Confidence bands for inverse regression models – with application to gel electrophoresis

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

[1]Fakultät für Mathematik

Ruhr-Universität Bochum, Germany

[2]Institut für Stochastik

Universität Karlsruhe, Germany

July 3, 2008

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 period 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. 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

[2]Address for correspondence: Dr. Hajo Holzmann, Universität Karlsruhe (TH), Institut für Stochastik, Englerstr. 2 D-76128 Karlsruhe, Germany, email: holzmann@stoch.uni-karlsruhe.de, Fon: +49/721/608 3272, Fax: +49/721/608 6066

1 Introduction

Suppose that at our disposal are observations (z_k, Y_k), $k = -n, \ldots, n$, 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 known function Ψ,

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

The recovery of θ from the data (z_k, Y_k) in model (1) is a statistical inverse problem (e.g. Mair & Ruymgaart, 1996) which is closely related to density deconvolution (e.g. Stefanski and Carroll, 1989; Fan, 1991; Delaigle & Gijbels, 2002). Note that in nonparametric deconvolution regression models, it is typically assumed (e.g. Cavalier & 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.

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 seperate 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 decide whether a specific adaptor protein binds to the wildtype of a neuronal receptor subunit but not to a mutant version of the receptor subunit.

From a technical point-of-view, an additional difficulty in model (1) is that the reconstruction of θ from $g = A\theta$ at any location x on the real line requires (asymptotically) information on

g on the full real line. Therefore, the design is chosen in the specific form with an additional sequence $a_n \to 0$, which ensures that the design points z_k will asymptotically exhaust $\mathbb{R}$. Further note that model (1) is closely related to nonparametric errors-in-variables regression model (cf. e.g. Fan and Truong, 1993, who determine convergence rates in a random design errors-in-variables model).

The purpose of this paper is to construct uniform confidence bands for the unknown function θ on a compact interval I, as well as to determine a uniform rate of convergence for certain estimators. 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 example, Neumann (1998) investigates bootstrap confidence bands for densities. In a regression context, Härdle (1989) constructed asymptotic confidence bands for M-smoothers. Eubank and Speckman (1993) for the Nadaraya-Watson estimator and Xia (1998) for local polynomial estimators, respectively, suggested confidence bands based on an explicit bias correction and not on undersmoothing. 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. Here, we shall extend these results to the case of inverse regression on the real line with convolution-type operators in model (1).

In the following we propose a kernel-type estimator in model (1). 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,

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

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

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

In Section 2 we introduce the basic assumptions and present our asymptotic results for constructing confidence bands. However, it is well-known that convergence in the resulting limit theorems is rather slow (Hall 1993). Therefore, we propose a bootstrap method based on the residual bootstrap in section 3. The performance of these 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. All proofs are deferred to an appendix.

2 Confidence bands for inverse regression

2.1 Regularity assumptions

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. To fix the notation, denote the Fourier transform of a function f by $\Phi_f(t) = \int_{\mathbb{R}} f(x) \exp(itx)\, dx$. 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 and the problem is 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 (1991) 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

$$\Phi_\Psi(\omega)\omega^\beta \to C_\epsilon, \qquad \omega \to \infty, \tag{3}$$

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 (3) 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

$$K^{(j)}(z;h) = \frac{1}{2\pi} \int_{\mathbb{R}} (-i\omega)^j e^{-i\omega z} \frac{\Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega, \quad 0 \leq j \leq p. \tag{4}$$

Now, if (3) holds, the deconvolution kernel $K^{(j)}(z;h)$ given in (4) 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

$$\begin{aligned} K^{(j)}(z) &= \frac{1}{2\pi C_\epsilon} \int_0^\infty (-i\omega)^j \exp(-i\omega z)\omega^\beta \Phi_k(\omega)\, d\omega \\ &\quad + \frac{1}{2\pi \overline{C_\epsilon}} \int_{-\infty}^0 (-i\omega)^j \exp(-i\omega z)|\omega|^\beta \Phi_k(\omega)\, d\omega, \end{aligned} \tag{5}$$

c.f. Bissantz *et al.* (2007). Note that the second term in (5) 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 (2), which is proportional to $\sigma^2/(nh^{2\beta+2j+1}a_n)$.

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 (5).

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 (5) 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 $g = K\theta$ satisfies

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

Assumption 2 B. is technical refinement of (3), indeed, the limit kernel $K^{(j)}$ is required to formulate it in the first place. For further discussion we refer to Bissantz et al. (2007).

2.2 Asymptotic confidence bands

In this section we 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 (2). 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

$$C_{K,1} = \frac{1}{2\pi|C_\epsilon|^2}\int_{\mathbb{R}} \omega^{2(\beta+j)}\Phi_k^2(\omega)\, d\omega, \quad C_{K,2} = \frac{\int_{\mathbb{R}} \omega^{2(\beta+j+1)}\Phi_k^2(\omega)\, d\omega}{\int_{\mathbb{R}} \omega^{2(\beta+j)}\Phi_k^2(\omega)\, d\omega}. \tag{6}$$

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

$$\max_{x\in[0,1]} |E\hat{\theta}_n^{(j)}(x) - \theta_n^{(j)}(x)| = O\left(h^{s-j-1} + h^{-(\beta+j+1)}a_n^r\right) \tag{7}$$

Note that in contrast to deconvolution density estimation, where the bias does not depend on the error density, the order in (7) 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.

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, $nh^{2(\beta+j)+1}a_n/\log(1/h) \to \infty$ 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 1. 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 $\log(1/h)\cdot(nh^{2(\beta+s)-1}a_n + nh^{-1}a_n^{1+2r}) \to 0$. 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 outside of some compact interval, or the interfering convolution function Ψ has exponential tails, such as the density of a Laplace distribution.

As a further corollary we obtain rates of uniform convergence of the estimator $\hat{\theta}^{(j)}$.

Corollary 3. *Let the assumptions of Theorem 1 be fulfilled. If additionally*

$$\frac{nh^{2\beta+2s-1}a_n}{\log h^{-1}} = O(1), \qquad \frac{nh^{-1}a_n^{2r+1}}{\log h^{-1}} = O(1), \tag{8}$$

then the estimator $\hat{\theta}^{(j)}$ has uniform convergence rate

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

3 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 N-N bootstrap 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.

In our indirect regression model with fixed design and homoscedastic errors, the standard choice is the residual bootstrap (Hall 1992). Therefore, in the following we propose a bootstrap procedure based on the residual bootstrap. Since the bootstrapping procedure requires the residuals, we shall concentrate on estimation of the function θ itself (and not its derivatives, for which additional estimation of θ would be required). Consider the residuals

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

where

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

and h has to be chosen for estimation of θ (and not its derivatives). When bootstrapping from the centered versions of the residuals $\tilde{\epsilon}_i$, due to boundary problems one typically does not use those residuals which are obtained at points z_i close to the boundary. Now, since under Assumption 1, the kernel k has compactly supported Fourier transform, it cannot itself have compact support. Nevertheless, typically k will be rapidly decreasing in the tails. Therefore, it is reasonable to use only those $\tilde{\epsilon}_i$ for which for some $\eta > 0$ with say $\eta > 2h$, $-a_n^{-1} + \eta \le z_i \le a_n^{-1} - \eta$, i.e. for the indices $-(n - \eta a_n n) \le i \le (n - \eta a_n n)$. For those i set

$$\hat{\epsilon}_i = \tilde{\epsilon}_i - \frac{1}{[2n(1 - \eta a_n)]} \sum_i \tilde{\epsilon}_i, \tag{9}$$

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 process $Z^{(j)}_{n,5}(x)$ is given by

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

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 $\sup_{x\in[0,1]} |Z^{(j),*}_n(x)|$, conditional on the original observations. The bootstrap confidence band for $E\hat{\theta}_n$ (and for θ in case of undersmoothing) is given by

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

Remark. A formal proof of the consistency of the residual bootstrap for constructing confidence bands remains an open problem. Neumann and Pohlzehl (1998) showed the consistency of the wild bootstrap for constructing confidence bands in (direct) nonparametric regression models. We suspect that a similar result for the wild bootstrap could be derived in our setting, however, simulations indicated that in our homoscedastic setup, the residual bootstrap outperforms the wild bootstrap. Therefore, we do not pursue this issue further.

4 Simulations

In this section we investigate the performance of the bootstrap confidence bands in a simulation study. In Sections 4.1 and 4.2, respectively, we discuss the simulation framework and the selection of a suitable bandwidth for the estimator on which the confidence bands are based. Then, in Section 4.3, we present the results of our simulation study, where we assume the convolution function ψ to be known, and finally, in Section 4.4, some results for the case of miss-specified convolution function ψ.

4.1 Simulation framework

We simulate from model (1), where the noise terms ε_k are i.i.d. centered normal with variance σ^2, and the design points are $z_k = \frac{k}{na_n}$, $k = -n, \ldots, n$ for certain samples sizes $2n+1$.
For the unknown regression function we take

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

Most of the region where these two functions significantly deviate from 0 is in the interval $[-4, 4]$, and we will hence use $a_n = 0.25$ in the subsequent simulations. In general, in an application, it is also recommendable to choose a_n such that the largest part of the region of the x-axis where the signal deviates significantly from 0 is captured, but to avoid an excessively small value a_n to avoid a large number of observations which are essentially just noise.
Moreover, in the main part of our simulations the convolution function is

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

where $\lambda = 3$. 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. eq. (9)).

4.2 Bandwidth selection

We now discuss the selection of the bandwidth h for estimator (2). It is well-known that the simulated coverage probabilities and confidence band areas are sensitive to a suitable selection of the bandwidth in the case of deconvolution problems (cf. Bissantz et al., 2007). Fig. 1 shows the simulated coverage probabilities and band areas for 90% nominal coverage probability in case of the Gaussian function of interest θ_1 (left) and the bimodal function θ_2 (right). In both cases the sample size is $2n + 1 = 201$, the standard deviation of the noise $\sigma = 0.1$ and $a_n = 0.25$, i.e. the observations sample the interval $[-4, 4]$. The effective coverage probability of the confidence band is significantly below the nominal coverage probability for bandwidths larger than approximately the 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 determing the largest bandwidth, for which the sup-distance between estimates for two subsequent bandwidth values is above a certain threshold value. This approach is

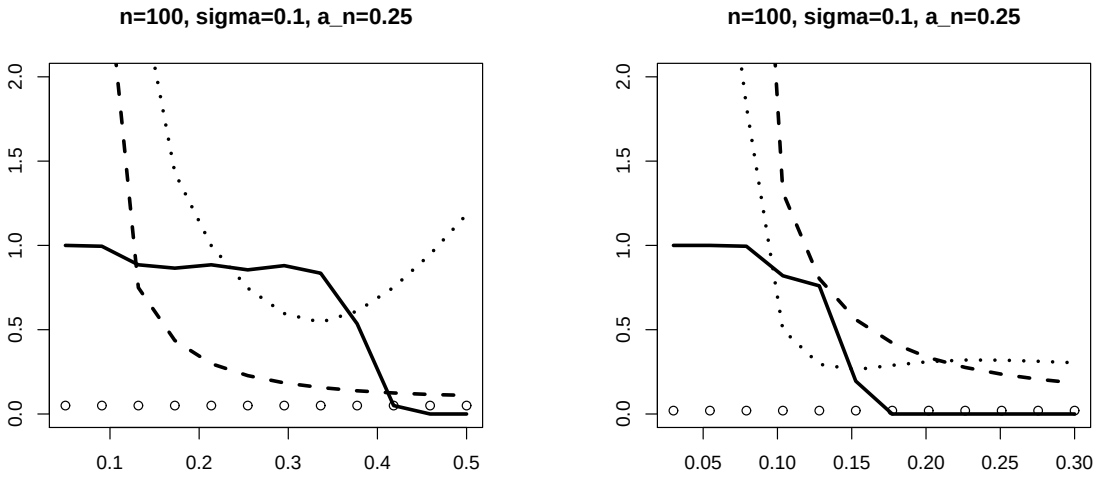

Figure 1: Average width and coverage probability of confidence bands with a nominal coverage probability of 90% for the Gaussian function θ_1 (left) and the bimodal function θ_2 (right). Solid lines represent simulated coverage probabilities from 200 simulations, dashed lines $1.5\times$ the mean area 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.

based on the observation that estimates computed with the spectral estimator (2) exhibit strongly increasing oscillations for bandwidths even only moderately below the L_∞-optimal bandwidth. 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.

Finally, Fig. 2 shows 90% nominal coverage probability confidence bands for the estimates $\hat{\theta}(x)$ of the Gaussian and the bimodal functions of interest θ_1 and θ_2, respectively, from $2n + 1 = 201$ observations based on data with $\sigma = 0.1$ and $a_n = 0.25$ and the bandwidth selected by the L_∞-optimal bandwidth selector. Note that the confidence bands are only valid for the interval $[0, 1]$, but have been continued throughout the plot.

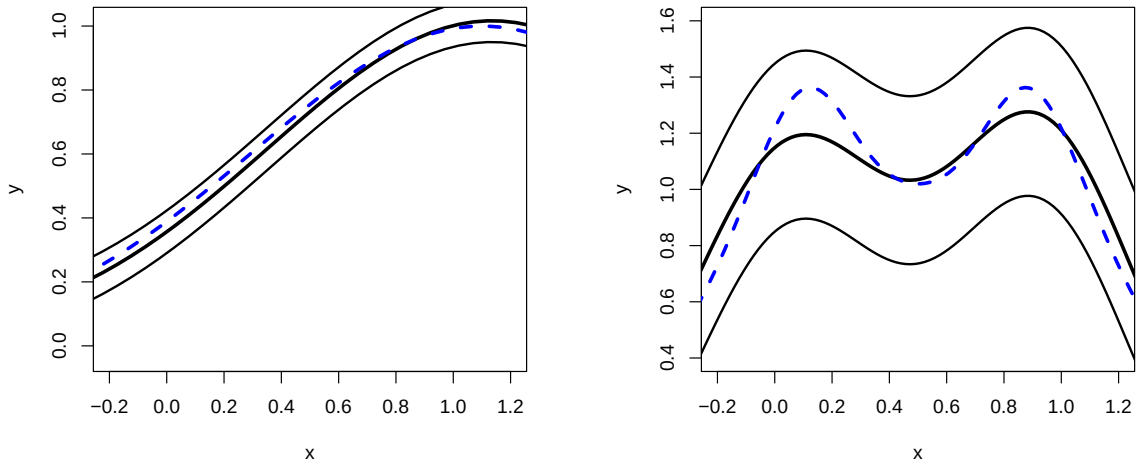

Figure 2: Estimate $\hat{\theta}_n(x)$ and associated 90% nominal coverage probability residual bootstrap confidence bands (solid lines) for the Gaussian function θ_1 (left) and the bimodal function θ_2 (right). Dashed lines represent the true functions θ_1 and θ_2, respectively.

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

In this Section we discuss the results of the main part of our simulation study, where we assume the convolution function ψ to be known as required for our asymptotic results in Section 2.2. For each combination of the parameters n, σ and regression functions θ_1, θ_2 we first determined a suitable bandwidth h for estimator (2) from the L_∞-bandwidth estimator outlined in the preceeding Section and detailed in Bissantz et al. (2007). Table 1 shows the results for simulations with the unimodal function 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 $g = A\theta$

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.

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.

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 $a_n = 0.25$.

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

4.4 Robustness and missspecification 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 consid-

ered 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. These are usually carried out to separate dna, rna or protein molecules according to their molecular weight and charge in the electrical field for analytical purposes or as preparative technique for the subsequent application of other techniques as for example mass spectrometry or PCR. A short summary of gel electrophoresis is as follows. A sample containing the molecules of interest in solution is applied to a plane gel of polyacrylamide and exposed to an electric field along the gel that drives migration of the molecules through pores in the gel. Thereby, small low weight molecules move faster through the pores than large molecules. According to their charge and weight migrating molecules are focused as a band at a certain distance from the starting point in the gel.

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. A sample containing

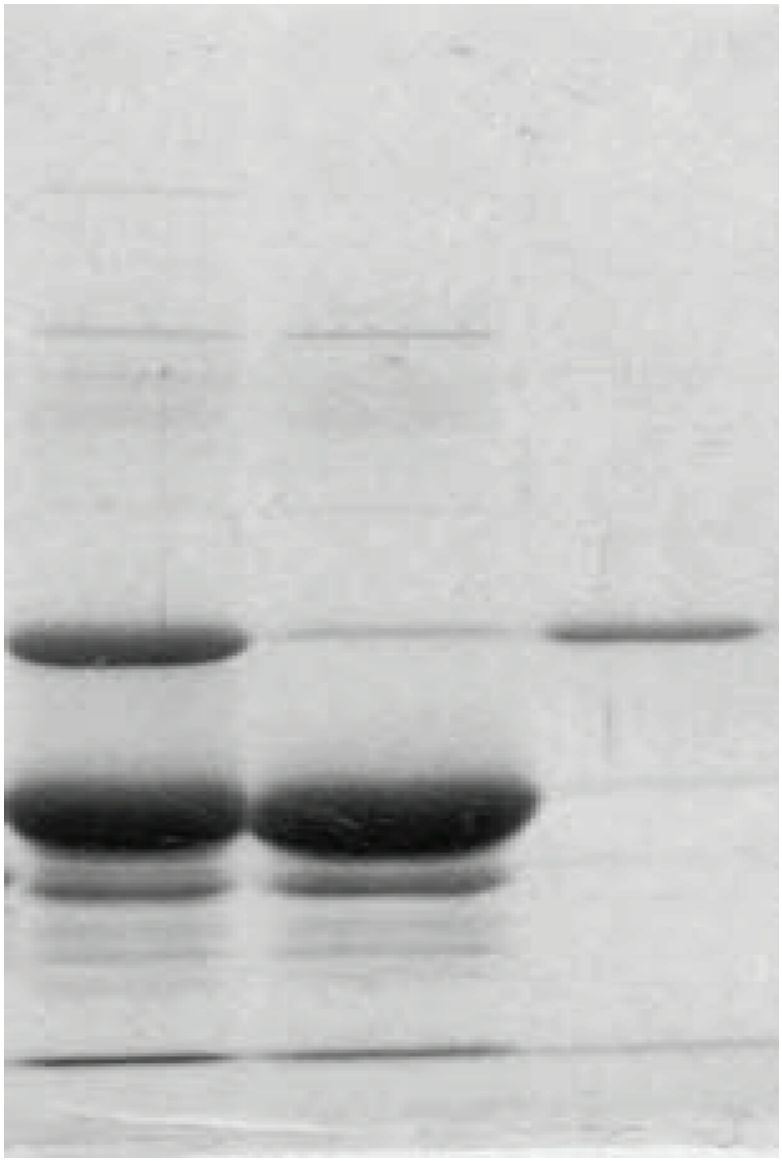

Figure 3: Result from a gel electrophoresis experiment with genetically engineered neuronal receptor subunits incubated with rat brain. Lanes are for wildtype receptor, mutante receptor and a standard molecule (from left to right).

the receptor and other protein bound to it in solution was applied on a 10% polyacrylamide gel and subjected to a electrical field of $1A/cm^2$ for 60 min. Negatively charged proteins moved from the starting point to the bottom of the gel. Thereby, the smaller receptor tail moves faster than the adaptor protein binding to it. The most intense band near the

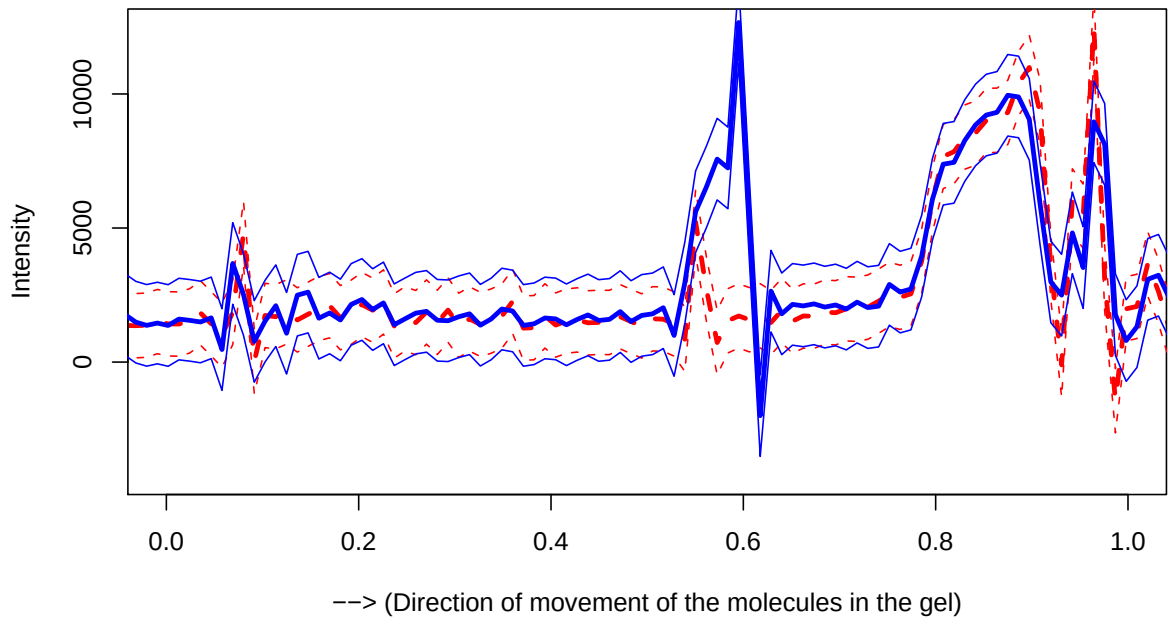

Figure 4: Intensity profiles and associated 90% bootstrap confidence bands for wildtype and mutante in the gel electrophoresis experiment discussed in the text. Solid lines show the distribution for the wildtype, and dashed lines for the mutante.

bottom of the gel depictures the receptor tail and the upper band is the binding adaptor protein. Intensity of the bands is according to the amount of protein in it. As one can see the wildtype receptor subunit binds a higher amount of adaptor protein than the mutated receptor subunit. Therefore, binding of the adaptor protein should occur specifically to a certain amino acid sequence in the wildtype receptor (left lane in Fig. 3) that was mutated in the middle lane. The weak band above the mutant receptor 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 right lane shows the band produced by a standard molecule of fixed weight. For our subsequent analysis we model the line-spread function $\Psi(x)$, which models the way a protein band is widended 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 (2). 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. A small simulation study with the bimodal test function used in

Section 4 was performed to validate the bootstrap confidence bands with regularization based on the ν-method numerically; indeed they are somewhat anticonservative with simulated coverage probabilities of 68%, 79% and 85% for nominal coverage probabilities 80%, 90% and 95%, respectively.

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 an amino acid sequence present in the wildtype receptor subunit but not in 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. We also discuss an application to a gel electrophoresis experiment.

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. Furthermore, the problem of (non-) adaptivity of confidence intervals or bands in nonparametric estimation problems (cf. Genovese and Wasserman 2008) will certainly apply to our indirect estimation problem as well. These can be overcome under certain shape restrictions (Dümbgen 2003). Therefore, the introduction of shape-restrictions in inverse problems seems to be particularly promising.

Acknowledgements

Melanie Birke and Nicolai Bissantz gratefully acknowledge support of the SFB 475 and from the BMBF project INVERS. Hajo Holzmann acknowledges financial support from the Claussen-Simon-Stiftung, and from the Landesstiftung Baden-Württemberg, "Juniorprofessorenprogramm". The authors would like to thank Holger Dette and Axel Munk for helpful discussions and Kathrin Bissantz for providing the biological data and for helping to analyse it.

References

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

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.

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

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

Genovese, C. and Wasserman, L. (2008), Adaptive confidence bands. *Ann. Statist*, to appear.

Härdle, W. (1989), Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.*, **29**, 163–179.

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.

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

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.

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

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*

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

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

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

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

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

Then

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

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 $j\ge 0$,

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

so that

$$|K^{(j)}(x;h)| = O(h^{-\beta}) \text{ uniformly in } x. \tag{10}$$

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 (10) 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,

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

□

We further introduce the processes

$$\begin{aligned} Z_{n,2}^{(j)}(x) &= h^{\beta-1/2}\int_0^{\frac{1}{a_n}} K^{(j)}\left(\frac{x-s}{h};h\right)dW(s), \\ Z_{n,3}^{(j)}(x) &= h^{-1/2}\int_0^{1/a_n} K^{(j)}\left(\frac{x-s}{h}\right)dW(s), \\ Z_{n,4}^{(j)}(x) &= h^{-1/2}\int_0^{\infty} K^{(j)}\left(\frac{x-s}{h}\right)dW(s), \end{aligned}$$

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

$$\begin{aligned} Z_{n,1}^{(j)}(x) \stackrel{d}{=}\ & n^{-1}a_n^{-3/2}h^{\beta-3/2}\sum_{r=1}^{n} K^{(j+1)}\left(\frac{x-z_r}{h};h\right)W\left(\frac{r}{n}\right) \\ & + h^{\beta-1/2}a_n^{1/2}\,K^{(j)}\left(\frac{x-z_n}{h};h\right)W\left(\frac{1}{n}\right) \end{aligned} \tag{11}$$

For the first term on the right in (11), using the modulus of continuity of Brownian motion on $[0,1]$ and (10), we get

$$\begin{aligned} & n^{-1}a_n^{-3/2}h^{\beta-3/2}\sum_{r=1}^{n} K^{(j+1)}\left(\frac{x-z_r}{h};h\right)W\left(\frac{r}{n}\right) \\ = \ & h^{\beta-3/2}a_n^{-3/2}\int_0^1 K^{(j+1)}\left(\frac{x-u/a_n}{h};h\right)W(u)\,du + O_p\left((h\,a_n)^{-3/2}\sqrt{\log(n)}/(\sqrt{n})\right). \end{aligned} \tag{12}$$

Further, for the integral in (12) 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 (11) 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 (12) yields the lemma. □

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

$$\|Z_{n,2}^{(j)} - Z_{n,3}^{(j)}\|_{[0,1]} = o_P((\log(n))^{-1/2}).$$

Proof. Estimate

$$\begin{aligned} \left|Z_{n,2}^{(j)}(x) - Z_{n,3}^{(j)}(x)\right| &= h^{-1/2}\left|\int_0^{1/a_n}\left(h^{\beta}K^{(j)}\left(\frac{x-s}{h};h\right) - K^{(j)}\left(\frac{x-s}{h}\right)\right)dW(s)\right| \\ &\leq h^{-3/2}\int_0^{1/a_n}\left|h^{\beta}K^{(j+1)}\left(\frac{x-s}{h};h\right) - K^{(j+1)}\left(\frac{x-s}{h}\right)\right||W(s)|ds \\ &\quad + h^{-1/2}\left|h^{\beta}K^{(j)}\left(\frac{x-1/a_n}{h};h\right) - K^{(j)}\left(\frac{x-1/a_n}{h}\right)\right||W(1/a_n)| \\ &= I + II \end{aligned}$$

For I we have that

$$\begin{aligned} I &\leq h^{-1/2}\int_{\mathbb{R}}\left|h^{\beta}K^{(j+1)}(u;h) - K^{(j+1)}(u)\right||W(x-hu)|du \\ &= O_P(h^{\delta}), \end{aligned}$$

from Assumption 2 B. and the law of the iterated logarithm for Brownian motion. For II we have from Assumption 2 C. and the law of the iterated logarithm for Brownian motion

$$II \leq O_P(h^{\delta}a_n^{-1/2}|\log\log a_n|^{1/2}).$$

□

Lemma 8. *Under Assumption 2.B*

$$\|Z_{n,3}^{(j)} - Z_{n,4}^{(j)}\|_{[0,1]} = o_p\big((\log(n))^{-1/2}\big).$$

Proof.

$$\begin{aligned}
\left|Z_{n,4}^{(j)}(x) - Z_{n,3}^{(j)}(x)\right| &= \left|h^{-1/2}\int_{1/a_n}^{\infty} K^{(j)}\left(\frac{x-s}{h}\right) dW(s)\right| \\
&\leq h^{-1/2}\int_{-\infty}^{(x-a_n^{-1})/h} \left|K^{(j+1)}(u)\, W(x-hu)\right| du \\
&\quad + \left|h^{-1/2}K^{(j)}\left(\frac{x-1/a_n}{h}\right) W(1/a_n)\right| \\
&= O_P\Big(h^{-1/2}\exp(-\frac{1}{a_n h})\Big).
\end{aligned}$$

by Assumption 2 D. and yet another application of the law of the iterated logarithm for the Wiener process. □

Proof of theorem 1: The theorem now follows from Lemmas 5-8 and an application of Theorem/Corollary A.1 in Bickel & Rosenblatt (1993) to the process $Z_{n,4}^{(j)}(x)$. □

Proof of (7). Due to Assumption 3, θ and hence also g is continuously differentiable with bounded derivative. Thus,

$$\begin{aligned}
E\hat{\theta}_n^{(j)}(x) &= \frac{1}{2\pi n h^{j+1} a_n}\sum_{r=-n}^{n} g(z_r)\int_{-\infty}^{\infty}(-i\omega)^j e^{-\frac{i\omega(x-z_r)}{h}}\frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)}d\omega \\
&= \frac{1}{2\pi h^{j+1}}\int_{-\infty}^{\infty}(-i\omega)^j e^{\frac{-i\omega x}{h}}\frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)}\Big(A_n(\omega) + O\big((na_n)^{-1}\big)\Big)d\omega,
\end{aligned}$$

where

$$A_n(\omega) = \int_{-1/a_n}^{1/a_n} e^{\frac{i\omega y}{h}} g(y)dy = \Phi_\Psi\left(\frac{\omega}{h}\right)\Phi_\theta\left(\frac{\omega}{h}\right) - \int_{(-\infty,-1/a_n]\cup[1/a_n,\infty)} e^{i\omega y/h} g(y)dy.$$

Therefore,

$$\begin{aligned}
&\theta^{(j)}(x) - E\left(\hat{\theta}_n^{(j)}(x)\right) \\
= \;& \frac{1}{2\pi h^{j+1}}\int_{-\infty}^{\infty}(-i\omega)^j e^{\frac{-i\omega x}{h}}\frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)}\cdot\Big(\int_{-\infty}^{-1/a_n} e^{\frac{i\omega y}{h}} g(y)dy + \int_{1/a_n}^{\infty} e^{\frac{i\omega y}{h}} g(y)dy\Big)d\omega \\
&+ \frac{1}{2\pi h^{j+1}}\int_{-\infty}^{\infty}(-i\omega)^j e^{\frac{-i\omega x}{h}}(1-\Phi_k(\omega))\Phi_\theta(\omega/h)\, d\omega + O\left(n^{-1}h^{-(j+\beta+1)}a_n^{-1}\right) \\
= \;& I + II + O\left(n^{-1}h^{-(j+\beta+1)}a_n^{-1}\right),
\end{aligned}$$

since

$$\frac{1}{2\pi h^{j+1}}\int_{-\infty}^{\infty}|\omega|^j\frac{|\Phi_k(\omega)|}{|\Phi_\Psi(\omega/h)|}d\omega \leq \frac{c}{2\pi h^{j+1}}\int_{-1}^{1}\frac{|\omega|^{\beta+j}}{h^\beta}d\omega = O(h^{-(\beta+j+1)}),$$

using Assumption 1. From Assumption 3.A

$$|I| = o\left(h^{s-j-1}\right).$$

Moreover, from Assumption 3.B

$$\Big|\int_{-\infty}^{-1/a_n} e^{\frac{iy\omega}{h}}g(y)dy + \int_{1/a_n}^{\infty} e^{\frac{iy\omega}{h}}g(y)dy\Big| \leq \int_{(-\infty,-1/a_n]\cup[1/a_n,\infty)}|g(y)|dy = O((1/a_n)^{-r}).$$

so that

$$\begin{aligned}|II| &\leq \frac{1}{2\pi h^{j+1}}\Big|\int_{-\infty}^{\infty}(-i\omega)^j e^{\frac{-i\omega x}{h}}\frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)}\Big|\Big|\int_{-\infty}^{-1/a_n} e^{\frac{iy\omega}{h}}g(y)dy + \int_{1/a_n}^{\infty} e^{\frac{iy\omega}{h}}g(y)dy\Big|d\omega,\\ &= O\left(h^{-(\beta+j+1)}a_n^r\right), \qquad (13)\end{aligned}$$

Collecting all the terms, the bias can be estimated as $O\left(h^{s-j-1}+h^{-(\beta+j+1)}a_n^r\right)$, uniformly in x. □

Proof of Corollary 2: Consider the processes

$$\begin{aligned}Z_{n,5}^{(j)}(x) &= \frac{n^{1/2}h^{\beta+j+1/2}a_n^{1/2}}{\hat\sigma}\left(\hat\theta_n^{(j)}(x) - E[\hat\theta_n^{(j)}(x)]\right),\\ Z_{n,6}^{(j)}(x) &= \frac{n^{1/2}h^{\beta+j+1/2}a_n^{1/2}}{\hat\sigma}\left(\hat\theta_n^{(j)}(x) - \theta^{(j)}(x)\right).\end{aligned}$$

From Theorem 1, $\|Z_n^{(j)}\|_{[0,1]} = O_P\left((\log(1/h))^{1/2}\right)$. Since

$$Z_n^{(j)}(t) - Z_{n,5}^{(j)}(t) = \frac{\hat\sigma-\sigma}{\hat\sigma}Z_n^{(j)}(t),$$

we conclude that $\|Z_n^{(j)}(x) - Z_{n,5}^{(j)}\|_{[0,1]} = o_P\left((\log(1/h))^{-1/2}\right)$. Moreover, using (7), we have uniformly for $x\in[0,1]$,

$$\begin{aligned}|Z_{n,5}^{(j)}(x) - Z_{n,6}^{(j)}(x)| &= \frac{n^{1/2}h^{\beta+j+1/2}a_n^{1/2}}{\hat\sigma}|E[\hat\theta_n^{(j)}(x)] - \theta^{(j)}(x)|\\ &= O_P(n^{1/2}h^{\beta+s-1/2}a_n^{1/2} + n^{1/2}h^{-1/2}a_n^{1/2+r}).\end{aligned}$$

Therefore the conclusions of Theorem 1 also remain valid for the process $Z_{n,6}^{(j)}(x)$, and corollary 2 follows from rearranging the terms.

□

Proof of Corollary 3: We split the supremum in two parts,

$$\begin{aligned}\sup_{x\in[0,1]}|\hat{\theta}^{(j)}(x)-\theta^{(j)}(x)| &\leq \sup_{x\in[0,1]}|\hat{\theta}^{(j)}(x)-\mathrm{E}[\hat{\theta}^{(j)}(x)]|+\sup_{x\in[0,1]}|\,\mathrm{E}[\hat{\theta}^{(j)}(x)]-\theta^{(j)}(x)| \\ &= A_{n,j}+B_{n,j}\end{aligned}$$

and estimate both parts separately. Note, that with the notation of Theorem 1 and

$$T_{n,j}=(2\log h^{-1})^{1/2}(||Z_n^{(j)}||_{[0,1]}-d_n)$$

we have

$$A_{n,j}=\Big(\frac{T_{n,j}}{(2\log h^{-1})^{1/2}}+d_n\Big)\Big(\frac{C_{K,1}}{nh^{2\beta+2j+1}a_n}\Big)^{1/2}.$$

Theorem 1 yields the convergence in distribution of T_n. Therefore $T_{n,j}=O_P(1)$ (see e.g. Shao, Tu, 1996, p. 449) and with the definition of d_n

$$A_{n,j}=O_P\Big(\frac{1}{nh^{2\beta+2j+1}a_n\log h^{-1}}\Big)^{1/2}+O_P\Big(\frac{\log h^{-1}}{nh^{2\beta+2j+1}a_n}\Big)^{1/2}=O_P\Big(\frac{\log n}{nh^{2\beta+2j+1}a_n}\Big)^{1/2}.$$

Now, it remains to estimate $B_{n,j}$. The estimate of the rate of the bias derived in the proof of corollary 2 shows, that the bias can be estimated uniformly in x by

$$\mathrm{E}[\hat{\theta}^{(j)}(x)]-\theta^{(j)}(x)=O_P\Big(h^{s-j-1}+h^{-(\beta+j+1)}a_n^r\Big)$$

and hence, with assumption (8) we have

$$B_{n,j}=O_P\Big(\frac{\log h^{-1}}{nh^{2\beta+2j+1}a_n}\Big)^{1/2}$$

which finishes the proof of Corollary 3. □