The Annals of Statistics
1994, Vol. 22, No. 2, 867–885

ON CURVE ESTIMATION BY MINIMIZING MEAN ABSOLUTE DEVIATION AND ITS IMPLICATIONS[1]

BY JIANQING FAN AND PETER HALL

University of North Carolina and Australian National University

The local median regression method has long been known as a robustified alternative to methods such as local mean regression. Yet, its optimal statistical properties are largely unknown. In this paper, we show via decision-theoretic arguments that a local weighted median estimator is the best least absolute deviation estimator in an asymptotic minimax sense, under L_1-loss. We also study asymptotic efficiency of the local median estimator in the class of all possible estimators. From a practical viewpoint our results show that local weighted medians are preferable to histogram estimators, since they enjoy optimality properties which the latter do not, under virtually identical smoothness assumptions on the underlying curve. Among smoothing methods that are adapted to functions with only one derivative, little is to be gained by using an estimator other than one based on the local median.

1. Introduction. Local weighted median regression methods have been considered as an effective robust nonparametric smoother. See, for example, Härdle and Gasser (1984), Tsybakov (1986), Truong (1989), Hall and Jones (1990), Chaudhuri (1991) and references therein. Yet, its asymptotic optimality properties are largely unknown. Tsybakov (1986) showed that the local weighted median estimator achieves optimal rates of convergence. However, it still remains unknown how efficient the local weighted median smoother is.

To motivate our study, let us first consider L_2-theory. Consider the regression model,

$$Y_i = g(x_i) + \varepsilon_i, \qquad E\varepsilon_i = 0, \qquad \operatorname{var}(\varepsilon_i) = 1,$$

where $\{\varepsilon_i\}$ are independent and identically distributed random variables, and $\{x_i\}$ are design points. The popular approaches in the vast nonparametric literature include the methods of kernel, splines (polynomial splines and smoothing splines) and orthogonal series. For details, we refer to the recent books by Härdle (1994) and Wahba (1990), among others. All of these procedures are locally linear, in the sense that the estimator admits the form $\widehat{g}(x) = \sum_i w_i Y_i$, where the weights are independent of Y_i. In other words, all popular regression methods in the literature are a subclass of the class of smoothers $\{\widehat{a}\}$ which

Received April 1992; revised November 1992.

[1]Supported by NSF Grant DMS-85-05550 and completed while authors visited the Mathematical Sciences Research Institute, Berkeley, California.

AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20.

Key words and phrases. Curve estimation, kernel, mean absolute deviation, median, minimax risk.

minimise

$$\sum_{i=1}^{n} w_i(Y_i - a)^2, \tag{1.1}$$

for weights w_i. Here the optimal weights are derived from a quadratic kernel function, the so-called Bartlett–Epanechnikov kernel, and the formula for w_i depends on x_i.

Different methods assign different weights. Among linear smoothers, Fan (1993) found a best linear procedure and proved that the local polynomial regression is an asymptotic linear minimax procedure. He also showed that there is little to gain if one attempts to use a nonlinear procedure. In other words, the class of linear smoothers is small enough for one to be able to identify the best, and is large enough for there to be little loss of efficiency by restricting attention to this subclass of estimators. See Ibragimov and Khas'minskii (1984), Donoho (1994) and Donoho and Liu (1991) for comprehensive studies of the minimax theory of both linear and nonlinear procedures. See also Brown and Low (1992).

The above motivations suggest that the best local median smoother should be optimal in the class of least absolute deviation (LAD) smoothers. To fix this idea, let $\widehat{\Theta}$ denote the class of all possible estimators $\widehat{\theta} = \widehat{\theta}_w$ that are obtained by minimising the weighted sum of absolute deviations [compare with (1.1)]

$$\sum_{i=1}^{n} w_i|Y_i - \theta|, \tag{1.2}$$

for an arbitrary choice of nonnegative weights $w = (w_1, \dots, w_n)$. Here, the weights do not depend on the random variables $\{Y_i\}$. Throughout the paper, we call $\widehat{\Theta}$ the class of LAD estimators.

To simplify our discussion, we consider estimating the median regression function $g(x)$ at the point $x = 0$. Define the class of unknown regression functions by

$$\mathcal{G}_a = \{g\colon\ |g(x) - g(0)| \le a|x| \ \text{ for all } x\},$$

where a is a positive constant. We limit ourselves to this class to avoid technicalities; possible extensions are discussed in Section 3. Let R_{LAD} denote the LAD minimax risk:

$$R_{\mathrm{LAD}} = \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{g \in \mathcal{G}_a} E_g|\hat{\theta} - g(0)|.$$

We then find an asymptotic formula for R_{LAD} for this highly nonlinear class of regression estimators and show that the local weighted median smoother with a triangular kernel (2.5) achieves the minimax lower bound. In other words, the local weighted median is the best LAD in an asymptotic minimax sense. (Henceforth, for the sake of simplicity, we shall omit the adjective "weighted" from this description.)

How much can one possibly gain by searching for a smoother other than the local median smoother? To answer this question, let

$$R_N = \inf_{\hat{\theta}} \sup_{g \in \mathcal{G}_a} E_g \left| \hat{\theta} - g(0) \right|$$

be the minimax risk, for estimating $g(0)$, among all possible procedures. Let $r_{N/L}$, defined by (2.8), be the ratio between the minimax risk of the nonlinear procedure and that of the linear procedure for the bounded normal mean problem, under L_1-loss. According to Donoho (1994), $r_{N/L} \geq 1/1.23$, obtained by Liu via numerical computation. Then, we show that

$$1 \geq \frac{R_N}{R_{\mathrm{LAD}}} \geq r_{N/L} \left(\frac{I_\varepsilon}{4f^2(0)} \right)^{1/3} \{1 + o(1)\}, \tag{1.3}$$

where $f(0)$ and I_ε are the density and the Fisher information of the random variable ε. The first factor can be interpreted as the loss of efficiency by using the "linear class" of estimators, and the second factor as the loss of efficiency by using the least absolution deviation method instead of local maximum likelihood. See Remark 2.3 for details. The second factor of (1.3) equals unity if the error distribution is double exponential. In this case, the local maximum likelihood and the local LAD method are the same.

In the above discussion of minimax risks, we keep the distribution of the error ε fixed. Since the least absolute deviation method does not use knowledge of the error distribution, it is also reasonable to consider the minimax risk which takes the error distribution to be a nuisance parameter. If we allow the error distribution to vary in such a way that $f(0) \geq f_0$, then we have

$$1 \geq \frac{R_N^*}{R_{\mathrm{LAD}}^*} \geq r_{N/L}\{1 + o(1)\} \geq 0.813\{1 + o(1)\}, \tag{1.4}$$

where R_N^* is defined similarly to R_N except that the class of unknown functions is now defined by

$$\{(g, f)\colon g \in \mathcal{G}_a,\ f(0) \geq f_0 \text{ and } f \text{ is equicontinuous at } 0\}. \tag{1.5}$$

In other words, in a more global minimax sense (which allows the error distribution to vary), there is little to be gained by using a nonlinear procedure. The justification is given in Remark 2.1.

This conclusion is of practical significance, but the data-analytic implications of our work extend beyond such results. The main practical competitor with the local median, under virtually identical smoothness assumptions and enjoying a broadly similar convergence rate, is the histogram estimator. However, since histogram methods do not enjoy the same optimality properties as those based on local medians, then the latter approach is preferable. In this sense, the relationship between local median estimators and histogram estimators is similar to that between local mean estimators and kernel estimators. Fan (1993) showed that, while kernel estimators enjoy more popular appeal than their counterparts based on local means, they do not enjoy the optimality properties of the latter. In practical work on curve estimation, the local median and local mean should be seen as significant competitors with histogram and kernel estimators, respectively.

Our minimax scenario is stimulated by the pioneering work of Ibragimov and Khas'minskii (1984), Donoho (1994), Donoho and Liu (1991) and Fan (1993). However, the current study is also very different. For example, quadratic loss was used in most of the above studies; linear procedures were singled out for specific consideration; and Gaussian white noise models were assumed. Although L_1-loss was also considered by Donoho (1994), the statistical procedure that he used is still linear. Yet, the LAD estimators studied in this paper are certainly nonlinear.

The paper is organised as follows. Section 2 gives the main results on minimax theory. Section 3 discusses the possible extensions and implications of our main results. The technical proofs are given in Section 4.

2. Bounds for mean absolute deviation. Assume that the pairs (x_i, Y_i), $1 \le i \le n$, are generated by the model

$$Y_i = g(x_i) + \varepsilon_i,$$

where $g \in \mathcal{G}_a$; the sequence $x_1, x_2, \ldots$ (which is taken to be fixed, i.e., is conditioned upon) represents a realisation of a sequence $X_1, X_2, \ldots$ of independent and identically distributed random variables from a population whose distribution function admits a bounded derivative d in a neighbourhood of the origin, satisfying

$$d \text{ is continuous at the origin, and } d(0) > 0; \tag{2.1}$$

and the ε_i's are independently and identically distributed, with a bounded density f satisfying

$$f \text{ is continuous at the origin, and } f(0) > 0. \tag{2.2}$$

The median of ε_i is zero. Let Z denote a standard normal random variable, and define

$$G(u) = u^{-1/3} E|Z + u|, \qquad u > 0.$$

Let $u_0 = 0.65735\ldots$ denote that value of u which minimises G, and let $\min_u G(u) = G(u_0) = 1.109064\ldots$ be the minimum value of the function $G(u)$. The value u_0 and $G(u_0)$ were computed via numerical integration. Set

$$h = n^{-1/3}\{2a^2 f(0)^2 d(0)/3\}^{-1/3} u_0^{2/3}. \tag{2.3}$$

We begin by describing a lower bound to the best worst-case convergence rate of LAD estimators of $g(0)$.

THEOREM 2.1. *Assume the conditions stated above, in particular* (2.1) *and* (2.2). *Then,*

$$\inf_{\widehat{\theta}\in\widehat{\Theta}} \sup_{g\in\mathcal{G}_a} E_g\left|\widehat{\theta}-g(0)\right| \geq \{1+o(1)\}n^{-1/3}a^{1/3}\{18f(0)^2 d(0)\}^{-1/3}G(u_0). \tag{2.4}$$

Next we describe an upper bound, which shows that the "optimal" mean absolute deviation suggested by Theorem 2.1 is asymptotically attained by the LAD estimator which uses the weights

$$w_i = \begin{cases} h - |x_i|, & \text{if } |x_i| \leq h, \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

where h is given by (2.3).

THEOREM 2.2. *Assume the conditions of Theorem* 2.1. *Let w be given by* (2.5), *and let $\widehat{\theta}_0$ denote the estimator constructed using that particular choice of weights. Then,*

$$\sup_{g\in\mathcal{G}_a} E_g\left|\widehat{\theta}_0-g(0)\right| \leq \{1+o(1)\}n^{-1/3}a^{1/3}\{18f(0)^2 d(0)\}^{-1/3}G(u_0). \tag{2.6}$$

Combining the inequalities (2.4) and (2.6), and noting that the lower and upper bounds are asymptotically equivalent, we see that in each case the inequality may be replaced by an equality. In other words, the local median smoother with the triangular kernel (2.5) is the best LAD estimators and the asymptotic LAD minimax risk is given by

$$R_{\text{LAD}} = \{1+o(1)\}n^{-1/3}a^{1/3}\{18f(0)^2 d(0)\}^{-1/3}G(u_0),$$

where $G(u_0) = 1.109064\ldots$. To our knowledge, this is the first time that minimax risk has been derived for the class of highly nonlinear smoothers.

The value $G(u_0)$ is indeed closely related to the bounded normal mean problem under L_1-loss. Let

$$\rho_L(t) = \inf_b \sup_{|\theta|\leq t} E|bZ-\theta|, \qquad \rho_N(t) = \inf_\delta \sup_{|\theta|\leq t} E|\delta(Z)-\theta|, \qquad Z\sim N(\theta,1), \tag{2.7}$$

be the linear minimax risk and minimax risk for estimating θ, knowing $|\theta|\leq t$. Set

$$r_{N/L} = \inf_t \rho_N(t)/\rho_L(t). \tag{2.8}$$

The similar quantities under L_2-loss have been extensively studied in the literature. See, for example, Levit (1980), Bickel (1981) and Donoho, Liu and MacGibbon (1990). Using the fact that $E|bZ-\theta|$ is convex and symmetric in

θ, its maximum over $\{|\theta| \le t\}$ must attain at $\theta = t$. This, together with some simple algebra, shows that

$$\sup_t \rho_L(t)t^{-1/3} = G(u_0)(2/3)^{2/3}(1/3)^{1/3}.$$

Thus, we can express R_{LAD} as

$$R_{\text{LAD}} = \{1 + o(1)\}\{3a/8f^2(0)d(0)n\}^{1/3} \sup_t \rho_L(t)t^{-1/3}. \tag{2.9}$$

We next examine the minimax lower bound among all possible estimators. Let us impose some condition on the density f of the random variable ε.

CONDITION 2.1. The function $\{f(x+t)/f(x)\}^{1/2}$ is differentiable in quadratic mean at $t = 0$ with the derivative function $v(x)$ [see Le Cam (1985), pages 574–578], that is,

$$E\{\sqrt{f(\varepsilon + t)/f(\varepsilon)} - 1 - tv(\varepsilon)\}^2 = o(t), \quad \text{as } t \to 0. \tag{2.10}$$

Condition 2.1 is used to ensure experiments under consideration converge to a Gaussian shift experiment, as defined by Le Cam (1985). See Lemma 4.3 for more details. As a consequence of (2.10), the Fisher information

$$I_\varepsilon \equiv 4Ev^2(\varepsilon) < \infty.$$

THEOREM 2.3. *If the density satisfies Condition* 2.1, *then we have the following minimax bound*:

$$R_N \ge \{1 + o(1)\}\{3a/2n\,d(0)I_\varepsilon\}^{1/3} \sup_{t>0} t^{-1/3}\rho_N(t). \tag{2.11}$$

Indeed, Theorem 2.3 is a special case of the following setup. Let $\mathcal{G}$ be a class of the median regression functions. Let the modulus of continuity of functional $g(0)$ be

$$\omega_{\mathcal{G}}(\varepsilon) = \sup\left\{|g_0(0) - g_1(0)|: g_i \in \mathcal{G},\ \|g_0 - g_1\| \le \varepsilon\right\},$$

where $\|\cdot\|$ is the usual L_2-norm. In many nonparametric setups such as Theorem 2.3 [see Donoho and Liu (1991) for the calculation of $\omega_{\mathcal{G}_a}(\cdot)$], the following condition holds.

CONDITION 2.2. $\mathcal{G}$ is convex and the extremal pair of $\omega_{\mathcal{G}}(\varepsilon\|H\|)$ can be chosen to satisfy

$$g_0(x) - g_1(x) = \varepsilon^p H(x/\varepsilon^{2q})\{1 + o(1)\}, \tag{2.12}$$

uniformly in x as $\varepsilon \to 0$ for a continuous function $H(\cdot)$ with a compact support, where $q = 1 - p$.

Note that (2.12) implies that $\|g_0 - g_1\| = \varepsilon\|H\|\{1 + o(1)\}$. Thus, Condition 2.2 implies that

$$\omega_{\mathcal{G}}(\|H\|\varepsilon) = |g_0(0) - g_1(0)|\{1 + o(1)\} = |H(0)|\varepsilon^p\{1 + o(1)\}. \tag{2.13}$$

We are now ready to state Theorem 2.4.

THEOREM 2.4. *Suppose that Conditions* 2.1 *and* 2.2 *hold. Let* $\ell(\cdot)$ *be a convex loss function satisfying* $\ell(tx) = t^r\ell(x)$ *for all* $t > 0$. *Then, we have the following minimax lower bound for estimating* $g(0)$ *in the class* $\mathcal{G}$ *under the loss function* $\ell(\cdot)$:

$$\begin{aligned} R_N(\mathcal{G}; \ell) &\equiv \inf_{\hat{g}} \sup_{g \in \mathcal{G}} E\ell(\hat{g} - g(0)) \\ &\geq \{1 + o(1)\}\left[\omega_{\mathcal{G}}\left(\{I_\varepsilon d(0) n\}^{-1/2}\right)2^{-q}\right]^r \sup_t t^{-qr}\rho_N(t; \ell), \end{aligned} \tag{2.14}$$

where $\rho_N(t; \ell)$ *is defined similarly to* (2.7) *except the loss now is* ℓ.

REMARK 2.1. The least absolute deviation method does not use the knowledge of error distribution. Thus, it is also appropriate to study the minimax risk among the class of unknown parameters defined by (1.5). Then, the LAD minimax risk is given by [see (2.6) and (2.9)]

$$R^*_{\text{LAD}} \leq \{1 + o(1)\}\{3a/8f_0^2 d(0)n\}^{1/3} \sup_t \rho_L(t)t^{-1/3}.$$

Now, if we take $f(x) = f_0 \exp(-2f_0|x|)$ in Theorem 2.3, then the minimax lower bound is

$$R^*_N \geq \{1 + o(1)\}\{3a/8f_0^2 d(0)n\}^{1/3} \sup_t \rho_N(t)t^{-1/3}.$$

Thus, (1.4) holds. There is not much gain (at most by a factor of 1.23) by using an estimator other than the local median estimator.

REMARK 2.2. Keeping the distribution of ε fixed, the ratio of the minimax risks is given by (1.3). [Compare (2.9) and (2.11).] The second factor can be interpreted as the loss of efficiency inherent in using the least absolute deviation estimator instead of using the maximum likelihood estimator:

$$\hat{g}_{\text{MLE}} = \arg\max_a \left\{ \sum_{i=1}^n w_i \log f(Y_i - a) \right\}. \tag{2.15}$$

The first factor can be explained as the loss of efficiency when restricting the estimator to the class of estimators obtained by minimising linear forms in the log-likelihood, as defined by (2.15).

REMARK 2.3. If we take the L_r-loss in Theorem 2.4, then

$$R_N(\mathcal{G},\ell) \geq \{1+o(1)\}\left(\omega_{\mathcal{G}}\left[\{I_\varepsilon d(0)n\}^{-1/2}\right]2^{-q}\right)^r \sup_t t^{-rq}\rho_N(t;\ell).$$

In particular, if we take $\mathcal{G} = \mathcal{G}_a$, then

$$\inf_{\hat\theta} \sup_{g\in\mathcal{G}_a} E|\hat\theta - g(0)|^r \geq \{3a/2d(0)I_\varepsilon\}^r \sup_{t>0} t^{-r/3}\rho_N(t;r) \quad \text{for } r\geq 1,$$

where

$$\rho_N(t;r) = \inf_\delta \sup_{|\theta|\leq t} E|\delta(Z)-\theta|^r, \qquad Z \sim N(\theta,1).$$

3. Discussions. We have derived the asymptotic minimax risk under L_1-loss for a specific class of smoothness $\mathcal{G}_a$. Yet, the minimax scenario does not end here. We mention here several possible extensions, to indicate the general phenomena behind the theory that we presented. The proofs of these conjectures are beyond the scope of this paper.

Convex loss function and the class of linear estimators. Consider again estimating $g(0)$ in the class of mean functions $\mathcal{G}_a$. Let ℓ be a convex function with a unique minimiser at 0. Let the function g be defined by

$$g(x) = \arg\min_b E\{\ell(Y-b) \mid X=x\}.$$

Let (X_i, Y_i), $i=1,\dots,n$, be a random sample from the population (X,Y). Define the class of "linear estimators" $\hat\theta = \hat\theta_w$ that are obtained by [compare with (1.1) and (1.2)]

$$\hat\theta = \arg\min_\theta \left\{\sum_{i=1}^n w_i\ell(Y_i-\theta)\right\}, \tag{3.1}$$

where w_i are nonnegative weights depending only on the covariates $\{X_i\}$. Here, the linearity refers to the fact that the deviation function is linear in the function $\ell(Y_i-\theta)$. In the L_2-case, this linearity entails that the resulting estimators depend linearly on $\{Y_j\}$. We would expect that the estimator (3.1) with triangular kernel defined by (2.5), where the bandwidth h minimises the corresponding maximum risk of the estimator, is an asymptotic "linear" minimax estimator and that there is little gain by using nonlinear procedures.

To make the above statement more precisely, suppose ℓ satisfies

$$\ell(t) = c|t|^r\{1+o(1)\} \quad \text{as } t\to 0.$$

With $Z\sim N(0,1)$, set

$$G_\ell(u) = E\{u^{-r/3}\ell(u+Z)\}, \quad u_\ell = \arg\min_u G_\ell(u).$$

Then, the asymptotic "linear minimax" risk under the convex loss function ℓ is expected to be

$$R_L = \{1 + o(1)\}\{2a\sigma_\ell^2/9n\,d(0)\}^{r/3}G_\ell(u_\ell),$$

and the optimal weight is given by (2.5) with $h = \{6u_0^2\sigma_\ell^2/d(0)\,na^2\}^{1/3}$, where with $\psi(t) = E[\ell\{Y - g(0) + t\} \mid X = 0]$,

$$\sigma_\ell^2 = E\Big(\big[\ell'\{Y - g(0)\}\big]^2 \Big| X = 0\Big)\Big/\{\psi''(0)\}^2.$$

This conjecture is stimulated by (2.9) together with the expressions, given by Tsybakov (1986) and Fan, Hu and Truong (1994), of the asymptotic bias and variances that arise in estimating g. We would also expect that Remark 2.1 holds with appropriate changes. For example, the constraint (1.5) should be changed to $\sigma_\ell^2 \le B$, instead of $f(0)$ bounded from below.

Other classes of constraints. Consider, for example,

$$\mathcal{G}_a^* = \{g\colon\ |g(x) - g(0) - g'(0)x| \le ax^2/2 \text{ for all } x\}.$$

Define the class of "linear smoothers" $\widehat{\alpha}$ which together with $\widehat{\beta}$ minimises

$$\sum_{i=1}^n w_i\ell(Y_i - \alpha - \beta X_i), \quad \text{or even} \quad \sum_{i=1}^n w_i\ell(Y_i - \alpha).$$

Of course, $\widehat{\alpha}$ estimates $g(0)$. We would expect that the optimal weight is Epanechnikov-kernel weight:

$$w_i = \left(1 - |X_i/h|^2\right)_+,$$

where h is chosen to minimise risk. We would also expect that one cannot much improve on the best linear smoother.

We finally remark that the constant factors in the asymptotic minimax risks (linear and nonlinear) given in Section 2 depend on the class of unknown functions only through the modulus of continuity as explained by Donoho (1994) and Donoho and Liu (1991).

4. Proofs.

PROOF OF THEOREM 2.1. We preface the proof with two lemmas. Their proofs are conceptually straightforward, although algebraically tedious, and may be found in Fan and Hall (1992).

LEMMA 4.1. *Let $w_1, \ldots, w_n$ be chosen to minimise $E|\widehat{g}(0) - g(0)|$, with*

$$g(x) = \begin{cases} a|x|, & \text{if } |x| \le c, \\ ac, & \text{if } |x| > c. \end{cases} \tag{4.1}$$

If $c > 0$ is taken sufficiently small and the conditions of the theorem are imposed on the design and error distributions, then, for each $\eta > 0$,

$$\left(\sum_{|x_i|>\eta} w_i\right) \Big/ \left(\sum_{i=1}^{n} w_i\right) \to 0$$

as $n \to \infty$.

LEMMA 4.2. *Fix $a, c > 0$, and define $g(x) = a|x|$ if $|x| \le c$, $g(x) = ac$ otherwise. Assume that the design variables x_i are as stipulated in Theorem 2.1. Let $\ell_n^{(1)}$ and $\ell_n^{(2)}$ denote, respectively, the infimum and supermum of $\sum w_i g(x_i)$, subject to $\sum w_i^2 = 1$. Let $\ell_n^{(1)} \le \ell_n \le \ell_n^{(2)}$. Then the maximum $m(\ell)$ of $\sum w_i$, subject to $\sum w_i g(x_i) = \ell$ and $\sum w_i^2 = 1$, satisfies*

$$n^{-1/3}\ell_n^{-1} m(\ell_n) \to 0 \quad \text{if } \ell_n \to \infty,$$
$$n^{-1/3} m(\ell_n) \to 0 \quad \text{if } \ell_n \to 0.$$

Let g be defined by (4.1), and let $b, c > 0$ be as defined in the proof of Lemma 4.1. Define $J_i(u) = \operatorname{sgn}\{\varepsilon_i + g(x_i) - g(0) + u\}$, and set $A \equiv E|\hat{\theta} - g(0)| = E|\hat{\theta}|$. Then

$$A = \int_0^{\infty} \left[P\left\{ \sum_{i=1}^{n} w_i J_i(u) \le 0 \right\} + P\left\{ \sum_{i=1}^{n} w_i J_i(-u) > 0 \right\} \right] du. \tag{4.2}$$

Assume that the weights w_i are chosen to minimize A when g is given by (4.1). In view of Lemma 4.1, $\delta \equiv \{\sum w_i g(x_i)\}/(\sum w_i) \to 0$ as $n \to \infty$. Define $\mu(u) \equiv \sum w_i E J_i(u)$, and let

$$V(u) \equiv \sum_{i=1}^{n} w_i \{J_i(u) - E J_i(u)\}. \tag{4.3}$$

Now $\mu(\delta v) = 2f(0)(1+v)\sum w_i g(x_i) - R_1(v)$, where, for each $v_0 > 0$,

$$\sup_{|v| \le v_0} |R_1(v)| = o\left\{ \sum_{i=1}^{n} w_i g(x_i) \right\}. \tag{4.4}$$

In this notation, for each $v_0 > 0$,

$$\begin{aligned} A &= \delta \int_0^{\infty} \left[P\{V(\delta v) \le -\mu(\delta v)\} + P\{V(-\delta v) > -\mu(-\delta v)\} \right] dv \\ &\ge \delta \int_0^{v_0} \left[P\left\{ V(\delta v) \le -2f(0)(1+v) \sum_{i=1}^{n} w_i g(x_i) + R_1(v) \right\} \right. \\ &\quad \left. + P\left\{ V(-\delta v) > -2f(0)(1-v) \sum_{i=1}^{n} w_i g(x_i) + R_1(-v) \right\} \right] dv. \end{aligned} \tag{4.5}$$

We may standardise the w_i's by asking that

$$\sum_{i=1}^{n} w_i^2 = 1. \tag{4.6}$$

This assumption will be imposed in all that follows. By passing to a subsequence of n-values, if necessary, we may assume without loss of generality that, for some $0 \leq \lambda \leq \infty$,

$$\sum_{i=1}^{n} w_i g(x_i) \to \lambda. \tag{4.7}$$

We claim that $\lambda = 0$ and $\lambda = \infty$ are both impossible. Let us first treat the case $\lambda = \infty$. Here, if $0 < v < 1$ then, since $E\{V(-\delta v)^2\}$ is uniformly bounded [indeed, by (4.6), is bounded by unity], the second probability in the integrand of (4.5) converges to unity. Therefore, $A \geq \{1 + o(1)\}\delta$. In the context of Lemma 4.2, let $\ell_n = \sum w_i g(x_i)$ for the particular choice of weights w_i that minimise A. Then δ exceeds $a\ell_n/m(\ell_n)$, where m is as in Lemma 4.2; and by that lemma, $n^{1/3}\ell_n/m(\ell_n) \to \infty$. Thus, $n^{1/3}A \to \infty$. However, Theorem 2.2 shows that it is possible to choose the weights w_i so that $A = O(n^{-1/3})$. Hence, the assumption $\lambda = \infty$ contradicts our assumption that the w_i's are chosen to minimise A when g is given by (4.1).

Next we show that $\lambda = 0$ is also impossible. If $v > 0$ then, writing $C_1 = \sup f$, we have

$$0 \leq EJ_i(\delta v) = 2\int_0^{g(x_i)+\delta v} f(y)\,dy \leq 2C_1\{g(x_i) + \delta v\},$$

with a similar inequality for $v < 0$. It follows that

$$|\mu(\delta v)| \leq \sum_{i=1}^{n} w_i |EJ_i(\delta v)| \leq 2C_1\big(1 + |v|\big)\sum_{i=1}^{n} w_i g(x_i).$$

Hence, if $v > 0, q(v) \equiv P\{V(\delta v) \leq -\mu(\delta v)\} + P\{V(\delta v) > \mu(-\delta v)\}$ satisfies

$$\begin{aligned} q(v) \geq\ & P\left\{V(\delta v) \leq -2C_1(1+v)\sum_{i=1}^{n} w_i g(x_i)\right\} \\ & + P\left\{V(-\delta v) > 2C_1(1+v)\sum_{i=1}^{n} w_i g(x_i)\right\}. \end{aligned} \tag{4.8}$$

Let $L_i = \operatorname{sgn}\{\varepsilon_i + g(x_i)\}$ and $W = \sum w_i(L_i - EL_i)$. Then,

$$\begin{aligned} \operatorname{var}\{V(\delta v) - W\} &\leq \sum_{i=1}^{n} w_i^2 E\{J_i(\delta v) - L_i\}^2 \\ &\leq \sum_{i=1}^{n} w_i^2 P\{|\varepsilon + g(x_i)| \leq \delta|v|\} \\ &\leq C_1\delta|v|\sum_{i=1}^{n} w_i^2 = C_1\delta|v|. \end{aligned} \tag{4.9}$$

If $\xi_1 > 0$ and

$$(4.10)\qquad |v| \le \xi_1 \left\{ C_1 \sum_{i=1}^{n} w_i\, g(x_i) \right\}^{-1},$$

then $\delta|v| \le \xi_1 C_1^{-1}$, and so $\operatorname{var}\{V(\delta v) - W\} \le \xi_1$. Therefore, defining $\xi = \xi_1^{1/3}$, we see that if (4.10) holds, then

$$\begin{aligned} P\{V(\delta v) < -\xi\} &\ge P(W \le -2\xi) - P\{|V(\delta v) - W| > \xi\} \\ &\ge P(W \le -2\xi) - \xi. \end{aligned}$$

Similarly, $P\{V(-\delta v) > \xi\} \ge P(W > 2\xi) - \xi$. Thus,

$$P\{V(\delta v) < -\xi\} + P\{V(-\delta v) > \xi\} \ge P(|W| > 2\xi) - 2\xi.$$

Combining this result with (4.8) we see that, for each $\xi > 0$,

$$(4.11)\qquad \int_0^\infty q(v)\,dv \ge \left\{P(|W| > 2\xi) - 2\xi\right\}\xi^3 \left\{ C_1 \sum_{i=1}^{n} w_i g(x_i) \right\}^{-1}.$$

To conclude our proof that $\lambda = 0$ is impossible, we may assume without loss of generality that, for some $0 \le \lambda' \le 1$ and $0 < \sigma \le 1$, $\max w_i \to \lambda'$ and $\operatorname{var}(W) \to \sigma^2$. [Otherwise, use a subsequence argument. If c is sufficiently small, then $\sigma \ne 0$. Assumption (4.6) guarantees that $\sigma \le 1$.] Suppose the indices i have been permuted, so that $\max w_i = w_n$. If $\lambda' = 0$, then Lindeberg's theorem may be used to prove that W is asymptotically normal $N(0, \sigma^2)$. Writing Z for a standard normal variable, and taking $\xi > 0$ to be so small that $q_0 \equiv \{P(\sigma|Z| > 2\xi) - 2\xi\}\xi^3 > 0$, we see from (4.11) that

$$(4.12)\qquad \delta \int_0^\infty q(v)\,dv \ge \{1 + o(1)\} q_0 \delta \left\{ C_1 \sum_{i=1}^{n} w_i g(x_i) \right\}^{-1} \sim q_0 \left(C_1 \sum_{i=1}^{n} w_i \right)^{-1}.$$

If $\lambda' > 0$, then

$$\begin{aligned} P(|W| > 2\xi) &\ge \inf_u P(w_n|K_n + u| > 2\xi) \\ &= \inf_u P\left[w_n \left|I\{\varepsilon + g(x_n) > 0\} + u\right| > \xi\right]. \end{aligned}$$

For $\xi > 0$ sufficiently small, the right-hand side above is bounded above 2ξ as $n \to \infty$. Taking ξ of this form and defining q_0 to equal the lim inf (as $n \to \infty$) of the right-hand side, minus 2ξ, we deduce once more from (4.9) that (4.10) holds. Now apply the second part of Lemma 4.2, with $\ell_n = \Sigma w_i g(x_i)$, to show that $n^{1/3}\Sigma w_i \to \infty$. Hence, $n^{1/3}A \to \infty$, which (as in the case $\lambda = \infty$) we know from Theorem 2.2 to be false. We may therefore assume that, in formula (4.7), $0 < \lambda < \infty$.

It now follows from (4.4), (4.5) and (4.7) that

$$A \geq \int_0^{v_0} \Big[P\{V(\delta v) \leq -2f(0)\lambda(1+v)+R_2(v)\} + P\{V(-\delta v) > -2f(0)\lambda(1-v)+R_2(-v)\}\Big]\,dv, \tag{4.13}$$

where $0 < \lambda < \infty$ and $\delta' \equiv \sup_{|v|\leq v_0}|R_2(v)| = o(1)$. Let W be as defined two paragraphs above. In view of (4.9),

$$\sup_{|v|\leq v_0} P\{|V(\delta v) - W| > \delta^{1/4}\} \leq \delta^{-1/2} \sup_{|v|\leq v_0} E\{V(\delta v) - W\}^2 \leq C_1 v_0 \delta^{1/2}.$$

Hence,

$$P\{V(\delta v) \leq -2f(0)\lambda(1+v)+R_2(v)\} \geq P\{W \leq -2f(0)\lambda(1+v) - (\delta^{1/4}+\delta')\} - C_1 v_0 \delta^{1/2},$$

uniformly in $|v| \leq v_0$. Similarly, $P\{V(\delta v) > -2f(0)\lambda(1-v)+R_2(-v)\}$ may be bounded below, whence by (4.13),

$$\begin{aligned} A &\geq \delta \int_0^{v_0} \Big[P\{W \leq -2f(0)\lambda(1+v) - (\delta^{1/4}+\delta')\} \\ &\qquad + P\{W > -2f(0)\lambda(1-v) - (\delta^{1/4}+\delta')\}\Big]\,dv - 2C_1 v_0 \delta^{3/2} \\ &= \delta \int_0^{v_0} \Big[P\{W \leq -2f(0)\lambda(1+v)\} + P\{W > -2f(0)\lambda(1-v)\}\Big]\,dv + o(\delta). \end{aligned}$$

Noting that $E(W^2)$ is bounded, and defining $W' = \{2f(0)\lambda\}^{-1}W$, we obtain

$$\begin{aligned} &\int_{v_0}^{\infty} \Big[P\{W \leq -2f(0)\lambda(1+v)\} + P\{W > -2f(0)\lambda(1-v)\}\Big]\,dv \\ &\qquad \leq E(W'+1)^2 \int_{v_0}^{\infty} v^{-2}\,dv \to 0, \end{aligned}$$

uniformly in n as $v_0 \to \infty$. Therefore,

$$\begin{aligned} A &\geq \{1+o(1)\}\delta \int_0^{\infty} \{P(W'+1 \leq -v) + P(W'+1 > v)\}\,dv \\ &= \{1+o(1)\}\delta E|W'+1| = \{1+o(1)\}EB, \end{aligned} \tag{4.14}$$

where

$$B = \left(\sum_{i=1}^{n} w_i\right)^{-1} \left|\{2f(0)\}^{-1} \sum_{i=1}^{n} w_i(L_i - EL_i) + \sum_{i=1}^{n} w_i g(x_i)\right|.$$

Finally, we show how to choose $w_1, \ldots, w_n$ to minimise B asymptotically, subject to $\sum w_i^2 = 1$, when g is given by (4.1). The method of infinitesimal calculus, with a Lagrange multiplier, provides the formula

$$w_i = \begin{cases} \alpha\{a\beta - g(x_i)\}, & \text{if } g(x_i) \le a\beta, \\ 0, & \text{otherwise,} \end{cases}$$

with $\alpha, \beta > 0$ to be selected. A little asymptotic analysis shows that, in order to attain a minimum, we must have $\beta = \beta(n) \to 0$ as $n \to \infty$, in which case $(\max w_i^2)/(\sum w_i^2) \to 0$, which implies that, since $\sum w_i^2 = 1$,

$$\sum_{i=1}^{n} w_i^2(L_i - EL_i) \to N(0,1)$$

in distribution. Therefore, if the random variables Z is standard normal,

$$\inf_w B \sim t_1^{-1} E\left|\{2f(0)\}^{-1}Z + t_2\right|,$$

where

$$t_1 = a\alpha \sum_{i=1}^{n} (\beta - |x_i|)I(|x_i| \le \beta) \sim na\alpha\beta^2\, d(0),$$

$$t_2 = a^2\alpha \sum_{i=1}^{n} (\beta - |x_i|)|x_i|I(|x_i| \le \beta) \sim \tfrac{1}{3}na^2\alpha\beta^3\, d(0).$$

The constraint $\sum w_i^2 = 1$ is equivalent to

$$1 = na^2\alpha^2 \sum_{i=1}^{n} (\beta - |x_i|)^2 I(|x_i| \le \beta) \sim \tfrac{2}{3}na^2\alpha^2\beta^3\, d(0).$$

Therefore, $\alpha = \{\frac{2}{3}na^2\beta^3\, d(0)\}^{-1/2}$, whence

$$t_1 \sim \{\tfrac{3}{2}\, d(0)\, n\beta\}^{1/2}, \qquad t_2 \sim a\{\tfrac{1}{6}\, d(0)\, n\beta^3\}^{1/2},$$

$$\inf_w B \sim \{\tfrac{3}{2}\, d(0)\}^{-1/2} n^{-1/3}(n\beta^3)^{-1/6} \times E\left|\{2f(0)\}^{-1}Z + a\{\tfrac{1}{6}\, d(0)\}^{1/2}(n\beta^3)^{-1/2}\right|.$$

Define $\gamma(s) \equiv s^{-1/6}E|\{2f(0)\}^{-1}Z + a\{\frac{1}{6}\, d(0)\}^{1/2}s^{1/2}|$, and let $s_0 > 0$ minimise γ. Taking $\beta \sim (n^{-1}s_0)^{1/3}$, we see that

$$\inf_w B \sim n^{-1/3}\{\tfrac{3}{2}\, d(0)\}^{-1/2}\gamma(s_0), \qquad \alpha \to \{\tfrac{2}{3}a^2 s_0^3\, d(0)\}^{-1/2}.$$

From this result and (4.14) we conclude that

$$A \ge \{1 + o(1)\}n^{-1/3}\{\tfrac{3}{2}\, d(0)\}^{-1/2}\gamma(s_0),$$

which is equivalent to (2.4). This completes the proof of Theorem 2.1. □

PROOF OF THEOREM 2.2. Let us standardise for scale in definition (2.5) of w_i by redefining

$$w_i = \Big\{(h - |x_i|)I(|x_i| \le h)\Big\}\Bigg\{\sum_{j=1}^{n}(h - |x_j|)^2 I(|x_j| \le h)\Bigg\}^{-1/2}$$

This definition will be used throughout the arguments below. Observe that if $g \in \mathcal{G}_a$, then,

$$\left(\frac{\partial}{\partial\theta}\right)\sum_{i=1}^{n} w_i|Y_i - \theta| = \sum_{i=1}^{n} w_i \operatorname{sgn}\{\varepsilon_i + g(x_i) - g(0) + g(0) - \theta\}$$
$$\begin{cases} \le -\sum_{i=1}^{n} w_i \operatorname{sgn}\{\varepsilon_i - a|x_i| + g(0) - \theta\} \\ \ge -\sum_{i=1}^{n} w_i \operatorname{sgn}\{\varepsilon_i + a|x_i| + g(0) - \theta\}. \end{cases}$$

Therefore, with $\widehat{\theta}_\pm$ denoting the solutions θ of the equations

$$\sum_{i=1}^{n} w_i \operatorname{sgn}\{\varepsilon_i \pm a|x_i| + g(0) - \theta\} = 0,$$

we have $\widehat{\theta}_- \le \widehat{\theta}_0 \le \widehat{\theta}_+$. It is straightforward to prove that $E_g\{\widehat{\theta}_\pm - g(0)\}^2$ is uniformly bounded, and so

$$\sup_{n \ge 1}\sup_{g \in \mathcal{G}_a} E_g\{\widehat{\theta}_0 - g(0)\}^2 < \infty. \tag{4.15}$$

Let $\eta > 0$ be a very small but fixed constant, and observe that

$$P_g\{|\widehat{\theta}_0 - g(0)| > \eta\} = P_g\{\widehat{\theta}_0 > g(0) - \eta\} + P_g\{\widehat{\theta}_0 \le g(0) + \eta\}$$
$$= P\Bigg\{\sum_{i=1}^{n} w_i J_i(-\eta) > 0\Bigg\} + P\Bigg\{\sum_{i=1}^{n} w_i J_i(\eta) \le 0\Bigg\}, \tag{4.16}$$

where $J_i(u)$ is as before. Bernstein's inequality may be used to prove that, for each $k > 0$, both of the probabilities in (4.16) equal $O(n^{-k})$ uniformly in $g \in \mathcal{G}_a$. Therefore,

$$\sup_{g \in \mathcal{G}_a} P_g\{|\widehat{\theta}_0 - g(0)| > \eta\} = O(n^{-k}), \tag{4.17}$$

for all $\eta, k > 0$. Combining (4.15) and (4.17), we obtain

$$\sup_{g \in \mathcal{G}_a} E_g\Big[|\widehat{\theta}_0 - g(0)| I\Big\{|\widehat{\theta}_0 - g(0)| > \eta\Big\}\Big]$$
$$\le \sup_{g \in \mathcal{G}_a}\Big[E_g\Big\{|\widehat{\theta}_0 - g(0)|^2\Big\} P_g\Big\{|\widehat{\theta}_0 - g(0)| > \eta\Big\}\Big]^{1/2} = O(n^{-k}),$$

for all $\eta, k > 0$. Thus, it suffices to prove Theorem 2.2 with $E_g|\widehat{\theta}_0 - g(0)|$ replaced by

$$m_{g,\eta} \equiv \int_0^{\eta} \Big[P_g\{\widehat{\theta}_0 > g(0) + u\} + P_g\{\widehat{\theta}_0 \le g(0) - u\}\Big]\,du,$$

for $\eta > 0$ arbitrarily small.

It follows that we may let $\eta = \eta(n) \to 0$ sufficiently slowly, and this we shall do below.

Set

$$\delta = \left(a\sum_{i=1}^n w_i|x_i|\right)\Big/\left(\sum_{i=1}^n w_i\right), \qquad \mu(u) = \sum_{i=1}^n w_i E J_i(u).$$

Define $V(u)$ by (4.3), and note that

$$m_{g,\eta} = \delta \int_0^{\eta/\delta} \Big[P_g\{V(\delta v) \le -\mu(\delta v)\} + P_g\{V(-\delta v) > -\mu(-\delta v)\}\Big]\,dv.$$

Now, $\sum w_i|x_i| \to C$, where $0 < C < \infty$, and so

$$\begin{aligned} \mu(\delta v) &= 2\sum_{i=1}^n w_i \int_0^{g(x_i) - g(0) + \delta v} f(y)\,dy \\ &= 2\sum_{i=1}^n w_i\{g(x_i) - g(0) + \delta v\} f(0) + o\big(1 + |v|\big), \end{aligned} \tag{4.18}$$

uniformly in $g \in \mathcal{G}_a$ and $|v| \le \eta/\delta$. The nonuniform version of the Berry–Esseen theorem may be used to prove that, if Z is a standard normal random variable,

$$\sup_{g\in\mathcal{G}_a}\ \sup_{|v|\le\eta/\delta}\ \sup_{-\infty<y<\infty}\ (1 + |y|)^3\big|P_g\{V(\delta v) \le y\} - P(Z \le y)\big|\,dv \to 0$$

as $n \to \infty$. Hence,

$$\begin{aligned} m_{g,\eta} &= \delta \int_0^{\eta/\delta} \Big[P\{Z \le -\mu(\delta v)\} + P\{Z > -\mu(-\delta v)\}\Big]\,dv \\ &\quad + o\left(\delta \int_0^{\eta/\delta} \Big[\{1 + |\mu(\delta v)|\}^{-3} + \{1 + |\mu(-\delta v)|\}^{-3}\Big]\,dv\right), \end{aligned} \tag{4.19}$$

uniformly in $g \in \mathcal{G}_a$.

In view of (4.18), $|\mu(\delta v)| \ge C_1|v|\sum w_i|x_i| \ge C_2|v|$ uniformly in $C_3 \le |v| \le \eta/\delta$, for $C_3 > 0$ sufficiently large. Therefore, the term o in (4.19) equals $o(\delta) = o(n^{-1/3})$. Hence, by (4.19),

$$m_{g,\eta} = \delta \int_0^{\eta/\delta} P\left[\left|Z - \sum_{i=1}^n w_i\{g(x_i) - g(0)\}\right| > av\sum_{i=1}^n w_i|x_i|\right]dv + o(\delta),$$

uniformly in $g \in \mathcal{G}_a$. Thus,

$$\sup_{g \in \mathcal{G}_a} m_{g,\eta} \sim \delta \int_0^\infty P\left(\left| Z + a \sum_{i=1}^n w_i |x_i| \right| > av \sum_{i=1}^n w_i |x_i| \right) dv$$
$$= E\left| \left(\sum_{i=1}^n w_i \right)^{-1} Z + \delta \right|.$$

From this point, the proof of Theorem 2.2, may be completed by using routine methods to develop an asymptotic approximation to the last-written expectation. □

PROOF OF THEOREM 2.3. The modulus of continuity for $g(0)$ in the class of $\mathcal{G}_a$ is

$$\omega_{\mathcal{G}_a}(\varepsilon) = 3^{1/3} a^{1/3} \varepsilon^{2/3}.$$

See Donoho and Liu (1991). Now, applying Theorem 2.4, we obtain the result. □

PROOF OF THEOREM 2.4. The first step consists of the following result, whose proof may be found in Fan and Hall (1992).

LEMMA 4.3. *Let $a_{n,i}$ be a sequence of constants such that*

$$\max_{1 \le i \le n} a_{n,i} \to 0 \quad \text{and} \quad \sum_{i=1}^n a_{n,i}^2 \to A > 0.$$

Then, based on the observations from the model $Y_i = a_{n,i}\theta + \varepsilon_i$, where $\{\varepsilon_i\}$ are i.i.d. with density f satisfying Condition 2.1, *we have*

$$\inf_{\hat\theta} \sup_{|\theta| \le 1/2} E\ell(\widehat\theta - \theta) \ge \{1 + o(1)\} \inf_{\delta} \sup_{|\theta| \le 1/2} E\ell\{\delta(Z) - \theta\}, \qquad Z \sim N(\theta, 1/AI_\varepsilon).$$

The above inequality is indeed equality when ℓ is bounded. If moreover, $\ell(tx) = t^r \ell(x)$, for all $t > 0$, then

$$\inf_{\hat\theta} \sup_{|\theta| \le 1/2} E\ell(\widehat\theta - \theta) \ge \{1 + o(1)\}(AI_\varepsilon)^{-r/2} \rho_N\{(AI_\varepsilon)^{1/2}, \ell\}.$$

The idea of the proof of Theorem 2.4 is to use the "hardest one-dimensional subproblem," due to Donoho and Liu (1991). For fixed g_0, g_1 in $\mathcal{G}$, let

$$g_\theta(x) = \theta g_0(x) + (1 - \theta) g_1(x) = g_1 + \theta(g_0 - g_1).$$

The convexity assumption ensures that $m_\theta \in \mathcal{G}$ for *all* $0 \le \theta \le 1$. Without loss of generality assume that $g_0(0) - g_1(0) \geq 0$. Thus,

$$
\begin{aligned}
R_N(\mathcal{G}, \ell) &\geq \inf_{\hat\theta} \sup_{0 \leq \theta \leq 1} E\ell\{\widehat{\theta} - g_\theta(0)\} \\
&= \{g_0(0) - g_1(0)\}^r \inf_{\hat\theta} \sup_{0 \leq \theta \leq 1} E\ell(\widehat{\theta} - \theta) \qquad (4.20)\\
&= \{g_0(0) - g_1(0)\}^r \inf_{\hat\theta} \sup_{|\theta| \leq 1/2} E\ell(\widehat{\theta} - \theta).
\end{aligned}
$$

The observations are based on the submodel $Y_i = g_\theta(X_i) + \varepsilon_i$ or equivalently based on $Y_i^* = Y_i - g_1(x_i) = \theta\{g_0(x_i) - g_1(x_i)\} + \varepsilon_i$. Now, take the pair g_1 and g_0 as in Condition 2.2 with $\varepsilon = \varepsilon_n$ to be determined later. Then, by the definition of g_1 and g_0 [see (2.13)], $|g_1(0) - g_0(0)| = \{1 + o(1)\}\omega_{\mathcal{G}}(\|H\|\varepsilon_n)$. Thus, expression (4.20) can be written as

$$
R_N(\mathcal{G}, \ell) \geq \{1 + o(1)\}\left\{\omega_{\mathcal{G}}(\|H\|\varepsilon_n)\right\}^r \inf_{\hat\theta} \sup_{|\theta| \leq 1/2} E\ell(\widehat{\theta} - \theta), \qquad (4.21)
$$

based on the observations $Y_i^* = a_i\theta + \varepsilon_i$, where $a_i = \varepsilon_n^p H(x_i/\varepsilon_n^{2q})$. Take $\varepsilon_n = cn^{-1/2}$ with a positive constant c. Then, it can easily be shown that

$$
\begin{aligned}
\sum_{i=1}^n a_i^2 &= \varepsilon_n^{2p}\left[n\, d(0)\, \varepsilon_n^{2q}\|H\|^2 + O\left\{\left(n\varepsilon_n^{2q}\right)^{1/2}\right\}\right] \\
&= \{1 + o(1)\}A,
\end{aligned}
$$

where $A = d(0)\|H\|^2 c^2$. By Lemma 4.3 and (4.21), we have

$$
R_N(\mathcal{G}, \ell) \geq \{1 + o(1)\}\left\{\omega_{\mathcal{G}}(\|H\|cn^{-1/2})\right\}^r (AI_\varepsilon)^{-r/2}\rho_N\{(AI_\varepsilon)^{1/2}/2; \ell\}. \qquad (4.22)
$$

Set $t = (AI_\varepsilon)^{1/2}/2 = \{d(0)I_\varepsilon\|H\|^2/4\}^{1/2}c$. By (2.13) and (4.22), we have

$$
R_N \geq \{1 + o(1)\}\left(\omega_{\mathcal{G}}\left[2\{I_\varepsilon\, d(0)\, n\}^{-1/2}\right]\right)^r t^{rp}(2t)^{-r}\rho_N(t; \ell). \qquad (4.23)
$$

Since (4.23) holds for all t, we have

$$
R_N \geq \{1 + o(1)\}\left(\omega_{\mathcal{G}}\left[\{I_\varepsilon\, d(0)\, n\}^{-1/2}\right]\Big/2^q\right)^r \sup_t t^{-rq}\rho_N(t; \ell).
$$

This completes the proof. □

Acknowledgments. This project was undertaken while the authors were visiting the Mathematical Sciences Research Institute, Berkeley, and they are most grateful for the hospitality and financial support of the Institute. The authors would like to thank Professor Le Cam for helpful discussions. Two referees and an Associate Editor also provided very helpful comments, which led to this significantly improved, revised version of the paper.

REFERENCES

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309.

BROWN, L. D. and LOW, M. G. (1992). Asymptotic equivalence of nonparametric regression and white noise. Unpublished manuscript.

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles and implications. *Ann. Statist.* **18** 1416–1437.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216.

FAN, J. and HALL, P. (1992). On curve estimation by minimizing mean absolute deviation, and its implications. Technical Report 036-92, Mathematical Sciences Research Institute, Berkeley, CA.

FAN, J., HU, T. C. and TRUONG, Y. K. (1994). Robust nonparametric function estimation. *Scand. J. Statist.* To appear.

HALL, P. and JONES, M. C. (1990). Adaptive *M*-estimation in nonparametric regression. *Ann. Statist.* **18** 1712–1728.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

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

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1984). On nonparametric estimation of a linear functional in a Gaussian white noise. *Theory Probab. Appl.* **29** 19–32.

LE CAM, L. (1985). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LEVIT, B. Y. (1980). On asymptotic minimax estimates of the second order. *Theory Probab. Appl.* **25** 552–568.

TRUONG, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–617.

TSYBAKOV, A. B. (1986). Robust reconstruction of functions by the local-approximation method. *Problems Inform. Transmission* **22** 133–146.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA A.C.T. 2601
AUSTRALIA