Useful results concerning analysis of both high-degree and also high-dimension local polynomial estimators are in Ruppert and Wand (1992).

A forthcoming manuscript by Fan, Gasser, Gijbels, Brockmann and Engel shows that the efficiency of the local polynomial regression fit is good even for the estimation of quite high derivatives and that the local polynomial fits yield minimax efficient linear smoothers for estimating the regression function as well as its derivatives. Moreover, it is seen that for all polynomial degrees and estimation of any derivative, the optimal kernel is still the familiar Epanechnikov kernel. This answer is much simpler than the complicated case wise solutions developed for kernel estimation in Gasser, Müller and Mammitzsch (1985), for example.

7.2 Open Questions

Here is a summary of the open problems discussed above.

1. What is the best way to compute local polynomial estimators?
2. Are local polynomials competitive with smoothing splines in terms of speed?
3. Which degree of polynomial should be used?
4. Is it really better to estimate derivatives by the appropriate coefficient, rather than by differentiating an estimator of the regression?

7.3 Closing Quote

As Theodor Gasser has said (in private conversation): "We have not found any disadvantages of the local polynomial method as yet. It should become a golden standard nonparametric technique."

ACKNOWLEDGMENT

Research of both authors was supported in part by NSF Grant DMS-92-03135.

Comment

Hans-Georg Müller

1. INTRODUCTION

The article by Hastie and Loader (H&L) clearly demonstrates the importance of choosing a good smoothing method in the nonparametric regression context. The authors provide important insights and further strengthen the case for the "Local Weighted Least Squares" (LWLS) method. This article is a continuation of the extensive discussion of Chu and Marron (1991) who compared various aspects of different kernel regression smoothers but did not include LWLS.

It can be argued that LWLS is a third type of kernel method, generalizing the Nadaraya–Watson (NW) approach. When discussing kernel smoothing, one may want to refer to a broader perspective which includes not only nonparametric regression as probably the most important application but also the estimation of density, spectral density, hazard, intensity, quantile density and other functions. It is then useful to have a general framework available which provides for the construction of kernels, boundary kernels and bandwidth selectors for a whole range of smoothing problems. Such a framework can be provided for "explicit" kernel methods, including Parzen–Rosenblatt kernel estimates in the density and Nadaraya–Watson, Priestley–Chao or Gasser–Müller kernel estimates in the regression context. It is not clear whether LWLS could be included in such a framework, as it is uniquely geared toward regression.

The LWLS method is of particular interest for change-point modeling (Section 5), owing to its extraordinary flexibility which allows, for instance, constructing local fits satisfying linear constraints within the local regression model. Moreover, many well-studied features of (global) linear model fits can be extended to local linear models, like testing of linear hypotheses, diagnostics, local goodness-of-fit, modeling of correlation structure and heteroscedasticity and so on. Along with the many desirable features demonstrated by H&L, this makes LWLS a very attractive option for smoothing.

We should not, however, overly rely on a single method for all possible nonparametric regression problems. It is clear that a fixed bandwidth LWLS method has problems with smoothing data like those presented in Figure 6 of H&L: the "holes" in the data may lead to inappropriate zero valued or undefined regression estimates. Window and bandwidth choices adapting to design nonuniformities are needed in such cases. This may lead to a fairly complicated smoother, so that some of the initial simplicity is lost for highly nonuni-

Hans-Georg Müller is Professor, Division of Statistics, University of California, Davis, California 95616.

form designs. While for instance nearest neighbor type bandwidth choices (Cleveland, 1979) are a possible approach to overcome some of these problems, they are difficult to apply in higher dimensions and may be asymptotically inefficient. In contrast, Gasser–Müller (GM) type smoothers $\hat{f}_{GM}(x, b)$ using a fixed bandwidth b have the desirable property that even as $b \to 0$, "holes" will not open up. Instead, for any fixed n, irrespective of the design,

$$\lim_{b \to 0} \hat{f}_{GM}(x, b) = \sum_{j=1}^{n} Y_j 1_{A_j}(x),$$

the r.h.s. being an "empirical regression function" with A_j's as defined in Section 6 of H&L, while for LWLS, $\hat{f}_{LS}(x, b) \to 0$ or is undefined as $b \to 0$, whenever x is not a design point.

Another cautionary note regarding "universal" regression smoothing relates to the dichotomy random versus fixed design regression. While in the former case, where the predictors are assumed to be i.i.d. random variables, LWLS outperforms both of the other (NW and GM) kernel type estimates asymptotically (Fan, 1992, 1993), this is not the case in the fixed design regression model, where for designs generated by regular design densities both LWLS and GM estimates are asymptotically equivalent; this equivalence extends to the boundary area as well (see Section 4 below).

In many specialized large sample applications with fixed designs like smoothing of pixelled image data or of "preaveraged" or "binned" data, where the data are averages formed over regular "bins" (compare Azari and Müller, 1992, or Härdle and Scott, 1992) the design is actually equidistant and in such situations the three kernel methods (including LWLS) behave nearly identically. Other criteria like personal preference or numerical efficiency are then relevant. For data which are not necessarily fixed design or are not coming from a regular design, LWLS (perhaps also related methods like IRENORE, see Section 7) are the method of choice. However, the user still needs to decide whether the data come from a fixed or random design whenever confidence statements and variance estimates are required, as these depend essentially on the type of design. An avoidance strategy to deal with this issue is to interpret the results even in the random design case "conditionally" on the predictors, which essentially amounts to assuming a fixed design.

2. LOCAL WEIGHTED LEAST SQUARES

After some earlier foundational work on LWLS, interest in the LWLS method was renewed recently by the findings of Fan (1992, 1993), and H&L's article will most certainly stimulate further interest.

This method is commonly motivated by shrinking the domain of a linear or polynomial regression to a local window. Another intuitive approach to the LWLS method is to start with a linear smoother $\hat{f}(x) = \Sigma l_j(x) Y_j$. How should one choose the weight functions $l_j(\cdot)$? Requiring that the method be "local," let $l_j(x) = 0$, if $| x_j - x | > b$, where b is a bandwidth (smoothing parameter). Further, requiring that polynomials up to order $(k-1)$ for some integer $k \geq 1$ are to be estimated without bias, let

$$\sum l_j(x)(x_j - x)^i = 0, \quad 0 < i < k,$$
$$\sum l_j(x) = 1.$$

Under these requirements, one could try to minimize the variance of $\hat{f}$. As $\operatorname{var}(\hat{f}(x)) \sim \Sigma l_j^2(x)$ for uncorrelated homoscedastic errors, the problem amounts to minimizing $\Sigma l_j^2(x)$ subject to the above linear constraints.

One obtains $\hat{f}_{LS}(x) = \hat{\beta}_0$, where

$$(1) \quad \begin{aligned}&(\hat{\beta}_0, \ldots, \hat{\beta}_{k-1}) \\ &\quad = \underset{(\beta_0, \ldots, \beta_{k-1})}{\arg\min} \sum_{j=1}^{n} G\left(\frac{x - x_j}{b}\right)\left\{Y_j - \sum_{l=0}^{k-1} \beta_l (x - x_j)^l\right\}^2,\end{aligned}$$

and $G(\cdot) = 1_{[-1,1]}$. This is an unweighted local least squares estimator, which has undesirable smoothness properties and is not optimal with regard to mean squared error (MSE). Using a more general "kernel" function $G \geq 0$ with support $[-1, 1]$, one obtains LWLS. This estimator can be viewed as a third type of kernel smoother; the close connection to kernel methods is also evident by deriving the NW estimator as a LWLS smoother, locally fitting constants.

What is a good choice of G? $G \equiv 1_{[-1,1]}$ is numerically efficient, as updating formulae are available for this case, but leads to aesthetically unappealing jittery estimates. From the MSE point of view, the optimal "kernel" function is the Bartlett–Priestley weight $G(x) = (1 - x^2)1_{[-1,1]}$. Choice of smoother "kernel" functions, such as $G(x) = (1 - x^2)^{\mu} 1_{[-1,1]}$, $\mu > 1$, will lead to smoother resulting curve estimates at the expense of increased variance. Choice of even k (i.e., fitting locally linear, cubic etc. polynomials rather than quartic, quadratic, etc.) corresponds to symmetric "equivalent" kernels (see Section 3) and therefore may be advantageous in terms of bias.

For nonnormal situations like Poisson regression, local generalized linear models are a natural extension [compare Staniswalis (1989) for local likelihood methods]. Any "global" regression routine which allows for case weight specification by the user can be "localized" by successively feeding the data into local moving windows plus providing case weights (the same weights as one would use for LWLS). This feature allows for great flexibility and convenience in designing and implementing particular variations of LWLS.

Note that LWLS weights $l_j(x)$ may also be an attractive choice to define a version of empirical conditional distribution functions $F_n(y \mid x) = \Sigma l_j(x) 1_{\{Y_j \le y\}}$ and corresponding M-estimators and estimators $\int \psi(y) dF_n(y \mid x) = \Sigma l_j(y) \psi(Y_j)$ of functionals $\Sigma \psi(y) dF(y \mid x)$ of conditional distributions, where for instance $\psi(y) = y$ corresponds to LWLS regression, $\psi(y) = (y - \mu)^2$, $\mu = \int y dF$ to variance function estimation.

3. "EQUIVALENT" KERNELS

Usually, the LWLS "kernel" G will be symmetric; it is not a "real" kernel function as it does not satisfy the usual moment conditions (of a second-order kernel). For regular (not necessarily uniform) designs generated by a design density, $\hat{f}_{LS}$ is asymptotically equivalent to a version of $\hat{f}_{GM}$ (Müller, 1987). The "equivalent" kernel employed by the "equivalent" $\hat{f}_{GM}$ is

$$K_{EQ} \equiv G p_{k-1}, \tag{2}$$

where p_{k-1} is a polynomial of degree $k - 1$ which is uniquely determined by the moment conditions

$$\int K_{EQ}(x) x^j dx = 0, \quad 0 < j < k, \quad \int K_{EQ}(x) dx = 1, \tag{3}$$

implying that K_{EQ} is a kernel of order k. Figure 1 demonstrates that $\hat{f}_{GM}$ with this kernel K_{EQ} and $\hat{f}_{LS}$ are indistinguishable also for finite samples for $G(x) = (1 - x^2) 1_{[-1,1]}$ and the "equivalent" Bartlett–Priestley–Epanechnikov kernel $K_{EQ}(x) = (3/4)(1 - x^2) 1_{[-1,1]}$. The weights $l_j(0.5)$ when estimating at $x = 0.5$ with $b = 0.5$, distributed to measurements at 201 equidistant design points in [0, 1] are seen to be identical for both methods.

What are the "equivalent" kernels when estimating derivatives? According to (2) above, extended to cover derivatives, one finds for the first derivative and $G = (1 - x^2) 1_{[-1,1]}$ the equivalent kernel $K_{EQ}(x) = (15/4)(x^3 - x) 1_{[-1,1]}$ (the "optimal" kernel for estimating the first derivative under minimal sign changes).

If one would consider the obvious alternative, namely, to estimate a derivative by direct differentiation of the weight functions l_j, that is, $\tilde{f}_{LS}^{(1)}(x) = \Sigma l_j^{(1)}(x) Y_j$, where $\hat{f}_{LS}(x) = \Sigma l_j(x) Y_j$, then this would correspond to applying the derivative of the Bartlett–Priestley–Epanechnikov kernel, which is the "minimum variance kernel" $K_{\mathrm{MinVar}}(x) = -(3/2)x \, 1_{[-1,1]}$. For the same reasons as discussed in Gasser and Müller (1984), it seems therefore to be indeed preferable to use $\hat{f}_{LS}^{(1)}(x) = \hat{\beta}_1(x)$ as advocated by H&L, rather than $\tilde{f}_{LS}^{(1)}(x)$.

Weights l_j for estimating $f^{(1)}$ (0.5) with bandwidth $b = 0.5$ on an equidistant design of 201 points in [0, 1] are shown for $\hat{f}_{GM}^{(1)}$ using K_{EQ} and K_{MinVar} and for $\hat{f}_{LS}^{(1)}$ in Figure 2. It is evident that the large sample equivalence between $\hat{f}_{GM}^{(1)}$ with K_{EQ} and $\hat{f}_{LS}^{(1)}$ materializes for this sample size.

4. BOUNDARY KERNELS

Even if one opts for LWLS as nonparametric regression technique with its "automatic" boundary adaptation feature, boundary kernels or one of the alternative methods by Rice (1984), Schuster (1985) or Hall and Wehrly (1991) (which can be expressed and understood in terms of "equivalent" boundary kernels) are needed for other curve estimation problems. For all these methods, including LWLS, the usual balance between bias and variance is disturbed near endpoints due to bias adaptation which increases variances (Müller, 1991, 1992; Müller and Zhou, 1991; Müller and Wang, 1992). Therefore, boundary adaptive bandwidth choices like stabilizing the window size at $2b$ (corresponding to varying bandwidths according to $b(x) = 2b - x$, $x \in [0, b)$ if 0 is a left endpoint) can lead to significant

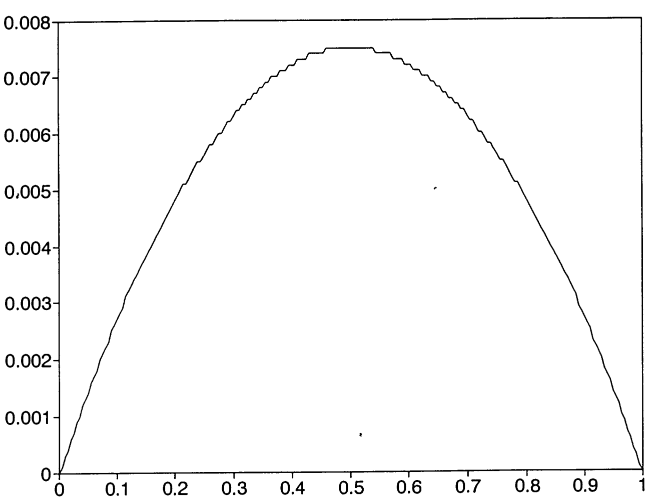

FIG. 1. *Weights distributed by LWLS (solid) and $\hat{f}_{GM}$ with equivalent kernel K_{EQ} for $k = 2$, when estimating at $x = 0.5$ and 201 points in the window. The two graphs coincide.*

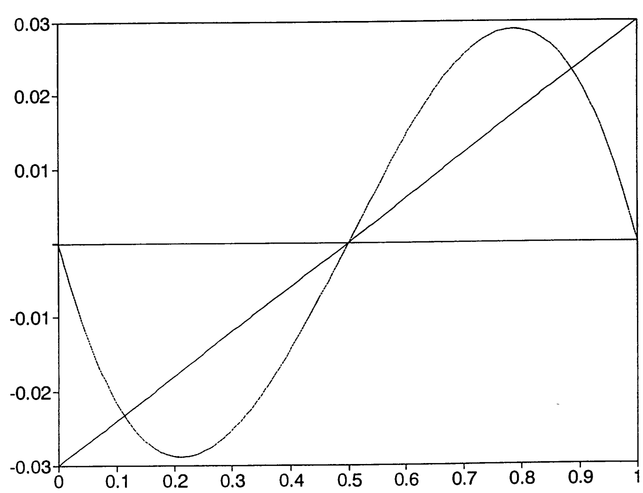

FIG. 2. *Same as Figure 1 but weights for first derivative. LWLS $\hat{f}_{LS}^{(1)}$ dotted, $\hat{f}_{GM}^{(1)}$ with K_{EQ} dashed, $\hat{f}_{GM}^{(1)}$ with K_{MinVar} solid. Dashed and dotted graphs coincide.*

improvements. Thus the boundary problem becomes a bandwidth problem for LWLS which cannot be ignored whenever structure near boundaries matters. In such cases, the "automatic" boundary adaptation feature of LWLS should not tempt the user to assume that the problem is solved.

Relation (2) can be extended to find "equivalent" boundary kernels. We replace $G = G1_{[-1,1]}$ in (2) by a truncated version $G_q = G1_{[-1,q]}$, where $q = x/b$, $x \in [0, b)$, and obtain then the resulting kernel for the case of fitting local lines, $K_{EQ,q}(x) = G_q(x)(\alpha_{0,q} + \alpha_{1,q}x)$. Here, $\alpha_{0,q}$ and $\alpha_{1,q}$ are calculated such that $K_{EQ,q}$ satisfies moment conditions (3) for $k = 2$. Explicit solutions are available. Kernels $K_{EQ,q}$ were suggested in Gasser and Müller (1979) as a simple and effective way to construct boundary kernels, and the "equivalence" property with LWLS was discussed in Lejeune (1985), Ruppert and Wand (1992).

Another class of "smooth optimum" boundary kernels $K_{SO,q}$ was derived in Müller (1991) for kernel of arbitrary orders k (and arbitrary derivatives) by extending a variational problem, the solution of which are MSE "optimal" kernels under sign change restrictions, to the boundary: for $k = 2$,

$$K_{SO,q}(x) = \frac{6(1+x)(q-x)}{(1+q)^3} \cdot \left\{1 + 5\left(\frac{1-q}{1+q}\right) + 10\frac{1-q}{(1+q)^2}x\right\}.$$

The unified framework within which these kernels are derived requires that smoothness conditions be imposed at both endpoints of the support of the kernel. Under this restriction, these kernels are "smooth optimal" in a certain sense. Without this restriction, they are not MSE optimal. In fact, the smoothness requirement at the endpoint increases the variation of these kernel functions and therefore the variance of resulting curve estimates.

This problem was recognized in Müller and Wang (1992), where a new class of boundary kernels based on expansions in orthogonal polynomials was proposed: for $k = 2$,

$$K_{MW,q}(x) = \frac{12}{(1+q)^4}(x+1) \cdot [x(1-2q) + (3q^2 - q + 1)/2].$$

As these kernels are not subject to smoothness constraints, they have less variation than boundary kernels $K_{SO,q}$ and therefore better MSE properties. The advantage of using boundary kernels $K_{MW,q}$ or $K_{SO,q}$ over $K_{EQ,q}$ is that closed formulas are available for all orders k, whereas for the construction of kernels $K_{EQ,q}$ for $k > 3$, a linear system of equations has to be solved numerically for each point in the boundary region where a function estimate is desired. This disadvantage does not apply in the case $k = 2$. Note that $K_{SO,1} \equiv K_{MW,1} \equiv K_{EQ,1}$ if $G(x) = (1 - x^2)^\mu 1_{[-1,1]}$ and $k = 2$, for any $\mu \geq 0$, which means that these boundary kernel constructions coincide in the "interior" and correspond there to a nonnegative polynomial kernel function (for $k = 2$). When comparing $\hat{f}_{LS}$ with $\hat{f}_{GM}$ in the boundary area, $\hat{f}_{GM}$ should be constructed with kernels $K_{EQ,q}$.

The weights $l_j(\cdot)$ distributed by $\hat{f}_{GM}$ when estimating in the boundary area using boundary kernels $K_{EQ,q}$, $K_{SO,q}$ and $K_{MW,q}$ and the corresponding weights distributed by LWLS are shown for $b = 0.5$, estimating at the endpoint $x = 0$ itself (corresponding to $q = 0$) in Figure 3 and at $x = 0.25$ (corresponding to $q = 0.5$) in Figure 4. It is obvious that the weights belonging to $\hat{f}_{GM}$ with $K_{EQ,q}$, and those of $\hat{f}_{LS}$ are almost indistinguishable, therefore also bias and variance properties of these two methods will be the same. Very close comes $\hat{f}_{GM}$ with boundary kernels $K_{MW,q}$, whereas weights of $\hat{f}_{GM}$ with boundary kernels $K_{SO,q}$ are substantially different and from a MSE point of view less desirable. Kernels $K_{MW,q}$ are easy to compute for various orders and derivatives and have good MSE properties. Kernels $K_{EQ,q}$ are a good choice for order $k = 2$.

Note that application of kernels $K_{SO,q}$ is "equivalent" to the following alternative LWLS procedure to handle boundaries: $\hat{f}_{LS,alt}(x) = \hat{f}_{LS}(b) + \hat{f}_{LS}^{(1)}(b)(x - b)$ for x in the left boundary area, $x \in [0, b)$.

5. REGRESSION FUNCTIONS WITH CHANGE-POINTS

Assume that the regression function f to be estimated has a change-point at τ in the form of a jump discontinuity

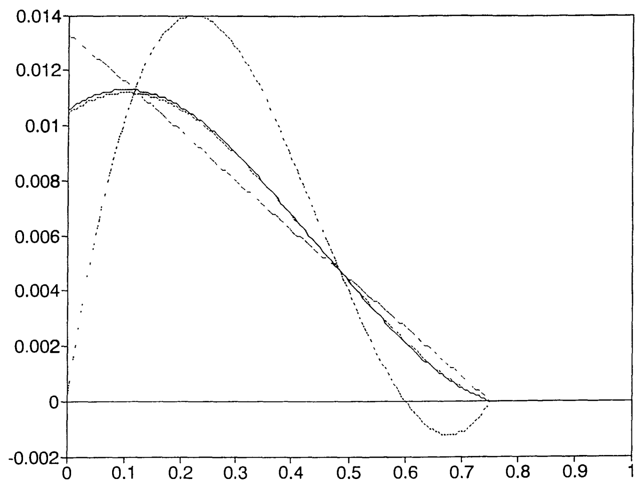

FIG. 3. *Boundary weights distributed by LWLS (solid), by $\hat{f}_{GM}$ with $K_{EQ,q}$ (dotted) i.e., the curve closest to the solid curve, $K_{MW,q}$ (dash/dot) i.e., the curve with the largest value at 0 and $K_{SO,q}$ (short dashed) i.e., the curve with the largest maximum and a value of 0 at 0, when estimating at left endpoint $x = 0$ ($q = 0$), $b = 0.5$.*

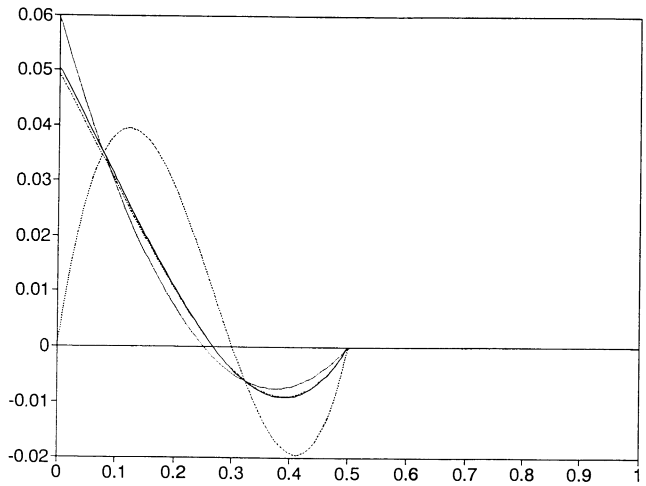

FIG. 4. *Same as Figure 3 when estimating at $x = 0.25 (q = 0.5)$. Note that the graph corresponding to weights distributed by $\hat{f}_{GM}$ with $K_{MW,\,0.5}$ has the largest value at 0 among the four graphs, while the graph corresponding to $\hat{f}_{GM}$ with $K_{SO,\,0.5}$ has a value of 0 at 0.*

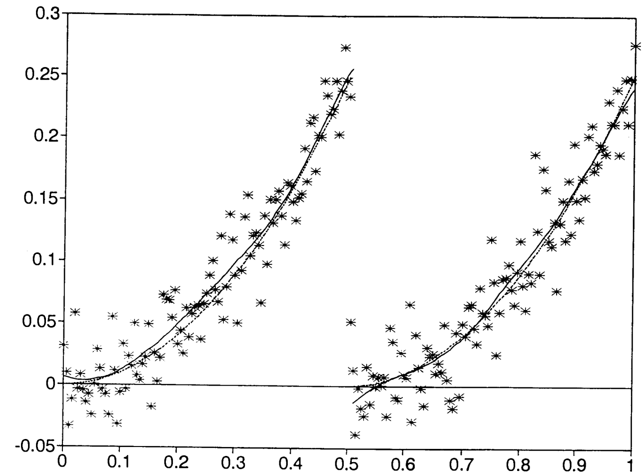

FIG. 5. *Estimation of the regression function in a change-point model (4). LWLS (solid) with $b = 0.1$, true curve (dashed) and scatterplot ($\sigma = 0.025$). Estimated change-point at $\hat{\tau} = 0.505$.*

(4) $$f(x) = f_1(x)1_{\{0 \le x < \tau\}} + f_2(x)1_{\{\tau \le x \le 1\}},$$

where f_1, f_2 are smooth functions with $|f_1(\tau) - f_2(\tau)| = \Delta_0 > 0$. Another possibility is a jump discontinuity in the derivative

(5) $$f^{(1)}(x) = f_1^{(1)}(x)1_{\{0 \le x < \tau\}} + f_2^{(1)}(x)1_{\{\tau \le x \le 1\}},$$

where $f_1^{(1)}$, $f_2^{(1)}$ are smooth with $|f_1^{(1)}(\tau) - f_2^{(1)}(\tau)| = \Delta_1 > 0$.

Undiscriminating application of smoothers to data coming from functions in models (4) or (5) will oversmooth the change-point and lead to inconsistent estimates of f at τ in model (4) and of $f^{(1)}$ in model (5). A two-step procedure using LWLS which works well for model (4) is as follows:

(6)
(a) Fit the linear four parameter model $[\beta_0 + \beta_1(x - t)]1_{\{t-b \le x \le t\}} + [\beta_3 + \beta_4(x - t)]1_{\{t \le x \le t+b\}}$ with change-point at t by LWLS within each window $[t - b, t + b]$, and find the estimate $\hat{\tau} = \arg\max_t |\hat{\beta}_3(t) - \hat{\beta}_0(t)|$ for τ. Obtain the curve estimate $\hat{f}$ as follows:
(b) If $\hat{\tau} \notin [t - b, t + b]$, fit the line $[\beta_0 + \beta_1(x - t)]$ by LWLS to obtain $\hat{f}(t) = \hat{\beta}_0(t)$. If $\hat{\tau} \in [t - b, t + b]$, fit the model $[\beta_0 + \beta_1(x - t)]1_{\{t-b \le x < \hat{\tau}\}} + [\beta_3 + \beta_4(x - t)]1_{\{\hat{\tau} \le x \le t+b\}}$ by LWLS and evaluate at t to obtain $\hat{f}(t)$.

For model (5) the procedure is analogous, except that one would fit in step (a) models $\beta_0 + \beta_1(x - t) + \beta_2(x - t)1_{\{t \le x \le t+b\}}$, define $\hat{\tau} = \arg\max_t |\hat{\beta}_2(t)|$ and fit in step (b) models $\beta_0 + \beta_1(x - t) + \beta_2(x - t)1_{\{\hat{\tau} \le x \le t+b\}}$ by LWLS, evaluating at t to obtain $\hat{f}(t)$.

This procedure is similar to proposals by Müller (1992a) where change-point estimates are based on differences of one-sided kernel smoothers, and by Eubank and Speckman (1992) who adopt a semiparametric approach. For the situation of a known change-point in model (4), compare also Eubank and Speckman (1991). An application to simulated data is shown for model (4) in Figure 5 for the function $f(x) = x^2 1_{[0,\,0.5]} + (x - 0.5)^2 1_{[0.5,\,1]}$ with $\Delta_0 = 0.25$ and for model (5) in Figure 6 for the function $f(x) = x^2 1_{[0,\,0.5]} + (x^2 - 0.25)1_{[0.5,\,1]}$ with $\Delta_1 = 1.0$ (error std 0.025, $n = 201$). In both cases, the underlying function including the change-point is seen to be well reproduced.

6. SPATIAL SMOOTHING

Some smoothing problems in spatial applications involve measurements which are not associated with fixed locations but rather with sets. An example are the Sudden Infant Death Syndrome (SIDS) data ana-

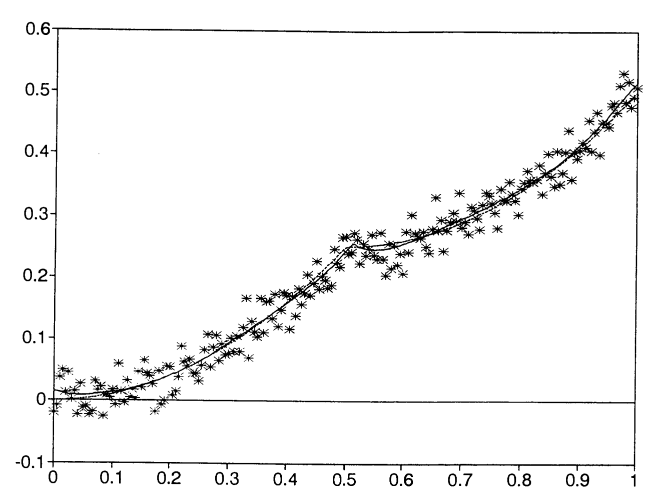

FIG. 6. *Same as Figure 5 in a change-point model (5). Estimated change-point at $\hat{\tau} = 0.51$.*

lyzed in Cressie (1991), where data are only available averaged over counties, with counties of quite irregular shapes. In this case the data are aptly described by the "empirical regression function" $\Sigma Y_j 1_{A_j}$, where sets A_j represent the counties. Obviously, GM type estimators are directly applicable to such data, while NW and LWLS estimators are not.

A generalized version of LWLS which is capable of handling data like this is as follows: consider $\tilde{f}_{LS}(x) = \tilde{\beta}_0(x)$, where

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \int [y - (\beta_0 + \beta_1^t(t - x))]^2 \cdot G\left(\frac{t_1 - x_1}{b}, \ldots, \frac{t_p - x_p}{b}\right) dF(x,y) \tag{7}$$

for any probability measure dF on $\mathbb{R}^p \times \mathbb{R}$, $p \geq 1$.

If F is the joint distribution of measurements (X_i, Y_i), $X_i \in \mathbb{R}^p$, $Y_i \in \mathbb{R}$, this yields the convolution of the true regression function with the bandwidth-scaled kernel K_{EQ} of (2). If $dF \equiv dF_n \equiv n^{-1} \Sigma\delta_{\{X_i, Y_i\}}$, the empirical measure, then $\tilde{f}_{LS}(x)$ is the ordinary LWLS estimator.

Let now (A_i) be a partition of the compact domain A into measureable sets $A_i \subset A$, $1 \leq i \leq n$. Choosing $dF = \Sigma[\lambda(A_i)/\lambda(A)]\{dF_{U,A_i} \times \delta_{Y_i}\}$, where F_{U,A_i} is the uniform distribution on A_i, δ_y an atom of mass 1 at y and λ the Lebesgue measure, then one obtains

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \frac{1}{\lambda(A)} \sum_{i=1}^{n} \int_{A_i} [y_i - \beta_0 + \beta_1(t - x)]^2 \cdot G\left(\frac{t_1 - x_1}{b}, \ldots, \frac{t_p - x_p}{b}\right) dx.$$

This special case of (7) yields the proposed LWLS type estimator $\tilde{f}_{LS}(x) = \tilde{\beta}_0(x)$ for this spatial smoothing problem.

7. OTHER SMOOTHING METHODS

As mentioned by H&L, other important and popular nonparametric regression estimators are smoothing and regression splines (see Eubank, 1988, or Wahba, 1990), and it would be of interest to see how these compare with the other methods. It follows from results of Silverman (1984) that for regular designs smoothing splines are asymptotically equivalent to GM estimators with certain "equivalent" kernels with noncompact supports and bandwidths which vary locally according to the design density; compare also Messer and Goldstein (1993) who investigated corresponding boundary kernels. We may therefore expect that at any fixed point, even for finite n, the behavior of smoothing splines will be closer to GM estimators than to the other kernel methods. The design-adaptive local bandwidth variation feature of smoothing splines is a bonus. For other smoothing methods, such local bandwidths variation can be implemented as well, but only at the expense of substantial additional conceptual and numerical complexity.

One advantage of controlled local bandwidth variation for kernel and LWLS estimates, however, is that this may include adaptations not only to locally varying design density but also to curvature and heteroscedasticity (Müller and Stadtmüller, 1987; Fan and Gijbels, 1992a).

Considering the random design case, the NW estimator which is a special case of LWLS has serious drawbacks as pointed out by H&L and Chu and Marron (1991). One of the more serious problems is that this estimator cannot reproduce straight lines as regression functions (in the univariate case) when the marginal density of the predictors X_i is nonuniform. An identity reproducing transformation, which is applicable to any nonparametric regression estimator, was introduced in Müller and Song (1991) to address this problem. A corresponding identity reproducing nonparametric regression estimator (IRENORE) derived from the NW estimator then has the same asymptotic MSE properties as LWLS in random designs. It is thus another approach which achieves the desirable MSE properties of LWLS.

ACKNOWLEDGMENT

Research supported by Air Force Grant AFOSR 89-0386.

Rejoinder

Trevor Hastie and Clive Loader

If the "smoothing community" includes the users of smoothers, then local regression has been popular for more than 10 years. In particular, Cleveland's (1979) implementation is widely used across a broad spectrum of disciplines and is appreciated for its simple but effective approach to boundary bias, local bandwidth