The Annals of Statistics
1992, Vol. 20, No. 2, 695–711

ON BOOTSTRAP CONFIDENCE INTERVALS IN NONPARAMETRIC REGRESSION[1]

By Peter Hall

University of Glasgow and Australian National University

Several authors have developed bootstrap methods for constructing confidence intervals in nonparametric regression. On each occasion a nonpivotal approach has been employed. Nonpivotal methods are still the overwhelmingly popular choice when statisticians use the bootstrap to compute confidence intervals, but they are not necessarily the most appropriate. In this paper we point out some of the theoretical advantages of pivoting. They include a reduction in the error of the bootstrap distribution function estimate, from $n^{-1/2}$ to $n^{-1}h^{-1/2}$ (where h denotes bandwidth); and a reduction in coverage error of confidence intervals, from either $n^{-1/2}h^{-1/2}$ or $n^{-1/2}h^{1/2}$ (depending on which nonpivotal method is used) to n^{-1}. Several comparisons are drawn with the case of nonparametric density estimation, where a pivotal approach also reduces errors associated with confidence intervals, but where the orders of magnitude of the respective errors are quite different from their counterparts for nonparametric regression.

1. Introduction. Among users of the bootstrap for constructing confidence intervals there has developed a debate over relative merits of pivotal and nonpivotal methods. A statistic is (asymptotically) pivotal if its large-sample distribution does not depend on unknowns. The discussion papers of DiCiccio and Romano [3], Hall [10] and Hinkley [13] describe the main issues in the debate. In simple problems, such as estimation of a mean, nonpivotal methods usually require subsidiary corrections if they are to achieve the accuracy of pivotal methods. However, it is still true that nonpivotal methods, devoid of corrections, enjoy by far the greatest following among statisticians. Without exception, published accounts of bootstrap methods for constructing confidence intervals in nonparametric regression use nonpivotal techniques (Härdle and Bowman [7], Härdle [5], Härdle and Marron [8]). In this paper we draw attention to some of the theoretical advantages of a pivotal approach in the context of nonparametric regression.

The confidence interval problem for nonparametric regression falls naturally into two parts, the first being construction of a confidence interval for the expected value of the estimator and the second involving bias correction. In

Received November 1989; revised October 1990.

[1]Work supported by Visiting Fellowship Research Grant from the U.K. Science and Engineering Research Council.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20.

Key words and phrases. Bootstrap, confidence interval, coverage error, density estimation, Edgeworth expansion, nonparametric regression, percentile method, percentile-t, simultaneous confidence interval.

particular, each of the papers [5, 7, 8] treats the confidence interval problem in this dichotomous way. We shall discuss both parts of the problem. The effect of bias depends very much on how bias is corrected and there are differing views among statisticians as to how this should be done. We shall treat two different approaches: explicit bias correction and undersmoothing.

To simplify exposition, we shall treat only the case of a single confidence interval, although the properties which we shall relate apply without change to any fixed number of confidence intervals. Thus, the advantages of pivotal methods are available for simultaneous confidence intervals. However, technology for generalizing the results to simultaneous confidence bands is still in its infancy and the level of detail provided in this paper is not yet available for that context.

To enable further simplification, we shall assume that the errors are homoscedastic. Identical results may be derived in the heteroscedastic case, provided the variance function admits a parametric model. However, one cannot obtain the same convergence rates if the error structure can only be modelled nonparametrically. The wild bootstrap suggested by Härdle [5] was developed to handle the latter case.

Before describing our results, it will be helpful to list, for the sake of comparison, the main properties of bootstrap methods in classical finite-parameter problems. There, the principal pivotal and nonpivotal methods are percentile-t and percentile, respectively. The reader is referred to DiCiccio and Romano [3], Hall [10] and Hinkley [13] for details. (i) The bootstrap estimates the distribution of a pivotal statistic with accuracy n^{-1} (in probability) and of a nonpivotal statistic with accuracy $n^{-1/2}$, where n is sample size. (ii) Use of pivotal methods to construct confidence intervals results in coverage errors of size n^{-1} for both one- and two-sided intervals. On the other hand, the coverage errors are $n^{-1/2}$ in the case of nonpivotal methods and one-sided intervals.

The analogues of these properties in the case of kernel-type nonparametric regression are as follows; we use h to denote bandwidth of the estimator. (i) The bootstrap estimates the distribution of a pivotal statistic with accuracy $n^{-1}h^{-1/2}$ and of a nonpivotal statistic with accuracy $n^{-1/2} + n^{-1}h^{-1/2}$. When $h \asymp n^{-1/5}$, as is typically the case in practice with second-order kernels, the respective errors are $n^{-9/10}$ and $n^{-1/2}$; the former, available from pivotal methods, is smaller. (ii) Use of pivotal methods to construct confidence intervals results in coverage errors of size n^{-1} for both one- and two-sided intervals. Coverage errors can be maintained at this level, even after bias correction. On the other hand, the errors are at least $(h/n)^{1/2}$ in the case of nonpivotal methods and one-sided intervals. When $h \asymp n^{-1/5}$, the respective errors are of size n^{-1} and $n^{-3/5}$ and so pivotal methods have an advantage once again. Furthermore, the coverage error can be as poor as $(nh)^{-1/2}$ in the case of some nonpivotal methods.

The striking aspect of these conclusions is that pivotal methods perform so well, achieving coverage accuracy of order n^{-1} even in the infinite-parameter problem of nonparametric regression. The reason is that the standard deviation estimate used to pivot the estimated regression mean is $\sqrt{n}$-consistent, although the mean estimate itself typically converges at a much slower rate.

This disparity between the convergence rates of numerator and denominator of the pivotal statistic is also responsible for several other unusual properties. In particular, Edgeworth expansions of distributions of studentized and nonstudentized versions of the regression mean agree to first order, that is, to order $(nh)^{-1/2}$. This property fails in most classical problems, for example, in the case of estimating a mean. It also fails in the case of nonparametric density estimation, where distributions of pivotal and nonpivotal statistics *differ* in terms of order $(nh)^{-1/2}$. For nonparametric regression, the first point of difference between the distributions of pivotal and nonpivotal statistics is in terms of $(h/n)^{1/2}$, which is smaller than $(nh)^{-1/2}$. However, the distributions agree in terms of order $(nh)^{-1}$, which is usually smaller than $(h/n)^{1/2}$.

Section 2 will describe the main arguments behind the conclusions drawn above and Section 3 will treat the case of bias correction. As many technical details as possible will be deferred to Sections 4 and 5, which will state formal Edgeworth expansions and provide proofs, respectively. During Section 2 we shall, where instructive, draw comparisons with the case of nonparametric density estimation, which may be treated using somewhat similar methods.

2. Main ideas. We take the regression model to be

$$Y_i = g(x_i) + e_i, \qquad 1 \le i \le n, \tag{2.1}$$

where g is an unknown smooth function, $x_1, \dots, x_n$ are design points confined to a given interval, and $e_1, \dots, e_n$ are independent and identically distributed errors with zero mean and variance σ^2. Our estimator of g, using kernel K and bandwidth h, is

$$\hat{g}(x) = \left[\sum_{i=1}^{n} Y_i K\left\{\frac{x - x_i}{h}\right\}\right] \bigg/ \left[\sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\}\right].$$

See Härdle [6] for an excellent account of the general properties of such estimators.

The variance of $\hat{g}(x)$ is

$$\gamma_x^2 = \operatorname{var} \hat{g}(x) = \sigma^2 \tau_x^2 \left[\sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\}\right]^{-2} \equiv \sigma^2 \beta_x^2, \tag{2.2}$$

say, where

$$\tau_x^2 = \sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\}^2.$$

Estimation of γ_x^2 demands that we estimate σ^2 and for that purpose we shall employ the difference method described by Rice [17], Gasser, Sroka and

Jennen-Steinmetz [4], Müller and Stadtmüller [15, 16] and Müller [14]. Specifically, let $\{d_j\}$ be a sequence of numbers with the properties

$$\sum d_j = 0,\ \sum d_j^2 = 1,\ d_j = 0 \text{ for } j < -m_1 \text{ or } j > m_2,\ d_{-m_1} d_{m_2} \neq 0,$$

where $m_1, m_2 \geq 0$. Put $m = m_1 + m_2$. Assume that the sample $\mathscr{X} = \{(x_i, Y_i), 1 \leq i \leq n\}$ has been ordered such that $x_1 < \cdots < x_n$. Then our variance estimator is

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=m_1+1}^{n-m_2} \left(\sum_j d_j Y_{i+j} \right)^2.$$

Our estimator of γ_x^2 is obtained by replacing σ^2 by $\hat{\sigma}^2$ in formula (2.2):

$$\hat{\gamma}_x^2 = \hat{\sigma}^2 \beta_x^2.$$

The "ordinary" and "studentized" versions of $\hat{g} - E\hat{g}$, both standardized for scale, are

$$S = S(x) = \frac{\hat{g}(x) - E\hat{g}(x)}{\gamma_x} = \frac{1}{\tau_x \sigma} \sum_{i=1}^{n} e_i K\left\{ \frac{x - x_i}{h} \right\}, \tag{2.3}$$

$$T = T(x) = \frac{\hat{g}(x) - E\hat{g}(x)}{\hat{\gamma}_x} = \frac{1}{\tau_x \hat{\sigma}} \sum_{i=1}^{n} e_i K\left\{ \frac{x - x_i}{h} \right\}, \tag{2.4}$$

respectively. Our conclusions about pivotal and nonpivotal forms of the bootstrap hinge on differences between the distributions of S and T. To describe these differences, write $\mu_j = E\{(e_1/\sigma)^j\}$ for the jth standardized moment of the error distribution and let Φ and ϕ denote the standard normal distribution and density functions, respectively. Then

$$\begin{aligned} P(S \leq u) = {} & \Phi(u) + (nh)^{-1/2} \mu_3 p_1(u) \phi(u) \\ & + (nh)^{-1} \{(\mu_4 - 3) p_2(u) + \mu_3^2 p_3(u)\} \phi(u) \\ & + O\{n^{-1} + (nh)^{-3/2}\}, \end{aligned} \tag{2.5}$$

$$\begin{aligned} P(T \leq u) = {} & \Phi(u) + (nh)^{-1/2} \mu_3 p_1(u) \phi(u) \\ & + (nh)^{-1} \{(\mu_4 - 3) p_2(u) + \mu_3^2 p_3(u)\} \phi(u) \\ & + (h/n)^{1/2} \mu_3 p_4(u) \phi(u) + O\{n^{-1} + (nh)^{-3/2}\}, \end{aligned} \tag{2.6}$$

where $p_1, \ldots, p_4$ are known polynomials whose coefficients are bounded and depend only on the design points, not at all on the error distribution. (Formulae for the polynomials will be given in Section 3.)

The first conclusion to be drawn from (2.5) and (2.6) is that the distributions of S and T agree to first and third order, that is, in terms of sizes

$(nh)^{-1/2}$ and $(nh)^{-1}$, but differ to second order, that is, in terms of size $(h/n)^{1/2}$. Indeed, the distribution of S does not contain terms of order $(h/n)^{1/2}$, whereas the distribution of T does. Interestingly, in the analogue of (2.4) and (2.5) for a nonparametric density estimator, the terms of size $(nh)^{-1/2}$ differ. That is, the distributions of ordinary and studentized forms of the estimator differ to first order (and also to second order). This property is the norm in classical applications of the bootstrap [10] and establishing it for density estimators is straightforward.

Next we examine bootstrap versions of S and T and of the formulae (2.5) and (2.6). Observe from (2.3) and (2.4) that the mean function g does not influence the distribution of S and enters the distribution of T only through $\hat{\sigma}$. It turns out that the effect of g and $\hat{\sigma}$ is relatively minor; in particular, g enters only the $O(n^{-1})$ term on the right-hand side of (2.6). Therefore, when using the bootstrap to approximate these distributions we shall, in effect, make the fictitious assumption that $g \equiv O$. To estimate the error distribution, first compute the simple residuals

$$\tilde{e}_i = Y_i - \hat{g}(x_i), \qquad i \in \mathscr{I},$$

where $\mathscr{I}$ is an appropriate set of indices (e.g., the set of i's such that x_i is not too close to the boundary of the interval on which inference is being conducted; see Härdle and Bowman [7]). Define n' to equal the number of elements of $\mathscr{I}$, let Σ'_i denote summation over $i \in \mathscr{I}$ and put

$$\bar{\tilde{e}} = n'^{-1} \sum_i{}' \tilde{e}_i, \qquad \hat{e}_i = \tilde{e}_i - \bar{\tilde{e}},$$

the latter being centered residuals. Conditional on the sample $\mathscr{X} = \{(x_i, Y_i), 1 \le i \le n\}$, draw a resample $\{e_1^*, \dots, e_n^*\}$ at random, with replacement, from $\{\hat{e}_i, i \in \mathscr{I}\}$. Define

$$\check{\sigma}^2 = \frac{1}{n'} \sum_i{}' \hat{e}_i^2, \qquad \hat{\sigma}^{*2} = \frac{1}{n-m} \sum_{j=m_1+1}^{n-m_2} \left(\sum_j d_j e_{i+j}^* \right)^2,$$

$$S^* = \frac{1}{\tau_x \check{\sigma}} \sum_{i=1}^{n} e_i^* K\left\{ \frac{x - x_i}{h} \right\},$$

$$T^* = \frac{1}{\tau_x \hat{\sigma}^*} \sum_{i=1}^{n} e_i^* K\left\{ \frac{x - x_i}{h} \right\}.$$

The conditional distributions of S^* and T^*, given $\mathscr{X}$, are good approximations to the distributions of S and T, respectively. Indeed, if we define

$$\hat{\mu}_j = E\left\{ (e_1^*/\check{\sigma})^j \middle| \mathscr{X} \right\} = \check{\sigma}^{-j} n'^{-1} \sum_i{}' \hat{e}_i^j,$$

then we have the following bootstrap analogues of (2.5) and (2.6):

$$P(S^* \le u|\mathscr{X}) = \Phi(u) + (nh)^{-1/2}\hat{\mu}_3 p_1(u)\phi(u) + (nh)^{-1}\{(\hat{\mu}_4 - 3)p_2(u) + \hat{\mu}_3^2 p_3(u)\}\phi(u) + O_p\{n^{-1} + (nh)^{-3/2}\}, \tag{2.7}$$

$$P(T^* \le u|\mathscr{X}) = \Phi(u) + (nh)^{-1/2}\hat{\mu}_3 p_1(u)\phi(u) + (nh)^{-1}\{(\hat{\mu}_4 - 3)p_2(u) + \hat{\mu}_3^2 p_3(u)\}\phi(u) + (h/n)^{1/2}\hat{\mu}_3 p_4(u)\phi(u) + O_p\{n^{-1} + (nh)^{-3/2}\}. \tag{2.8}$$

The polynomials $p_1, \dots, p_4$ in (2.7) and (2.8) are exactly as they were in (2.5) and (2.6).

In virtually all cases of interest, $(nh)^{-3/2}$ is of smaller order than n^{-1}; consider, for example, the most common circumstance where h is of size $n^{-1/5}$. Therefore the remainders $O\{n^{-1} + (nh)^{-3/2}\}$ in (2.5)–(2.8) are, in reality, $O(n^{-1})$. Furthermore, it is usually the case that $\hat{\mu}_j = \mu_j + O_p(n^{-1/2})$; see Theorem 2.1 at the end of this section. We may therefore deduce from (2.5)–(2.8) that

$$P(S^* \le u|\mathscr{X}) - P(S \le u) = O_p(n^{-1}h^{-1/2}),$$

$$P(T^* \le u|\mathscr{X}) - P(T \le u) = O_p(n^{-1}h^{-1/2}).$$

These are precise rates of convergence, not simply upper bounds, because the convergence rate of $\hat{\mu}_3$ to μ_3 is precisely $n^{-1/2}$. Thus, the exact rate of convergence of bootstrap approximations to the distributions of S and T is $n^{-1}h^{-1/2}$, which is slightly poorer than the rate n^{-1} found in classical finite parameter problems, but slightly better than the corresponding rate $(nh)^{-1}$ found in applications of the bootstrap to density estimation.

We are now in a position to explain why bootstrap methods for pivotal and nonpivotal approaches admit the different properties announced in Section 1. The pivotal percentile-t method approximates the distribution of T by that of T^*, and as we have just shown, the resulting error is of size $n^{-1}h^{-1/2}$. On the other hand, nonpivotal techniques, such as the percentile method which appears to be almost universally used in practice, approximate the distribution of

$$U = \frac{1}{\beta_x}\{\hat{g}(x) - E\hat{g}(x)\} = \frac{1}{\tau_x}\sum_{i=1}^{n} e_i K\left\{\frac{x - x_i}{h}\right\}$$

by the conditional distribution of

$$U^* = \frac{1}{\tau_x}\sum_{i=1}^{n} e_i^* K\left\{\frac{x - x_i}{h}\right\}. \tag{2.9}$$

To appreciate the significant errors which arise in this approximation, observe from (2.5) and (2.7) that

$$P(U^* \le u|\mathscr{X}) - P(U \le u) = P(S^* \le u/\check{\sigma}|\mathscr{X}) - P(S \le u/\sigma) \tag{2.10}$$
$$= \Phi(u/\check{\sigma}) - \Phi(u/\sigma) + O_p(n^{-1}h^{-1/2}).$$

[Here we have used the fact that $\check{\sigma} - \sigma = O_p(n^{-1/2})$; see Theorem 2.1 below.] Since $\check{\sigma} - \sigma$ is of precise size $n^{-1/2}$, then $\Phi(u/\check{\sigma}) - \Phi(u/\sigma)$ is of precise size $n^{-1/2}$. It therefore follows from (2.10) that the bootstrap approximation to the distribution of U will have errors of at least $n^{-1/2}$, as stated in Section 1.

Despite the poor performance of nonpivotal methods, they do a little better than their counterparts in the related problem of nonparametric density estimation. There, bootstrap approximations to the distributions of pivotal and nonpivotal statistics are in error by $(nh)^{-1}$ and $(nh)^{-1/2}$, respectively. The error of $n^{-1/2}$ in the nonparametric regression case is a little better than the error $(nh)^{-1/2}$ for density estimation and the improvement is due to the fact that variance in nonparametric regression can be estimated $\sqrt{n}$ consistently. (In some respects the case of density estimation may be viewed as a straightforward generalization of the classical case, with nh in the former replacing n in the latter.)

It remains to explain the coverage accuracy properties announced in Section 1. We begin with the case of the pivotal percentile-t method. Given a probability level α, define $\hat{t}_\alpha, t_\alpha, z_\alpha$ to be the solutions of the equations

$$P\big(T^* \le \hat{t}_\alpha|\mathscr{X}\big) = P(T \le t_\alpha) = \Phi(z_\alpha) = \alpha.$$

If we knew the exact distribution of T, then we would compute t_α and take

$$\mathscr{T}_\alpha = \big(\hat{g}(x) - \hat{\gamma}_x t_\alpha, \infty\big) \tag{2.11}$$

as an α-level confidence interval for $E\hat{g}(x)$. Its coverage is precisely $P(T \le t_\alpha) = \alpha$. Generally the distribution of T is unknown, and in such cases we might replace t_α in (2.11) by its bootstrap estimate, $\hat{t}_\alpha$. The resulting interval $\hat{\mathscr{T}}_\alpha = (\hat{g}(x) - \hat{\gamma}_x\hat{t}_\alpha, \infty)$ covers $E\hat{g}(x)$ with probability $P(T \le \hat{t}_\alpha)$ and has coverage error

$$\delta_\alpha = P\big(T \le \hat{t}_\alpha\big) - \alpha.$$

To determine the size of this error, observe from (2.6), (2.8) and the fact that $\hat{\mu}_j - \mu_j = O_p(n^{-1/2})$, that

$$P(T^* \le u|\mathscr{X}) - P(T \le u) \tag{2.12}$$
$$= (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(u)\phi(u) + O_p(n^{-1} + n^{-3/2}h^{-1}).$$

It would always be the case in practice that h was of larger order than $n^{-1/2}$ and hence that $n^{-3/2}h^{-1}$ was of smaller order than n^{-1}. Thus by (2.12),

$$\hat{t}_\alpha - t_\alpha = (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(z_\alpha) + O_p(n^{-1}). \tag{2.13}$$

Therefore the coverage error which we seek is

$$\delta_\alpha = P\big\{T - (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(z_\alpha) \le t_\alpha\big\} - P(T \le t_\alpha) + O(n^{-1}).$$

(A rigorous proof of this result follows lines described in Section 4.) An Edgeworth expansion of the distribution of $T - (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(u)$ may be developed in the same manner as was the expansion for T at (2.6) and it will be found that the expansions differ only in terms of order n^{-1}. Therefore $\delta_\alpha = O(n^{-1})$. That is, the coverage errors of confidence intervals constructed by the percentile-t method are of order n^{-1}. (The argument above checks this in the case of one-sided intervals and similarly it may be proved for two-sided intervals.) Interestingly, the coverage error is larger, of order $(nh)^{-1}$, in the circumstance of nonparametric density estimation.

The case of nonpivotal methods, such as percentile, may be treated similarly. But there, a significant contributor to coverage error of a one-sided interval is the term of order $(h/n)^{1/2}$ by which the expansions (2.5) and (2.6) differ. This difference has a detrimental impact on coverage accuracy. In more detail, define U^* as at (2.9) and let $\hat{u}_\alpha, u_\alpha$ be the solutions of

$$P(U^* \le \hat{u}_\alpha | \mathscr{X}) = P(U \le u_\alpha) = \alpha.$$

A slight variant of the argument leading to (2.13) shows that

$$\hat{\sigma}^{-1}\hat{u}_\alpha - \sigma^{-1}u_\alpha = (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(z_\alpha) + O_p(n^{-1}).$$

One form of the percentile-method α-level confidence interval for $E\hat{g}(x)$ is $(\hat{g}(x) - \beta_x \hat{u}_\alpha, \infty)$, which has coverage probability

$$\begin{aligned} P(U \le \hat{u}_\alpha) &= P\big(T \le \hat{\sigma}^{-1}\hat{u}_\alpha\big) \\ &= P\big\{T - (nh)^{-1/2}(\hat{\mu}_3 - \mu_3)p_1(z_\alpha) \le \sigma^{-1}u_\alpha\big\} + O(n^{-1}) \\ &= P\big(T \le \sigma^{-1}u_\alpha\big) + O(n^{-1}) \\ &= \alpha + (h/n)^{1/2}\mu_3 p_4(z_\alpha) + O(n^{-1}). \end{aligned}$$

Therefore the coverage error of this percentile method confidence interval is of order $(h/n)^{1/2}$, which is generally larger than n^{-1}.

There is a second percentile method, called the backwards method in [10], and it may be checked that this alternative approach leads to even worse errors—of size $(nh)^{-1/2}$. In the case of nonparametric density estimation, both of the percentile methods produce coverage errors of size $(nh)^{-1/2}$. The fact that they produce errors of different orders in the case of nonparametric regression is due once again to the peculiar nature of the statistic T, having a numerator and denominator with different convergence rates.

We close this section by showing that $\hat{\mu}_j - \mu_j = O_p(n^{-1/2})$ for general $j \ge 3$. We assume that K is bounded and compactly supported and that K, the bandwidth h and the index set $\mathscr{I}$ are chosen so that for some $\varepsilon > 0$,

$$\max_{i \in \mathscr{I}} E\{\hat{g}(x_i) - g(x_i)\}^{2j} = O(n^{-j\{(1/2)+\varepsilon\}}). \tag{2.14}$$

This condition is quite mild. For example, suppose K is a second order kernel, that the bandwidth h is of size $n^{-1/5}$, that g has two bounded derivatives, that $E(e_1^{2j}) < \infty$ and that $\mathscr{I}$ is chosen so that x_i for $i \in \mathscr{I}$, is bounded away from either end of the interval of estimation, as proposed by Härdle and Bowman [7]. Then the left-hand side of (2.14) is of order $n^{-4j/5}$ and so (2.14) is certainly satisfied.

THEOREM 2.1. *If $E(e_1^{2j}) < \infty$ and (2.14) holds and if the index set $\mathscr{I}$ contains at least εn elements for some $\varepsilon > 0$ and all large n, then $\hat{\mu}_j - \mu_j = O_p(n^{-1/2})$.*

PROOF. It suffices to show that the stated conditions imply

$$n'^{-1} \sum_i{}' \hat{e}_i^j = E(e_1^j) + O_p(n^{-1/2}) \tag{2.15}$$

for $j \geq 2$. Put $\bar{e} = n'^{-1}\Sigma_i' e_i$, $\eta_i = \hat{g}(x_i) - g(x_i) - n'^{-1}\Sigma_k'\{\hat{g}(x_k) - g(x_k)\}$, $\Delta_k = n'^{-1}\Sigma_i'(e_i - \bar{e})^{j-k}(-\eta_i)^k$. Since $\hat{e}_i = e_i - \bar{e} - \eta_i$, then

$$n'^{-1} \sum_i{}' \hat{e}_i^j = \sum_{k=0}^{j} \binom{j}{k} \Delta_k. \tag{2.16}$$

It is easily proved that $\Delta_0 - E(e_1^j) = O_p(n^{-1/2})$. For $k \geq 2$,

$$\begin{aligned}|\Delta_k| &\leq \left[\left\{n'^{-1} \sum_i{}' (e_i - \bar{e})^{2(j-k)}\right\} n'^{-1} \sum_i{}' \eta_i^{2k}\right]^{1/2} \\ &= O_p(n^{-(k/2)\{(1/2)+\varepsilon\}}) = O_p(n^{-(1/2)-\varepsilon}),\end{aligned}$$

where we have used the fact that

$$\begin{aligned}\max_{i \in \mathscr{I}} E(\eta_i^{2k}) &= O\left[\max_{i \in \mathscr{I}} E\{\hat{g}(x_i) - g(x_i)\}^{2k}\right] \\ &= O\left[\max_{i \in \mathscr{I}} E\{\hat{g}(x_i) - g(x_i)\}^{2j}\right]^{k/j} = O(n^{-k\{(1/2)+\varepsilon\}}).\end{aligned}$$

It may be proved after a little algebra that $E(\Delta_1^2) = o(n^{-1})$. The desired result (2.15) follows on combining the estimates from (2.16) down. □

3. The effect of bias. Let $\hat{t}_\alpha$ denote the solution of the equation $P(T^* \leq \hat{t}_\alpha | \mathscr{X}) = \alpha$ and define $\hat{\mathscr{T}}_\alpha = (\hat{g}(x) - \hat{\gamma}_x \hat{t}_\alpha, \infty)$. As noted in Section 2, $\hat{\mathscr{T}}_\alpha$ may be regarded as a confidence interval for $E\hat{g}(x)$, with coverage error $O(n^{-1})$. If a confidence interval for $g(x)$ is to be based on $\hat{\mathscr{T}}_\alpha$, then a bias correction may be required, or it may be necessary to choose the bandwidth h so that bias is not excessive. In the present section, we shall apply the work in Section 2 to the problem of bias-corrected confidence intervals. We shall address two distinct points—the influence of bias correction on the position of interval endpoint and the effect on coverage accuracy.

Bias may be defined by $b(x) = E\hat{g}(x) - g(x)$. In order to be specific about properties of bias, let us initially take the kernel K to be a known symmetric density function, such as Epanechnikov's kernel (Härdle [6], Section 3.1). For such kernels it may be proved that if, for example, the design points are regularly spaced and if g has four continuous derivatives, then

$$b(x) = \tfrac{1}{2}k_2h^2g''(x) + O\{h^4 + (nh)^{-1}\} \tag{3.1}$$

as $h \to 0$ and $n \to \infty$, where $k = \int u^2K(u)\,du$. If $\hat{b}(x)$ denotes an estimator of $b(x)$, then $\hat{\mathscr{I}}'_\alpha = (\hat{g}(x) - \hat{\gamma}_x\hat{t}_\alpha - \hat{b}(x), \infty)$ is a bias-corrected version of $\hat{\mathscr{I}}_\alpha$.

One approach to bias correction is to estimate the dominant term on the right-hand side of (3.1) by constructing an estimator $\tilde{g}''(x)$ of $g''(x)$. If, in the notation of Härdle ([6], Section 4.5), we use a kernel estimator of order $(k, p) = (2, r + 2)$; and if the bandwidth of the estimator is chosen appropriately and g has $r + 2$ derivatives; then $\tilde{g}''(x) - g''(x) = O_p(n^{-r/(2r+5)})$ ([6], Section 4.5). Taking $\hat{b}(x) = \frac{1}{2}k_2h^2\hat{g}''(x)$ as our bias estimator, we see from (3.1) that

$$\hat{b}(x) - b(x) = O_p\{h^4 + (nh)^{-1} + n^{-2r/(2r+5)}\}. \tag{3.2}$$

The terms of order $h^4 + (nh)^{-1}$ in this formula are genuinely of that size, not of smaller order, as may be shown by a longer argument. This implies that the error in the bias correction discussed above is at least of the same size [viz., $(nh)^{-1}$] as the skewness correction term provided by the bootstrap. To appreciate why, note that $\hat{\gamma}_x \sim \text{const.}(nh)^{-1/2}$ and, by (2.8), $\hat{t}_\alpha = z_\alpha - (nh)^{-1/2}\hat{\mu}_3p_1(z_\alpha) + o_p\{(nh)^{-1/2}\}$, where $z_\alpha = \Phi^{-1}(\alpha)$. Therefore,

$$\hat{\gamma}_x\hat{t}_\alpha = \hat{\gamma}_xz_\alpha - (nh)^{-1/2}\hat{\gamma}_x\hat{\mu}_3p_1(z_\alpha) + o_p\{(nh)^{-1}\}. \tag{3.3}$$

The term $\hat{\gamma}_xz_\alpha$ on the right-hand side of (3.3) represents the quantity which we would use instead of $\hat{\gamma}_x\hat{t}_\alpha$ to construct the confidence intervals $\hat{\mathscr{I}}_\alpha$ and $\hat{\mathscr{I}}'_\alpha$, if we employed the normal approximation rather than the bootstrap. The second term on the right-hand side of (3.3) denotes the skewness correction provided by the bootstrap and is of size $(nh)^{-1}$.

An alternative approach to bias correction is to simply ignore the effect of bias and interpret $\hat{\mathscr{I}}_\alpha$ as a confidence interval for $g(x)$ rather than $E\hat{g}(x)$. This is tantamount to taking $\hat{b}(x) \equiv 0$ in the definition of $\hat{\mathscr{I}}'_\alpha$. Since bias is of size h^2 and $\hat{\gamma}_x\hat{t}_\alpha$ is of size $(nh)^{-1/2}$, then ignoring bias will not affect the asymptotic level of the confidence interval if and only if h^2 is of smaller order than $(nh)^{-1/2}$. For this to happen, h must be of smaller order than $n^{-1/5}$, that is, the regression estimator must be undersmoothed by an order of magnitude relative to the optimal amount of smoothing for point estimation of g. If we follow this route, then formula (3.3) should be replaced by

$$\hat{b}(x) - b(x) = -b(x) = O(h^2).$$

Since the skewness correction term provided by the bootstrap is of size $(nh)^{-1}$ [see (3.3)], then the skewness correction dominates the bias correction if and only if h^2 is of smaller order than $(nh)^{-1}$, that is, h is of smaller order than $n^{-1/3}$.

The influence of bias correction on coverage accuracy is more subtle and is perhaps best treated by considering a more general context where the kernel

K is of order s:

$$\int u^j K(u)\,du = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \le j \le s-1, \\ k_s \ne 0, & \text{if } j = s. \end{cases}$$

See Härdle ([6], Section 4.5); the symmetric-density kernel considered earlier is an example of the case $s = 2$. Using an sth order kernel K and assuming that m has s continuous derivatives, formula (3.1) should be replaced by

$$b(x) = (s!)^{-1} k_s(-h)^s g^{(s)}(x) + o(h^s) + O\{(nh)^{-1}\}. \tag{3.4}$$

To estimate $b(x)$, suppose we first estimate $g^{(s)}(x)$ as accurately as possible, assuming that g has $r + s$ derivates in all. This can be done using a kernel of order $(k, p) = (s, r + s)$, in the notation of Härdle ([6], Section 4.5), and produces an estimator $\tilde{g}^{(s)}(x)$ with error of size $\delta = n^{-r/(2r+2s+1)}$. Therefore the bias estimator $\hat{b}(x) = (s!)^{-1}k_s(-h)^s\tilde{g}^{(s)}(x)$ is in error by terms of at least $h^s\delta$. [The figure $h^s\delta$ does not take into account additional errors arising from failure to estimate the $o(h^s) + O\{(nh)^{-1}\}$ terms on the right-hand side of (3.4), and so the actual error may be larger.] Using this estimator of bias, we see that the confidence interval $\hat{\mathscr{T}}'_\alpha$ covers $g(x)$ with probability

$$\alpha_1 = P\{\hat{g}(x) - \hat{\gamma}\hat{t}_\alpha - \hat{b}(x) \le g(x)\} = P(T \le \hat{t}_\alpha + R_1),$$

where $R_1 = \{\hat{b}(x) - b(x)\}/\hat{\gamma}_x$. Since R_1 is at least of size $(nh)^{1/2}h^s\delta$, then $\alpha_1 = P(T \le \hat{t}_\alpha) + \Delta_1$, where, to a conservative approximation, $\Delta_1 \asymp (nh)^{1/2}h^s\delta$.

A second approach, based on the undersmoothing method described two paragraphs earlier, is to take K to be an $(r + s)$th order kernel, thereby utilizing at the first opportunity all the smoothness assumptions made about f, and to put $\hat{b}(x) \equiv 0$. In this case, the bias-corrected interval $\hat{\mathscr{T}}'_\alpha$ is identical to $\hat{\mathscr{T}}_\alpha$, with coverage probability

$$\alpha_2 = P\{\hat{g}(x) - \hat{\gamma}\hat{t}_\alpha \le g(x)\} + P(T \le \hat{t}_\alpha + R_2),$$

where $R_2 = -b(x)/\hat{\gamma}_2$. Now, R_2 is of size $(nh)^{1/2}h^{r+s}$ and so $\alpha_2 = P(T \le \hat{t}_\alpha) + \Delta_2$, where $\Delta_2 \asymp (nh)^{1/2}h^{r+s}$.

The final step in this argument is to determine the appropriate h for both these approaches. We know from Section 2 that $P(T \le \hat{t}_\alpha) = \alpha + O(n^{-1})$. Hence, the total coverage error is of size at least $\beta_1 = n^{-1} + (nh)^{1/2}h^s\delta$ when K is an sth order kernel and bias is corrected explicitly; and of size $\beta_2 = n^{-1} + (nh)^{1/2}h^{r+s}$ when K is an $(r + s)$th order kernel and no bias correction is made. Bearing in mind that $s = n^{-r/(2r+2s+1)}$, we see that $\beta_1 = n^{-1} + (hn^{1/(2r+2s+1)})^{s+(1/2)}$. Therefore, h should be of order at least $n^{-(s+(1/2))^{-1}-(2r+2s+1)^{-1}}$ if the coverage accuracy of the interval $\hat{\mathscr{T}}'_\alpha$ is to be at most $O(n^{-1})$ and if bias is corrected explicitly. A similar argument shows that h should be of order at most $n^{-1/(2r+2s+1)}$ if bias is corrected by undersmoothing.

4. Edgeworth expansions. In this section we provide explicit regularity conditions under which the expansions (2.5)–(2.8) are available. Section 5 will outline proofs of those results.

Throughout we assume the regression model (2.1) and use notation introduced in Section 2. Note in particular the definitions of τ_x, S and T. We begin by defining the polynomials $p_1, \dots, p_4$ appearing in (2.5)–(2.8). Put

$$a_1 = \frac{(nh)^{1/2}}{\tau_x^3} \sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\}^3, \qquad a_2 = \frac{(nh)^2}{\tau_x^4} \sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\}^4,$$

$$b = -\frac{1}{2(nh)^{1/2}\tau_x} \sum_{i=1}^{n} K\left\{\frac{x - x_i}{h}\right\},$$

$$p_1(u) = -\tfrac{1}{6}a_1(u^2 - 1), \qquad p_2(u) = -\tfrac{1}{24}a_2 u(u^2 - 3)^4,$$

$$p_3(u) = -\tfrac{1}{72}a_1^2 u(u^4 - 10u^2 + 15), \qquad p_4(u) = -bu^2.$$

Note that $p_1, \dots, p_4$ depend on neither the regression mean g nor the difference sequence $\{d_j\}$. Assume that (a) the design points x_i are confined to a given compact interval I and are either regularly spaced on I or represent the first n random observations of a distribution whose density is bounded away from zero on I (in the former case we should really notate x_i as x_{ni}, indicating a double array, but we suppress the additional subscript), (b) the kernel K is bounded, compactly supported and nondegenerate, (c) the bandwidth h satisfies $n^{-1+\varepsilon} \le h = h(n) \le n^{-\varepsilon}$ for some $\varepsilon > 0$ and all large n, (d) the error distribution is nonsingular and satisfies $E(e_1) = 0$ and $E(|e_1|^C) < \infty$ for some finite but sufficiently large $C > 0$ [whose choice depends on the value of ε in (c)], (e) x is an interior point of the interval I.

THEOREM 4.1. *Under conditions* (a)–(e) *and for the above definitions of* $p_1, \dots, p_4$, *expansions* (2.5) *and* (2.6) *obtain uniformly in* $-\infty < u < \infty$. *If, in addition,* (2.14) *holds and* $\mathcal{J}$ *contains at least* εn *elements for some* $\varepsilon > 0$, *then expansions* (2.7) *and* (2.8) *obtain uniformly in* u.

The coefficients a_1, a_2 and b appearing in the polynomials are bounded as $n \to \infty$. Indeed, if condition (a) holds with $I = [0, 1]$ and if we take f to be either the design density (if the design is obtained randomly) or the constant function 1 (if the design points are regularly spaced), then

$$a_1 \sim f(x)^{-1/2}\left(\int K^2\right)^{-3/2} \int K^3, \qquad a_2 \sim f(x)^{-1}\left(\int K^2\right)^{-2} \int K^4,$$

$$b \sim -\tfrac{1}{2} f(x)^{1/2}\left(\int K^2\right)^{-1/2}.$$

5. Proofs. Derivation of an expansion for S is relatively straightforward, since S is a sum of independent random variables. There are some complicating features in the case of T, due to the fact that (a) the random variable $\hat{\sigma}$

used to standardize T is a sum of $(2m + 1)$-dependent variables, not a sum of independent variables, and (b) the numerator and denominator of T have different convergence rates. We shall concentrate our attention on T and derive (2.6). Following that proof we shall outline the modifications necessary to obtain (2.8).

It may be assumed without loss of generality that $\sigma^2 = 1$. We begin by describing an argument which identifies the form of the Edgeworth expansion, up to terms of order n^{-1}. Following that, we shall make the details mathematically rigorous.

Define $m = m_1 + m_2$ and $\tau_x^2 = \sum_i K\{(x - x_i)/h\}^2$, extend the sequence $\{e_j\}$ to $\{e_{-m_1}, \dots, e_0, \dots, e_{n+m_2}\}$ and put

$$\tilde{\sigma}^2 = \frac{1}{n-m} \sum_{i=m_1+1}^{n-m_2} \Big(\sum_j d_j e_{i+j} \Big)^2, \qquad \tilde{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^{n} \Big(\sum_j d_j e_{i+j} \Big)^2.$$

Then $\hat{\sigma}^2 - \tilde{\sigma}^2 = O_p(n^{-1})$ (see Lemma 5.3) and clearly $\tilde{\sigma}^2 - \tilde{\sigma}_1^2 = O_p(n^{-1})$, whence it follows that $T = T_1 + O_p(n^{-1})$, where

$$T_1 = \frac{1}{\tau_x} \left[\sum_{i=1}^{n} e_i K\left\{ \frac{x - x_i}{h} \right\} \right] \left[1 - \frac{1}{2n} \sum_{i=1}^{n} \left\{ \Big(\sum_j d_j e_{i+j} \Big)^2 - 1 \right\} \right].$$

Hence, Edgeworth expansions of T and T_1 agree up to terms of order n^{-1}.

To evaluate the expansions of T_1, put $a_1' = \mu_3 a_1$, $a_2' = (\mu_4 - 3) a_2$ and $b' = \mu_3 b$. We claim that, if sufficiently many moments of e_1 are finite,

$$E(T_1) = (h/n)^{1/2} b', \qquad \operatorname{var}(T_1) = 1 + O(n^{-1}),$$
$$E(T_1 - ET_1)^3 = (nh)^{-1/2} a_1' + (h/n)^{1/2} 6b' + O(n^{-1}),$$
$$E(T_1 - ET_1)^4 - 3(\operatorname{var} T_1)^2 = (nh)^{-1} a_2' + O(n^{-1}),$$

and fifth and higher cumulants of T_1 are of order $n^{-1} + (nh)^{-3/2}$. These results follow from Lemma 5.1 below, whose proof is straightforward. Therefore the characteristic function of T_1 equals

$$\begin{aligned} & e^{-t^2/2} \exp\Big[(nh)^{-1/2} \tfrac{1}{6} a_1' (it)^3 + (h/n)^{1/2} b' \{ it + (it)^3 \} \\ & \qquad + (nh)^{-1} \tfrac{1}{24} a_2' (it)^4 + O\{ n^{-1} + (nh)^{-3/2} \} \Big] \\ & = e^{-t^2/2} \Big[1 + (nh)^{-1/2} \tfrac{1}{6} a_1' (it)^3 + (h/n)^{1/2} b' \{ it + (it)^3 \} \\ & \qquad + (nh)^{-1} \{ \tfrac{1}{24} a_2' (it)^4 + \tfrac{1}{72} (a_1')^2 (it)^6 \} + O\{ n^{-1} + (nh)^{-3/2} \} \Big]. \end{aligned}$$

Inverting this Fourier–Stieltjes transform, we deduce the claimed form of the expansion of T_1 (and hence of T).

LEMMA 5.1. *If $r \geq 1$ is an integer and $E(e_1^{4r}) < \infty$, then*

$$E(T_1 - ET_1)^r = E(T_2^r) + r(r-1) E(T_1) E(T_2^{r-1}) + O(n^{-1}).$$

Next we prove rigorously that the Edgeworth expansion of the distribution of T exists and that it must have the form claimed earlier up to terms or order $n^{-1} + (nh)^{-3/2}$.

Lemma 5.2 below shows that if we replace $\hat{\sigma}^2$ by $\tilde{\sigma}^2$ in the formula for T, we commit an error of only $O(n^{-5/4})$ in the Edgeworth expansion. Therefore, it suffices to prove the theorem for T replaced by

$$T' = \frac{1}{\tau_x \tilde{\sigma}} \sum_{i=1}^{n} e_i K\left\{\frac{x - x_i}{h}\right\} \tag{5.1}$$
$$= (nh)^{1/2} \tau_x^{-1} \left\{1 + (n-m)^{-1/2} W_{n2}\right\}^{-1/2} W_{n1},$$

where

$$W_{n1} = \frac{1}{(nh)^{1/2}} \sum_{j=1}^{n} e_j K\left\{\frac{x - x_j}{h}\right\},$$

$$W_{n2} = \frac{1}{(n-m)^{1/2}} \sum_{j=m_1+1}^{n-m_2} \left\{\left(\sum_k d_k e_{j+k}\right)^2 - 1\right\}.$$

The first of these two random variables is a sum of independent components, the second is a sum of $(2m+1)$-dependent, identically distributed components. The vector (W_{n1}, W_{n2}) is asymptotically distributed as (W_1, W_2), a normal vector with the same mean and variance as (W_{n1}, W_{n2}). We shall extend this result by establishing an Edgeworth expansion for the distribution of (W_{n1}, W_{n2}), in which the first term is the distribution of (W_1, W_2) and the remainder is of size $(nh)^{-(r+1)/2}$, where $r \geq 0$ is an arbitrarily large integer. The terms in the expansions are all polynomials multiplied by the density of (W_1, W_2) and the coefficients of the polynomials are of orders $(nh)^{-i/2} h^j$, where $1 \leq i \leq r$ and $0 \leq j \leq r$. The existence of the desired expansion of the distribution of T' now follows from formula (5.1), expressing T' as a function of (W_{n1}, W_{n2}), and elementary Taylor expansion as in Bhattacharya and Ghosh ([1], page 444). That the Edgeworth expansion must agree with the one identified earlier is a consequence of the usual relationship between cumulants of a function of two variables W_{n1}, W_{n2} and cumulants of W_{n1}, W_{n2} themselves, as deduced by Taylor expansion (e.g., [1], page 444f).

Before writing down the Edgeworth expansion of the distribution of (W_{n1}, W_{n2}) we need a little notation. Given 2-vectors $\mathbf{z} = (z_1, z_2)$ and $\nu = (\nu_1, \nu_2)$, where ν_1 and ν_2 are nonnegative integers, define $\mathbf{z}^\nu = z_1^{\nu_1} z_2^{\nu_2}$, $\|\mathbf{z}\| = (z_1^2 + z_2^2)^{1/2}$ (the latter for real z_1, z_2), $\nu! = \nu_1! \nu_2!$ and $|\nu| = \nu_1 + \nu_2$. For $1 \leq k \leq s$, let $\Sigma'_{(s,k)}$ denote summation over k-vectors $\mathbf{j} = (j_1, \ldots, j_k)$ of positive integers such that $j_1 + \cdots + j_k = s$ and let $\Sigma''_{(j,k)}$ denote summation over k-sequences of 2-vectors $\{\nu_l = (\nu_{l1}, \nu_{l2}): 1 \leq l \leq k\}$ such that $|\nu_l| = j_l + 2$, $1 \leq l \leq k$. Let $\mathbf{t} = (t_1, t_2)$ denote a 2-vector with real components and define the differential operator $D^\nu = (\partial/\partial t_1)^{\nu_1} (\partial/\partial t_2)^{\nu_2}$. Write β, β_n for the characteristic functions of (W_1, W_2), (W_{n1}, W_{n2}), respectively. Put

$$Q_s(\mathbf{z}) = \beta(-i\mathbf{z}) \sum_{k=1}^{s} (1/k!) \sum_{(s,k)}{}' \sum_{(\mathbf{j},k)}{}'' \prod_{l=1}^{k} \left[(\mathbf{z}^{\nu_l}/\nu_l!) D^{\nu_l} \log \beta_n(\mathbf{t})|_{\mathbf{t}=\mathbf{0}}\right].$$

Let q_s denote the density on $\mathbb{R}^2$ of the signed measure whose Fourier–Stieltjes transform is $Q_s(i\mathbf{t})$. [Then $q_s = p_s\psi$, where ψ is the density of (W_1, W_2) and p_s is a polynomial of degree $3s$ in two variables, even for even s and odd for odd s. In particular, $q_0 = \psi$.] Let $\mathscr{A}$ denote the class of all convex sets $A \subseteq \mathbb{R}^2$.

LEMMA 5.2. *Let $r \geq 0$ be an integer, assume that $E(e_1^{2(r+3)}) < \infty$ and that the distribution of e_1 is nonsingular, and assume the conditions of Theorem 3.1 on the kernel K, on the design and on the bandwidth h. Then*

$$\sup_{A \in \mathscr{A}} \left| P\{(W_{n1}, W_{n2}) \in A\} - \int_A \sum_{s=0}^{r} q_s(\mathbf{x})\, d\mathbf{x} \right| = O\{(nh)^{-(r+1)/2}\}.$$

Since we assumed that for some $\varepsilon > 0$, $h > n^{-1+\varepsilon}$, then $(nh)^{-(r+1)/2}$ may be made less than n^{-C} for any given $C > 0$ simply by choosing r sufficiently large. Therefore Lemma 5.2 provides an Edgeworth expansion with a remainder of order n^{-C} for any C which we care to select, at the expense of a moment condition which depends on choice of C.

Lemma 5.2 may be established in the usual way, via a "smoothing" or kernel function (e.g., [1, page 208 ff], but without the necessity for any truncation since we are not striving to be economical about moment assumptions), using arguments of Heinrich [12]. In this work, the role of Cramer's condition is filled by the property that for each $\varepsilon > 0$ there exists a $C = C(\varepsilon) > 0$ such that

$$\sup_{|t_1|+|t_2|>\varepsilon} \left| E\left(\exp\left[it_1 \sum_{j=1}^{n} e_j K\left\{\frac{x - x_j}{h}\right\} + it_2 \sum_{j=m_1+1}^{n-m_2} \left(\sum_k d_k e_{j+k}\right)^2\right]\right)\right| \tag{5.2}$$
$$= O\{\exp(-Cnh)\}.$$

This completes the proof of (2.6). The proof of (2.8) is similar, the main point of difference being that the error distribution is now the discrete distribution which places mass n'^{-1} at each of the points $\hat{e}_j$, $j \in \mathscr{I}$. This distribution does not enjoy the nonsingularity property, but that assumption was used only during the proof of (5.2). The analogue of (5.2) does not hold, although the key step in the proof of that result does have a valid counterpart:

$$\left| E\left(\exp\left[it_1 \sum_{j=1}^{n} e_j^* K\left\{\frac{x - x_j}{h}\right\} + it_2 \sum_{j=m_1+1}^{n-m_2} \left(\sum_k d_k e_{j+k}^*\right)^2\right] \middle| \mathscr{X}\right)\right|$$
$$\equiv |\hat{\psi}_n(t_1, t_2)| \leq E\left\{\left|\prod_{l=1}^{l_0} \hat{\chi}(\hat{a}_l, t_2)\right| \middle| \mathscr{X}\right\},$$

where $\hat{\chi}(u_1, u_2) = E\{\exp(iu_1 e_1^* + iu_2 e_1^{*2}) | \mathscr{X}\}$,

$$\hat{a}_l = t_1 K\left\{\frac{x - x_j}{h}\right\} + 2t_2 \sum_{j=j_l-m_2}^{j_l+m_1} d_{j_l-j}\left(\sum_{k \neq j_l-j} d_k e_{j+k}^*\right).$$

The trick now is to substitute this result directly into analytical estimates involving the smoothing or kernel function and argue as in ([9], pages 1442–1443), to establish the bootstrap version of Lemma 5.1. Note that, while the bootstrap counterpart of (5.2) is not valid, it is true that for each $\varepsilon > 0$, there exists $\varepsilon' > 0$ such that with probability 1, for all sufficiently large n,

$$\sup_{\varepsilon < |u_1| + |u_2| \leq n^{1/2}/\log n} |\hat{\chi}(u_1, u_2)| \leq 1 - \varepsilon'$$

(compare Csörgő [18], page 130) and that for any $\lambda > 0$,

$$\sup_{u_1, u_2} P\left\{|\hat{\chi}(u_1, u_2) - \chi(u_1, u_2)| > n^{-1/2} \log n\right\} = O(n^{-\lambda}).$$

Acknowledgment. I am grateful to two referees for their helpful comments, which have significantly improved presentation.

REFERENCES

[1] Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

[2] Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

[3] DiCiccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals (with discussion). *J. Roy. Statist. Soc. Ser. B* **50** 338–354.

[4] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

[5] Härdle, W. (1989). Resampling for inference from curves. *Proc. 47th Session of the Internat. Statist. Inst.*

[6] Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

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

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

[9] Hall, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14** 1431–1452.

[10] Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 237–249.

[11] Heinrich, L. (1982). A method for the derivation of limit theorems for sums of m-dependent random variables. *Z. Wahrsch. Verw. Gebiete* **60** 501–515.

[12] Heinrich, L. (1984). Nonuniform estimates and asymptotic expansions of the remainder in the central limit theorem for m-dependent random variables. *Math. Nachr.* **115** 7–20.

[13] Hinkley, D. V. (1988). Bootstrap methods. *J. Roy. Statist. Soc. Ser. B* **50** 321–337.

[14] Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.* **46**. Springer, New York.

[15] Müller, H.-G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–635.

[16] MÜLLER, H.-G. and STADTMÜLLER, U. (1988). Detecting dependencies in smooth regression models. *Biometrika* **75** 639–650.

[17] RICE, J. (1984). Bandwidth choice for nonparametric kernel regression. *Ann. Statist.* **12** 1215–1230.

[18] CSÖRGŐ, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.* **9** 130–144.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA ACT 2601
AUSTRALIA