

THE INSTITUTE  
OF STATISTICS

THE UNIVERSITY OF  
NORTH CAROLINA



MINIMAX EFFICIENCY OF LOCAL POLYNOMIAL  
FIT ESTIMATORS AT BOUNDARIES

by

M.Y. Cheng

J. Fan

J.S. Marron

May 1993

Mimeo Series #2098  
DEPARTMENT OF STATISTICS  
Chapel Hill, North Carolina

Mimeo M. Y. Cheng, J. Fan  
Series J. S. Marron  
2098 Minimax Efficiency of  
Local Polynomial Fit  
Estimators at Boundaries

Name	Date

Department of Statistics Library

# Minimax Efficiency of Local Polynomial Fit Estimators at Boundaries

Ming-Yen Cheng, Jianqing Fan, and J. S. Marron  
Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27514-3260

MAY 10, 1993

## Abstract

Many popular curve estimators based on smoothing have difficulties caused by boundary effects. These effects are visually very disturbing in practice, and can play a dominant role in theoretical analysis. Local linear estimators are known to give very good visual performance at boundaries and to have an asymptotic rate of convergence which is as good as interior points. We show that these estimators are also optimal in the much deeper sense of best possible constant coefficient, using minimax lower bound results. This shows local linear estimators must be at least as efficient as the much more complicated "optimal boundary kernels". Parallel results for both regression and density estimation are presented. The results are extended to estimation of derivatives, since this is vital to applications in plug-in bandwidth selection.

## 1 Introduction

Nonparametric curve estimation methods make no assumptions on the shape of the curves of interest and hence allow flexible modeling of the data. If the support of the true curve has important boundaries then most such methods

give estimates that are severely biased in regions near the end points. This boundary problem affects 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 reduce its effect. Gasser & Müller (1979), Gasser, Müller, & Mammitzsch (1985), Granovsky & Müller (1991), Müller (1991a), and Müller (1991b) discuss using boundary kernels to correct this problem for the conventional kernel estimators. Rice (1984) suggests a linear combination of two kernel estimators with different bandwidths to reduce the bias. Schuster's (1985) mirror image estimator in density estimation "folds back" the probability mass that extends beyond the support. The estimator introduced in Hall & Wehrly (1991) is essentially a more sophisticated regression version of Schuster's approach. Djojosingito & Speckman (1992) approach boundary bias reduction based on a finite-dimensional projection in Hilbert space. Boundary effects for smoothing splines are discussed in Rice & Rosenblatt (1981). Eubank & Speckman (1991) also provide some boundary correction methods. Fan & Gijbels (1992) point out that the local linear regression smoother adapts to boundary estimation automatically. Moreover, unlike most other methods, the local linear regression smoother does not require knowledge of the location of the endpoints.

The main purpose of this article is to show that a local linear regression estimator is asymptotically efficient even in the deep sense of constant coefficient for estimating regression functions at endpoints in a minimax sense. A similar result for the density estimation setting is also presented. This result settles the important question of how local linear estimators compare with "optimal boundary kernels", by showing the former must be at least as efficient. We feel this gives local linear estimators an important advantage, because they are also ( i ) easier to interpret ( ii ) much easier to implement ( iii ) appear far faster to compute ( to factors of 100, see Fan & Marron (1992). )

We review some regression smoothers and the boundary adaptive property of the local linear estimator in section 2. Then the minimax efficiency of the local linear regression smoother is discussed in section 3. Modification of the method for the purpose of density estimation and boundary performance of the resulting estimator are investigated in section 4. Analogous boundary efficiency of that estimator follows immediately. Then, in section 5, we calculate the relative efficiency of the local linear estimator with the Gaussian

kernel, and also of Rice's boundary adjusted estimator. In sections 6 and 7, we extend the results to higher order local polynomial fitting for estimating function derivatives.

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 and Hall (1992), Donoho and Johnstone (1992), Brown and Low (1993), Efroimovich (1993), Fan (1993), 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, the minimax problem at a boundary point has not been studied 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.

## 2 Some Regression Smoothers

Suppose  $(X_1, Y_1), \dots, (X_n, Y_n)$  is a random sample from a population  $(X, Y)$  with density function  $f(x, y)$ . Our goal is to estimate the regression function  $m(x) = E(Y|X = x)$ . There are a number of estimators proposed for this purpose in the literature. Two of the simplest, but most widely studied are the Nadaraya-Watson and Gasser-Müller kernel estimators, see Chu and Marron (1991) for references and discussion.

Recently many nice properties of the local linear regression estimator have been presented, see Stone (1977) and Fan (1992). To understand this estimator, let  $\hat{a}$  and  $\hat{b}$  minimize the following weighted sum of squares

$$\sum_{i=1}^n (Y_i - a - b(X_i - x))^2 K\left(\frac{X_i - x}{h}\right). \quad (1)$$

The local linear regression estimator, defined to be  $\hat{a}$ , can be written as

$$\hat{m}_{LL}(x) = \frac{\sum_{i=1}^n \omega_i Y_i}{\sum_{i=1}^n \omega_i},$$

with

$$\omega_i = [S_{n,2} - (X_i - x)S_{n,1}] K\left(\frac{X_i - x}{h}\right),$$

where

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

Let  $f_X(\cdot)$  be the marginal density of  $X$ . In applications, there are often boundaries present in the support of  $f_X$ ; e.g.  $f_X$  is supported in  $[0, 1]$ . In that case most regression estimators need to be modified. Otherwise they are seriously biased in the boundary regions even to the extent of a slower rate of convergence. Fan and Gijbels (1992) show that the local linear smoother does not share this shortcoming; it retains the same rate of convergence at the boundaries as in the interior. Denote the conditional variance of  $Y$  on  $X$  by  $\sigma^2(x) = \text{Var}(Y|X = x)$ . The behavior of the estimator  $\widehat{m}_{LL}$  at the left endpoints  $x = ch, c > 0$  is restated in the following theorem. Some conditions are needed:

(A1)  $f_X(\cdot), m''(\cdot)$ , and  $\sigma(\cdot)$  are bounded on  $[0, 1]$  and right continuous at the point zero.

(A2)  $\limsup_{|u| \rightarrow \infty} |K(u)u^5| < \infty$ .

**Theorem 1** (Fan and Gijbels, 1992) *Suppose conditions (A1) and (A2) hold. Then the conditional mean squared error of the estimator  $\widehat{m}_{LL}$  at the boundary point  $x = ch$  is given by*

$$\left\{ \frac{1}{4} [m''(0+) \alpha_K(c)]^2 h^4 + \frac{\beta_K(c) \sigma^2(0+)}{nh f_X(0+)} \right\} (1 + o_p(1)),$$

where

$$\alpha_K(c) = \frac{s_{2,c}^2 - s_{1,c}s_{3,c}}{s_{2,c}s_{0,c} - s_{1,c}^2}, \beta_K(c) = \frac{\int_{-c}^{\infty} (s_{2,c} - us_{1,c})^2 K(u) du}{(s_{2,c}s_{0,c} - s_{1,c}^2)^2},$$

with  $s_{l,c} = \int_{-c}^{\infty} u^l K(u) du, l = 0, 1, 2, 3$ .

Thus this estimator maintains the same asymptotic mean squared error rate everywhere, including the boundaries. Fan (1992) also shows that it attains the minimum risk within a certain class of estimators at interior points. An interesting question is whether it achieves some optimal risk even at the boundaries. For the left boundary at  $x = 0$ , this turns out to give a whole family of different minimax problems indexed by  $c \in [0, 1]$  in the representation  $x = ch$ . The most important of these, and also the simplest to analyze, is  $x = 0$ , so only that case is considered here.

### 3 Linear Regression Smoothers

Kernel estimators and  $\widehat{m}_{LL}(x)$ , and most other estimators are weighted averages of the responses. Such estimators are called linear smoothers in the literature.

**Definition 1** *A linear smoother has the form*

$$\widehat{m}_L(x) = \sum_{i=1}^n W_i(X_1, \dots, X_n) Y_i.$$

This class of smoothers also includes the popular smoothing spline and the orthogonal series estimators. The linear smoother  $\widehat{m}_L(x)$  has conditional risk

$$\begin{aligned} R(m(x), \widehat{m}_L(x)) &\equiv E \left[ (\widehat{m}_L(x) - m(x))^2 \mid X_1, \dots, X_n \right] \\ &= \left[ \sum_{i=1}^n W_i m(X_i) - m(x) \right]^2 + \sum_{i=1}^n W_i^2 \sigma^2(X_i) \\ &\geq \frac{m^2(x)}{1 + \sum_{i=1}^n m^2(X_i) / \sigma^2(X_i)}. \end{aligned} \quad (2)$$

The above inequality is validated by the following lemma from Fan (1992).

**Lemma 1** *Let  $a = (a_1, \dots, a_n)'$  and  $w = (w_1, \dots, w_n)'$  be  $n$ -dimensional real vectors. Then,*

$$\min_w \left[ (w'a - b)^2 + \sum_{i=1}^n c_i w_i^2 \right] = \frac{b^2}{1 + \sum_{i=1}^n a_i^2 / c_i}, \quad (3)$$

*and the minimizer is  $w_i = \frac{b}{1 + \sum_{i=1}^n a_i^2 / c_i} a_i / c_i$ .*

Apparently the risk and hence the minimized risk depends on the regression function. Furthermore for even a nonsense estimator there always exists one regression function that makes it the most favorable. Thus we restrict to some class of regression functions and compare estimators by their worst

risk over the class. A class of joint densities which reflects the idea of “ $m$  is twice right differentiable at  $x = 0$ ” is

$$C_2 = \left\{ f(\cdot, \cdot) : |m(y) - m(0) - m'(0)y| \leq \frac{C}{2}y^2, f \text{ and } \sigma \text{ satisfy condition (A1)} \right\}.$$

Define the minimax risk of linear smoothers for the class  $C_2$  as

$$R_{0,L}(n, C_2) = \inf_{\widehat{m}_L} \sup_{\text{linear } f \in C_2} E \left( [\widehat{m}_L(0) - m(0)]^2 | X_1, \dots, X_n \right).$$

Then the best linear smoother for the left end point is the one that achieves this minimax risk. Proof for the following theorem is given in the appendix.

**Theorem 2** *Assume that  $\sigma(\cdot)$  is bounded away from  $\infty$ . Then*

$$R_{0,L}(n, C_2) = 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{4/5} (1 + o_p(1)), \quad (4)$$

and the best linear smoother is given by  $\widehat{m}_{LL}(0)$  with kernel  $(1 - u)I_{[0,1]}(u)$  and  $h = \left( \frac{480\sigma^2(0)}{C^2nf_X(0)} \right)^{1/5}$ .

**Remark 3.1.** If the kernel function is symmetric about zero then there are on average only half as many observations as at the interior used in estimating  $m(0)$ . In order for the bias to be comparable, the weights of the responses are doubled. Hence the variances at the boundary are about four times larger. This reflects in the best possible risks; the constant multiplier in (4) is exactly four times of that for (4.3) in Fan (1992).

**Remark 3.2.** A similar argument leads to a best linear smoother for estimating  $m(1)$  which is  $\widehat{m}_{LL}(1)$  with kernel  $(1 + u)I_{[-1,0]}(u)$  and  $h = \left( \frac{480\sigma^2(1)}{C^2nf_X(1)} \right)^{1/5}$ .

## 4 Local Linear Fit for Density Estimation

Nonparametric smoothing techniques are widely used in estimation of density functions. Boundary effects occur in this application as well. We apply the technique of local linear fitting to density estimation, by regressing on bin

counts, with the intention of matching its excellent performance in regression. Suppose  $X_1, \dots, X_n$  is a random sample from a population with density function  $f$  which is supported in  $[0, 1]$ . Let  $\{t_1, \dots, t_g\}$  be a set of bin centers in  $[0, 1]$ , that can be thought of as fixed design points. Then the count of observations nearby each bin center provides information about the density there. More precisely, take  $t_j = (j - \frac{1}{2})b$ ,  $j = 1, \dots, g$ , where  $b > 0$  is the bin width. Define the bin count around  $t_j$  as  $c_j = \sum_{i=1}^n I_{[t_j - b/2, t_j + b/2]}(X_i)$ , for  $j = 1, \dots, g$ . The law of large numbers states that  $n^{-1}b^{-1}c_j$  is approximately  $b^{-1} \int_{t_j - b/2}^{t_j + b/2} f(u) du \approx f(t_j)$ . Thus it is reasonable to estimate  $f(x)$  by a local linear fit to the bin counts:

$$\min \sum_{j=1}^g \left( n^{-1}b^{-1}c_j - a - b(t_j - x) \right)^2 K \left( \frac{t_j - x}{h} \right).$$

This results in the estimator

$$\hat{f}_L(x) = n^{-1}b^{-1} \sum_{j=1}^g \omega_j c_j / \sum_{j=1}^g \omega_j,$$

with

$$\omega_j = [S_2 - (t_j - x)S_1] K \left( \frac{t_j - x}{h} \right),$$

where

$$S_l = \sum_{j=1}^g K \left( \frac{t_j - x}{h} \right) (t_j - x)^l, l = 1, 2.$$

Convenient assumptions for asymptotic analysis are:

(A3)  $K^{(l)}$  is bounded and absolutely integrable with finite second moment on its support, for  $l = 0, 1, 2$ .

(A4)  $f$  and its first two derivatives are bounded.

Under conditions (A3) and (A4), for  $x = ch$ ,  $c \geq 0$ , we can show

$$MSE(\hat{f}_L(x)) = \left\{ \frac{h^4}{4} (f''(0+) \alpha_K(c))^2 + \frac{1}{nh} f(0+) \beta_K \right\} (1 + o(1)), \quad (5)$$

as  $n \rightarrow \infty$ ,  $nh \rightarrow \infty$ , and  $b = o(h)$ . It may be worthwhile to mention that when  $x$  is an interior point the mean squared error of this estimator is asymptotically the same as that of the ordinary kernel density estimator with kernel  $K$ .

Next we discuss some optimal properties of this estimator at the left endpoint 0, analogous result for the right edge 1 follows easily from a symmetric augmentation of the method. Note that  $\hat{f}_L(x)$  is a weighted average of the bin counts which are linear functions of the data. Hence the estimator has the form  $\hat{f}_L(x) = \frac{1}{n} \sum_{i=1}^n \psi(X_i, x)$ . The conventional kernel density estimator is of the same form. Estimators of this type are called linear and may be expressed as

$$\hat{f}_\psi(x) = \int \psi(t, x) d\hat{F}_n(t),$$

where  $\hat{F}_n$  is the empirical distribution function of the sample. A direct analog of the “smooth class”  $C_2$  in section 3 is the class of density functions supported in  $[0, 1]$

$$C_{M,2} = \{f : |f| \leq M, |f(x) - f(0) - f'(0)x| \leq \frac{C}{2}x^2 \},$$

where  $C$  and  $M$  are some fixed positive constants. This class of functions is also considered in Donoho and Liu (1991). Denote the minimax mean squared error over linear estimators of  $f(0)$  for this class as

$$R_{0,L}(n, C_{M,2}) = \inf_{\psi} \sup_{f \in C_{M,2}} E \left( \hat{f}_\psi(0) - f(0) \right)^2.$$

The estimator  $\hat{f}_\psi(x)$  has mean squared error

$$\begin{aligned} E \left( \hat{f}_\psi(0) - f(0) \right)^2 &= \left( \int \psi(t, 0) f(t) dt - f(0) \right)^2 + \frac{1}{n} \text{Var}(\psi(X_1, 0)) \\ &\geq \left( \int \psi(t, 0) f(t) dt - f(0) \right)^2 + \frac{1}{n} \int \psi^2(t, 0) f(t) dt. \end{aligned}$$

For any  $f_1, f_2 \in C_{M,2}$

$$\sup_{f \in C_{M,2}} E \left( \hat{f}_\psi(0) - f(0) \right)^2 \geq \frac{1}{4} \left\{ \left[ \int \psi(f_1 - f_2) - (f_1(0) - f_2(0)) \right]^2 + \frac{2}{n} \int \psi^2(f_1 + f_2) \right\}.$$

Minimizing this over all  $\psi$  yields

$$\inf_{\psi} \sup_{f \in C_{M,2}} E \left( \hat{f}_\psi(0) - f(0) \right)^2 \geq \frac{1}{4} \frac{(f_1(0) - f_2(0))^2}{1 + \frac{n}{2} \int \frac{(f_1 - f_2)^2}{f_1 + f_2}}. \quad (6)$$

The above minimization follows from

**Lemma 2** Suppose  $\xi \geq 0$  and  $\eta$  are functions on  $R^1$ , and  $b$  is a constant then

$$\min_{\psi} \left( \int \psi \eta - b \right)^2 + \int \psi^2 \xi = \frac{b^2}{1 + \int \frac{\eta^2}{\xi}},$$

and the minimum is attained when  $\psi = \frac{b}{1 + \int \frac{\eta^2}{\xi}} \frac{\eta}{\xi}$ .

**Theorem 3** With the definitions of  $C_{M,2}$  and  $R_{0,L}(n, C_{M,2})$  given in the above,

$$R_{0,L}(n, C_{M,2}) = 3 \cdot 15^{-1/5} \left( \frac{\sqrt{CM}}{n} \right)^{4/5} (1 + o(1)),$$

and the best linear estimator is the local linear fit estimator with kernel weight function  $(1-u)I_{[0,1]}(u)$  and  $h = \left( \frac{480M}{C^2 n} \right)^{1/5}$ .

Proof of this theorem is given in the appendix.

## 5 Relative Efficiency

Now that the best possible risk for linear estimators is known, it can be used as a base line for measuring performances of such estimators. We will discuss this for the regression setup since the conclusions for density estimation are the same as those for uniform design regression setting. Define the efficiency of a linear estimator  $\hat{T}(0)$  of  $m(0)$  as

$$\text{Eff}(\hat{T}(0)) = \lim_{n \rightarrow \infty} \left( \frac{R_{0,L}(n, C_2)}{\sup_{m \in C_2} E(\hat{T}(0) - m(0))^2} \right)^{5/4}.$$

The power 5/4 puts efficiency on the traditional and interpretable “sample size scale” since both numerator and denominator have asymptotic rate of convergence  $n^{-4/5}$ .

In the previous sections we have shown that the local linear estimator is 100% efficient with the kernel weight function  $K_0 = (1-u)I_{[0,1]}(u)$ . The Gaussian density function is often used in kernel smoothing methods since it make the estimators visually more pleasant; i.e. smooth and without

undesired angles. The local linear estimator with Gaussian kernel is easily shown to have efficiency

$$\left( \frac{3 \cdot 15^{-1/5}}{\frac{5}{4} \left( \frac{\pi-4}{\pi-2} \right)^{2/5} \left( \frac{\sqrt{\pi}(\pi+1-\sqrt{8})}{(\pi-2)^2} \right)^{4/5}} \right)^{5/4} \approx 0.9802$$

at edge points. Hence there is very small lost of efficiency while gaining the visual benefit by using the Gaussian kernel in local linear fitting.

Another important boundary corrected estimator whose efficiency considered here is Rice's (1984) modification. This method linearly combines two conventional kernel estimators with different bandwidths to achieve the same bias order, i.e.  $h^2$ , as in the interior. It is a Nadaraya-Watson kernel estimator itself and as noted in the paper Fan (1992) its efficiency is 0 when the design points are non-uniform. The reason is that its bias contains an extra term involving  $m'(0)$  which can be arbitrarily large. Hence consider the case of uniform design. For Rice's estimator, it is not known which kernel function combined with what bandwidth ratio, i.e. the ratio of the two bandwidths in the combination, will give the best performance. Here the relative efficiencies the well known Epanechnikov kernel and the popular Gaussian kernel are considered. The best bandwidth ratio for the Gaussian kernel at  $x = 0$  is shown by straightforward calculation to be 1 which results in the effective kernel

$$\varphi_1(u) = 2(2 - u^2)\varphi(u),$$

and this estimator has efficiency

$$\left( \frac{3 \cdot 15^{-1/5}}{\frac{5}{4} \left( \frac{11}{4\sqrt{\pi}} \right)^{4/5}} \right)^{5/4} \approx 0.9783.$$

Using the Epanechnikov kernel, the best bandwidth ratio can be shown to be  $1 + \sqrt{5/2}$  with efficiency

$$\left( \frac{3 \cdot 15^{-1/5}}{\frac{5}{4} \left( \frac{18(8\sqrt{10}+24)^2}{125(2+\sqrt{10})^3} \right)^{4/5}} \right)^{5/4} \approx 0.9517.$$

From these numbers we can conclude that the Rice modification is also an excellent method of boundary adjustment for regression with uniform design or density estimation.

## 6 Derivative Estimation

There are situations where estimating derivatives of curves is the interest. An important case is for pilot estimators in plug-in methods of bandwidth selection. Extending the method of local linear fitting to higher polynomial fitting provides a solution to this problem. Fan, Gasser, Gijbels, Brockmann, and Engels (1992) study its properties and show that the local polynomial fit is minimax efficient among all linear estimators for estimating derivatives at an interior point. We demonstrate minimax efficiency at boundaries in this section. The formulation will be in the regression setting and similar results for density derivatives are obtained with little extra effort.

The local polynomial derivative estimator may be motivated as follows. Suppose the curve under investigation is smooth up to  $(p + 1)$ -th derivative. Then via Taylor's polynomial approximation

$$m(z) \approx \sum_{j=1}^p \frac{m^{(j)}(x)}{j!} (z - x)^j, \quad (7)$$

within a neighborhood of  $x$ . This suggests a local polynomial regression

$$\sum_{i=1}^n \left( Y_i - \sum_{j=1}^p b_j (X_i - x)^j \right)^2 K \left( \frac{X_i - x}{h} \right). \quad (8)$$

Let  $\hat{b}_j(x)$ ,  $j = 0, 1, \dots, p$  be the solution of the least squares problem. Then from (7) it is appropriate to estimate  $m^{(\nu)}(x)$  by  $\widehat{m}_\nu(x) = \nu! \hat{b}_\nu(x)$ . It is shown in Fan et al (1993) that

$$\hat{b}_\nu(x) = \sum_{i=1}^n W_\nu^n \left( \frac{X_i - x}{h} \right) Y_i, \quad (9)$$

where

$$W_\nu^n(t) = e_\nu^T S_n^{-1} (1, ht, \dots, h^p t^p)^T K(t) \quad (10)$$

with

$$e_\nu = (0, \dots, 0, 1, 0, \dots, 0)^T,$$

$$S_{n,j}(x) = \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (X_i - x)^j, j = 0, 1, \dots, 2p, \text{ and } S_n = (S_{n,i+j-2}).$$

Here  $W_\nu^n$  is the weight function for estimating the  $\nu^{\text{th}}$  derivative of the regression function generated from the local polynomial fitting. Expression (9) reveals that  $\hat{b}_\nu(x)$  is very much like a kernel estimator except that the "kernel"  $W_\nu^n$  depends on the sample. However,

$$S_{n,j} = ES_{n,j}(1 + o_p(1)) = nh^{j+1}f_X(x)s_j(1 + o_p(1)),$$

where  $s_j = \int_{-\infty}^{\infty} t^j K(t) dt$  for an interior point  $x$ . Hence with  $S = (s_{i+j})_{0 \leq i, j \leq p}$  and from (10),

$$W_\nu^n(t) \approx \frac{1}{nh^{\nu+1}f_X(x)} e_\nu^T S^{-1}(1, t, \dots, t^p)^T K(t).$$

Therefore, by (9),

$$\hat{b}_\nu(x) \approx \frac{1}{nh^{\nu+1}f_X(x)} \sum_{i=1}^n K_\nu^*\left(\frac{X_i - x}{h}\right) Y_i, \quad (11)$$

where

$$K_{\nu,p+1}^*(t) = e_\nu^T S^{-1}(1, t, \dots, t^p)^T K(t) = \sum_{l=0}^p S^{\nu l} t^l K(t), \quad (12)$$

with  $S^{-1} \equiv (S^{jl})_{0 \leq j \leq p, 0 \leq l \leq p}$ . The equivalent kernel  $K_{\nu,p+1}^*$  satisfies the higher order kernel conditions

$$\int t^q K_{\nu,p+1}^*(t) dt = \delta_{\nu,q}, \text{ for } 0 \leq \nu, q \leq p. \quad (13)$$

When  $x$  is a boundary point the equivalent kernel differs from  $K_\nu^*$  only in the matrix  $S$ . Suppose  $x = ch$ ,  $c \geq 0$ , then  $S_{n,j} = ES_{n,j}(1 + o_p(1)) = nh^{j+1}f_X(0)s_{j,c}(1 + o_p(1))$ , where  $s_{j,c} = \int_{-c}^{\infty} u^j K(u) du$ . Then the boundary equivalent kernel is

$$K_{\nu,c}^*(t) = e_\nu^T S_c^{-1}(1, t, \dots, t^p)^T K(t), \quad (14)$$

where  $S_c = (s_{i+j-2,c})_{1 \leq i,j \leq p+1}$ . The conditional MSE at the boundary point  $x = ch$ , for  $n \rightarrow \infty, h \rightarrow 0, nh \rightarrow \infty$ , is

$$E \left\{ \left( \hat{b}_\nu(x) - \frac{m^{(\nu)}(x)}{\nu!} \right)^2 \middle| X_1, \dots, X_n \right\} \\ \stackrel{p}{\approx} \left\{ \left( \int_{-c}^{\infty} t^{p+1} K_{\nu,c}^*(t) dt \right)^2 \left( \frac{m^{(p+1)}(0)}{(p+1)!} h^{p+1-\nu} \right)^2 + \frac{\sigma^2(0)}{nh^{2\nu+1} f_X(0)} \int_{-c}^{\infty} K_{\nu,c}^{*2}(t) dt \right\} \quad (15)$$

Here  $\stackrel{p}{\approx}$  means asymptotically equivalent in probability sense. The proof is referred to Fan, Gasser, Gijbels, Brockmann, and Engels (1992). The conventional kernel method uses kernels for derivative estimation which are very hard to interpret. This problem becomes even worse at boundary regions. Comparatively, the local polynomial fitting naturally provides interpretable and effective estimators, e.g. in the sense of rate of convergence, and requires no additional complicated boundary modifications.

**Remark 6.1.** The local polynomial fitting technique is applicable to estimating derivatives of a density function. Notation is the same as in section 4, simply replace the  $\{Y_i\}_{1 \leq i \leq n}$  and  $\{X_i\}_{1 \leq i \leq n}$  by  $\{n^{-1}b^{-1}c_i\}_{1 \leq i \leq g}$  and  $\{t_i\}_{1 \leq i \leq g}$ , respectively. And, e. g. at the boundaries, the resulting estimator has mean squared error asymptotically equal to the quantity in (15) with  $m^{(p+1)}(0)$  replaced by  $f^{(p+1)}(0)$  and  $\frac{\sigma^2(0)}{f_X(0)}$  replaced by  $f(0)$ .

## 7 Minimax Efficiency of Derivative Estimators

Now some minimax theory for general derivative estimation at endpoints is developed. We shall focus on optimizing over the class of linear estimators is developed. We discuss this for the right endpoint  $x = 0$ . Suppose  $f_X$  and  $\sigma$  are continuous at 0, with  $f_X(0) > 0$  and  $\sigma(0) < \infty$ , let

$$C_{p+1} = \left\{ m : m \text{ has support } [0, \infty), \left| m(z) - \sum_{j=0}^p \frac{m^{(j)}(0)}{j!} z^j \right| \leq C \frac{|z|^{p+1}}{(p+1)!} \right\}. \quad (16)$$

Define

$$R_{0,L}(\nu) = \inf_{\widehat{T}_\nu} \sup_{\text{linear } m \in C_{p+1}} E \left\{ \left( \widehat{T}_\nu - m^{(\nu)}(0) \right)^2 \middle| X_1, \dots, X_n \right\}$$

as the linear minimax risk for estimating  $m^{(\nu)}(0)$ . A quantity closely related to the  $R_{0,L}(\nu)$  is the modulus of continuity

$$\omega_{p+1,\nu}(\epsilon) = \sup \left\{ \left| m_1^{(\nu)}(0) - m_0^{(\nu)}(0) \right| : m_0, m_1 \in C_{p+1}, \|m_1 - m_0\| = \epsilon \right\}, \quad (17)$$

see Donoho and Liu (1991) and Fan (1992a). Let

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

and

$$\begin{aligned} \theta_{\nu,p} = & \left( \frac{2p+3}{2\nu+1} \right) \left( \frac{(p+\nu+2)!}{(p-\nu+1)!\nu!} \right)^2 \left( \frac{r}{2(p+\nu+2)} \right)^r \left( \frac{(p+1)!C}{(2p+3)!} \right)^{2s} \\ & \times \left( \frac{\sigma^2(0)}{nf_X(0)} \right)^r. \end{aligned} \quad (18)$$

**Theorem 4** *The modulus of continuity is given by*

$$\begin{aligned} \omega_{p+1,\nu}(\epsilon) = & \epsilon^r \left( \frac{(p+\nu+2)!}{\nu!(p-\nu+1)!(2\nu+1)} \right) \left( \frac{2C(p+1)!}{(2p+2)!} \right) \\ & \times \left( \frac{(2\nu+1)(2p+3)}{2(p+\nu+2)} \right)^{r/2} (1 + o(1)). \end{aligned}$$

**Theorem 5** *The linear minimax risk for estimating the  $\nu^{\text{th}}$  derivative of the regression function at its right endpoint is*

$$R_{0,L}(\nu) = \theta_{\nu,p}(1 + o_p(1)).$$

**Theorem 6** *Let  $\widehat{m}_\nu(0)$  be the estimator resulting from a local polynomial fit of order  $p$  with the kernel function  $K_0(u) = (1-u)I_{[0,1]}(u)$ . Then it is the best linear estimator for  $m^{(\nu)}(0)$  in the sense that*

$$\frac{R_{0,L}(\nu)}{\sup_{m \in C_{p+1}} E \left\{ \left( \widehat{m}_\nu(0) - m^{(\nu)}(0) \right)^2 \middle| X_1, \dots, X_n \right\}} \xrightarrow{p} 1.$$

Moreover, its equivalent kernel is

$$K_{\nu, p+1}^{opt}(t) = \sum_{j=0}^{p+1} \lambda_j t^j I_{[0,1]}(t),$$

where

$$\lambda_j = \frac{(-1)^{j+\nu} (p+j+1)! (p+\nu+2)!}{j! 2^\nu! (p-\nu)! (p-j+1)! (j+\nu+1)!}, j = 0, 1, \dots, p+1.$$

**Remark 7.1.** Suppose the condition  $|m| \leq M$  for some positive constant is added to the definition of the class  $C_{p+1}$  in (16). Then it can be shown that the analogous linear minimax MSE for estimating the  $\nu^{th}$  derivative of a density function  $f \in C_{p+1}$  at boundaries is asymptotically

$$\left(\frac{2p+3}{2\nu+1}\right) \left(\frac{(p+\nu+2)!}{(p-\nu+1)! \nu!}\right)^2 \left(\frac{rM}{2(p+\nu+2)n}\right)^r \left(\frac{(p+1)!C}{(2p+3)!}\right)^{2s}.$$

Furthermore, the estimator constructed from a local polynomial fit of order  $p$  with kernel  $K_0 = (1-u)I_{[0,1]}(u)$  has mean squared error asymptotically equal to the linear minimax MSE.

**Remark 7.2.** Although in section 5 we noticed that the Rice (1984) modification is highly efficient in estimating the functions, we don't know whether a similar implementation for derivative estimation will retain this property or not. But, at least we can say that it requires a lot of effort to do so and it has to be done for each  $\nu$  separately. One merit of the local polynomial fit is that the derivative estimators are produced easily from its one-time-for-all- $\nu$  least squares fitting.

#### APPENDIX.

I. *Proof of Theorem 2.* Since  $X_1, \dots, X_n$  are i.i.d.,

$$\sum_{i=1}^n m^2(X_i)/\sigma^2(X_i) = nEm^2(X_1)/\sigma^2(X_1) + O_p\left(\sqrt{nEm^4(X_1)/\sigma^4(X_1)}\right).$$

From this and (2) we have

$$R(m(0), \widehat{m}_L(0)) \geq \frac{m^2(0)}{1 + nEm^2(X_1)/\sigma^2(X_1) + O_p\left(\sqrt{nEm^4(X_1)/\sigma^4(X_1)}\right)}. \quad (19)$$

Take  $m_0(y) = \frac{b_n}{4} \left[ 1 - \frac{3\sqrt{C}y}{b_n} + \frac{2Cy^2}{b_n^2} \right] I_{[0,1]} \left( \frac{\sqrt{C}y}{b_n} \right)$ , where  $b_n = \left( \frac{480\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{1/5}$ . Note that  $m_0 \in C_2$  and  $b_n$  maximizes (21) below. Now,

$$\begin{aligned}
Em_0^2(X_1)/\sigma^2(X_1) &= \frac{b_n^4}{16} \int \left[ 1 - \frac{3\sqrt{C}y}{b_n} + \frac{2Cy^2}{b_n^2} \right]^2 I_{[0,1]} \left( \frac{\sqrt{C}y}{b_n} \right) \frac{f_X(y)}{\sigma^2(y)} dy \\
&= \frac{b_n^5}{16} \int \left[ 1 - 3\sqrt{c}z + 2Cz^2 \right]^2 I_{[0,1]}(\sqrt{C}z) \frac{f_X(b_n z)}{\sigma^2(b_n z)} dz \\
&= \frac{b_n^5 f_X(0)}{16\sigma^2(0)} \int \left[ 1 - 3\sqrt{c}z + 2Cz^2 \right]^2 I_{[0,1]}(\sqrt{C}z) dz (1 + o_p(1)) \\
&= \frac{b_n^5 f_X(0)}{120\sqrt{C}\sigma^2(0)} (1 + o_p(1)). \tag{20}
\end{aligned}$$

From (19), (20), and the fact that  $Em^4(X_1)/\sigma^4(X_1) = O(b_n^8)$ ,

$$\begin{aligned}
R(m_0(0), \widehat{m}_L(0)) &\geq \frac{b_n^4/16}{1 + \frac{nb_n^5 f_X(0)}{120\sqrt{C}\sigma^2(0)} (1 + o_p(1))} \tag{21} \\
&= 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{4/5} (1 + o_p(1)).
\end{aligned}$$

Hence

$$R_{0,L}(n, C_2) \geq 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{4/5} (1 + o_p(1)). \tag{22}$$

On the other hand, let  $\widehat{m}_0$  be the local linear smoother with kernel  $K_0(u) = (1-u)I_{[0,1]}(u)$ ,

$$\begin{aligned}
R(m(0), \widehat{m}_0(0)) &\leq \frac{h^4}{4} \alpha_{K_0}^2(0) C^2 + \frac{1}{nh} \beta_{K_0}(0) \frac{\sigma^2(0)}{f_X(0)} + o_p \left( h^4 + \frac{1}{nh} \right) \\
&= 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{4/5} (1 + o_p(1)).
\end{aligned}$$

The last equality holds with  $h = \left( \frac{\sigma^2(0)\beta_{K_0}(0)}{nC^2\alpha_{K_0}^2(0)f_X(0)} \right)^{1/5}$ . Therefore,

$$R_{0,L}(n, C_2) \leq \sup_{C_2} R(m(0), \widehat{m}_0(0)) \leq 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}\sigma^2(0)}{nf_X(0)} \right)^{4/5} (1 + o_p(1)). \tag{23}$$

The result follows from (22) and (23).

## II. Proof of lemma 2.

$$\left(\int \psi \eta - b\right)^2 + \int \psi^2 \xi \geq \min_t \left( \min_{\int \psi \eta = t} (t - b)^2 + \int \psi^2 \xi \right).$$

But under the constraint  $\int \psi \eta = t$ ,

$$\left(\int \psi^2 \xi\right) \left(\int \frac{\eta^2}{\xi^2} \cdot \xi\right) \geq \left(\int \psi \cdot \frac{\eta}{\xi} \cdot \xi\right)^2 = t^2,$$

where equality holds when  $\psi = \frac{t}{\int \frac{\eta^2}{\xi}} \frac{\eta}{\xi}$ . Hence,

$$\left(\int \psi \eta - b\right)^2 + \int \psi^2 \xi \geq \min_t \left( (t - b)^2 + \frac{t^2}{\int \frac{\eta^2}{\xi}} \right) = \frac{b^2}{1 + \int \frac{\eta^2}{\xi}},$$

where the minimum is attained by  $t = \frac{b}{1 + \left(\int \frac{\eta^2}{\xi}\right)^{-1}}$ .

## III. Proof of theorem 3.

Let  $f_1(x) = g_0(x) + g_n(x) - c_n$ ,  $f_2(x) = g_0(x) - g_n(x) + c_n$ , where

$$g_0(x) = \left[ -\frac{(M - \delta)^2}{2} x + (M - \delta) \right] I_{[0, \frac{2}{M - \delta}]}(x),$$

$$g_n(x) = \frac{b_n^2}{2} \left( \frac{1}{2} - \frac{3\sqrt{C}x}{2b_n} + \frac{Cx^2}{b_n^2} \right) I_{[0, \frac{b_n}{\sqrt{C}}]}(x),$$

with  $b_n = \left(\frac{480\sqrt{CM}}{n}\right)^{1/5}$ ,  $c_n = \int g_n = \frac{b_n^3}{24\sqrt{C}}$ , and

$$\max \left( \frac{b_n^2}{4}, \frac{b_n^2}{32} + c_n \right) \leq \delta \leq \min \left( \frac{M^2 b_n}{2\sqrt{C}}, M - \frac{M^2 b_n}{\sqrt{C}} \right).$$

Then  $f_1, f_2 \in C_{2,M}$  and from (6) we have

$$\inf_{\psi} \sup_{f \in C_{M,2}} \mathbb{E} \left( \widehat{f}_{\psi}(0) - f(0) \right)^2 \geq \frac{1}{4} \frac{4g_n(0)^2}{1 + \frac{n}{2} \int \frac{4g_n^2}{2g_0}}$$

$$\begin{aligned}
&= \frac{g_n(0)^2}{1 + \frac{n}{g_0(0)} \int g_n^2} (1 + o(1)) \\
&= \frac{b_n^4/16}{1 + \frac{n}{M} \frac{b_n^5}{120\sqrt{C}}} (1 + o(1)) \\
&= 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}M}{n} \right)^{4/5} (1 + o(1)). \tag{24}
\end{aligned}$$

But suppose  $\widehat{f}_{\psi_0}$  is the local linear fit with kernel  $K_0(u) = (1 - u)I_{[0,1]}(u)$ . Then by (5)

$$\begin{aligned}
\inf_{\psi} \sup_{f \in \mathcal{C}_{M,2}} E \left( \widehat{f}_{\psi}(0) - f(0) \right)^2 &\leq \sup_{f \in \mathcal{C}_{M,2}} E \left( \widehat{f}_{\psi_0}(0) - f(0) \right)^2 \\
&\leq \frac{h^4}{4} \alpha_{K_0}^2(0) C^2 + \frac{1}{nh} \beta_{K_0}(0) M + o \left( h^4 + \frac{1}{nh} \right) \\
&= 3 \cdot 15^{-1/5} \left( \frac{\sqrt{C}M}{n} \right)^{4/5} (1 + o(1)). \tag{25}
\end{aligned}$$

The last equality is valid with  $h = \left( \frac{M\beta_{K_0}(0)}{C^2\alpha_{K_0}^2(0)n} \right)^{1/5}$ . Combining (24) and (25) finishes the proof of the theorem.

#### IV. Proof of theorems 4 - 6.

Denote  $K_{\nu,p+1}^{opt}$  to be the equivalent kernel of  $\widehat{m}_{\nu}(0) = \nu! \widehat{b}_{\nu}(0)$ , given by (14), with  $K(u) = K_0(u) = (1 - u)I_{[0,1]}(u)$ . We prove theorems 4 - 6 based on the norm  $\|K_{\nu,p+1}^{opt}\|$  and the  $(p+1)^{th}$  moment of  $K_{\nu,p+1}^{opt}$ . The calculation of these quantities and the function  $K_{\nu,p+1}^{opt}$  is very technically involved and appears in subsection V. First we construct an upper bound for the linear minimax risk  $R_{0,L}(\nu)$ ,

$$\begin{aligned}
&\sup_{m \in \mathcal{C}_{p+1}} E \left\{ \left( \widehat{m}_{\nu}(0) - m^{(\nu)}(0) \right)^2 \middle| X_1, \dots, X_n \right\} \\
&\leq \left( \frac{C}{(p+1)!} \int_0^1 t^{p+1} K_{\nu,p+1}^{opt}(t) dt \right)^2 h^{2(p+1-\nu)} + \int_0^1 K_{\nu,p+1}^{opt 2}(t) dt \frac{\sigma^2(0)}{nf_X(0)} h^{-(2\nu+1)} \\
&\equiv A_1 h^{2(p+1-\nu)} + A_2 h^{-(2\nu+1)}.
\end{aligned}$$

Take  $h = \left( \frac{(2\nu+1)A_2}{2(p+1-\nu)A_1} \right)^{1/(2p+3)} n^{-1/(2p+3)}$  which minimizes the above quantity, then

$$\begin{aligned} & \sup_{m \in C_{p+1}} E \left\{ \left( \widehat{m}_\nu(0) - m^{(\nu)}(0) \right)^2 \middle| X_1, \dots, X_n \right\} \\ & \leq A_1^s A_2^r (2p+3)(2\nu+1)^{-s} [2(p+1-\nu)]^{-r} (1 + o_p(1)) \\ & = r^{-r} s^{-s} \left( \frac{C}{(p+1)!} \right)^{2s} \left( \frac{\sigma^2(0)}{n f_X(0)} \right)^r \left( \int_0^1 t^{p+1} K_{\nu, p+1}^{opt}(t) dt \right)^{2s} \|K_{\nu, p+1}^{opt}\|^{2r} (1 + o_p(1)). \end{aligned}$$

From (33) and (35) the above expression equals

$$\begin{aligned} & \left( \frac{2(p+1-\nu)}{2p+3} \right)^{-r} \left( \frac{2\nu+1}{2p+3} \right)^{-s} \left( \frac{C}{(p+1)!} \right)^{2s} \left( \frac{\sigma^2(0)}{n f_X(0)} \right)^r \\ & \times \left( \frac{(p+\nu+2)!(p+1)!^2}{\nu!(2p+3)!(p-\nu+1)!} \right)^{2s} \left( \frac{2(p+\nu+2)(p+\nu+1)!^2}{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2} \right)^r \\ & = \theta_{\nu, p}, \end{aligned} \tag{26}$$

as defined in (18). Hence

$$R_{0,L}(\nu) \leq \theta_{\nu, p}(1 + o_p(1)).$$

To establish a lower bound for  $R_{0,L}(\nu)$ , if  $f \in C_{p+1}$ , take

$$m_1(x) = \delta^{p+1} f(x/\delta), m_0(x) = -m_1(x),$$

where  $\delta = \left( \frac{\epsilon^2}{4\|f\|^2} \right)^{\frac{1}{2p+3}}$ . Obviously  $m_0, m_1 \in C_{p+1}$ , and

$$\|m_1 - m_0\|^2 = 4\delta^{2p+3} \|f\|^2 = \epsilon^2.$$

Therefore the modulus of continuity defined in (17) satisfies

$$\omega_{0,\nu}(\epsilon) \geq |m_1^{(\nu)}(0) - m_0^{(\nu)}(0)| = 2 |f^{(\nu)}(0)| \left( \frac{\epsilon^2}{4\|f\|^2} \right)^{\frac{p+1-\nu}{2p+3}}. \tag{27}$$

Now, let

$$a = \left( \frac{C}{(p+1)! |\lambda_{p+1}|} \right)^{1/(p+1)},$$

and

$$f(x) = \begin{cases} K_{\nu,p+1}^{opt}(ax) & , \text{ if } 0 \leq x \leq 1. \\ 0 & , \text{ otherwise.} \end{cases}$$

Then

$$\|f\|^2 = \frac{\|K_{\nu,p+1}^{opt}\|^2}{a}, f^{(\nu)}(0) = a^\nu \nu! \lambda_\nu,$$

and (27) becomes

$$\begin{aligned} \omega_{0,\nu}(\epsilon) &\geq 2\nu! |\lambda_\nu| \left( \frac{C}{(p+1)! |\lambda_{p+1}|} \right)^s \|K_{\nu,p+1}^{opt}\|^{-r} \left( \frac{\epsilon^2}{4} \right)^{\tau/2} \\ &= 2\nu! \left( \frac{(p+\nu+1)!^2 (p+\nu+2)}{\nu!^3 (p-\nu)!^2 (p-\nu+1)(2\nu+1)} \right) \left( \frac{C(p+1)!^2 \nu! (p-\nu)! (p+\nu+2)}{(p+1)! (2p+2)! (p+\nu+1)!} \right)^s \\ &\quad \times \left( \frac{(2\nu+1)(2p+3)}{2(p+\nu+2)} \right)^{\tau/2} \left( \frac{\nu! (p-\nu)!}{(p+\nu+1)!} \right)^\tau \left( \frac{\epsilon}{2} \right)^\tau \\ &= \epsilon^\tau \left( \frac{(p+\nu+2)!}{\nu! (p-\nu+1)! (2\nu+1)} \right) \left( \frac{2C(p+1)!}{(2p+2)!} \right) \left( \frac{(2\nu+1)(2p+3)}{2(p+\nu+2)} \right)^{\tau/2}. \end{aligned}$$

Applying Theorem 6 of Fan (1993), we have

$$\begin{aligned} R_{0,L}(\nu) &\geq r^\tau s^s \left[ \nu! \lambda_\nu \left( \frac{C}{(p+1)! |\lambda_{p+1}|} \right)^s \|K_{\nu,p+1}^{opt}\|^{-r} \left( \frac{\sigma^2(0)}{nf_X(0)} \right)^{\tau/2} \right]^2 \\ &= \left( \frac{C}{(p+1)!} \right)^{2s} \left( \frac{\sigma^2(0)}{nf_X(0)} \right)^\tau \frac{r^\tau s^s \nu!^2 \lambda_\nu^2}{\lambda_{p+1}^{2s} \|K_{\nu,p+1}^{opt}\|^{2r}}. \end{aligned}$$

Equations (34) and (35) give

$$\begin{aligned} R_{0,L}(\nu) &\geq r^\tau s^s \nu!^2 \left( \frac{C}{(p+1)!} \right)^{2s} \left( \frac{\sigma^2(0)}{nf_X(0)} \right)^\tau \left( \frac{\nu! (p-\nu)! (p+1)!^2}{(p+\nu+1)! (2p+2)!} \right)^{2s} \\ &\quad \times \left( \frac{(p+\nu+1)!^2 (p+\nu+2)}{\nu!^3 (p-\nu)!^2 (2\nu+1)(p-\nu+1)} \right)^2 \left( \frac{(2\nu+1)(2p+3) \nu!^2 (p-\nu)!^2}{2(p+\nu+2)(p+\nu+1)!^2} \right)^\tau \\ &= \theta_{r,\nu} \end{aligned}$$

given in (18).

In summary, ( i ) since the upper and lower bounds are the same, we prove theorem 4 and 5, ( ii ) the maximum risk of  $\widehat{m}_\nu(0)$  is given in (26) and the first part of theorem 6 follows immediately, the second part has been shown in (34). Thus we complete the proof.

V. Calculation of the function  $K_{\nu,p+1}^{opt}$ , its norm and  $(p+1)^{th}$  moment:

The famous Legendre polynomials are defined as

$$P_n(x) = \frac{d^n}{dx^n} ((1+x)(1-x))^n, \quad -1 \leq x \leq 1, \quad n = 0, 1, 2, \dots$$

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

$$Q_n(y) = \frac{d^n}{dy^n} (y(1-y))^n \equiv \sum_{j=0}^n q_{n,j} y^j \quad n = 0, 1, 2, \dots$$

Then

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

Explicitly,

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

So,

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

Let  $K_{\nu,p+1}^{opt}$  denote the equivalent kernel of  $\widehat{m}_\nu(0) = \nu! \widehat{b}_\nu(0)$ , given by (14), with  $K(u) = K_0(u) = (1-u)I_{[0,1]}(u)$ . Since  $K_{\nu,p+1}^{opt}$  is a polynomial of order  $(p+1)$ , we can write

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

The coefficients  $a_i$  can be determined by the moment properties in (13). Let  $\beta = \int_0^1 x^{p+1} K_{\nu, p+1}^{opt}(x) dx$ . Then

$$a_i \|Q_i\|^2 = \int_0^1 Q_i(x) K_{\nu, p+1}^{opt}(x) dx = \begin{cases} 0 & , \text{ if } 0 \leq i < \nu. \\ \nu! q_{i, \nu} & , \text{ if } \nu \leq i \leq p. \\ \nu! q_{p+1, \nu} + q_{p+1, p+1} \beta & , \text{ if } i = p+1. \end{cases} \quad (29)$$

Therefore, from (28) and (29),

$$\begin{aligned} \frac{1}{\nu!} K_{\nu, p+1}^{opt}(x) &= \sum_{i=0}^p q_{i, \nu} \frac{(2i+1)}{i!^2} Q_i(x) + \frac{(2p+3)}{(p+1)!^2} (q_{p+1, \nu} + \frac{q_{p+1, p+1}}{\nu!} \beta) Q_{p+1}(x) \\ &= \sum_{i=0}^p q_{i, \nu} \frac{(2i+1)}{i!^2} \sum_{j=0}^i q_{i, j} x^j + \frac{(2p+3)}{(p+1)!^2} (q_{p+1, \nu} + \frac{q_{p+1, p+1}}{\nu!} \beta) Q_{p+1}(x) \\ &= \sum_{j=0}^p \left( \sum_{i=j \vee \nu}^p q_{i, \nu} \frac{(2i+1)}{i!^2} q_{i, j} \right) x^j + \frac{(2p+3)}{(p+1)!^2} (q_{p+1, \nu} + \frac{q_{p+1, p+1}}{\nu!} \beta) Q_{p+1}(x). \quad (30) \end{aligned}$$

Here,

$$\sum_{i=j \vee \nu}^p q_{i, \nu} \frac{(2i+1)}{i!^2} q_{i, j} = \frac{(-1)^{j+\nu}}{j!^2 \nu!^2} \sum_{i=j \vee \nu}^p \frac{(i+\nu)!(2i+1)(j+i)!}{(i-\nu)!(i-j)!}$$

(Note:  $(2i+1) = \{(i+j+1)(i+\nu+1) - (i-j)(i-\nu)\} / (j+\nu+1)$ .)

$$\begin{aligned} &= \frac{(-1)^{j+\nu}}{j!^2 \nu!^2 (j+\nu+1)} \sum_{i=j \vee \nu+1}^p \left( \frac{(i+\nu+1)!(j+i+1)!}{(i-\nu)!(i-j)!} - \frac{(i+\nu)!(j+i)!}{(i-\nu-1)!(i-j-1)!} \right) \\ &\quad + \frac{(-1)^{j+\nu}}{j!^2 \nu!^2} \frac{((j \vee \nu) + \nu)!(2(j \vee \nu) + 1)(j + (j \vee \nu))!}{((j \vee \nu) - \nu)!((j \vee \nu) - j)!} \\ &= \frac{(-1)^{j+\nu}}{j!^2 \nu!^2 (j+\nu+1)} \frac{(p+\nu+1)!(j+p+1)!}{(p-\nu)!(p-j)!}. \quad (31) \end{aligned}$$

Also, since  $Q_i(1) = \frac{d^i}{dy^i} (y(1-y))^i \Big|_{y=1} = (-1)^i i!$  and  $K_{\nu, p+1}^{opt}(1) = 0$  (see (14))

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

This is the same as

$$\sum_{i=\nu}^p \frac{(2i+1)}{i!^2} \nu! q_{i,\nu} (-1)^i i! + \frac{(2p+3)}{(p+1)!^2} (\nu! q_{p+1,\nu} + q_{p+1,p+1} \beta) (-1)^{p+1} (p+1)! = 0. \quad (32)$$

The first term is

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

(Note:  $(2i+1) = (i+\nu+1) - (i-\nu)$ .)

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

Thus equation (32) yields

$$\beta = \frac{(-1)^{\nu+p} (p+\nu+2)! (p+1)!^2}{\nu! (2p+3)! (p-\nu+1)!}. \quad (33)$$

Combining this with equations (30) and (31) we have

$$K_{\nu,p+1}^{opt}(x) = \sum_{j=0}^{p+1} \lambda_j x^j,$$

where

$$\lambda_j = \frac{(-1)^{j+\nu} (p+j+1)! (p+\nu+2)!}{j!^2 \nu! (p-\nu)! (p-j+1)! (j+\nu+1)!}, j = 0, 1, \dots, p+1. \quad (34)$$

Since the polynomials  $\{Q_i\}$  are orthogonal,

$$\begin{aligned} \|K_{\nu,p+1}^{opt}\|^2 &= \sum_{i=0}^{p+1} a_i^2 \|Q_i\|^2 = \sum_{i=\nu}^p \nu!^2 q_{i,\nu}^2 \frac{2i+1}{i!^2} + \frac{(2p+3)}{(p+1)!^2} (\nu! q_{p+1,\nu} + q_{p+1,p+1} \beta)^2 \\ &= \sum_{i=\nu}^p \frac{(2i+1)(i+\nu)!^2}{\nu!^2 (i-\nu)!^2} + \frac{(2p+3)}{(p+1)!^2} (\nu! q_{p+1,\nu} + q_{p+1,p+1} \beta)^2 \end{aligned}$$

From (33), (Note:  $(2i+1) = \{(i+\nu+1)^2 - (i-\nu)^2\} / (2\nu+1)$ .)

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

$$\begin{aligned}
& + \frac{(p + \nu + 1)!^2}{(2p + 3)\nu!^2(p - \nu)!^2} \\
& = \frac{2(p + \nu + 2)(p + \nu + 1)!^2}{(2\nu + 1)(2p + 3)\nu!^2(p - \nu)!^2}. \tag{35}
\end{aligned}$$

## References

- [1] Brown, L. D. and Low, M. G. (1993) Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, to appear.
- [2] Chu, C. K. and Marron, J. S. (1991) Choosing a kernel regression estimator. *Statistical Science*, 6, No. 4 404-436.
- [3] Djojosugito, R. A. and Speckman, P. L. (1992) Boundary bias correction in nonparametric density estimation. *Commun. Statist. - Theory Meth.*, 21, 69-88.
- [4] Donoho, D. L. and Johnstone, I. (1992) Ideal spatial adaptation by wavelet shrinkage. Technical report 400, Stanford University.
- [5] Donoho, D. L. and Liu, R. C. (1991) Geometrizing rate of convergence III. *Ann. Statist.*, 19, 668-701.
- [6] Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990) Minimax risk over hyper rectangles, implications. *Ann. Statist.*, 18, 1416-1437.
- [7] Eubank, R. L. and Speckman, P. (1991) A bias reduction theorem with applications in nonparametric regression. *Scand. J. Stat.* 211-222.
- [8] Efroimovich, S. (1993) On nonparametric regression for i.i.d. observations in general setting. *Ann. Statist.*, to appear.
- [9] Fan, J. (1992) Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* 87, 998-1004.
- [10] Fan, J. (1993) Local linear regression smoothers and their minimax efficiency. *Ann. Statist.*, 21, 196-216.

- [11] Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engels, J. (1992) Exploiting a standard of nonparametric regression via local polynomial fitting, manuscript.
- [12] Fan, J. & Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *Ann. Stat.* 20, 2008-2036.
- [13] Fan, J. and Hall, P. (1993) On curve estimation by minimizing mean absolute deviation and its implications. *Ann. Statist.*, to appear.
- [14] Fan, J. and Marron, J. S. (1992) Fast implementations of nonparametric curve estimators. Institute of Statistics, University of North Carolina at Chapel Hill, Mimeo series #2093
- [15] Gasser, T. & Müller, H. G. (1979) Kernel estimation of regression function., In *Lecture Notes in Mathematics*, 757, 23-68, New York: Springer.
- [16] Gasser, T. , Müller, H. G., & Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *J. R. Statist. Soc. B*, 47, No.2, 238-252.
- [17] Granovsky, B. L. and Müller, H. G. (1991) Optimizing kernel methods: A unifying variational principle. *International Statistical Review*, 59, No. 3, 373-388.
- [18] 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.
- [19] Müller, H. G. (1991a) Smooth optimal kernel estimators near endpoints. *Biometrika*, 78, No. 3, 521-530.
- [20] Müller, H. G. (1991b) On the boundary kernel method for nonparametric curve estimation near endpoints, unpublished manuscript.
- [21] Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in  $L_2$ . *Ann. Statist.*, 13, 984-997.
- [22] Rice, J. (1984) Boundary modification for nonparametric regression. *Commun. Statist. A*, 13, 893-900.

- [23] Rice, J. & Rosenblatt, M. (1981) Integrated mean squared error of a smoothing spline. *J. Approx. Theory*, 33, 353-365.
- [24] Stone, C. J. (1977) Consistent nonparametric regression, *Ann. Statist.*, 5, No. 4, 595-645.