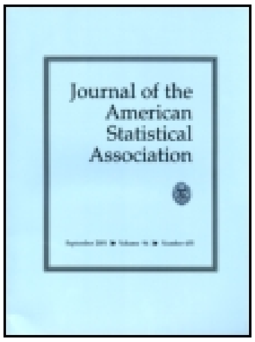



Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

# Confidence Bands in Nonparametric Regression

R. L. Eubank & P. L. Speckman

**To cite this article:** R. L. Eubank & P. L. Speckman (1993) Confidence Bands in Nonparametric Regression, Journal of the American Statistical Association, 88:424, 1287-1301

**To link to this article:** https://doi.org/10.1080/01621459.1993.10476410

Published online: 27 Feb 2012.

# Confidence Bands in Nonparametric Regression

R. L. EUBANK and P. L. SPECKMAN*

New bias-corrected confidence bands are proposed for nonparametric kernal regression. These bands are constructed using only a kernel estimator of the regression curve and its data-selected bandwidth. They are shown to have asymptotically correct coverage properties and to behave well in a small-sample study. One consequence of the large-sample developments is that Bonferroni-type bands for the regression curve at the design points also have conservative asymptotic coverage behavior with no bias correction.

KEY WORDS: Bias correction; Bonferroni bands; Cross-validation; Kernel estimators.

## 1. INTRODUCTION

Nonparametric function estimators provide important data analytic tools. Many of their properties have been thoroughly investigated and are now well understood. Unfortunately, techniques for constructing interval estimates to accompany these estimators have been slower to develop. Current methods for constructing confidence bands or regions in this setting have some practical shortcomings that will be discussed subsequently. In this article we develop new methodology for interval estimation in nonparametric regression.

Consider the situation where responses $y_1, \ldots, y_n$ are obtained at equally spaced design points $t_r = r/n$, $r = 1, \ldots, n$. The $y_r$ and $t_r$ are related under the model

$$y_r = \mu(t_r) + \varepsilon_r, \qquad r = 1, \ldots, n, \tag{1}$$

where the $\varepsilon_r$ are independent identically distributed random variables with mean 0 and common variance $\sigma^2$ and $\mu$ is an unknown smooth regression curve. The problem to be addressed is the construction of confidence bands for $\mu$. Specifically, given $\alpha \in (0, 1)$ and an estimator $\hat{\mu}$ for $\mu$, we want to find a bound $l_\alpha$ such that

$$P(\sup_{0 \le t \le 1} |\hat{\mu}(t) - \mu(t)| \le l_\alpha) \ge 1 - \alpha, \tag{2}$$

at least in large samples.

One of the difficulties with finding solutions to (2) is that nonparametric estimators for $\mu$ are biased. Consequently, confidence band procedures must deal with estimator bias to ensure that the bands are correctly centered and proper coverages are attained.

To illustrate the problems created by estimator bias, consider the case of a kernel estimator for $\mu$ of the form

$$\mu_\lambda(t) = \frac{1}{n\lambda} \sum_{r=1}^{n} y_r K\left(\frac{t - t_r}{\lambda}\right), \tag{3}$$

with $\lambda > 0$ the bandwidth and $K$ a kernel function supported on $[-1, 1]$. Classically, confidence bands to accompany $\mu_\lambda$ have been based on the presumption of an approximate standard normal distribution for $T_\lambda = [\mu_\lambda(t) - \mu(t)]/\sqrt{\operatorname{var} \mu_\lambda(t)}$. This provides pointwise confidence intervals for $\mu(t)$ that can then be used along with Bonferroni's inequality to obtain, for example, simultaneous confidence intervals for the values of the regression curve at the design points. The problem with this approach is that $T_\lambda$ may not have the assumed distribution. To see this, express $T_\lambda$ as

$$T_\lambda = \frac{\mu_\lambda(t) - E\mu_\lambda(t)}{\sqrt{\operatorname{var} \mu_\lambda(t)}} + \frac{E\mu_\lambda(t) - \mu(t)}{\sqrt{\operatorname{var} \mu_\lambda(t)}}. \tag{4}$$

The first term in the sum (4) is asymptotically standard normal under weak conditions. Thus for $T_\lambda$ to have the desired limiting distribution, we need $\lambda$ to be chosen in a way that ensures the second term in the sum converges to 0. Bandwidths that minimize local or global mean squared error (MSE) measures for $\mu_\lambda$ will not have this property, because they balance the bias and standard error of the estimator in a way that forces both to decay at the same rate. The same will be true for bandwidths obtained from most of the common data-driven bandwidth selectors. As a result, pointwise confidence intervals obtained from $T_\lambda$ cannot be expected to work in practice. Although Bonferroni bands with stochastic bandwidths may still satisfy (2), this conclusion cannot be reached on the basis of asymptotic normal theory for $T_\lambda$.

The failure of normal approximations to provide valid confidence bands has fostered a number of alternative approaches. Several authors have studied the limiting distribution of (a suitably recentered and rescaled version of) $M = \sup_{0 \le t \le 1} |\hat{\mu}(t) - \mu(t)|$ for various estimators $\hat{\mu}$ of $\mu$. One of the earliest results of this type is due to Bickel and Rosenblatt (1973) for kernel density estimators. Extensions of the Bickel and Rosenblatt work to kernel regression were done by Johnston (1982) and Härdle (1989), and similar results for nearest-neighbor estimators were derived by Révész (1979) and Bjerve, Doksum, and Yandell (1985). Unfortunately, the confidence bands obtained from these articles rely on deterministic, suboptimal choices for smoothing parameters that make the bias of $\hat{\mu}$ negligible relative to its standard error.

Knafl, Sacks, and Ylvisaker (1982, 1985) and Hall and Titterington (1988) have developed confidence bands for $\mu$ based on large-sample upper bounds for the size of $M$. Their approach involves the approximation of $M$ by $M_k$

* R. L. Eubank is a Professor, Department of Statistics, Texas A&M University, College Station, TX 77843-3143. P. L. Speckman is Associate Professor, Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211. A portion of Eubank's research was supported under National Science Foundation Grant DMS-9024879.


Journal of the American Statistical Association
December 1993, Vol. 88, No. 424, Theory and Methods

$= \max_{1 \le r \le k} |\hat{\mu}(\xi_r) - \mu(\xi_r)|$ for $0 = \xi_1 < \cdots < \xi_k = 1$, a grid of points in [0, 1]. To use these bands in practice, one must have a priori bounds on the magnitude of the bias for $\hat{\mu}$ over each subinterval $[\xi_j, \xi_{j+1}]$, as well as a choice for $k$.

Hall and Titterington (1988) also established the best rate of decay for confidence band lengths. For example, for the class of regression functions with two continuous derivatives, they showed that no confidence band with uniform $100(1 - \alpha)\%$ coverage over this class can have length converging to 0 faster than does a constant multiple of $(\log n/n)^{2/5}$. The bands that they derived are seen to attain this rate if $k$ is chosen correctly.

Bootstrap confidence bands based on kernel estimators of $\mu$ have been studied by Härdle and Bowman (1988) and Härdle and Marron (1991). The Härdle–Marron bands involve a partitioning of [0, 1] into subintervals and require selection of an additional bandwidth used for a bias correction in their bootstrapping process. If this auxiliary bandwidth is chosen correctly, then the bands are found to be asymptotically valid in the sense that (2) holds for large $n$.

Wahba (1983) and Nychka (1988, 1990) considered confidence intervals for the $\mu(t_r)$ based on Bayesian considerations and smoothing spline estimators. The resulting intervals are valid in an average coverage sense. More precisely, if the coverage probabilities at the $t_r$ are averaged over the design, then this average will tend to be close to the nominal level. But coverages at any particular design point can depart appreciably from the desired level, and the intervals will tend to undercover where the estimation bias is largest. A more classical frequentist confidence band for smoothing spline estimators was proposed by Cox (1986). His method relies in part on an estimate of the norm of the bias.

Simultaneous asymptotic confidence regions for the values of $\mu$ at the design points were given by Li (1989). His regions are $n$-dimensional ellipsoids whose projections on their coordinate axes can be used to construct a parallel of confidence bands for the $\mu(t_r)$. The lengths of these "bands" have the displeasing property of actually growing with $n$.

Despite the difficulties involved, one might still hope for a relatively simple solution to the confidence band problem. Ideally, such a procedure would be entirely data-driven and based directly on the estimator that is being fitted to the data. In the next section we show that these goals are attainable for the kernel estimator (3). The bands we derive require only the estimator and a data-driven choice for its bandwidth $\lambda$. No additional tuning parameters are involved, and the bands can be produced directly, without further iteration or data reuse, given the value to be used for $\lambda$.

The large-sample properties of our confidence bands and Bonferroni-type bands are discussed in Section 2. Their small sample behavior is then investigated in Section 3 via a Monte Carlo experiment. Results are summarized and some closing comments are made in Section 4.

## 2. BIAS-CORRECTED CONFIDENCE BANDS

In this section confidence bands are derived for model (1). We deal initially with the simple case where $\mu$ is twice continuously differentiable and periodic in that

$$\mu^{(j)}(0) = \mu^{(j)}(1), \qquad j = 0, 1, 2, \tag{5}$$

and discuss how this condition might be removed subsequently. When (5) holds, one can periodically extend the data to compute $\mu_\lambda(t)$ for values of $t$ within $\lambda$ of 0 or 1 and thereby avoid problems with boundary effects. We begin with some preliminary discussions.

Assume that $\mu$ is twice continuously differentiable and that $K$ in (3) is a square-integrable kernel satisfying $\int_{-1}^{1} K(u)\,du = 1$, $\int_{-1}^{1} uK(u)\,du = 0$, and $\int_{-1}^{1} u^2K(u)\,du \neq 0$. Then it is known that if $\lambda \to 0$ as $n \to \infty$ in such a way that $n\lambda \to \infty$,

$$E\mu_\lambda(t) \sim \mu(t) + \lambda^2 B\mu''(t) \tag{6}$$

and

$$\operatorname{var} \mu_\lambda(t) \sim \frac{\sigma^2 V^2}{n\lambda}, \tag{7}$$

where

$$B = \int_{-1}^{1} u^2K(u)\,du \Big/ 2 \tag{8}$$

and

$$V^2 = \int_{-1}^{1} K(u)^2\,du. \tag{9}$$

When (5) holds, (6) and (7) may be combined to obtain approximations to various global performance measures for $\mu_\lambda$. For example, the average MSE for $\mu_\lambda$ is

$$R(\lambda) = n^{-1} \sum_{r=1}^{n} E(\mu(t_r) - \mu_\lambda(t_r))^2$$

$$\sim \frac{\sigma^2 V^2}{n\lambda} + \lambda^4 B^2 \int_0^1 \mu''(t)^2\,dt. \tag{10}$$

From (10), we see that the minimizer of $R(\lambda)$ is approximately

$$\lambda_0 = n^{-1/5}\left[\sigma^2 V^2 \Big/ 4B^2 \int_0^1 \mu''(t)^2\,dt\right]^{1/5}, \tag{11}$$

which plays an important role in the sequel.

To use $\mu_\lambda$ in practice, it is necessary to have a choice for the bandwidth $\lambda$. Available data-driven methods for choosing $\lambda$ include minimization of the generalized cross-validation criterion. In our situation of equally spaced design points, this criterion takes the form $n^{-1} \sum_{r=1}^{n} (y_r - \mu_\lambda(t_r))^2 / (1 - K(0)/n\lambda)^2$. In what follows we will use $\hat{\lambda}$ to denote some generic bandwidth estimator. Our only condition is that

$$(\hat{\lambda} - \lambda_0)/\lambda_0 = O_p(n^{-1/10}). \tag{12}$$

Rice (1984) and Härdle, Hall, and Marron (1988) have shown that (12) is satisfied by many bandwidth estimators, including the one obtained from generalized cross-validation.

The confidence bands we propose to use have the form

$$\{\mu: \sup_{0\le t\le 1} |\mu_{\hat\lambda}(t) - b_{\hat\lambda}(t) - \mu(t)| \le l_\alpha\}, \tag{13}$$

where $b_{\hat\lambda}(t)$ is a bias correction and $l_\alpha$ is an appropriately chosen bound. Motivated by (6), we take

$$b_{\hat\lambda}(t) = \hat\lambda^2 B\hat\mu''(t), \tag{14}$$

with $\hat\mu''$ as an estimator for the second derivative of $\mu$. The specific form we use for $\hat\mu''$ is

$$\hat\mu''(t) = \frac{1}{n\tilde\lambda^3}\sum_{r=1}^{n} y_r K^*\left(\frac{t-t_r}{\tilde\lambda}\right), \tag{15}$$

with $\tilde\lambda = \hat\lambda^{5/7}$ and $K^*$ a square-integrable kernel supported on $[-1, 1]$ that satisfies $\int_{-1}^{1} u^j K^*(u)\,du = 0$, $j = 0, 1$, and $\int_{-1}^{1} u^2 K^*(u)\,du = 2$. The conditions on $K^*$ imply that (15) is just a kernel estimator for $\mu''$ with bandwidth $\hat\lambda^{5/7}$ (cf. Müller 1988). This choice for the bandwidth allows us to make the bias correction totally data-driven. Assuming (12), it also provides a bandwidth that decays at the optimal rate $n^{-1/7}$ for estimating $\mu''$ when it is differentiable.

The choice of $l_\alpha$ in (13) is a consequence of the following theorem, whose proof is given in the Appendix.

*Theorem.* Assume that (5) and (12) hold and that (a) the $\varepsilon$'s possess more than 19 absolute moments; (b) $\mu''$ satisfies a Lipschitz condition of order $\nu > 0$; and (c) $K$ and $K^*$ are twice continuously differentiable. Define

$$l_\alpha^* = \frac{\sigma V}{\sqrt{n\hat\lambda}}\left\{\sqrt{-2\log\hat\lambda} + \frac{1}{\sqrt{-2\log\hat\lambda}}[C + x_\alpha]\right\} \tag{16a}$$

with

$$C = \log\left(\frac{1}{2\pi}\left[\int_{-1}^{1} K'(u)^2\,du \Big/ \int_{-1}^{1} K(u)^2\,du\right]^{1/2}\right) \tag{16b}$$

and

$$x_\alpha = -\log\left(\frac{-\log(1-\alpha)}{2}\right). \tag{16c}$$

Then

$$\lim_{n\to\infty} P(\sup_{0\le t\le 1} |\mu_{\hat\lambda}(t) - b_{\hat\lambda}(t) - \mu(t)| \le l_\alpha^*) = 1 - \alpha. \tag{17}$$

To use the bound $l_\alpha^*$ provided by (16) in practice, we typically will have to estimate $\sigma^2$. This can be done without altering the conclusion of the theorem if we use any of various $\sqrt{n}$-consistent estimators for $\sigma^2$ that can be found in, for example, Gasser, Sroka, and Jennen-Steinmetz (1986), Hall, Kay, and Titterington (1990), or Hall and Marron (1990). Letting $\hat\sigma$ be any of these estimators, we define

$$l_\alpha = \hat\sigma l_\alpha^* / \sigma \tag{18a}$$

to obtain the asymptotic $100(1-\alpha)\%$ confidence band

$$\mu_{\hat\lambda}(t) - b_{\hat\lambda}(t) \pm l_\alpha \quad \forall\, t \in [0, 1]. \tag{18b}$$

The length of the band (18) behaves like $2\sigma V \times \sqrt{2\log n/5n\hat\lambda} = O_p(\sqrt{\log n}/n^{2/5})$, for large $n$, as a result of (12). Thus the length converges to 0 only slightly slower than does the $(\log n/n)^{2/5}$ rate of Hall and Titterington (1988). To obtain this latter rate, Hall and Titterington chose a smoothing parameter designed to minimize the length of their intervals, rather than MSE. For kernel estimation, their approach would translate into use of a bandwidth with order $(\log n/n)^{1/5}$. As a result, we conjecture that their optimal rate of decay for the band length cannot be obtained using a bandwidth of order $n^{-1/5}$ (e.g., one that satisfies (12)) and that $\sqrt{\log n}/n^{2/5}$ is the best possible rate for our situation.

A somewhat surprising conclusion that follows from the theorem is that Bonferroni-type bands for the $\mu(t_r)$ are asymptotically conservative even though they do not explicitly account for bias. These bands take the form

$$\mu_{\hat\lambda}(t_r) \pm l_{B\alpha}, \qquad r = 1, \ldots, n, \tag{19a}$$

with

$$l_{B\alpha} = \frac{\hat\sigma V Z_{\alpha/2n}}{\sqrt{n\hat\lambda}} \tag{19b}$$

for $Z_{\alpha/2n}$, the $100(1-\alpha/2n)$th percentile of the standard normal distribution. To see why (19) works, observe that

$$P(\max_{1\le r\le n} |\mu_{\hat\lambda}(t_r) - \mu(t_r)| \le l_{B\alpha})$$
$$\ge P(\sup_{0\le t\le 1} |\mu_{\hat\lambda}(t) - b_{\hat\lambda}(t) - \mu(t)| \le l_{B\alpha} - \sup_{0\le t\le 1} |b_{\hat\lambda}(t)|).$$

So the Bonferroni bands will have asymptotic coverage at least $1-\alpha$ if $l_{B\alpha} - \sup_{0\le t\le 1}|b_{\hat\lambda}(t)|$ is larger than $l_\alpha$ with probability tending to 1. But this is true because $l_{B\alpha}$ behaves like $\sigma V\sqrt{2\log n}/\sqrt{n\hat\lambda}$, $l_\alpha$ behaves like $\sigma V\sqrt{(2/5)\log n}/\sqrt{n\hat\lambda}$, and, using arguments in the Appendix, $\sup_{0\le t\le 1}|b_{\hat\lambda}(t)| = O_p(\hat\lambda^2) = O_p(n^{-2/5})$.

Our conclusions about the bands (18)–(19) remain valid under various conditions other than the ones we have stated. For example, condition (5) can be removed if interest is restricted to confidence bands over a subinterval $[a, b]$ with $a > 0$ and $b < 1$. This also allows us to avoid estimator edge effects. It is unclear whether our results extend to estimators using boundary corrections such as boundary kernels, although we would expect this to be the case. Indeed, it is even unknown whether (12) holds when boundary kernels are used.

The requirement of more than 19 moments for the error distribution is undoubtedly stronger than necessary. This condition is an artifice of the method of proof of Lemma 4, and there seems to be no way to relax it using our approach. We conjecture that the theorem remains valid if one assumes only that the $\varepsilon$'s have finite variance.

To conclude this section, let us detail our specific proposal for confidence bands that will be used in the next section.

The bias-corrected bands we use take the form

$$\mu_{\hat\lambda}(t) - b_{\hat\lambda}(t) \pm l_\alpha, \qquad (20a)$$

where

$$l_\alpha = \frac{\hat\sigma V_{1n}}{\sqrt{n\hat\lambda}}\left\{\sqrt{-2\log\hat\lambda} + \frac{1}{\sqrt{-2\log\hat\lambda}}[C + x_\alpha]\right\} \qquad (20b)$$

for

$$C = \log\left(\frac{1}{2\pi}\left[\int_{-1}^{1} K'(u)^2\,du \Big/ \int_{-1}^{1} K(u)^2\,du\right]^{1/2}\right), \qquad (20c)$$

$$x_\alpha = -\log\left(\frac{-\log(1-\alpha)}{2}\right), \qquad (20d)$$

$$b_{\hat\lambda}(t) = \hat\lambda^2 B \frac{1}{n\hat\lambda^{15/7}} \sum_{r=1}^{n} y_r K^*\left(\frac{t - t_r}{\hat\lambda^{5/7}}\right), \qquad (20e)$$

$$B = \int_{-1}^{1} u^2 K(u)\,du \Big/ 2, \qquad (20f)$$

and

$$V_{1n}^2 = \sum_{|r|\le n\hat\lambda}\left[\frac{1}{n\hat\lambda}K\left(\frac{r}{n\hat\lambda}\right) + \frac{B}{n\hat\lambda^{1/7}}K^*\left(\frac{r}{n\hat\lambda^{5/7}}\right)\right]^2. \qquad (20g)$$

Here $\hat\sigma$ can be any $\sqrt{n}$-consistent estimator of $\sigma$, and $K^*$ can be any kernel suitable for second-derivative estimation. In our simulations we used the Hall et al. (1990) estimator for $\sigma^2$:

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

Our choice for $K$ was the Epanechnikov kernel $K(u) = .75(1 - u^2)I_{[-1,1]}(u)$, with $I$ denoting the indicator function; for $K^*$ we used $K^*(u) = (105/32)(-5u^4 + 6u^2 - 1)I_{[-1,1]}(u)$ from Gasser, Müller, and Mammitzsch (1985).

Notice that (16a) and (20b) differ in the use of $V_{1n}$ rather than $V$. The idea here is that when we construct our bias-corrected bands, we are really estimating $\mu$ by $\mu_\lambda + b_\lambda = \sum w_{r\lambda} y_r$ for weights $w_{r\lambda}$. These weights are what are used to compute $V_{1n}$. Although using $V_{1n}$ can be shown to be asymptotically equivalent to using $V$, we have found that (20b) provides substantial finite sample improvements.

The other bands we considered were the Bonferroni bands, which we implemented as

$$\mu_{\hat\lambda}(t_r) \pm l_{B\alpha}, \qquad r = 1, \ldots, n \qquad (21a)$$

for

$$l_{B\alpha} = \frac{\hat\sigma V Z_{\alpha/2n}}{\sqrt{n\hat\lambda}}, \qquad (21b)$$

with $Z_{\alpha/2n}$ as the $100(1 - \alpha/2n)$th percentile of the standard normal distribution.

The question now arises as to which of the two types of bands is preferable in practice. The Bonferroni bands (21) are simpler and are commonly used in practice. Thus they would be preferred if their coverage properties and lengths are similar to the bands (20). We address this issue in the next section through a simulation study.

## 3. FINITE SAMPLE PROPERTIES

To assess the performance of the bands (18)–(19) in practice, we conducted a Monte Carlo experiment. Data were generated from model (1) using normal errors and two choices for the regression curve:

$$\mu_1(t) = \sin^2(2\pi t)$$

and

$$\mu_2(t) = e^{-32(t-.5)^2}.$$

Sample sizes $n$ = 50, 100, 200, 300, 400, 600, and 800 were used, and $\sigma$ was chosen to be .05, .10, or .20. These values of $\sigma$ were selected to give random errors of typical size near 5%, 10%, or 20% of the magnitude of the curves. Confidence bands were then investigated for $\alpha$ = .01, .05, .10, and .20.

Plots of typical data sets from the simulation corresponding to each regression function for $n$ = 100 and $\sigma$ = .10 are shown in Figures 1a and 1b. Plotted along with the data are the true regression function and a kernel estimator with bandwidth selected by generalized cross-validation. The fitted curves exhibit a common feature of kernel estimators in the way they tend to overestimate troughs and underestimate peaks of $\mu_1$ and $\mu_2$. Figures 2a and 2b are plots of the Bonferroni and bias-corrected bands with $\alpha$ = .10 for $\mu_1$ and $\mu_2$, showing how bias-corrected bands shift in the appropriate fashion to compensate for the bias of the kernel estimator at peaks and valleys of the regression functions. Figures 3a and 3b plot the bias-corrected bands for $\mu_1$ and $\mu_2$ with $\alpha$ = .10 and $\alpha$ = .01. Notice that there is relatively little difference in band length for the different choices of $\alpha$.

We generated 5,000 replicate samples for each experimental setting, using a different seed for every case. Kernel estimates were fit to the data sets, with $\lambda$ selected by generalized cross-validation. Bands (20) and (21) were then computed for $\alpha$ = .01, .05, .10, and .20 using the Hall et al. (1990) estimator of $\sigma$. The proportion of times all the $\mu(t_r)$, $r = 1, \ldots, n$, fell inside the bands was recorded, as was the average (over replications) band half-lengths $l_\alpha$ and $l_{B\alpha}$. In addition, empirical coverage probabilities were computed for three other intervals:

$$\mu_{\hat\lambda}(t) - \frac{1}{n\hat\lambda}\sum_{r=1}^{n}\mu(t_r)K\left(\frac{t - t_r}{\hat\lambda}\right) \pm l_\alpha^*,$$

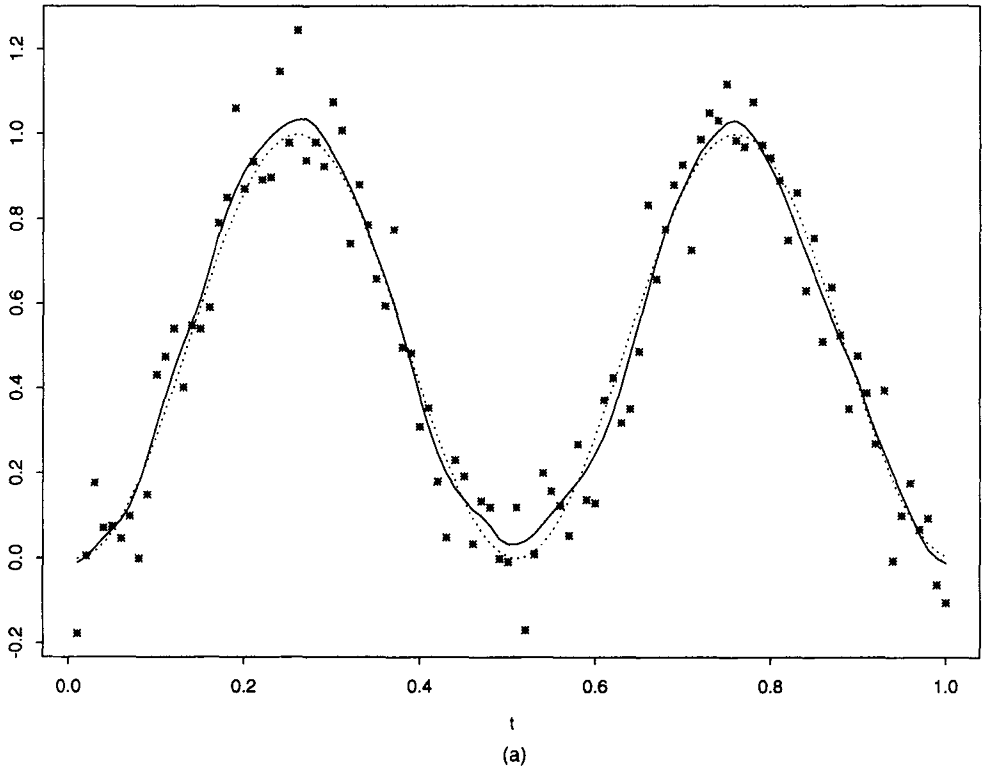
$$\mu_{\hat\lambda}(t) \pm l_\alpha^*,$$

and

$$\mu_{\hat\lambda}(t) \pm \frac{\hat\sigma}{\sigma} l_\alpha^*.$$

These additional three bands represent coverage with no bias when $\sigma$ is known ("truth"), coverage when bias is ignored but $\sigma$ is known ("sigma known"), and coverage when bias is ignored but $\sigma$ is estimated ("sigma estimated").

We summarize some of the results of the simulation through a series of plots. Figures 4a and 4b present the empirical coverage probabilities for the particular case $\sigma$ = .10



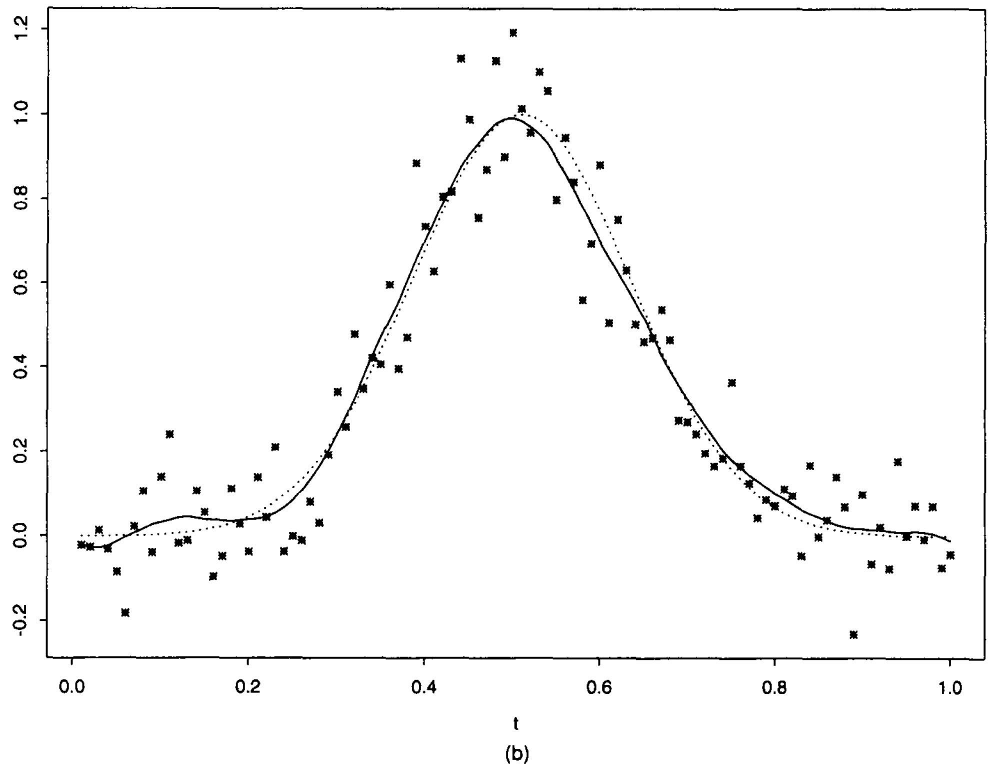




*Figure 1. Data with (a) regression curve $\mu_1$, $n = 100$, and $\sigma = .10$; (b) regression curve $\mu_2$, $n = 100$, and $\sigma = .10$. ····· $\mu(t)$; ——— $\hat{\mu}(t)$.*

and $\alpha = .10$ as a function of sample size for $\mu_1$ and $\mu_2$. Estimated standard errors for the coverage probabilities are plotted as error bars. Figures 5–8 show the results of the entire simulation for the bias-corrected and Bonferroni bands in all combinations of $\sigma$ and $\alpha$ for the two regression functions. (For simplicity, the error bars are suppressed.) In Figures 9a and 9b, the confidence band half lengths $l_\alpha$ and $l_{B\alpha}$ are plotted for $\sigma = .10$ with $\alpha = .10$ and $\alpha = .01$. For these plots, standard errors range from .00075 to .00004; hence they are omitted.

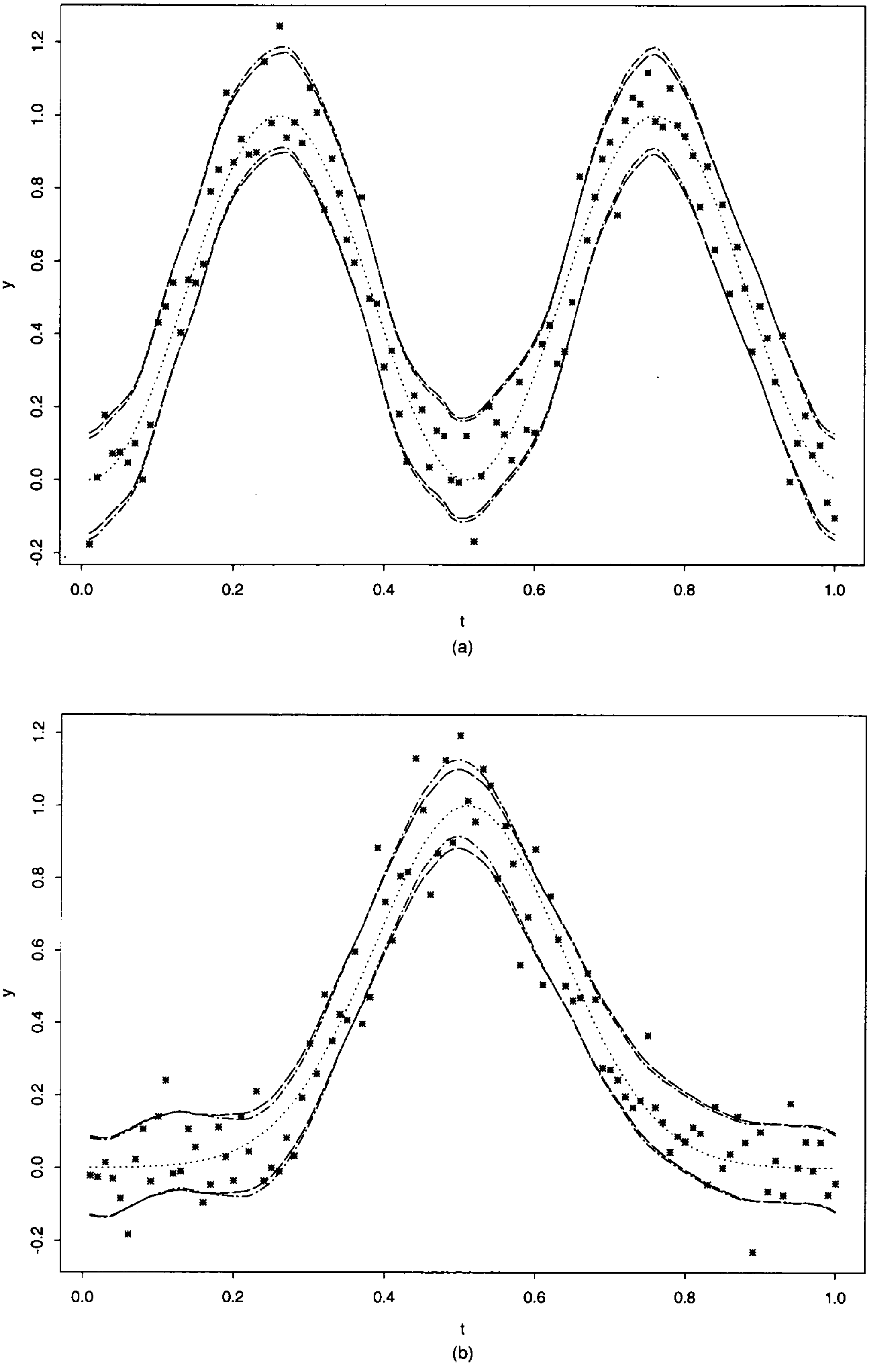



*Figure 2. 90% bias-corrected and Bonferroni bands for the data in (a) Figure 1a and (b) Figure 1b.* ····· *$\mu(t)$*; —·— *bias-corrected;* — — *Bonferroni.*

From Figures 4a and 4b, it is evident that the Bickel–Rosenblatt approximation for the maximum random error ("truth") is quite good even when $\lambda$ is chosen from the sample. This observation remained true for other combinations of $\sigma$ and $\alpha$ not shown here. The "sigma known" case demonstrates that bias in $\mu_\lambda$ is a serious problem and leads to substantial undercoverage. The Hall et al. (1990) estimator of $\sigma$ appeared to have a positive bias, which improved the

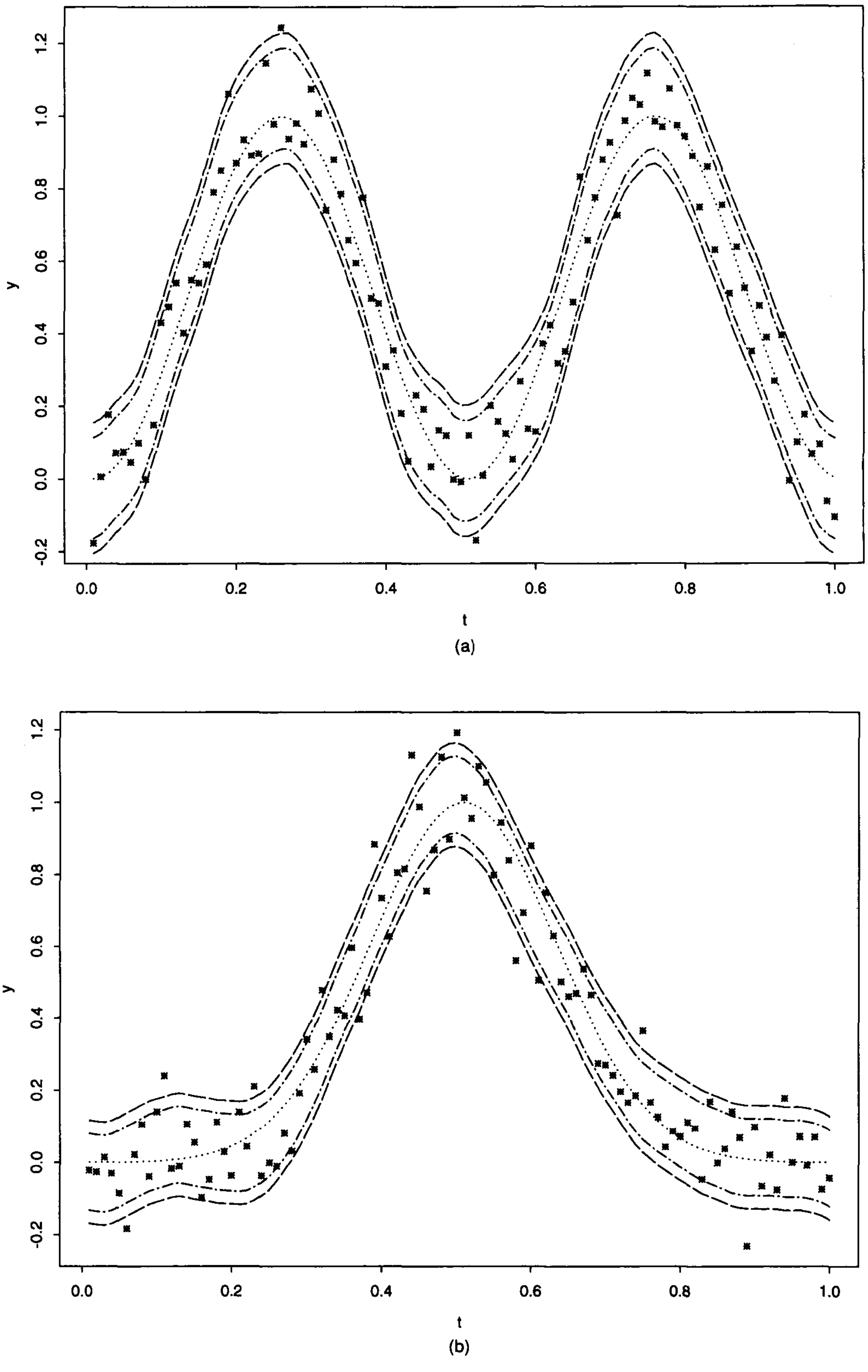
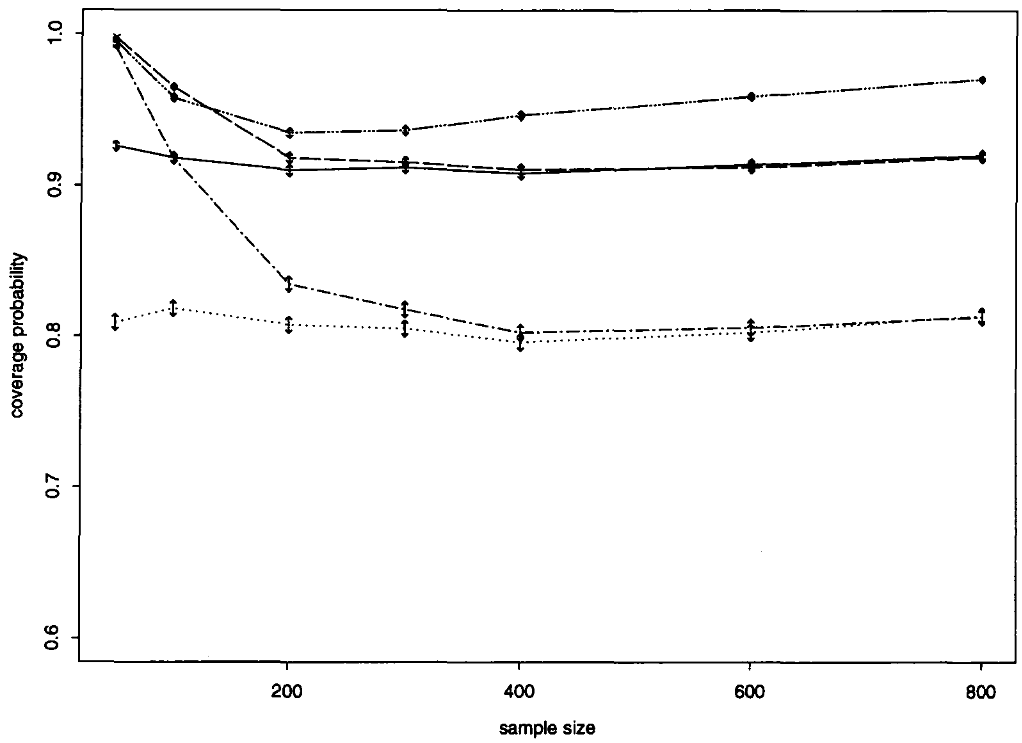


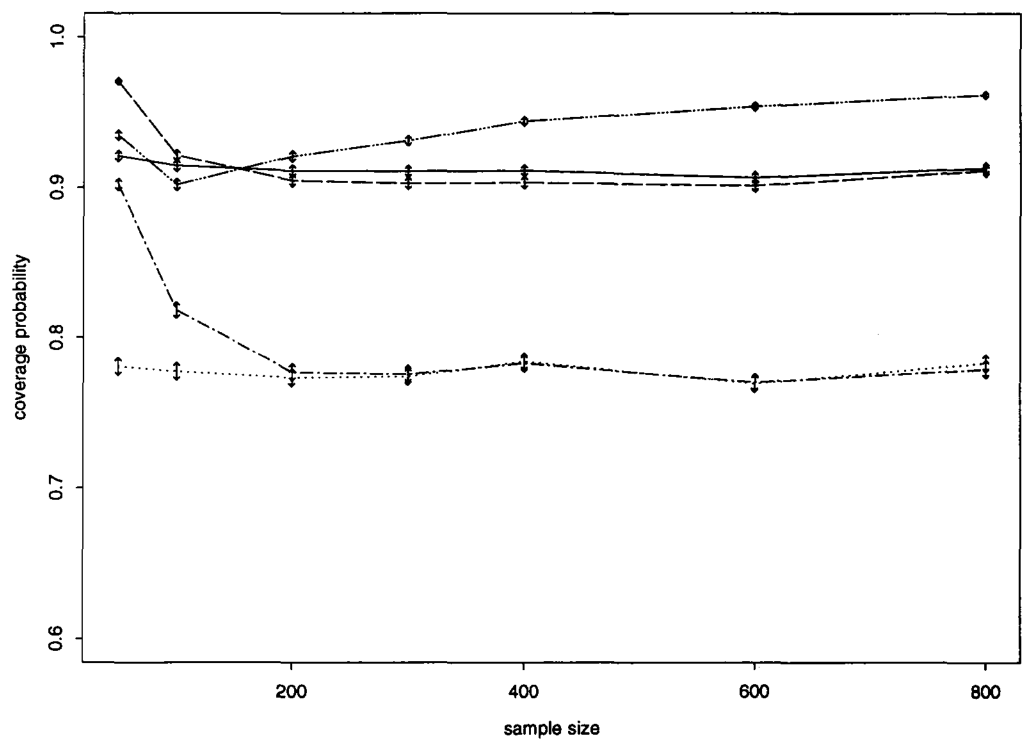

*Figure 3. 90% and 99% bias-corrected bands for the data in (a) Figure 1a and (b) Figure 1b.* · · · · · $\mu(t)$; — · — · $\alpha = .10$; — — $\alpha = .01$.

coverage probabilities even when the bias in $\mu_\lambda$ was ignored (the "sigma estimated" case) for samples of 50 or 100. But, for larger samples, the bias of $\hat{\sigma}$ becomes small, and the effect of the bias in $\mu_\lambda$ predicted by the asymptotics clearly takes hold.

Examination of Figures 5–8 clarifies the relationship between the bias-corrected and Bonferroni bounds. In all the cases, except possibly for estimating $\mu_2$ with $\sigma = .20$ and samples less than 200, the bias-correction technique appears to be quite effective. In the remaining cases, coverage prob-




(a)




(b)

*Figure 4. Empirical coverage probabilities with standard errors for (a) regression curve $\mu_1$, $\sigma$ = .10, and $\alpha$ = .10, and (b) regression curve $\mu_2$, $\sigma$ = .10, and $\alpha$ = .10. ——— truth; · · · · · $\sigma$ known; — · — · $\sigma$ estimated; — — bias-corrected; — · · · — Bonferroni.*

abilities were close to or exceeded the nominal probabilities. On the other hand, the Bonferroni method tends to provide coverage at least as large as the nominal in all cases. But for large samples, the bound appears to be needlessly conservative, with all coverage probabilities approaching 1.

In contrast to what was predicted by our large-sample theory, Figures 9a and 9b demonstrate that the Bonferroni bands can be shorter than the bias-corrected bands. This occurred for small samples with small $\alpha$ and, as a result, is not entirely surprising. The differences between $l_\alpha$ and $l_{B\alpha}$ will be manifest

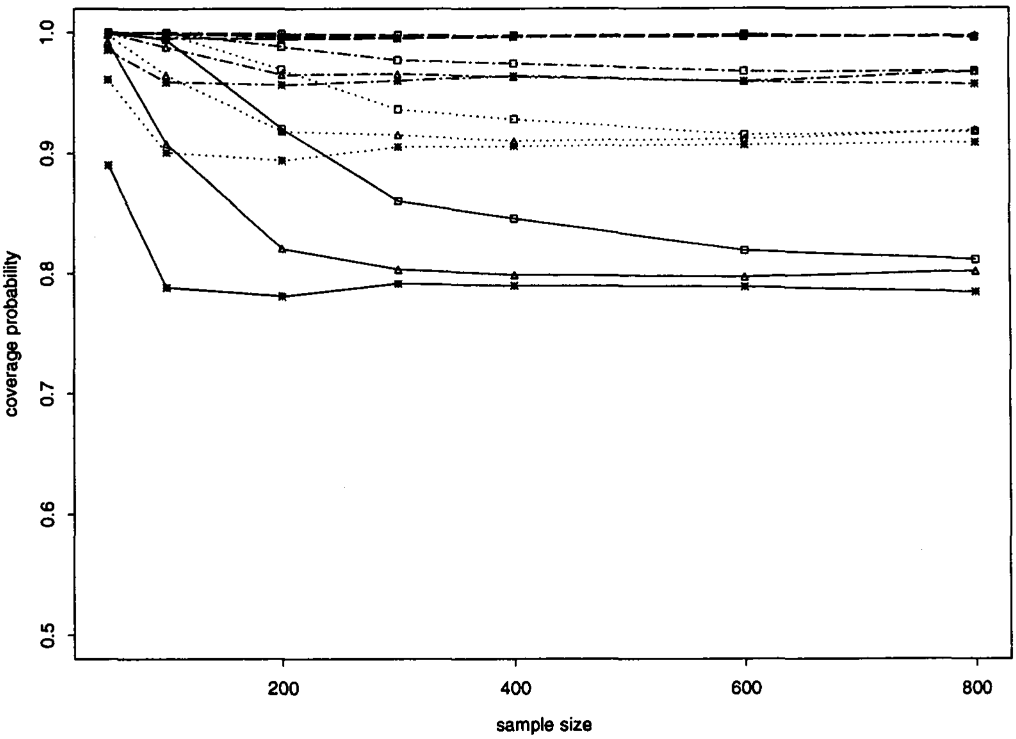



*Figure 5. Bias correction method empirical coverage probabilities for regression curve $\mu_1$. □ ——— $\alpha = .20$, $\sigma = .05$; □ · · · · · $\alpha = .10$, $\sigma = .05$; □ — · — · $\alpha = .05$, $\sigma = .05$; □ — — $\alpha = .01$, $\sigma = .05$; △ ——— $\alpha = .20$, $\sigma = .10$; △ · · · · · $\alpha = .10$, $\sigma = .10$; △ — · — · $\alpha = .05$, $\sigma = .10$; △ — — $\alpha = .01$, $\sigma = .10$; • ——— $\alpha = .20$, $\sigma = .20$; * · · · · · $\alpha = .10$, $\sigma = .20$; * — · — · $\alpha = .05$, $\sigma = .20$; * — — $\alpha = .01$, $\sigma = .20$.*

at $\sqrt{\log n}$ rates. There are also several constants in these bounds, some of which depend on $\alpha$. For sample sizes greater than 200, the asymptotics appear to take effect, and the Bonferroni bands are larger.

Small differences in band length can cause big differences in coverage. We have found that obtaining near-nominal coverage probabilities is a delicate matter. Seemingly minor modifications to the algorithm (such as the choice of $\tilde{\lambda}$ and

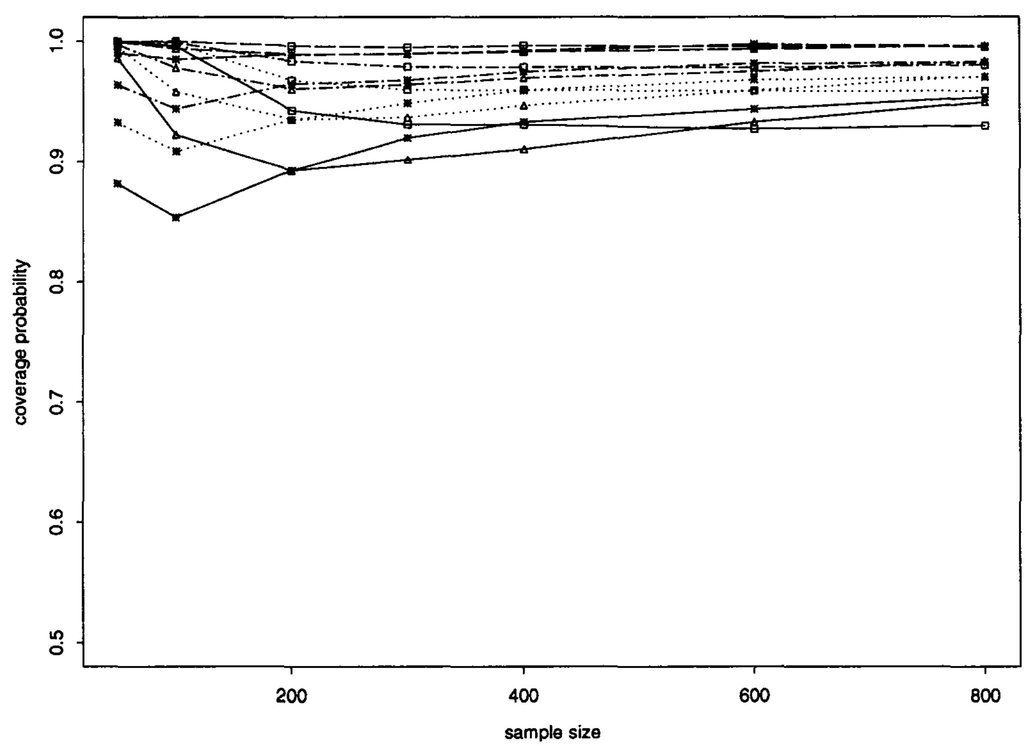



*Figure 6. Bonferroni method empirical coverage probabilities for regression curve $\mu_1$. □ ——— $\alpha = .20$, $\sigma = .05$; □ · · · · · $\alpha = .10$, $\sigma = .05$; □ — · — · $\alpha = .05$, $\sigma = .05$; □ — — $\alpha = .01$, $\sigma = .05$; △ ——— $\alpha = .20$, $\sigma = .10$; △ · · · · · $\alpha = .10$, $\sigma = .10$; △ — · — · $\alpha = .05$, $\sigma = .10$; △ — — $\alpha = .01$, $\sigma = .10$; • ——— $\alpha = .20$, $\sigma = .20$; * · · · · · $\alpha = .10$, $\sigma = .20$; * — · — · $\alpha = .05$, $\sigma = .20$; * — — $\alpha = .01$, $\sigma = .20$.*

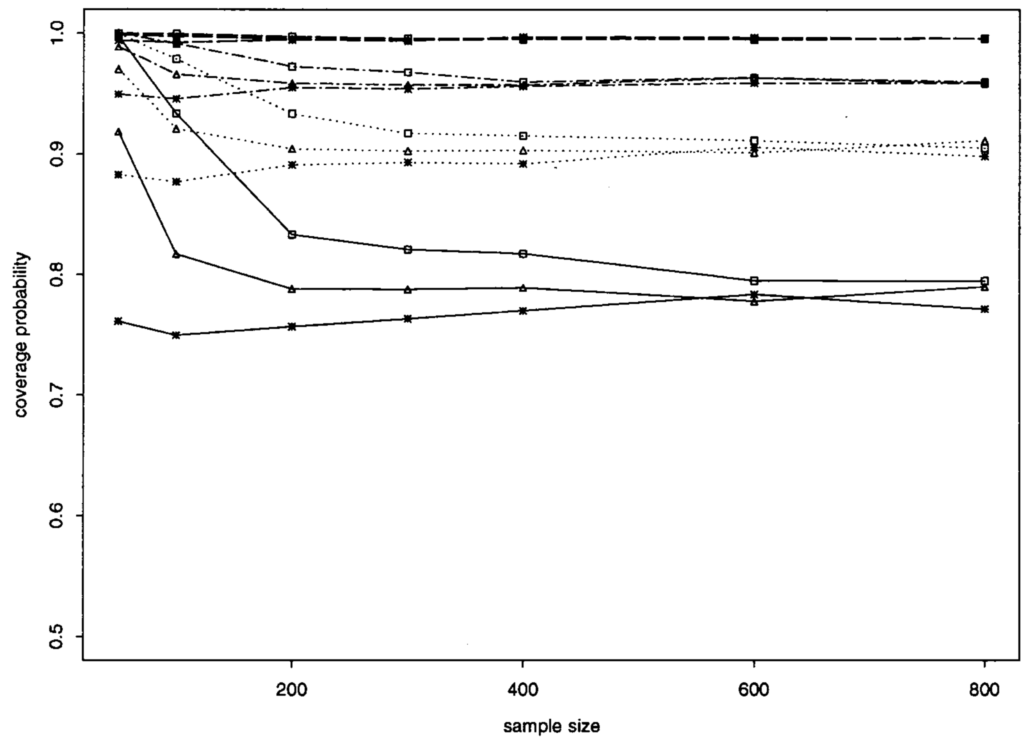



*Figure 7. Bias correction method empirical coverage probabilities for regression curve $\mu_2$.* □ ——— $\alpha = .20$, $\sigma = .05$; □ · · · · · $\alpha = .10$, $\sigma = .05$; □ — · — · $\alpha = .05$, $\sigma = .05$; □ — — $\alpha = .01$, $\sigma = .05$; △ ——— $\alpha = .20$, $\sigma = .10$; △ · · · · · $\alpha = .10$, $\sigma = .10$; △ — · — · $\alpha = .05$, $\sigma = .10$; △ — — $\alpha = .01$, $\sigma = .10$; * ——— $\alpha = .20$, $\sigma = .20$; * · · · · · $\alpha = .10$, $\sigma = .20$; * — · — · $\alpha = .05$, $\sigma = .20$; * — — $\alpha = .01$, $\sigma = .20$.

computation of $V_{1n}$) can have substantial impact on the results. This may be due in part to the nature of both $l_\alpha$ and $l_{B\alpha}$. It is easy to see that as $n \to \infty$, $l_\alpha / l_{\alpha'} \to 1$ and $l_{B\alpha} / l_{B\alpha'} \to 1$ for any choice $0 < \alpha, \alpha' < 1$! In practice, this means that there is relatively little difference in the lengths of the confidence bands for different $\alpha$. This behavior is quite ev-

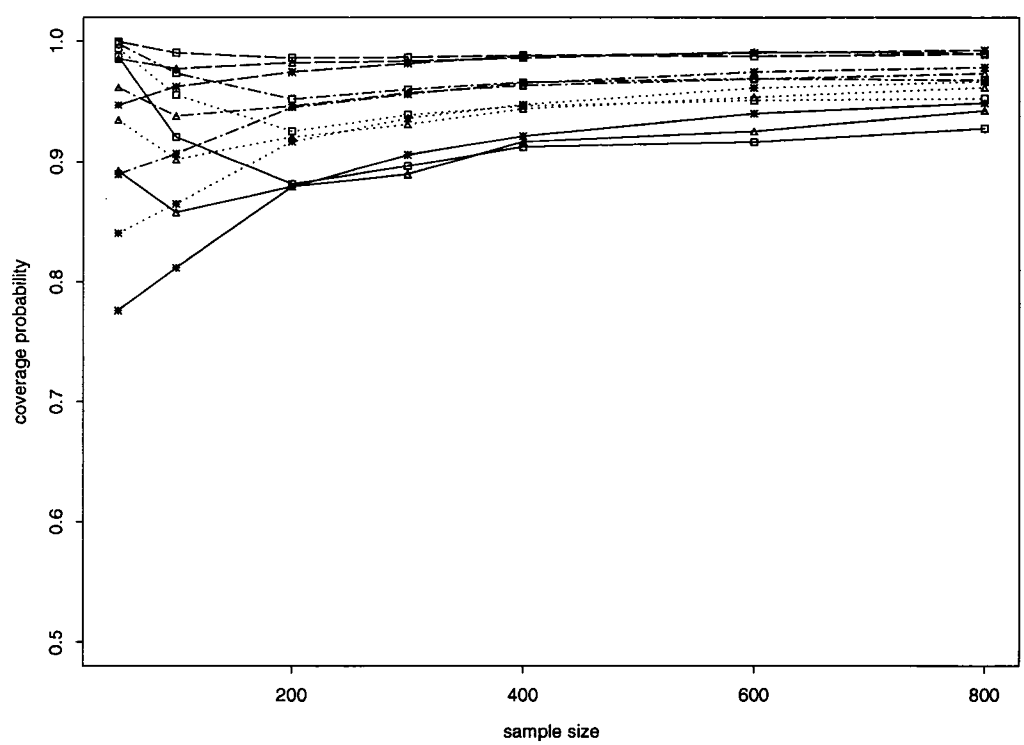
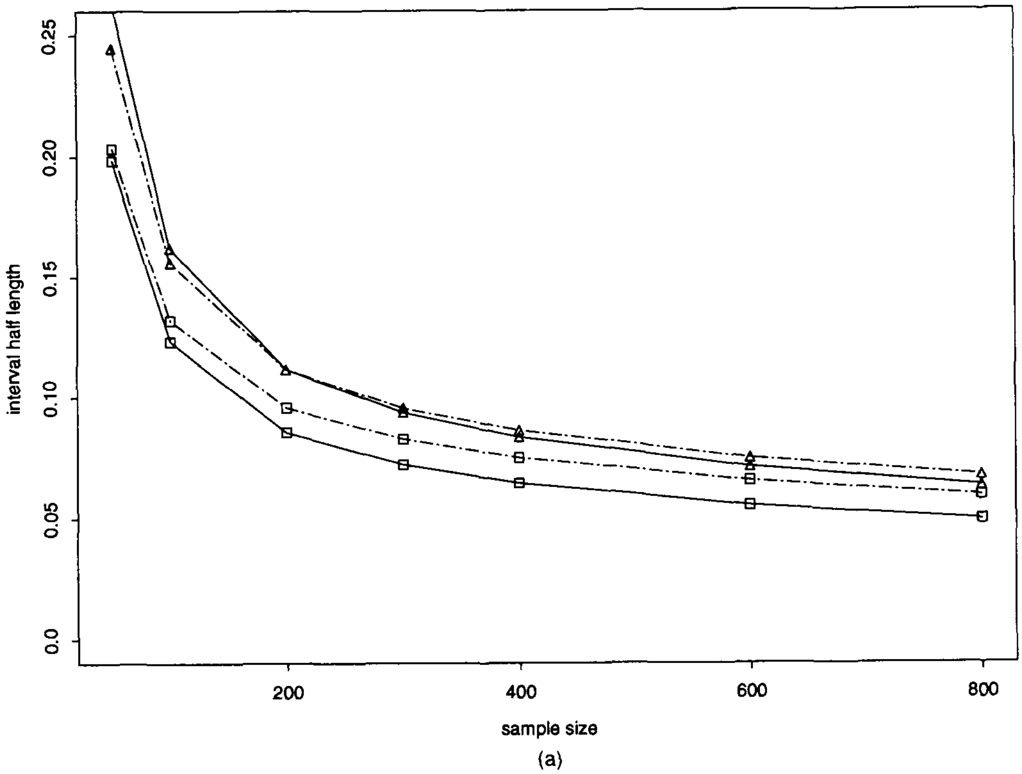



*Figure 8. Bonferroni method empirical coverage probabilities for regression curve $\mu_2$.* □ ——— $\alpha = .20$, $\sigma = .05$; □ · · · · · $\alpha = .10$, $\sigma = .05$; □ — · — · $\alpha = .05$, $\sigma = .05$; □ — — $\alpha = .01$, $\sigma = .05$; △ ——— $\alpha = .20$, $\sigma = .10$; △ · · · · · $\alpha = .10$, $\sigma = .10$; △ — · — · $\alpha = .05$, $\sigma = .10$; △ — — $\alpha = .01$, $\sigma = .10$; * ——— $\alpha = .20$, $\sigma = .20$; * · · · · · $\alpha = .10$, $\sigma = .20$; * — · — · $\alpha = .05$, $\sigma = .20$; * — — $\alpha = .01$, $\sigma = .20$.



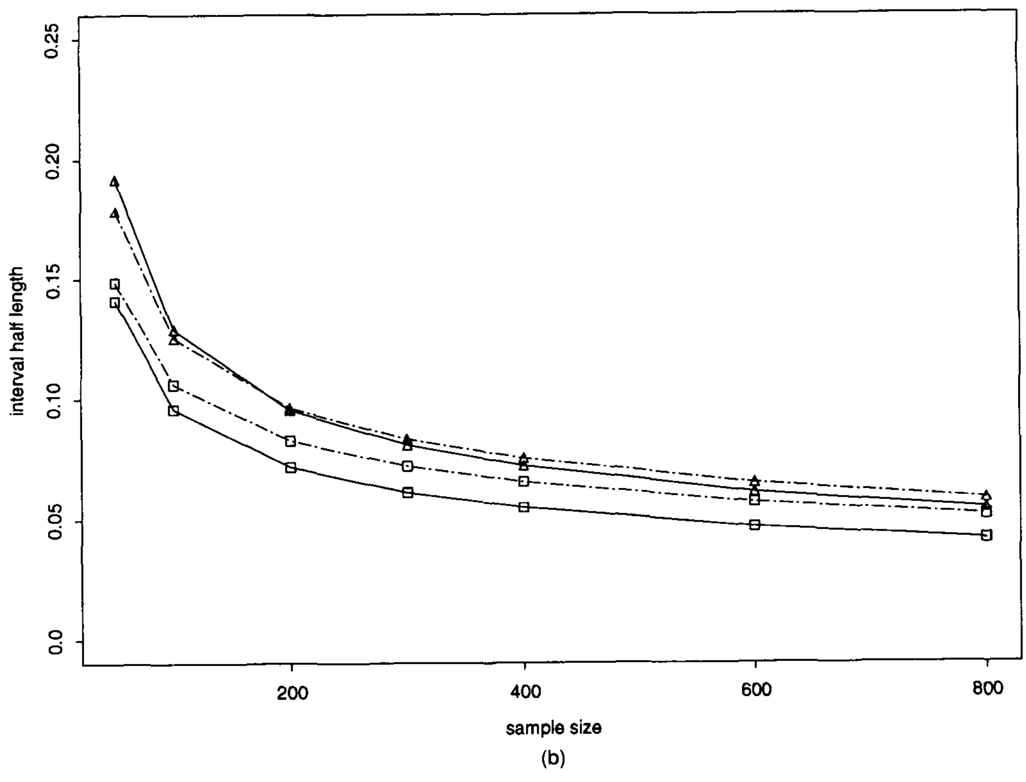




*Figure 9. Confidence band half-lengths with standard errors for (a) regression curve $\mu_1$ and $\sigma = .10$ with $\alpha = .10$ and $\alpha = .01$, and (b) regression curve $\mu_2$ and $\sigma = .10$ with $\alpha = .10$ and $\alpha = .01$. □ ——— bias-corrected, $\alpha = .10$; △ ——— bias-corrected, $\alpha = .01$; □ — · — · Bonferroni, $\alpha = .10$; △ — · — · Bonferroni, $\alpha = .01$.*

ident in Figures 3a and 3b, where the 90% and 99% confidence bands appear quite close together.

All of the foregoing results were obtained under the simplifying assumptions of a periodic regression curve with equally spaced points. Although our theory at this time is not directly applicable for relaxing these assumptions, we have some limited Monte Carlo evidence suggesting that the results presented here should hold in greater generality. Data were generated from model (1) using normal errors with

$$\mu(t) = x + e^{-32(t-.5)^2}$$

to introduce boundary effects, and Gasser–Müller boundary kernels were used for computing $\mu_\lambda(t)$ and the second-derivative estimate for the bias correction. Ordinary "leave-one-out" cross validation was used in place of generalized cross-validation. Because ordinary cross-validation is much more computationally expensive than generalized cross-validation, we were unable to replicate the entire Monte Carlo study. But the limited data we have is reassuring. With equally spaced points, a sample size of 100, $\sigma = .10$, and 5,000 replications, the Bonferroni method had empirical coverage probabilities of .895, for $\alpha = .20$, .935 for $\alpha = .10$, .959 for $\alpha = .05$, and .989 for $\alpha = .01$. The bias correction technique produced empirical coverage probabilities .816 for $\alpha = .20$, .922 for $\alpha = .10$, .968 for $\alpha = .05$, and .998 for $\alpha = .01$. A second simulation with unequally spaced design points

$$t_i = \left(\frac{i - .5}{n}\right)^{1/3}$$

resulted in empirical coverage probabilities .945, .970, .980, and .998 for the Bonferroni method and .901, .964, .990, and 1.000 for the bias-correction method.

## 4. SUMMARY AND CONCLUSIONS

In this article we have studied the properties of two data-driven confidence bands for nonparametric regression curve estimation: a new bias-corrected method and the classical Bonferroni band. Both bands have been shown to have the correct asymptotic coverage behavior, and the Bonferroni band was found to be generally conservative in a simulation study. The Bonferroni band has the advantage of being simple to compute, and it is centered around the standard kernel estimator. No explicit bias correction appears necessary, at least asymptotically. On the other hand, the bias-corrected bands had near-nominal coverage probabilities in all cases for samples even as small as 100, and they were conservative in almost all other cases.

Several important areas for extension of the work in this article have been suggested by the referees and associate editor. These include the situations where we allow for random designs or nonconstant variances in model (1). It seems likely that the conclusion of our theorem can be extended to random $t$'s using results such as those of Johnson (1982) and Härdle (1989) under the proviso that something like (12) can be assumed for the bandwidth estimator. We are unaware of any work that might justify this assumption at present. To extend our work to include nonconstant variances, it seems reasonable to use variance function estimation techniques such as those of Silverman (1985), Gasser et al. (1986), Müller and Stadtmüller (1987), or Cox and O'Sullivan (1990) to replace $\hat{\sigma}$ in (20)–(21) by an estimator $\hat{\sigma}(t)$ of the entire variance curve. This will create new difficulties, such as the problem of selecting a smoothing parameter for the variance curve estimator. In addition, it may not be reasonable to assume that the standard bandwidth selectors for $\mu_\lambda$ will produce a bandwidth satisfying (12) if the errors are heteroscedastic (cf. Andrews 1989). As a result, we cannot recommend the use of our type of methodology in cases of random designs or heteroscedastic errors at this time.

A referee has also pointed out that it seems unreasonable to construct Bonferroni bands at every design point when the sample is large. For a large sample, it would be better to, for example, work with only a subset of the design that was sufficient to give good resolution in plots of the bands. This would reduce the length of the Bonferroni bands and thereby make them less conservative and more competitive with the bias-corrected bands. But this proposal introduces another degree of subjectivity (in effect, another smoothing parameter): the choice of the number of points in the subset. We have not explored this approach but pose it as a problem for future study.

In conclusion, we note that the results presented here apply to the case of a periodic regression curve and equally spaced design points. An open problem is the extension of this work to boundary-corrected estimators and nonuniform designs. Our limited simulation experience is positive, and we expect that a version of our theorem will remain valid if one uses boundary kernels and the Gasser–Müller estimator in such cases, provided that changes in estimator variances over the design are taken into account.

## APPENDIX: PROOF OF THE THEOREM

We begin by establishing some notation that will be used in the proof. Assume without loss of generality that $\sigma = 1$ and $V = 1$. Then set

$$Z_\lambda(t) = \sqrt{n\lambda} \sum_{r=1}^{n} \varepsilon_r \frac{1}{n\lambda} K\left(\frac{t - t_r}{\lambda}\right),$$

$$B_\lambda(t) = \sqrt{n\lambda}\left\{ \sum_{r=1}^{n} \mu(t_r) \frac{1}{n\lambda} K\left(\frac{t - t_r}{\lambda}\right) - \mu(t)\right\},$$

$$a_\lambda = \sqrt{-2 \log \lambda},$$

and

$$d_\lambda = a_\lambda + C/a_\lambda,$$

and for any continuous function $f$ on [0, 1] define

$$\|f\| = \sup_{0 \le t \le 1} |f(t)|.$$

Using this notation, we see that the theorem will be proved when we show that

$$P(a_{\hat{\lambda}}[\|Z_{\hat{\lambda}} + B_{\hat{\lambda}} - \sqrt{n\hat{\lambda}}b_{\hat{\lambda}}\| - d_{\hat{\lambda}}] \le x) \rightarrow e^{-2e^{-x}} \quad \text{(A.1)}$$

for all real $x$.

The basic idea behind the proof of (A.1) is to write

$$a_{\hat{\lambda}}[\|Z_{\hat{\lambda}} + B_{\hat{\lambda}} - \sqrt{n\hat{\lambda}}b_{\hat{\lambda}}\| - d_{\hat{\lambda}}] = a_{\lambda_0}[\|Z_{\lambda_0}\| - d_{\lambda_0}] + e_n, \quad \text{(A.2)}$$

with $\lambda_0$ defined in (11). We show that $a_{\lambda_0}[\|Z_{\lambda_0}\| - d_{\lambda_0}]$ has the required limiting distribution (A.1) in Lemma 1. The remaining lemmas establish that $e_n = o_p(1)$.

*Lemma 1.* $P(a_{\lambda_0}[\|Z_{\lambda_0}\| - d_{\lambda_0}] \le x) \rightarrow e^{-2e^{-x}}$.

*Proof.* By theorem 2.2.4 of Csörgő and Révész (1981), there exists a Wiener process $W_1$ on $[0, \infty)$ such that

$$|S_r - W_1(r)|/\delta_r \overset{\text{a.s.}}{=} O(1), \tag{A.3}$$

with $\delta_r = (r \log\log r)^{1/4}(\log r)^{1/2}$ and $S_r = \sum_{j=1}^r \varepsilon_j$. Thus define

$$Z_{\lambda_0}^{(1)}(t) = (n\lambda_0)^{-3/2} \sum_{r=1}^{n-1} K'\left(\frac{t-t_r}{\lambda_0}\right) W_1(r) + K\left(\frac{t-1}{\lambda_0}\right) W_1(n) \Big/ \sqrt{n\lambda_0}$$

and note that if $S_0 \equiv 0$,

$$\begin{aligned} Z_{\lambda_0}(t) &= (n\lambda_0)^{-1/2} \sum_{r=1}^{n} K\left(\frac{t-t_r}{\lambda_0}\right)(S_r - S_{r-1}) \\ &= (n\lambda_0)^{-1/2}\left\{\sum_{r=1}^{n-1}\left[\frac{t_{r+1}-t_r}{\lambda_0}\right]K'\left(\frac{t-t_r}{\lambda_0}\right)S_r + \frac{1}{2}\sum_{r=1}^{n-1}\left[\frac{t_{r+1}-t_r}{\lambda_0}\right]^2 K''(\xi_r)S_r\right\} \\ &\quad + K\left(\frac{t-1}{\lambda_0}\right)S_n \Big/ \sqrt{n\lambda_0} \\ &= (n\lambda_0)^{-3/2}\sum_{r=1}^{n-1} K'\left(\frac{t-t_r}{\lambda_0}\right)S_r + \frac{(n\lambda_0)^{-5/2}}{2}\sum_{r=1}^{n-1}K''(\xi_r)S_r \\ &\quad + K\left(\frac{t-1}{\lambda_0}\right)S_n \Big/ \sqrt{n\lambda_0} \end{aligned}$$

for mean values $\xi_r$, $r = 1, \ldots, n-1$. Consequently,

$$\begin{aligned} Z_{\lambda_0}(t) - Z_{\lambda_0}^{(1)}(t) &= (n\lambda_0)^{-3/2}\sum_{r=1}^{n-1} K'\left(\frac{t-t_r}{\lambda_0}\right)[S_r - W_1(r)] \\ &\quad + K\left(\frac{t-1}{\lambda_0}\right)[S_n - W_1(n)] \Big/ \sqrt{n\lambda_0} \\ &\quad + \frac{(n\lambda_0)^{-5/2}}{2}\sum_{r=1}^{n-1}K''(\xi_r)S_r. \end{aligned} \tag{A.4}$$

Because of (A.3), the first term in (A.4) is bounded in magnitude by

$$\begin{aligned} &(n\lambda_0)^{-3/2}\max_{1\le r\le n}|S_r - W_1(r)| \sum_{r=1}^{n-1}\left|K'\left(\frac{t-tr}{\lambda_0}\right)\right| \\ &\quad = O_p(\delta_n n^{-6/5})n\left[\int_0^1\left|K'\left(\frac{t-s}{\lambda_0}\right)\right|ds + O\left(\frac{1}{n}\right)\right] = O_p(\delta_n n^{-2/5}), \end{aligned}$$

because $n\lambda_0 = O(n^{4/5})$. Also, $|S_n - W_1(n)|/\sqrt{n\lambda_0} = O_p(\delta_n n^{-2/5})$.

To handle the last term in (A.4), we observe that this quantity is bounded in magnitude by a constant multiple of $\sum_{r=1}^n |S_r|/n^2$. But $E\,\sum_{r=1}^n |S_r| \le \sum_{r=1}^n \sqrt{\operatorname{var}(S_r)} = \sum_{r=1}^n \sqrt{r} = O(n^{3/2})$. Thus, combining all our estimates, we can conclude that $a_{\lambda_0}\|Z_{\lambda_0} - Z_{\lambda_0}^{(1)}\| = o_p(1)$, because $a_{\lambda_0} = O(\sqrt{\log n})$.

To finish the proof, we show that $a_{\lambda_0}[\|Z_{\lambda_0}^{(1)}\| - d_{\lambda_0}]$ and $a_{\lambda_0}[\|Z_{\lambda_0}^{(2)}\| - d_{\lambda_0}]$ have the same limiting distribution with

$$Z_{\lambda_0}^{(2)}(t) = \int_0^{1/\lambda_0} K\left(\frac{t}{\lambda_0} - s\right) dW(s)$$

and $W$ the usual Wiener process on $[0, 1]$. For this purpose, note that

$$\begin{aligned} Z_{\lambda_0}^{(1)}(t) &\overset{d}{=} n^{-1}\lambda_0^{-3/2}\sum_{r=1}^{n-1}K'\left(\frac{t-t_r}{\lambda_0}\right)W\left(\frac{r}{n}\right) + K\left(\frac{t-1}{\lambda_0}\right)W(1)\Big/\sqrt{\lambda_0} \\ &= \lambda_0^{-3/2}\int_0^1 K'\left(\frac{t-u}{\lambda_0}\right)W(u)\,du + K\left(\frac{t-1}{\lambda_0}\right)W(1)\Big/\sqrt{\lambda_0} \\ &\quad + O_p(n^{3/10}\sqrt{\log n}/\sqrt{n}), \end{aligned}$$

using the modulus of continuity of the Wiener process. We have used the notation "$\overset{d}{=}$" here to indicate that two quantities have the same distribution.

Now integrate by parts to see that

$$\begin{aligned} &\lambda_0^{-3/2}\left\{\int_0^1 K'\left(\frac{t-u}{\lambda_0}\right)W(u)\,du + \lambda_0 K\left(\frac{t-1}{\lambda_0}\right)W(1)\right\} \\ &\quad \overset{d}{=} \lambda_0^{-1}\left\{\int_0^1 K'\left(\frac{t-u}{\lambda_0}\right)W\left(\frac{u}{\lambda_0}\right)du + \lambda_0 K\left(\frac{t-1}{\lambda_0}\right)W\left(\frac{1}{\lambda_0}\right)\right\} \\ &\quad \overset{d}{=} Z_{\lambda_0}^{(2)}(t). \end{aligned}$$

To complete the proof, apply corollary A.1 of Bickel and Rosenblatt (1973) to $Z_{\lambda_0}^{(2)}$.

*Note Added In Proof.* We recently became aware that Stadtmüller (1986) used the technique of Lemma 1 to obtain similar results.

It remains to show that $e_n$ is $o_p(1)$ in (A.2). We can express $e_n$ as

$$\begin{aligned} e_n &= (a_{\hat\lambda} - a_{\lambda_0})[\|Z_{\lambda_0}\| - d_{\lambda_0}] + a_{\hat\lambda}(d_{\lambda_0} - d_{\hat\lambda}) \\ &\quad + a_{\hat\lambda}[\|Z_{\hat\lambda}\| - \|Z_{\lambda_0}\|] + a_{\hat\lambda}[\|Z_{\hat\lambda} + B_{\hat\lambda} - \sqrt{n\hat\lambda}b_{\hat\lambda}\| - \|Z_{\hat\lambda}\|]. \end{aligned}$$

Each term in this sum is dealt with in turn by the lemmas that follow.

*Lemma 2.* $(a_{\hat\lambda} - a_{\lambda_0})[\|Z_{\lambda_0}\| - d_{\lambda_0}] = o_p(1)$.

*Proof.* Under our assumptions, $|a_{\hat\lambda} - a_{\lambda_0}| = O_p(n^{-1/10}/\sqrt{\log n})$ and, from Lemma 1, $\|Z_{\lambda_0}\| - d_{\lambda_0} = O_p(1/\sqrt{\log n})$.

*Lemma 3.* $a_{\hat\lambda}(d_{\lambda_0} - d_{\hat\lambda}) = o_p(1)$.

*Proof.* The result follows on noting that $a_{\hat\lambda} = O_p(\sqrt{\log n})$ and $d_{\lambda_0} - d_{\hat\lambda} = O_p(n^{-1/10}/\sqrt{\log n})$.

*Lemma 4.* $a_{\hat\lambda}|\|Z_{\hat\lambda}\| - \|Z_{\lambda_0}\|| = o_p(1)$.

*Proof.* Because $a_{\hat\lambda} = O_p(\sqrt{\log n})$ and $|\|Z_{\hat\lambda}\| - \|Z_{\lambda_0}\|| \le \|Z_{\hat\lambda} - Z_{\lambda_0}\|$, it suffices to show that $\|Z_{\hat\lambda} - Z_{\lambda_0}\| = O_p(n^{-\theta})$ for some $\theta > 0$. To prove this, we require two steps. First, we show that

$$\|Z_{\hat\lambda} - Z_{\lambda_0}\| - \max_{1\le r\le n}|Z_{\hat\lambda}(t_r) - Z_{\lambda_0}(t_r)| = O_p(n^{-\theta}). \tag{A.5}$$

Second, we show that $\max_{1\le r\le n}|Z_{\hat\lambda}(t_r) - Z_{\lambda_0}(t_r)| = O_p(n^{-\theta})$.

Set $t_0 = 0$, $t_{n+1} = 1$ and let $I_r = [t_{r-1}, t_r]$, $r = 1, \ldots, n$. Then

$$\begin{aligned} \|Z_{\hat\lambda} - Z_{\lambda_0}\| &= \max_r \sup_{t\in I_r}|Z_{\hat\lambda}(t) - Z_{\lambda_0}(t)| \\ &\le \max_r|Z_{\hat\lambda}(t_r) - Z_{\lambda_0}(t_r)| + \max_r\sup_{t\in I_r}|Z_{\hat\lambda}(t) - Z_{\hat\lambda}(t_r)| \\ &\quad + \max_r\sup_{t\in I_r}|Z_{\lambda_0}(t) - Z_{\lambda_0}(t_r)|. \end{aligned}$$

But for any $\lambda = O_p(n^{-1/5})$ and $t \in I_r$, there are mean values $\xi_{jr}$ such that

$$|Z_\lambda(t) - Z_\lambda(t_r)| \le \frac{1}{\sqrt{n\lambda}} \sum_{j=1}^{n} |K'(\xi_{jr})| \left|\frac{t - t_r}{\lambda}\right| |\varepsilon_j|$$
$$\le \sup_u |K'(u)| O_p(\max_j |\varepsilon_j| n^{-2/5}).$$

Here we have used the finite support property of the kernel, which ensures that there are only $O(n\lambda)$ terms in the sum. Because the $\varepsilon$'s have $19 + \zeta$ absolute moments for some $\zeta > 0$, it is known that $\max_j |\varepsilon_j|$ cannot grow at a faster rate in probability than $n^{1/(19+\zeta)}$. Thus (A.5) has been verified.

To deal with $\max_r |Z_{\hat\lambda}(t_r) - Z_{\lambda_0}(t_r)|$, we first point out that, due to (12), it suffices to restrict our study to the set of bandwidths $\Lambda_n = \{\lambda: |\lambda - \lambda_0|/\lambda_0 \le n^{-\gamma}\}$ for any $\gamma < 1/10$. The proof uses a partitioning argument for $\Lambda_n$ similar to that of Härdle et al. (1988).

Set

$$T(t; \lambda, \lambda') = \sum_{r=1}^{n} \varepsilon_r \left[K\left(\frac{t - t_r}{\lambda}\right) - K\left(\frac{t - t_r}{\lambda'}\right)\right].$$

Then

$$Z_{\hat\lambda}(t) - Z_{\lambda_0}(t) = \frac{1}{\sqrt{n\hat\lambda}} T(t; \hat\lambda, \lambda_0) + \left(\frac{1}{\sqrt{n\hat\lambda}} - \frac{1}{\sqrt{n\lambda_0}}\right)\sqrt{n\lambda_0} Z_{\lambda_0}(t).$$

The last term is $O_p(n^{-1/10}\sqrt{\log n})$ by Lemma 1 and (12). Thus we must show that $n^{-2/5}\max_r |T(t_r; \hat\lambda, \lambda_0)| = O_p(n^{-\theta})$. To accomplish this, we prove that for some $\theta > 0$,

$$n^{-2/5} \max_r \sup_{\lambda \in \Lambda_n} |T(t_r; \lambda, \lambda_0)| = O_p(n^{-\theta}).$$

Partition $\Lambda_n$ into $n^\rho$ subsets $\Lambda_{jn}, j = 1, \ldots, n^\rho$, with the centerpoint $\lambda_j$ of $\Lambda_{jn}$ satisfying $|\lambda - \lambda_j|/\lambda_0 \le n^{-1}$ if $\lambda \in \Lambda_{jn}$. Then

$$\sup_{\lambda \in \Lambda_n} |T(t_r; \lambda, \lambda_0)| \le \max_j |T(t_r; \lambda_j, \lambda_0)| + \max_j \sup_{\lambda \in \Lambda_{jn}} |T(t_r; \lambda, \lambda_j)|.$$

Now

$$\left|K\left(\frac{t - t_r}{\lambda}\right) - K\left(\frac{t - t_r}{\lambda'}\right)\right| \le \sup |uK'(u)|\, |\lambda - \lambda'|/\min(\lambda, \lambda').$$

Thus, $\sup_{\lambda \in \Lambda_{jn}} |T(t_r; \lambda, \lambda_j)| \le \max_r |\varepsilon_r| O(n^{-1/5})$ uniformly in $j$ and $r$. Consequently,

$$n^{-2/5} \max_r \max_j \sup_{\lambda \in \Lambda_{jn}} |T(t_r; \lambda, \lambda_j)| = O_p(n^{-3/5} \max_r |\varepsilon_r|).$$

Finally, use the Bonferroni and Markov inequalities to see that for any $\tau > 0$,

$$P(n^{-2/5} \max_r \max_j |T(t_r; \lambda_j, \lambda_0)| > \tau)$$
$$\le \sum_{r=1}^{n} \sum_{j=1}^{n^\rho} \frac{E|n^{-2/5}T(t_r; \lambda_j, \lambda_0)|^s}{\tau^s}. \quad \text{(A.6)}$$

By Theorem 2 of Whittle (1960)

$$E|n^{-2/5}T(t_r; \lambda_j, \lambda_0)|^s$$
$$\le C_\varepsilon \left\{n^{-4/5} \sum_{k=1}^{n} \left(K\left(\frac{t_r - t_k}{\lambda_j}\right) - K\left(\frac{t_r - t_k}{\lambda_0}\right)\right)^2\right\}^{s/2} = O(n^{-s\gamma}),$$

with $C_\varepsilon$ a constant depending only on $s$ and $E|\varepsilon_1|^s$. This bound is seen to be uniform in $r$ which shows that the bound (A.6) is $O(n^{\rho+1-s\gamma})$. One may check that $\rho = 1 - \gamma$, so the bound converges to zero for $s > (2/\gamma) - 1$. As $\gamma < 1/10$ is arbitrary, the result follows if the $\varepsilon$'s have $19 + \zeta$ moments for any $\zeta > 0$.

Because

$$|\,\|Z_{\hat\lambda} + B_{\hat\lambda} - \sqrt{n\hat\lambda} b_{\hat\lambda}\| - \|Z_{\hat\lambda}\|\,|$$
$$\le \|B_{\hat\lambda} - \sqrt{n\hat\lambda} b_{\hat\lambda}\|$$
$$\le \|B_{\hat\lambda} - B_{\lambda_0}\| + \|\sqrt{n\hat\lambda} b_{\hat\lambda} - \sqrt{n\lambda_0} b_{\lambda_0}\| + \|B_{\lambda_0} - \sqrt{n\lambda_0} b_{\lambda_0}\|,$$

the proof of the theorem will be completed when we show each term in the upper bound is $O_p(n^{-\theta})$. Again we establish this through a set of lemmas.

*Lemma 5.* $\|B_{\hat\lambda} - B_{\lambda_0}\| = O_p(n^{-\theta})$ for some $\theta > 0$.

*Proof.* Define $\Lambda_n$ as in Lemma 4 and note that for $\lambda \in \Lambda_n$,

$$B_\lambda(t)/\sqrt{n\lambda} = \lambda^2 B\mu''(t) + O(n^{-(2+\nu)/5}) \quad \text{(A.7)}$$

uniformly in $t$. Thus for $\hat\lambda \in \Lambda_n$, $B_{\hat\lambda}(t) - B_{\lambda_0}(t) = O_p(n^{-\gamma}) + O(n^{-\nu/5})$.

*Lemma 6.* $\|\sqrt{n\hat\lambda} b_{\hat\lambda} - \sqrt{n\lambda_0} b_{\lambda_0}\| = O_p(n^{-\theta})$ for some $\theta > 0$.

*Proof.* We can write

$$[\sqrt{n\hat\lambda} b_{\hat\lambda}(t) - \sqrt{n\lambda_0} b_{\lambda_0}(t)]/B = \sum_{r=1}^{n} w_r(t)\varepsilon_r + \sum_{r=1}^{n} w_r(t)\mu(t_r),$$

with

$$w_r(t) = \frac{\sqrt{n\hat\lambda}\hat\lambda^2}{n\tilde\lambda^3} K^*\left(\frac{t - t_r}{\tilde\lambda}\right) - \frac{\sqrt{n\lambda_0}\lambda_0^2}{n\tilde\lambda_0^3} K^*\left(\frac{t - t_r}{\tilde\lambda_0}\right)$$

for $\tilde\lambda_0 = \lambda_0^{5/7}$. The last term in this expression can be handled using arguments similar to those for Lemma 5. The first term can be expressed as

$$\sqrt{\tilde\lambda} \sum_{r=1}^{n} \left[K^*\left(\frac{t - t_r}{\tilde\lambda}\right) - K^*\left(\frac{t - t_r}{\tilde\lambda_0}\right)\right]\varepsilon_r \Big/ \sqrt{n}$$
$$+ (\sqrt{\tilde\lambda} - \sqrt{\tilde\lambda_0}) \sum_{r=1}^{n} K^*\left(\frac{t - t_r}{\tilde\lambda_0}\right)\varepsilon_r \Big/ \sqrt{n}.$$

The two terms in this sum can now be treated using parallels of the arguments for Lemmas 4 and 1. We omit the details.

*Lemma 7.* $\|B_{\lambda_0} - \sqrt{n\lambda_0} b_{\lambda_0}\| = O_p(n^{-\theta})$.

*Proof.* This result is a consequence of (A.7) and arguments similar to those used in proving Lemma 1.

*[Received February 1992. Revised October 1992.]*

## REFERENCES


Andrews, D. K. W. (1989), "Asymptotic Optimality of Generalized $C_p$, Cross-Validation and Generalized Cross-Validation in Regression With Heteroskedastic Errors," Cowles Foundation Discussion Paper 906, Yale University.

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics,* 1, 1071–1095.

Bjerve, S., Doksum, K. A., and Yandell, B. S. (1985), "Uniform Confidence Bands for Regression Based on a Simple Moving Average," *Scandinavian Journal of Statistics,* 12, 159–169.

Cox, D. D. (1986), "Approximation Theory of Method of Regularization Estimators: Applications," in *Contemporary Mathematics Volume 59: Function Estimates,* ed. J. S. Marron, Providence, RI: American Mathematical Society, pp. 105–124.

Cox, D. D., and O'Sullivan, F. (1990), "Analysis of Penalized Likelihood-Type Estimators With Application to the Generalized Smoothing Problem in Sobolev Spaces," *The Annals of Statistics,* 18, 1676–1695.

Csörgő, M., and Révész, P. (1981), *Strong Approximations in Probability and Statistics,* New York: Academic Press.

Gasser, T., Müller, H.-G., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society,* Ser. B, 47, 238–252.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986), "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika,* 73, 625–633.
Hall, P., Kay, J. W., and Titterington, D. M. (1990), "Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression," *Biometrika,* 77, 521–528.
Hall, P., and Marron, S. (1990), "On Variance Estimation in Nonparametric Regression," *Biometrika,* 77, 415–419.
Hall, P., and Titterington, D. M. (1988), "On Confidence Bands in Nonparametric Density Estimation and Regression," *Journal of Multivariate Analysis,* 27, 228–254.
Härdle, W. (1989), "Asymptotic Maximal Deviation of *M*-Smoothers," *Journal of Multivariate Analysis,* 29, 163–179.
——— (1990), *Applied Nonparametric Regression,* Cambridge, U.K.: Cambridge University Press.
Härdle, W., and Bowman, A. W. (1988), "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association,* 83, 102–110.
Härdle, W., Hall, P., and Marron, J. S. (1988), "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?" (with discussion), *Journal of the American Statistical Association,* 83, 86–95.
Härdle, W., and Marron, S. (1991), "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *The Annals of Statistics,* 19, 778–796.
Knafl, G., Sacks, J., and Ylvisaker, D. (1982), "Model Robust Confidence Intervals," *Journal of Statistical Planning and Inference,* 6, 319–334.
——— (1985), "Confidence Bands for Regression Functions," *Journal of the American Statistical Association,* 80, 682–692.
Johnston, G. J. (1982), "Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates," *Journal of Multivariate Analysis,* 12, 402–414.
Li, K. C. (1989), "Honest Confidence Regions for Nonparametric Regression," *The Annals of Statistics,* 17, 1001–1008.
Müller, H.-G. (1988), *Nonparametric Regression Analysis of Longitudinal Data,* New York: Springer-Verlag.
Müller, H.-G., and Stadtmüller, U. (1987), "Estimation of Heteroscedasticity in Regression Analysis," *The Annals of Statistics,* 15, 610–625.
Nychka, D. (1988), "Bayesian Confidence Intervals for Smoothing Splines," *Journal of the American Statistical Association,* 83, 1134–1143.
——— (1990), "The Average Posterior Variance of a Smoothing Spline and a Consistent Estimate of the Average Squared Error," *The Annals of Statistics,* 18, 415–428.
Rice, J. (1984), "Bandwidth Choice for Nonparametric Regression," *The Annals of Statistics,* 12, 1215–1230.
Révész, P. (1979), "On the Nonparametric Estimation of the Regression Function," *Problems of Control and Information Theory,* 8, 297–302.
Silverman, B. (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 47, 1–52.
Stadtmüller, U. (1986), "Asymptotic Properties of Nonparametric Curve Estimates," *Periodica Mathematica Hungarica,* 17(2), 83–108.
Wahba, G. (1983), "Bayesian 'Confidence Intervals' for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society,* Ser. B, 45, 133–150.
Whittle, P. (1960), "Bounds for the Moments of Linear and Quadratic Forms in Independent Variables," *Theory of Probability and Its Applications,* 5, 302–305.