The Annals of Statistics
1993, Vol. 21, No. 1, 179–195

A NEW CLASS OF KERNELS FOR NONPARAMETRIC CURVE ESTIMATION

BY KAREN MESSER[1] AND LARRY GOLDSTEIN[2]

California State University and University of Southern California

We introduce a new class of variable kernels which depend on the smoothing parameter b through a simple scaling operation, and which have good integrated mean square error (IMSE) convergence properties. These kernels deform "automatically" near the boundary, eliminating boundary bias. Computational formulas are given for all orders of kernel in terms of exponentially damped sines and cosines. The kernel is a computationally convenient approximation to a certain Green's function, with the resulting kernel estimate closely related to a smoothing spline estimate.

1. Introduction. The problem of choosing an appropriate kernel function $K_b(x,t)$ arises in the setting of nonparametric curve fitting: One would like to estimate an unknown function $f(x)$ or its derivatives, where f may be for example a regression function, density function, link function or a dose response curve. Here, b is a smoothing parameter which is chosen by the user, and goes to 0 as the number of observations goes to infinity. Nothing is assumed to be known about f except that it belongs to some smoothness class, for example f is continuous or has a specified number p of continuous derivatives. We shall assume f is to be estimated on $[0,1]$. Most of the estimation methods proposed in this context involve local smoothing; the estimate $\hat{f}(x)$ is an average of neighboring observations, using some weight function which depends on the number of observations and may depend on x. In the case of a kernel estimator the weight function is chosen explicitly and is often a scaled and translated version of some fixed kernel, a "translation" kernel of the form $K_b(x,t) = k((x-t)/b)$. It is advantageous to allow the shape of the kernel to depend on x in order to correct for bias effects at the boundary, giving rise to a more complicated "variable" kernel $K_b(x,t)$. It is the choice of $K_b(x,t)$ which is the subject of this paper. In general, the kernel K should be chosen to exploit the smoothness of f; thus we are looking for a class or hierarchy of kernels, one kernel corresponding to each order of smoothness p.

We present a new class of variable kernel functions which are easy to compute, and which have a simplifying scaling property, such that each kernel

Received July 1989; revised February 1992.
[1]Partially supported by NSF Grant DMS-91-08037.
[2]Partially supported by NSF Grant DMS-90-05833
AMS 1991 *subject classifications*. 62G07, 62J02.
Key words and phrases. Nonparametric curve estimation, kernel, boundary bias, Green's function.

has good convergence properties over the appropriate smoothness class of functions f. These kernels are closely related to the equivalent kernel of a smoothing spline, as in Speckman (1981), Cox (1983) and Silverman (1984), and so for want of a better term we call them pseudo-spline kernels. They have several attractive features:

1. The kernels deform smoothly near the boundary in such a way as to correct for boundary bias.
2. The kernel can be evaluated by a simple scaling operation on a fixed function.
3. The scaling function $K(x, t)$ is available in closed form for all orders of kernel, and for estimating any derivative. It is a sum of exponentially damped trigonometric polynomials.
4. A small simulation study shows that these kernels have good finite sample properties. The kernel of order p behaves like a translation kernel of order $2p$, to which a smooth boundary correction of order p has been applied.

Several authors have investigated the "optimal" choice of kernel K under varying optimality criteria, beginning with Epanechnikov (1969), who studied the question for density estimation at a point and who considered only positive kernels. Rosenblatt (1971) and Benedetti (1977) investigated the question further in the regression context. More recently, Gasser and Müller (GM) (1979, 1984) and Gasser, Müller and Mammitzsch (GMM) (1985) have introduced two classes of optimal kernels, which minimize the variance and the integrated mean squared error, respectively, among kernels which satisfy certain moment conditions such as in (3). A good discussion is in Müller (1988).

In order to focus attention on the kernel, in this paper we consider a simple regression model. One observes pairs (x_i, y_i), $i = 1, \dots, n$, with $x_i = i/n$ and

$$y_i = f(x_i) + \varepsilon_i. \tag{1}$$

The ε_i are independent with $E(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$. The unknown function f has p continuous derivatives on $[0, 1]$.

The kernel estimate $\hat{f}$ of f with kernel $K_b(x, t)$ and bandwidth b is defined as

$$\hat{f}(x) = \frac{1}{nb} \sum_1^n y_i K_b(x, x_i).$$

Here, b must be chosen to go to zero at the proper rate as $n \to \infty$.

We sketch a standard argument for convergence of the estimate $\hat{f}(x)$ to $f(x)$ in the case where $p = 2$; see GM (1979) for details and considerable refinement. Consider the expected value of the estimate $E\hat{f}$ and suppose for simplicity K_b is a translation kernel. Any of various quadrature arguments followed by a change of variable and a Taylor expansion in t of $f(x + bt)$ gives an approximation

$$(2)\qquad \begin{aligned} E\hat{f}(x) &\approx b^{-1}\int_0^1 k((x-t)b^{-1})f(t)\,dt \\ &= \int_{(x-1)/b}^{x/b} k(t)\left\{f(x) - btf^{(1)}(x) + \tfrac{1}{2}(bt)^2 f^{(2)}(x) + R(bt)\right\}dt. \end{aligned}$$

Here $R(bt)$ is the remainder term in the Taylor expansion.

Now suppose the kernel $k(t)$ has compact support $[-\tau, \tau]$, and satisfies the usual symmetry and moment conditions:

$$(3)\qquad \begin{aligned} &\int_{-\tau}^{\tau} t^j k(t)\,dt = 1 - j, \qquad j = 0, 1, \\ &\left|\int_{-\tau}^{\tau} t^2 k(t)\,dt\right| = \alpha, \qquad \alpha < \infty. \end{aligned}$$

Then, so long as $(x-1)/b < -\tau < \tau < x/b$, and ignoring the quadrature error, we have easily that $|E\hat{f}(x) - f(x)| = (1/2)\alpha b^2 f^{(2)}(x) + o(b^2)$. Similar considerations give $\mathrm{Var}(\hat{f}(x)) \approx \int k(t)^2\sigma^2/(nb)$. Using these relations, we may choose b to minimize the asymptotic MSE of the estimate at the point x. As $n \to \infty$, if b is chosen optimally $\hat{f}(x)$ will converge to $f(x)$ with the optimal order of convergence for $p = 2$ as established in Stone (1980). Appropriately modifying the moment conditions (3) gives, by a similar argument, a kernel suitable for estimating $f^{(\nu)}$, the νth derivative of f, for $f \in C^p$.

To see the need for a variable kernel, consider x near the boundary. Then the convergence argument for the bias breaks down: The first moment as in (3) is now truncated, and the second term in (2) correspondingly fails to drop out. Thus the bias of the estimate for x within τb of the boundary of [0, 1] is of larger order than the bias in the interior. It can be shown that these boundary effects in fact dominate the IMSE, and a less than optimal rate of convergence results. In addition, it is not unusual for these boundary effects to smear over a substantial portion of the interval. [See, e.g., Eubank (1988) and the references therein.]

There are several approaches to the boundary problem in the literature. Rice (1983) provides a simple and effective boundary correction which can be easily computed for an arbitrary translation kernel $k(s)$. The corrected kernel is a linear combination of p copies of the original kernel evaluated at different bandwidths, thus introducing p new smoothing parameters to be chosen by the user. Eubank and Speckman (1991) have proposed a general bias removing approach which gives rise to several boundary correction schemes. GM (1979, 1984) and GMM (1985) propose a modification of the moment conditions (3) in which the kernel satisfies for each x in the boundary region

$$\int_{(x-1)/b}^{\tau} k(t)t\,dt = 0.$$

Then the corresponding terms in the expansion (2) drop out as before.

The present paper and the approach of GM share the advantage of presenting closed form formulas for the resulting boundary corrected weight function. Such formulas can be computationally convenient, and are theoretically tractable if the kernel estimator is to be used in a more complicated model than that considered here. Closed form formulas for boundary corrected minimum variance kernels as in GM (1979) have been recently given in Müller (1991) in terms of ultraspherical polynomials. Closed form formulas for the pseudospline kernel are given in Section 2.

In this paper, rather than begin with the moment conditions (3) we exploit the somewhat different properties of the Green's function of a smoothing spline estimator, that is, the equivalent kernel of the smoothing spline literature [Speckman (1981), Cox (1983) and Silverman (1984)]. Our kernel is thus closely related to the equivalent kernel of a smoothing spline estimator. The special case $p = 2$ was presented in Messer (1991).

The true "equivalent kernel" would give rise to an estimator with desirable properties, as detailed in Section 4; however it depends in a rather complicated way upon the smoothing parameter b. In particular, the Green's function $G_b(x, t)$ does not satisfy the scaling property we require of a kernel, that is, $G_b(x, t)$ cannot be obtained from $G_1(x, t)$ by a simple scaling operation. The construction of $G_b(x, t)$ requires the solution of a set of $2p \times 2p$ linear equations for each value of b. The key idea of this paper is to exploit a certain symmetry in the construction of the Green's function, approximating it by the kernel $K_b(x, t)$ which retains the asymptotic properties of $G_b(x, t)$ and allows b to enter as a scaling parameter.

The motivation for choosing an equivalent smoothing spline kernel is the hope that the kernel estimate will inherit many of the properties of the corresponding spline estimate. For the uniform design considered here, and $0 \le \nu \le p$, the kernel estimate of $f^{(\nu)}$ will have the same asymptotic bias and variance, to first order, as the spline estimate considered in Rice and Rosenblatt (1983); see Messer (1991). In some cases the pseudospline kernel estimate may be computationally more convenient than the corresponding spline estimate, especially with large data sets. A kernel estimator is sometimes more convenient from a theoretical point of view as well, especially if it satisfies the scaling relation 2. For nonuniform designs a comparison to a spline estimator is more difficult.

How does the pseudospline kernel compare to existing kernels in the literature? For many boundary corrected kernels the asymptotic integrated mean squared error of the estimator can be shown to be determined by the behavior of the kernel for x away from the boundary. For many kernels, if f has p derivatives and the kernel is chosen to estimate the νth derivative of f, the variance of the estimate can be seen to be asymptotic to the quantity $(\sigma^2/nb^{2\nu+1})V(K)$, where $V(K) \sim \int K_1(0.5, t)^2\, dt$. Similarly, the bias will be asymptotically proportional to $b^{p-\nu}B(K)$, where $B(K)$ is often of the form $B(K) \sim \int K_1(0.5, t)t^p\, dt$. The integrals are taken over the support of the kernel. The asymptotic IMSE can then be seen to depend on the kernel through the functional

$$T = \left(V^{p-\nu}B^{2p+1}\right)^{2/(2p+1)}.$$

(See GM or GMM for details.)

The bias is of a somewhat different form for the pseudospline kernel, and in fact is $o(b^p)$; see Lemma 6.2 and (11). The spline kernel of order p suitable for estimating the νth derivative may be viewed as containing an interior translation-invariant component which is of order $m = 2p + \nu$ at any fixed interior point. That is, at an interior point if the function f actually has $m + 1$ derivatives, the asymptotic bias is proportional to $b^{m-\nu}B(K)$, where $|B(K)| \sim \int k^{(\nu)}(t)t^m\,dt = m!$ and where $k(t)$ is the translation component of the kernel as given in (6). (See Proposition 3.4.) To this, boundary correction terms of order p are added which decay exponentially away from the boundary, and which bring the order of the kernel down to p over the entire interval. If we compare the asymptotic bias and variance constants $V(K)$ [given in (10)] and $B(K)$ for the spline and the GM optimal kernels, an interesting comparison is between the spline kernel of order p and the GM kernel of order $m = 2p + \nu$. In this case, which ignores boundary effects, the GM kernels, being given as the solution of a variational problem, do somewhat better in terms of asymptotic IMSE. The corresponding constants may be found for the GM kernel in GMM, Table 2, page 243. For example, for $p = 2$, $\nu = 0$; $p = 2$, $\nu = 1$, respectively, we obtain $T = 0.6227$, 2.33, respectively, for the spline kernel and $T = 0.6199$, 2.168 for the GM optimal kernel. Results of a small numerical study are reported in Section 5, where the pseudospline kernel is seen to compare favorably in finite samples both with and without boundary correction.

The remainder of the paper is organized as follows. In Section 2 explicit computational formulas are presented. In Section 3 theorems are stated concerning properties of the kernel and rates of convergence of the kernel estimator. Section 4 presents the connection between our kernel and the equivalent kernel of a smoothing spline. Section 5 presents numerical comparisons. Section 6 contains proofs.

2. Formula for the kernel. In this section we present the general formula for the kernel of order p, and give the explicit formula for $p = 2$ and 3.

2.1. *The general formula.* For $0 \le j \le 2p - 1$, let the $2p$th roots of -1 be given by $r_j = \exp(i\pi(2j+1)/2p)$ and define the $2p \times 1$ column vector

$$\psi_b(t) = \left\langle e^{ib^{-1}tr_0}, \dots, e^{ib^{-1}tr_{p-1}}, e^{ib^{-1}(1-t)r_0}, \dots, e^{ib^{-1}(1-t)r_{p-1}} \right\rangle'.$$

We shall denote the components of $\psi_b(t)$ by $\psi_{b,j}$, $0 \le j \le 2p - 1$.

The kernel is made up of two parts, one translation invariant and the other comprised of terms that deal with boundary effects. The translation invariant part $k(t)$ of the kernel is the Fourier transform of $(\sqrt{2\pi}(x^{2p} + 1))^{-1}$, that is,

$(2\pi)^{-1}\int_{-\infty}^{\infty} e^{-itx}(x^{2p}+1)^{-1}\,dx$. Hence

$$k(t) = \frac{-1}{2p}\sum_{j=0}^{p-1} ir_j\psi_{1,j}(|t|) \tag{4}$$

and

$$k^{(l)}(t) = \frac{-1}{2p}\sum_{j=0}^{p-1}(ir_j)^{l+1}\psi_{1,j}(t) \quad \text{for } t > 0.$$

For p even this becomes

$$k(t) = p^{-1}\sum_{j=0}^{p/2-1} e^{-|t|\mathrm{Im}(r_j)}\left\{\mathrm{Im}(r_j)\cos\left(\mathrm{Re}(r_j)|t|\right) + \mathrm{Re}(r_j)\sin\left(\mathrm{Re}(r_j)|t|\right)\right\}$$

and for p odd

$$k(t) = (2p)^{-1}e^{-|t|} + p^{-1}\sum_{j=0}^{p/2-3/2} e^{-|t|\mathrm{Im}(r_j)}\left\{\mathrm{Im}(r_j)\cos\left(\mathrm{Re}(r_j)|t|\right) + \mathrm{Re}(r_j)\sin\left(\mathrm{Re}(r_j)|t|\right)\right\}.$$

Next we give the boundary terms of the kernel. They are constructed from $\phi(t) = \langle \phi_p(t), \dots, \phi_{2p-1}(t)\rangle'$, a $p \times 1$ vector of the homogeneous solutions of the differential equation (14) of Section 4 with $b = 1$, which satisfy the boundary conditions

$$\left[\phi^{(p)}(0), \dots, \phi^{(2p-1)}(0)\right] = I, \tag{5}$$

where I is the $p \times p$ identity matrix. The required vector $\phi(t) = [C, 0_{p\times p}]\psi_1(t)$, where the $p \times p$ matrix C is given by $C = L^{-1}\Lambda^{-p}$, with $\Lambda = i\,\mathrm{diag}\langle r_0, \dots, r_{p-1}\rangle$ and L the Vandermonde matrix $L = [\mathbf{1}, \Lambda\mathbf{1}, \dots, \Lambda^{p-1}\mathbf{1}]$. Here $\mathbf{1} = \langle 1, \dots, 1\rangle'$.

The equation for the kernel $K_b(x,t)$ is

$$\begin{aligned} K_b(x,t) &= b^{-1}k\left(\frac{x-t}{b}\right) \\ &\quad + b^{-1}\sum_{j=1}^{p}(-1)^{j+1}\left\{\phi_{2p-j}\left(\frac{t}{b}\right)k^{(2p-j)}\left(\frac{x}{b}\right)\right. \\ &\qquad \left. +\phi_{2p-j}\left(\frac{1-t}{b}\right)k^{(2p-j)}\left(\frac{1-x}{b}\right)\right\}. \end{aligned} \tag{6}$$

2.2. *Examples for* $p = 2$ *and* 3. The general formula may be simplified in any particular case by the use of trigonometric identities. For example, with $p = 2$ let $b' = \sqrt{2}\,b$, and let

$$\Phi(u,v) = e^{-u}(\cos(u) - \sin(u) + 2\cos(v)).$$

Then

$$K_b(x,t) = (2b')^{-1}e^{-|x-t|/b'}\left(\sin\frac{|x-t|}{b'} + \cos\left(\frac{x-t}{b'}\right)\right)$$
$$+(2b')^{-1}\left(\Phi\left(\frac{x+t}{b'},\frac{x-t}{b'}\right) + \Phi\left(\frac{1-x}{b'}+\frac{1-t}{b'},\frac{1-x}{b'}-\frac{1-t}{b'}\right)\right).$$

We next present the functions appearing in (6) for $p = 2$ and $p = 3$. For $p = 2$,

$$k(t) = \frac{1}{2\sqrt{2}}e^{-(1/\sqrt{2})|t|}\left\{\left(\cos\left(\frac{1}{\sqrt{2}}t\right) + \sin\left(\frac{1}{\sqrt{2}}|t|\right)\right\},$$

$$k^{(2)}(t) = \frac{1}{2\sqrt{2}}e^{-(1/\sqrt{2})|t|}\left\{-\cos\left(\frac{1}{\sqrt{2}}t\right) + \sin\left(\frac{1}{\sqrt{2}}|t|\right)\right\},$$

$$k^{(3)}(t) = \frac{1}{2}e^{-(1/\sqrt{2})|t|}\cos\left(\frac{1}{\sqrt{2}}t\right)\operatorname{sgn}(t),$$

$$\phi_2(t) = -e^{-(1/\sqrt{2})t}\left\{-\cos\left(\frac{1}{\sqrt{2}}t\right) + \sin\left(\frac{1}{\sqrt{2}}t\right)\right\},$$

$$\phi_3(t) = \sqrt{2e}^{-(1/\sqrt{2})t}\cos\left(\frac{1}{\sqrt{2}}t\right).$$

For $p = 3$,

$$k(t) = \frac{1}{6}\left(e^{-|t|} + e^{-(|t|/2)}\left\{\cos\left(\frac{\sqrt{3}}{2}t\right) + \sqrt{3}\sin\left(\frac{\sqrt{3}}{2}|t|\right)\right\}\right),$$

$$k^{(3)}(t) = \frac{1}{6}\left(-e^{-|t|} + e^{-(|t|/2)}\left\{\cos\left(\frac{\sqrt{3}}{2}t\right) + \sqrt{3}\sin\left(\frac{\sqrt{3}}{2}|t|\right)\right\}\right),$$

$$k^{(4)}(t) = \frac{1}{6}\left(e^{-|t|} + e^{-(|t|/2)}\left\{\cos\left(\frac{\sqrt{3}}{2}t\right) - \sqrt{3}\sin\left(\frac{\sqrt{3}}{2}t\right)\right\}\right),$$

$$k^{(5)}(t) = -\frac{1}{6}\left(e^{-|t|} + 2e^{-(|t|/2)}\cos\left(\frac{\sqrt{3}}{2}t\right)\right),$$

$$\phi_3(t) = -e^{-t} + e^{-(t/2)}\left\{\frac{2}{\sqrt{3}}\sin\left(\frac{\sqrt{3}}{2}|t|\right)\right\},$$

$$\phi_4(t) = -e^{-t} + e^{-(t/2)}\left\{-\cos\left(\frac{\sqrt{3}}{2}t\right) + \sqrt{3}\sin\left(\frac{\sqrt{3}}{2}t\right)\right\},$$

$$\phi_5(t) = -e^{-t} + e^{-(t/2)}\left\{-\cos\left(\frac{\sqrt{3}}{2}t\right) + \frac{1}{\sqrt{3}}\sin\left(\frac{\sqrt{3}}{2}t\right)\right\}.$$

3. Properties of the kernel. In this section we give the properties of the kernels defined in Section 2, and of the kernel estimator of the introduction. We begin with notation, and then give two theorems which bound the asymptotic bias functional of the kernel and evaluate the asymptotic variance functional. Finally we apply these theorems to bound the asymptotic IMSE for the kernel estimator of the simple regression model considered in the introduction. Proofs are given in Section 6.

Mixed partial derivatives $\partial^{i+j}K_b(x,t)/\partial x^i \partial t^j$ are denoted as $K_b^{(i,j)}(x,t)$. We shall need the norms $\|h\|_\infty = \sup_{x\in[0,1]}|h(x)|$ and $\|h\|_{(\infty,p)} = \sum_{j=0}^p \|h^{(j)}\|_\infty$. In what follows, $C(\cdot)$ will denote a positive constant which depends only on its argument and which is not necessarily the same at each occurrence; $o(1)$ will denote a term which depends only on n and b and which goes to 0 as $n \to \infty$. We shall always take $0 < b \leq 1$.

The following theorem bounds the asymptotic bias functional of the kernel.

THEOREM 3.1. *For the kernel* $K_b(x,t)$ *as given in* (6),

$$\left|\int_0^1 K_b^{(\nu,0)}(x,t) f(t)\, dt - f^{(\nu)}(x)\right| \leq b^{p-\nu}\|f\|_{(\infty,p)}\left(\sum_{j=0}^{p-1} \frac{1}{p\,\mathrm{Im}(r_j)}\right)(1+o(1)), \tag{7}$$

$$\int_0^1 \left\{\int_0^1 K_b^{(\nu,0)}(x,t) f(t)\, dt - f^{(\nu)}(x)\right\}^2 dx \leq C(p,\nu,f)\, b^{2(p-\nu)} o(1), \tag{8}$$

for all $b > 0$, $f \in C^p[0,1]$ *and* $0 \leq \nu \leq p$.

If $f^{(p+1)}$ *exists on* $[0,1]$ *and satisfies* $\|f^{(p+1)}\|_\infty \leq \|f\|_{(\infty,p)}$, *then the right-hand side of* (8) *becomes* $C(p,\nu) b^{2(p-\nu)+1}\|f\|_{(\infty,p)}$.

REMARK 1. Notice that the first bound is independent of x for $x \in [0,1]$. Hence, there is no boundary bias to first order.

REMARK 2. The second bound combined with Theorem 3.2 below implies that, at any particular $f \in C^p$, the kernel estimator behaves like a kernel of order at least $p+1$ in that it improves on the optimal order of convergence established in Stone (1980). Of course, super efficient decay does not hold uniformly over the appropriate space of functions.

Next we give a theorem which is used to establish the asymptotic variance of the kernel estimator, apart from quadrature error.

THEOREM 3.2.

$$\int_0^1\int_0^1 \left(K_b^{(\nu,j)}(x,t)\right)^2 dt\, dx = b^{-(2(\nu+j)+1)}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{x^{\nu+j}}{1+x^{2p}}\right)^2 dx\right)(1 + C(p)O(b)), \tag{9}$$

where

$$(10)\quad \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{x^s}{1+x^{2p}}\right)^2 dx = \frac{(2p-2s-1)}{2p^2}\left(\frac{\sin(\pi(2s+1)/2p)}{1-\cos(\pi(2s+1)/p)}\right).$$

REMARK. The limiting value $p = \infty$ in (10) yields $1/((2s+1)\pi)$.

The following theorem establishes the asymptotic bias and variance of the kernel estimator by a simple application of the two previous theorems.

THEOREM 3.3. *Consider the regression model as given in* (1). *Let* $b = b(n)$ *depend on* n *in such a way that* $b \to 0$, $b^{p+2}n \to \infty$ *as* $n \to \infty$. *Let the estimate* $\widehat{f^{(\nu)}}(x)$ *be given by*

$$\widehat{f^{(\nu)}}(x) = \frac{1}{n}\sum_{j=1}^{n} y_j K_{b(n)}^{(\nu,0)}(x, x_j).$$

Then

$$(11)\quad \int_0^1\left(E\widehat{f^{(\nu)}}(x) - f^{(\nu)}(x)\right)^2 dx = C(p,\nu,f)\,b^{2(p-\nu)}o(1)$$

and

$$(12)\quad \int_0^1 \operatorname{Var}\left(\widehat{f^{(\nu)}}(x)\right) dx = \frac{\sigma^2}{nb^{2\nu+1}}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{x^\nu}{1+x^{2p}}\right)^2 dx\right)(1 + C(p)o(1)).$$

The following proposition is useful in comparing the spline kernel to existing moment kernels in the literature.

PROPOSITION 3.4.

$$\int_{-\infty}^{\infty} t^j k^{(\nu)}(t)\,dt = \begin{cases} 0, & 0 < j < 2p+\nu,\ j \neq \nu, \\ (-1)^\nu(\nu)!, & j = \nu, \\ (-1)^{p+\nu+1}(2p+\nu)!, & j = 2p+\nu. \end{cases}$$

4. The Green's function. In this section we present the connection between the kernel $K_b(x,t)$ and the Green's function of a corresponding smoothing spline estimator. Consider the "continuous" version of the smoothing spline minimization problem for given f and equally spaced design points x_i,

$$(13)\quad \min_h \int_0^1 (f-h)^2 + b^{2p}\int_0^1 (h^{(p)})^2$$

and the associated Euler equations

$$(-1)^p b^{2p} h^{(2p)} + h = f, \tag{14}$$

$$0 = h^{(j)}(0) = h^{(j)}(1), \qquad j = p, \dots, 2p-1. \tag{15}$$

It can be shown that the solution to (13) exists and is unique [see e.g., Cox (1983)], and hence that the solution is determined by the system (14), (15). It can then be shown that the solution to the latter is determined by the unique Green's function $G_b(x,t)$, that is, the minimizing h is given by

$$h(x) = \int_0^1 G_b(x,t) f(t)\, dt.$$

The equivalent kernel of the literature is exactly $G_b(x,t)$. However, $G_b(x,t)$ depends in a complicated way on the smoothing parameter b, and in particular does not satisfy the scaling property we require of a kernel, that is, $G_b(x,t)$ cannot be obtained by a simple scaling operation from $G_1(x,t)$.

We briefly sketch the construction of the exact Green's function [see, e.g., Speckman (1981) and Coddington and Levinson (1955)]. First the "fundamental solution" is obtained: a Green's function for (14) so long as f satisfies stringent boundary conditions. Next, the fundamental solution is extended to the space of functions $C^p([0,1])$, by modifying it to satisfy boundary conditions (15) when considered as a function of x for fixed t, $x \neq t$; $x, t \in [0,1]$. This construction requires the solution of a $2p \times 2p$ system of linear equations for each value of b.

The construction of the kernel $K_b(x,t)$ is related. We begin with the fundamental solution $b^{-1}k(b^{-1}(x-t))$, and modify it by adding the proper linear combination of solutions to the homogeneous equation (14) for $b = 1$ so that $K_1(x,t)$ is nearly $G_1(x,t)$. The exact relation may be found in the proof of Theorem 4.1. Then $K_b(x,t)$ is obtained by a simple scaling operation on $K_1(x,t)$, as can be seen from (6). In some sense, when $K_b(x,t)$ is near the left boundary (i.e., for x near 0), we allow the right boundary to be unconstrained and vice versa. We then exploit the exponential decay of the homogeneous solutions to (14) to obtain the following bound on the difference between $K_b(x,t)$ and $G_b(x,t)$:

THEOREM 4.1. *For* $0 < b \leq 1$

$$\sup_{0 \leq x, t \leq 1} \left| K_b^{(\nu,0)}(x,t) - G_b^{(\nu,0)}(x,t) \right| \leq C(p,\nu) b^{-(\nu+1)} e^{-(b^{-1}\sin(\pi/2p))}$$

The proof is given in Section 6.

5. Numerical results. A small numerical study was done to compare the pseudospline kernel with the GM optimal kernels in finite samples, both with boundary effects and without. In the first case we consider the behavior of the estimate over an interior region following the example of GM Tables 4–6. In the interior the spline kernel of order p suitable for estimating the νth derivative behaves as a moment kernel of order $2p + \nu$; see Proposition 3.4 and the discussion at the end of the introduction. Hence the spline kernel of nominal order p was compared to the GM kernel of order $2p + \nu$.

As in GM Tables 4–6, we used the smooth test function $f(x) = 2 - 2x + 3\exp(-100(x - 0.5))$ with $2n$ equispaced observations over the interval $[-0.5, 1.5]$, and evaluated the optimal IMSE for $\sigma^2 = 0.4$ over the interior interval $[0, 1]$. This avoids boundary effects; as in GM, both kernels were used without boundary correction. We used a simple Simpson's quadrature rule rather than the more complicated GM quadrature rule, and replicated the results in GM Table 4 in several cases for verification. Programming was done in MATLAB. Comparisons of the IMSE evaluated at the optimal bandwidth were done for the cases $p = 2, 4$; $n = 25, 100$; $\nu = 0$ and $p = 4$, $n = 100$, $\nu = 2$. In all cases the optimal IMSE were within a few percent of each other, with the spline kernel having the smaller IMSE despite having a larger IMSE asymptotically. Presumably for larger sample sizes the comparison would reverse.

The more interesting comparison is perhaps with boundary correction. We used the shifted function $f(x) = 2 - 2x + 3\exp(-100x)$, with equispaced observations within $[0, 1]$. Now the function has a bump truncated at the left boundary. The GM boundary corrected kernel as in Müller (1991) Table 2, for $k = \mu = 2$, $\nu = 0$ was compared to the pseudospline kernel for $p = 2$. [The GM kernel is the common quartic kernel $15/16(1 - x^2)^2$, boundary corrected.] For sample sizes $n = 25, 40, 50, 100, 200$, the ratio of the pseudospline IMSE to the GM IMSE was 64%, 69%, 73%, 75% and 74%, respectively.

6. Proofs of Theorems. We shall use notation and results from Rudin (1973) (R) on Fourier transforms and differential equations. As in R, we take as measure $dm(t)$ on R, where $dm(t) = (\sqrt{2\pi})^{-1}\,dt$ and where dt is Lebesgue measure. We correspondingly redefine $k(t)$ to have the factor $\sqrt{2\pi}$ in front for this section only, and in keeping with the notation in R shall use $\hat{f}$ to denote the Fourier transform of f. The L^2 norm is taken to be $\|h\|_2^2 = \int_{-\infty}^{\infty} h^2(t)\,dm(t)$. We first give the following useful proposition which is a simple consequence of our definitions.

Proposition 6.1. *Let* $k_b(x) = b^{-1}k(xb^{-1})$. *Then*

$$\left|k_b^{(j)}(x)\right| \le C(j, p)\, b^{-(j+1)} e^{-\sin(\pi/2p)|x/b|}$$

for all $j \ge 0$. *A similar bound holds for* $(b^{-1}\phi_l(t/b))^{(j)}$.

In order to establish Theorem 3.1 we shall need the following representation of the asymptotic bias:

Lemma 6.2.

$$\begin{aligned} &\int_0^1 K_b^{(\nu,0)}(x,t)(t)\,dt - f^{(\nu)}(x) \\ &\quad = -b^{2p}\int_0^1 K_b^{(\nu,p)}(x,t)\, f^{(p)}(t)\,dt + O\big(b^{-\nu} e^{-b^{-1}\sin(\pi/2p)}\big)\|f\|_{(\infty,p)}, \end{aligned} \tag{16}$$

where the term $O(\cdot)$ *depends only on its argument.*

PROOF. Consider the differential equation (14) as an equation over the space $\mathscr{D}'(R)$ of distributions on R [R (6.7)]. By the usual Fourier techniques, one can show that $k(x-t)$ is a fundamental solution for (14) for $b=1$, that is, that $k * f$ solves (14) for $f \in \mathscr{D}(R)$, the space of test functions on R, where $(k * f)(x) = \int_{-\infty}^{\infty} k(x-t) f(t)\, dm(t)$ as in R. It follows that

$$\int_{-\infty}^{\infty} k(x-t) f(t)\, dm(t) = f(x) + (-1)^{p+1} \int_{-\infty}^{\infty} k(x-t) f^{(2p)}(t)\, dm(t). \tag{17}$$

By specializing (17) to test functions f with support on [0, 1], using the usual properties of the fundamental solution k and integrating by parts $2p$ times we obtain

$$\int_0^1 k(x-t) f(t)\, dm(t) = f(x) + (-1)^{p+1} \left\{ \int_0^1 k^{(2p)}(x-t) f(t)\, dm(t) + \Delta k^{(2p-1)}(0) f(x) \right\}.$$

With $b=1$, considering functions f with $f(x)=0$ and functions f which are zero outside an arbitrarily small neighborhood of x, and then applying the chain rule for arbitrary b yields

$$(-1)^p b^{2p} (k_b)^{(2p)}(x) + k_b(x) = 0, \qquad x \in (0,1], \tag{18}$$

$$(-1)^p b^{2p} \Delta (k_b)^{(2p-1)}(0) = 1, \tag{19}$$

where $\sqrt{2\pi}\, \Delta f(0) = \lim_{x \to 0^+} f(x) - \lim_{x \to 0^-} f(x)$.

Now, for $f \in C^p[0,1]$, the expression $\int_0^1 k_b(x-t) f(t)\, dm(t) - f(x)$ contains boundary terms such as $b^{j-1} k^{(2p-j)}(x/b) f^{(j-1)}(0)$ which are canceled by adding the boundary correction terms. In particular, one can demonstrate

$$\begin{aligned} \int_0^1 K_b(x,t) f(t)\, dm(t) - f(x) &= -b^{2p} \int_0^1 K_b^{(0,p)}(x,t) f^{(p)}(t)\, dm(t) \\ &\quad + (-1)^{p+1} \sum_{j=1}^{p} \sum_{m=1}^{p} b^{m-1} \phi_{2p-j}^{(2p-m)}\left(\frac{1}{b}\right) \\ &\quad \times \left\{ (-1)^{j+m} k^{(2p-j)}\left(\frac{x}{b}\right) f^{(m-1)}(1) \right. \\ &\qquad \left. + (-1)^{j+1} k^{(2p-j)}\left(\frac{1-x}{b}\right) f^{(m-1)}(0) \right\}. \end{aligned}$$

Differentiating both sides of the preceding equation with respect to x ν times and bounding the double sum using Proposition 6.1 establishes the lemma. □

In the following theorems we shall often bound cross terms by use of inequalities of the kind $(a+b)^2 \le 2(a^2+b^2)$ without further comment.

PROOF OF THEOREM 3.1. Statement (7) follows readily from Lemma 6.2 and Proposition 6.1, noting that

$$b^{-p}e^{-b^{-1}\sin(\pi/2p)}$$

is bounded for $b \in (0,\infty)$. To establish (8), we see from Lemma 6.2 that it is enough to consider the following integral:

$$b^{4p}\int_0^1\left\{\int_0^1 K_b^{(\nu,p)}(x,t)\,f^{(p)}(t)\,dm(t)\right\}^2 dx. \tag{20}$$

First, we shall consider the contribution from the boundary terms in the above integral, and show that it is of order $O(b^{2(p-\nu)+1})$. Consider the contribution from a typical boundary term of the form

$$\begin{aligned} b^{2(p-\nu)}&\int_0^1\left(b^{-1}\int_0^1 k^{(2p-j+\nu)}\left(\frac{x}{b}\right)\phi_{2p-j}^{(p)}\left(\frac{t}{b}\right)f^{(p)}(t)\,dm(t)\right)^2 dx \\ &= b^{2(p-\nu)}\left(\int_0^1\left(k^{(2p-j+\nu)}\left(\frac{x}{b}\right)\right)^2 dx\right)\left(b^{-1}\int_0^1 \phi_{2p-j}^{(p)}\left(\frac{t}{b}\right)f^{(p)}(t)\,dm(t)\right)^2. \end{aligned}$$

The left-hand integral is bounded by $C(p)b$, and the right-hand integral is easily seen to be bounded. Hence the contribution from the boundary terms may be majorized by $C(p,f)b^{2(p-\nu)+1}$.

To demonstrate (8) it now suffices to show that the contribution to (20) from the translation invariant term is of order $o(b^{2(p-\nu)})$. Extend f to R by taking $f(t)=0$ for $t \notin [0,1]$. Then we may write the contribution to (20) from the translation invariant term and bound it using Parseval's relation and the smoothness properties of k as follows:

$$\begin{aligned} b^{2(p-\nu)}&\int_0^1\left\{\int_{-\infty}^{\infty} b^{-1}k^{(\nu+p)}\left(\frac{x-t}{b}\right)f^{(p)}(t)\,dm(t)\right\}^2 dx \\ &\le b^{2(p-\nu)}\sqrt{2\pi}\left\|\widehat{\left(k_b^{(\nu+p)} * f^{(p)}\right)}\right\|_2^2 \\ &= b^{2(p-\nu)}\sqrt{2\pi}\int_{-\infty}^{\infty}\left(\widehat{k^{(\nu+p)}}(bu)\widehat{f^{(p)}}(u)\right)^2 dm(u) \\ &= b^{2(p-\nu)}\sqrt{2\pi}\int_{-\infty}^{\infty}\left(\frac{(bu)^{\nu+p}}{1+(bu)^{2p}}\right)^2\widehat{f^{(p)}}(u)^2\,dm(u), \end{aligned} \tag{21}$$

the last by recalling that $\hat{k}(t) = (t^{2p}+1)^{-1}$.

We shall show that the above integral tends to 0 as $b \to 0$. Note that $f^{(p)}$, and therefore $\widehat{f^{(p)}}$ are in $L_2(R)$, and that $\sup_{-\infty<u<\infty}|u^{\nu+p}/(1+u^{2p})| < \infty$.

Hence, given ε we may choose A depending only on f such that

$$\int_{|u|>A}\left(\frac{(bu)^{\nu+p}}{1+(bu)^{2p}}\right)^2\widehat{f^{(p)}}(u)^2\,dm(u)\le\frac{\varepsilon}{2}.$$

Then

$$\begin{aligned}\int_{-\infty}^{\infty}&\left(\frac{(bu)^{\nu+p}}{1+(bu)^{2p}}\right)^2\widehat{f^{(p)}}(u)^2\,dm(u)\\&\le\int_{|u|\le A}\left(\frac{(bu)^{\nu+p}}{1+(bu)^{2p}}\right)^2\widehat{f^{(p)}}(u)^2\,dm(u)+\frac{\varepsilon}{2}\\&\le b^{2(\nu+p)}\|f^{(p)}\|_\infty^2C(p,\nu,A)+\frac{\varepsilon}{2}.\end{aligned}$$

The last may be made less than ε for $b<b_0$, where b_0 depends on p, ν and f. This establishes (8).

To prove the last statement of the theorem, suppose that $f^{(p+1)}$ exists on $[0,1]$ and is bounded in absolute value by $\|f\|_{(p,\infty)}$. In this case we may integrate by parts in (21) to obtain

$$\begin{aligned}b^{2(p-\nu)}\int_0^1\Bigg\{&\int_{-\infty}^{\infty}k^{(\nu+p-1)}\left(\frac{x-t}{b}\right)f^{(p+1)}(t)\,dm(t)\\&-k^{(p+\nu-1)}\left(\frac{x-1}{b}\right)f^{(p)}(1)+k^{(p+\nu-1)}\left(\frac{x}{b}\right)f^{(p)}(0)\Bigg\}^2\,dx.\end{aligned}$$

By a change of variable we may bound the above by $C(p)\|f\|_{(\infty,p)}b^{2(p-\nu)+1}$. This establishes that the contribution to (20) from the translation invariant term is $O(b^{2(p-\nu)+1})$. □

PROOF OF THEOREM 3.2. The contribution from the boundary terms may be shown to be $O(b)$ times that of the translation invariant term by an argument similar to that in the proof of Theorem 3.1. Statement (9) follows from Parseval's relation and a change in variable. A contour integration then establishes (10). □

PROOF OF THEOREM 3.3. Consider first the bias of the estimate. In this model, we have easily

$$\begin{aligned}E\left(\widehat{f^{(\nu)}}(x)\right)&=n^{-1}\sum_{j=1}^{n}K_{b(n)}^{(\nu,0)}(x,x_j)\,f(x_j)\\&=\int_0^1K_{b(n)}^{(\nu,0)}(x,t)\,f(t)\,dt+\mathscr{Q}(x),\end{aligned}$$

where the quadrature error $\mathscr{Q}(x)$ is bounded in absolute value by $\|f\|_{(\infty,p)}C(p,\nu)n^{-1}b^{-\nu-2}$. Hence in order to establish (11) it is enough to consider the integral approximation to the bias, and then (11) follows immediately from Theorem 3.1 (8).

Statement (12) follows similarly, where the approximation error is $O(n^{-2}b^{-2\nu-3})$. □

PROOF OF PROPOSITION 3.4. It follows from the remarks preceding (17) that $h(x) = \int_{-\infty}^{\infty} k(x-t)t^j\,dt$ is a solution to (14) for $b = 1$ and $f(x) = x^j$. Substituting for $h(x)$ in (14), differentiating under the integral and integrating by parts yields $h(x) = x^j$ for $0 \le j \le 2p-1$, and $h(x) = x^{2p} + (-1)^{p+1}(2p)!$ for $j = 2p$. Integrating by parts yields $\int_{-\infty}^{\infty} k^{(\nu)}(x-t)t^j\,dt = \int_{-\infty}^{\infty} k(x-t)(t^j)^{(\nu)}\,dt$ and substituting $x = 0$ completes the proof. □

PROOF OF THEOREM 4.1. For $p \le j,\ k \le 2p-1$, let $\theta_k(t)$ be the solution to the homogeneous equation (14) with scaling parameter there set equal to 1 and satisfying boundary conditions

$$\theta_k^{(j)}(0) = \delta_{k,j}, \qquad \theta_k^{(j)}\left(\frac{1}{b}\right) = 0 \tag{22}$$

and let $\theta(t) = \langle \theta_p(t), \dots, \theta_{2p-1}(t)\rangle'$ be the corresponding $p \times 1$ vector. The Green's function for (14) and (15) is given by [Coddington and Levinson (1955) and Speckman (1981)].

$$\begin{aligned} G_b(x,t) &= b^{-1}k(b^{-1}(s-t)) + b^{-1}\sum_{j=1}^{p}(-1)^{j+1} \\ &\quad \times \left\{\theta_{2p-j}\left(\frac{t}{b}\right)k^{(2p-j)}\left(\frac{x}{b}\right) + \theta_{2p-j}\left(\frac{1-t}{b}\right)k^{(2p-j)}\left(\frac{1-x}{b}\right)\right\}. \end{aligned} \tag{23}$$

Comparing (6) to (23), and using Proposition 6.1 to verify that $k^{(2p-j+\nu)}$ is bounded, we see that it is sufficient to establish the bound

$$\sup_{0 \le t \le 1}\left|\theta_k\left(\frac{t}{b}\right) - \phi_k\left(\frac{t}{b}\right)\right| \le C(p,\nu)e^{-b^{-1}\sin(\pi/2p)}. \tag{24}$$

Let ψ_b, Λ and L be as in Section 2. As the components of ψ_b form a basis for the solution space of (14), we have

$$\theta\left(\frac{t}{b}\right) = [C_b, B_b]\psi_b(t)$$

for some pair of $p \times p$ matrices C_b, B_b. Using

$$\psi_b^{(j)}(t) = b^{-j}\begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{bmatrix}\psi_b(t),$$

we have from (22)

$$\left[\theta^{(p)}(0),\theta^{(p+1)}(0),\dots,\theta^{(2p-1)}(0),\theta^{(p)}\left(\frac{1}{b}\right),\dots,\theta^{(2p-1)}\left(\frac{1}{b}\right)\right]$$

$$= [I_p,0] = [C_b,B_b]\begin{bmatrix}\Lambda & 0\\ 0 & -\Lambda\end{bmatrix}^p\left[\psi_b(0),\dots,\begin{bmatrix}\Lambda & 0\\ 0 & -\Lambda\end{bmatrix}^{p-1}\psi_b(0),\psi_b(1),\right.$$

$$\left.\dots,\begin{bmatrix}\Lambda & 0\\ 0 & -\Lambda\end{bmatrix}^{p-1}\psi_b(1)\right].$$

Letting $D_b = \operatorname{diag}(e^{ir_0 b^{-1}},\dots,e^{ir_{p-1}b^{-1}})$ and $S = (-1)^p\operatorname{diag}(1,-1,\dots,(-1)^p)$, we simplify the above to find that $[C_b, B_b]$ is the solution to

$$[C_b,B_b]\begin{bmatrix}\Lambda^p & 0\\ 0 & \Lambda^p S\end{bmatrix}\begin{bmatrix}I & D_b\\ D_b & I\end{bmatrix}\begin{bmatrix}L & 0\\ 0 & L\end{bmatrix} = [I_p,0].$$

Solving, $[C_b, B_b] = [L^{-1}(I - D_b^2)^{-1}\Lambda^{-p}, -L^{-1}(I - D_b^2)^{-1}D_b\Lambda^{-p}S]$.

Recall that $\phi(t/b) = [C,0]\psi_b(t) = [L^{-1}\Lambda^{-p},0]\psi_b(t)$. Therefore, $\theta(t/b) - \phi(t/b) = [C_b - C, B_b]\psi_b(t)$ with $\psi_b(t)$ bounded, and (24) will follow by demonstrating that $\|(I - D_b^2)^{-1} - I\|$ and $\|D_b\|$ may be bounded by $C(p)e^{-b^{-1}\sin(\pi/2p)}$. For $\|D_b\|$ this is immediate, as $\|D_b\| = |e^{-b^{-1}ir_0}| = e^{-b^{-1}\sin(\pi/2p)}$.

We have

$$(I - D_b^2)^{-1} - I = \operatorname{diag}\left(\frac{e^{-b^{-1}i2r_0}}{1 - e^{-b^{-1}i2r_0}},\dots,\frac{e^{-b^{-1}i2r_{p-1}}}{1 - e^{-b^{-1}i2r_{p-1}}}\right)$$

and therefore

$$\left\|(I - D_b^2)^{-1} - I\right\| \le \left|\frac{e^{-b^{-1}i2r_0}}{1 - e^{-b^{-1}i2r_0}}\right| \le \frac{e^{-b^{-1}2\sin(\pi/2p)}}{1 - e^{-b^{-1}2\sin(\pi/2p)}} \le C(p)e^{-b^{-1}\sin(\pi/2p)},$$

where $C(p)$ may be taken to be $e^{-\sin(\pi/2p)}/(1 - e^{-2\sin(\pi/2p)})$ for $0 < b \le 1$. □

REFERENCES

BENEDETTI, J. K. (1977). On the nonparametric estimation of regression functions. *J. Roy. Statist. Soc. Ser. B* **39** 248–253.

CODDINGTON, E. and LEVINSON, N. (1955). *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.

COX, D. (1983). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

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

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

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

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. and MÜLLER, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 197–211.

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

MESSER, K. (1991). A comparison of a spline estimate to its "equivalent" kernel estimate. *Ann. Statist.* **19** 817–829.

MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.* **46**. Springer, New York.

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

RICE, J. (1983). Boundary modification for kernel regression. *Comm. Statist. A—Theory Methods* **13** 893–900.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1841.

RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

SILVERMAN, B. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** 898–916.

SPECKMAN, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. Technical report, Dept. Statistics, Univ. Missouri, Columbia.

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

DEPARTMENT OF MATHEMATICS
CALIFORNIA STATE UNIVERSITY
FULLERTON, CALIFORNIA 92634-4080

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113