The Annals of Statistics
1990, Vol. 18, No. 4, 1712–1728

ADAPTIVE *M*-ESTIMATION IN NONPARAMETRIC REGRESSION

BY PETER HALL AND M. C. JONES

Australian National University and IBM Thomas J. Watson Research Center

A method for robust nonparametric regression is discussed. We consider kernel M-estimates of the regression function using Huber's ψ-function and extend results of Härdle and Gasser to the case of random designs. A practical adaptive procedure is proposed consisting of simultaneously minimising a cross-validatory criterion with respect to both the smoothing parameter and a robustness parameter occurring in the ψ-function. This method is shown to possess a theoretical asymptotic optimality property, while some simulated examples confirm that the approach is practicable.

1. Introduction and summary. Methods for nonparametric regression are popular nowadays [e.g., Eubank (1988), Härdle (1989)] and estimating the regression function robustly in this context is often desirable. We consider one appealing approach to robust nonparametric regression based on concatenating the kernel method for smoothing and the M-estimation approach to robust estimation. Specifically, the kernel M-estimate of the regression mean $m(x)$ based on data (x_i, Y_i), $1 \le i \le n$, is the solution $\hat{m}(x)$ to the equation $H\{x, \hat{m}(x)\} = 0$, where

$$H\{x, \theta(x)\} = \sum_{i=1}^{n} a_i(x)\Psi\{Y_i - \theta(x)\},\tag{1.1}$$
$$a_i(x) = (nh)^{-1}K\{h^{-1}(x - x_i)\}.$$

Here, K is the kernel function (often a symmetric density function) and $h > 0$ is the smoothing parameter, or window width. The function Ψ, often odd and monotone, controls the robustness properties of $\hat{m}(x)$. Note that if $\Psi(x) = x$ were chosen, we would recover the (nonrobust) Nadaraya–Watson kernel regression estimate [Eubank (1988), page 115, Härdle (1989), page 21].

An important paper on kernel M-estimation is Härdle and Gasser (1984); see also Härdle (1984) and Härdle and Tsybakov (1988). Härdle and Gasser showed that these estimates have many of the advantages typically associated with robust inferences. In particular, they studied mean squared error properties of kernel M-estimates and displayed a neat factorisation of the result into terms attributable to the smoothing and another term attributable to robustness. The asymptotic bias of $\hat{m}(x)$ does not depend on Ψ in any way, while the asymptotic variance of $\hat{m}(x)$ is proportional to $\beta(\Psi) = \tau_2(\Psi)/\tau_1^2(\Psi)$, where $\tau_1(\Psi) = E\{\Psi'(\varepsilon)\} > 0$ and $\tau_2(\Psi) = \operatorname{var}\{\Psi(\varepsilon)\}$. Here, expectation and variance are taken over the distribution of the generic residual ε. Moreover, the

Received February 1989; revised November 1989.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62F35.
Key words and phrases. Cross-validation, Huber's ψ-function, kernel estimation, robust smoothing.

question of choosing Ψ to minimise $\beta(\Psi)$ is precisely the classic problem of efficiency in choice of Ψ in M-estimation [Huber (1981)]. In Section 2, we extend Härdle and Gasser's (1984) results which hold for the fixed x-design case to the random design situation (heuristically for quite general monotone Ψ, rigorously for the specific Ψ class discussed later).

For the sake of definiteness and simplicity, we are especially concerned with the use of Huber's (1964, 1981) ψ-function

$$\psi(u) = \psi_c(u) = \begin{cases} 1, & \text{for } u > c, \\ u/c, & \text{for } -c < u \le c, \\ -1, & \text{for } u \le -c, \end{cases}$$

as Ψ in (1.1). Note that with probability 1, $\psi'(\varepsilon)$ is well defined and nonnegative. The function ψ affords a range of potential robust estimators, varying from the (nonrobust) kernel mean estimate when $c = \infty$ to a kernel median estimate as $c \to 0$, with intermediate values of c corresponding to a class of compromises between the two extremes. The optimal choice of c for minimising $\beta(\psi)$ is $c = \infty$ when ε has the normal distribution and is $c = 0$ when ε is double exponential. To encompass both these extremes properly, however, it is also useful to consider use of

$$c\psi_c(u) = \begin{cases} c, & \text{for } u > c, \\ u, & \text{for } -c < u \le c, \\ -c, & \text{for } u \le -c, \end{cases}$$

in (1.1). Of course, the value of the resulting estimator is no different whether ψ or $c\psi$ is used. It is convenient, though, to work directly with ψ if we wish to allow the limiting case $c = 0$, and to work with $c\psi$ if we wish to allow $c = \infty$; in this way, we can treat the full range $0 \le c \le \infty$.

There are, therefore, two values, h and c, to be specified to implement this kernel M-estimation procedure. If the true mean function were known, we could choose the pair h, c that yields the minimum value of the average squared difference between m and $\hat{m}$ at the datapoints,

$$S(h, \psi) = n^{-1} \sum{}' \{\hat{m}(x_i) - m(x_i)\}^2,$$

where Σ' denotes summation over those x_i's which lie in a given interval (s, t). This restriction to a central set (s, t) of x_i's is made to avoid what can be awkward problems with end effects. Since S is unknown, we estimate it and choose h, c to minimise this estimate of S. The argument give in Section 3 shows that an estimate based on the well-known cross-validatory principle proves suitable [see formula (3.1)]. Of course, choosing h by cross-validation is a familiar technique in nonparametric regression [Eubank (1988), Härdle (1989)]; moreover, choosing appropriate M-estimates adaptively is popular in both robust location estimation and parametric robust regression. It is a novel aspect of the current study that h and c are selected simultaneously by a data-based procedure. Our aim in this paper is to show that such a method is both valid and practicable, at least for moderate to large samples. An advantage of our cross-validation method over alternatives, such as plug-in methods, is that cross-validation is purely data-based, and does not require the selection of subsidiary smoothing parameters. An advantage of plug-in methods is that they generally have better convergence rates relative to the smoothing parameter which minimizes mean squared error.

In Section 4 we comment on computational considerations; the computational burden of this cross-validatory approach is not as great as one might suppose. We go on to describe some simulated examples of this procedure which indicate that the method has practical potential. Proofs are deferred to Sections 5 and 6.

It is possible to develop generalizations of our results in several directions. One is to the case where the function ψ is chosen from a larger parametric class than the one-parameter family considered here. Another is to the case of a ψ function which is determined nonparametrically. This circumstance could arise in the context of robust inference with asymmetric errors, when ψ would be a functional of the unknown design density. The latter could be estimated parametrically, and the estimate substituted into the formula for ψ.

2. Squared error properties. Assume that the data can be modelled by

$$Y_i = m(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

where $x_1, \dots, x_n, \varepsilon_1, \dots, \epsilon_n$ are independent random variables. The x_i's have common (univariate) density g and the ε_i's are symmetrically distributed about zero with common continuous distribution function F and density $f = F'$. We conduct inference conditional on the x_i's. Let $-\infty < s < t < \infty$ and take $\mathscr{H} = \{h\colon n^{-1+\eta} \le h \le n^{-\eta}\}$, where $0 < \eta < \frac{1}{2}$ is arbitrary but fixed. We work only with h's in this set; this is reasonable because the asymptotically optimal h (in the L_2 sense) has the form $cn^{-\alpha}$ for a constant c and $0 < \alpha < 1$. We also assume that K satisfies

$$K \ge 0, \qquad \int K = 1, \qquad \int zK(z)\,dz = 0, \tag{2.1}$$

K vanishes outside a compact set and K is Hölder continuous.

THEOREM 2.1. *Assume condition* (2.1) *on* K; *that for some* $\delta > 0$, g *and* m *are bounded and Hölder continuous on* $(s-\delta, t+\delta)$; *that* $g > 0$ *on* $[s,t]$; *and that* f *is bounded on* $(-\infty, \infty)$ *and is continuous in a neighbourhood of the origin. Then*

$$\hat{m}(x) - m(x) = H\{x, m(x)\}\{\tau_1 g(x) + \delta_1(x)\}^{-1} + (nh)^{-1/2}\delta_2(x), \tag{2.2}$$

where for each $c_0 > 0$,

$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \sup_{s\le x\le t} \{|\delta_1(x)| + |\delta_2(x)|\} \to 0 \quad a.s. \tag{2.3}$$

If we replace ψ *by* $c\psi$ *in the definition of* H *and assume additionally that all* *moments of ε are finite, then* (2.2) *holds for* $0 \le c \le \infty$ *and in place of* (2.3),

$$\sup_{0\le c\le\infty}\ \sup_{h\in\mathscr{H}}\ \sup_{s\le x\le t}\ \{|\delta_1(x)| + |\delta_2(x)|\} \to 0 \quad a.s.$$

The dichotomous form of Theorem 2.1 reflects the two interpretations of ψ which are necessary to cope with both the extreme cases $c = 0$ and $c = \infty$.

Define

$$B_1 = \kappa_2(\tau_2/\tau_1^2)(t-s) \quad \text{and} \quad B_2 = \tfrac{1}{4}\kappa_1^2\int_s^t \gamma(x)^2 g(x)^{-1}\,dx.$$

THEOREM 2.2. *Assume condition* (2.1) *on* K; *that for some* $\delta > 0$, g *and* m *have two bounded continuous derivatives on* $(s - \delta, t + \delta)$; *that* $g > 0$ *on* $[s, t]$; *and that* f *is bounded on* $(-\infty, \infty)$ *and is continuous in a neighbourhood of the origin. Then*

$$S(h,\psi) = B_1(nh)^{-1} + B_2h^4 + \delta_3, \tag{2.4}$$

where for each $c_0 > 0$,

$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \{(nh)^{-1} + h^4\}^{-1}|\delta_3| \to 0 \quad a.s. \tag{2.5}$$

If we replace ψ *by* $c\psi$ *in the definition of* H *and assume additionally that all moments of* ε *are finite, then in place of* (2.5),

$$\sup_{0\le c\le\infty}\ \sup_{h\in\mathscr{H}}\ \{(nh)^{-1} + h^4\}^{-1}|\delta_3| \to 0 \quad a.s.$$

Formula (2.4) is the basis for the claim in Section 1 that the asymptotic variance of $\hat{m}$ depends on Ψ only through the ratio $\beta = \tau_2/\tau_1^2$ and that the asymptotic bias is independent of Ψ. It implies that the minimum of $S(h,\psi)$ is asymptotic to a constant multiple of $(\beta/\eta)^{4/5}$ and that the asymptotic minimum is achieved with $h = \{B_1(4B_2n)^{-1}\}^{1/5}$.

3. Cross-validation. Define

$$U(h,\psi) = n^{-1}\sum_i{}'\hat{m}(x_i)m(x_i), \qquad \hat{U}(h,\psi) = n^{-1}\sum_i{}'\hat{m}_i(x_i)m(x_i)Y_i,$$

where $\hat{m}_i(x)$ solves $H_i\{x, \hat{m}_i(x)\}$ and $H_i(x,\theta) = \sum_{j\ne i}a_j(x)\psi(Y_j - \theta)$.

THEOREM 3.1. *Assume the conditions of Theorem* 2.2 *and in addition that* $f(0) > 0$. *Then there exists* $\eta' > 0$, *depending on* η, *such that*

$$\hat{U}(h,\psi) = U(h,\psi) + n^{-1}\sum_i{}'m(x_i)\varepsilon_i + \delta_4,$$

where

$$\sup_{n^{-\eta'}\le c\le\infty}\ \sup_{h\in\mathscr{H}}\ \{(nh)^{-1} + h^4\}^{-1}|\delta_4| \to 0 \quad a.s.$$

The basic cross-validatory criterion is

$$\mathrm{CV}(h,\psi) = n^{-1} \sum_i{}' \hat{m}(x_i)^2 - 2\hat{U}(h,\psi);$$

compare Härdle (1984). To assess the efficacy of cross-validation, note that $\mathrm{CV}(h,\psi) = S(h,\psi) - V - 2\delta_4$, where $V = n^{-1}\Sigma_i' m(x_i)^2 + 2n^{-1}\Sigma_i' m(x_i)\varepsilon_i$ does not depend on c or h. Theorems 2.2 and 3.3 imply that if $(\hat{h},\hat{\psi})$ minimises $\mathrm{CV}(h,\psi)$ in the range $n^{-\eta'} \le c \le \infty$ and $h \in \mathscr{H}$, then $S(\hat{h},\hat{\psi})/\{\inf_{0\le c\le\infty} \inf_{h\in\mathscr{H}} S(h,\psi)\} \to 1$ a.s. This demonstrates that cross-validation results in asymptotic minimisation of $S(h,\psi)$.

The criterion $\mathrm{CV}(h,\psi)$ is not the only cross-validatory one which could be used. In numerical work it is convenient to replace the sum $\Sigma_i' \hat{m}(x_i)^2$ in the formula for CV by $\Sigma_i' \hat{m}_i(x_i)^2$. This gives the alternative criterion

$$\mathrm{CV}_1(h,\psi) = n^{-1} \sum_i{}' \hat{m}_i(x_i)^2 - 2\sum_i{}' \hat{m}_i(x_i) Y_i.$$

Thus, minimising CV_1 is equivalent to minimising

$$\mathrm{CV}_2(h,\psi) = n^{-1} \sum_i{}' \{Y_i - \hat{m}_i(x_i)\}^2. \tag{3.1}$$

The latter criterion is obtainable from the general principles for cross-validation discussed by Stone (1974, 1977). Furthermore, minimisations of CV and CV_2 are asymptotically equivalent, since it may be proved that under the conditions of Theorem 3.1,

$$\sup_{n^{-\eta'}\le c\le c_0} \sup_{h\in\mathscr{H}} \{(nh)^{-1} + h^4\}^{-1} |\mathrm{CV}(h,\psi) - \mathrm{CV}_1(h,\psi)| \to 0 \quad \text{a.s.}$$

4. Simulated examples. We computed estimators using the criterion CV_2 and iteratively reweighted least squares (IRLS): From a current estimate $\hat{m}^{(p-1)}(x)$, say, at a point x, update estimated weights $w_i^{(p-1)}$ by

$$w_i^{(p)} = w_i^{(p-1)} a_i(x) \psi\big(Y_i - \hat{m}^{(p-1)}(x)\big) / \big(Y_i - \hat{m}^{(p-1)}(x)\big),$$

$1 \le i \le n$, and solve the associated weighted least squares problem (which is trivial) to obtain $\hat{m}^{(p)}(x)$, then iterate to convergence at step P, say. (If IRLS converges, it does so to the required solution; on fewer than 5% of occasions did we have trouble with convergence.) Moreover, because the solution $\hat{m} = \hat{m}^{(P)}$ also solves a weighted least squares problem (with weights $w_i^{(P)}$, $1 \le i \le n$), the function CV_2 is given by

$$\mathrm{CV}_2(h,\psi) = n^{-1} \sum_i{}' \left[(1 - h_i)^{-1} \{Y_i - \hat{m}^{(P)}(x_i)\}\right]^2,$$

where $h_i = w_i^{(P)}/\Sigma_{j=1}^n w_j^{(P)}$. Thus the apparent need to actually drop out each x_i in turn to obtain $m_i(x_i)$ is circumvented. These ideas are not new: see, for example, O'Sullivan (1988).

Our simulations used the Epanechnikov kernel and were based on the model

$$(4.1) \qquad Y_i = 80\left(x_i - \tfrac{1}{2}\right)^2 + \varepsilon_i, \qquad 1 \le i \le n.$$

We took 200 x_i's from the uniform distribution on [0, 1] and concentrated our interest on the central portion $[s, t] = [\frac{1}{4}, \frac{3}{4}]$. The effective sample size was, therefore, approximately 100. If the distribution of ε is normal $N(0,1)$, then the bivariate cross-validation function $\mathrm{CV}(h, \psi_c)$ is typically a long valley in the c direction, sloping gently downwards as c increases. This indicates a slight preference for the mean fit ($c = \infty$). A similar result is obtained if the error distribution is double exponential, except that here the valley slopes gently in the other direction, suggesting a median-like fit ($c = 0$). This case is illustrated in Figure 1. Here and in most other circumstances at which we have looked, CV_2 is often quite flat as a function of c, reflecting a relative indifference to which value of c is employed. However, bowl-shaped CV_2 functions can be achieved by using error distributions which are contrived to have both small $f(0)$ and large error variance, or by drawing very large samples. There can sometimes be anomalies of the sort often seen in cross-validatory approaches; in particular, there is sometimes a deep global minimum of the criterion at an overly small value of h or a dramatic plunge of the criterion as $c \to 0$. We regard both as artefacts of the method which can be ignored in practice.

5. Proofs for Section 2. In this section, we give proofs of Theorems 2.1 and 2.2. The major respect in which our arguments differ from those for related problems, for example in Stone (1984), is that uniformity is here required over $\psi = \psi_c$ as well as h. For any fixed $0 < c_1 < c_2 < \infty$, the range $c_1 \le c \le c_2$ is relatively easily handled because ψ_c varies smoothly there. Indeed, there exists a constant C, depending only on c_1 and c_2, such that

$$(5.1) \qquad \sup_{-\infty < u < \infty} |\psi_c(u) - \psi_{c'}(u)| \le C|c - c'|,$$

whenever $c_1 \le c,\ c' \le c_2$. The situation is rather different when $c_1 = 0$, for then the bound (5.1) fails to hold. Indeed,

$$\sup_{-\infty < u < \infty} |\psi_c(u) - \psi_0(u)| = 1,$$

when $c \ne 0$. Therefore, we shall pay particular attention to establishing the first parts of Theorems 2.1 and 2.2, i.e., results (2.3) and (2.5). Under the additional assumption that all moments of ε are finite, and replacing ψ by $c\psi$ in the definition of H, it is relatively easy to establish versions of (2.3) and (2.5) when the supremum over $0 \le c \le c_0$ is taken instead over $c_0 \le c \le \infty$. This is because of the smooth way in which the limit $c = \infty$ is approached; a truncation argument, with $c \le n^{\lambda}$ and $c \ge n^{\lambda}$ treated separately, may be used there. Hence we shall not devote effort to the range $c_0 \le c \le \infty$.

A cornerstone of our proofs is the lattice argument, which we spell out in the proof of Lemma 5.1 and invoke on other occasions. Details of the argument are similar in all applications, and so they are given only for Lemma 5.1.

The following notation is needed. Define $\zeta(u) = E\{\psi(\varepsilon + u)\}$,

$$\chi(x,\theta) = \sum_{i=1}^{n} a_i(x)\zeta\{m(x_i) - \theta\},$$

$$\chi'(x,\theta) = (\partial/\partial\theta)\chi(x,\theta),$$

$$A(x,\theta) = H(x,\theta) - \chi(x,\theta).$$

Note that $\tau_1 = E\{\psi'(\varepsilon)\} = \zeta'(0)$. So as to more clearly motivate our technical arguments, we first give the main points of the proofs of Theorem 2.1 and 2.2. These are followed by statements and proofs of the requisite Lemmas 5.1–5.6. In those results, the phrase under the conditions of Theorem 2.1 [2.2] refers to the conditions stated just prior to (2.2) [respectively, (2.4)].

PROOF OF THEOREM 2.1. Define $\theta = m(x)$ and $\hat{\theta} = \hat{m}(x)$. Now,

$$0 = H(x,\hat{\theta}) = A(x,\hat{\theta}) + \chi(x,\hat{\theta}),$$

$$\chi(x,\hat{\theta}) = \chi(x,\theta) + (\hat{\theta} - \theta)\chi'(x,\hat{\theta}^*),$$

where $\hat{\theta}^*$ lies between θ and $\hat{\theta}$. Therefore

$$\begin{aligned}\hat{\theta} - \theta &= -\{A(x,\hat{\theta}) + \chi(x,\theta)\}/\chi'(x,\hat{\theta}^*)\\ &= \{H(x,\theta) + \delta_3(x)\}/\{\tau_1 g(x) + \delta_1(x)\},\end{aligned} \tag{5.2}$$

where $\delta_1(x) = -\chi'(x,\hat{\theta}^*) - \tau_1 g(x)$ and $\delta_3(x) = A(x,\hat{\theta}) - A(x,\theta)$. Hence it suffices to prove the version of (2.2) which has δ_2 replaced by $(nh)^{1/2}\delta_3$.

Write $\sup^\dagger$ to denote the supremum over $0 \le c \le c_0$, $h \in \mathscr{H}$ and $s \le x \le t$. If we show that

$$\sup{}^\dagger|\hat{\theta} - \theta| \to 0 \quad \text{a.s.}, \tag{5.3}$$

then it follows from Lemma 5.2 that $\sup^\dagger|\delta_1(x)| \to 0$. By (5.5) of Lemma 5.1 and Lemma 5.3, we have $\sup^\dagger\{|A(x,\hat{\theta})| + |\chi(x,\theta)|\} = O(n^{-\eta})$ for some $\eta > 0$, and so by (5.2), $\sup^\dagger|\hat{\theta} - \theta| = O(n^{-\eta})$. It then follows by (5.6) of Lemma 5.1 that $\sup^\dagger(nh)^{1/2}|\delta_3| \to 0$. This completes the proof of Theorem 2.1, except for the necessity of checking (5.3).

To prove (5.3), note that since $H(x,\theta)$ is nonincreasing in θ, then $\hat{\theta} > \theta + \eta$ implies $O = H(x,\hat{\theta}) \le H(x,\theta + \eta) = A(x,\theta + \eta) + \chi(x,\theta + \eta)$. Therefore,

$$P(\hat{\theta} - \theta > \eta) \le P\{A(x,\theta + \eta) > -\chi(x,\theta + \eta)\}. \tag{5.4}$$

Now,

$$\chi(x,\theta + \eta) = \sum_{i=1}^{n} a_i(x)\zeta\{m(x_i) - m(x) - \eta\} \le \zeta(\eta_n - \eta)\sum_{i=1}^{n} a_i(x),$$

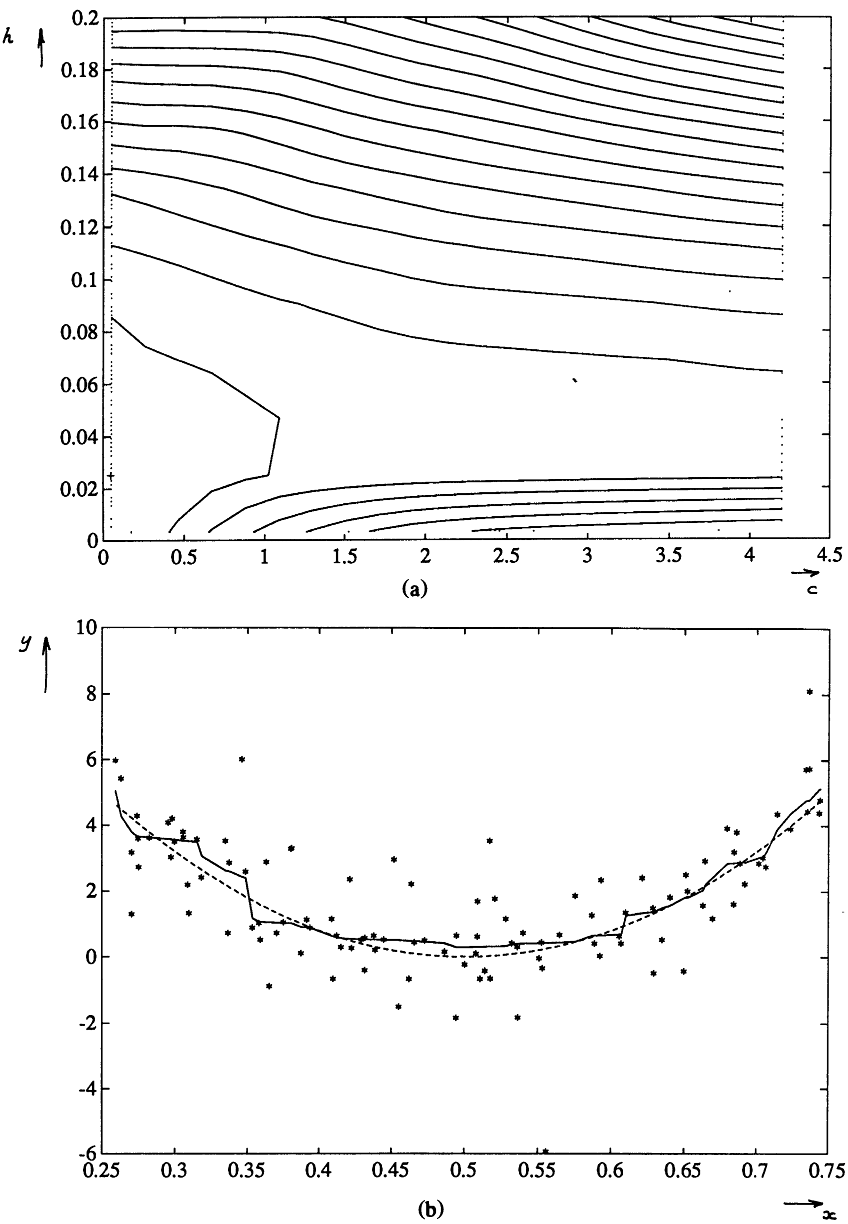

FIG. 1. (a) *A contour plot of* $\mathrm{CV}_2(h, \psi_c)$ *based on the data described under* (b). *There are* 10 *equispaced values of h from* 0.00325 *to* 0.2 *inclusive and* 21 *equispaced values of c from* 0.05 *to* 4.2 *inclusive.* (b) *The stars represent those datapoints whose x-values lie in* $[\frac{1}{4}, \frac{3}{4}]$ (*from an original sample of* 200 $U[0,1]$ x_i*'s) and whose y's arise from the model* (4.1) *with residual distribution the double exponential. The dotted curve is the true mean, m; the solid curve is the estimate* $\hat{m}$ *corresponding to* $(c, h) = (0.05, 0.0225)$ *chosen by reference to* CV_2 (*here,* $c = 0.05$ *is effectively the same as* $c = 0$).

where K vanishes outside $(-C_1, C_1)$ and

$$\eta_n = \sup_{h \in \mathscr{H}} \sup_{s<x<t} \sup_{y:\,|x-y|\le C_1 h} |m(x) - m(y)|.$$

Continuity of m on $(s-\delta, t+\delta)$ guarantees that $\eta_n \to 0$ as $n \to \infty$. If n is so large that $\eta_n \le \frac{1}{2}\eta$, then

$$\zeta(\eta_n - \eta) \le \zeta(-\tfrac{1}{2}\eta) \le -C_2 = -\inf_{0\le c\le c_0} |\zeta(-\tfrac{1}{2}\eta)|,$$

with $C_2 > 0$. Combining the estimates from (5.4) down, we see that

$$\begin{aligned} P(\hat{\theta} - \theta > \eta) &\le P\left\{A(x, \theta + \eta) > C_2 \sum_{i=1}^{n} a_i(x)\right\} \\ &\le P\{A(x, \theta+\eta) > \tfrac{1}{2}C_2 g(x)\} \\ &\quad + P\left\{\left|\sum_{i=1}^{n} a_i(x) - g(x)\right| > \tfrac{1}{2} g(x)\right\}. \end{aligned}$$

Application of Bernstein's or Rosenthal's inequality to bound the two probabilities on the right-hand side now shows that for all $\lambda > 0$, $\sup^{\dagger} P\{\hat{m}(x) - m(x) > \eta\} = O(n^{-\lambda})$. An identical argument produces the same bound if $\hat{m}, m$ are interchanged in this statement. Therefore $\sup^{\dagger} P\{|\hat{m}(x) - m(x)| > \eta\} = O(n^{-\lambda})$. Use of the lattice argument now gives (5.3). □

Proof of Theorem 2.2. Define $\chi(x) = \chi\{x, m(x)\}$, $\mu(x) = E\{\chi(x)\}$, $H(x) = H\{x, m(x)\}$, $A(x) = A\{x, m(x)\}$, $B(x) = \chi(x) - \mu(x)$, $S = n^{-1}\Sigma'_i\{\hat{m}(x_i) - m(x_i)\}^2$, $T = \tau_1^{-2} n^{-1} \Sigma'_i H(x_i)^2 g(x_i)^{-2}$. By Theorem 2.1,

$$\sup_{0\le c\le c_0} \sup_{h\in\mathscr{H}} |S - T| / \{T + (nh)^{-1}\} \to 0 \quad \text{a.s.},$$

and so it suffices to prove the version of (2.4)–(2.5) in which S is replaced by T. To obtain that result, observe that $T = \tau_1^{-2}\{T_1 + T_2 + T_3 + 2(T_4 + T_5 + T_6)\}$, where T_1 through T_6 are, respectively, $n^{-1}\Sigma'_i A(x_i)^2 g(x_i)^{-2}$, $n^{-1}\Sigma'_i \mu(x_i)^2 g(x_i)^{-2}$, $n^{-1}\Sigma'_i B(x_i)^2 g(x_i)^{-2}$, $n^{-1}\Sigma'_i A(x_i)B(x_i)g(x_i)^{-2}$, $n^{-1}\Sigma'_i A(x_i)\mu(x_i)g(x_i)^{-2}$, $n^{-1}\Sigma'_i B(x_i)\mu(x_i)g(x_i)^{-2}$. Lemmas 5.4, 5.5, 5.6 show that T_1, T_2, T_3 are asymptotic to $(nh)^{-1}C_1$, $h^4 C_2$ and 0, respectively, a.s. and uniformly in c and h, where $C_1 = \kappa_2\tau_2(t-s)$, $C_2 = (\frac{1}{4})(\kappa_1\tau_1)^2 \int_{s<x<t} \gamma(x)^2 g(x)^{-1}\,dx$. Similarly, it may be proved that T_4, T_5, T_6 equal $o\{(nh)^{-1} + h^4\}$ a.s., uniformly in c and h. Therefore, $T = \tau_1^{-2}\{(nh)^{-1}C_1 + h^4 C_2\} + o\{(nh)^{-1} + h^4\}$ uniformly in c and h, which proves the theorem. □

Lemma 5.1. *Let* $-\infty < \alpha < \beta < \infty$ *and* $\delta > 0$. *Under the conditions of Theorem* 2.1,

$$\sup_{0\le c\le c_0} \sup_{h\in\mathscr{H}} (nh)^{1/2} \sup_{s\le x\le t,\, \alpha\le\theta\le\beta} |A(x,\theta)| = O(n^{\delta}) \quad a.s., \tag{5.5}$$

$$\sup_{0\le c\le c_0} \sup_{h\in\mathscr{H}} (nh)^{1/2} \sup_{s\le x\le t,\, \alpha\le\theta\le\beta} \sup_{\theta_2:\,|\theta_1-\theta_2|\le n^{-\delta}} |A(x,\theta_1) - A(x,\theta_2)| \to 0 \quad a.s. \tag{5.6}$$

PROOF. We derive only (5.6). Let $\mathscr{C} \subseteq [0, c_0]$, $\mathscr{H}' \subseteq \mathscr{H}$, $\mathscr{S} \subseteq [s, t]$ and $\mathscr{I} \subseteq [\alpha, \beta]$ be the lattice subsets with the same endpoints as the respective containing intervals, each consisting of n^a regularly spaced elements, for arbitrarily large but fixed $a > 0$. Write sup* for the supremum over $c \in \mathscr{C}$, $h \in \mathscr{H}'$, $x \in \mathscr{S}$, $\theta_1 \in \mathscr{I}$, and $\theta_2 \in \mathscr{I}$ such that $|\theta_1 - \theta_2| \leq n^{-\delta}$. We shall prove shortly that

$$\sup{}^*(nh)^{1/2}|A(x, \theta_1) - A(x, \theta_2)| \to 0 \quad \text{a.s.} \tag{5.7}$$

Given this result, the following facts may be used to prove that (5.7) continues to hold if sup* is replaced by the supremum over all $c \in [0, c_0]$, $h \in \mathscr{H}$, $x \in (s, t)$, $\theta_1 \in (\alpha, \beta)$ and $\theta_2 \in (\alpha, \beta)$ such that $|\theta_1 - \theta_2| \leq n^{-\delta}$; Hölder continuity of $\chi(x, \theta)$ in θ; Hölder continuity of H in h;

$$|A(x, \theta_1) - A(x, \theta)| \leq |A(x, \theta_1) - A(x, \theta_2)| + 2\{\chi(x, \theta_1) - \chi(x, \theta_2)\},$$

for $\theta_1 \leq \theta \leq \theta_2$;

$$\begin{aligned} &|H_{c_1}(x, \theta_1) - H_{c_2}(x, \theta)| \\ &\quad \leq C(c_2 - c_1)^{1/2} \sum_{i=1}^{n} a_i(x) \\ &\qquad + 2 \sum_{i=1}^{n} a_i(x)\left[I\{|Y_i - \theta| \leq (c_2 - c_1)^{1/2}\} - P\{|Y_i - \theta| \leq (c_2 - c_1)^{1/2}|\}\right], \end{aligned}$$

for $0 \leq c_1 \leq c_2$. The lemma follows.

It remains to prove (5.7). Write $P(\cdot|\mathscr{X})$ and $E(\cdot|\mathscr{X})$ to denote probability and expectation conditional on the x_i's. Define $\psi_1(u) = I(u > c)$, $\psi_2(u) = I(u \leq -c)$, $\psi_3(u) = c^{-1}uI(-c < u \leq c)$, $H_j(x, \theta) = \sum_{i=1}^{n} a_i(x)\psi_j(x_i - \theta)$, $\chi_j = E(H_j|\mathscr{X})$ and $A_j = H_j - \chi_j$. Then $A = A_1 - A_2 + A_3$, and so it suffices to derive (5.7) for A replaced by A_j, in each of the cases $j = 1, 2, 3$. By judicious use of Markov's and Rosenthal's inequalities, first prove that for all $\lambda > 0$,

$$\sup P\left\{(nh)^{1/2}|A_j(x, \theta_1) - A_j(x, \theta_2)| > n^{-\delta/4}\middle|\mathscr{X}\right\} = O(n^{-\lambda}) \quad \text{a.s.}, \tag{5.8}$$

where the supremum is over $0 \leq c \leq c_0$, $h \in \mathscr{H}$, $x \in [s, t]$, $\theta_1 \in [\alpha, \beta]$ and $\theta_2 \in [\alpha, \beta]$ such that $|\theta_1 - \theta_2| \leq n^{-\delta}$. Put

$$\mathscr{E}_n = \left\{\sup{}^*(nh)^{1/2}|A_j(x, \theta_1) - A_j(x, \theta_2)| > n^{-\delta/4}\right\}.$$

It follows from (5.8) that for all $\lambda > 0$, $P(\mathscr{E}_n|\mathscr{X}) = O(n^{-\lambda})$ a.s., which entails $P(\mathscr{E}_n \text{ i.o.}) = 0$. The latter implies (5.7). □

LEMMA 5.2. *Let $\{\eta_n\}$ denote a sequence of positive constants converging to zero. Under the conditions of Theorem* 2.1,

$$\sup_{0 \leq c \leq c_0} \sup_{h \in \mathscr{H}} (nh)^{1/2} \sup_{s \leq x \leq t} \sup_{|\theta - m(x)| \leq \eta_n} |\chi'(x, \theta) + \tau_1 g(x)| \to 0 \quad a.s.$$

PROOF. Define $B(\theta) = \sup_{0 \le c \le c_0} |\zeta'(\theta) - \tau_1(c)|$,

$$\eta'_n = \sup_{h \in \mathscr{H}} \sup_{s \le x \le t} \sup_{y:\,|x-y| \le C_1 h} |m(x) - m(y)|, \qquad \eta''_n = \eta_n + \eta'_n.$$

Then $\eta'_n \to 0$ as $n \to \infty$, $B(\theta) \to 0$ as $\theta \to 0$, and

$$|\chi'(x,\theta) + \tau_1(c)g(x)| \le B(\eta''_n) \sum_{i=1}^{n} a_i(x) + \Big\{ \sup_{0 \le c \le c_0} |\tau_1(c)| \Big\} \left| \sum_{i=1}^{n} a_i(x) - g(x) \right|.$$

The lemma follows from these results and the fact that

$$\sup_{h \in \mathscr{H}} \sup_{s \le x \le t} \left| \sum_{i=1}^{n} a_i(x) - g(x) \right| \to 0 \quad \text{a.s.},$$

the latter being proved by standard techniques from nonparametric density estimation. □

LEMMA 5.3. *Let* $-\infty < \alpha < \beta < \infty$. *Under the conditions of Theorem* 2.1 *and for some* $\delta > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_{s \le x \le t} \sup_{\alpha \le \theta \le \beta} |\chi(x,\theta)| = O(n^{-\delta}) \quad a.s.$$

PROOF. Use Bernstein's or Rosenthal's inequality to show that for all δ, $\lambda > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_{s \le x \le t} \sup_{\alpha \le \theta \le \beta} P\big\{|\chi(x,\theta) - E\chi(x,\theta)| > (nh)^{-1/2} n^{\delta}\big\} = O(n^{-\lambda}).$$

Now apply the lattice argument to demonstrate that for all $\delta > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} (nh)^{1/2} \sup_{s < x < t} \sup_{\alpha < \theta < \beta} |\chi(x,\theta) - E\chi(x,\theta)| = O(n^{-\delta}) \quad \text{a.s.} \tag{5.9}$$

Finally, prove by elementary calculus that for some $\delta > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} h^{-\delta} \sup_{s \le x \le t} \sup_{\alpha \le \theta \le \beta} |E\chi(x,\theta)| = O(1). \tag{5.10}$$ □

LEMMA 5.4. *Under the conditions of Theorem* 2.2,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} \left| h \sum_i{}' A\{x_i, m(x_i)\}^2 g(x_i)^{-2} - \kappa_2 \tau_2 (t-s) \right| \to 0 \quad a.s. \tag{5.11}$$

PROOF. Let $\mathscr{C}$, $\mathscr{H}'$ and $\mathscr{S}$ be as in the proof of Lemma 5.1. Arguing as in that proof we see that it suffices to establish (5.11) when the suprema on the left-hand side are taken over $c \in \mathscr{C}$ and $h \in \mathscr{H}'$. To derive this modified result, we would write $A = A_1 - A_2 + A_3$ as in the argument just preceding

(5.8) and treat separately each component of $A^2 = (A_1 - A_2 + A_3)^2$. However, for notational simplicity, we shall work directly with A^2. Thus, we prove that for some $C, \delta > 0$,

$$P\left[\sup_{c\in\mathscr{C}}\sup_{h\in\mathscr{H}}\left|h\sum_i{}' A\{x_i, m(x_i)\}^2 g(x_i)^{-2} - \kappa_2\tau_2(t-s)\right| > Cn^{-\delta} \text{ i.o.}\right] = 0. \tag{5.12}$$

Define

$$\Delta_j(x) = \psi\{m(x_j) - m(x) + \varepsilon_j\} - \zeta\{m(x_j) - m(x)\},$$

$$D_1(x) = \sum_j a_j(x)^2\left[\Delta_j(x)^2 - E\left\{\Delta_j(x)^2\middle|\mathscr{X}\right\}\right],$$

$$D_2(x) = \sum_j a_j(x)^2\left[E\left\{\Delta_j(x)^2\middle|\mathscr{X}\right\} - E\left\{\Delta_j(x)^2\right\}\right],$$

$$D_3(x) = \sum_j E\left\{a_j(x)^2\Delta_j(x)^2\right\} - (nh)^{-1}\kappa_2\tau_2 g(x),$$

$$D_4(x) = 2\sum\sum_{1\le j<k\le n} a_j(x)a_k(x)\,\Delta_j(x)\,\Delta_k(x).$$

Then $A\{x, m(x)\}^2 - (nh)^{-1}\kappa_2\tau_2 g(x) = \sum_{1\le l\le 3} D_l(x) + 2D_4(x)$. By judicious use of Rosenthal's and Markov's inequalities we may prove that

$$\sup_{0<c\le c_0}\sup_{h\in\mathscr{H}}\sup_{s\le x\le t} P\{nh|D_l(x)| > n^{-\delta}|\mathscr{X}\} = O(n^{-\lambda}) \quad \text{a.s.}, l = 1, 2, 3,$$

$$\sup_{0<c\le c_0}\sup_{h\in\mathscr{H}} P\left\{h\left|\sum_i{}' D_4(x_i)g(x_i)^{-2}\right| > n^{-\delta}\middle|\mathscr{X}\right\} = O(n^{-\lambda}) \quad \text{a.s.},$$

for all $\lambda > 0$. Combining the last three results and using the lattice argument, we may deduce that

$$\sup_{0<c\le c_0}\sup_{h\in\mathscr{H}} P\left(h\left|\sum_i{}'\left[A\{x_i, m(x_i)\}^2 g(x_i)^{-2} - (nh)^{-1}\kappa_2\tau_2 g(x_i)^{-1}\right]\right| > 5n^{-\delta}\middle|\mathscr{X}\right) = O(n^{-\lambda}) \quad \text{a.s.} \tag{5.13}$$

But for all $0 < \delta < \frac{1}{2}$ and $\lambda > 0$, Markov's inequality implies

$$P\left\{\left|n^{-1}\sum_i{}' g(x_i)^{-1} - (t-s)\right| > n^{-\delta}\right\} = O(n^{-\lambda}). \tag{5.14}$$

The desired result (5.12) follows from (5.13) and (5.14). □

LEMMA 5.5. *Define* $\mu(x) = E[\chi\{x, m(x)\}]$. *Under the conditions of Theorem* 2.2,

$$\sup_{0\le c\le c_0}\sup_{h\in\mathscr{H}}\left|n^{-1}h^{-4}\sum_i{}'\mu(x_i)^2 g(x_i)^{-2} - \tfrac{1}{4}(\kappa_1\tau_1)^2\int_s^t \gamma(x)^2 g(x)^{-1}\,dx\right| \to 0 \quad a.s.$$

The proof is by elementary calculus and the law of large numbers.

LEMMA 5.6. *Define* $\chi(x) = \chi\{x, m(x)\}$. *Under the conditions of Theorem* 2.2,

$$\sup_{0\le c\le c_0}\sup_{h\in\mathscr{H}} n^{-1}h^{-4}\sum_i{}'\{\chi(x_i) - \mu(x_i)\}^2 g(x_i)^{-2} \to 0 \quad a.s.$$

PROOF. Put $\chi_1(x) = \sum_{j\le n} a_j(x)\{m(x_j) - m(x)\}$, $\mu_1(x) = E\{\chi_1(x)\}$. Then

$$\chi(x) = \chi_1(x) + \delta_1(x)h^2\sum_{j=1}^n a_j(x),$$

$$\mu(x) = \mu_1(x) + \delta_2(x)h^2\sum_{j=1}^n E\{a_j(x)\},$$

where $\sup\{|\delta_1(x)| + |\delta_2(x)|\} \to 0$. Hence

$$n^{-1}\sum_i{}'\{\chi(x_i) - \mu(x_i)\}^2 g(x_i)^{-2} = U + o(U + h^4),$$

uniformly in c and h, where

$$U = n^{-1}\sum_i{}'\{\chi_1(x_i) - \mu_1(x_i)\}^2 g(x_i)^{-2}$$

$$\le \Big\{\inf_{s\le x\le t} g(x)\Big\}^{-2}\sup_{s\le x\le t}\{\chi_1(x) - \mu_1(x)\}^2 = O(n^{-1+\delta}),$$

for all $\delta > 0$, uniformly in c and h. (Use Markov's inequality and the lattice argument.) The lemma is immediate. □

6. Proof of Theorem 3.1. We shall confine our argument to an outline, the details being similar to those of the proofs of Theorems 2.1 and 2.2. By way of notation, recall that $\mathscr{X} = \{x_1, x_2, \cdots\}$ and $H_i(x,\theta) = \sum_{j\ne i} a_j(x)\psi(Y_j - \theta)$, and put $\chi_i(x,\theta) = E\{H_i(x,\theta)|\mathscr{X}\} = \sum_{j\ne i} a_j'(x)\zeta\{m(x_j) - \theta\}$, $A_i(x,\theta) = H_i(x,\theta) - \chi_i(x,\theta)$. Observe that

$$\hat{U}(h,\psi) = U(h,\psi) + n^{-1}\sum_i{}'m(x_i)\varepsilon_i$$

$$+ n^{-1}\sum_i{}'\{\hat{m}_i(x_i) - \hat{m}(x_i)\}m(x_i)$$

$$+ n^{-1}\sum_i{}'\{\hat{m}_i(x_i) - m(x_i)\}\varepsilon_i.$$

Lemmas 6.2 and 6.3 treat, respectively, the third and fourth terms on the right-hand side and show that those terms equal $o\{(nh)^{-1} + h^4\}$ uniformly in c and h with probability one. The theorem is immediate. □

LEMMA 6.1. *Under the conditions of Theorem* 3.1, *we have for some* $u > 0$,

$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \sup_i{}'|\hat{m}_i(x_i) - m(x_i)| = O(n^{-u}) \quad a.s., \tag{6.1}$$
$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \sup_i{}'|\hat{m}(x_i) - m(x_i)| = O(n^{-u}) \quad a.s.$$

PROOF. We establish only (6.1). Put $\theta_i = m(x_i)$ and $\hat{\theta}_i = \hat{m}_i(x_i)$. If $\hat{\theta}_i > \theta_i + \delta$, then, since K is nonnegative and ψ is nondecreasing, $H_i(x_i, \theta_i + \delta) \ge 0$. Hence for each c, h, i and δ, $P(\hat{\theta}_i > \theta_i + \delta|\mathscr{X}) = P\{A_i(x_i, \theta_i + \delta) \ge -\chi_i(x_i, \theta_i + \delta)|\mathscr{X}\}$. Taking $\delta = n^{-u}$, we may now prove that for some $C > 0$,

$$P(\hat{\theta}_i > \theta_i + \delta|\mathscr{X}) \le P\left\{A_i(x_i, \theta_i + \delta) \ge Cn^{-2u}\sum_{j\ne i} a_j(x_i)\middle|\mathscr{X}\right\}. \tag{6.2}$$

Put $\hat{g}(x) = \sum_{j\le n} a_j(x)$. We may show by methods standard for density estimators that

$$\sup_{h\in\mathscr{H}}\ \sup_{s\le x\le t} |\hat{g}(x) - g(x)| = O(n^{-v}) \quad \text{a.s.},$$

for some $v > 0$. Hence by Markov's inequality applied to (6.2),

$$\sup_i{}' P(\hat{\theta}_i > \theta_i + \delta|\mathscr{X})\Big/E\left[\{A_i(x_i, \theta_i + \delta)\}^{2l}|\mathscr{X}\right] = O(n^{4lu}),$$

for each $l \ge 1$. Application of Rosenthal's inequality yields

$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}} (nh)^l \sup_i{}' E\left[\{A_i(x_i, \theta_i + \delta)\}^{2l}|\mathscr{X}\right] = O(1) \quad \text{a.s.},$$

with probability one. Therefore, if u is chosen sufficiently small,

$$\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \sup_i{}' P(\hat{\theta}_i > \theta_i + \delta|\mathscr{X}) = O(n^{-\lambda}) \quad \text{a.s.},$$

for all $\lambda > 0$. Using the lattice argument we may now deduce that, with $\delta = n^{-u}$,

$$P\left[\sup_{0\le c\le c_0}\ \sup_{h\in\mathscr{H}}\ \sup_i{}'\{\hat{m}_i(x_i) - m(x_i)\} > \delta \text{ infinitely often}\right] = 0.$$

A similar proof may be used to establish the same identity with $\hat{m}_i(x_i) - m(x_i)$ replaced by $m(x_i) - \hat{m}_i(x_i)$. The lemma follows from these two results. □

LEMMA 6.2. *Under the conditions of Theorem* 3.1 *and for* $c_n = n^{-\eta'}$ *with* η' *sufficiently small*,

$$\sup_{c_n\le c\le c_0}\ \sup_{h\in\mathscr{H}} h\left|\sum_i{}'\{\hat{m}_i(x_i) - \hat{m}(x_i)\}m(x_i)\right| \to 0 \quad a.s.$$

PROOF. Since $\psi = \psi_c$ is a continuous function of $c > 0$, then it suffices to take the supremum over rational c. In this event $\psi'(u)$ exists for all irrational u and all values c under consideration. This means, for example, that with probability one, $\psi'(\varepsilon_j)$ is well-defined for $j \geq 1$ and all values of c under consideration. Bearing that in mind, define $\hat{\theta}_i = \hat{m}_i(x_i)$, $\hat{\omega}_i = \hat{m}(x_i)$, $\hat{d}_i = \psi(Y_i - \hat{\omega}_i)$, $d_i = \psi(\varepsilon_i)$, $b_i = \sum_{j \neq i} a_j(x_i)\psi'(\varepsilon_j)$, $\beta_i = \tau_1 \sum_{j \neq i} a_j(x_i)$,

$$\hat{b}_i = \sum_{j \neq i} a_j(x_i) \int_0^1 \psi'\{Y_j - \hat{\theta}_i - t(\hat{\omega}_i - \hat{\theta}_i)\}\, dt.$$

Since $-(\hat{\omega}_i - \hat{\theta}_i)\hat{b}_i = H_i(x_i, \hat{\omega}_i) = (nh)^{-1}K(0)\hat{d}_i$, then

$$\begin{aligned} nh(\hat{\omega}_i - \hat{\theta}_i) = K(0)\{&d_i\beta_i^{-1} + d_i(b_i^{-1} - \beta_i^{-1}) \\ &+ (\hat{d}_i - d_i)b_i^{-1} + \hat{d}_i(\hat{b}_i^{-1} - b_i^{-1})\}. \end{aligned} \tag{6.3}$$

It may be shown by arguments similar to those used to prove Lemma 5.1 that with probability one,

$$\sup_{c_n \leq c \leq c_0} \sup_{h \in \mathscr{H}} n^{-1} \left| \sum_i{}' d_i \beta_i^{-1} m(x_i) \right| \to 0, \tag{6.4}$$

$$\sup_{c_n \leq c \leq c_0} \sup_{h \in \mathscr{H}} n^{-1} \sum_i{}' \left| d_i(b_i^{-1} - \beta_i^{-1}) m(x_i) \right| \to 0, \tag{6.5}$$

$$\sup_{c_n \leq c \leq c_0} \sup_{h \in \mathscr{H}} n^{-1} \sum_i{}' \left| (\hat{d}_i - d_i) b_i^{-1} m(x_i) \right| \to 0, \tag{6.6}$$

$$\sup_{c_n \leq c \leq c_0} \sup_{h \in \mathscr{H}} n^{-1} \sum_i{}' \left| \hat{d}_i(\hat{b}_i^{-1} - b_i^{-1}) m(x_i) \right| \to 0, \tag{6.7}$$

provided $c_n = n^{-\eta'}$ and η' is sufficiently small. Combining (6.3)–(6.7) we deduce the lemma. □

LEMMA 6.3. *Under the conditions of Theorem* 3.1 *and for* $c_n = n^{-\eta'}$ *with* η' *sufficiently small*,

$$\sup_{c_n \leq c \leq c_0} \sup_{h \in \mathscr{H}} (h + n^{-1}h^{-4}) \left| \sum_i{}' \{\hat{m}_i(x_i) - m(x_i)\} \varepsilon_i \right| \to 0 \quad a.s.$$

PROOF. Define $\hat{\theta}_i = \hat{m}_i(x_i)$, $\theta_i = m(x_i)$,

$$\tilde{b}_i = \sum_{j \neq i} a_j(x_i) \int_0^1 \psi'\{Y_j - \theta_i - t(\hat{\theta}_i - \theta_i)\}\, dt.$$

Since $\hat{\theta}_i - \theta_i = H(x_i, \theta_i)\tilde{b}_i^{-1}$, then it suffices to prove that

$$\sup_{n^{-\eta'} \leq c \leq c_0} \sup_{h \in \mathscr{H}} h \left| \sum_i{}' H_i(x_i, \theta_i) \tilde{b}_i^{-1} \varepsilon_i \right| \to 0 \quad \text{a.s.} \tag{6.8}$$

Define $U_{ij} = \theta_j - \theta_i$ if $a_j(x_i) \neq 0$ and $U_{ij} = 0$ otherwise, and put $\mathbf{U}_i = (U_{i1}, \dots, U_{in})$, $V_i = \theta_i - \hat{\theta}_i$, $M_i = \mu_i(\mathbf{U}_i, V_i)$ and $D_i = d_i(\mathbf{U}_i, V_i)$. Note that $\tilde{b}_i = B_i(\mathbf{U}_i, V_i)$ and $\sup_i'(\max_{j \le n}|U_{ij}| + |V_i|) \to 0$. Hence by (6.8),

$$\sup_{n^{-\eta'} \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_i{}' (nh)^{1/2} |\tilde{b}_i - M_i| = O(n^{\eta' + \Delta}) \quad \text{a.s.} \tag{6.9}$$

It may be proved as for Lemma 6.1 that for some $u > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_i{}' |M_i - \tau_1 g(x_i)| = O(n^{-u}) \quad \text{a.s.} \tag{6.10}$$

Since $\tilde{b}_i = M_i + D_i$, then results (6.9) and (6.10) together imply that if η' and Δ are chosen so small that $\eta' + \Delta < \frac{1}{2}\eta$ and $l \ge 1$,

$$\sup_{n^{-\eta'} \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_i{}' (nh)^l \left| \tilde{b}_i^{-1} - \sum_{k=0}^{2l-1} (-1)^k D_i^k M_i^{-(k+l)} \right| = O(n^{2l(\eta' + \Delta)}) \quad \text{a.s.}$$

From this formula and the fact that for each $\delta > 0$,

$$\sup_{0 \le c \le c_0} \sup_{h \in \mathscr{H}} \sup_i{}' \left\{ (nh)^{-1} + h^4 \right\}^{-1/2} |H_i(x_i, \theta_i)| = O(n^{\delta}) \quad \text{a.s.},$$

we may prove that

$$\begin{aligned} &\sup_{n^{-\eta'} \le c \le c_0} \sup_{h \in \mathscr{H}} h \left| \sum_i{}' H_i(x_i, \theta_i) \tilde{b}_i^{-1} \varepsilon_i \right| \\ &\quad \le 2l \sup_{0 \le k \le 2l-1} \sup_{n^{-\eta'} \le c \le c_0} \sup_{h \in \mathscr{H}} h |\xi(k)| + o(1) \quad \text{a.s.}, \end{aligned} \tag{6.11}$$

where $\xi(k) = \sum_i' H_i(x_i, \theta_i) d_i(\mathbf{U}_i, V_i)^k \mu_i(\mathbf{U}_i, V_i)^{-(k+1)} \varepsilon_i$. It may be proved after tedious moment calculations that for each k,

$$\sup_{n^{-\eta'} \le c \le c_0} \sup_h h |\xi(k)| \to 0 \quad \text{a.s.} \tag{6.12}$$

The desired result (6.8) follows from (6.11) and (6.12). □

Acknowledgments. We are extremely grateful to Barbara La Scala for performing the computations associated with this paper and to two referees and an Associate Editor for very helpful comments. This work was done while the second author held a visiting fellowship at the Australian National University funded by the Mathematical Sciences Research Visitors Program of that university.

REFERENCES

EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.

HÄRDLE, W. (1984). How to determine the bandwidth of some linear smoothers in practice. In *Robust and Nonlinear Time Series Analysis*. (J. Franke, W. Härdle and D. Martin, eds.). *Lecture Notes in Statist.* **26**. Springer, New York.

HÄRDLE, W. (1989). *Applied Nonparametric Regression*. Cambridge Univ. Press, New York.

HÄRDLE, W. and GASSER, T. (1984). Robust non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.

HÄRDLE, W. and TSYBAKOV, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16** 120–135.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

O'SULLIVAN, F. (1988). Robust smoothing. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **8** 170–173. Wiley, New York.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12** 1285–1297.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 111–147.

STONE, M. (1977). Asymptotics for and against cross-validation. *Biometrika* **64** 29–35.

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

FACULTY OF MATHEMATICS
THE OPEN UNIVERSITY
WALTON HALL
MILTON KEYNES MK76AA
UNITED KINGDOM