The Annals of Statistics
1997, Vol. 25, No. 4, 1691–1708

ON AUTOMATIC BOUNDARY CORRECTIONS[1]

BY MING-YEN CHENG, JIANQING FAN AND J. S. MARRON

*National Chung Cheng University,
University of North Carolina and University of North Carolina*

Many popular curve estimators based on smoothing have difficulties caused by boundary effects. These effects are visually disturbing in practice and can play a dominant role in theoretical analysis. Local polynomial regression smoothers are known to correct boundary effects automatically. Some analogs are implemented for density estimation and the resulting estimators also achieve automatic boundary corrections. In both settings of density and regression estimation, we investigate best weight functions for local polynomial fitting at the endpoints and find a simple solution. The solution is universal for general degree of local polynomial fitting and general order of estimated derivative. Furthermore, such local polynomial estimators are best among all linear estimators in a weak minimax sense, and they are highly efficient even in the usual linear minimax sense.

1. Introduction. Nonparametric curve estimation methods make no assumptions on the functional form of the curves of interest and hence allow flexible modeling of the data. If the support of the true curve is bounded then most nonparametric methods give estimates that are severely biased in regions near the endpoints. This boundary problem affects the global performance visually and also in terms of a slower rate of convergence in the usual asymptotic analysis. It has been recognized as a serious problem and many works are devoted to reducing the effects. Gasser and Müller (1979), Gasser, Müller and Mammitzsch (1985), Granovsky and Müller (1991) and Müller (1991) discuss boundary kernel methods. Rice (1984) suggests a linear combination of two kernel estimators with different bandwidths to reduce the bias. Schuster's (1985) mirror image density estimator folds back the probability mass that extends beyond the support. The estimator introduced in Hall and Wehrly (1991) is essentially a more sophisticated regression version of Schuster's approach. Djojosugito and Speckman (1992) approach boundary bias reduction based on a finite-dimensional projection in a Hilbert space. Boundary effects for smoothing splines are discussed in Rice and Rosenblatt (1981). Eubank and Speckman (1991) also provide some boundary correction methods.

The above methods provide effective boundary correction, but the more effective ones tend to be quite complicated. This discourages their widespread

Received August 1995; revised October 1996.

[1]This research is part of Ming-Yen Cheng's dissertation under the supervision of Professors J. Fan and J. S. Marron at the University of North Carolina. Research is partially supported by NSF Grants DMS-92-03135 and DMS-95-04414 and NSA Grant 96-1-0015.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62C20.

Key words and phrases. Boundary correction, data binning, local polynomial fit, minimax risk, weak minimaxity.

use (since implementation is a nontrivial exercise) and also entails difficult analysis. A simpler and more direct approach to boundary correction [see Fan and Gijbels (1992) and Hastie and Loader (1993) for insightful discussion] is based on local polynomial fitting. The simplicity comes because boundary adaption is “automatic” in the sense that no explicit correction is needed.

While local polynomial fitting is becoming widely accepted because of its simplicity, an important question is how well they compare to other boundary adjustment methods in terms of efficiency. One approach to this issue would be careful analysis of all of the earlier proposals, with a detailed comparison of their properties. But this would be a tedious task because:

1. there are so many proposals;
2. a number of them are quite complicated and thus difficult to work with, both numerically and analytically.

This paper presents results which address the issue of how the boundary adjustments that are implicit to the local polynomial smoother compare with other boundary adjustments in a simple and clean way. We show in a minimax sense that no linear estimator (which includes not only the proposals cited above, but others that could be devised later) can have better asymptotic mean squared error performance than that of the local polynomial smoother. This avoids the need for detailed analysis of complicated methods, since it makes it clear that they cannot be substantially better in terms of efficiency. The local polynomial smoother is already the choice of many because of its simplicity, and we show that there is no loss in terms of efficiency in this choice.

Given the good performance of local polynomial methods in regression, it is natural to look for analogs in density estimation. In the density estimation context, the data falls on the real line and there are neither design points nor responses. Data binning produces a regression type context where local polynomial fitting can be applied to the bin counts. The resulting estimators of the density and its derivatives do attain automatic boundary corrections. Lejeune and Sarda (1992) and Jones (1993) discuss another approach to local polynomial density estimators: a polynomial is fitted to the empirical density function by minimizing a locally weighted L_2-distance. Our estimators are approximate binned versions of those [see Cheng (1997) for details].

We also study best weight functions for the local polynomial fitting at the boundary. We focus on solving the problem when estimating at the endpoints since that is the most important case. The answer is surprisingly simple in the sense that a particular weight function is the best for all cases of local polynomial fitting, independent of the degree and the order of the derivative being estimated. This is the key to showing that the optimal local polynomial estimators are best among all linear estimators in a weak minimax sense. Also, they are highly efficient even in a more conventional linear minimax sense. We show that the above-mentioned properties hold in both density estimation and regression settings.

Nonparametric minimax problems are interesting and challenging. Recent advancements in this area can be found in, for example, Nussbaum (1985), Donoho and Liu (1991), Fan (1993), Donoho and Johnstone (1994), Fan and Hall (1994), Brown and Low (1996), Efroimovich (1996), Brown, Low and Zhao (1997) and references therein. Most articles focus either on the minimax risk of estimating a whole function or on that of estimating a function at interior points. However, minimax problems at a boundary point have not been studied as often, and the methods used here are different from that at an interior point. In particular, we handle the "effective optimal kernel" through a representation in terms of Legendre polynomials.

This article is organized as follows. In Section 2, asymptotic mean squared errors of local polynomial estimators at boundary points are summarized for both settings of regression and density estimation. Optimal weighting for local polynomial fitting and weak minimax efficiency are discussed in Section 3. Section 4 investigates a general minimax problem and minimax optimal kernels. Section 5 contains some concluding remarks. Proofs are in Section 6.

2. Mean squared errors at boundary points. In this section we briefly discuss asymptotic mean squared errors of local polynomial estimators at boundary points in regression and density estimation contexts.

2.1. *Regression setting.* Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are an i.i.d. sample from a bivariate population (X, Y). The regression function is

$$f(x) = E(Y|X = x).$$

The local polynomial estimator of $f^{(\nu)}(x)$, the νth derivative of f at x, obtained from fitting a pth-degree local polynomial based on the nonnegative weight function K and bandwidth $h > 0$, is

$$\widehat{f^{(\nu)}}(x) = \sum_{i=1}^{n} W^{K}_{p+1,\nu}\left(\frac{X_i - x}{h}\right) Y_i, \tag{1}$$

where $W^{K}_{p+1,\nu}(t) = \nu! e_\nu^{\mathrm{T}} S_n^{-1}(1, ht, \ldots, h^p t^p)^T K(t)$. Here, e_ν is a unit vector whose $(\nu + 1)$th element is 1 and

$$S_n = (S_{n,\, i+j-2})_{1\le i,\, j\le p+1}$$

with

$$S_{n,l} = \sum_{j=1}^{n} (X_j - x)^l K\left(\frac{X_j - x}{h}\right), \qquad l = 0, 1, \ldots, 2p.$$

Details of the above calculation can be found in Ruppert and Wand (1994) and Fan et al. (1997).

Let $g_X(\cdot)$ be the marginal density of X and let $\sigma^2(\cdot)$ be the conditional variance of Y given X; that is, $\sigma^2(x) = \mathrm{Var}(Y|X = x)$. In addition, we assume that the support of g_X is $[0, 1]$. Consider the boundary point $x = 0$. Denote by $S_j = \int_0^{+\infty} t^j K(t)\, dt$, $j = 0, 1, \ldots, 2p$, and $S = (S_{i+j-2})_{0\le i,\, j\le p+1}$. Put

$$K^*_{p+1,\nu}(t) = e_\nu^T S^{-1}(1, t, \ldots, t^p)^T K(t) I_{[0,\infty)}(t), \tag{2}$$

which is the equivalent kernel for estimating the νth derivative at the point $x = 0$. The function $K^*_{p+1,\nu}$ satisfies the following moment conditions:

$$\int_0^\infty t^q K^*_{p+1,\nu}(t)\,dt = e_\nu^T S^{-1} S e_q = \delta_{\nu,q} \quad \text{for } 0 \le \nu, q \le p. \tag{3}$$

CONDITION 1. (i) Functions $g_X(\cdot)$, $f^{(p+1)}(\cdot)$ and $\sigma^2(\cdot)$ are bounded on $[0, 1]$ and right continuous at 0.

(ii) The weight function $K(\cdot)$ has a bounded support.

Under Condition 1, if $\sigma^2(0) < \infty$ and $g_X(0) > 0$, then the conditional mean squared error of $\widehat{f^{(\nu)}}(0)$ is

$$\begin{aligned} &E\Big[\Big(\widehat{f^{(\nu)}}(0) - f^{(\nu)}(0)\Big)^2 \Big| X_1, \dots, X_n\Big] \\ &\quad \approx_p \left(\frac{\nu! f^{(p+1)}(0)}{(p+1)!} \int t^{p+1} K^*_{p+1,\nu}(t)\,dt\right)^2 h^{2(p+1-\nu)} \\ &\qquad + \frac{\nu!^2 \sigma^2(0)}{n h^{2\nu+1} g_X(0)} \int K^*_{p+1,\nu}(t)^2\,dt, \end{aligned} \tag{4}$$

as $n \to \infty$, $h \to 0$ and $nh^{2\nu+1} \to \infty$. See Fan et al. (1997). Here "$\approx_p$" means the random variables are asymptotically the same in probability.

2.2. *Density estimation setting.* There are several methods of adapting the ideas of local polynomial fitting to the setting of density estimation. The following discusses one approach briefly and the details are referred to Cheng (1997). Suppose that $X_1, \dots, X_n$ are an i.i.d. sample from a population following a density function f supported on $[0, 1]$. For each sample size n, choose a binwidth $b > 0$ (depending on n) and let $t_j = (j - \frac{1}{2})b$, for $j = 1, \dots, G$ with $G = [1/b]$. Define the bin count at t_j as

$$c_j = \sum_{i=1}^n I_{[t_j - b/2,\, t_j + b/2)}(X_i).$$

The local polynomial estimator of $f^{(\nu)}(x)$, denoted as $\widehat{f^{(\nu)}}(x)$, based on the weight function K and bandwidth $h > 0$ is defined as the $(\nu+1)$th coefficient of the local polynomial fit to the data $\{(t_j, n^{-1}b^{-1}c_j),\ j = 1, \dots, G\}$.

CONDITION 2. (i) The lth derivative of K is bounded on its support, $l = 0, 1, \dots, p$.

(ii) The density function f and its first $p+1$ derivatives are bounded.

Proof of the following theorem can be found in Cheng (1994).

THEOREM 1. *Suppose that Condition* 2 *holds. Then*

$$E\left(\widehat{f^{(\nu)}}(0)\right) = f^{(\nu)}(x) + \frac{\nu! f^{(p+1)}(0)}{(p+1)!}\left(\int t^{p+1} K^*_{p+1,\nu}(t)\,dt\right) h^{p+1-\nu} + o(h^{p+1-\nu}), \tag{5}$$

and

$$\operatorname{Var}\left(\widehat{f^{(\nu)}}(0)\right) = \frac{\nu!^2 f(0)}{nh^{2\nu+1}} \int K^*_{p+1,\nu}(t)^2\,dt + o\left(\frac{1}{nh^{2\nu+1}}\right), \tag{6}$$

as $n \to \infty$, $h \to 0$, $nh^{2\nu+1} \to \infty$ *and* $b/h \to 0$.

Following Fan and Gijbels (1992) and Ruppert and Wand (1994), the local polynomial fitting also adapts automatically to boundary regions for the density estimation setting.

REMARK 1. Lejenue and Sarda (1992) and Jones (1993) proposed fitting local polynomials to the empirical density function. The resulting estimator of $f^{(\nu)}(0)$ is a kernel estimator with the kernel $K^*_{p+1,\nu}$. All the results for the density estimation setting given in this article apply to that estimator as well.

3. Optimal weight function and weak minimaxity. Local polynomial estimators are intuitive and achieve boundary corrections automatically. An interesting question is what would be an optimal weighting scheme at the boundary regions. The most important case is when $x = 0$ and it is the situation considered here. We will discuss the problem in the regression setting. The optimal weight function is closely related to the weak minimax problem defined in (15). For simplicity, we omit asymptotically negligible terms throughout this article.

3.1. *Optimal boundary weighting scheme.* For any nonnegative weight function K, minimizing the right-hand side of (4) with respect to h, we obtain the best asymptotic mean squared error

$$\gamma_{p+1,\nu}\left(T_{p+1,\nu}(K)\right)^{2/(2p+3)} (f^{(p+1)}(0))^{2s} \left(\frac{\sigma^2(0)}{g_X(0)n}\right)^r,$$

where

$$T_{p+1,\nu}(K) \equiv \left|\int t^{p+1} K^*_{p+1,\nu}(t)\,dt\right|^{2\nu+1} \left(\int K^*_{p+1,\nu}(t)^2\,dt\right)^{p+1-\nu}, \tag{7}$$

$$\gamma_{p+1,\nu} = \nu!^2 r^{-r} s^{-s} (p+1)!^{-2s} \tag{8}$$

and

$$r = \frac{2(p+1-\nu)}{2p+3}, \qquad s = \frac{2\nu+1}{2p+3}.$$

Note that the asymptotically optimal mean squared error depends on the weight function K only through the quantity $T_{p+1,\nu}(K)$. Next, we find the best weight function for the local polynomial at the left boundary point; that is,

$$\min_{\substack{K\geq 0\\ K \text{ Lipschitz continuous}}} T_{p+1,\nu}(K), \tag{9}$$

where by Lipschitz continuity we mean that there exists a constant C such that $|K(x)-K(y)| \leq C|x-y|$. The solution turns out to be the triangular weight function $K_0(t) = (1-t)I_{[0,1]}(t)$ for all p and ν.

THEOREM 2. *For any p and ν, the triangular weight function $K_0(t) = (1-t)I_{[0,1]}(t)$ minimizes $T_{p+1,\nu}(K)$ among all nonnegative and Lipschitz continuous functions. Furthermore, with $\gamma_{p+1,\nu}$ defined by* (8),

$$T_{p+1,\nu}(K_0) = \left(\frac{\theta_{p+1,\nu}}{\gamma_{p+1,\nu}}\right)^{(2p+3)/2}, \tag{10}$$

where

$$\theta_{p+1,\nu} = \left(\frac{2p+3}{2\nu+1}\right)\left(\frac{(p+\nu+2)!}{(p-\nu+1)!\nu!}\right)^2\left(\frac{r}{2(p+\nu+2)}\right)^r\left(\frac{(p+1)!}{(2p+3)!}\right)^{2s}. \tag{11}$$

Let the resulting equivalent kernel [see (2)] of K_0 be

$$K^{\text{opt}}_{p+1,\nu}(t) \equiv \nu! e_\nu^T S^{-1}(1, t, \ldots, t^p)^T K_0(t) = \sum_{j=0}^{p+1} \lambda_j t^j \mathrm{I}_{[0,1]}(t). \tag{12}$$

The coefficients λ_j in $K^{\text{opt}}_{p+1,\nu}$ and the $(p+1)$th moment and L_2-norm [hence the value of $T_{p+1,\nu}(K_0)$] can be computed explicitly. In the Appendix we will show that

$$\lambda_j = \frac{(-1)^{j+\nu}(p+j+1)!(p+\nu+2)!}{j!^2\nu!(p-\nu)!(p-j+1)!(j+\nu+1)}, \qquad j = 0, 1, \ldots, p+1, \tag{13}$$

$$\int t^{p+1} K^{\text{opt}}_{p+1,\nu}(t)\,dt = \frac{(-1)^{\nu+p}(p+\nu+2)!(p+1)!^2}{\nu!(2p+3)!(p-\nu+1)!}$$

and

$$\int K^{\text{opt}}_{p+1,\nu}(t)^2\,dt = \frac{2(p+\nu+2)(p+\nu+1)!^2}{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2}.$$

From Theorem 1, it is easy to see that the problem of finding best boundary weight functions for local polynomial density derivative estimation is the same as (9).

3.2. *Weak minimaxity.* Given a constant C, $0 < C < \infty$, the condition

$$\left| f(z) - \sum_{j=0}^{p} \frac{f^{(j)}(0)}{j!} z^j \right| \le C \frac{|z|^{p+1}}{(p+1)!}, \qquad z \in [0,1] \tag{14}$$

reflects the idea that "the $(p+1)$th right derivative of f at zero is bounded by C." For a finite integer D, $D \ge 2$, define a set of regression functions

$$C_{p+1}^F = \{f_0, f_1, \dots, f_D\},$$

where f_0 and f_1, satisfying (14), are special functions given in an earlier draft of this paper [Cheng, Fan and Marron (1993)], $f_2 = Cx^{p+1}/(p+1)!$ and f_j is any function on $[0,1]$ such that $|f_j^{(p+1)}(0)| \le C$, $j = 3, \dots, D$. The linear minimax risk when estimating $f^{(\nu)}(0)$ over the class C_{p+1}^F is defined as

$$R_{p+1,\nu}(n, C_{p+1}^F) = \inf_{\widehat{f_L^{(\nu)}}(0) \text{ linear}} \sup_{f \in C_{p+1}^F} E\Big[\Big(\widehat{f_L^{(\nu)}}(0) - f^{(\nu)}(0)\Big)^2 \big| X_1, \dots, X_n\Big]. \tag{15}$$

We call $R_{p+1,\nu}(n, C_{p+1}^F)$ a weak linear minimax risk for estimating $f^{(\nu)}(0)$.

It can be shown [see Cheng, Fan and Marron (1993)] that

$$R_{p+1,\nu}(n, C_{p+1}^F) = \gamma_{p+1,\nu} C^{2s} \left(\frac{\sigma^2(0)}{n g_X(0)}\right)^r (T_{p+1,\nu}(K_0))^{2/(2p+3)}, \tag{16}$$

and that the local polynomial fit of order p with the triangular weight function K_0 and bandwidth

$$h_0 = \left(\frac{(p-\nu+1)(2p+3)!^2 \sigma^2(0)}{(p+\nu+2)(p+1)!^2(2p+3) g_X(0) C^2 n}\right)^{1/(2p+3)}$$

achieves this asymptotic minimax risk. Thus, the weak linear minimax risk is closely related to the optimal boundary weight function K_0, and, in a weak minimax sense, local polynomial fitting based on the triangular weight function provides the best possible boundary correction method.

4. Minimax optimal kernel and minimaxity. We discuss boundary linear minimax risks and the corresponding optimal kernels for a more general class of regressions. Define the following class of regression functions:

$$C_{p+1} = \{\text{regression functions } f \text{ that satisfy (14)}\}.$$

Then the linear minimax risk for estimating $f^{(\nu)}(0)$ over the class C_{p+1} is given as

$$R_{p+1,\nu}(n, C_{p+1}) = \inf_{\widehat{f_L^{(\nu)}}(0) \text{ linear}} \sup_{f \in C_{p+1}} E\Big[\Big(\widehat{f_L^{(\nu)}}(0) - f^{(\nu)}(0)\Big)^2 \big| X_1, \dots, X_n\Big]. \tag{17}$$

In the following sections we discuss, for general values of p and ν, bounds and values of the linear minimax risk and minimax optimal boundary kernels.

4.1. *Minimax optimal boundary kernels.* In definition (17), the design density $g_X(\cdot)$ is fixed and can be regarded as known. If $g_X(0)$ is known and uniform around the point zero, a higher order kernel estimator of $f^{(\nu)}(0)$ is

$$\widehat{f}_h(0) = \frac{\sum_{i=1}^n K_{p+1,0}(X_i/h)Y_i}{\sum_{i=1}^n K_{p+1,0}(X_i/h)},$$
$$\widehat{f_h^{(\nu)}}(0) = \frac{\nu!}{nh^{\nu+1}g_X(0)} \sum_{i=1}^n K_{p+1,\nu}\left(\frac{X_i}{h}\right)\{Y_i - \tilde{f}(0)\} \quad \text{for } \nu > 0,$$

where $h > 0$ and the boundary kernel $K_{p+1,\nu}$ satisfies

$$\int_0^\infty t^j K_{p+1,\nu}(t)\,dt = \delta_{\nu,q}, \qquad q = 0, 1, \dots, p,$$
$$\int_0^\infty t^{p+1} K_{p+1,\nu}(t)\,dt \neq 0. \tag{18}$$

Here $\tilde{f}(0)$ is a kernel regression estimator defined as

$$\tilde{f}(0) = \frac{1}{nhg_X(0)} \sum_{i=1}^n K_0\left(\frac{X_i}{h}\right)Y_i,$$

for the given bandwidth h and the uniform kernel K_0. It can be shown that, as $n \to \infty$, $h \to 0$ and $nh^{2\nu+1} \to \infty$,

$$\begin{aligned}\sup_{f\in C_{p+1}} &E\left[\left(\widehat{f_h^{(\nu)}}(0) - f^{(\nu)}(0)\right)^2 \middle| X_1, \dots, X_n\right] \\ &\le \left(\frac{\nu! C}{(p+1)!}\right)^2 \left(\int_0^\infty \left|t^{p+1}K_{p+1,\nu}(t)\right| dt\right)^2 h^{2(p+1-\nu)} \\ &\quad + \frac{\nu!^2\sigma^2(0)}{nh^{2\nu+1}g_X(0)} \int_0^\infty K_{p+1,\nu}(t)^2\,dt.\end{aligned} \tag{19}$$

Minimizing (19) with respect to h, we obtain

$$\sup_{f\in C_{p+1}} E\left[\left(\widehat{f_h^{(\nu)}}(0) - f^{(\nu)}(0)\right)^2 \middle| X_1, \dots, X_n\right] \le \xi_{p+1,\nu}(Q_{p+1,\nu}(K_{p+1,\nu}))^{2/(2p+3)},$$

where

$$Q_{p+1,\nu}(K_{p+1,\nu}) = \left(\int_0^\infty \left|t^{p+1}K_{p+1,\nu}(t)\right| dt\right)^{2\nu+1} \left(\int_0^\infty K_{p+1,\nu}(t)^2\,dt\right)^{p+1-\nu} \tag{20}$$

and

$$\xi_{p+1,\nu} = \gamma_{p+1,\nu} C^{2s} \left(\frac{\sigma^2(0)}{ng_X(0)}\right)^r.$$

Then, since kernel estimators are linear, for any boundary kernel $K_{p+1,\nu}$ satisfying (18),

$$R_{p+1,\nu}(n, C_{p+1}) \le \xi_{p+1,\nu}(Q_{p+1,\nu}(K_{p+1,\nu}))^{2/(2p+3)}.$$

Therefore, if $\overline{K}_{p+1,\nu}$ minimizes $Q_{p+1,\nu}(K_{p+1,\nu})$ among all $K_{p+1,\nu}$ satisfying (18), then we have an upper bound for the linear minimax risk

$$R_{p+1,\nu}(n, C_{p+1}) \le \xi_{p+1,\nu}\Big(Q_{p+1,\nu}\Big(\overline{K}_{p+1,\nu}\Big)\Big)^{2/(2p+3)}. \tag{21}$$

Such a kernel is called a minimax optimal boundary kernel.

For any nonnegative integer p and $\nu = 0, 1, \dots, p$, define a boundary kernel $\overline{K}_{p+1,\nu}$ as

$$\overline{K}_{p+1,\nu}(t) = \left(\sum_{j=0}^{p} c_j t^j - \theta t^{p+1}\right)_+ - \left(\sum_{j=0}^{p} c_j t^j + \theta t^{p+1}\right)_-, \tag{22}$$

where c_j, $j = 0, \dots, p$, and θ are determined by

$$\begin{aligned} &\int_0^\infty t^j \overline{K}_{p+1,\nu}(t)\,dt = \delta_{j,\nu}, \qquad j = 0, \dots, p, \\ &\int_0^\infty t^{p+1} \overline{K}_{p+1,\nu}(t)\,dt = \theta \neq 0. \end{aligned} \tag{23}$$

The following theorem on the minimax optimal boundary kernels is a direct generalization of Theorem 1 in Sacks and Ylvisaker (1981), so we omit its proof.

THEOREM 3. *For any nonnegative integer* p *and* $\nu = 0, 1, \dots, p$, *any minimax optimal boundary kernel is determined by* (22) *and* (23).

Therefore, minimax optimal boundary kernels are defined implicitly and have no closed forms. Figure 1 shows, for some values of p and ν, a numerical approximation of the rescale of $\overline{K}_{p+1,\nu}$ whose support is the interval $[0, 1]$.

4.2. *Minimax risks.* More about the linear minimax risk is investigated in this section. First, combining (16), (21) and the fact that $C^F_{p+1} \subseteq C_{p+1}$, we have the following asymptotic upper and lower bounds for the value of $R_{p+1,\nu}(n, C_{p+1})$.

THEOREM 4. *For any nonnegative integer* p *and* $\nu = 0, 1, \dots, p$, *asymptotically*,

$$\begin{aligned} \xi_{p+1,\nu}\Big(Q_{p+1,\nu}\Big(\overline{K}_{p+1,\nu}\Big)\Big)^{2/(2p+3)} &\ge R_{p+1,\nu}(n, C_{p+1}) \\ &\ge \xi_{p+1,\nu}\big(T_{p+1,\nu}(K_0)\big)^{2/(2p+3)}. \end{aligned}$$

We conjecture the first inequality in Theorem 4 is indeed an equality:

$$R_{p+1,\nu}(n, C_{p+1}) = \xi_{p+1,\nu}\Big(Q_{p+1,\nu}\Big(\overline{K}_{p+1,\nu}\Big)\Big)^{2/(2p+3)}(1 + o(1)). \tag{24}$$

Note that the minimax optimal kernels $\overline{K}_{p+1,\nu}$ are defined implicitly. Hence the upper bounds given in Theorem 4 have no explicit formulae. However,

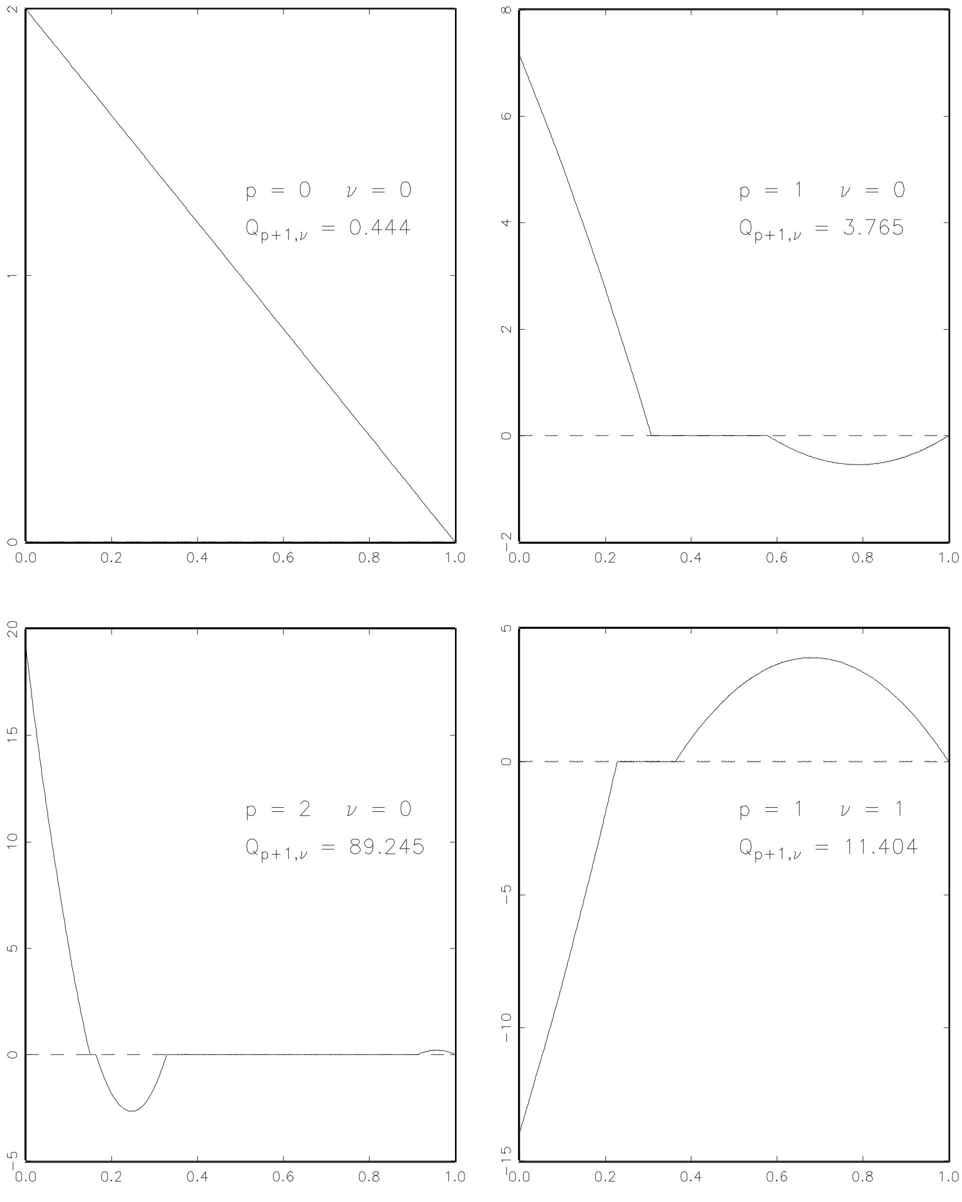

FIG. 1. *The optimal minimax kernel* $\overline{K}_{p+1,\nu}$ *defined in* (22) *and* (23) *is rescaled to* $[0,1]$ *and plotted by the solid line. The quantity* $Q_{p+1,\nu}(\overline{K}_{p+1,\nu})$ *is denoted as* $Q_{p+1,\nu}$. *Note that* $Q_{p+1,\nu}(c^{-(\nu+1)}\overline{K}_{p+1,\nu}(\cdot/c)) = Q_{p+1,\nu}(\overline{K}_{p+1,\nu}(\cdot))$ *for any* $c > 0$. *Upper left,* $p = 0$, $\nu = 0$; *upper right,* $p = 1$, $\nu = 0$; *lower left,* $p = 2$, $\nu = 0$; *lower right,* $p = 1$, $\nu = 1$.

since $(\nu!)^{-1}K^{\mathrm{opt}}_{p+1,\nu}$ satisfies (18) and $\overline{K}_{p+1,\nu}$ is a minimax optimal kernel, we have the following corollary.

COROLLARY 5. *For any nonnegative integer* p *and* $\nu = 0, 1, \dots, p$, *asymptotically*,

$$\xi_{p+1,\nu}\left(Q_{p+1,\nu}\left(\frac{1}{\nu!}K^{\mathrm{opt}}_{p+1,\nu}\right)\right)^{2/(2p+3)} \geq R_{p+1,\nu}(n, C_{p+1}) \geq \xi_{p+1,\nu}\big(T_{p+1,\nu}(K_0)\big)^{2/(2p+3)}.$$

Note that explicit expressions for the asymptotic upper and lower bounds in Corollary 5 are available from (10) and (13). The upper bound in Corollary 5 can be obtained via the local polynomial fitting with the triangular kernel $K_0 = (1-t)I_{[0,1]}(t)$. In the most common case where $p = 1$ and $\nu = 0$, $K^{\mathrm{opt}}_{2,0}(t) = 6(1-3t+2t^2)I_{[0,1]}(t)$ and Corollary 5 gives

$$2.2441 \approx \frac{3\cdot 15^{1/5}}{2^{6/5}} \geq \left(\frac{n g_X(0)}{\sqrt{C}\sigma^2(0)}\right)^{4/5} R_{2,0}(n, C_2) \geq 3\cdot 15^{-1/5} \approx 1.7454.$$

5. Concluding remarks. On the issue of boundary corrections, we explored the possibility of extending local polynomial fitting regression techniques to the density estimation setting. Various optimization problems were investigated and connections among them were discussed.

For local polynomial estimation, a best weighting scheme, minimizing $T_{p+1,\nu}(K)$ over all nonnegative Lipschitz continuous functions K, is simply the triangular weight function and, even more surprisingly, this statement holds for general p and ν. We also showed that best local polynomial estimators are 100% efficient in a weak linear minimax sense. Explicit formulae for the weak linear minimax risks, equivalent kernels of the triangular weight scheme, its norm and moments are available.

More general linear minimax risks are also investigated. They associate with problems of minimizing $Q_{p+1,\nu}(K_{p+1,\nu})$ among all boundary kernels $K_{p+1,\nu}$. The solutions, and hence the linear minimax risk, are implicitly defined and can only be approximated numerically. Fortunately, we can combine weak and general linear minimax results and obtain explicit upper and lower bounds.

In minimax senses, kernel estimators with minimax optimal kernels are better than local polynomial estimators yielding weak minimax efficiency. How much is the gain of going from weak minimax to general minimax optimal kernels? Consider the most important case of $p = 1$ and $\nu = 0$. The minimax optimal kernel after being rescaled to $[0, 1]$ is approximately $(-12.31t^2 - 19.44t + 7.128)_+ - (12.31t^2 - 19.44t - 7.128)_-$ and $Q_{2,0}(\overline{K}_{2,0}) \approx 3.7646$. A best linear estimator, in a weak minimax sense, of $f(0)$ is the local linear estimator with

TABLE 1

Minimax efficiencies $R_{p+1,\nu}$ and the ratio $r_{p+1,\nu}$. (In each cell, the first value is $R_{p+1,\nu}$ and the second value is $r_{p+1,\nu}$)

	$\nu = 0$		$\nu = 1$		$\nu = 2$	
$p = 0$	1	1				
$p = 1$	0.9464	0.8217	0.9740	0.9707		
$p = 2$	0.8583	0.7399	0.6922	0.4937	0.6582	0.6834

weight K_0. Its relative efficiency, to the minimax optimal kernel estimator, is

$$\left(\frac{Q_{2,0}(\overline{K}_{2,0})}{Q_{2,0}(K_{2,0}^{\text{opt}})}\right)^{2/5} \approx \frac{3.7646^{2/5}}{(6/5)(15/2)^{1/5}} \approx 0.9464.$$

More generally, let $R_{p+1,\nu}$ be the ratio of (24) to the upper bound in Corollary 5, and let $r_{p+1,\nu}$ be the ratio of the lower bound in Theorem 4 to the upper bound in Theorem 4. Table 1 shows a few values of the ratios.

As noted above, the best local polynomial estimators are preferable to the minimax optimal kernel estimators for the following reasons. First, local polynomial fitting is far more intuitive than estimators based on minimax optimal kernels which have funny forms. Second, local polynomial estimators with best weight function attain weak minimax efficiencies. Third, a best weight function is the simple triangular function and it holds for any p and ν. Finally, best local polynomial estimators are highly efficient in the usual minimax sense.

6. Proofs.

PROOF OF THEOREM 2. Let $B(K) = \int t^{p+1} K^*_{p+1,\nu}(t)\,dt$ and $V(K) = \int K^*_{p+1,\nu}(t)^2\,dt$. Then, with $K_c(\cdot) = c^{-1}K(\cdot/c)$, we have

$$B(K_c) = c^{p+1-\nu} B(K) \quad \text{and} \quad T_{p+1,\nu}(K_c) = T_{p+1,\nu}(K).$$

Therefore, minimizing $T_{p+1,\nu}(K)$ among all nonnegative and Lipschitz-continuous K is equivalent to

$$\min V(K) \tag{25}$$

subject to $B(K) = B(K_0)$, $K \geq 0$ and Lipschitz continuous.

We first of all show that the solution to (25) exists. Suppose that $\{K_n\}$ is a sequence of functions satisfying the side conditions of (25) such that

$$\lim_{n\to\infty} V(K_n) = \inf V(K).$$

Then, by the weak topology of $L^2[0,\infty)$, there exists a subsequence $\{n_j\}$ such that

$$\lim_{j\to\infty} K_{n_j} = K^0 \quad \text{in } L^2[0,\infty),$$

for some $K^0 \in L^2[0, \infty)$. By using the same argument as establishing a similar relationship between almost sure convergence and convergence in probability, there exists a further subsequence that converges to K^0 almost everywhere. By the Lipschitz continuity of the function K_{n_j}, K^0 is Lipschitz continuous (if necessary redefine the values of K^0 on a set with measure zero). Clearly K^0 satisfies the side conditions of (25) and is a solution to (25).

Next, we determine the solution K^0. Let $K^0_{p+1,\nu}$ be the equivalent kernel (2) induced by K^0. Denote the support of K^0 as A. Let P_0 be a polynomial of order $(p+1)$ on A which minimizes

$$\int_A \left(K^0_{p+1,\nu}(t) - \sum_{j=0}^{p+1} \theta_j t^j \right)^2 dt$$

among all $(\theta_0, \theta_1, \ldots, \theta_{p+1})$. Write

$$K^0_{p+1,\nu}(\cdot) = P_0(\cdot) + r_0(\cdot). \tag{26}$$

Then

$$\int_A t^q r_0(t)\, dt = 0, \qquad q = 0, 1, \ldots, p+1. \tag{27}$$

Let

$$K^*_\eta(t) = \big(K^0_{p+1,\nu}(t) - \eta r_0(t)\big) I_A(t).$$

Then, by the definition of $K^0_{p+1,\nu}$, when η is small enough, K^*_η has at most p roots in A.

We now show that K^*_η is an equivalent kernel (2) induced by some nonnegative Lipschitz continuous weight function. Let $\tau_1, \ldots, \tau_d$, $d \le p$, be the roots of K^*_η in A. Put

$$P(t) = (-1)^\kappa \prod_{j=1}^{d} (t - \tau_j), \qquad K_\eta(t) = K^*_\eta(t)/P(t),$$

where $\kappa = 0$ or 1 so that K_η is nonnegative. Then

$$K^*_\eta(t) = P(t) K_\eta(t) I_A(t).$$

Write the equivalent kernel (2) of K_η as $Q(t)K_\eta(t)I_A(t)$ with $Q(t)$ a polynomial of order p. Thus, by (3) and (27),

$$\begin{aligned} 0 &= \int_A (K^*_\eta(t) - Q(t)K_\eta(t))(P(t) - Q(t))\, dt \\ &= \int_A K_\eta(t)(P(t) - Q(t))^2\, dt. \end{aligned}$$

This implies that $P(t) = Q(t)$ for all $t \in A$ and hence K^*_η is induced by the nonnegative Lipschitz continuous kernel K_η. Note that, by (27), K_η also satisfies the constraint $B(K_\eta) = B(K_0)$. From

$$V(K_0) \le V(K_\eta),$$

or, by (26), equivalently

$$\int_A (P_0(t) + r_0(t))^2\,dt \le \int_A (P_0(t) + (1-\eta)r_0(t))^2\,dt,$$

and (27), we obtain that

$$\int_A r_0(t)^2\,dt \le 0.$$

Thus, we conclude that $K^0_{p+1,\nu}(t) = P_0(t)$, a polynomial of degree $p+1$ in A. Hence K^0 is a linear function in A. Lipschitz continuity of K^0 requires that $A = [0, c]$ for some positive constant c and $K^0(t) = (1 - t/c)I_{[0,c]}(t)$. The side condition $B(K^0) = B(K_0)$ implies $c = 1$, namely, $K^0 = K_0$.

APPENDIX

Calculation of the function $K^{\mathrm{opt}}_{p+1,\nu}$, its norm and $(p+1)$th moment. The Legendre polynomials on the interval $[-1, 1]$ are defined as

$$P_n(x) = \frac{d^n}{dx^n}((1+x)(1-x))^n, \qquad -1 \le x \le 1,\ n = 0, 1, 2, \ldots.$$

The linear transformation $y = (x+1)/2$ of these polynomials yields an orthogonal system with respect to Lebesgue measure on $[0, 1]$. Write

$$Q_n(y) = \frac{d^n}{dy^n}(y(1-y))^n \equiv \sum_{j=0}^{n} q_{n,j}\, y^j, \qquad 0 \le y \le 1,\ n = 0, 1, 2, \ldots.$$

Then

$$\begin{aligned} \|Q_n\|^2 &= \int_0^1 Q_n^2(y)\,dy = (-1)^n \int_0^1 y^n(1-y)^n \frac{d^{2n}}{dy^{2n}}(y(1-y))^n\,dy \\ &= \int_0^1 y^n(1-y)^n(2n)!\,dy = (2n)!\frac{n!^2}{(2n+1)!} = \frac{n!^2}{2n+1}. \end{aligned} \tag{28}$$

Explicitly,

$$Q_n(y) = \frac{d^n}{dy^n}\sum_{j=0}^{n}\binom{n}{j}(-y)^j y^n = \sum_{j=0}^{n}\binom{n}{j}(-1)^j\frac{(n+j)!}{j!}y^j.$$

So,

$$q_{n,j} = \binom{n}{j}(-1)^j\frac{(n+j)!}{j!}, \qquad n = 0, 1, \ldots, p+1,\ j = 0, 1, \ldots, n.$$

Since $K^{\mathrm{opt}}_{p+1,\nu}$ is a polynomial of order $(p+1)$ [see (12)], we can write

$$K^{\mathrm{opt}}_{p+1,\nu}(x) = \sum_{i=0}^{p+1} a_i Q_i(x).$$

The coefficients a_i can be determined by the moment properties in (3). Denote

$$\beta = \int_0^1 x^{p+1} K^{\text{opt}}_{p+1,\nu}(x)\,dx. \tag{29}$$

Then

$$\begin{aligned} a_i\|Q_i\|^2 &= \int_0^1 Q_i(x) K^{\text{opt}}_{p+1,\nu}(x)\,dx \\ &= \begin{cases} 0, & \text{if } 0 \le i < \nu, \\ \nu! q_{i,\nu}, & \text{if } \nu \le i \le p, \\ \nu! q_{p+1,\nu} + q_{p+1,p+1}\beta, & \text{if } i = p+1. \end{cases} \end{aligned} \tag{30}$$

Therefore, from (28) and (30),

$$\begin{aligned} \frac{1}{\nu!} K^{\text{opt}}_{p+1,\nu}(x) &= \sum_{i=0}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} Q_i(x) \\ &\quad + \frac{(2p+3)}{(p+1)!^2}\left(q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta\right) Q_{p+1}(x) \\ &= \sum_{i=0}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} \sum_{j=0}^{i} q_{i,j} x^j \\ &\quad + \frac{(2p+3)}{(p+1)!^2}\left(q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta\right) Q_{p+1}(x) \\ &= \sum_{j=0}^{p}\left(\sum_{i=j\vee\nu}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} q_{i,j}\right) x^j \\ &\quad + \frac{(2p+3)}{(p+1)!^2}\left(q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta\right) Q_{p+1}(x). \end{aligned} \tag{31}$$

Here,

$$\begin{aligned} \sum_{i=j\vee\nu}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} q_{i,j} &= \frac{(-1)^{j+\nu}}{j!^2\nu!^2} \sum_{i=j\vee\nu}^{p} \frac{(i+\nu)!(2i+1)(j+i)!}{(i-\nu)!(i-j)!} \\ &\qquad \left[\text{Note: } (2i+1) = \frac{\{(i+j+1)(i+\nu+1) - (i-j)(i-\nu)\}}{j+\nu+1}\right] \\ &= \frac{(-1)^{j+\nu}}{j!^2\nu!^2(j+\nu+1)} \sum_{i=j\vee\nu+1}^{p} \left(\frac{(i+\nu+1)!(j+i+1)!}{(i-\nu)!(i-j)!} \right. \\ &\qquad \left. - \frac{(i+\nu)!(j+i)!}{(i-\nu-1)!(i-j-1)!}\right) \\ &\quad + \frac{(-1)^{j+\nu}}{j!^2\nu!^2} \frac{((j\vee\nu)+\nu)!(2(j\vee\nu)+1)(j+(j\vee\nu))!}{((j\vee\nu)-\nu)!((j\vee\nu)-j)!} \\ &= \frac{(-1)^{j+\nu}}{j!^2\nu!^2(j+\nu+1)} \frac{(p+\nu+1)!(j+p+1)!}{(p-\nu)!(p-j)!}. \end{aligned} \tag{32}$$

Also, since

$$Q_i(1) = \frac{d^i}{dy^i}(y(1-y))^i\Big|_{y=1} = (-1)^i i!$$

and $K^{\text{opt}}_{p+1,\nu}(1) = 0$ [see (12)],

$$K^{\text{opt}}_{p+1,\nu}(1) = \sum_{i=0}^{p+1} a_i Q_i(1) = \sum_{i=0}^{p+1} a_i (-1)^i i! = 0.$$

This is the same as

$$\begin{aligned} &\sum_{i=\nu}^{p} \frac{(2i+1)}{i!^2}\nu! q_{i,\nu}(-1)^i i! \\ &\quad + \frac{(2p+3)}{(p+1)!^2}(\nu! q_{p+1,\nu} + q_{p+1,p+1}\beta)(-1)^{p+1}(p+1)! = 0. \end{aligned} \tag{33}$$

The first term is

$$\begin{aligned} &\sum_{i=\nu}^{p} \frac{(2i+1)}{i!^2}\nu! \binom{i}{\nu}(-1)^\nu \frac{(i+\nu)!}{\nu!}(-1)^i i! \\ &= \frac{(-1)^\nu}{\nu!}\sum_{i=\nu}^{p}\frac{(-1)^i(i+\nu)!(2i+1)}{(i-\nu)!} \\ &\qquad\qquad [\text{Note: } (2i+1) = (i+\nu+1) + (i-\nu)] \\ &= \frac{(-1)^\nu}{\nu!}\left[\sum_{i=\nu}^{p}\frac{(-1)^i(i+\nu+1)!}{(i-\nu)!} + \sum_{i=\nu+1}^{p}\frac{(-1)^i(i+\nu)!}{(i-\nu-1)!}\right] \\ &= \frac{(-1)^{\nu+p}(p+\nu+1)!}{\nu!(p-\nu)!}. \end{aligned}$$

Thus (33) yields

$$\beta = \frac{(-1)^{\nu+p}(p+\nu+2)!(p+1)!^2}{\nu!(2p+3)!(p-\nu+1)!}. \tag{34}$$

Combining this with (31) and (32) we have

$$K^{\text{opt}}_{p+1,\nu}(x) = \sum_{j=0}^{p+1} \lambda_j x^j I_{[0,1]}(x),$$

where λ_j, $j = 0, 1, \dots, p+1$, are given in (13). Since the polynomials $\{Q_i\}$ are orthogonal,

$$\begin{aligned} \|K^{\text{opt}}_{p+1,\nu}\|^2 &= \sum_{i=0}^{p+1} a_i^2 \|Q_i\|^2 \\ &= \sum_{i=\nu}^{p} \nu!^2 q_{i,\nu}^2 \frac{2i+1}{i!^2} + \frac{(2p+3)}{(p+1)!^2}\big(\nu! q_{p+1,\nu} + q_{p+1,p+1}\beta\big)^2 \end{aligned} \tag{35}$$

$$= \sum_{i=\nu}^{p} \frac{(2i+1)(i+\nu)!^2}{\nu!^2(i-\nu)!^2} + \frac{(2p+3)}{(p+1)!^2}\left(\nu! q_{p+1,\,\nu} + q_{p+1,\,p+1}\beta\right)^2$$

$$\left[\text{From (34) and noticing that } (2i+1) = \frac{\{(i+\nu+1)^2 - (i-\nu)^2\}}{2\nu+1}\right]$$

$$= \frac{1}{\nu!^2(2\nu+1)} \sum_{i=\nu+1}^{p} \left\{ \frac{(i+\nu+1)!^2}{(i-\nu)!^2} - \frac{(i+\nu)!^2}{(i-\nu-1)!^2} \right\}$$

$$+ \frac{(2\nu)!^2(2\nu+1)}{\nu!^2} + \frac{(p+\nu+1)!^2}{(2p+3)\nu!^2(p-\nu)!^2}$$

$$= \frac{2(p+\nu+2)(p+\nu+1)!^2}{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2}.$$

Acknowledgments. The authors are grateful to Lawrence D. Brown, David Ruppert and two referees for important and helpful suggestions. We thank Mr. Jing-Ting Zhang for helping us find the optimal minimax kernels.

REFERENCES

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

BROWN, L., LOW, M. and ZHAO, L. (1997). Anomalous behavior of fixed parameter asymptotic rates of convergence in nonparametric function estimation problems. Unpublished manuscript.

CHENG, M. Y. (1994). On boundary effects of smooth curve estimators. Ph.D. dissertation, Dept. Statistics, Univ. North Carolina, Chapel Hill.

CHENG, M. Y. (1997). Boundary-aware estimators of integrated density derivative products. *J. Roy. Statist. Soc. Ser. B* **59** 191–203.

CHENG, M. Y., FAN, J. and MARRON, J. S. (1993). Minimax efficiency of local polynomial fit estimators at boundaries. Mimeo Series 2098, Inst. Statist., Univ. North Carolina, Chapel Hill.

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

DJOJOSUGITO, R. A. and SPECKMAN, P. L. (1992). Boundary bias correction in nonparametric density estimation. *Comm. Statist. Theory Methods* **21** 69–88.

DONOHO, D. L. and JOHNSTONE, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

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 over hyper rectangles, implications. *Ann. Statist.* **18** 1416–1437.

EFROIMOVICH, S. (1996). On nonparametric regression for i.i.d. observations in general setting. *Ann. Statist.* **24** 1126–1144.

EPANECHNIKOV, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.* **14** 153–158.

EUBANK, R. L. and SPECKMAN, P. (1991). A bias reduction theorem with applications in nonparametric regression. *Scand. J. Statist.* **18** 211–222.

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

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

FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036.

FAN, J. and HALL, P. (1994). On curve estimation by minimizing mean absolute deviation and its implications. *Ann. Statist.* **22** 867–885.

FAN, J. and MARRON, J. S. (1994). Fast implementations of nonparametric curve estimators. *J. Comput. Graph. Statist.* **3** 35–56.

FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. and ENGELS, J. (1997). Local polynomial fitting: a standard for nonparametric regression. *Ann. Inst. Statist. Math.* **49** 79–99.

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

GASSER, T., MÜLLER, H. G. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

GRANOVSKY, B. L. and MÜLLER, H. G. (1991). Optimizing kernel methods: a unifying variational principle. *Internat. Statist. Rev.* **59** 373–388.

HALL, P. and WEHRLY, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *J. Amer. Statist. Assoc.* **86** 665–672.

HASTIE, T. and LOADER, C. (1993). Local regression: automatic kernel carpentry. *Statist. Sci.* **8** 120–143.

JONES, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* **3** 135–146.

LEJEUNE, M. and SARDA, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.* **14** 457–471.

MÜLLER, H. G. (1991). Smooth optimal kernel estimators near endpoints. *Biometrika* **78** 521–530.

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

RICE, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13** 893–900.

RICE, J. and ROSENBLATT, M. (1981). Integrated mean squared error of a smoothing spline. *J. Approx. Theory* **33** 353–365.

RUPPERT, D. and WAND, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

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

SCHUSTER, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* **14** 1123–1136.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

MING-YEN CHENG
INSTITUTE OF MATHEMATICAL STATISTICS
NATIONAL CHUNG CHENG UNIVERSITY
TAIWAN

JIANQING FAN
J. S. MARRON
DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260
E-MAIL: jfan@stat.unc.edu