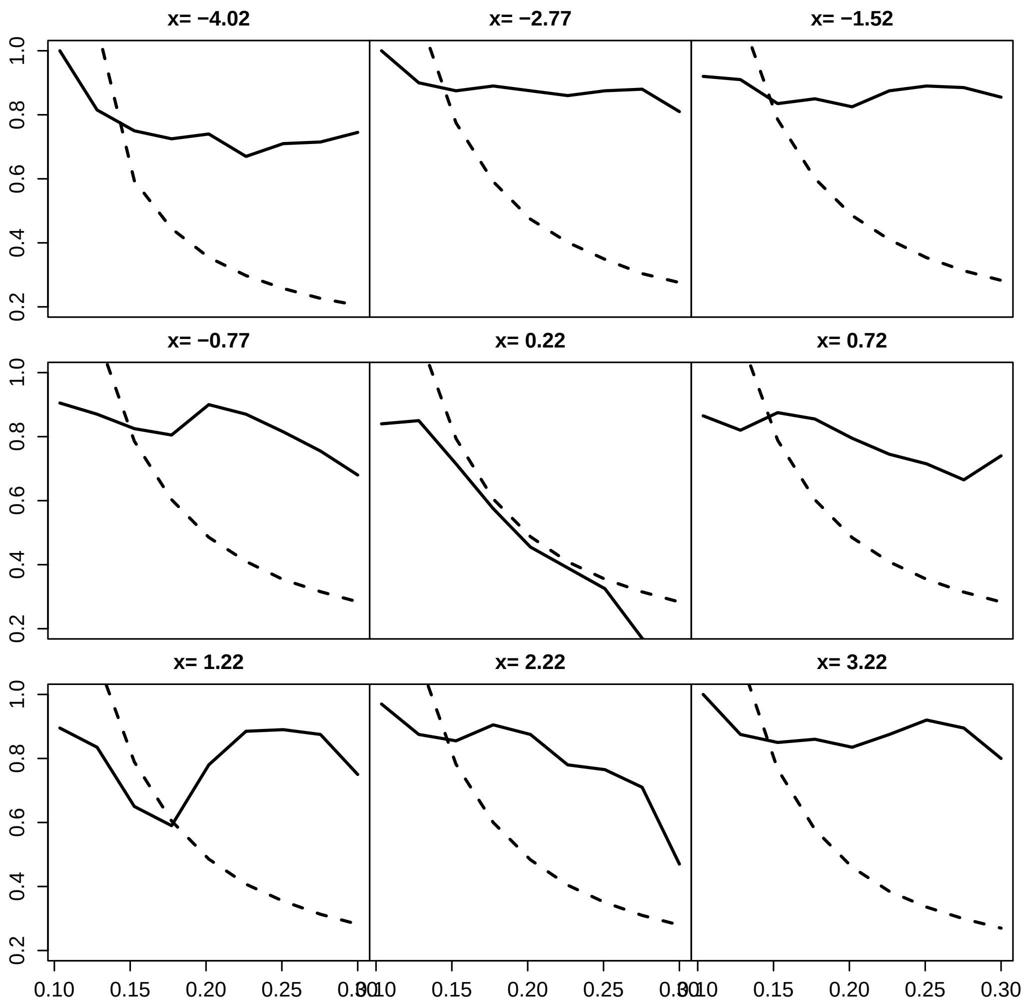

Figure 2: Simulated effective coverage probabilities and interval length for (residual) bootstrap confidence intervals with 90% nominal coverage probability at a number of locations along the x-axis, and in dependence of the bandwidth h. The true function is the bimodal function $\theta_2(x)$, and the simulation parameters are $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$. Solid lines indicate the effective coverage probability, and dashed lines the effective confidence interval length (multiplied by a factor of 3).

5.1 Simulation framework and bandwidth selection

Our simulations are based on the assumption that the observations follow model (1), i.e.

$$Y_k = (K\theta)(z_k) + \varepsilon_k, \quad k = -n, \dots, n,$$

where A is the convolution operator on $\mathbb{R}$ which causes the convolution of the function of interest θ with the convolution function ψ, and the noise is (centered) normally distributed with variance σ^2. The design points are $z_k = \frac{k}{na_n}$. The convolution function ψ is given by the Laplace density function

$$\psi(x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

where $\lambda = 3$ and the functions of interest are

$$\begin{aligned} \theta_1(x) &= e^{-\frac{(x-1.1)^2}{2\cdot 0.64}} \quad \text{and} \\ \theta_2(x) &= e^{-\frac{(x-0.2)^2}{2\cdot 0.09}} + 1.2 \cdot e^{-\frac{(x-0.85)^2}{2\cdot 0.04}}. \end{aligned}$$

We have performed simulations for a number of combinations of the parameters sample size $2n+1$ and noise variance σ^2. The design parameter $a_n = 0.25$ was selected such that the interval $\left[-\frac{1}{a_n}, \frac{1}{a_n}\right]$ includes most of the $x-$axis where the functions θ_1 and θ_2 deviate significantly from 0. In all cases we generated randomly 200 data sets according to model (1) and then performed 400 replications of a residual bootstrap to determine confidence intervals for $\theta(x)$. To this end we constructed the sampling distribution for the residuals from all observations $|z_k| \le \frac{1}{a_n} - 2.01h$. Confidence intervals were determined at 39 equidistant points along the x-axis.

Figures 1 and 2 show the effective (simulated) coverage probability and size of the confidence intervals at some of the considered locations along the x-axis in dependence of the bandwidth h. Similar as for uniform confidence bands (e.g. Bissantz et al. 2007a, and Birke et al., 2008) the coverage probability decreases with increasing bandwidth h, due to the increasing bias of the estimator which has to be corrected by the bootstrap. On the other hand, for decreasing bandwidth, the variability of the estimates increases significantly, which results in a large area of the confidence intervals. Note that the bias problem is less obvious at positions close to the boundary of the coverage region of the design points $[-1/a_n, 1/a_n]$, which is due to the small curvature and the proximity of the function values to 0 of the true functions θ_1 and θ_2 there. However, in general the bandwidth choice is fundamental to a proper determination of the confidence intervals. We use the L_∞-motivated bandwidth selection algorithm introduced in Bissantz et al. (2007a) and applied in Birke et al. (2008) for the selection of bandwidth. However, whereas these authors aimed at an undersmoothing bandwidth there and, in consequence, chose the bandwidth somewhat smaller than "approximately L_∞-optimal", we also bootstrap the bias here, and do not need to achieve undersmoothing. The L_∞-motivated bandwidth estimator consists in evaluating the estimator $\hat{f}_n(x;h)$ at equidistantly spaced bandwidth, and choosing among these the largest bandwidth, where the supremum of the differences between the estimators for two adjacent bandwidth steps exceeds a certain threshold. Since we do not necessarily need undersmoothing here, we choose a smaller tuning parameter τ for the bandwidth selection than in the uniform confidence bands case. Values $\tau \approx 1.5$ in combination with 12 bandwidth steps involved, covering an order of magnitude in bandwidth value, turned out to be a good choice in simulations. Here we proceed by first selecting a

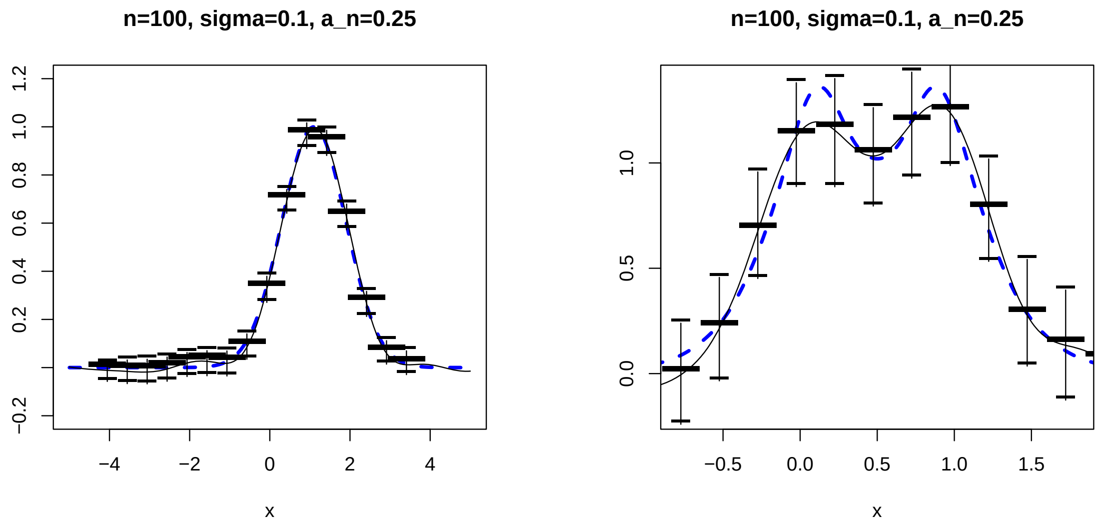

Figure 3: Typical estimates $\hat{\theta}_n(x)$ (solid line) and associated 95% nominal coverage probability residual bootstrap confidence intervals for the Gaussian function (left) and bimodal function (right) at a number of positions along the $x-$axis. Dashed lines indicate the true functions θ_1 and θ_2, respectively.

bandwidth for each combination of parameters θ_i, $i = 1, 2$, n, σ and a_n from analyzing a small number of random datasets with the L_∞-optimal bandwidth estimator. Then we keep this bandwidth fixed for all subsequent simulations with the same set of parameters.

Figure 3 shows typical estimates of θ_1 and θ_2 for sample size $2n+1 = 201$, $\sigma = 0.1$ and $a_n = 0.1$ together with some of the associated pointwise 95%-confidence intervals. To give the plots a clearer lay-out we only show confidence intervals at every second x-position considered in the simulations. Here, the confidence intervals were computed as

$$c_n^*(x) = \left(2\hat{\theta}_n(x) - \vartheta^*_{n,0.975}(x)\,,\, 2\hat{\theta}_n(x) - \vartheta^*_{n,0.025}(x)\right),$$

where $\vartheta^*_{n,0.025}(x)$ and $\vartheta^*_{n,0.975}(x)$ are the 0.025 and 0.975-quantile of the bootstrapped distribution $\hat{\theta}^*_n(x)$ at the respective position.

5.2 Simulation results

We now discuss the results of the simulations. Table 1 provides mean and standard deviation of the simulated coverage probabilities and confidence interval lengths at 16 locations covering the interval $[-1, 3]$ on the $x-$axis. This interval has been selected such that it represents well the region where most of the signal of both functions of interest θ_1 and θ_2 is.
From the Tables, we make the following conclusions. The confidence intervals perform well for both functions of interest with respect to the coverage probabilities, being only slightly smaller than their nominal values. In particular the fact that this is already true for sample

θ	n	σ	a_n	80% nominal cov. Cov.prob. [%]	Length ($\times$100)	90% nominal cov. Cov.prob. [%]	Length ($\times$100)	95% nominal cov. Cov.prob. [%]	Length ($\times$100)
θ_1	100	0.1	0.25	78 ± 3	7.3 ± 0.1	88 ± 3	9.4 ± 0.1	93 ± 2	11.2 ± 0.1
	100	0.5	0.25	74 ± 3	26.5 ± 0.3	86 ± 2	34.0 ± 0.4	92 ± 2	40.5 ± 0.4
	1000	0.1	0.25	77 ± 3	2.8 ± 0.03	88 ± 2	3.7 ± 0.04	93 ± 2	4.4 ± 0.04
	1000	0.5	0.25	76 ± 2	10.3 ± 0.1	87 ± 2	13.2 ± 0.1	93 ± 1	15.8 ± 0.1
θ_2	100	0.1	0.25	75 ± 3	28.9 ± 0.1	86 ± 2	37.0 ± 0.1	92 ± 2	44.1 ± 0.1
	1000	0.1	0.25	73 ± 7	11.5 ± 0.1	85 ± 7	14.8 ± 0.2	91 ± 5	17.6 ± 0.2

Table 1: Mean and standard deviation of the simulated coverage probabilities and confidence interval length for the Gaussian function θ_1 and the bimodal function θ_2 determined from the properties of confidence intervals computed at 16 equidistant positions approximately covering the interval $[-1, 3]$.

Setting	80% nominal cov. Cov.prob. [%]	Length ($\times$100)	90% nominal cov. Cov. prob. [%]	Length ($\times$100)	95% nominal cov. Cov.prob. [%]	Length ($\times$100)
Standard dev. of ψ 5% underestimated	71 ± 4	7.9 ± 0.1	83 ± 3	10.2 ± 0.1	89 ± 3	12.2 ± 0.2
Standard dev. of ψ 5% overestimated	75 ± 5	6.9 ± 0.1	85 ± 4	8.9 ± 0.1	91 ± 3	10.6 ± 0.1
Laplace, miss-spec. as Gaussian	73 ± 3	8.8 ± 0.1	83 ± 4	11.3 ± 0.2	89 ± 3	13.5 ± 0.2
Gaussian, miss-spec. as Laplace	72 ± 6	7.4 ± 0.1	83 ± 5	9.5 ± 0.1	89 ± 4	11.3 ± 0.1

Table 2: Mean and standard deviation of the simulated coverage probabilities and confidence interval length for the Gaussian function determined from the properties of confidence intervals computed at 16 equidistant positions approximately covering the interval $[-1, 3]$ for some scenarios of miss-specification of the convolution function ψ.

size $2n + 1 = 201$ for the bimodal function indicates a good performance of the bootstrap procedure w.r.t. bias correction. However, for the bimodal function this comes at the price of significantly larger confidence intervals than for the more simple unimodal function (note the different noise levels used for the functions of interest).
Finally, in Table 2 we show the results of some simulations where the convolution function ψ is miss-specified. This is often the case in practical applications if the distribution of errors of the measurements of the covariate x is imperfectly known. The first two rows give the results from simulations where the parameter λ in the convolution function ψ is over- respectively underestimated by 5%. The simulations discussed in the third and fourth row show results of simulations where the shape of the convolution function ψ is miss-specified. Here, in the first case we assume that the convolution function ψ is in fact given by the density of a Laplace distribution but specified as the density of a normal distribution with same variance 2/9 in the estimator. In the second case (shown in the fourth row) the Laplace and normal density are exchanged. The results of the simulations in Table 2 indicate that our confidence intervals are not very sensitive on such kinds of modest miss-specification of the convolution function ψ, since a comparison with the corresponding results in the first line of Table 1 reveals only a reduction of the effective (simulated) coverage probabilities of about $2 - 6\%$ points, and increase in confidence interval length of $\approx 10\%$.

Acknowledgements

Financial support of DFG grant SFB 475 and the BMBF under project "INVERS" 03BIPAH4 is gratefully acknowledged.

6 Appendix: Proofs

Proof of Theorem 1.
To prove asymptotic normality of the estimator (2) we show that condition (4) holds for the weights defined in (3) (cf. Eubank, 1999). First, we have

$$\max_{\mathbf{k}\in\{-n,\dots,n\}^d} \left| \frac{1}{Nh^{j+d}a_n^d} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} \frac{e^{-i\langle \omega, (\mathbf{x}-z_{\mathbf{k}})\rangle/h} \Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega \right| = O\left(\frac{1}{Na_n^d h^{\beta+j+d}} \right),$$

using assumption 1.

For the denominator, we have the following result from Parseval's equality and Assumption 1

$$\sum_{\mathbf{r}\in\{-n,\ldots,n\}^d} w_{\mathbf{j},\mathbf{r},n}^2(\mathbf{x}) = \sum_{\mathbf{r}\in\{-n,\ldots,n\}^d} \frac{1}{N^2h^{2j+2d}a_n^{2d}} \left(\int_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} \frac{1}{(2\pi)^d} \frac{e^{-i\langle \omega, (\mathbf{x}-z_{\mathbf{r}})\rangle/h} \Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega \right)^2$$

$$= \frac{1}{Nh^{2j+d}a_n^d} \left(\int_{[-1/(ha_n),1/(ha_n)]^d} \left(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} (-i\omega)^{\mathbf{j}} \frac{e^{-i\langle \omega, (\mathbf{x}/h-\mathbf{s})\rangle} \Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega \right)^2 d\mathbf{s} + O\left(\frac{1}{(na_n)^d} \right) \right)$$

$$\asymp \frac{1}{Nh^{2j+d}a_n^d} \int_{\mathbb{R}^d} \Bigg(\underbrace{\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} (-i\omega)^{\mathbf{j}} \frac{e^{-i\langle \omega, (\mathbf{x}/h-\mathbf{s})\rangle} \Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega}_{=:G(-s)} \Bigg)^2 d\mathbf{s}\ [1+o(1)]$$

$$= \frac{1}{Nh^{2j+d}a_n^d} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} |\Phi_G(\omega)|^2 d\omega\ [1+o(1)]$$

$$= \frac{1}{Nh^{2j+d}a_n^d} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} \frac{||\omega||^{2\mathbf{j}} |\Phi_k(\omega)|^2}{|\Phi_\Psi(\omega/h)|^2} d\omega\ [1+o(1)]$$

$$\sim \frac{1}{Nh^{2\beta+2j+d}a_n^d} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} ||\omega||^{2\beta+2j} |\Phi_k(\omega)|^2 d\omega.$$

Hence, we have

$$\frac{\max_{\mathbf{k}\in\{-n,\ldots,n\}^d} |w_{\mathbf{j},\mathbf{k},n}(\mathbf{x})|}{\left(\sum_{\mathbf{r}\in\{-n,\ldots,n\}^d} w_{\mathbf{j},\mathbf{r},n}^2(\mathbf{x}) \right)^{1/2}} = O\left(\frac{N^{-1}a_n^{-d}h^{-\beta-j-d}}{N^{-1/2}h^{-\beta-j-d/2}a_n^{-d/2}} \right) = O\left(\frac{1}{\sqrt{Nh^d a_n^d}} \right).$$

Therefore, condition (4) is fulfilled and with the central limit theorem in Eubank (1999),

$$N^{1/2}h^{\beta+j+d/2}a_n^{d/2} \left(\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) - E\hat{\theta}_n^{(\mathbf{j})}(\mathbf{x}) \right)$$

is asymptotically normal with asymptotic variance

$$\frac{\sigma^2}{(2\pi)^d} \int_{\mathbb{R}^d} \omega^{2\beta}\omega^{2j} |\Phi_k(\omega)|^2 d\omega.$$

□

Proof of Lemma 3.

First note that

$$
\begin{aligned}
E\left[\hat{\theta}_n^{(j)}(\mathbf{x})\right] - \theta^{(j)}(\mathbf{x}) &= -\frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} e^{\frac{-i\langle \omega, \mathbf{x}\rangle}{h}} (1-\Phi_k(\omega))\Phi_\theta(\omega/h)\, d\omega \\
&\quad - \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} e^{\frac{-i\langle \omega, \mathbf{x}\rangle}{h}} \frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)} \cdot \Big(\int\limits_{([-1/a_n, 1/a_n]^d)^c} e^{\frac{i\langle \omega, \mathbf{y}\rangle}{h}} g(\mathbf{y}) d\mathbf{y} \Big) d\omega \\
&\quad + O\big((N a_n^d h^{j+d})^{-1}\big) \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} (-i\omega)^{\mathbf{j}} e^{\frac{-i\langle \omega, \mathbf{x}\rangle}{h}} \frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)} d\omega \\
&= A + B + C \cdot O\big((N a_n^d h^{j+d})^{-1}\big). \qquad (5)
\end{aligned}
$$

For A we get with assumption 2

$$
\begin{aligned}
|A| &\leq \frac{1}{(2\pi)^d h^{j+d}} \int_{\mathbb{R}^d} ||\omega||^j |1-\Phi_k(\omega)||\Phi_\theta(\omega/h)| d\omega \\
&\leq \frac{2}{(2\pi)^d h^{j+d}} \int_{([-b,b]^d)^c} ||\omega||^j |\Phi_\theta(\omega/h)| d\omega \\
&= \frac{2}{(2\pi)^d} \int_{([-b/h, b/h]^d)^c} \frac{1}{||\eta||^{s-j-1}} ||\eta||^{s-1} |\Phi_\theta(\eta)| d\eta \\
&\leq \frac{2h^{s-j-1}}{b^{s-j-1}(2\pi)^d} \int_{([-b/h, b/h]^d)^c} ||\eta||^{s-1} |\Phi_\theta(\eta)| d\eta = o(h^{s-j-1})
\end{aligned}
$$

since $b/h \to \infty$ for $h \to 0$ and assumption 3.A holds. For B we obtain with assumptions 3.B and 1

$$
\begin{aligned}
|B| &\leq \frac{1}{(2\pi)^d h^{j+d}} \int_{\mathbb{R}^d} ||\omega||^j \frac{|\Phi_k(\omega)|}{|\Phi_\psi(\omega/h)|} d\omega \Big(\int_{([-1/a_n, 1/a_n]^d)^c} \frac{1}{||\mathbf{y}||^r} |g(\mathbf{y})| \; ||y||^r d\mathbf{y} \Big) \\
&\leq O\Big(\frac{a_n^r}{h^{\beta+j+d}}\Big) \int_{\mathbb{R}^d} ||\omega||^{\beta+j} \frac{|\Phi_k(\omega)|}{||\omega/h||^\beta \; |\Phi_\psi(\omega/h)|} d\omega \\
&\leq O\Big(\frac{a_n^r}{h^{\beta+j+d}}\Big) \int_{[-1,1]^d} ||\omega||^{\beta+j} d\omega = O\Big(\frac{a_n^r}{h^{\beta+j+d}}\Big) = o(h^{s-j-1}).
\end{aligned}
$$

The third term can be estimated quite similarly by

$$
|C| = O\Big(\frac{1}{h^\beta}\Big)
$$

such that we get with (5)

$$
b_{\theta,n}(x) = o(h^{s-j-1}) + O\Big(\frac{1}{N a_n^d h^{\beta+j+d}}\Big) = o(h^{s-j-1}).
$$

□

Proof of Theorem 5.

We apply a conditional central limit theorem (see Corollary 3.1 of Hall and Heyde, 1980) to the estimator

$$
\hat{\theta}_n^{(j)\,*}(x) = \sum_{r=-n}^{n} Y_r^* w_{j,r,n}(x)
$$

with the same deterministic weights $w_{j,r,n}(x)$ as in (3). If the bootstrap residuals are generated as described above, they are, conditionally on $\mathcal{Y}$, independent identically distributed and satisfy

$$\mathrm{E}_*[\varepsilon_r^*] = \frac{1}{2n+1}\sum_{k=-n}^{n}(\hat{\varepsilon}_k - \bar{\hat{\varepsilon}}) = 0 \tag{6}$$

$$\mathrm{E}_*[(\varepsilon_r^*)^2] = \frac{1}{2n+1}\sum_{k=-n}^{n}(\hat{\varepsilon}_k - \bar{\hat{\varepsilon}})^2 \xrightarrow{P} \sigma^2 \tag{7}$$

From Theorem 1 we see that the variance of the estimator, after standardization, is

$$\mathrm{Var}\Big(\Big(\sum_{r=-n}^{n} w_{j,r,n}^2(x)\Big)^{-1/2}\hat{\theta}_n^{(j)}(x)\Big) = \sigma^2.$$

We have to show that the conditional variance of $\left(\sum_{r=-n}^{n} w_{j,r,n}^2(x)\right)^{-1/2}\hat{\theta}_n^{(j)\,*}(x)$ converges to σ^2. Conditionally on the sample $\mathcal{Y}$, the random variables $Y_{-n}^*,\ldots,Y_n^*$ are independent. The expectation of Y_r^*, conditionally on $\mathcal{Y}$, is $\hat{g}(z_r)$ and the conditional variance is

$$\begin{aligned}
&\mathrm{Var}_*\Big(\Big(\sum_{r=-n}^{n} w_{j,r,n}^2(x)\Big)^{-1/2}\hat{\theta}_n^{(j)\,*}(x)\Big)\\
&= \Big(\sum_{r=-n}^{n} w_{j,r,n}^2(x)\Big)^{-1}\sum_{r=-n}^{n} w_{j,r,n}^2(x)\,\mathrm{E}[(Y_r^* - \hat{g}(z_r))^2|Y_{-r+1}^*,\ldots,Y_{r-1}^*]\\
&= \Big(\sum_{r=-n}^{n} w_{j,r,n}^2(x)\Big)^{-1}\sum_{r=-n}^{n} w_{j,r,n}^2(x)\,\mathrm{E}_*[(\varepsilon_r^*)^2]\\
&= \left(\frac{1}{2n+1}\sum_{r=-n}^{n}(\hat{\varepsilon}_r - \bar{\hat{\varepsilon}})^2\right) \to \sigma^2
\end{aligned}$$

because of (7) and with

$$s_n^2(x) = \sigma^2\sum_{r=-n}^{n} w_{j,r,n}^2(x)$$

and

$$X_{n,r} = w_{j,r,n}(x)\frac{Y_r^* - \hat{g}(z_r)}{s_n(x)}$$

the first condition of Corollary 3.1 of Hall and Heyde (1980), that is

$$\sum_{r=-n}^{n}\mathrm{E}[X_{n,r}^2|Y_{-r+1}^*,\ldots,Y_{r-1}^*] \xrightarrow{P} 1$$

is fulfilled.
In the second step we verify the conditional Lindeberg condition

$$\sum_{r=-n}^{n}\mathrm{E}\left[X_{n,r}^2 I\{|X_{n,r}|>\delta\}\,\middle|\,Y_{-r+1}^*,\ldots,Y_{r-1}^*\right] \xrightarrow{P} 0. \tag{8}$$

To this end note that $\mathrm{E}_*[Y_r^*] = \hat{g}(z_r)$ and $Y_r^* - \hat{g}(z_r) = \varepsilon_r^*$. With these notations, the left-hand side of (8) is

$$\frac{1}{s_n^2(x)} \sum_{r=-n}^{n} w_{j,r,n}^2(x) \, \mathrm{E}_* \left[\varepsilon_r^{*2} I \left\{ |\varepsilon_r^*| > \delta \frac{s_n(x)}{|w_{j,r,n}(x)|} \right\} \right] \quad = \quad \frac{\sum_{r=-n}^{n} w_{j,r,n}^2(x) \Lambda_{n,r}}{\sum_{r=-n}^{n} w_{j,r,n}^2(x)}.$$

The Lindeberg condition is fulfilled if $\Lambda_{n,r} \xrightarrow{P} 0$ uniformly in r but with $s_n(x)/\max_{j=-n}^{n} |w_{j,r,n}(x)| = c_n \to \infty$ because of condition (4) there is

$$\Lambda_{n,r} \quad \leq \quad \mathrm{E}_* \left[\varepsilon_r^{*2} I \left\{ |\varepsilon_r^*| > \delta c_n \right\} \right] = \int_{(-\infty, -\delta c_n]} \varepsilon_r^{*2} dP_{r,*} + \int_{[\delta c_n, \infty)} \varepsilon_r^{*2} dP_{r,*} \tag{9}$$

where $P_{r,*}$ denotes the conditional measure of ε_r^* given $\mathcal{Y}$. The whole integral

$$\int_{(-\infty,\infty)} \varepsilon_r^{*2} dP_{r,*} = \mathrm{E}_*[\varepsilon_r^*]$$

exists and converges in probability to σ^2 (see (7)) and therefore the tail integrals in (9) converge to 0 in probability uniformly in r because, conditionally on $\mathcal{Y}$, the residuals ε_r^*, $r = -n, \ldots, n$ are identically distributed. We now obtain

$$(s_n(x))^{-1} \left(\hat{\theta}_n^*(x) - \mathrm{E}_*[\hat{\theta}_n^*(x)] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

and it remains to check if the difference between $\mathrm{E}\, \hat{\theta}_n^{(j)}(x) - \theta^{(j)}(x)$ and $\mathrm{E}_*[\hat{\theta}_n^{(j)\,*}(x)] - \hat{\theta}_n^{(j)}(x)$ converges in probability to 0. With (6) and (7) we obtain with a similar computation as in the proof of Lemma 3 for the difference between the two biases

$$\begin{aligned} \mathrm{E}[\hat{\theta}_n^{(j)}(x)] - \theta^{(j)}(x) - (\mathrm{E}_*[\hat{\theta}_n^{(j)\,*}(x)] - \hat{\theta}(x)) \quad &= \quad o_P(h^{s-j-1}) + O_P\Big(\frac{a_n^r}{h^{\beta+j+1}} \Big) + O_P\Big(\frac{1}{nh^{\beta+j+1} a_n} \Big) \\ &= \quad o_P\Big(\frac{1}{\sqrt{nh^{2\beta+2j+1} a_n}} \Big). \end{aligned}$$

□

References

Birke, M., Bissantz, N. and Holzmann, H. (2008), Confidence bands for inverse regression models – with application to gel electrophoresis. *Working paper, Ruhr-Universität Bochum and Universität Karlsruhe.*

Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007a), Nonparametric confidence bands in deconvolution density estimation, *J. R. Stat. Soc. Ser. B* **69** 483–506.

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.

Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inv. Prob.* **24**, 034009.

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

Delaigle, A. and Gijbels, I. (2002) Estimation of integrated squared density derivatives from a contaminated sample. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 869–886.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Eubank, R.L. (1999). *Nonparametric regression and spline smoothing.* Second edition. Statistics: Textbooks and Monographs, 157. Marcel Dekker, Inc., New York

Fan, J. (1991), On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.

Härdle, W. and Bowman, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102-110.

Hall, P. (1993), On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation. *J. R. Stat. Soc. Ser. B* **55**, 291–304.

Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application.* Probability and Mathematical Statistics, Academic Press, New York.

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

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255-285.

Munk, A., Bissantz, N., Wagner, T. and Freitag, G. (2005), On difference based variance estimation in nonparametric regression when the covariate is high dimensional, *J. Roy. Statist. Soc. Ser. B* **67**, 19–41.

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

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1295.