The Annals of Statistics
1993, Vol. 21, No. 1, 196–216

LOCAL LINEAR REGRESSION SMOOTHERS AND THEIR MINIMAX EFFICIENCIES[1]

BY JIANQING FAN

University of North Carolina

In this paper we introduce a smooth version of local linear regression estimators and address their advantages. The MSE and MISE of the estimators are computed explicitly. It turns out that the local linear regression smoothers have nice sampling properties and high minimax efficiency—they are not only efficient in rates but also nearly efficient in constant factors. In the nonparametric regression context, the asymptotic minimax lower bound is developed via the heuristic of the "hardest one-dimensional subproblem" of Donoho and Liu. Connections of the minimax risk with the modulus of continuity are made. The lower bound is also applicable for estimating conditional mean (regression) and conditional quantiles for both fixed and random design regression problems.

1. Introduction. Nonparametric regression provides a useful diagnostic tool for data analysis. A useful mathematical model is to think of estimating a regression function

$$m_f(x_0) = E_f(Y|X = x_0),$$

based on a random sample of data $(X_1, Y_1), \dots, (X_n, Y_n)$ from an unknown joint density $f(\cdot, \cdot)$. For convenience, we will suppress the dependence of the regression function $m_f(\cdot)$ on f. Popular kernel methods for estimating $m(\cdot)$ include the Nadaraya–Watson [Nadaraya (1964) and Watson (1964)] and the Gasser–Müller [Gasser and Müller (1979)] estimators. This paper focuses on studying the asymptotic properties of the local linear regression smoothers. One motivation of introducing this class of estimators is that they repair the drawbacks of the Nadaraya–Watson and Gasser–Müller estimators. See Fan (1992) and Chu and Marron (1991) for additional discussion.

Another important motivation of studying the local linear smoother is to find (nearly) precise minimax risk in the regression setup. With an optimal choice of kernel and bandwidth, the estimator provides a good upper bound on the minimax risk. The lower bound is derived by using the heuristic of the "hardest one dimensional subproblem". In particular, a geometric quantity—modulus of continuity [Donoho (1990) and Donoho and Liu (1991)] is involved in both the lower and upper bound. We show that the minimax

Received May 1990; revised January 1992.

[1]Supported by NSF Grants DMS-90-05905, DMS-91-13527 and DMS-92-03135.

AMS 1991 *subject classifications*. Primary 62G20; secondary 62G05, 62F35.

Key words and phrases. Local linear smoothers, hardest one-dimensional subproblem, minimax risk, modulus of continuity, nonparametric regression.

lower bound is nearly sharp for the following two cases:

1. Bounded two-derivative constraints [see (3.4)].
2. Bounded Lipschitz constraints (see the example in Section 5.1).

These minimax results, on the other hand, give theoretical supports to the intuitively appealing method—local linear smoothers. We would expect, but have not yet shown, that such a lower bound is nearly sharp for other constraints.

We decompose the difficulty of nonparametric regression into two parts: constraints on the regression function itself and constraints on marginal densities and conditional variances. It turns out that the upper bound of the conditional variances and the lower bound of the marginal densities are strongly related to minimax risks. An important application of the lower bound is to determine the efficiency of a regression estimator (see Section 5.1). Even though our attention is focused on random design problems whose marginal densities are also unknown, the lower bound is also applicable for both fixed and random design problems whose marginal distributions are known.

Our approach on the lower bound is related to other work in the literature and in particular the work in white-noise models and density estimation models. See Section 5.2 for further references. What seems innovative in our approach is the use of normal submodels to avoid the technicalities of convergence of experiments [Le Cam (1985)].

The paper is organized as follows. Section 2 introduces local linear smoothers, whose mean squared error (MSE) and mean integrated squared error (MISE) are computed in Section 3. We use the risks of these regression estimators as upper bounds of the minimax risks. The minimax problems are studied in Section 4, paying particular attention to the lower bound. Potential applications of the lower bound are discussed in Section 5. Proofs are deferred until Section 6.

2. Local linear smoothers.

Let us extend the idea of local linear regression. A similar idea can be found in Stone (1977), Cleveland (1979), Lejeune (1985) and Müller (1987). Assume that we know that the second derivative of $m(x)$ exists. Our proposal is to construct a smooth version of a local polynomial: finding a and b to minimize

$$\sum_1^n \left(Y_j - a - b(x_0 - X_j)\right)^2 K\left(\frac{x_0 - X_j}{h_n}\right), \tag{2.1}$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth. Let $\hat{a}$ and $\hat{b}$ be the solution to the weighted least squares problem (2.1). Simple calculation yields

$$\hat{a} = \sum_1^n w_j Y_j \Big/ \sum_1^n w_j,$$

with w_j defined by (2.3). For a technical reason (to avoid zero in the denominator), we use $\hat{m}(x_0)$ to estimate the regression function $m(x_0)$:

$$\hat{m}(x_0) = \sum_1^n w_j Y_j \Big/ \left(\sum_1^n w_j + n^{-2} \right), \tag{2.2}$$

where

$$w_j = K\left(\frac{x_0 - X_j}{h_n} \right) \left(s_{n,2} - (x_0 - X_j) s_{n,1} \right), \tag{2.3}$$

with

$$s_{n,l} = \sum_1^n K\left(\frac{x_0 - X_j}{h_n} \right) (x_0 - X_j)^l, \qquad l = 0, 1, 2. \tag{2.4}$$

A nice feature of estimator (2.2) is that the weight w_j satisfies

$$\sum_1^n (x_0 - X_j) w_j = 0. \tag{2.5}$$

This property ensures that the bias of the estimator does not depend on the derivatives of the marginal density. To see this, we note that by (2.5),

$$E\hat{m}(x_0) = m(x_0) + E \frac{\sum_1^n \left[m(X_j) - m(x_0) - m'(x_0)(X_j - x_0) \right] w_j - n^{-2} m(x_0)}{\left(\sum_1^n w_j + n^{-2} \right)}.$$

If we do Taylor expansions for $m(X_j)$ at point x_0, the second term is of order $O(h_n^2)$, as effective design points have order $(X_j - x_0)^2 = O(h_n^2)$. Thus no derivative of $f_X(\cdot)$ is involved in the preceding calculation (rigorous proof can be found in the proof of Theorem 1).

We refer to estimator (2.2) as a local linear regression smoother for the reasons that it is derived by using a local linear approximation and that it is linear in the response. It will become clear in Section 3 that the local linear smoother has important sampling properties: It adapts to both random and fixed designs and to a variety of design densities $f_X(\cdot)$. Moreover, the best local linear smoother is the best linear smoother in an asymptotic minimax sense (Theorem 5). The local linear smoother also has good finite sampling and design-adaptation properties. See the simulations and discussions in Fan (1992) for details.

Let us briefly mention how the previous idea can be extended to the case where $m(x)$ has a bounded kth derivative. The idea is exactly the same except replacing the linear polynomial in (2.1) by a $(k-1)$-order polynomial. In particular, when $m(x)$ has one derivative, one finds the minimizer of

$$\sum_1^n (Y_j - a)^2 K\left(\frac{x_0 - X_j}{h_n} \right), \tag{2.6}$$

and the resulting estimator is the Nadaraya–Watson estimator. In other words, we use this estimator when the unknown regression function has only a bounded derivative.

3. Asymptotic properties. We now discuss the asymptotic properties of estimator (2.2). Assumptions are as follows.

CONDITION 1.

(i) The regression function $m(\cdot)$ has a bounded second derivative.

(ii) The marginal density $f_X(\cdot)$ of X satisfies $|f(x) - f(y)| \le c|x - y|^{\alpha}$, for $0 < \alpha < 1$, and $f_X(x_0) > 0$.

(iii) The conditional variance $\sigma^2(x) = \mathrm{Var}(Z|X = x)$ is bounded and continuous.

(iv) The kernel $K(\cdot)$ is a bounded and continuous density function satisfying

$$\int_{-\infty}^{\infty} K(y)\,dy = 1, \qquad \int_{-\infty}^{\infty} yK(y)\,dy = 0, \qquad \int_{-\infty}^{\infty} y^2K(y)\,dy \neq 0,$$

$$\int_{-\infty}^{\infty} y^{2r}K(y) < \infty \qquad \text{for } r = 1, 2, \dots .$$

Note that the conditions on $K(\cdot)$ are imposed for the convenience of technical arguments and can be relaxed.

THEOREM 1. *Under Condition* 1, *if* $h_n = dn^{-\beta}$, $0 < \beta < 1$, *then estimator* (2.2) *has the MSE*

$$\begin{aligned} E(\hat{m}(x_0) - m(x_0))^2 &= \frac{1}{4}\left(m''(x_0)\int_{-\infty}^{\infty} u^2K(u)\,du\right)^2 h_n^4 \\ &\quad + \frac{1}{nh_n} f^{-1}(x_0)\sigma^2(x_0)\int_{-\infty}^{\infty} K^2(u)\,du \\ &\quad + o\left(h_n^4 + \frac{1}{nh_n}\right). \end{aligned} \tag{3.1}$$

Let $w(\cdot)$ be a bounded weight function with a compact support $[a, b]$. Then the MISE can be obtained as follows.

THEOREM 2. *Under Condition* 1, *if* $f_X(\cdot)$ *is bounded away from* 0 *on the interval* $[a, b]$, *then the MISE is given by*

$$\begin{aligned} E&\int_{-\infty}^{\infty} (\hat{m}(x) - m(x))^2 w(x)\,dx \\ &= \frac{1}{4}\left[\int_{-\infty}^{\infty} u^2K(u)\,du\right]^2 \int_{-\infty}^{\infty} (m''(x))^2 w(x)\,dx\, h_n^4 \\ &\quad + \frac{1}{nh_n}\int_{-\infty}^{\infty} \frac{\sigma^2(x)}{f_X(x)} w(x)\,dx \int_{-\infty}^{\infty} K^2(u)\,du + o\left(h_n^4 + \frac{1}{nh_n}\right). \end{aligned} \tag{3.2}$$

Simple algebra yields the optimal bandwidth for MISE (3.2)

$$h_{\text{opt}} = \left(\frac{\int_{-\infty}^{\infty} f^{-1}(x)\sigma^2(x)w(x)\,dx \int_{-\infty}^{\infty} K^2(u)\,du}{\left[\int_{-\infty}^{\infty} u^2K(u)\,du\right]^2 \int_{-\infty}^{\infty} (m''(x))^2 w(x)\,dx} \right)^{1/5} n^{-1/5}.$$

We now state a uniform convergence result of Theorem 1.

THEOREM 3. *If the kernel satisfies Condition* 1 *and* $h_n = dn^{-\beta}$, $0 < \beta < 1$, *then*

$$\begin{aligned} &\sup_{f \in \mathscr{C}_2} E_f(\hat{m}(x_0) - m(x_0))^2 \\ &\quad \le \frac{1}{4} C^2 \left(\int_{-\infty}^{\infty} u^2 K(u)\,du \right)^2 h_n^4 \\ &\qquad + \frac{B}{nh_n b} \int_{-\infty}^{\infty} K^2(u)\,du + o\left(h_n^4 + \frac{1}{n_n} \right), \end{aligned} \tag{3.3}$$

where, with C, C^*, B, b, c *and* α *being positive constants,*

$$\begin{aligned} \mathscr{C}_2 = \{ f(\cdot,\cdot)\colon &|m(x) - m(x_0) - m'(x_0)(x - x_0)| \\ &\le C(x - x_0)^2/2,\ |m(x_0)| \le C^* \} \\ \cap \{ f(\cdot,\cdot)\colon &\sigma^2(x) \le B,\ f_X(x_0) \ge b,\ |f_X(x) - f_X(y)| \le c|x - y|^\alpha \}. \end{aligned} \tag{3.4}$$

REMARK 1. The Lipschitz condition in (3.4) is imposed only for a technical reason in the development of the upper bound. The constants α and c will not be involved in the following discussion. The uniform convergence will be used in Section 4 where minimax risk is evaluated [see Theorem 4 and (4.3)]. In MSE terms, by (4.3), local linear smoother (2.2) with h_n minimizing (3.3) has minimax efficiency

$$\begin{aligned} &\left(\frac{R(n, \mathscr{C}_2)}{\max_{f \in \mathscr{C}_2} E(\hat{m}(x_0) - m(x_0))^2} \right)^{1/2} \\ &\qquad \ge 0.529 \left(\int_{-\infty}^{\infty} u^2 K(u)\,du \right)^{-1/5} \left(\int_{-\infty}^{\infty} K^2(u)\,du \right)^{-2/5}, \end{aligned}$$

where $R(n, \mathscr{C}_2) = \inf_{\hat{T}_n} \sup_{f \in \mathscr{C}_2} E_f(\hat{T}_n - m(x_0))^2$ is the minimax risk. For example, estimator (2.2) with the Epanečnikov and normal kernel has at least efficiency 89.6% and 87.8%, respectively.

4. Asymptotic minimax theory. It is well known that estimator (2.2) is optimal in terms of rates of convergence [see Stone (1980)]. More precisely, it is not possible to improve the rate $n^{-4/5}$ uniformly in $\mathscr{C}_2$ defined by (3.4). In other words, the minimax risk

$$R(n, \mathscr{C}_2) = \inf_{\hat{T}_n} \sup_{f \in \mathscr{C}_2} E_f\big(\hat{T}_n - m(x_0)\big)^2 \asymp n^{-4/5},$$

where " $\asymp$ " means that both sides have the same order. That is, *only the rate* of the asymptotic minimax risk is known. Naturally, one would ask how far away from optimal is the constant factor of the local linear smoother. In this section we are going to show that estimator (2.2) is nearly optimal in constant factors as well. Moreover, we will show that it is the best linear smoother in a large class of linear methods. These results are new in nonparametric regression context. Indeed, without using the local linear smoother, it is not easy to give a precise evaluation of the minimax risk $R(n, \mathscr{C}_2)$.

4.1. *An upper bound of minimax risk.* An obvious upper bound of $R(n, \mathscr{C}_2)$ is (3.3). Minimizing the right-hand side of (3.3) yields an optimal choice of bandwidth and kernel function:

$$h_n^{(1)} = \left(\frac{15B}{bC^2 n}\right)^{1/5}, \qquad K_0(x) = \frac{3}{4}[1 - x^2]_+. \tag{4.1}$$

Substituting them into (3.3) yields a minimax upper bound:

$$R(n, \mathscr{C}_2) \le \frac{3}{4} 15^{-1/5} C^{2/5} \left(\frac{B}{bn}\right)^{4/5} (1 + o(1)). \tag{4.2}$$

The right-hand side of (4.2) is the risk of the estimator $\hat{m}^*(x_0)$ defined by (2.2) with bandwidth and kernel given by (4.1).

THEOREM 4. *An upper bound of the asymptotic minimax risk is given by* (4.2). *Moreover, the estimator $\hat{m}^*(x_0)$ has asymptotic minimax efficiency at least* 89.6%:

$$\frac{R(n, \mathscr{C}_2)}{\sup_{f \in \mathscr{C}_2} E_f\big(\hat{m}^*(x_0) - m_f(x)\big)^2} \ge 0.896^2 + o(1).$$

The last statement in Theorem 4 will be verified in following sections, where a more general theory for the lower bound is developed. Combining the two statements in Theorem 4 yields the minimax risk:

$$1 + o(1) \ge \frac{R(n, \mathscr{C}_2)}{\frac{3}{4} 15^{-1/5} C^{2/5} (B/bn)^{4/5}} \ge 0.896^2 + o(1). \tag{4.3}$$

Theorem 4 proves that the estimator $\hat{m}^*(x_0)$ is nearly an asymptotic minimax estimator. The following theorem shows that it is also an asymptotic linear minimax estimator. To fix the idea, call an estimator $\hat{m}_L(x_0)$ linear if it is a weighted average of Y_j's:

$$\hat{m}_L(x) = \sum_{j=1}^{n} W_j(X_1, \dots, X_n) Y_j. \tag{4.4}$$

Evidently, the local linear regression smoother $\hat{m}^*(x_0)$ is a linear smoother. Let the minimax risk of linear smoothers be

$$R_L(n, \mathscr{C}_2) = \inf_{\hat{m}_L \text{ linear}} \sup_{f \in \mathscr{C}_2} E(\hat{m}_L(x_0) - m(x_0))^2.$$

Then we have the following result.

THEOREM 5. *The linear minimax risk is given by*

$$R_L(n, \mathscr{C}_2) = \frac{3}{4} 15^{-1/5} C^{2/5} \left(\frac{B}{bn} \right)^{4/5} (1 + o(1)), \tag{4.5}$$

and the estimator $\hat{m}^*(x_0)$ is the asymptotic best linear smoother in the sense that

$$R_L(n, \mathscr{C}_2) \Big/ \sup_{f \in \mathscr{C}_2} E(\hat{m}^*(x_0) - m(x_0))^2 \to 1.$$

4.2. *Modulus of continuity.* Connections of modulus continuity with both upper and lower bounds for nonparametric density models and Gaussian white models have been extensively studied in the literature. See Donoho (1990), Donoho and Liu (1991), Donoho and Nussbaum (1990), among others. However, in a nonparametric regression context the connections appear to be new.

Assume more generally that we wish to estimate $m_f(x_0) = E_f(Y|X = x_0)$ with a nonparametric constraint $f \in \mathscr{F}$. For convenience of discussion, assume that $\mathscr{F} = \mathscr{F}_m \cap \mathscr{F}_{b,B}$ [compare (3.4)], where $\mathscr{F}_m$ contains constraints on m and $\mathscr{F}_{b,B}$ imposes constraints on marginal densities and conditional variance:

$$\mathscr{F}_{b,B} = \{ f(\cdot, \cdot) \colon f_X(x_0) \ge b, \sigma^2(x) \le B, |f_X(x) - f_X(y)| \le c|x - y|^{\alpha} \}. \tag{4.6}$$

Note that the Lipschitz condition $|f_X(x) - f_X(y)| \le c|x - y|^{\alpha}$ is used only for technical arguments in the upper bound and hence the constants c and α are not related to the upper and lower bound. Indeed, in the lower bound development below this condition will not be used.

Define the modulus of continuity at a point x_0 over $\mathscr{F}_m$ by

$$\omega_{\mathscr{F}_m}(\varepsilon) = \sup\{ |m_1(x_0) - m_0(x_0)| \colon m_j \in \mathscr{F}_m, \|m_1 - m_0\| \le \varepsilon \}, \tag{4.7}$$

where $\|\cdot\|$ is the usual L_2-norm on $L_2(-\infty, \infty)$. In nonparametric applications, one typically has

$$\omega_{\mathscr{F}_m}(\varepsilon) = A\varepsilon^p (1 + o(1)) \quad \text{as } \varepsilon \to 0, \, p \in (0, 1),$$

and the extremal pair is attained at $m_0(\cdot)$ and $m_1(\cdot)$ satisfying

$$(4.8)\quad m_1(x) - m_0(x) = \varepsilon^p H\left(\frac{x_0 - x}{\varepsilon^{2q}}\right)(1 + o(1)) \quad \text{uniformly in } x \text{ as } \varepsilon \to 0,$$

where $q = 1 - p$ and $H(\cdot)$ is a bounded and continuous function.

DEFINITION. A functional $m_f(x_0)$ is regular on $\mathscr{F}_m$ with exponent p, if the extremal pair of modulus of continuity (4.7) exists and has form (4.8).

As an illustration, consider the constraint $\mathscr{C}_2$. A similar computation can also be found in Donoho and Liu (1991). In this case, $\mathscr{C}_2 = \mathscr{D}_2 \cap \mathscr{F}_{b,B}$ where

$$(4.9)\quad \mathscr{D}_2 = \left\{m(\cdot): |m(x) - m(x_0) - m'(x_0)(x - x_0)| \le C(x - x_0)^2/2\right\}.$$

Let us determine the modulus function for the class $\mathscr{D}_2$:

$$\omega_{\mathscr{D}_2}(\varepsilon) = \sup\{|m_1(x_0) - m_0(x_0)|: \|m_1 - m_0\| \le \varepsilon,\ m_0 \in \mathscr{D}_2,\ m_1 \in \mathscr{D}_2\}.$$

First, by Lemma 7 of Donoho and Liu (1991), the extremal pair can be chosen of the form: $m_0 = m$ and $m_1 = -m$. Thus

$$\omega_{\mathscr{D}_2}(\varepsilon) = 2\sup\{|m(x_0)|: \|m(\cdot)\| \le \varepsilon/2,\ m \in \mathscr{D}_2\}.$$

It follows that $\omega_{\mathscr{D}_2}$ is the inverse function of

$$\varepsilon(\omega) = 2\inf\{\|m(\cdot)\|: |m(x_0)| = \omega/2,\ m \in \mathscr{D}_2\}.$$

A solution to the last problem is obviously the function $m^*(\cdot)$ which is equal to $\omega/2$ at x_0 and descends to 0 as rapidly as possible:

$$m^*(x) = \left[\omega - C(x - x_0)^2\right]_+/2.$$

The L_2-norm of $m^*(\cdot)$ is given by

$$\int_{-\infty}^{\infty} (m^*(x))^2\,dx = \frac{4}{15}C^{-1/2}\omega^{5/2}.$$

This implies

$$\varepsilon(\omega) = 2\sqrt{\frac{4}{15}C^{-1/2}\omega^{5/2}}.$$

Hence

$$(4.10)\quad \omega_{\mathscr{D}_2}(\varepsilon) = (15/16)^{2/5}C^{1/5}\varepsilon^{4/5}.$$

The extremal pair is attained at $m_0 = m^*$ and $m_1 = -m^*$ with

$$\begin{aligned} m^*(x) &= 2^{-1}\left[(15/16)^{2/5}C^{1/5}\varepsilon^{4/5} - C(x - x_0)^2\right]_+ \\ &= 2^{-1}C^{1/5}\varepsilon^{4/5}\left[(15/16)^{2/5} - C^{4/5}\left(\frac{x - x_0}{\varepsilon^{2/5}}\right)^2\right]_+. \end{aligned}$$

Hence condition (4.8) holds with $p = 4/5$ and

$$H(x) = C^{1/5}\left[(15/16)^{2/5} - C^{4/5}x^2\right]_+.$$

The relationship between the modulus of continuity and the upper bound (4.2) can be expressed as

$$R(n,\mathscr{C}_2) \le \frac{q}{4}\omega_{\mathscr{D}_2}^2\left(2\sqrt{\frac{pB}{nbq}}\right)(1+o(1)) \tag{4.11}$$

$$= \frac{p^p q^q}{4}\omega_{\mathscr{D}_2}^2\left(2\sqrt{\frac{B}{nb}}\right)(1+o(1)), \tag{4.12}$$

where $p = 4/5$, the exponent of the modulus of continuity, and $q = 1 - p$.

4.3. *Heuristics of the hardest one-dimensional subproblem.* Let us turn our attention away from the specific constraint $\mathscr{C}_2$ toward a general constraint $\mathscr{F} = \mathscr{F}_m \cap \mathscr{F}_{b,B}$. Assume that $m(x_0)$ (suppress the dependence on f) is regular on $\mathscr{F}_m$ with exponent p. Consider the nonparametric minimax risk

$$R(n,\mathscr{F}) = \inf_{\hat{T}_n \text{ measurable}} \sup_{f\in\mathscr{F}} E_f\left(\hat{T}_n - m(x_0)\right)^2.$$

Assume that $\mathscr{F}_m$ is convex so that

$$m_\theta(x) = (1-\theta)m_0(x) + \theta m_1(x) \in \mathscr{F}_m, \qquad \theta\in[0,1], \tag{4.13}$$

where m_0 and m_1 are an extremal pair of the modulus of continuity $\omega_{\mathscr{F}_m}(2\sqrt{pB/(nbq)})$ [compare (4.11)]. Thus there exists a family of joint densities $\mathscr{F}_0 \equiv \{f_\theta\colon \theta\in[0,1]\}$ such that

$$E_{f_\theta}(Y|X=x) = m_\theta(x).$$

An obvious lower bound on $R(n,\mathscr{F})$ is

$$\begin{aligned} R(n,\mathscr{F}) &\ge R(n,\mathscr{F}_0)\\ &= |m_0(x_0) - m_1(x_0)|^2 \inf_{\hat{T}_n \text{ measurable}} \sup_{0\le\theta\le 1} E\left(\hat{T}_n - \theta\right)^2\\ &= \omega_{\mathscr{F}_m}^2\left(2\sqrt{\frac{pB}{nbq}}\right)\inf_{\hat{T}_n}\sup_{|\theta|\le 1/2} E\left(\hat{T}_n-\theta\right)^2(1+o(1)). \end{aligned} \tag{4.14}$$

The last equality holds since m_0 and m_1 are the extremal pair of the modulus. Thus we have reduced the full nonparametric problem to a one-dimensional subproblem (estimating θ from the parametric family $\mathscr{F}_0$) and made the connection of the lower bound with the modulus of continuity.

Relevant information on the second factor of (4.14) is estimating a bounded normal mean from a normal model. See Bickel (1981), Ibragimov and Khas'minskii (1984), Donoho, Liu and MacGibbon (1990), among others. Consider observing the real-valued random variable $Y \sim N(\theta,\sigma^2)$; the objective is to estimate θ knowing that θ is bounded: $|\theta| \le \tau$. The minimax risk for this problem is denoted by

$$\rho_N(\tau, \sigma) = \inf_{\hat{T} \text{ measurable}} \sup_{|\theta| \le \tau} E\big(\hat{T}_n(Y) - \theta\big)^2, \tag{4.15}$$

which has a simple relation:

$$\rho_N(\tau, \sigma) = \sigma^2 \rho_N(\tau/\sigma, 1).$$

Similarly, minimax affine risk is

$$\rho_A(\tau, \sigma) \equiv \inf_{a,b} \sup_{|\theta| \le \tau} (a + bY - \theta)^2 = \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}.$$

However, there is no closed form for ρ_N, but a simple inequality is available:

$$0.8 \le \eta_\varepsilon \equiv \rho_N(1/2, \varepsilon)/\rho_A(1/2, \varepsilon) \le 1 \tag{4.16}$$

[Donoho, Liu and MacGibbon (1990)].

We would expect that the second factor of (4.14) is (see Section 4.4 for details)

$$\inf_{\hat{T}_n} \sup_{|\theta| \le 1/2} E\big(\hat{T}_n(X_1, Y_1, \dots, X_n, Y_n) - \theta\big)^2 \approx \rho_N\left(\frac{1}{2}, \sqrt{\frac{q}{4p}}\right).$$

If we show that

$$\liminf_{n \to \infty} \inf_{\hat{T}_n} \sup_{|\theta| \le 1/2} E\big(\hat{T}_n(X_1, Y_1, \dots, X_n, Y_n) - \theta\big)^2 \ge \rho_N\left(\frac{1}{2}, \sqrt{\frac{q}{4p}}\right), \tag{4.17}$$

then (4.14) leads to

$$\begin{aligned} R(n, \mathscr{F}) &\ge \rho_N\left(\frac{1}{2}, \sqrt{\frac{q}{4p}}\right) \omega^2_{\mathscr{F}_m}\left(2\sqrt{\frac{pB}{qbn}}\right)(1 + o(1)) \\ &= \xi_p \frac{p^p q^q}{4} \omega^2_{\mathscr{F}_m}\left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1)), \end{aligned}$$

where (4.16) was used in the last expression and $\xi_p = \eta_{\sqrt{q/4p}}$.

Comparing the last display with (4.12), we have given a nearly sharp evaluation of the asymptotic minimax risk for the class of constraint $\mathscr{C}_2$. In that case, $p = 4/5$ and a better evaluation is available: $\xi_{4/5} \ge 1/1.243$ [see Table 1 of Donoho and Liu (1991)]. This proves the second conclusion of Theorem 4 and it remains to verify (4.17).

4.4. *Modulus continuity and minimax lower bound.* To validate (4.17), we consider a normal submodel:

$$f_\theta(x, y) = \frac{1}{\sqrt{2\pi B}} \exp\big(-(y - m_\theta(x))^2/B\big) g(x), \tag{4.18}$$

where $g(x)$ is a marginal density, and m_θ was defined by (4.13). We make an assumption on the richness of $\mathscr{F}$.

Richness of joint densities. There exists a bounded density g with $g(x_0) = b$ *such that the normal submodel* (4.18) *is in the class of constraint* $\mathscr{F} = \mathscr{F}_m \cap \mathscr{F}_{b,B}$.

Based on the normal submodel (4.18), a sufficient statistic for θ would be

$$\hat{\delta}_n \equiv \sum_1^n \big(Y_j - m_0(X_j)\big)\big(m_1(X_j) - m_0(X_j)\big),$$

(4.19)

$$\hat{\sigma}_n^2 \equiv \sum_1^n \big(m_0(X_j) - m_1(X_j)\big)^2.$$

Thus considering statistics based on $\hat{\delta}_n$ and $\hat{\sigma}_n^2$ would be good enough for estimating the unknown parameter θ. Note that conditioning on $X_1, \dots, X_n$,

$$\hat{\delta}_n / \hat{\sigma}_n^2 \sim N\big(\theta, B/\hat{\sigma}_n^2\big).$$

Hence definition (4.15) gives

$$\inf_{\hat{T}_n} \sup_{|\theta| \le 1/2} E\Big(\big(\hat{T}_n - \theta\big)^2 | X_1, \dots, X_n\Big) = \rho_N\left(\frac{1}{2}, \frac{\sqrt{B}}{\hat{\sigma}_n}\right). \tag{4.20}$$

Recall that m_0 and m_1 are the extremal pair of $\omega_{\mathscr{F}_m}(\varepsilon_n)$ with $\varepsilon_n = 2\sqrt{pB/(nbq)}$. Regularity condition (4.8) leads to

$$m_0(x) - m_1(x) = \varepsilon_n^p H\left(\frac{x_0 - x}{\varepsilon_n^{2q}}\right)(1 + o(1)) \quad \text{and}$$

(4.21)

$$\int_{-\infty}^{\infty} (m_0 - m_1)^2 \, dx = \varepsilon_n^2.$$

The last two displays imply $\int_{-\infty}^{\infty} H^2(x)\, dx = 1$. Note that by (4.21),

$$\begin{aligned} E\hat{\sigma}_n^2 &= n\int_{-\infty}^{\infty} \big(m_0(x) - m_1(x)\big)^2 g(x)\, dx \\ &= ng(x_0)\varepsilon_n^2 \int_{-\infty}^{\infty} H^2(x)\, dx (1 + o(1)) \\ &= \frac{4pB}{q} + o(1) \end{aligned}$$

and

$$\begin{aligned} \operatorname{Var}\big(\hat{\sigma}_n^2\big) &\le n\int_{-\infty}^{\infty} \big(m_0(x) - m_1(x)\big)^4 g(x)\, dx \\ &\le \varepsilon_n^{2p} \sup_x H^2(x) \left[n\int_{-\infty}^{\infty} \big(m_0(x) - m_1(x)\big)^2 g(x)\, dx\right] \\ &\to 0 \quad \text{as } n \to \infty. \end{aligned}$$

These two facts demonstrate that

$$\hat{\sigma}_n \to_P \sqrt{\frac{4pB}{q}}. \tag{4.22}$$

Heuristically (rigorous proof is given in Section 6), (4.20) and (4.22) entail

$$\inf_{\hat{T}_n} \sup_{|\theta| \le 1/2} E(\hat{T}_n - \theta)^2 \approx E\rho_N\left(\frac{1}{2}, \frac{\sqrt{B}}{\hat{\sigma}_n}\right) \approx \rho_N\left(\frac{1}{2}, \sqrt{\frac{q}{4p}}\right).$$

This together with (4.20) validate (4.17).

THEOREM 6. *Let* $\mathscr{F}_m$ *be convex and* $\mathscr{F} = \mathscr{F}_m \cap \mathscr{F}_{b,B}$ *be rich. If* $m_f(x_0)$ *is regular on* $\mathscr{F}_m$ *with exponent* p, *then a minimax lower bound is given by*

$$R(n, \mathscr{F}) \ge \xi_p \frac{p^p q^q}{4} \omega_{\mathscr{F}_m}^2 \left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1)), \tag{4.23}$$

where $\mathscr{F}_{b,B}$ *and* $\xi_p = \eta_{\sqrt{q/4p}}$ *were defined by* (4.6) *and* (4.16), *respectively. The left-hand side of* (4.23), *omitting the factor* ξ_p, *is also a linear minimax lower bound.*

REMARK 2. The result of Theorem 4 holds also for a random-design regression problem whose marginal density is known to be $g(\cdot)$ with $g(x_0) = b$. The reason is that in the lower bound development, the marginal density was fixed all the time. The lower bound is also applicable for fixed designs with design points $x_i = G(i/n)$ and $G' = g$, since previous arguments were conditioned on covariates $X_1, \dots, X_n$.

REMARK 3. Suppose that we wish to estimate a conditional quantile $Q_r(x_0)$ defined by [see Truong (1989)]

$$P\{Y \le Q_r(x_0) | X = x_0\} = r,$$

based on a random sample of size n. Then for normal submodel (4.18),

$$Q_r(x_0) = m_\theta(x_0) + z_r\sqrt{B},$$

where $z_r = \Phi^{-1}(r)$ and $\Phi(\cdot)$ is the standard normal cdf. Thus estimating $Q_r(x_0)$ in the normal submodel is as difficult as estimating $m_\theta(x_0)$. This yields a lower bound:

$$\inf_{\hat{T}_n} \sup_{f \in \mathscr{F}} E_f(\hat{T}_n - Q_r(x_0))^2 \ge \xi_p \frac{p^p q^q}{4} \omega_{\mathscr{F}_m}^2 \left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1)).$$

However, it remains unknown how sharp this lower bound is for estimating conditional quantiles.

5. Discussion. The minimax lower bound is derived via the heuristic of hardest one-dimensional subproblem. We have shown that such a bound is indeed *nearly sharp* for a two-bounded-derivative constraint. Analysis of minimax upper bounds for other constraints goes beyond the intent of this paper, but provides interesting topics for future research.

5.1. *Nearly sharp lower bound.* We have shown that a minimax lower bound is [$\xi_p \geq 0.8 = 0.894^2$, by (4.16)]

$$R(n, \mathscr{F}) \geq 0.894^2 \frac{p^p q^q}{4} \omega^2_{\mathscr{F}_m}\left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1)). \tag{5.1}$$

If one can find an estimator such that its maximum risk is no larger than

$$\frac{p^p q^q}{4} \omega^2_{\mathscr{F}_m}\left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1)), \tag{5.2}$$

then such an estimator has at least a minimax efficiency 89.4% in a sense similar to Theorem 4, and consequently the lower bound is nearly sharp. With such a sharp minimax lower bound, we can compute the efficiency as follows:

$$\text{Efficiency of an estimator} \geq \left(\frac{\text{Minimax lower bound (5.1)}}{\text{Maximum MSE of the estimator}}\right)^{1/2}. \tag{5.3}$$

Two-bounded-derivative constraints $\mathscr{C}_2$ are not the only examples that the upper bound (5.2) holds. We conjecture that a general theory can be made if one makes connections with white-noise models as Donoho and Liu (1991) did in density estimation. Let us give another example in which the minimax upper bound (5.2) holds.

EXAMPLE (Bounded Lipschitz constraints). Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. from a joint density $f \in \mathscr{C}_1 = \mathscr{D}_1 \cap \mathscr{F}_{b,B}$ with

$$\mathscr{D}_1 = \{m(\cdot): |m(x) - m(y)| \leq C|x - y|, \forall\, x, y \in \mathbb{R}\}.$$

A similar machinery from (4.9) to (4.10) yields the modulus of continuity:

$$\omega_{\mathscr{D}_1}(\varepsilon) = 3^{1/3} C^{1/3} \varepsilon^{2/3},$$

and $m(x_0)$ is regular on $\mathscr{D}_1$. This together with Theorem 6 leads to

$$R(n, \mathscr{C}_1) \geq \xi_{2/3} 3^{-1/3} \left(\frac{BC}{bn}\right)^{2/3} (1 + o(1)),$$

where $\xi_{2/3} = 1/1.178 = 0.92^2$ by Donoho and Liu (1991). On the other hand, exhibiting the maximum risk of the estimator

$$\sum_1^n \left(1 - \left|\frac{x_0 - X_j}{h_n^{(2)}}\right|\right)_+ Y_j \Big/ \sum_1^n \left[\left(1 - \left|\frac{x_0 - X_j}{h_n^{(2)}}\right|\right)_+ + \frac{1}{n}\right], \tag{5.4}$$

with $h_n^{(2)} = (3B/bC^2 n)^{1/3}$ [corresponding to estimator (2.6) with $K(x) =$

$(1 - |x|)_+]$ yields an upper bound:

$$R(n, \mathscr{C}_1) \le 3^{-1/3}\left(\frac{BC}{bn}\right)^{2/3}(1 + o(1)), \tag{5.5}$$

that is, (5.2) holds. In summary, we have the following result.

THEOREM 7. *Under the constraint* $\mathscr{C}_1$, *the minimax risk is bounded by*

$$0.92^2 + o(1) \le \frac{R(n, \mathscr{C}_1)}{3^{-1/3}(BC/bn)^{2/3}} \le 1 + o(1).$$

Moreover, estimator (5.4) *has asymptotic minimax efficiency at least* 92%.

5.2. *Relation to other work.* Vast literature has been devoted in analyzing the behavior of the Nadaraya–Watson and the Gasser–Müller regression estimators. Drawbacks of these estimators are eliminated via introducing a new class of estimators.

Previous work on minimax regression problems has mainly focused on determining optimal rates of convergence [Stone (1980)]. Local polynomial regression estimators were used in Stone (1980) to determine the rates of convergence. To analyze constant factors, we extend the idea of local polynomial regression estimators.

A closely related idea for minimax bounds is the work of Donoho (1990) and Donoho and Liu (1991), where white-noise and density estimation models are emphasized. What seems innovative in our approach is the decomposition of nonparametric constraints into two parts: $\mathscr{F}_m$ and $\mathscr{F}_{b,B}$, and the use of normal submodels to avoid technicalities of convergence of experiments.

Other efforts in finding minimax risks in the regression setup include Sacks and Ylvisaker (1978) and Li (1982) who find minimax linear estimates for fixed designs under some specific constraints, and Nussbaum (1985) and Low (1993) where the attention is mostly focused on some specific global problems. In particular, Sacks and Ylvisaker (1978) offer a method of solving the linear minimaxity issues. This paper attempts to give a general theory for understanding minimax nonparametric regression and provides an insight to this problem. In the density estimation setup, contributions include Sacks and Ylvisaker (1981), Efroimovich and Pinsker (1982), Sacks and Strawderman (1982), Birgé (1987) and Donoho and Liu (1991). There is also a long history in finding minimax risks for Gaussian white-noise models and other related problems. See Pinsker (1980), Ibragimov and Khas'minskii (1984), Brown and Liu (1989), Donoho and Johnstone (1989), Donoho, Liu and MacGibbon (1990), Donoho and Nussbaum (1990), among others.

6. Proof. Theorems 1–3 can be proved along the same lines. The proof of Theorem 3 is more involved and requires more details. For this reason we decide only to prove Theorem 3.

PROOF OF THEOREM 3. First of all, estimator (2.2) has the MSE

$$E(\hat{m}(x_0) - m(x_0))^2 = E\left(\frac{\sum_1^n (Y_j - m(x_0))w_j}{\sum_1^n w_j + n^{-2}}\right)^2$$
$$+ n^{-4}m^2(x_0)E\left(\sum_1^n w_j + n^{-2}\right)^{-2} \tag{6.1}$$
$$- 2n^{-2}m(x_0)E\frac{\sum_1^n (Y_j - m(x_0))w_j}{\left(\sum_1^n w_j + n^{-2}\right)^2}.$$

Denote $Z_n = O_r(a_n)$, if $\sup_{f \in \mathscr{C}_2} E|Z_n|^r = O(a_n^r)$. A similar meaning extends to $o_r(a_n)$. Obvious operations include

$$O_r(a_n)O_r(b_n) = O_{r/2}(a_n b_n) \qquad \text{(Cauchy–Schwarz inequality)}$$

and

$$Z_n = EZ_n + O_r\left(\left(E|Z_n - EZ_n|^r\right)^{1/r}\right). \tag{6.2}$$

We also use o and O to denote the order of magnitude uniformly in $f \in \mathscr{C}_2$. For example, expression (6.3) means that

$$\sup_{f \in \mathscr{C}_2} |Es_{n,l} - nh_n^{l+1}f(x_0)s_l| = O(h_n^\alpha).$$

Then it is easy to show, by using the method of the kernel density estimate, that with $s_{n,l}$ defined by (2.4),

$$Es_{n,l} = nh_n^{l+1}f_X(x_0)s_l(1 + O(h_n^\alpha)), \qquad l = 0,1,2, \tag{6.3}$$

and that [see (6.2)] for an integer $r > 0$,

$$\frac{1}{nh_n^{l+1}}s_{n,l} = \frac{1}{nh_n^{l+1}}Es_{n,l} + O_r\left(\frac{1}{\sqrt{nh_n}}\right)$$
$$= f_X(x_0)s_l + O_r\left(h_n^\alpha + \frac{1}{\sqrt{nh_n}}\right), \qquad l = 0,1,2, \tag{6.4}$$

where $s_l = \int_{-\infty}^{\infty} u^l K(u)\,du$ and, in particular, $s_0 = 1$ and $s_1 = 0$. A direct consequence of (6.4) is that

$$\sum_1^n w_n = s_{n,0}s_{n,2} - (s_{n,1})^2 = n^2h_n^4 s_2 f_X^2(x_0)\left(1 + O_r\left(h_n^\alpha - \frac{1}{\sqrt{nh_n}}\right)\right). \tag{6.5}$$

Next, let $W_n = (\sum_1^n w_j + n^{-2})/(n^2h_n^4)$ and $W = s_2 f_X^2(x_0)$. We are going to show that

$$\frac{1}{W_n} = \frac{1}{W} + o_4(1). \tag{6.6}$$

To see this, we first note that

$$E\left(\frac{W}{W_n}-1\right)^4 = E\frac{(W_n-W)^4}{W_n^4}1_{\{|W_n-W|\le W/2\}} + E\frac{(W_n-W)^4}{W_n^4}1_{\{|W_n-W|>W/2\}}$$
$$\le (W/2)^{-4}E(W_n-W)^4 + n^{16}E(W_n-W)^4 1_{\{|W_n-W|>W/2\}}$$
$$\equiv A_n + B_n,$$

where the fact that $W_n \ge n^{-4}$ was used in the second term. Next (6.5) assures that $A_n = o(1)$ and that

$$B_n \le n^{16}(W/2)^{-r}E(W_n-W)^{r+4} = O\left(n^{16}\left(h_n^\alpha + \frac{1}{\sqrt{nh_n}}\right)^{r+4}\right) = o(1),$$

by choosing a sufficiently large r. Thus (6.6) holds. A direct consequence of (6.6) is that the second term of (6.1) has order

$$n^{-4}m^2(x_0)E\left(\sum_1^n w_j + n^{-2}\right)^{-2} = m^2(x_0)(nh_n)^{-8}E\frac{1}{W_n^2} = O\big((nh_n)^{-8}\big).$$

This and (6.1) lead to

$$E(\hat{m}(x_0)-m(x_0))^2 = E\left(\frac{\Sigma_1^n[Y_j-m(x_0)]w_j}{\Sigma_1^n w_j + n^{-2}}\right)^2 + o\big((nh_n)^{-4}\big),$$

and the conclusion follows if we show that the main term

$$E\left(\frac{\Sigma_1^n[Y_j-m(x_0)]w_j}{\Sigma_1^n w_j + n^{-2}}\right)^2 \tag{6.7}$$

has bias and variance decomposition (3.3).

Conditioning on covariates X_j, $j=1,\dots,n$ and then using mean and variance decomposition, we have

$$E\left(\frac{\Sigma_1^n[Y_j-m(x_0)]w_j}{\Sigma_1^n w_j + n^{-2}}\right)^2 = E\left(\frac{\Sigma_1^n[m(X_j)-m(x_0)]w_j}{\Sigma_1^n w_j + n^{-2}}\right)^2 + E\frac{\Sigma_1^n\sigma^2(X_j)w_j^2}{(\Sigma_1^n w_j + n^{-2})^2}. \tag{6.8}$$

Let $R(X_j) = m(X_j) - m(x_0) - m'(x_0)(X_j - x_0)$. Then (2.5) leads to

$$\sum_1^n[m(X_j)-m(x_0)]w_j = \sum_1^n R(X_j)w_j$$
$$= \sum_1^n R(X_j)K\left(\frac{x_0-X_j}{h_n}\right)s_{n,2} - \sum_1^n R(X_j)(x_0-X_j)K\left(\frac{x_0-X_j}{h_n}\right)s_{n,1}. \tag{6.9}$$

By a standard argument [see (6.2)], for $l = 0, 1$, we have

$$\frac{1}{nh_n^{3+l}} \sum_1^n R(X_j)(X - x_0)^l K\left(\frac{x_0 - X_j}{h_n}\right) = h_n^{-3-l} E(m(X) - m(x_0) - m'(x_0)(X - x_0)) \times (X - x_0)^l K\left(\frac{x_0 - X}{h_n}\right) + o_8(1)$$

and

$$\sup_{f \in \mathscr{C}_2} \left| h_n^{-3-l} E(m(X) - m(x_0) - m'(x_0)(X - x_0))(X - x_0)^l K\left(\frac{x_0 - X}{h_n}\right) \right|$$

$$\begin{aligned} &\le \frac{C}{2} \sup_{f \in \mathscr{C}_2} E \left| \frac{X - x_0}{h_n} \right|^{l+2} \frac{1}{h_n} K\left(\frac{x_0 - X}{h_n}\right) \\ &= \frac{C}{2} \sup_{f \in \mathscr{C}_2} \int_{-\infty}^{\infty} |y|^{l+2} K(y) f_X(x_0 - h_n y)\, dy \\ &= O(1). \end{aligned} \tag{6.10}$$

Substituting the last two displays into (6.9) and using (6.4), we have

$$\sum_1^n (m(X_j) - m(x_0)) w_j = n^2 h_n^6 f_X(x_0) s_2 S_n + o_4(n^2 h_n^6),$$

where

$$S_n = h_n^{-3} E(m(X) - m(x_0) - m'(x_0)(X - x_0)) K\left(\frac{x_0 - X}{h_n}\right).$$

It is concluded from (6.6) that

$$E\left(\frac{\sum_1^n (m(X_j) - m(x_0)) w_j}{\sum_1^n w_j + n^{-2}}\right)^2 = \left(\frac{S_n}{f_X(x_0)}\right)^2 h_n^4 + o(h_n^4). \tag{6.11}$$

By (6.10), we have

$$\begin{aligned} &\sup_{f \in \mathscr{C}_2} |S_n / f_X(x_0)| \\ &\quad \le \frac{C}{2} s_2 + \sup_{f \in \mathscr{C}_2} \int_{-\infty}^{\infty} y^2 K(y) |f_X(x_0 - h_n y) - f_X(x_0)| / f_X(x_0)\, dy \\ &\quad = \frac{C}{2} s_2 + o(1). \end{aligned}$$

Hence (6.11) entails that

$$\sup_{f \in \mathscr{C}_2} E\left(\frac{\sum_1^n (m(X_j) - m(x_0))w_j}{\sum_1^n w_j + n^{-2}}\right)^2 \le \left(\frac{Cs_2}{2}\right)^2 h_n^4 + o(h_n^4).$$

To complete the proof, we need only to compute the second term of (6.8). Note that

$$\begin{aligned} \sum_1^n \sigma^2(X_j) w_j^2 &= \sum_1^n \sigma^2(X_j) K^2\left(\frac{x_0 - X_j}{h_n}\right) \\ &\quad \times \left(s_{n,2}^2 - 2(x_0 - X_j)s_{n,2}s_{n,1} + (x_0 - X_j)^2 s_{n,1}^2\right) \\ &\le B\sum_1^n K^2\left(\frac{x_0 - X_j}{h_n}\right)\left|s_{n,2}^2 - 2(x_0 - X_j)s_{n,2}s_{n,1} + (x_0 - X_j)^2 s_{n,1}^2\right|. \end{aligned} \tag{6.12}$$

A standard argument [see (6.2)] yields

$$\begin{aligned} &\frac{1}{nh_n^{l+1}} \sum_1^n K^2\left(\frac{x_0 - X_j}{h_n}\right)(x_0 - X_j)^l \\ &\quad = f_X(x_0)\int_{-\infty}^{\infty} u^l K^2(u)\, du + o_4(1), \qquad l = 0, 1, 2. \end{aligned} \tag{6.13}$$

Since $s_1 = 0$ the dominant term of (6.12) is its first term:

$$\sum_1^n K^2\left(\frac{x_0 - X_j}{h_n}\right) s_{n,2}^2.$$

Consequently, combination of (6.4), (6.12) and (6.13) gives

$$\sum_1^n \sigma^2(X_j) w_j^2 \le n^3 h_n^7 B f_X^3(x_0) s_2^2 \int_{-\infty}^{\infty} K^2(u)\, du (1 + o_2(1)),$$

and we conclude from (6.6) that

$$\begin{aligned} \sup_{f \in \mathscr{C}_2} E \frac{\sum_1^n \sigma^2(X_j) w_j^2}{\left(\sum_1^n w_j + n^{-2}\right)^2} &\le \sup_{f \in \mathscr{C}_2} \left(\frac{n^3 h_n^7 B f_X^3(x_0) s_2^2 \int K^2(u)\, du}{n^4 h_n^8 s_2^2 f_X^4(x_0)}\right)(1 + o(1)) \\ &\le \frac{B}{nh_n b}\int_{-\infty}^{\infty} K^2(u)\, du (1 + o(1)). \end{aligned}$$

This completes the proof. □

PROOF OF THEOREM 5. Since (4.2) also supplies an upper bound for $R_L(n, \mathscr{C}_2)$, it suffices to show that

$$R_L(n, \mathscr{C}_2) \ge \frac{3}{4} 15^{-1/5} C^{2/5} \left(\frac{B}{bn}\right)^{4/5} (1 + o(1)). \tag{6.14}$$

For a linear smoother (4.4), by Lemma 1 of Fan (1992),

$$E\Big[(\hat{m}_L(x_0) - m(x_0))^2 | X_1, \dots, X_n\Big]$$
$$= \left[\sum_{j=1}^{n} W_j m(X_j) - m(x_0)\right]^2 + \sum_{j=1}^{n} W_j^2 \sigma^2(X_j)$$
$$\geq \frac{m^2(x_0)}{1 + \sum_{j=1}^{n} m^2(X_j)/\sigma^2(X_j)}.$$

Thus, by Jessen's inequality,

$$E(\hat{m}_L(x_0) - m(x_0))^2 \geq \frac{m^2(x_0)}{1 + nEm^2(X_1)/\sigma^2(X_1)}. \tag{6.15}$$

Specifically, take a submodel $f_0(\cdot, \cdot) \in \mathscr{C}_2$ such that $m_0(y) = (b_n^2/2)[1 - C(y - x_0)^2/b_n^2]_+$, $\sigma_0^2(\cdot) = B$ and $f_X(x_0) = b$, where $b_n = (15\sqrt{C}B/bn)^{1/5}$ maximizes (6.16). Then it is easy to verify that

$$E_{f_0} m_0^2(X_1)/\sigma_0^2(X_1) = \frac{b_n^4}{4B}\int_{-\infty}^{\infty}\left[1 - C(y - x_0)^2/b_n^2\right]_+^2 f_X(y)\,dy$$
$$= \frac{4bb_n^5}{15\sqrt{C}B}(1 + o(1)).$$

Substituting this into (6.15), we have

$$\begin{aligned} R_L(n, \mathscr{C}_2) &\geq \frac{b_n^4/4}{1 + \left(4bnb_n^5/15\sqrt{C}B\right)(1 + o(1))} \\ &= \frac{3}{4}15^{-1/5}\left(\frac{\sqrt{C}B}{bn}\right)^{4/5}(1 + o(1)). \end{aligned} \tag{6.16}$$

This verifies (6.14) and completes the proof. □

Proof of Theorem 6. We need only to prove (4.17). First of all, let $\pi(\theta)$ be a least favorable prior for problem (4.15) with $\tau = 0.5$ and $\sigma = \sqrt{q/4p}$, namely,

$$\rho_N\left(0.5, \sqrt{q/4p}\right) = \inf_T E_\theta E\left(\hat{T}(Y) - \theta\right)^2, \qquad Y \sim N\left(\theta, \sqrt{q/4p}\right). \tag{6.17}$$

Denote the Bayes risk with the prior π for normal model $X \sim N(\theta, \sigma^2)$ by

$$B_\pi(\sigma) = \inf_{\hat{T}_n} E_\theta E_X\left(\hat{T}_n(X) - \theta\right)^2.$$

Then (6.17) can be expressed as

$$B_\pi\left(\sqrt{q/4p}\right) = \rho_N\left(0.5, \sqrt{q/4p}\right). \tag{6.18}$$

Next, let us turn our attention back to problem (4.17) with n i.i.d. observations $\{(X_i, Y_i)\}$ from (4.18). By sufficiency,

$$\begin{aligned}
&\inf_{\hat{T}_n} \sup_{|\theta| \le 0.5} E\big(\hat{T}_n(X_1, Y_1, \dots, X_n, Y_n) - \theta\big)^2 \\
&\quad = \inf_{\hat{T}_n^*} \sup_{|\theta| \le 0.5} E\Big(\hat{T}_n^*\big(\hat{\delta}_n, \hat{\sigma}_n\big) - \theta\Big)^2 \\
(6.19) \qquad &\quad \ge \inf_{\hat{T}_n^*} E_\theta E\Big(\hat{T}_n^*\big(\hat{\delta}_n, \hat{\sigma}_n\big) - \theta\Big)^2 \\
&\quad \ge E_{\hat{\sigma}_n} \inf_{\hat{T}_n} E_\theta E_{\hat{\delta}_n}\Big[\Big(\hat{T}_n\big(\hat{\delta}_n, \hat{\sigma}_n\big) - \theta\Big)^2 \Big| \hat{\sigma}_n\Big],
\end{aligned}$$

where $\hat{\sigma}_n$ and $\hat{\delta}_n$ were defined by (4.19). Given $\hat{\sigma}_n$, $\hat{\delta}_n/\hat{\sigma}_n \sim N(\theta, B/\hat{\sigma}_n^2)$. Thus, by (6.19),

$$\inf_{\hat{T}_n} \sup_{|\theta| \le 0.5} E\big(\hat{T}_n - \theta\big)^2 \ge E_{\hat{\sigma}_n} B_\pi\big(\sqrt{B}/\hat{\sigma}_n\big).$$

Note that $B_\pi(\cdot)$ is bounded by $1/4$ (as $|\theta| \le 0.5$) and continuous. The dominated convergence theorem, (4.22) and (6.18) yield

$$E_{\hat{\sigma}_n} B_\pi\big(\sqrt{B}/\hat{\sigma}_n\big) \to B_\pi\big(\sqrt{q/4p}\big) = \rho_N\big(0.5, \sqrt{q/4p}\big).$$

This together with (6.19) entail

$$\liminf_{n \to \infty} \inf_{\hat{T}_n} \sup_{|\theta| \le 0.5} E\big(\hat{T}_n - \theta\big)^2 \ge \rho_N\big(0.5, \sqrt{q/4p}\big),$$

as was to be shown. □

Acknowledgments. The author gratefully acknowledges Professors D. L. Donoho, J. S. Marron and Y. K. Truong for many useful discussions. Thanks also go to two referees and an Associate Editor for their helpful comments and constructive suggestions.

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.

BIRGÉ, L. (1987). Estimating a density under the order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012.

BROWN, L. D. and LIU, R. C. (1989). A sharpened inequality for the hardest affine subproblem. Unpublished manuscript.

CHU, C. K. and MARRON, J. S. (1991). Choosing a kernel regression estimator (with discussion). *Statist. Sci.* **6** 404–436.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

DONOHO, D. L. (1990). Statistical estimation and optimal recovery. Technical Report 214, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and JOHNSTONE, I. (1989). Minimax risk over l_p-balls. Technical Report 204, Dept. Statistics, Univ. California, Berkeley.

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

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

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

EFROIMOVICH, S. Y. and PINSKER, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission* **18** 175–189.

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

GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. *Smoothing Techniques for Curve Estimation. Lectures Notes in Math.* **757** 23–68. Springer, New York.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1984). On nonparametric estimation of values of a linear functional in a Gaussian white noise (in Russian). *Teor. Veroyatnost. i Primenen.* **29** 19–32.

JENNEN-STEINMETZ, C. and GASSER, T. (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc.* **83** 1084–1089.

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

LEJEUNE, M. (1985). Estimation non-paramètrique par noyaux: Règression polynomiale mobile. *Rev. Statist. Appl.* **33** 43–68.

LI, K. C. (1982). Minimaxity of the method of regularization on stochastic processes. *Ann. Statist.* **10** 937–942.

LOW, M. (1993). Renormalizing upper and lower bounds for integrated risk in the white noise model. *Ann. Statist.* **21**. To appear.

MÜLLER, H. G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*.**82** 231–238.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 52–68.

SACKS, J. and STRAWDERMAN, W. (1982). Improvements on linear minimax estimates. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta, and J. O. Berger, eds.) **2** 287–304. Academic, New York.

SACKS, J. and YLVISAKER, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6** 1122–1137.

SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.

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

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

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

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

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