Confidence bands for inverse regression models

Melanie Birke[1], Nicolai Bissantz[1] and Hajo Holzmann[2]

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

[2]Fachbereich Mathematik und Informatik
Philipps-Universität Marburg

January 11, 2011

Abstract

We construct uniform confidence bands for the regression function in inverse, homoscedastic regression models with convolution-type operators. Here, the convolution is between two non-periodic functions on the whole real line rather than between two periodic functions on a compact interval, since the former situation arguably arises more often in applications. First, following Bickel and Rosenblatt [*Ann. Statist.* **1**, 1071–1095] we construct asymptotic confidence bands which are based on strong approximations and on a limit theorem for the supremum of a stationary Gaussian process. Further, we propose bootstrap confidence bands based on the residual bootstrap and prove consistency of the bootstrap procedure. A simulation study shows that the bootstrap confidence bands perform reasonably well for moderate sample sizes. Finally, we apply our method to data from a gel electrophoresis experiment with genetically engineered neuronal receptor subunits incubated with rat brain extract.

Keywords: Confidence bands, Inverse problems, Deconvolution, Rates of convergences, Nonparametric Regression, Bootstrap

1 Introduction

When dealing with inverse problems with stochastic noise, statistical methods have become a standard tool. There are two main approaches, either by Bayesian methods (Bertero et al., 2009; Kaipio and Somersalo, 2005) or by using methods from nonparametric curve estimation (Mair and Ruymgaart, 1996; Cavalier, 2008; Bissantz et al., 2007). For the latter method, the main research focus has been on constructing estimators which achieve the minimax risk of estimation. Further inferential methods, in particular the construction of confidence intervals and confidence bands, are much less developed. The purpose of this paper is therefore to construct asymptotic and bootstrap uniform confidence bands in a deconvolution regression model on the real line, and to apply the methods to data from a gel-electrophoresis experiment. Suppose that at our disposal are observations (z_k, Y_k), $k = -n, \dots, n$, from the model

$$Y_k = (A\theta)(z_k) + \epsilon_k, \tag{1}$$

where $z_k = k/(na_n)$, $a_n \to 0$ for $n \to \infty$ are fixed design points, the ϵ_k's are i.i.d. errors with $E\epsilon_k = 0$, $E\epsilon_k^2 = \sigma^2$, $E\epsilon_k^4 < \infty$, and A is a linear, one-to-one convolution operator with some

[2]Address for correspondence: Dr. Hajo Holzmann, Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerweinstr. D-35032 Marburg, Germany email: holzmann@mathematik.uni-marburg.de, Fon: + 49 6421 2825454

known function Ψ,

$$(A\theta)(z) = \int_{\mathbb{R}} \Psi(z-t)\,\theta(t)\,dt.$$

for the unknown regression function θ in model (1).
Note that in nonparametric deconvolution regression models, it is typically assumed (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. In general, for reconstruction problems of astronomical and biological images from telescopic and microscopic imaging devices which involves deconvolution, the assumption of periodicity of both θ and Ψ is often unrealistic, since the object of interest (for example a galaxy, say, or one single tissue cell) is not periodic. Neither is Ψ in such cases - rather it is a function (called the "point-spread-function") which is quite well localized around 0 in many cases. Hence (1) provides a more appropriate model in this context. Model (1) is closely related to density deconvolution (e.g. Fan, 1991a; Bissantz and Holzmann, 2008) and also to nonparametric errors-in-variables regression, in particular with Berkson errors (cf. Delaigle, Hall and Qiu 2006). For further discussion of our model and relations to the errors in variables model, see Section 2.
A specific application where the data can be modelled (approximately) by a one-dimensional convolution operator is polyacrylamide gel electrophoresis. Here, the task is to separate a mixture of molecules (nuclein acids or proteins) according to their different molecular masses. However, random effects such as diffusion in the gel result in a widening of these bands, which complicates separation of bands of proteins with very similar masses. We will use bootstrap confidence bands for deconvolution in order to conclude that a specific adaptor protein binds to the wildtype of a neuronal receptor subunit but not to a mutant version.
In a pioneering work, Bickel and Rosenblatt (1973) constructed confidence bands for a density function of i.i.d. observations, based on the asymptotic distribution of the supremum of a centered kernel density estimator. Since then, their method has been further developed both in the density estimation and also in a regression framework. For density estimation, Neumann (1998) constructs bootstrap confidence bands, and Giné and Nickl (2010) construct adaptive asymptotic bands over generic sets. In a regression context, asymptotic confidence bands were constructed by Eubank and Speckman (1993) for the Nadaraya-Watson estimator and by Xia (1998) for local polynomial estimators. Bootstrap confidence bands for nonparametric regression were proposed by Hall (1993), Neumann and Polzehl (1998) and by Claeskens and van Keilegom (2003). For the statistical inverse problem of deconvolution density estimation, Bissantz et al. (2007) constructed asymptotic and bootstrap confidence bands, while Lounici and Nickl (2010) obtain nonasymptotic conservative bands by using concentration inequalities. It is of major interest to obtain confidence bands for further inverse problems, and in this paper we make a first step in this direction for model (1). Note that while the operator in (1) is the same as in density deconvolution, the references above show that the construction of confidence bands in regression models is quite different.
When constructing confidence bands or even confidence intervals for $\theta(x)$, one has to deal with the bias. In a few of the above papers, notably in Eubank and Speckman (1993) and in Xia (1998), the bias is explicitly estimated and substracted, so that the rate-optimal bandwidth can still be used. However, estimating the bias involves estimating higher order derivatives, and the practical benefit of explicit bias correction is often quite small. Therefore, most of the above approaches rely on undersmoothing, where the bandwidth is chosen smaller than the rate optimal bandwidth, and the standard deviation then dominates the bias. Some theoretical justification why undersmooting is to be preferred when constructing confidence

intervals is given in Hall (1992). In this paper, we also restrict ourselves to undersmoothing.

The outline of the paper is as follows. In Section 2, we discuss the relation of our model to the Berkson errors in variables model and propose a kernel deconvolution estimator. In Section 3 we introduce the basic assumptions and present our asymptotic results for constructing confidence bands. Since it is well-known that convergence in the resulting limit theorems is rather slow (Hall 1993), we propose a bootstrap method based on the residual bootstrap in Section 3.2 and prove its consistency. The performance of the bootstrap confidence bands is investigated in a simulation study in Section 4. In Section 5 we use bootstrap confidence bands to analyze the results from a gel electrophoresis experiment with genetically engineered neuronal receptor subunits. Some proofs are given in the appendix, complete proofs can be found in Birke et al. (2010).

2 Methodology

Model (1) is related to the errors in variables model with Berkson errors, where (Z_k, Y_k) are observed which satisfy

$$Y_k = \theta(X_k) + \epsilon_k, \qquad X_k = Z_k + \delta_k, \tag{2}$$

and $Z_k, \delta_k, \epsilon_k$ are independent random variables with $E\epsilon_k = 0$ (cf. Delaigle et al. 2006). In (2) the aim is to estimate θ. Thus, $\theta(x) = E\big(Y_k | X_k = x\big)$ is related to the observed regression function $g(z) = E\big(Y_k | Z_k = z\big)$ via $g(z) = (\theta * \psi)(z)$, where ψ is the density of $-\delta_k$.
Hence model (1) is similar to a fixed-design Berkson errors-in-variables model, where the effect of the additional noise δ_k on θ is observed on average as the quantity $(\psi * \theta)(z)$. Methodologically, our model is closer to the classical nonparametric regression model with deconvolution as in Cavalier and Tsybakov (2002), where the blurring of θ does not necessarily arise through uncertainty about the design points.
In the following we propose a kernel-type estimator in model (1). To fix the notation, denote the Fourier transform of a function f by $\Phi_f(t) = \int_{\mathbb{R}} f(x) \exp(itx)\, dx$. Suppose that θ is p-times continuously differentiable for some $p \geq 0$. Under the assumption that $\Phi_\Psi(\omega) \neq 0$ for all $\omega \in \mathbb{R}$ and that Φ_k, the Fourier transform of the kernel k (which integrates to 1) has compact support, the kernel deconvolution estimator for the j^{th} derivative of θ, given by,

$$\hat{\theta}_n^{(j)}(x) = \hat{\theta}_{n,h}^{(j)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-i\omega)^j e^{-i\omega x} \Phi_k(h\omega) \frac{\hat{\Phi}_g(\omega)}{\Phi_\Psi(\omega)} d\omega, \quad 0 \leq j \leq p, \tag{3}$$

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

$$\widehat{\Phi}_g(\omega) = \frac{1}{n a_n} \sum_{r=-n}^{n} Y_r e^{i\omega z_r}.$$

From deconvolution density estimation, it is well-known that the optimal rate at which θ can be estimated depends on the smoothness of θ as well as on the smoothness of the convolution function Ψ, or equivalently on the tail properties of its Fourier transform. Roughly speaking, Ψ is ordinary smooth and hence the inverse problem is mildly ill-posed if the Fourier transform $|\Phi_\Psi(t)|$ decays at a polynomial rate as $t \to \infty$, in which case the optimal rate for estimating θ is also of polynomial order. In contrast, if $|\Phi_\Psi(t)|$ decays at an exponential rate as $t \to \infty$, Ψ

is supersmooth, the problem is called severely ill-posed and the optimal convergence rate for θ is typically only of logarithmic order. For details in the density estimation context see Fan (1991a) and Pensky and Vidakovic (1999), among others. In the following we shall restrict ourselves to ordinary smooth Ψ, which yields a mildly ill-posed problem in model (1).

More specifically, we shall assume that

$$\Phi_\Psi(\omega)\omega^\beta \to C_\epsilon, \qquad \omega \to \infty, \tag{4}$$

for some $\beta \geq 0$ and $C_\epsilon \in \mathbb{C} \setminus \{0\}$. Note that this implies that $\Phi_\Psi(\omega)|\omega|^\beta \to \bar{C}_\epsilon$, $\omega \to -\infty$. For example, if Ψ is the density of a Laplace distribution, we have $\Phi_\Psi(\omega) = 1/(1+\omega^2)$, and assumption (4) holds with $\beta = 2$ and $C_\epsilon = 1$

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

$$\hat{\theta}_n^{(j)}(x) = \frac{1}{nh^{j+1}a_n} \sum_{r=-n}^{n} Y_r K^{(j)}\left(\frac{x - z_r}{h}; h\right),$$

where the deconvolution kernel $K^{(j)}(z;h)$ is given by

$$K^{(j)}(z;h) = \frac{1}{2\pi}\int_{\mathbb{R}} (-i\omega)^j e^{-i\omega z} \frac{\Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega, \quad 0 \leq j \leq p. \tag{5}$$

Now, if (4) holds, the deconvolution kernel $K^{(j)}(z;h)$ given in (5) has a simple asymptotic form. In fact, from the dominated convergence theorem,

$$h^\beta K^{(j)}(z;h) \to K^{(j)}(z), \quad h \to 0,$$

where

$$\begin{aligned} K^{(j)}(z) &= \frac{1}{2\pi C_\epsilon}\int_0^\infty (-i\omega)^j \exp(-i\omega z)\omega^\beta \Phi_k(\omega)\, d\omega \\ &\quad + \frac{1}{2\pi\overline{C_\epsilon}}\int_{-\infty}^0 (-i\omega)^j \exp(-i\omega z)|\omega|^\beta \Phi_k(\omega)\, d\omega, \end{aligned} \tag{6}$$

c.f. Fan (1991b). Note that the second term in (6) is the complex conjugate of the first, so that $K^{(j)}(z)$ is in fact real-valued. This shall allow us e.g. to obtain an explicit asymptotic formula for the pointwise variance of the estimator (3), which turns out to be proportional to $\sigma^2/(nh^{2\beta+2j+1}a_n)$.

3 Confidence bands for inverse regression

When constructing confidence bands, we shall concentrate on the case when Ψ and θ have non-compact support, in which we let $a_n \to 0$. Results for compactly supported functions, where $a_n = a$ can be chosen as an appropriate constant, can be derived analoguously.

3.1 Asymptotic confidence bands

In the following we list our exact assumptions which are required subsequently.

Assumption 1. The Fourier transform Φ_k of k is symmetric, three times differentiable and supported on $[-1,1]$, $\Phi_k(\omega) = 1$ for $\omega \in [-c,c]$, $c > 0$, and $|\Phi_k(\omega)| \leq 1$.

Assumption 2. A. $\int_{\mathbb{R}} |K^{(j+1)}(z;h)||z|^{3/2}(\log\log^+|z|)^{1/2}\,dz = O(h^{-\beta})$, where $\log\log^+|z| = 0$ if $|z| < e$, and $\log\log^+|z| = \log\log|z|$, otherwise.

B. For some $\delta > 0$,

$$\int_{\mathbb{R}} |h^\beta K^{(j+1)}(z;h) - K^{(j+1)}(z)||z|^{1/2}(\log\log^+|z|)^{1/2}\,dz = O(h^{1/2+\delta}),$$

where $K^{(j+1)}$ is given in (6).

C. Uniformly in z,

$$|h^\beta K^{(j+1)}(z;h) - K^{(j+1)}(z)| = O(h^{1/2+\delta}).$$

D. The limit kernel $K^{(j)}(z)$ in (6) has exponentially decreasing tails.

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

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

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

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

Remark 1 (*Discussion of the Assumptions).* Assumptions 1, 2 A. and B. and 3 A. also occur in Bissantz et al. (2007), for a detailed discussion see Remark 1 in that paper. In particular, Assumption 2 B. is a technical refinement of (4), indeed, the limit kernel $K^{(j)}$ is required to formulate it in the first place. Assumption 2 C. can also be checked by using (14) from Bissantz et al. (2007). As for Assumption 2 D., this amounts to exponentially decreasing tails of the second derivative of the kernel k if $\beta = 2$ and $C_\epsilon = 1$ (e.g. Laplace noise), and $j = 0$.
Finally, Assumption 3 B. is satisfied if $r > 1$ and

$$\int_{\mathbb{R}} |\theta(z)||z|^r\,dz < \infty, \qquad \int_{\mathbb{R}} |\Psi(z)||z|^r\,dz < \infty,$$

since using $|u+t|^r \le 2^r(|u|^r + |t|^r)$,

$$\begin{aligned}\int_{\mathbb{R}} |g(z)||z|^r\,dz &\le \int_{\mathbb{R}} |\theta(u)| \int_{\mathbb{R}} |u+t|^r \Psi(t)\,dt\,du \\ &\le 2^r \int_{\mathbb{R}} |\theta(u)||u|^r du + 2^r \int_{\mathbb{R}} |\theta(u)|\,du \int_{\mathbb{R}} |t|^r \Psi(t)\,dt\end{aligned}$$

We now construct asymptotic confidence bands for θ on compact intervals for ordinary smooth Ψ and, as a byproduct, determine rates of uniform convergence of the estimator (3). To facilitate a concise presentation we formulate the results for the interval $[0,1]$, however the generalization to $[a,b] \subset \mathbb{R}$ is straightforward (by affine transformation). Similarly as in Bickel and Rosenblatt (1973) we shall investigate the distribution of the supremum of the process

$$Z_n^{(j)}(x) = \frac{n^{1/2}h^{\beta+j+1/2}a_n^{1/2}}{\sigma}\left(\hat\theta_n^{(j)}(x) - E[\hat\theta_n^{(j)}(x)]\right),\ x \in [0,1].$$

Let $\|\cdot\|_I$ denote the sup-norm on an interval $I \subset \mathbb{R}$. Next we state our main limit theorem.

Theorem 1. *Let Assumption 1-3 hold,* $a_n \to 0$, $h^{2\delta}\log(n)/a_n \to 0$, $h^3 a_n^3 n/\log(n)^2 \to \infty$. *Then, for* $0 \le j \le p$,

$$P\left((2\log(1/h))^{1/2}(\|Z_n^{(j)}\|_{[0,1]}/C_{K,1}^{1/2} - d_n) \le \kappa\right) \to \exp(-2\exp(-\kappa)),$$

where

$$d_n = \left(2\log(1/h)\right)^{1/2} + \frac{\log\left(\frac{1}{2\pi}C_{K,2}^{1/2}\right)}{\left(2\log(1/h)\right)^{1/2}},$$

and

$$C_{K,1} = \frac{1}{2\pi|C_\epsilon|^2}\int_{\mathbb{R}} \omega^{2(\beta+j)}\Phi_k^2(\omega)\,d\omega, \quad C_{K,2} = \frac{\int_{\mathbb{R}} \omega^{2(\beta+j+1)}\Phi_k^2(\omega)\,d\omega}{\int_{\mathbb{R}} \omega^{2(\beta+j)}\Phi_k^2(\omega)\,d\omega}. \tag{7}$$

In order to construct confidence bands for $\theta^{(j)}$ we have to deal with the bias of $\hat{\theta}_n^{(j)}$. In the appendix we show that

$$\max_{x\in[0,1]} |E\hat{\theta}_n^{(j)}(x) - \theta_n^{(j)}(x)| = O\left(h^{s-j-1} + h^{-(\beta+j+1)}a_n^r\right) \tag{8}$$

Note that in contrast to deconvolution density estimation, where the bias does not depend on the error density, the order in (8) does depend on the index β of Φ_Ψ. As a consequence, the additional bias term decays to zero the slower (if it converges at all), the larger β is. However, by requiring that r in Assumption 3 is sufficiently large, we have that $h^{-(\beta+j+1)}a_n^r = o\left(h^{s-j-1}\right)$. This holds e.g. for convolution with a Laplace density and if θ is a function of compact support or exponential decay of its tails. This condition appears to be rather natural for practical applications, where the signal θ to be reconstructed is of limited extend in space or time, e.g. in microscopic or telescopic imaging, to mention only a few examples.
Next we give uniform confidence bands for the problem under consideration. To this end, assume that $\hat{\sigma}^2$ is an estimator of the variance σ^2 with rate $o_P((\log(1/h))^{-1})$ (cf. e.g. Munk et al., 2005), where h is the bandwidth used to estimate θ. For correcting the bias, as discussed in the introduction we shall rely on undersmoothing.

Corollary 2. *Let* $\hat{\sigma}^2$ *be an estimator of* σ^2 *with convergence rate* $o_P((\log(1/h))^{-1})$. *Under the assumptions of Theorem 1, if* $nh^{2(\beta+j)+1}a_n/\log(1/h) \to \infty$ *and* $\log(1/h)\cdot(nh^{2(\beta+s)-1}a_n + nh^{-1}a_n^{1+2r}) \to 0$, *we have*

$$P\left(\hat{\theta}_n^{(j)}(x) - b_n(x,\kappa) \le \theta^{(j)}(x) \le \hat{\theta}_n^{(j)}(x) + b_n(x,\kappa) \text{ for all } t \in [0,1]\right) \to \exp(-2\exp(-\kappa)),$$

where

$$b_n(x,\kappa) = \left(\frac{\hat{\sigma}^2 C_{K,1}}{nh^{2(\beta+j)+1}a_n}\right)^{1/2}\left(\frac{\kappa}{(2\log(1/h))^{1/2}} + d_n\right).$$

Remark 2. The width of the bands is $(\log(1/h)/nh^{2(\beta+j)+1}a_n)^{1/2}$. Hence the first condition in Corollary 2 ensures that this width converges to zero.
Undersmoothing in order to correct for the bias requires that, as $n \to \infty$ and $a_n, h \to 0$ we need to have $(\log(1/h))^{-1}\cdot(nh^{2(\beta+s)-1}a_n + nh^{-1}a_n^{1+2r}) \to 0$. These two conditions can be met simultaneously since $s > p+1$ and $j \le p$. As discussed previously, this can e.g. be achieved if the signal θ has compact support, or exponentially decaying tails, or if the interfering convolution function Ψ has exponential tails, such as the density of a Laplace distribution.

Further we obtain that under the assumptions of Theorem 1, if additionally $\frac{nh^{2\beta+2s-1}a_n}{\log h^{-1}} = O(1)$ and $\frac{nh^{-1}a_n^{2r+1}}{\log h^{-1}} = O(1)$, then the estimator $\hat{\theta}^{(j)}$ has uniform convergence rate

$$\sup_{x\in[0,1]} |\hat{\theta}^{(j)}(x) - \theta^{(j)}(x)| = O_P\Big(\frac{\log h^{-1}}{nh^{2\beta+2j+1}a_n}\Big)^{1/2}, \tag{9}$$

for the proof of (9)see the appendix.

3.2 Bootstrap confidence bands

It is well known, both from simulations as well as from theoretical investigations (Hall 1993), that the rate of convergence in Theorem 1 is rather slow and hence that the resulting confidence bands perform rather poorly in terms of coverage probability. Therefore, bootstrapping is a popular alternative to construct confidence bands. For direct density estimation, Hall (1993) investigated the rate of convergence for the simple nonparametric bootstrap (i.e. drawing n out of the n observations with repetitions) via the Edgeworth expansion. Neumann (1998) constructed direct strong approximations for the bootstrap process. For indirect density estimation, Bissantz et al. (2007) used a simple argument via strong approximation of the empirical process to show consistency of the bootstrap procedure. In the context of regression, Neumann and Pohlzehl (1998) used the wild bootstrap in a heteroscedastic regression model allowing both fixed and random design, and Claeskens and van Keilegom (2003) used the smooth bootstrap (for the actual observations, not the residuals) for homoscedastic likelihood regression models with random design. Both prove consistency of the resulting bootstrap procedures, with arguments relying on the strong approximation of the bootstrap processes.

For regression models with fixed design, these design points should be fixed during the bootstrapping procedure as well. Hence, the simple nonparametric bootstrap (drawing with replacement from the pairs (z_k, Y_k)) is inappropriate. Instead, one aims at bootstrapping from the distribution of the errors ϵ_i. If these are homoscedastic (as is the case in our model), one bootstraps from the residuals, which is called the residual bootstrap (Hall 1992). Although this approach is problematic in heteroscedastic models (cf. e.g. Neumann and Polzehl 1998), it is the method of choice in homoscedastic models. Therefore, in the following we propose a bootstrap procedure based on the residual bootstrap.

Since the bootstrapping procedure requires the residuals, we shall, in addition to the target function $\theta^{(j)}$, have to estimate the function g (or equivalently θ). Consider the residuals

$$\tilde{\epsilon}_i = Y_i - \hat{g}_n(z_i),$$

where

$$\hat{g}_n(x) = \big(A\hat{\theta}_n\big)(x) = \frac{1}{n\, a_n\, \tilde{h}} \sum_{r=-n}^{n} Y_r k\Big(\frac{x - z_r}{\tilde{h}}\Big),$$

and $\tilde{h}$ has to be chosen for estimation of θ (and not its derivatives). Due to boundary bias of the estimator $\hat{g}_n(x)$, one has to exclude those residuals which are to close to the boundary of the observational interval (see e.g. Hall, 1992 or Härdle and Bowman, 1988). For our situation, one has to observe that first, due to Assumption 1 the kernel k has compactly supported Fourier transform and hence cannot itself have compact support. Nevertheless,

typically k will be rapidly decreasing in the tails. Thus, instead of excluding points at a rate of $\tilde{h}$, we shall exclude points at a rate η such that $\tilde{h} = o(\eta)$. Second, we shall only exclude $o(n)$ observations. Since in a fixed interval there are only $na_n = o(n)$ of these, we could keep the η at a fixed distance from the boundary of the observational interval. Summarizing, for η with

$$\eta = O(1), \qquad \tilde{h} = o(\eta), \tag{10}$$

we use only those $\tilde{\epsilon}_i$ for which $-a_n^{-1} + \eta \le z_i \le a_n^{-1} - \eta$, or those i with $-(n - \eta a_n n) \le i \le (n - \eta a_n n)$ and set

$$\hat{\epsilon}_i = \tilde{\epsilon}_i - \frac{1}{2\lfloor n(1-\eta a_n)\rfloor} \sum_i \tilde{\epsilon}_i, \tag{11}$$

where the sum is taken over $-(n - \eta a_n n) \le i \le (n - \eta a_n n)$. Now draw with replacement from the $\hat{\epsilon}_i$ a bootstrap sample of residuals $\epsilon^*_{-n}, \ldots, \epsilon^*_n$. A bootstrap approximation to the processes $Z_n^{(j)}(x)$ or $Z_{n,6}^{(j)}(x)$ (in the appendix) is given by

$$Z_n^{(j)*}(x) = \frac{n^{1/2} h^{\beta+j+1/2} a_n^{1/2}}{\hat{\sigma}^*} \sum_{r=-n}^{n} \epsilon_r^* \, K^{(j)}\left(\frac{x - z_r}{h}; h\right),$$

where $\hat{\sigma}^*$ is computed as $\hat{\sigma}$ but from the bootstrap observations $Y_i^* = \hat{g}_n(z_i) + \epsilon_i^*$. Let $q^*_{1-\alpha}$ denote the α-quantile of $\sup_{x\in[0,1]} |Z_n^{(j),*}(x)|$, conditional on the original observations. The bootstrap confidence band for $E\hat{\theta}^{(j)}$ (and for $\hat{\theta}^{(j)}$ in case of undersmoothing) is given by

$$\Big[\hat{\theta}^{(j)}(x) - \frac{\hat{\sigma}\, q^*_{1-\alpha}}{n^{1/2} h^{\beta+j+1/2} a_n^{1/2}}, \hat{\theta}^{(j)}(x) + \frac{\hat{\sigma}\, q^*_{1-\alpha}}{n^{1/2} h^{\beta+j+1/2} a_n^{1/2}}\Big], \qquad x \in [0,1].$$

We now state the weak consistency of the residual bootstrap in this situation. To this end from Neumann and Polzehl (1998) we adopt the notation

$$U_n = \tilde{O}(V_n, \gamma_n) \text{ if } \mathrm{P}(|U_n| > CV_n) \le C\gamma_n$$

for $n \ge 1$ and some constant $C < \infty$. In Theorem 3 we derive a rate of convergence in the $\tilde{O}$ sense of the difference of $\{(\hat{\theta}^{(j)}(x) - \theta^{(j)}(x))/\hat{\sigma}\}_{x\in[0,1]}$ and $\{Z_n^{(j)*}(x)/nh^{2\beta+2j+1}a_n)^{1/2}\}_{x\in[0,1]}$.

Theorem 3. *Under the assumptions of Theorem 1, if the $\{\epsilon_i^*\}_{i=-n}^n$ are defined as above for η chosen as in (10), there exists a version of $\{\epsilon_i\}$ and conditionally on $\mathcal{Y} = \{Y_{-n}, \ldots, Y_n\}$ a version of $\{\epsilon_i^*\}$ on a rich enough probability space such that*

$$\sup_{x\in[0,1]} \left| \frac{1}{\hat{\sigma}}(\hat{\theta}^{(j)}(x) - \theta^{(j)}(x)) - \frac{1}{(nh^{2\beta+2j+1}a_n)^{1/2}} Z_n^{(j)*}(x) \right| = \tilde{O}\left(\frac{n^\delta \log(h^{-1})}{nh^{\beta+j+1}a_n} + h^{s-j-1}, n^{-\lambda}\right)$$

for some arbitrary small $\delta > 0$ and some $\lambda < \infty$.

The proof is similar to that of Theorem 2.1 in Neumann and Polzehl (1998) and is defered to the Appendix. Theorem 3 together with Corollary 2 yield the consistency of the residual bootstrap for constructing confidence bands. Note, that the method is not restricted to the homoscestic case or to the residual bootstrap. Therefore, a similar result will hold if we use wild bootstrap.

4 Simulations

We simulate from model (1), where the ε_k are independent $N(0, \sigma^2)$-distributed. For the unknown regression function we consider

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

These two functions essentially deviate from 0 in the interval $[-4, 4]$, and we will hence use $a_n = 0.25$ in the subsequent simulations. In the main part of our simulations the convolution function is $\psi(x) = \lambda\, e^{-\lambda|x|}/2$, with $\lambda = 3$. Thus, its scale is of similar magnitude as those of the regression functions θ_1 and θ_2.
In all simulations we determined the actual coverage probability and confidence band area from 200 randomly generated data sets according to model (1). For each of these random datasets, uniform confidence bands for the function $\theta(x)$ on the interval of interest $[0, 1]$ were determined from a residual bootstrap with 400 replications. Here, the sampling distribution for the residuals was estimated from the re-centered residuals computed for observations with $|z_k| \leq \frac{1}{a_n} - 2.01h$ (cf. (11)).

4.1 Bandwidth selection

We now discuss the selection of the bandwidth h for estimator (3). Fig. 1 shows the simulated coverage probabilities and band areas for 90% nominal coverage probability for the two test functions θ_1 (left) and θ_2 (right). In both cases the sample size is $2n + 1 = 201$ and $\sigma = 0.1$. The effective coverage probability of the confidence band is significantly below the nominal coverage probability for bandwidths larger than approximate L_∞-optimal bandwidth, which can be determined from the figure as the location of the minimum of the mean sup-distance between estimates and the true functions θ_1 and θ_2, respectively. This effect is due to the increase in the bias with increasing bandwidth, which results in a decrease in coverage propability. On the other hand, the mean area of the bootstrap confidence bands increases strongly with decreasing bandwidth due to the increasing variance. Hence, a suitable choice of bandwidth is the *largest* bandwidth, for which the effective coverage probability still matches its nominal value, at least approximately. Fig. 1 indicates that a suitable bandwidth is slightly smaller than the L_∞-optimal bandwidth, which is consistent with the idea of undersmoothing.

Estimation of the L_∞-optimal bandwidth (or some slightly smaller value) is not straightforward, as the true function $\theta(x)$ used to produce the dotted curves in Fig. 1 is obviously not known in practice. However, a suitable choice for the bandwidth is possible with the L_∞-based bandwidth selector introduced by Bissantz et al. (2007) for the density deconvolution case. In short, its idea is to replace the problem of determining the bandwidth with smallest mean sup-distance between estimates and true (and in practice unknown) function θ by the problem of determining the largest bandwidth, for which the sup-distance between estimates for two subsequent bandwidth values is above a certain threshold value, where we used $\tau = 5$ in our simulations, which results in somewhat undersmoothing bandwidth choices. In our simulations it turned out that considering 12 different bandwidths (indicated by the small circles in Fig. 1), covering an order of magnitude in value, is sufficient to allow for satisfying confidence band properties, as discussed below.
As an illustration, Fig. 2 shows 90% nominal coverage probability confidence bands for the estimates $\hat{\theta}(x)$ on $[0, 1]$ of θ_1 and θ_2, respectively, from $2n + 1 = 201$ observations based on

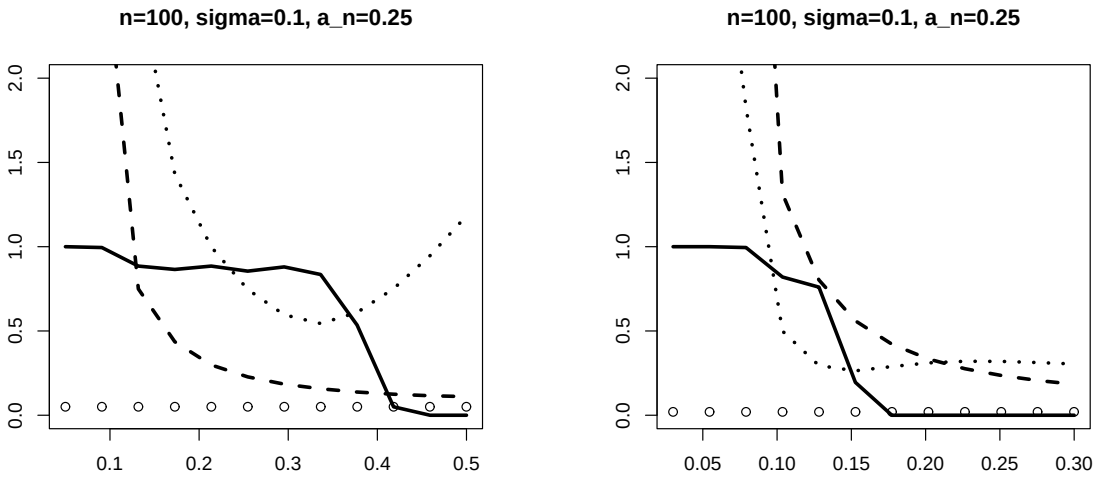

Figure 1: Average width and coverage probability of confidence bands with a nominal coverage probability of 90% for the Gaussian function θ_1 (left) and the bimodal function θ_2 (right). Solid lines represent simulated coverage probabilities from 200 simulations, dashed lines $1.5\times$ the mean area of these bootstrap confidence bands on the interval of interest $[0, 1]$, and dotted lines the mean sup-distance between estimates and the true functions θ_1 and θ_2, respectively. In the case of θ_1 the sup-distance has been multiplied by 10. Finally, circles indicate the bandwidth values considered to select the bandwidth for the subsequent simulations.

data with $\sigma = 0.1$ and $a_n = 0.25$ and the bandwidth selected by the L_∞-optimal bandwidth selector.

4.2 Simulated coverage probabilities and confidence band areas for bootstrap confidence bands

In this section we present the results on coverage probabilities and confidence band areas for bootstrap confidence bands. For each combination of the parameters n, σ and regression functions θ_1, θ_2 we first determined a suitable bandwidth h for estimator (3) from the L_∞-bandwidth estimator. Table 1 shows the results for simulations with the unimodal function

n	σ	80% nominal cov.		90% nominal cov.		95% nominal cov.	
		Cov. prob.	Width	Cov. prob.	Width	Cov. prob.	Width
100	0.5	79.0	0.216	87.5	0.258	95.0	0.293
100	0.1	78.0	0.085	88.5	0.100	94.0	0.113
1000	0.5	79.5	0.139	90.5	0.163	95.5	0.184
1000	0.1	76.5	0.041	86.5	0.048	92.5	0.054

Table 1: Simulated coverage probabilities and confidence band widths for the Gaussian function θ_1.

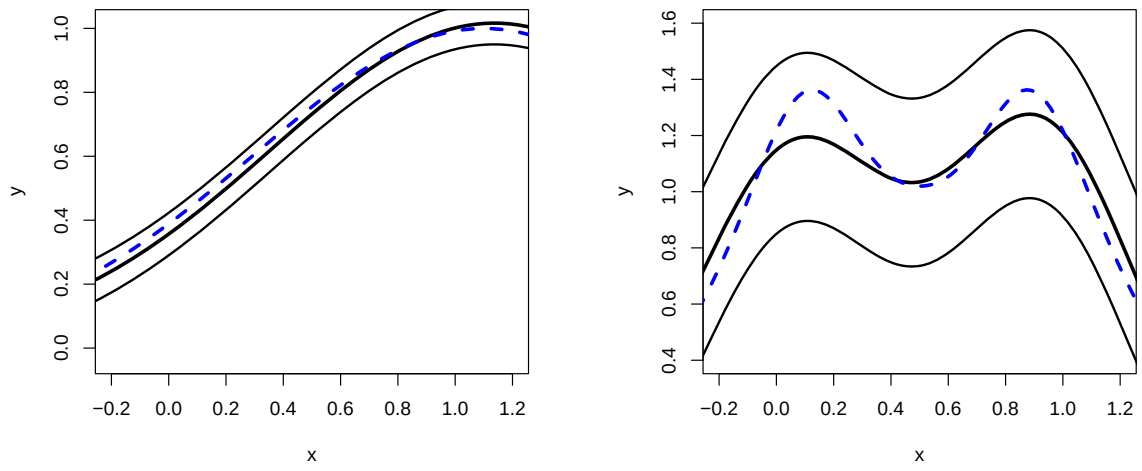

Figure 2: Estimate $\hat{\theta}_n(x)$ and associated 90% nominal coverage probability residual bootstrap confidence bands (solid lines) for the Gaussian function θ_1 (left) and the bimodal function θ_2 (right). Dashed lines represent the true functions θ_1 and θ_2, respectively.

n	σ	80% nominal cov.		90% nominal cov.		95% nominal cov.	
		Cov. prob.	Width	Cov. prob.	Width	Cov. prob.	Width
100	0.1	63.0	0.231	76.5	0.268	82.0	0.299
100	0.02	56.0	0.081	75.5	0.093	83.5	0.104
1000	0.1	74.5	0.131	90.5	0.152	95.0	0.169
1000	0.02	79.5	0.054	91.5	0.062	96.5	0.069

Table 2: Simulated coverage probabilities and confidence band widths for the bimodal function θ_2.

θ_1. The confidence bands perform rather well with respect to the coverage probabilities and the confidence band widths, which are significantly smaller for sample size $2n + 1 = 2001$ than for $2n+1 = 201$. Moreover, the bands for $a_n = 0.25$ are narrower by a factor of nearly 2 than for $a_n = 0.1$ which is due to the fact that a smaller value of a_n implies a larger interval covered by the design points. In consequence, the number of observations within the interval of interest $[0, 1]$ effectively decreases. On the other hand, determination of the empirical Fourier transform of $g = A\theta$ benefits from a larger interval covered by the design points, which implies that some trade-off has to be made in order to fix a_n in practical applications. Now turn to Table 2 which shows the results obtained from simulations with the bimodal function θ_2. The bands do not perform as well as for the unimodal function θ_1, particularly for sample size $2n + 1 = 201$, since the shorter scale of variation of θ_2 along the $x-$axis implies a stronger impact of bias at given bandwidth. This also implies that some simulations performed with $a_n = 0.1$ (not shown) produced unsatisfactory results. However, for a suitably chosen value of a_n the confidence bands appear useful for the bimodal function θ_2, even for

Setting	80% nominal cov.		90% nominal cov.		95% nominal cov.	
	Cov.prob.	Width	Cov.prob.	Width	Cov.prob.	Width
5% underestimated	76.0	0.071	84.5	0.084	90.0	0.094
5% overestimated	79.0	0.060	89.0	0.071	92.5	0.080
Lap., miss-specified as Gauss.	78.0	0.090	86.5	0.106	92.0	0.120
Gauss., miss-specified as Lap.	74.5	0.086	87.0	0.101	92.0	0.114

Table 3: Simulated coverage probabilities and confidence band widths for various settings of miss-specifications for the convolution density ψ. In all cases, $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$.

the smaller of the sample sizes considered, as is also indicated by Fig. 2.

4.3 Robustness and misspecification of ψ

In practical applications, the convolution function ψ is often not fully known. Hence, in the final part of the simulations we have considered some typical cases of miss-specification of the function ψ:

- The width (or standard deviation if ψ is a density) of the convolution function ψ may be miss-specified. Hence we performed simulations where the standard deviation of ψ is over- or underestimated by 5%, respectively.
- The geometric shape of the function ψ may only be approximately known. We considered both the case that ψ is in fact Gaussian with variance 2/9, i.e. the errors of x are normally distributed, but specified as Laplace with same variance in the data analysis, and the reverse case, where ψ is the Laplace density but miss-specified as Gaussian with same variance 2/9.

Table 3 shows the results of these simulations. Whereas in all of these miss-specifiation scenarios our asymptotic theory for the confidence bands does not hold, the simulation results are quite satisfactory with simulated coverage probabilities close to their nominal values and confidence band width about $20 - 80\%$ larger than for ψ correctly specified (cf. the results in Table 1). Hence, the bootstrap confidence bands appear to be well-suited for practical applications, as soon as the convolution function ψ is at least approximately known.

5 Gel electrophoresis of genetically engineered neuronal receptor subunits

5.1 Experimental setup

In this section we apply our methods to data from a gel electrophoresis experiment usually carried out in molecular biology to separate dna, rna or protein molecules according to their molecular weight, e. g. for the subsequent application of other techniques as for example mass spectrometry or PCR. In our case, a sample containing proteins of interest was applied to a plane gel of polyacrylamide and exposed to an electric field along the gel. According to their weight migrating molecules are focused as a bands visible on the gel.

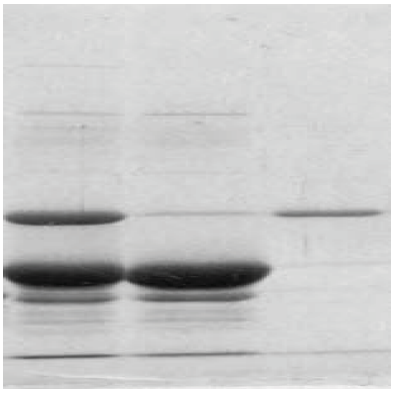

Figure 3: Result from a gel electrophoresis experiment with genetically engineered neuronal receptor subunits incubated with rat brain. Lanes (running in horizontal direction) are for wildtype receptor, mutante receptor and a standard molecule (from left to right).

Figure 3 shows the result of a gel electrophoresis of genetically engineered neuronal receptor subunits incubated with rat brain extract to capture other proteins that specifically bind to the wildtype (left lane) but not to the mutated receptor (middle lane). The right lane shows a standard, mono-constituent sample of the adapter protein. Here, the smaller receptor tail moves faster than the adaptor protein binding to it. Therefore, the receptor band can be found near the bottom of the gel whereas the upper band is the binding adaptor protein. Intensity of the bands is according to the amount of protein in it. The wildtype receptor subunit binds a higher amount of adaptor protein than the mutant. Therefore, binding of the adaptor protein is specific to the wildtype receptor but not to the mutant (left lane in Fig. 3) in the middle lane. There is a weak band above the mutant receptor which appears to be slightly offset the migration height of the adaptor protein, and may therefore be due to some different molecule. In the sample containing the wildtype receptor this band may be overlayed by the band of the adaptor protein binding to the receptor. However, all bands in this experiment show a certain width. This is due to random effects such as diffusion, that affects all molecules in the solution, and furthermore due to the unavoidable biodegradation of proteins over time, which results in molecules of masses very close, but not identical, to the original protein. In order to make a firm conclusion if the weak line in the mutante probably is offset in position (and hence differs in molecular mass from the adaptor protein), this broadening of the lines has to be removed.

5.2 Statistical model and analysis

In our subsequent analysis we model the data as follows. Since the large extension of the protein bands perpendicular to the movement of the molecules is due to the width of the lane where the solution was applied at the starting point of the gel, we integrate the signal for each sample along this direction. The resulting profile (along the direction of movement) can then be closely modeled by a one-dimensional convolution of the form (1), where the covariable x is the distance from the starting point of a lane to the position under consideration, and the response is the signal integrated orthogonal to the direction of the x-coordinate.
As mentioned above, the bottom lane shows the band produced by a standard molecule of

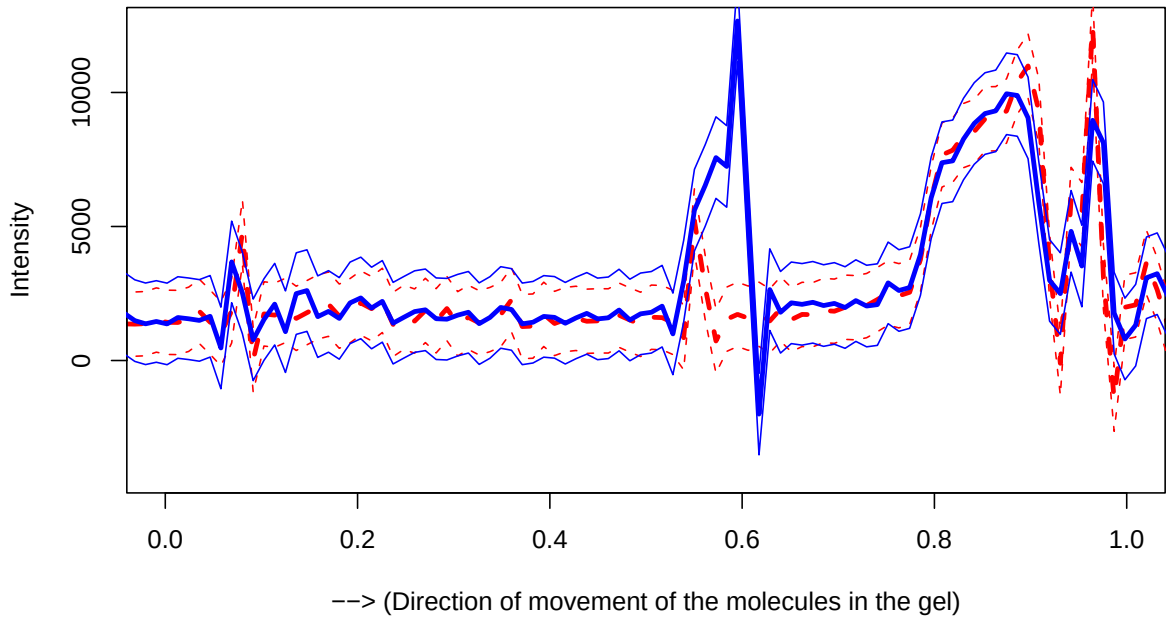

Figure 4: Intensity profiles and associated 90% bootstrap confidence bands for wildtype and mutante in the gel electrophoresis experiment discussed in the text. Solid lines show the distribution for the wildtype, and dashed lines for the mutante.

fixed weight. For our subsequent analysis we model the line-spread function $\Psi(x)$, which models the way a protein band is broadened due to random effects as the density of a Laplace distribution. This assumption appears safe given the simulation results shown in Section 4, which show that the bootstrap confidence bands are robust against a moderate miss-specification of the convolution function $\Psi(x)$. We estimate the parameter λ of $\Psi(x)$ from the profile of the standard molecule, which was to this end again integrated perpendicular to the direction of movement.

Fig. 4 shows estimates of the profiles for the wildtype receptor and the mutated receptor, together with 90%-bootstrap confidence bands from 100 bootstrap replications.

In order to compute the estimator we used a ν-method with $\nu = 1$ and 40 iterations. Note that this corresponds to a slightly different form of the estimator (3). Indeed, the regularization of the inverse $1/\Phi_\Psi$ is not achieved by multiplication with a function of compact support leading to $\Phi_K(h\omega)/\Phi_\Psi(\omega)$, but rather by using a general regularization approach $F(\Phi_\Psi; \alpha)(\omega)$ which converges to $1/\Phi_\Psi(\omega)$ as $\alpha \to 0$ (for further details see Bissantz et al. 2007). The reason is that regularization by the ν method performs better for capturing the steep peaks in the regression function as shown in Fig. 4. ¿From the deconvolved profiles it is straightforward to conclude that the weak band above the mutated receptor is clearly offset from the strong band visible above the wildtype receptor, whereas other bands in the profiles are not offset which excludes an inhomogeneous electric force (and hence speed of molecular motion) as explanation for this offset. Hence, we conclude that the weak band corresponds

to a protein molecule different from the intense band of the adaptor protein binding to the wildtype receptor. Very probably the molecule resulting in the weak band above the receptor subunit mutant is also present as a weak band above the wildtype receptor, but this band is overlayed by the more intense band of the adaptor protein. From these results we conclude that the adaptor protein specifically binds to the wildtype receptor subunit but not to the mutant subunit.

6 Conclusions

In order to assess the precision of statistical estimators, it is essential to construct accompanying confidence intervals or even confidence bands. In this paper, we introduced a kernel-type estimator for a noisy nonparametric regression problem, which requires an additional deconvolution, and construct a uniform confidence band for such an estimator.
Generally speaking, such deconvolution techniques should find broad application in the reconstruction of images from fluorescence microscopy at the nanoscale. These experiments invariably include the observation of inherently stochastic phenomena with substantial measurement error. This measurement error is often ignored in practice leaving some experimental conclusions in doubt.
Constructing confidence intervals and bands is a well-studied problem in direct nonparametric regression and density estimation problems, but there are few examples for inverse estimation problems. Therefore, extensions of our results to other models such as positron emission tomography should be studied in the future. A further promising extension is to introduce shape restrictions for the target functions as in Dümbgen (2003).

Acknowledgements

Melanie Birke and Nicolai Bissantz gratefully acknowledge support of the SFB 823 and from the BMBF project INVERS. Hajo Holzmann acknowledges financial support from the Deutsche Forschungsgemeinshaft, grant Ho 3260/3-1 and the Claussen-Simon-Stiftung. The authors would like to thank Holger Dette and Axel Munk for helpful discussions.

References

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

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

Birke, M., Bissantz, N. and Holzmann, H. (2010), Confidence bands for inverse regressions models: Technical details. Supplementary material.

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

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

Bissantz, N. and Holzmann, H. (2008), Statistical inference for inverse problems. *Inverse Problems* **24**, doi: 10.1088/0266-5611/24/3/034009.

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

Claeskens, G. and van Keilegom, I. (2003), Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, **31**, 1852–1884.

Delaigle, A., Hall, P. and Qiu, P. (2006), Nonparametric methods for solving the Berkson errors-in-variables problem. *J. R. Statist. Soc. B*, **68**, 201–220.

Dümbgen, L. (2003), Optimal confidence bands for shape-restricted curves. *Bernoulli*, **9**, 423–449.

Eubank, R. L. and Speckman, P. L. (1993), Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.*, **88**, 1287–1301.

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

Fan, J. (1991b), Asymptotic normality for deconvolution kernel density estimators. *Sankhya Ser. A*, **53**, 97–110.

Giné, E. and Nickl, R. (2010), Confidence bands in density estimation. *Ann. Statist.*, to appear.

Hall, P. (1992) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist* **20**, 675–694.

Hall, P. (1992), On bootstrap confidence intervals in nonparametric regression. *Ann Statist* **20**, 695-711.

Hall, P. (1993), On Edgeworth expansion and bootstrap confidence bands in nonparametric regression. *J. R. Statist. Soc. B*, **55**, 291–304.

Halmos, P. R. (1963), What does the spectral theorem say? *Amer. Math. Monthly* **70**, 241–247.

Kaipio, J. and Somersalo, E. (2005), *Statistical and Computational Inverse Problems*. Springer, Berlin.

Lounici, K. and Nickl, R. (2010), *Global uniform risk bounds for wavelet deconvolution estimators*. Preprint, University of Cambridge.

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

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. R. Statist. Soc. B* **67**, pp. 19–41.

Neumann, M. H. (1998), Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26**, 2014–2048.

Neumann, M. H. and Polzehl, J. (1998), Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Statist.*, **9**, 307–333.

Pensky, M. and Vidakovic, B. (1999), Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27**, 2033–2053.

Xia, Y. (1998), Bias-corrected confidence bands in nonparametric regression. *J. R. Statist. Soc. B*, **60**, 797–811.

7 Appendix: Proofs

Proof of Theorem 1: Following Bickel and Rosenblatt (1973) and Eubank and Speckman (1993), our proof is based on an approximation of $Z_n^{(j)}$ by a Gaussian process which does not depend on the true regression function θ. We shall use the following strong approximation result for sums of i.i.d. random variables.

Lemma 4. *(Csörgo and Revesz, 1981) There exists a Wiener process W_1 on $[0,\infty)$ such that*

$$|S_n - W_1(n)| = O(\delta_n) \text{ a.s.},$$

where $\delta_n := (n\log\log(n))^{1/4}(\log(n))^{1/2}$, $S_n = \sum_{j=1}^n \varepsilon_j$ and $\varepsilon_1, \varepsilon_2, \ldots$ i.i.d. with $E[\varepsilon_j] = 0$, $E[\varepsilon_j^2] = 1$ and $E[\varepsilon_j^4] < \infty$.

To keep the proof more transparent we split the approximation of the process $Z_n^{(j)}(t)$ into several steps, assume $\sigma^2 = 1$, and consider only the observations $r = 1, \ldots, n$. The desired results then immediately follow from repeating the same steps for the observations $r = -n, \ldots, 0$. Note that

$$Z_n^{(j)}(x) = n^{1/2}h^{\beta+1/2}a_n^{1/2}\sum_{r=1}^n \frac{1}{nha_n}\varepsilon_r K^{(j)}\left(\frac{x-z_r}{h}; h\right)$$

and let

$$Z_{n,1}^{(j)}(x) = \frac{n^{-3/2}h^{\beta-3/2}}{a_n^{3/2}}\sum_{r=1}^n K^{(j+1)}\left(\frac{x-z_r}{h}; h\right)W_1(r) + n^{-1/2}h^{\beta-1/2}a_n^{-1/2}K^{(j)}\left(\frac{x-z_n}{h}; h\right)W_1(n).$$

Lemma 5. *Under Assumptions 1 and 2.A*

$$\|Z_n^{(j)} - Z_{n,1}^{(j)}\|_{[0,1]} = o_p\left((\log(n))^{-1/2}\right).$$

Proof. Setting $S_0 = 0$, from a Taylor expansion we have with intermediate points $\xi_r \in [(x-z_r)/h, (x-z_{r+1})/h]$ that

$$\begin{aligned}
Z_n^{(j)}(x) &= n^{-1/2}h^{\beta-1/2}a_n^{-1/2}\sum_{r=1}^n K^{(j)}\left(\frac{x-z_r}{h}; h\right)(S_r - S_{r-1}) \\
&= n^{-1/2}h^{\beta-1/2}a_n^{-1/2}\left\{\sum_{r=1}^{n-1}\left(\frac{z_{r+1}-z_r}{h}\right)K^{(j+1)}\left(\frac{x-z_r}{h}; h\right)S_r\right. \\
&\quad \left. - \frac{1}{2}\sum_{j=1}^{n-1}\left(\frac{z_{r+1}-z_r}{h}\right)^2 K^{(j+2)}(\xi_r; h)\, S_r\right\} + n^{-1/2}h^{\beta-1/2}a_n^{-1/2}K^{(j)}\left(\frac{x-z_n}{h}; h\right)S_n \\
&= n^{-3/2}h^{\beta-3/2}a_n^{-3/2}\sum_{r=1}^{n-1}K^{(j+1)}\left(\frac{x-z_r}{h}; h\right)S_r - 2^{-1}\,n^{-5/2}h^{\beta-5/2}a_n^{-5/2}\sum_{r=1}^{n-1}K^{(j+2)}(\xi_r; h)\, S_r \\
&\quad + n^{-1/2}h^{\beta-1/2}a_n^{-1/2}K^{(j)}\left(\frac{x-z_n}{h}; h\right)S_n
\end{aligned}$$

Taking the difference of $Z_n^{(j)}(x)$ and $Z_{n,1}^{(j)}(x)$ we estimate

$$\begin{aligned}
|Z_n^{(j)}(x) - Z_{n,1}^{(j)}(x)| &= \left|n^{-3/2}h^{\beta-3/2}a_n^{-3/2}\sum_{r=1}^{n-1}K^{(j+1)}\left(\frac{x-z_r}{h}; h\right)(S_r - W_1(r))\right| \\
&+ \left|2^{-1}\,n^{-5/2}h^{\beta-5/2}a_n^{-5/2}\sum_{j=1}^{n-1}K^{(j+2)}(\xi_r; h)\, S_r\right| \\
&+ \left|n^{-1/2}h^{\beta-1/2}a_n^{-1/2}K^{(j)}\left(\frac{x-z_n}{h}; h\right)(S_n - W_1(n))\right|. \\
&= I + II + III.
\end{aligned}$$

Then

$$\begin{aligned} I &\leq n^{-3/2}h^{\beta-3/2}a_n^{-3/2}\max_{1\leq u\leq n}|S_u - W_1(u)|\sum_{r=1}^{n-1}\left|K^{(j+1)}\left(\frac{x-z_r}{h};h\right)\right| \\ &= O_p(\delta_n n^{-1/2}h^{\beta-3/2}a_n^{-1/2})\Big(\int_0^{1/a_n}\left|K^{(j+1)}\left(\frac{x-s}{h};h\right)\right|ds + O\left(h^{-\beta}(na_n)^{-1}\right)\Big) \\ &= O_p\Big((\log(\log(n)))^{1/4}(\log(n))^{1/2}\big(n^{-1/4}h^{-1/2}a_n^{-1/2} + n^{-5/4}h^{-3/2}a_n^{-3/2}\big)\Big), \end{aligned}$$

since by Assumption 2,

$$\int_0^{1/a_n}\left|K^{(j+1)}\left(\frac{x-s}{h};h\right)\right|ds = h\int_0^{1/(ha_n)}\left|K^{(j+1)}\left(\frac{x}{h}-s;h\right)\right|ds = O(h^{1-\beta})$$

and for every $j \geq 0$,

$$\left|h^\beta K^{(j)}(x)\right| = \left|\frac{h^\beta}{2\pi}\int_{\mathbb{R}}(-i\omega)^j e^{-i\omega x}\frac{\Phi_k(\omega)}{\Phi_\Psi(\omega/h)}d\omega\right| \leq \frac{1}{\pi C_\varepsilon}\int|\omega|^{j+\beta}|\Phi_k(\omega)|d\omega = C^* < \infty,$$

so that

$$|K^{(j)}(x;h)| = O(h^{-\beta}) \text{ uniformly in } x. \tag{12}$$

Further, we have

$$II = \frac{n^{-5/2}h^{\beta-5/2}a_n^{-5/2}}{2}\sum_{r=1}^{n-1}K^{(j+2)}(\xi_r)S_r = O_p(n^{-1}h^{-5/2}a_n^{-5/2}),$$

by using (12) and

$$E\sum_{r=1}^{n}|S_r| \leq \sum_{r=1}^{n}\sqrt{\mathrm{Var}(S_r)} = \sum_{r=1}^{n}\sqrt{r} = O\left(n^{3/2}\right).$$

Finally,

$$\begin{aligned} III &= O\Big(n^{-1/2}h^{-1/2}a_n^{-1/2}|S_n - W_1(n)|\Big) \\ &= O_P\left(n^{-1/4}h^{-1/2}(\log(\log(n)))^{1/4}(\log(n))^{1/2}a_n^{-1/2}\right). \end{aligned}$$

□

We further introduce the processes

$$\begin{aligned} Z_{n,2}^{(j)}(x) &= h^{\beta-1/2}\int_0^{\frac{1}{a_n}}K^{(j)}\left(\frac{x-s}{h};h\right)dW(s), \\ Z_{n,4}^{(j)}(x) &= h^{-1/2}\int_0^{\infty}K^{(j)}\left(\frac{x-s}{h}\right)dW(s), \end{aligned}$$

Lemma 6. *Under Assumptions 1 and 2,*

$$Z_{n,1}(x) \stackrel{d}{=} Z_{n,2}(x) + O_p\left((h\,a_n)^{-3/2}\sqrt{\log(n)}/(\sqrt{n})\right).$$

Proof. We have

$$\begin{aligned} Z_{n,1}^{(j)}(x) &\stackrel{d}{=} n^{-1}a_n^{-3/2}h^{\beta-3/2}\sum_{r=1}^{n} K^{(j+1)}\left(\frac{x-z_r}{h};h\right)W\left(\frac{r}{n}\right) \\ &\quad +h^{\beta-1/2}a_n^{1/2}\,K^{(j)}\left(\frac{x-z_n}{h};h\right)W\left(\frac{1}{n}\right) \end{aligned} \tag{13}$$

For the first term on the right in (13), using the modulus of continuity of Brownian motion on $[0,1]$ and (12), we get

$$\begin{aligned} & n^{-1}a_n^{-3/2}h^{\beta-3/2}\sum_{r=1}^{n} K^{(j+1)}\left(\frac{x-z_r}{h};h\right)W\left(\frac{r}{n}\right) \\ = \;& h^{\beta-3/2}a_n^{-3/2}\int_0^1 K^{(j+1)}\left(\frac{x-u/a_n}{h};h\right)W(u)\,du + O_p\left((h\,a_n)^{-3/2}\sqrt{\log(n)}/(\sqrt{n})\right). \end{aligned} \tag{14}$$

Further, for the integral in (14) we compute

$$\begin{aligned} & h^{\beta-3/2}a_n^{-3/2}\int_0^1 K^{(j+1)}\left(\frac{x-u/a_n}{h};h\right)W(u)\,du \\ \stackrel{d}{=} \;& h^{\beta-3/2}a_n^{-1}\int_0^1 K^{(j+1)}\left(\frac{x-u/a_n}{h};h\right)W(u/a_n))\,du \\ = \;& h^{\beta-3/2}\int_0^{1/a_n} K^{(j+1)}\left(\frac{x-u}{h};h\right)W(u)\,du \end{aligned}$$

and recollecting the second term on the right in (13) and changing scale as well to $W(1/a_n)$, we get

$$\begin{aligned} & h^{\beta-3/2}\int_0^{1/a_n} K^{(j+1)}\left(\frac{x-u}{h};h\right)W(u)\,du + h^{\beta-1/2}K^{(j)}\left(\frac{x-z_n}{h};h\right)W\left(\frac{1}{a_n}\right) \\ = \;& h^{\beta-1/2}\int_0^{1/a_n} K^{(j)}\left(\frac{x-s}{h};h\right)dW(s) = Z_{n,2}^{(j)}(x). \end{aligned}$$

which together with the remainder estimate in (14) yields the lemma. □

Using Assumptions 2 B., C. and the law of the iterated logarithm, one shows that

Lemma 7. *Under Assumptions 2 B., C.,*

$$\|Z_{n,2}^{(j)} - Z_{n,4}^{(j)}\|_{[0,1]} = o_P((\log(n))^{-1/2}).$$

Proof of theorem 1: The theorem now follows from Lemmas 5-7 and an application of Theorem/Corollary A.1 in Bickel and Rosenblatt (1993) to the process $Z_{n,4}^{(j)}(x)$. □