Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Simultaneous bootstrap confidence bands in nonparametric regression

Michael H. Neumann & Jörg Polzehl

Nonparametric Statistics, Vol. 9, pp. 307–333
Reprints available directly from the publisher
Photocopying permitted by license only

SIMULTANEOUS BOOTSTRAP CONFIDENCE BANDS IN NONPARAMETRIC REGRESSION

MICHAEL H. NEUMANN* and JÖRG POLZEHL

Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin

(Received 8 August 1996; In final form 16 December 1996)

In the present paper we construct asymptotic confidence bands in non-parametric regression. Our assumptions cover unequal variances of the observations and nonuniform, possibly considerably clustered design. The confidence band is based on an undersmoothed local polynomial estimator. An appropriate quantile is obtained via the wild bootstrap. We derive certain rates (in the sample size n) for the error in coverage probability, which improves on existing results for methods that rely on the asymptotic distribution of the maximum of some Gaussian process. We propose a practicable rule for a data-dependent choice of the band-width. A small simulation study illustrates the possible gains by our approach over alternative frequently used methods.

Keywords: Nonparametric regression; confidence bands; bootstrap; local polynomial estimator; strong approximation

1991 Mathematics Subject Classification: Primary: 62G07; Secondary: 62G09, 62G15

1. INTRODUCTION

Whenever we have a nonparametric curve estimate, confidence bands are an important means to get an impression about the accuracy that can be expected for the particular estimator. Such bands seem to be much more informative than pointwise confidence intervals, when one has to decide if some feature of the estimated curve should be

*Corresponding author.

considered as structure of the unknown function or should be explained due to random fluctuations of the estimate.

There already exists a long list on previous attempts on this subject, most of them are mentioned in the bibliography in Eubank and Speckman (1993). Much work in this area was stimulated by a paper by Bickel and Rosenblatt (1973), who primarily derived confidence bands for kernel density estimators. Additionally, they provided a useful technical result on the distribution of the maximum of certain Gaussian processes, which are stationary after centering, and serve as limits of the deviation process of kernel estimators if the sample size tends to infinity.

In the random design model, Liero (1982) for the Nadaraya-Watson kernel estimator, Johnston (1982) for the Yang estimator and Härdle (1989) for M-smoothers established confidence bands based on the limiting distribution of the deviation process. There exist similar results by Major (1973) for histogram estimators, Révész (1979) and Bjerve, Doksum and Yandell (1985) for nearest neighbor estimators. All of these authors used undersmoothing to make the effect of bias negligible.

A different approach was used in Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988) and Sun and Loader (1994), who constructed conservative confidence bands without undersmoothing, but on the basis of the prior knowledge of upper bounds for the roughness of the regression m.

Bootstrap methods were used in this context by Härdle and Bowman (1988) for pointwise confidence intervals and Härdle and Marron (1991) for the construction of a fixed number of simultaneous error bars. Bootstrap techniques were also proposed by Faraway (1990) in regression with i.i.d. errors for bandwidth choice and construction of confidence intervals. However, there was no rigorous result proved for the performance of confidence bands. Hall (1993) investigated the bootstrap for getting confidence bands in density estimation more closely and obtained that it provides much better asymptotic results than the approach based on the asymptotic limit distribution. An interesting comparison of the small sample properties of various methods was made by Loader (1993).

The latest development on asymptotically correct confidence bands that came to our attention is the paper by Eubank and Speckman

(1993). These authors argued that methods which rely on undersmoothing are difficult to apply in practice, since there does not exist any natural guideline how to define an asymptotically undersmoothed bandwidth in a reasonable way for a fixed sample size n. Instead of pure undersmoothing they produced an estimator with asymptotically negligible bias by a two-step method due to adding a bias corrector to the initial estimator. It turns out that the estimators at both stages can be furnished with natural, MSE-optimal bandwidths, which makes the application of usual bandwidth selectors possible.

For specific equations of inference there often exist specific methods that can be more powerful than an inference based on confidence bands. For example, Raz (1990) proposes a test for the presence of a regression effect. Generally, results on confidence bands could also be reformulated in terms of tests to assess the presence of some regression effect. However, confidence bands are a general purpose technique which is intended for giving a visual impression on the adequacy and variability of the nonparametric estimate.

In the present paper we start with a fixed design model as Eubank and Speckman (1993) did, and we improve some of the shortcomings of that paper that were already noticed by these authors. In particular, we admit heteroscedastic errors and nonuniform design, which result in a considerably nonstationary process as limit of the deviation process of our estimator. In view of the possibly considerably irregular design we apply the local polynomial estimator, which was first used in nonparametric regression by Stone (1977). A good account of the literature in this field is given in Ruppert and Wand (1994). It was shown in Fan (1992) that local polynomial estimators share the advantages of the Nadaraya-Watson estimator and the Gasser-Müller estimator both for random and regular nonuniform design. An important improvement upon the method of Eubank and Speckman is, that we use undersmoothing rather than a subsequent bias correction to obtain an asymptotically centered pivotal quantity. For the closely related problem of pointwise confidence intervals it is well-known that, if both approaches exploit the same amount of smoothness of the curve, undersmoothing outperforms explicit bias correction. This was rigorously proved in Hall (1991b) for confidence intervals for a density, Hall (1992) for intervals in regression with i.i.d. errors and Neumann (1992) for regression with heteroscedastic errors.

Further, we also include the boundary region of the estimator, which can be quite large in practical applications with finite sample size.

We do not know if our confidence bands can be appropriately modified to employ exact asymptotic results as given in Bickel and Rosenblatt (1973) or Qualls and Watanabe (1972) for essentially stationary Gaussian processes to determine a proper quantile in our situation. To find an appropriate quantile for the error process we use the wild bootstrap, which was already implicitly contained in Wu (1986) and later applied to nonparametric regression by Härdle and Mammen (1993). In distinction to all of the above mentioned papers we are able to derive a rate of nearly $(nh)^{-1/2}$ for the decay of the error in coverage probability. In contrast, it was shown in Hall (1991a) that the approach using the asymptotic limit distribution leads to a much slower decay of the coverage error at the rate of $(\log n)^{-1}$, which seems to be a strong argument in favor of our new method.

We propose a practicable data-driven rule to determine the bandwidth for our undersmoothing estimator.

2. THE METHOD AND THE MAIN RESULT

Throughout this paper we consider the model

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where the errors ε_i are independent, but not necessarily identically distributed with $E\varepsilon_i = 0$, $E\varepsilon_i^2 = v(x_i)$, obeying

$$(A_E) \quad 0 < v_{\inf} \leq v(x_i) \leq v_{\sup} < \infty, \; E|\varepsilon_i|^M \leq C(M) < \infty \text{ for all } i, M.$$

For the nonrandom design points $x_i = x_i(n)$ we assume that there exist constants $0 < C_1 \leq C_2 < \infty$ with

$$(A_D) \quad \begin{aligned} C_1(n(b-a) - \log n) \leq \; & \#\{i | x_i \in [a, b)\} \\ & \leq C_2(n(b-a) + \log n) \quad \text{for all} \quad 0 \leq a < b \leq 1. \end{aligned}$$

We adopt (A_D) in our fixed design model rather than the frequently assumed "regular design", i.e. $\int_0^{x_i} f(t)dt = i/n$ for some probability density f, because it also includes cases with considerably more irregular, clustered designs. The following remark shows that also the often considered case of "random design" is covered by our assumption.

Remark 1 Assume that the design points x_i are realizations of i.i.d. random variables with density f supported on $[0, 1]$, $0 < \inf_{x\in[0,1]} f(x) \leq \sup_{x\in[0,1]} f(x) < \infty$. Then (A_D) is satisfied with a probability exceeding $1 - O(n^{-\lambda})$ for arbitrary λ and appropriately chosen C_1, C_2.

The proof of this and the other results is contained in Section 6. In the following we assume

$$(A_S) \qquad m \in C^r[0, 1].$$

We apply a p-th order local polynomial estimator $\hat{m}(x)$ of $m(x)$, which is given as $\hat{a}_0(x, Y_1, \ldots, Y_n)$, where $\hat{a} = (\hat{a}_0, \ldots, \hat{a}_p)'$ minimizes

$$M_x = \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big)\,(Y_i - a_0 - a_1(x - x_i) - \cdots - a_p(x - x_i)^p)^2. \tag{2.2}$$

This even provides a simple solution of the usual boundary problem. We assume that K is a continuous nonnegative function with $K(x) > 0$ iff $|x| < 1$. It is clear that

$$\hat{m}(x) = \sum w_j(x) Y_j = W'_x \underline{Y} = [(D'_x K_x D_x)^{-1} D'_x K_x \underline{Y}]_1, \tag{2.3}$$

where $\underline{Y} = (Y_1, \ldots, Y_n)'$,

$$D_x = \begin{pmatrix} 1 & \frac{x-x_1}{h} & \cdots & (\frac{x-x_1}{h})^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{x-x_n}{h} & \cdots & (\frac{x-x_n}{h})^p \end{pmatrix},$$

$$K_x = \mathrm{Diag}\Big[K\Big(\frac{x - x_1}{h}\Big), \ldots, K\Big(\frac{x - x_n}{h}\Big)\Big].$$

It can be shown by similar calculations as in Ruppert and Wand (1994) that

$$\operatorname{var}(\widehat{m}(x)) = O((nh)^{-1}) \tag{2.4}$$

and

$$E\widehat{m}(x) - m(x) = O(h^{\min\{r,p+1\}}) \tag{2.5}$$

hold uniformly in $x \in [0,1]$. Let $k = \min\{r, p+1\}$.

In the present paper we consider confidence bands of the form

$$I_x = [\widehat{m}(x) - t(x), \widehat{m}(x) + t(x)], \tag{2.6}$$

and we intend to determine such values of $t(x)$ that the property

$$P(m(x) \in I_x \quad \text{for all} \quad x \in [0,1]) \longrightarrow 1 - \alpha \tag{2.7}$$

is satisfied for some prescribed α, $0 < \alpha < 1$.

In the special case of i.i.d. errors ε_i Eubank and Speckman (1993) approximated the process $\{(\widehat{m}(x) - m(x))/\sqrt{\operatorname{var}(\widehat{m}(x))}\}_{x\in[0,1]}$ by some stationary Gaussian process and determined the asymptotic $(1-\alpha)$-quantile of the maximum of the absolute value of the latter process by a result of Bickel and Rosenblatt (1973). According to Hall (1991a), this yields a uniform confidence band with an error in coverage probability of order $O((\log n)^{-1})$.

In our considerable inhomogeneous situation due to unequal variances, nonequidistant design and the inclusion of the boundary region we do not know if one can use any available result on the maximum of the limiting process to get an analytic expression for an asymptotically correct $t(x)$. Therefore we use the simple idea of bootstrap. In avoiding the approximation step for the distribution of the maximal deviation of some Gaussian process we hope to get a better coverage accuracy for the confidence band. Because of the heteroscedastic errors, we apply the wild bootstrap proposed by Wu (1986) and applied to nonparametric regression by Härdle and Mammen (1993). Starting from the residuals

$$\hat{\varepsilon}_i = Y_i - \widehat{m}(x_i),$$

we draw, conditioned on $\underline{Y} = (Y_1, \ldots, Y_n)'$, independent random variables ε_i^* with zero mean, variances $\hat{\varepsilon}_i^2$ and appropriately bounded higher order moments. For simplicity we restrict ourselves to either

(i) $\varepsilon_i^* \sim N(0, \hat{\varepsilon}_i^2)$

or

(ii) $P(\varepsilon_i^* = -\hat{\varepsilon}_i) = P(\varepsilon_i^* = +\hat{\varepsilon}_i) = 1/2.$

Now we attempt to mimic the stochastic part $\widehat{m}_0(x) = \sum w_j(x)\varepsilon_j$ of the process $\{\widehat{m}(x) - m(x)\}_{x\in[0,1]}$ by

$$\widehat{m}_0^*(x) = \sum w_j(x)\varepsilon_j^*.$$

If we compare the cumulative distribution functions of two random variables, then we can expect that they are close to each other, if the difference between the random variables is small with a high probability. Because of the frequent use of this fact we formalize it by introducing the following notion.

Definition 2.1 Let $\{Y_n\}$ and $\{Z_n\}(Z_n \geq 0$ *a.s.*) be sequences of random variables, and let $\{\gamma_n\}$ be a sequence of positive reals. We write

$$Y_n = \tilde{O}(Z_n, \gamma_n),$$

if

$$P(|Y_n| > CZ_n) \leq C\gamma_n$$

holds for $n \geq 1$ and some $C < \infty$.

This notion differs obviously from the usual O_p, which would provide a similar property for $n \geq n_0$ and an arbitrary constant γ instead of $C\gamma_n$ on the right-hand side. As a rule, for arbitrary $\delta, \lambda > 0$ we can conclude under sufficiently strong moment conditions on the distributions of the errors by Markov's and Whittle's inequalities (for the latter see Whittle (1960)) that

$$(a_n)'\underline{\varepsilon} = \tilde{O}(n^\delta \|a_n\|, n^{-\lambda}) \tag{2.8}$$

and

$$\underline{\varepsilon}' A_n \underline{\varepsilon} - E\underline{\varepsilon}' A_n \underline{\varepsilon} = \tilde{O}(n^{\delta}\sqrt{\mathrm{tr}(A_n A_n')}, n^{-\lambda}) \tag{2.9}$$

hold uniformly over $a_n \in \mathbb{R}^{\alpha}$ and arbitrary $(n \times n)$-matrices A_n, where $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$. Furthermore, we obtain similar assertions for random quantities a_n and A_n, if they vary in classes with an algebraically bounded entropy; for details see Lemma A.1 in Neumann (1995).

The following lemma shows how $\tilde{O}$ can be used to prove the closeness of two random variables.

LEMMA 2.1 *Let $\{X_n\}$ be a sequence of random variables with $P(X_n \in [a,b]) \leq C((b-a)c_n + \gamma_{n1})$ for arbitrary a, b. Further, we assume $Y_n = \tilde{O}(\gamma_{n2}, \gamma_{n3})$. Then*

$$P(X_n + Y_n < t) = P(X_n < t) + O(\gamma_{n1} + c_n \gamma_{n2} + \gamma_{n3})$$

holds uniformly in $t \in (-\infty, \infty)$.

The proof of this lemma follows immediately from the inequalities

$$\begin{aligned} P(X_n < t - C\gamma_{n2}) - P(|Y_n| > C\gamma_{n2}) &\leq P(X_n + Y_n < t) \\ &\leq P(X_n < t + C\gamma_{n2}) + P(|Y_n|) > C\gamma_{n2}). \end{aligned}$$

Now we establish a quite useful result on the pathwise approximation of the regression process $\{\widehat{m}(x) - m(x)\}_{x\in[0,1]}$ by $\{\widehat{m}_0^*(x)\}_{x\in[0,1]}$. Note that thereby we relate the *conditional* distributions $\mathscr{L}(\varepsilon_i^*|\underline{Y})$ with the *full* distributions $\mathscr{L}(\varepsilon_i)$. Here and throughout the paper we use the convention that $\delta > 0$ denotes an arbitrarily small and $\lambda < \infty$ an arbitrarily large constant.

THEOREM 2.1 *Assume $(A_D), (A_E), (A_S)$. Let the ε_i^*'s be defined as above, conditioned on $\underline{Y}$. Then there exist versions of $\{\varepsilon_i\}$ and $\{\varepsilon_i^*\}$ on an appropriate joint probability space such that*

$$\sup_{x\in[0,1]} \{|\widehat{m}(x) - m(x)) - \widehat{m}_0^*(x)|\} = \tilde{O}(n^{\delta}(nh)^{-1} + h^k, n^{-\lambda}).$$

holds on a set $(\varepsilon_1, \ldots, \varepsilon_n) \in \Omega_0$, with $P(\Omega_0) \geq 1 - O(n^{-\lambda})$.

2.1. Confidence Bands of Uniform Size

Let t^*_α be the $(1-\alpha)$-quantile of the (random) distribution of the quantity

$$U^*_{n0} = \sup_{x \in [0,1]} \{|\widehat{m}^*_0(x)|\}, \tag{2.10}$$

which is introduced to mimic

$$U_n = \sup_{x \in [0,1]} \{|\widehat{m}(x) - m(x)|\}.$$

We already know from Theorem 2.1 that the process $\widehat{m}(x) - m(x)$ is pathwise close to the conditional (conditioned on $\underline{Y}$) process $\widehat{m}^*_0(x)$ on an appropriate probability space. The following lemma provides a lower bound for probabilities that $\sup_x\{|\widehat{m}^*_0(x)|\}$ falls into small intervals. These two results will finally lead to an estimate of the error in coverage probability of the proposed confidence bands.

LEMMA 2.2 *Assume* (A_E), (A_D) *and* (A_S). *Then*

$$P\left(\sup_{x \in [0,1]} \{|\widehat{m}^*_0(x)|\} \in [a,b]\right) = O((b-a)(nh)^{1/2}(\log n)^{1/2} + n^\delta; (nh)^{-1/2}).$$

The following theorem establishes an upper bound for the error in coverage probability of the uniform confidence band of size t^*_α around $\widehat{m}(x)$.

THEOREM 2.2 *Assume* (A_D), (A_E), (A_S). *Then*

$$\begin{aligned} &P(m(x) \in [\widehat{m}(x) - t^*_\alpha, \widehat{m}(x) + t^*_\alpha] \textit{ for all } x \in [0,1]) \\ &\qquad = 1 - \alpha + O(n^\delta (nh)^{-1/2} + (nh)^{1/2}(\log n)^{1/2} h^k). \end{aligned}$$

2.2. Confidence Bands of Variable Size

A reasonable alternative to the above approach are confidence bands, whose size is proportional to an estimate of the standard deviation of

$\widehat{m}(x)$. The residuals $\hat{\varepsilon}_i$ can be also used to estimate $v(x) = \text{var}(\widehat{m}(x))$ by

$$\hat{v}(x) = \sum_j w_j^2(x)\hat{\varepsilon}_j^2. \tag{2.11}$$

Let t_α^{**} be the $(1-\alpha)$-quantile of the (random) distribution of the quantity

$$T_{n0}^* = \sup_{x\in[0,1]} \left\{ |\widehat{m}_0^*(x)| / \sqrt{\hat{v}(x)} \right\}, \tag{2.12}$$

which mimics

$$T_n = \sup_{x\in[0,1]} \left\{ |\widehat{m}(x) - m(x)| / \sqrt{\hat{v}(x)} \right\}.$$

On first sight, the treatment of these quantities seems to be more involved because of the random denominator. However, the next lemma shows that $\hat{v}(x)$ approximates $v(x)$ uniformly well such that the considerations can be reduced to linear forms.

LEMMA 2.3 *Assume* (A_D), (A_E), (A_S). *Then*

$$\sup_{x\in[0,1]} \{|\hat{v}(x) - v(x)|\} = \tilde{O}(n^\delta (nh)^{-3/2}, n^{-\lambda}).$$

This lemma entails that $|\widehat{m}(x) - m(x)|/\sqrt{\hat{v}(x)}$ and $\widehat{m}_0^*(x)/\sqrt{\hat{v}(x)}$ can be well approximated by $|\widehat{m}(x) - m(x)|/\sqrt{v(x)}$ and $\widehat{m}_0^*(x)/\sqrt{v(x)}$, respectively, which yields the following assertion.

THEOREM 2.3 *Assume* (A_D), (A_E), (A_S). *Then*

$$P\Big(m(x) \in [\widehat{m}(x) - \sqrt{\hat{v}(x)}t_\alpha^{**} \widehat{m}(x) + \sqrt{\hat{v}(x)}t_\alpha^{**}] \quad \textit{for all} \quad x \in [0,1] \Big)$$
$$= 1 - \alpha + O\Big(n^\delta (nh)^{-1/2} + (nh)^{1/2}(\log n)^{1/2} h^k \Big).$$

It follows that the rate for the coverage probability is nearly optimized by the choice

$$h \asymp n^{-1/(k+1)}.$$

On the other hand, it is known for kernel estimators that all commonly used bandwidth selectors are designed to minimize the risk, usually the mean squared error, of the estimator. Such a bandwidth would be of order $n^{-1/(2k+1)}$ in our case, and their use would lead to a nonvanishing error in coverage probability. A practicable and heuristically motivated method to determine an appropriate bandwidth is discussed in the next section.

In view of Remark 1, for random design the assertion (A_D) of the theorem holds conditioned on $\underline{X} = (X_1, \ldots, X_n)'$ with a probability exceeding $1 - O(n^{-\lambda})$. Hence, the unconditioned error in coverage probability will be of the same order as given in the above theorem.

3. A PRACTICABLE RULE FOR THE BANDWIDTH CHOICE

The usual difficulty with undersmoothing in applications is, that all commonly used bandwidth selection techniques are closely connected to the optimization of the mean squared error of the estimator. It turns out that these methods balance bias and standard deviation in such a way that they tend to zero at the same rate. Hence, they are not immediately applicable for confidence bands.

To provide some motivation for our following proposal, we urge the reader, to compare first local polynomial estimators of different regularity. Every inclusion of an additional term in the local polynomials to be fitted could also be interpreted as a refinement of the former local polynomial estimator; see also Jones and Foster (1993), who mentioned the equivalence of local polynomial fitting and the use of higher order kernels. Keeping this idea in mind, we can choose h mean-squared-error optimal for some local polynomial estimator $\tilde{m}$ of lower regularity. For example, we could apply cross-validation to select a bandwidth $\tilde{h}_{cv}$.

As a further refinement, we may additionally adjust the variance of $\hat{m}$ to the magnitude of the variance of the local polynomial estimator $\tilde{m}$, of lower regularity by using a bandwidth $\hat{h} = C\tilde{h}_{cv}$. In case of a local quadratic $\hat{m}$, a local linear $\tilde{m}$ and Triweight Kernel this is achieved by setting $C = 1.9$, see Table I of Fan and Gijbels (1995) for a more detailed explanation.

Note that essentially the same idea was used by Eubank and Speckman (1993) for confidence bands based on bias correction.

Although the confidence band is centered around a bias corrected estimator of higher regularity than the initial estimator, the authors proposed to choose the bandwidth MSE-optimal for the latter one.

Now we turn to the effect of the randomness of such a data-driven bandwidth to the error in coverage probability. To get some feeling for this effect, we state first a simple lemma.

LEMMA 3.1 *Assume* (A_D), (A_E), (A_S), $h \asymp n^{-\gamma}$ *and* $\hat{h} - h = O_P(n^{-\mu})$. *Then*

$$\widehat{m}_{\hat{h}}(x) - \widehat{m}_h(x) = O_P(n^{\gamma-\mu}(n^{\delta}(nh)^{-1/2} + h^k)).$$

The more important question however is, whether our procedure remains consistent in the case of a randomly selected bandwidth. Of course, we could try to mimic this randomness also by the bootstrap. This seems to make the method even more involved, and the effect is also not immediately clear. The following proposition provides an upper bound for the coverage accuracy with random bandwidth $\hat{h}$.

PROPOSITION 3.1 *Assume* (A_D), (A_E), (A_S), $h \asymp n^{-\gamma}$ *and* $P(|\hat{h} - h| \geq Cn^{-\mu}) \leq Cn^{-\lambda}$. *Then*

(i) $$P\big(m(x) \in [\widehat{m}_{\hat{h}}(x) - t_{\alpha}^{*}, \widehat{m}_{\hat{h}}(x) + t_{\alpha}^{*}] \text{ for all } x \in [0,1]\big)$$
$$= 1 - \alpha + O\Big(n^{\delta}(nh)^{-1/2} + (nh)^{1/2}(\log n)^{1/2}h^k$$
$$+ n^{\gamma-\mu}(n^{\delta} + (nh)^{1/2}(\log n)^{1/2}h^k) + n^{-\lambda}\Big),$$

(ii) $$P\Big(m(x) \in [\widehat{m}_{\hat{h}}(x) - \sqrt{\hat{v}(x)}t_{\alpha}^{**}, \widehat{m}_{\hat{h}}(x) + \sqrt{\hat{v}(x)}t_{\alpha}^{**}] \text{ for all } x \in [0,1]\Big)$$
$$= 1 - \alpha + O\Big(n^{\delta}(nh)^{-1/2} + (nh)^{1/2}(\log n)^{1/2}h^k$$
$$+ n^{\gamma-\mu}(n^{\delta} + (nh)^{1/2}(\log n)^{1/2}h^k) + n^{-\lambda}\Big).$$

In view of these results, each randomly chosen bandwidth $\hat{h}$ with $(\hat{h} - h)/h = \tilde{O}(n^{-\kappa}, n^{-\lambda})$ for some nonrandom bandwidth h and $\kappa > 0$ leads to a confidence band with asymptotically correct coverage probability.

4. SIMULATIONS

In this section we report details of a simulation study carried out to evaluate the performance of our bootstrap confidence bands. The simulation is based on model (2.1) with regression function

$$m(x) = \exp(-32(x - .5)^2). \quad (4.1)$$

This function was also used by Eubank and Speckman (1993) in their simulations. We restricted ourselves to sample sizes $n = 100$ and $n = 500$ and three variance structures, homogeneous variance $\sigma_i = 0.1$, weak heteroscedasticity $\sigma_i = 0.05 + .1m(x_i)$ and strong heteroscedasticity $\sigma_i = 0.01 + .2m(x_i)$.

We compared four confidence bands,

- the confidence band

$$I_x^T = [\widehat{m}_h(x) - \sqrt{\hat{v}(x)}t_\alpha^{**}, \widehat{m}_h(x) + \sqrt{\hat{v}_h(x)}t_\alpha^{**}], \quad (4.2)$$

- the confidence band of uniform size

$$I_x^U = [\widehat{m}_h(x) - t_\alpha^*, \widehat{m}_h(x) + t_\alpha^*] \quad (4.3)$$

- a modification of I_x^U for homoscedastic situations

$$I_x^W = [\widehat{m}_h(x) - \sqrt{w(x)}t_\alpha^{w*}, \widehat{m}_h(x) + \sqrt{w(x)}t_\alpha^{w*}]. \quad (4.4)$$

Here t_α^{w*} is the $(1 - \alpha)$-quantile of the distribution of

$$W_{n0}^* = \sup_{x\in[0,1]} \{|\widehat{m}_0^*(x)|/\sqrt{w(x)}\} \quad (4.5)$$

and $w(x) = \sum_j w_j^2(x)$ is a factor proportional to the variance of the local polynomial regression smoother at design point x.

- The confidence band proposed by Eubank and Speckman (1993)

$$I_x^{ES} = [\widehat{m}_{h_0}(x) - h_0^2 B\widehat{m}_{h_1}''(x) - l_\alpha^*, \widehat{m}_{h_0}(x) - h_0^2 B\widehat{m}_{h_1}''(x) + l_\alpha^*]. \quad (4.6)$$

In the case of the first three intervals $\widehat{m}_h$ is obtained using a local quadratic regression smoother with Triweight kernel $K(x) =$

$35/32(1-x^2)^3 I_{[-1,1]}$. In case of the confidence band proposed by Eubank and Speckman (1993) we used a Nadaraya-Watson kernel estimate with Epanechnikov kernel $K(x) = 3/4(1-x^2)^2 I_{[-1,1]}$ to estimate $m(x)$ and the second derivative of the Triweight kernel $K^*(x) = 105/16(-5x^4 + 6x^2 - 1)I_{[-1,1]}$ to estimate m'' which coincides with their choice for equidistant design. In order to avoid edge effects we used a simple edge correction (reflection at -1 and 1). The width l^*_α of the interval is given as

$$l^*_\alpha = \hat{\sigma} V_{1n}\left\{\sqrt{-2\log h_0} + \frac{1}{\sqrt{-2\log h_0}}\left[C - \log\frac{-\log(1-\alpha)}{2}\right]\right\}, \tag{4.7}$$

where $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n-2}(.809\,Y_i - .5\,Y_{i+1} - .309\,Y_{i+2})^2$ and

$$V_{1n}^2 = \sum_{|r| \le nh_1}\left[\frac{1}{nh_0}K\left(\frac{r}{nh_0}\right) - \frac{h_0^2 B}{nh_1^3}K^*\left(\frac{r}{nh_1}\right)\right]^2, \tag{4.8}$$

with $B = \int_{-1}^{1} x^2 K(x)dx$ and $h_1 = h_0^{5/7}$. The bandwidth h_0 was chosen by cross-validation. In case of the first three intervals we specified the bandwidth h as $h = 2.1\tilde{h}_{cv}$, with $\tilde{h}_{cv}$ the bandwidth minimizing the cross-validation criterion for the local linear regression smoother. The factor 2.1 was chosen to obtain similar variances for our local quadratic regression smoother and the bias corrected estimator of Eubank and Speckman. This leads to confidence bands of comparable size. The bandwidth is of order $O(n^{1/5})$ compared to the optimal order $O(n^{1/9})$ of the bandwidth of the local quadratic regression smoother used to estimate $m(x)$. Table I shows the results of the simulation.

In the case of homogeneous variance we observe a good compliance with the theoretical coverage probabilities for all four confidence bands. Results are best for I_x^W which takes homoscedasticity and variance structure of the estimate into account. The bands I_x^T and I_x^U behave slightly worse in the case of $n = 100$. This is mainly due to the variability of the variance estimate in the first case and the inhomogeneous variance of the estimate in the second case.

With increasing heteroscedasticity of the errors we still observe a good performance of the bootstrap bands in terms of coverage probability, while the coverage error of the Eubank/Speckman bands

TABLE I Coverage probabilities and mean area of confidence bands. Results obtained by 1000 simulations

		n = *100*				*n* = *100*			
σ	*level*	I_x^T	I_x^U	I_x^W	I_x^{ES}	I_x^T	I_x^U	I_x^W	I_x^{ES}
.1	.75	0.686	0.692	0.793	0.702	0.703	0.714	0.736	0.660
		(0.189)	(0.212)	(0.163)	(0.167)	(0.083)	(0.115)	(0.081)	(0.088)
	.9	0.854	0.832	0.936	0.876	0.861	0.866	0.889	0.853
		(0.240)	(0.252)	(0.193)	(0.196)	(0.097)	(0.145)	(0.092)	(0.101)
	.95	0.903	0.86	0.976	0.934	0.927	0.91	0.947	0.923
		(0.275)	(0.275)	(0.213)	(0.217)	(0.106)	(0.164)	(0.100)	(0.111)
	.975	0.936	0.892	0.981	0.970	0.962	0.938	0.969	0.961
		(0.304)	(0.294)	(0.231)	(0.237)	(0.115)	(0.179)	(0.107)	(0.120)
	.99	0.955	0.921	0.991	0.991	0.981	0.963	0.986	0.987
		(0.337)	(0.316)	(0.253)	(0.263)	(0.126)	(0.197)	(0.116)	(0.133)
$.05+.1m(x)$	.75	0.714	0.734	0.747	0.542	0.702	0.721	0.72	0.461
		(0.167)	(0.186)	(0.181)	(0.158)	(0.070)	(0.091)	(0.090)	(0.081)
	.9	0.854	0.899	0.911	0.709	0.888	0.879	0.888	0.604
		(0.209)	(0.229)	(0.228)	(0.184)	(0.081)	(0.109)	(0.110)	(0.094)
	.95	0.898	0.948	0.954	0.781	0.93	0.941	0.941	0.717
		(0.238)	(0.259)	(0.260)	(0.203)	(0.089)	(0.120)	(0.123)	(0.102)
	.975	0.933	0.966	0.97	0.861	0.968	0.968	0.968	0.801
		(0.264)	(0.286)	(0.290)	(0.222)	(0.097)	(0.131)	(0.135)	(0.111)
	.99	0.954	0.978	0.978	0.929	0.983	0.986	0.986	0.886
		(0.294)	(0.318)	(0.325)	(0.246)	(0.106)	(0.144)	(0.149)	(0.122)
$.01+.2m(x)$	.75	0.596	0.734	0.734	0.403	0.665	0.701	0.701	0.318
		(0.154)	(0.224)	(0.236)	(0.173)	(0.062)	(0.110)	(0.116)	(0.091)
	.9	0.797	0.901	0.901	0.551	0.829	0.879	0.878	0.446
		(0.189)	(0.296)	(0.312)	(0.202)	(0.072)	(0.137)	(0.144)	(0.104)
	.95	0.868	0.946	0.946	0.637	0.912	0.936	0.936	0.538
		(0.213)	(0.346)	(0.366)	(0.223)	(0.078)	(0.155)	(0.163)	(0.114)
	.975	0.920	0.962	0.962	0.709	0.95	0.963	0.963	0.627
		(0.236)	(0.393)	(0.416)	(0.243)	(0.085)	(0.172)	(0.181)	(0.124)
	.99	0.950	0.972	0.972	0.771	0.974	0.981	0.981	0.732
		(0.262)	(0.451)	(0.478)	(0.270)	(0.093)	(0.192)	(0.201)	(0.137)

increases significantly. The confidence bands I_x^T turn out best with respect to the area of the band, while not paying attention to the heterogeneous variance in the case of I_x^U and I_x^W results in larger areas of the confidence bands.

We also studied the effect of the bandwidth used in $\hat{m}$ choosing $h = C\tilde{h}_{cv}$ with $C = 1$ and $C = 1.4$. The results in terms of coverage probability are similar for $n = 100$ and better for $n = 500$, the effect being strongest for I_x^T and growing with heteroscedasticity. The areas of the confidence bands are increasing with decreasing h mainly reflecting the increasing variance of $\hat{m}$.

5. EXAMPLE: SUNLIGHT SCATTERING IN THE ATMOSPHERE

We illustrate our procedures using data from an experiment on sunlight scattering in the atmosphere carried out by Bellver (1987). The data were analyzed by Cleveland (1993), are part of the Trellis library of Splus and are also available from the S-archive of *Statlib*. The relation of particulate concentration (variable concentration) and the angle (degrees) at which polarization of sunlight vanishes (babinet point) is investigated.

We restrict out interest to the interval [16,85] for variable concentration, not considering the border regions where the design is sparse. We choose the bandwidth for the interval of interest according to our proposal in Section 3, i.e. selecting a bandwidth $\tilde{h}_{cv}$ for a local linear smoother by cross-validation and using the bandwidth $\hat{h} = C\tilde{h}_{cv}$ for the local quadratic estimate. The constant C is chosen to give the same asymptotic variance for both the local quadratic and local linear smoothers.

We decided to use a global bandwidth because data-driven global bandwidths usually converge with faster rates to a nonrandom bandwidth than local ones. Since we do not mimic the bandwidth selection step by the bootstrap, we think that this is more reliable. Moreover, on the subinterval [16,85] a global bandwidth seems applicable.

Figure 1 shows confidence bands I_x^W and I_x^T with approximate coverage probability 9. Additionally we display the fit obtained by Cleveland (1993) using loess on the cubic root of concentration with smoothing parameter $span = 1/3$.

We see that the confidence band I_x^T is of similar shape but slightly widened. This results from the additional variability originating from estimation of $\sigma(x)$. As our theory and simulations show, both bands have approximately correct coverage probability, with I_x^T allowing for a shape depending on $\sigma(x)$. In our nearly homoscedastic example we would therefore clearly prefer I_x^W. We display I_x^T only for comparison.

Note that the local quadratic fit shown in Figure 1 is chosen to construct the confidence bands rather than to be an optimal point estimate of $m(x)$. Especially values of concentration larger than 85 would require much larger (local) bandwidths.

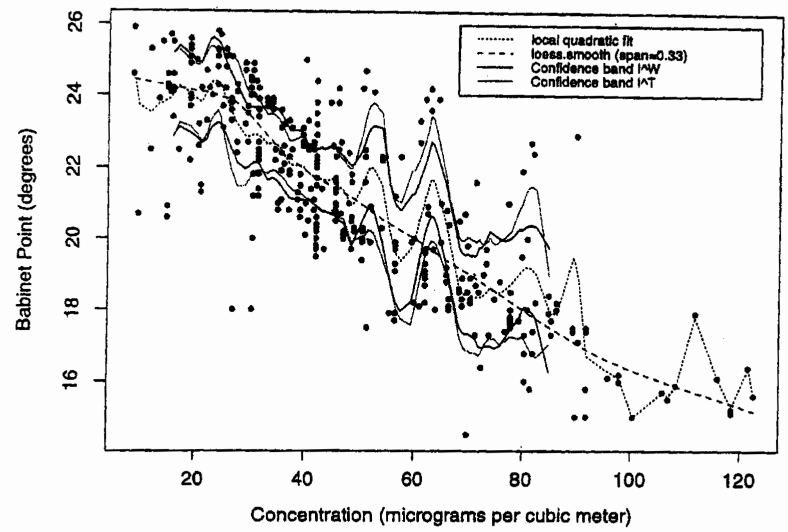

FIGURE 1 Sunlight scattering in the atmosphere, .9-confidence bands, local quadratic and loess fits.

Comparing lower and upper bounds of our confidence band I_x^W displayed in Figure 1, we see the presence of structure for concentration values around 60 that is not explained by the loess fit used in Cleveland (1993). Each curve fully covered by the confidence band shows at least a substantial change of slope in this region.

Although there may be more specific procedures to test for special effects, our confidence bands give an impression of the overall behavior of the regression curve by providing a region that contains $m(x)$ with a certain probability.

6. PROOFS

Proof of Remark 1 W.l.o.g. we prove this assertion for the simplest case $X_i \sim U[0, 1]$, i.e. $f \equiv 1$. The general case follows then immediately by the transformation $X_i = F^{-1}(Y_i)$, where F is the c.d.f. of X_i and $Y_i \sim U[0, 1]$ are independent random variables. Because of our assumption $0 < \inf f(x) \leq \sup f(x) < \infty$, we have that $0 < \inf\{\frac{d}{dx}F^{-1}(x)\} \leq \sup\{\frac{d}{dx}F^{-1}(x)\} < \infty$, which provides the assertion in the general case.

Let $G_n(t)=\sqrt{n}(F_n(t)-t)$, where $F_n(t)=n^{-1}\sum 1(\xi_i \leq t), \xi_1, \ldots, \xi_n \sim U[0,1]$ are independent. Applying Corollary 1 on p. 622 in Shorack and Wellner (1986) with $a = C(\log n)\, n^{-1}$, $b = \delta = 1/2$ and $\lambda = 3/\sqrt{2}\delta\sqrt{an}$, we obtain

$$P\left(\sup_{a \leq d-c \leq b} \frac{|G_n(d) - G_n(c)|}{\sqrt{d-c}} > \lambda\right) \leq \frac{24}{a\delta^3} \exp\left(-(1-\delta)^5 \frac{\lambda^2}{2}\right) = n^{-\lambda}, \tag{6.1}$$

if C is chosen sufficiently large.

Let now

$$|G_n(d) - G_n(c)| \leq \lambda\sqrt{d-c}.$$

We distinguish two cases.

If $d - c \geq a$, then

$$\begin{aligned}\left|\int_c^d dF_n - \int_c^d dF\right| &= n^{-1/2}|G_n(d) - G_n(c)| \\ &\leq n^{-1/2}\lambda\sqrt{d-c} \\ &= O\left(\sqrt{a}\sqrt{d-c}\right) = O(d-c).\end{aligned}$$

If $d - c < a$, then

$$\int_c^d dF < a = O(n^{-1}\log n) \tag{6.2}$$

and

$$\begin{aligned}\int_c^d dF_n \leq \int_c^{c+a} dF_n &= n^{-1/2}(G_n(c+a) - G_n(c)) + \int_c^{c+a} dF \\ &= O(n^{-1}\log n),\end{aligned}$$

which completes the proof. □

Proof of Theorem 2.1 In order to prove the assertion we will introduce new random variables $\xi_j \sim N(0, \operatorname{var}(\varepsilon_j))$, as well as their analogues in the bootstrap world $\xi_j^* \sim N(0, \operatorname{var}(\varepsilon_j^*))$, whose relation to the ε_j's is described below. We split up

$$\begin{aligned}(\widehat{m}(x) - m(x)) - \widehat{m}_0^*(x) &= [E\widehat{m}(x) - m(x)] + \sum_j w_j(x)(\varepsilon_j - \xi_j) \\ &\quad + \sum_j w_j(x)(\xi_j - \xi_j^*) + \sum_j w_j(x)(\xi_j^* - \varepsilon_j^*) \\ &= R_1(x) + \cdots + R_4(x). \end{aligned} \tag{6.3}$$

By (2.5) we immediately get

$$\sup_{x\in[0,1]} \{|R_1(x)|\} = O(h^k). \tag{6.4}$$

To estimate the terms $R_2(x)$ and $R_4(x)$ we will exploit well-known strong approximation results for partial sums of the ε_j's and ε_j^*'s, respectively. Further, we obtain an estimate for $R_3(x)$ by finding an appropriate pairing of the ξ_j's and ξ_j^*'s, which provides a close connection of the corresponding partial sum processes. This is done by relating the ξ_j's and the ξ_j^*'s to the same Wiener process. Accordingly, we will have three stochastic objects, $\{\varepsilon_j\}$, $\{\varepsilon_j^*\}$ and $W(t)$ which are appropriately connected on a sufficiently rich probability space.

Define the partial sum processes

$$S_j = \sum_{i\le j} \varepsilon_i \text{ and } \tilde{S}_j = \sum_{i\le j} \xi_i.$$

Then we have by Corollary 4 in Sakhanenko (1991, p. 76), that we can find a pairing of the ξ_j's and ξ_j^*'s, on a sufficiently rich probability space, such that

$$\max\nolimits_{1\le j\le n}\{|S_j - \tilde{S}_j|\} = \tilde{O}(n^\delta, n^{-\lambda}), \tag{6.5}$$

which implies by $w_j(x) - w_{j+1}(x) = O((nh)^{-2})$ that

$$\begin{aligned}\sup_x\{|R_2(x)|\} &= \sup_x\left\{\left|\sum w_j(x)(\varepsilon_j - \xi_j)\right|\right\} \\ &\le \sup_x\left\{\sum_{j=1}^{n-1} |w_j(x) - w_{j+1}(x)||S_j - \tilde{S}_j| + |w_n(x)||S_n - \tilde{S}_n|\right\} \\ &= \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda}). \end{aligned} \tag{6.6}$$

Next we intend to connect the ξ_j's with the ξ_j^*'s in such a way that the corresponding regression processes $\sum w_j(x)\xi_j$ and $\sum w_j(x)\xi_j^*$ are close to each other. We will get this linkage via a Wiener process $W(t)$, which is simultaneously used to define the ξ_j's and ε_j^*'s. From here on, all arguments for the ξ_j^*'s and ε_j^*'s are conditioned on $\underline{Y} = (Y_1, \ldots, Y_n)'$.

In contrast to the sequences $\{\varepsilon_j\}$ and $\{\xi_j\}$ of random variables, which have different distributions but matching variances, the sequences $\{\xi_j\}$ and $\{\xi_j^*\}$ consist of random variables from a convolution-invariant family but with different variances. The following heuristics suggests an appropriate cons-truction of the pairing of the ξ_j's and ξ_j^*'s. First, observe that it is more power-ful to connect two Gaussian random variables multiplicatively than additively: If $Z_1 \sim N(0, \sigma_1^2)$ and $Z_2 \sim N(0, \sigma_2^2)$, $\sigma_1 < \sigma_2$, then $\tilde{Z}_2 = (\sigma_2/\sigma_1)Z_1 \sim N(0, \sigma_2^2)$ is closer to Z_1 (it holds $\tilde{Z}_2 - Z_1 = O_P(\sigma_2 - \sigma_1)$) than $\widehat{Z}_2 = Z_1 + Z_3$ with $Z_3 \sim N(0, \sigma_2^2 - \sigma_1^2)$ independent of Z_1 (here $\widehat{Z}_2 - Z_1 = O_P(\sqrt{\sigma_2 - \sigma_1}\sqrt{\sigma_2 + \sigma_1})$).

This suggests to use the same stretch of $W(t)$ for defining both ξ_j and ξ_j^*. On the other hand, we are primarily interested in a close connection of $\sum w_j(x)\xi_j$ and $\sum w_j(x)\xi_j^*$. The (random) variances $\hat{\varepsilon}_j^2$ of ξ_j^* are different from but nearly unbiased for v_j. This indicates the possibility to get a closer link of the two regression processes by an appropriate connection of certain sums of the ξ_j's and ξ_j^*'s To make an optimal compromise between these two conflicting heuristics possible, we decompose the error vectors $\underline{\xi} = (\xi_1, \ldots, \xi_n)'$ and $\underline{\xi}^* = (\xi_1^*, \ldots, \xi_n^*)'$ into $\Delta \asymp h^{-1}$ packages of length $d_j \asymp nh$, respectively, that is

$$\underline{\xi} = (\xi_{11}, \ldots, \xi_{1d_1}, \ldots, \xi_{\Delta 1}, \ldots, \xi_{\Delta d_\Delta})'.$$

($\underline{\xi}^*$ is splitted up analogously).

Let $v_{jk} = E\xi_{jk}^2$, $v_{jk}^* = E\xi_{jk}^{*2}$ and $w_{jk}(x) = w_l(x)$, if l corresponds to (j, k). Further, let $V_j = \sum_{k=1}^{d_j} v_{jk}$, $V_j^* = \sum_{k=1}^{d_j} v_{jk}^*$ $(j = 1, \ldots, \Delta)$. We define

$$t_{jk} = \sum_{l \le k} v_{jl}, \quad t_{jk}^* = \sum_{l \le k} v_{jl}^*,$$

$$s_{jk} = (j-1) + t_{jk}/V_j, \quad s_{jk}^* = (j-1) + t_{jk}^*/V_j^*.$$

By interpolation with Brownian bridges we build a Wiener process $W(t)$ such that

$$\xi_{jk} = V_j^{1/2}(W(s_{jk}) - W(s_{j,k-1})).$$

Now we define, conditioned on $\underline{Y}$, independent random variables $\xi_{jk}^* \sim N(0,\ v_{jk}^*)$ as

$$\xi_{jk}^* = V_j^{*1/2}(W(s_{jk}^*) - W(s_{j,k-1}^*)).$$

To see what this construction implies for the regression processes, we decompose $\sum_{j,k} w_{jk}(x)[\xi_{jk} - \xi_{jk}^*]$ in a "coarse structure" term

$$\Delta_1(x) = \sum_j (V_j^{1/2} - V_j^{*1/2}) \sum_k w_{jk}(x)(W(s_{jk}^*) - W(s_{j,k-1}^*))$$

and a "fine structure" term

$$\Delta_2(x) = \sum_j V_j^{1/2} \sum_k w_{jk}(x)[(W(s_{jk}) - W(s_{j,k-1})) - (W(s_{jk}^*) - W(s_{j,k-1}^*))].$$

We can easily show that

$$\max_{j,k}\{|t_{jk} - t_{jk}^*|\} = \tilde{O}(n^\delta (nh)^{1/2},\ n^{-\lambda}), \tag{6.7}$$

which implies $V_j \asymp V_j^* \asymp nh$ and

$$\max_j\{|V_j^{1/2} - V_j^{*1/2}|\} = \max_j \left\{ \frac{|V_j - V_j^*|}{V_j^{1/2} + V_j^{*1/2}} \right\} = \tilde{O}(n^\delta,\ n^{-\lambda}).$$

Therefore we have

$$\sup_x\{|\Delta_1(x)|\} = \tilde{O}(n^\delta (nh)^{-1},\ n^{-\lambda}). \tag{6.8}$$

We rewrite

$$\begin{aligned}\Delta_2(x) &= \sum_j V_j^{1/2} \sum_k w_{jk}(x) \left[\int_{s_{j,k-1}}^{s_{jk}} d\,W(t) - \int_{s_{j,k-1}^*}^{s_{jk}^*} d\,W(t) \right] \\ &= \sum_j V_j^{1/2} \int_{j-1}^{j} [w_t - w_t^*] d\,W(t),\end{aligned}$$

where $w_t = w_{j,k}(x)$, if $t \in (s_{j,k-1}, s_{jk}]$, and $w_t^* = w_{j,k}(x)$, if $t \in (s_{j,k-1}^*, s_{jk}^*]$.

By (6.7) and $w_{j,k}(x) - w_{j,k+1}(x) = O((nh)^{-2})$ we acquire $\sup_t\{|w_t - w_t^*|\} = \tilde{O}(n^\delta (nh)^{-3/2}, n^{-\lambda})$, which implies that

$$\Delta_2(x) = \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda}). \tag{6.9}$$

From (6.8) and (6.9) we obtain that

$$\sup_x\{|R_3(x)|\} = \sup_x\left\{|\sum_j w_j(x)(\xi_j - \xi_j^*)|\right\} = \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda}). \tag{6.10}$$

Finally, we can find such a pairing of the ξ_j^*'s with the ξ_j^*'s that

$$\sup_x\{|R_4(x)|\} = \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda}) \tag{6.11}$$

holds, which completes the proof. □

Proof of Lemma 2.2 We prove first

$$\begin{aligned} P&\left(\sup_{x\in[0,1]}\left\{|\sum_j w_j(x)\xi_j|\right\} \in [a,b]\right) \\ &= O((b-a)(nh)^{1/2}(\log n)^{1/2} + n^\delta (nh)^{-1/2}). \end{aligned} \tag{6.12}$$

First, we split the interval [0,1] into Δ subintervals, Δ even, $1/(4h) \le \Delta < 1/(2h)$.

Define

$$Z_i = \sup_{x\in\Delta_i}\left\{\sum w_j(x)\xi_j\right\}, \quad Z_i^- = \inf_{x\in\Delta_i}\left\{\sum w_j(x)\xi_j\right\},$$

where $\Delta_i = [(i-1)/\Delta, i/\Delta)$.

Let p_n, p_{n1}, p_{n1}^-, and p_{n2}^- denote the densities of $\sup_{x\in[0,1]}\{|\sum w_j(x)\xi_j|\}$, $\max_{i \text{ odd}}\{Z_i\}$, $\min_{i \text{ odd}}\{Z_i^-\}$, $\max_{i \text{ even}}\{Z_i\}$ and $\min_{i \text{ even}}\{Z_i^-\}$, respectively. (Their existence follows by Theorem 1 in Tsirel'son

(1975).) Because of $p_{nj}(t) = p_{nj}^-(-t)$, $j = 1, 2$, we have

$$p_n(t) \leq 2p_{n1}(t) + 2p_{n2}(t).$$

W.l.o.g. we derive an upper estimate for $p_{n1}(t)$.

Let $j \leq \Delta$ be any odd number and let $\underset{\sim}{\xi}_j = (\xi_{j1}, \ldots, \xi_{jd_j})'$ be the subvector of those random variables from $(\xi_1, \ldots, \xi_n)'$, which are needed to compute $\widehat{m}_0(x) = \sum_l w_l(x)\xi_l$ on the interval Δ_j.

Let $e_j = \tilde{e}_j / \|\tilde{e}_j\|$, $\tilde{e}_j = (v_{j1}^{-1/2}, \ldots, v_{jd_j}^{-1/2})'$ and $\Sigma_j = \operatorname{cov}(\underset{\sim}{\xi}_j)$. It is easy to see that $\underset{\sim}{\xi}_j$ can be disintegrated into independent summands $\Sigma_j^{1/2} e_j e_j' \Sigma_j^{-1/2} \underset{\sim}{\xi}_j = 1\|\tilde{e}_j\|^{-1} e_j' \Sigma_j^{-1/2} \underset{\sim}{\xi}_j$ and $\Sigma_j^{1/2}(I - e_j e_j')\Sigma_j^{-1/2}\underset{\sim}{\xi}_j$. We decompose $\widehat{m}_0(x)$ correspondingly as

$$\widehat{m}_0(x) = \widehat{m}_{01}(x) + \widehat{m}_{02}(x),$$

where, because of $\sum_k w_{jk}(x) = 1$ for all $x \in \Delta_j$,

$$\widehat{m}_{01}(x) = \sum_k w_{jk}(x) e_j' \Sigma_j^{-1/2} \underset{\sim}{\xi}_j = \|\tilde{e}_j\|^{-1} e_j' \Sigma_j^{-1/2} \underset{\sim}{\xi}_j \sim N(0, \|\tilde{e}_j\|^{-2}.)$$

Let $m_{j1} = \widehat{m}_{01}(x)$ for any $x \in \Delta_j$ and $m_{j2} = \sup_{x \in \Delta_j}\{\widehat{m}_{02}(x)\}$. Since $\widehat{m}(x)$ uses only observations Y_j with $|x - x_j| \leq h$, we get that $Z_1, \ldots, Z_{\Delta-1}$ are independent. It is clear that $(m_{11}, \ldots, m_{\Delta-1,1})$ is independent of $m_2^{\text{odd}} = (m_{12}, m_{32}, \ldots, m_{\Delta-1,2})$. Hence, we have for the conditional distribution of $Z = \max_{j \text{ odd}}\{Z_j\}$ that

$$\begin{aligned}
P(Z \geq t | m_2^{\text{odd}}) &= P(Z_1 \geq t | m_2^{\text{odd}}) + P(Z_1 < t,\, Z_3 \geq t | m_2^{\text{odd}}) + \cdots \\
&\quad + P(Z_1 < t, \ldots, Z_{\Delta-3} < t,\, Z_{\Delta-1} \geq t | m_2^{\text{odd}}) \\
&= P(m_{11} \geq t - m_{12}) + P(m_{11} < t - m_{12})\, P(m_{31} \geq t - m_{32}) \\
&\quad + \cdots + P(m_{11} < t - m_{12}) \cdots P(m_{\Delta-3,1} < t - m_{\Delta-3,2}) \\
&\quad \times P(m_{\Delta-1,1} \geq t - m_{\Delta-1,2}),
\end{aligned} \tag{6.13}$$

which implies for the conditional density of Z

$$\begin{aligned}
p_{Z|m_2^{\text{odd}}}(t) &= \frac{d}{dt}\{-P(Z \geq t | m_2^{\text{odd}})\} \\
&\leq p_{m_{11}}(t - m_{12}) + p_{m_{31}}(t - m_{32}) P(m_{11} < t - m_{12}) + \cdots \\
&\quad + p_{m_{\Delta-1,1}}(t - m_{\Delta-1,2}) P(m_{11} < t - m_{12}, \ldots, m_{\Delta-3,1} \\
&\quad < t - m_{\Delta-3,2}).
\end{aligned} \tag{6.14}$$

Since $m_{j1} \sim N(0, \|\tilde{e}_j\|^{-2})$, it is easy to see that

$$p_{m_{j1}}(s) \leq P(m_{j1} \geq s)\|\tilde{e}_j\|\left(C + \sqrt{c \log n}\right) + Cn^{-c/2},$$

which implies by (6.13) and (6.14)

$$\begin{aligned} P_{Z|m_2^{\text{odd}}}(t) &\leq P(Z \geq t|m_2^{\text{odd}}) \max_j \{\|\tilde{e}_j\|\} (C + \sqrt{c \log n}) + C\Delta n^{-c/2} \\ &= O\left((nh)^{1/2} \sqrt{\log n}\right). \end{aligned}$$

Integration over all possible realizations of m_2^{odd} provides an upper estimate for $p_{n1}(t)$. By analogous considerations for $p_{n2}(t)$ we obtain (6.12). The assertion of the lemma follows in conjunction with (6.10) and (6.11). □

Proof of Theorem 2.2 By Theorem 2.1 we obtain that

$$\begin{aligned} |U_n - U_{n0}^*| &\leq \sup_{x\in[0,1]} \{|(\widehat{m}(x) - m(x)) - \widehat{m}_0^*(x)|\} \\ &= \tilde{O}(n^{\delta}(nh)^{-1} + h^k, n^{-\lambda-1}) \end{aligned}$$

holds on an appropriate probability space, which yields (by Lemmas 2.1 and 2.2

$$\begin{aligned} \sup_t\{|P(U_n < t) - P(U_{n0}^* < t|\underline{Y})|\} = \tilde{O}(n^{\delta}(nh)^{-1/2} \\ + h^k(nh)^{1/2}(\log n)^{1/2}, n^{-\lambda}) \end{aligned} \tag{6.15}$$

uniformly in a certain set $\underline{Y} \in \Omega_0$ with $P(\bar{\Omega}_0) = O(n^{-\lambda})$. This implies in particular

$$\begin{aligned} P(U_n < t)|_{t=t_\alpha^*} &= P(U_{n0}^* < t_\alpha^*|\underline{Y}) + \tilde{O}(n^{\delta}(nh)^{-1/2} \\ &\quad + h^k(nh)^{1/2} (\log n)^{1/2}, n^{-\lambda}) \\ &= 1 - \alpha + \tilde{O}(n^{\delta}(nh)^{-1/2} \\ &\quad + h^k(nh)^{1/2} (\log n)^{1/2}, n^{-\lambda}), \end{aligned} \tag{6.16}$$

again for $\underline{Y} \in \Omega_0$. Integrating over t_α^* we obtain the assertion. □

Proof of Lemma 2.3 From $\hat{\varepsilon}_i^2 = \varepsilon_i^2 - 2\varepsilon_i(\widehat{m}(x_i) - m(x_i)) + (\widehat{m}(x_i) - m(x_i))^2$ we immediately acquire for fixed x that

$$\begin{aligned}\hat{v}(x) - v(x) &\leq \left|\sum_j w_j^2(x)[\varepsilon_j^2 - v_j]\right| \\ &\quad + \left|\sum_j w_j^2(x)[(\widehat{m}(x_i) - m(x_i))^2 - 2\varepsilon_i(\widehat{m}(x_i) - m(x_i))]\right| \\ &= \tilde{O}(n^\delta (nh)^{-3/2}, n^{-\lambda}).\end{aligned}$$

Making this approximation for $\hat{v}(x) - v(x)$ on a sufficiently fine grid on [0,1] we prove the assertion. □

Proof of Theorem 2.3 Using the approximations

$$\frac{\widehat{m}(x) - m(x)}{\sqrt{\hat{v}(x)}} = \frac{\widehat{m}(x) - m(x)}{\sqrt{v(x)}} + \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda})$$

and

$$\widehat{m}_0^*(x)/\sqrt{\hat{v}(x)} = \widehat{m}_0^*(x)/\sqrt{v(x)} + \tilde{O}(n^\delta (nh)^{-1}, n^{-\lambda})$$

we can prove this assertion analogously to Theorem 2.2. □

Proof of Lemma 3.1 From $w_j(x, \hat{h}) - w_j(x, h) = O((\hat{h} - h)h^{-1}(nh)^{-1}$ and Lemma A.1 in Neumann (1995) we get that

$$\begin{aligned}\sum (w_j(x, \hat{h}) - w_j(x, h))\varepsilon_j &= O_P(n^\delta (\hat{h} - h)h^{-1}(nh)^{-1/2}) \\ &= O_P(n^\delta n^{\gamma-\mu}(nh)^{-1/2}).\end{aligned}$$

Using Taylor series expansion of m and the fact that $\bar{w}_j(x) = \frac{d}{dh}\{w_j(x, h)\} = O(h^{-1}(nh)^{-1})$ we obtain, for some ρ_j between x and x_j, that

$$\sum \bar{w}_j(x)(m(x_j) - m(x)) = \sum \bar{w}_j(x)\frac{m^{(k)}(\rho_j)}{k!}(x_j - x)^k = O(h^{k-1}),$$

which yields

$$\sum (w_j(x, \hat{h}) - w_j(x, h))m(x_j) = O_P((\hat{h} - h)h^{k-1}) = O_P(n^{\gamma-\mu} h^k). \quad \square$$

Proof of Proposition 3.1 According to the proof of Lemma 3.1 we get

$$|T_n(\hat{h}) - T_n(h)| \leq \sup_x\{|\widehat{m}_{\hat{h}}(x) - \widehat{m}_h(x)|\} = \tilde{O}(n^{\gamma-\mu}(n^\delta(nh)^{-1/2} + h^k))$$

and, by similar considerations,

$$t^*_\alpha(\hat{h}) - t^*_\alpha(h) = \tilde{O}(n^{\gamma-\mu}(n^\delta(nh)^{-1/2} + h^k)),$$

which proves the assertion in conjunction with Theorem 2.2. $\square$

Acknowledgement

We thank Professor V. Konakov for reference to the papers by Sakhanenko and Tsirel'son. Further, we thank an Associate Editor and an anonymous referee for helpful comments.

References

Bellver, C. (1987) Influence of particulate pollution on the positions of neutral points in the sky in Seville (Spain). *Atmospheric Environment*, **21**, 699–702.

Bickel, P. and Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1096.

Bjerve, S., Doksum, K. A. and Yandell, B. S. (1985) Uniform confidence bounds for regression based on a simple moving average. *Scand. J. Statist.*, **12**, 159–169.

Cleveland, W. S. (1993) Visualizing data. Hobard Press, Summit, New Jersey.

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

Fan, J. (1992) Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998–1004.

Fan, J. and Gijbels, I. (1995) Adaptive order polynomial fitting: bandwidth robustification and bias reduction. *J. of Computational and Graphical Statistics*, **4**, 213–227.

Faraway, J. (1990) Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Statist. Comput. Simul.*, **37**, 37–44.

Hall, P. (1991a) On the distribution of suprema. *Probab. Theory Rel. Fields*, **89**, 447–455.

Hall, P. (1991b) Edgeworth expansions for nonparametric density estimators, with applications, *Statistics*, **22**, 215–232.

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

Hall, P. and Titterington, D. M. (1988) On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.*, **27**, 228–254.

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

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

Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.

Härdle, W. and Marron, J. S. (1991) Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778–796.

Johnston, G. J. (1982) Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.*. **12**, 402–414.

Jones, M. C. and Foster, P. J. (1993) Generalized jackknifing and higher order kernels. *Nonparametric Statist.*, **3**, 81–94.

Knafl, J., Sacks, J. and Ylvisaker, D. (1985) Confidence bands for regression functions. *J. Amer. Statist. Assoc.*, **80**, 683–691.

Liero, H. (1982) On the maximal deviation of the kernel regression function estimate. *Math. Operationsforsch. Statist., Ser. Statist.*, **13**, 171–182.

Loader, C. R. (1993) Nonparametric regression, confidence bands and bias correction. *Proceedings of the Interface Between Statistics and Computer Science,* 131–136.

Major, P. (1973) On non-parametric estimation of the regression function. *Studia Sci. Math. Hungar.*, **8**, 347–361.

Neumann, M. H. (1992) Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure. Statistics, to appear.

Neumann, M. H. (1995) Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist.*, **23**, 1937–1959.

Qualls, C. and Watanabe, H. (1972) *Ann. Math. Statist.*, **43**, 580–596.

Raz, J. (1990) Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach (Corr: V85 p. 1182). *J. Amer. Statist. Assoc.*, **85**, 132–138.

Révész, P. (1979) On the nonparametric estimation of the regression function. *Problems of Control and Information Theory*, **8**, 297–302.

Ruppert, D. and Wand, M. P. (1994) Multivariate locally weighted least squares regression. *Ann. Statist.*, **22**, 1346–1370.

Sakhanenko, A. I. (1991) On the accuracy of normal approximation in the invariance principle. *Siberian Advances in Mathematics*, **1**, 58–91.

Shorack, G. R. and Wellner, J. A. (1986) Empirical Processes with Applications to Statistics. Wiley, New York.

Stone, C. J. (1977) Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–620.

Sun, J. and Loader, C. R. (1994) Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.*, **22**, 1328–1345.

Tsirel'son, B. S. (1975) The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.*, **20**, 847–856.

Whittle, P. (1960) Bounds for the moments of linear and quadratic forms in independent variables. *Theory Prob. Appl.*, **5**, 302–305.

Wu, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.*, **14**, 1261–1343.