The Annals of Statistics
2000, Vol. 28, No. 5, 1339–1355

DATA SHARPENING METHODS FOR BIAS REDUCTION IN NONPARAMETRIC REGRESSION

BY EDWIN CHOI, PETER HALL AND VALENTIN ROUSSON[1]

Australian National University, Australian National University and CSIRO and Australian National University

We consider methods for kernel regression when the explanatory and/or response variables are adjusted prior to substitution into a conventional estimator. This "data-sharpening" procedure is designed to preserve the advantages of relatively simple, low-order techniques, for example, their robustness against design sparsity problems, yet attain the sorts of bias reductions that are commonly associated only with high-order methods. We consider Nadaraya–Watson and local-linear methods in detail, although data sharpening is applicable more widely. One approach in particular is found to give excellent performance. It involves adjusting both the explanatory and the response variables prior to substitution into a local linear estimator. The change to the explanatory variables enhances resistance of the estimator to design sparsity, by increasing the density of design points in places where the original density had been low. When combined with adjustment of the response variables, it produces a reduction in bias by an order of magnitude. Moreover, these advantages are available in multivariate settings. The data-sharpening step is simple to implement, since it is explicitly defined. It does not involve functional inversion, solution of equations or use of pilot bandwidths.

1. Introduction. The term "data sharpening" refers to methods for preprocessing data so that, when they are substituted into a conventional estimator, performance is improved relative to what it would be if the raw data were employed. Of course, there are often other ways of improving performance, usually involving changing the construction of the estimator. Data sharpening recognizes that the conventional estimator had certain advantages, as well as shortcomings, and attempts to retain the former, even enhancing them, while adjusting the data so as to overcome the latter. It also provides insight into the reasons for poor performance of standard methods—we must understand deficiencies in order to rectify them.

In this paper we suggest data sharpening methods for reducing the bias, and in some instances improving performance in other respects, of standard nonparametric estimators of a regression mean. The methods are available in a multivariate setting and involve simple, explicitly defined adjustments to the data. We explore a range of different approaches, applicable to Nadaraya–Watson and local-linear estimators, and conclude that one of them in

Received March 1999; revised May 2000.

[1]Supported by Grant 81NE-54413 from the Swiss National Science Foundation.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62H05.

Key words and phrases. Bandwidth, curse of dimensionality, design sparsity, explanatory variables, kernel methods, local-linear estimator, local-polynomial methods, Nadaraya–Watson estimator, response variables, smoothing.

particular, designed for the local-linear case, has advantages that make it very attractive. It involves slightly shifting the design points (i.e., the explanatory variables) so that they become more concentrated in places where the original design density had been relatively low and more sparse where the concentration had been high. As a result, the estimator computed from the sharpened data is less susceptible to difficulties arising from design sparsity. At the same time, the estimator has an order of magnitude less bias [in fact, $O(h^4)$ rather than $O(h^2)$, where h denotes bandwidth] than the original local-linear estimator, except in the near vicinity of the boundary and has the same order of bias as the local-linear approach close to the boundary. The order of variance is unchanged.

Conventional approaches to bias reduction, based for example on high-order polynomials or high-order kernels, suffer more from design sparsity than the techniques that they are endeavouring to improve. In particular, high-order local polynomial methods involve more parameters than local linear methods and so demand more data in the local neighborhood for adequate fitting. As a result, they do not realize their theoretical gains unless sample size is relatively large; the requirement for more data in the local neighborhood is manifested in greater susceptibility to the curse of dimensionality. This is not the case for our bias-reduced estimator $\hat{g}_{\mathrm{LL},1}$, which performs relatively well for small samples and is the most significant of the new estimators suggested in this paper. In the case of that estimator, our method works simultaneously to reduce bias and improve resistance against design sparsity.

More generally than this particular approach, we explore a variety of different forms of data sharpening. We show that asymptotic performance of the Nadaraya–Watson estimator can be made equal to that of the local-linear method by moving design points closer together in places where the design density is high and further apart where it is low. The resulting estimator may be shown to be first-order equivalent to its local-linear counterpart. The fact that it is clearly more vulnerable to design sparsity than the original Nadaraya–Watson estimator provides an explanation of the relatively poor performance of the local-linear method in this respect.

A second approach, which involves adjusting response variables rather than explanatory variables, reduces the bias of the Nadaraya–Watson estimator by an order of magnitude; and a third method, which tweaks both explanatory and response variables before substituting them back into the conventional Nadaraya–Watson estimator, achieves bias reductions of the same order as the second method, but in a different manner. It was a version of the latter technique, although in the local-linear case, that we advocated earlier.

All our methods have counterparts for other approaches to nonparametric regression, for example for the convolution-type methods suggested by Priestley and Chao or Gasser and Müller [see, e.g., Wand and Jones (1995), pages 130–135]. We choose the Nadaraya–Watson and local linear methods for detailed analysis because they are arguably of greatest methodological interest. Indeed, there is a significant recent literature on techniques for improving the asymptotic performance of the Nadaraya–Watson estimator to that of

local linear methods. It includes work of Müller and Song (1993) and Mammen and Marron (1997) on "identity reproducing regression" or "mass centred smoothing," which involves applying data-dependent functional transformations to reduce the bias of the Nadaraya–Watson estimator. Müller (1997) has suggested an alternative technique which achieves the same end and is based on adjusting the density estimator in the denominator of the Nadaraya–Watson estimator. Hall and Presnell (1999) have considered a weighted bootstrap approach to making the Nadaraya–Watson estimator unbiased for linear functions; this is the principal source of the bias problems from which it suffers. Related contributions for other estimator types include those suggested by Hougaard (1988) and Hougaard, Plum and Ribel (1989).

Data-sharpening methods for density estimation have been discussed by Choi and Hall (1999), although in implementation they are different from those considered here. Methods that involve data shifting for purposes other than resubstitution into an estimator include those suggested by Boswell (1983), Fwu, Tapia and Thompson (1981) and Jones and Stewart (1997).

A potential drawback of our data-sharpening estimators is that, unlike local-linear methods, they are not unbiased for linear functions. As a result they may experience difficulties when estimating an approximately linear function that is particularly steep. Also, while the data-sharpening idea can be applied generally, some theoretical analysis or numerical experimentation would be needed to determine the best approach in any given setting.

2. Methodology.

2.1. *Definitions of basic estimators.* Suppose independent data vectors (X_i, Y_i) for $1 \le i \le n$ are generated from a multivariate distribution. The X_i's are assumed to be p-vectors, the Y_i's are scalars, and we wish to estimate the expected value of Y given X; that is, $g(x) = E(Y|X = x)$. Nadaraya–Watson and local-linear estimators of g are

$$\hat{g}_{\mathrm{NW}}(x) = \frac{\sum_i Y_i K_i(x)}{\sum_i K_i(x)}, \qquad \hat{g}_{\mathrm{LL}}(x) = e_1^T \{X(x)^T W(x) X(x)\}^{-1} X(x)^T W(x) Y, \tag{2.1}$$

respectively, where $K_i(x) = K\{(x - X_i)/h\}$, e_1 is the $(p+1)$-vector with 1 in the first position and 0 elsewhere, $W = \mathrm{diag}(K_i)$, $Y = (Y_1, \ldots, Y_n)^T$,

$$X(x) = \begin{pmatrix} 1 & (X_1 - x)^T \\ \vdots & \vdots \\ 1 & (X_n - x)^T \end{pmatrix},$$

K is a nonnegative kernel (a function of p variables), and h is a bandwidth. For discussion of local-linear smoothing for multivariate design, see, for example, Ruppert and Wand (1994) and Fan and Gijbels (1996).

Both $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{LL}}$ have biases of order h^2, but the bias of the latter is typically less in absolute value than that of $\hat{g}_{\mathrm{NW}}$, at least in minimax terms;

see Fan (1993). On the other hand, $\hat{g}_{\mathrm{NW}}$ is generally more resistant to sparse design than $\hat{g}_{\mathrm{LL}}$, not least because (unlike $\hat{g}_{\mathrm{LL}}$) it never takes the form of a nonzero number divided by zero. It also has the advantage of being always nonnegative if the Y_i's are nonnegative. In the context of local polynomial fitting, $\hat{g}_{\mathrm{LL}}$ is more resistant to sparse design than estimators that are based on fitting polynomials of second or third degree, which have biases of order h^4. This is particularly true in multivariate cases. We suggest data-sharpening methods for reducing the bias of $\hat{g}_{\mathrm{NW}}$ to that of $\hat{g}_{\mathrm{LL}}$, to first order, or reducing it to order h^4, and for reducing the bias of $\hat{g}_{\mathrm{LL}}$ to order h^4.

2.2. *Data-sharpening versions of* $\hat{g}_{\mathrm{NW}}$. Our first application of data sharpening to $\hat{g}_{\mathrm{NW}}$ is based on moving design variables a little closer together in places where the design density is high and a little further apart in places where it is low, so as to overcome difficulties caused by inaccurate estimation of the design density in both the numerator and the denominator of $\hat{g}_{\mathrm{NW}}$. (Estimation of the design density is implicit in the numerator and explicit in the denominator.) These difficulties are particularly familiar in simpler problems of density estimation, where they lead to density estimators being too low in peaks and too high in troughs.

This approach has some similarities to methods suggested for nonparametric density estimation by Choi and Hall (1999). However, in the latter case a reduced bandwidth must be used for the sharpening step; by way of contrast, when sharpening data for $\hat{g}_{\mathrm{NW}}$, the same bandwidth is used throughout.

Specifically, put

$$\widehat{X}_j = \frac{\sum_i X_i K_i(X_j)}{\sum_i K_i(X_j)}. \tag{2.2}$$

The first data-sharpening version of $\hat{g}_{\mathrm{NW}}$, which we denote by $\hat{g}_{\mathrm{NW},1}$, has the definition of $\hat{g}_{\mathrm{NW}}$ at (2.1) except that we replace the design points X_i by their sharpened form. That is,

$$\hat{g}_{\mathrm{NW},1}(x) = \frac{\sum_i Y_i \widehat{K}_i(x)}{\sum_i \widehat{K}_i(x)},$$

where $\widehat{K}_i(x) = K\{(x - \widehat{X}_i)/h\}$.

Our second version leaves the explanatory variables unchanged and instead adjusts the response variables. Specifically, put $\widehat{Y}_j = \hat{g}_{\mathrm{NW}}(X_j)$ and $\widetilde{Y}_j = 2Y_j - \widehat{Y}_j$, and let $\hat{g}_{\mathrm{NW},2}$ have the definition of $\hat{g}_{\mathrm{NW}}$, except that each Y_i is replaced by $\widetilde{Y}_i$,

$$\hat{g}_{\mathrm{NW},2}(x) = \frac{\sum_i \widetilde{Y}_i K_i(x)}{\sum_i K_i(x)}.$$

A third version alters both explanatory and response variables, as follows. Put $\breve{Y}_j = 2Y_j - \hat{g}_{\mathrm{LL}}(X_j)$, where we may choose to drop diagonal terms when

defining $\hat{g}_{\mathrm{LL}}(X_j)$. Then $\hat{g}_{\mathrm{NW},3}$ has the definition of $\hat{g}_{\mathrm{NW}}$, except that each (X_i, Y_i) is replaced by $(\widehat{X}_i, \breve{Y}_i)$,

$$\hat{g}_{\mathrm{NW},3}(x) = \frac{\sum_i \breve{Y}_i \widehat{K}_i(x)}{\sum_i \widehat{K}_i(x)}.$$

We shall argue in Section 4 that $\hat{g}_{\mathrm{NW},1}$ has properties similar to those of $\hat{g}_{\mathrm{LL}}$, while $\hat{g}_{\mathrm{NW},2}$ and $\hat{g}_{\mathrm{NW},3}$ have biases of order h^4, rather than h^2, except in the close vicinity of boundaries. All three estimators have asymptotic variances of order $(nh^p)^{-1}$.

2.3. *Data-sharpening versions of* $\hat{g}_{\mathrm{LL}}$. To construct our first sharpened version of $\hat{g}_{\mathrm{LL}}$ we move the X_i's in the direction opposite to that suggested in Section 2.2, and similarly translate the response variables Y_i, in order to counteract the major contributions to bias from both the numerator and denominator of $\hat{g}_{\mathrm{LL}}$. Specifically, put $\widetilde{X}_j = 2X_j - \widehat{X}_j$, where $\widehat{X}_j$ is as in Section 2.2, and let $\widetilde{Y}_j$ be as in Section 2.2. (Again, we may drop diagonal terms when defining $\widetilde{X}_j$ and $\widetilde{Y}_j$.) Then, $\hat{g}_{\mathrm{LL},1}$ has the definition of $\hat{g}_{\mathrm{LL}}$ at (2.1), except that we replace the data pairs (X_i, Y_i) by their sharpened form $(\widetilde{X}_i, \widetilde{Y}_i)$ throughout. That is,

$$\hat{g}_{\mathrm{LL},1}(x) = e_1^T \{\widetilde{X}(x)^T \widetilde{W}(x) \widetilde{X}(x)\}^{-1} \widetilde{X}(x)^T \widetilde{W}(x) \widetilde{Y},$$

where $\widetilde{W} = \mathrm{diag}(\widetilde{K}_i)$, $\widetilde{K}_i(x) = K\{(x - \widetilde{X}_i)/h\}$, $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)^T$ and

$$\widetilde{X}(x) = \begin{pmatrix} 1 & (\widetilde{X}_1 - x)^T \\ \vdots & \vdots \\ 1 & (\widetilde{X}_n - x)^T \end{pmatrix}.$$

The transformation that takes X_j to $\widetilde{X}_j$ moves the explanatory variables away from the mode or modes of their distribution, and so counteracts any problems that might be caused by design sparsity. This can be particularly beneficial in multivariate settings, where the "curse of dimensionality" can lead to greater difficulties through data sparsity than in one dimension. Panel (a) of Figure 1 illustrates this point. Arrows there show the way in which design points have been moved closer together in a place of relative sparsity, and also the corresponding changes that have been made to response variables.

Our second data-sharpened local linear estimator is the analogue of $\hat{g}_{\mathrm{NW},2}$. Let $\breve{Y}_i$ be as in Section 2.2. Then, $\hat{g}_{\mathrm{LL},2}$ has the definition of $\hat{g}_{\mathrm{LL}}$ except that each Y_i is replaced by $\breve{Y}_i$:

$$\hat{g}_{\mathrm{LL},2}(x) = e_1^T \{X(x)^T W(x) X(x)\}^{-1} X(x)^T W(x) \breve{Y},$$

where $\breve{Y} = (\breve{Y}_1, \ldots, \breve{Y}_n)^T$. Contrary to $\hat{g}_{\mathrm{LL},1}$, this form of data sharpening leaves the explanatory variables unchanged, as shown in panel (b) of Figure 1. Both $\hat{g}_{\mathrm{LL},1}$ and $\hat{g}_{\mathrm{LL},2}$ have biases of order h^4, rather than h^2, except in the

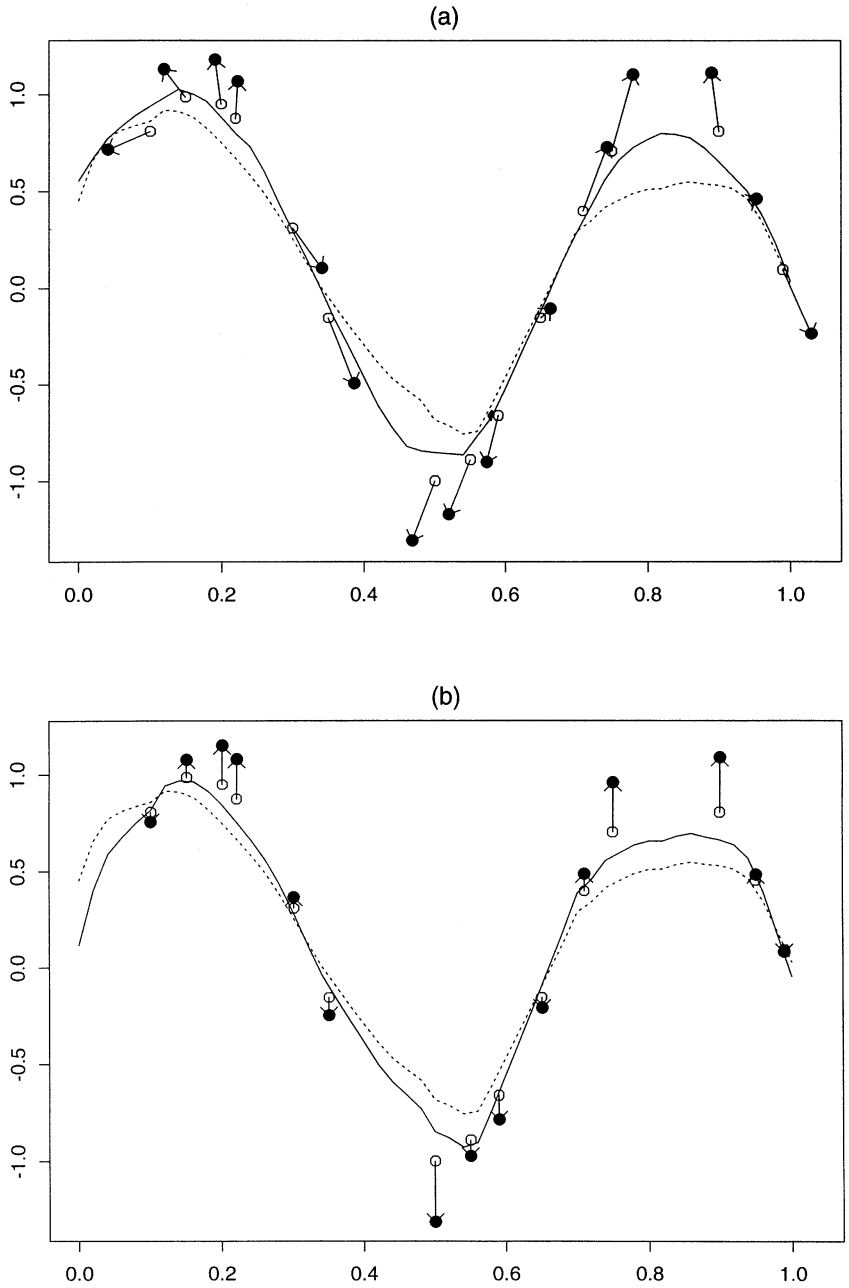

FIG. 1. *Geometric effects of data sharpening. Panels* (a) *and* (b) *indicate original data* (*white points*) *and sharpened data* (*black points*) *used for* $\hat{g}_{\mathrm{LL},1}$ *and* $\hat{g}_{\mathrm{LL},2}$, *respectively. For clarity, the original data contain no noise. Arrows indicate movements produced by the sharpening algorithm. The dotted and solid lines superimposed are the local linear estimates obtained from the original and the sharpened data, respectively. We observe that the latter curves are closer to the white points which represent the true curve. Additionally, the data sharpening illustrated in panel* (a) *has alleviated sparsity problems by moving design points closer together.*

close vicinity of boundaries; and in the case of $\hat{g}_{\mathrm{LL},2}$, bias is of order h^3 at boundaries.

Methods such as $\hat{g}_{\mathrm{NW},2}$ and $\hat{g}_{\mathrm{LL},2}$ lend themselves readily to iteration, enabling methods of arbitrarily high order to be obtained by repeatedly sharpening data prior to substitution into $\hat{g}_{\mathrm{NW}}$ or $\hat{g}_{\mathrm{LL}}$. It is not difficult to see that the size of variance remains unchanged at $(nh^p)^{-1}$, while bias reduces to $O(h^{2(l+1)})$ away from the boundary, in the case of l-fold sharpened data. Boundary bias is of order h^{l+1} in the case of $\hat{g}_{\mathrm{NW},2}$, and h^{l+2} for $\hat{g}_{\mathrm{LL},2}$. Numerical performance of methods based on iteration will be addressed in the next section.

The estimator $\hat{g}_{\mathrm{LL},2}$ has an obvious analogue in the context of a general kth-degree local polynomial estimator $\hat{q}$. There the sharpened response variables $\check{Y}_j$ should be redefined as $2Y_j - \hat{q}(X_j)$ and the sharpened version of $\hat{q}$ defined by constructing the estimator for data pairs $(X_i, \check{Y}_i)$ instead of (X_i, Y_i). Bias is reduced by an order of magnitude, relative to that for the unsharpened estimator, and variance is inflated only by a constant factor.

3. Numerical performance. In our Monte Carlo study, where we assess the finite-sample performance of our methods, we consider $p = 1$ and address a linear-with-Gaussian-peak target function defined by

$$g(x) = g_k(x) = 2 - 5x + 2.5k \exp\{-200k(x - 0.5)^2\}. \tag{3.1}$$

The parameter k, as it increases, sharpens the peak and makes estimators of g_k more prone to suffer bias problems. So as not to confound general estimator performance with edge effects, in most parts of our study we distributed design points beyond the ends of the estimation interval, which we took to be $\mathscr{I} = [0, 1]$. Nevertheless, we discuss edge effects. In a real-data example at the end of the section we consider the case $p = 2$.

We compared the data-sharpened estimators $\hat{g}_{\mathrm{NW},1}$, $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{NW},3}$, $\hat{g}_{\mathrm{LL},1}$ and $\hat{g}_{\mathrm{LL},2}$ with conventional local-polynomial methods, and with a multiplicative bias reduction technique suggested by Linton and Nielsen (1994) and Jones, Linton and Nielsen [(1995), Section 4.2]. The high-degree local-polynomial fits that we considered included the local quadratic estimator $\hat{g}_{\mathrm{quad}}$ and the local cubic estimator $\hat{g}_{\mathrm{cub}}$. A problem with the multiplicative bias reduction estimator is that its denominator can take values close or equal to zero, even when design data are relatively plentiful. In consequence it is not surprising to note that this method was not found competitive with other approaches with respect to squared error properties. Also, the sharpened versions $\hat{g}_{\mathrm{NW},1}$ and $\hat{g}_{\mathrm{NW},3}$ of $\hat{g}_{\mathrm{NW}}$ suffer particularly from data sparseness for relatively small samples and were found not to perform as well as the other data-sharpening methods. Thus, for the sake of brevity we only present in what follows the results for $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{LL},1}$, $\hat{g}_{\mathrm{LL},2}$, $\hat{g}_{\mathrm{quad}}$ and $\hat{g}_{\mathrm{cub}}$ along with the benchmarks $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{LL}}$.

The results discussed below are for the case of the biweight kernel, sample size $n = 100$, and Normal $\mathrm{N}(0, 0.5^2)$ errors. The design density was chosen to be $\alpha B(4, q) + (1 - \alpha)U(\mathscr{I})$ for $\alpha \in [0, 1]$ and $q = 1, 2, 4$, where $B(p, q)$

and $U(\mathscr{I})$ denote the beta distribution with parameters p and q and the uniform distribution on $\mathscr{I}$, respectively. For the case $q = 4$, a large value of α corresponds to dense design towards the centre of $\mathscr{I}$ and sparse design at the edges. The cases $q = 1$ and $q = 2$ correspond, respectively, to a J-shaped density and to an asymmetric unimodal density (with mode near the right end of $\mathscr{I}$) for large values of α. When continuing design points beyond the ends of $\mathscr{I}$ we took the density to be constant (i.e., uniformly distributed stochastic design).

Figure 2 depicts median integrated squared error (ISE) curves for various densities and estimators and for $k = 2$ in the target function g_k at (3.1). We considered median ISE rather than mean ISE since the latter, due to the random denominators, do not necessarily exist. Thus these curves were constructed by taking the median of 1000 replications of ISEs at each value of h. We employed a grid of 51 logarithmically equally spaced bandwidths, and for each bandwidth in the grid we evaluated the corresponding smooths at 201 equispaced points in $\mathscr{I}$. The trapezoidal rule was used to calculate integrated squared error.

For small values of h the curve estimates were not defined since there were insufficient design points in the bandwidth window. This problem was especially noticeable for high-order local-polynomial methods. We determined, for each estimator and for different values of α and q, the smallest bandwidths $h_{\min}$ for which the curve estimates could be calculated at each of the 201 grid points of $\mathscr{I}$ in 90% of cases. We noted that as α increased, $\hat{g}_{\mathrm{LL},1}$ had the smallest value of $h_{\min}$ for all three values of q, indicating its greater robustness against sparsity problems. The vertical lines in Figure 2 denote $h_{\min}$ for each estimator type. Note that $h_{\min}$ coincides for $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{NW},2}$ and for $\hat{g}_{\mathrm{LL}}$ and $\hat{g}_{\mathrm{LL},2}$. Median ISE curves were computed between the respective values of $h_{\min}$ and 0.35. Further information about the effects of data sparsity on computability of different estimators is given in Figure 3, where, for different estimator types and two different designs, the percentage of samples for which the estimator can be calculated is graphed against bandwidth. The estimators $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{LL},1}$ can be calculated more often than the others, with $\hat{g}_{\mathrm{LL},1}$ doing better than $\hat{g}_{\mathrm{NW}}$ in cases where design sparsity is a particular problem at boundaries.

To produce the curves in Figure 2 we resolved data sparsity problems by considering only those samples for which the estimates could be calculated at all grid points in $\mathscr{I}$ for their respective $h_{\min}$. Panels (a), (b), (c) and (d) in Figure 2 display results in the cases $\alpha = 0$ and $(\alpha, q) = (1, 4), (1, 2)$ and (1,1), respectively. Note that in each of these panels we present only the better of $\hat{g}_{\mathrm{quad}}$ and $\hat{g}_{\mathrm{cub}}$ for the sake of picture clarity. For uniform design, $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{LL},1}$ and $\hat{g}_{\mathrm{LL},2}$ are competitive with high-order-local-polynomial methods, as indicated in panel (a). In particular, $\hat{g}_{\mathrm{LL},1}$ performs very well. The performance of $\hat{g}_{\mathrm{quad}}$ and $\hat{g}_{\mathrm{cub}}$, however, deteriorates markedly when design is nonuniform, as is evident from panels (b), (c) and (d). For the case $(\alpha, q) = (1, 4)$, $\hat{g}_{\mathrm{NW},2}$ and $\hat{g}_{\mathrm{LL},1}$ perform best, achieving the smallest minimum values of median ISE curves. In the case $(\alpha, q) = (1, 2)$, for which the peak in the regression

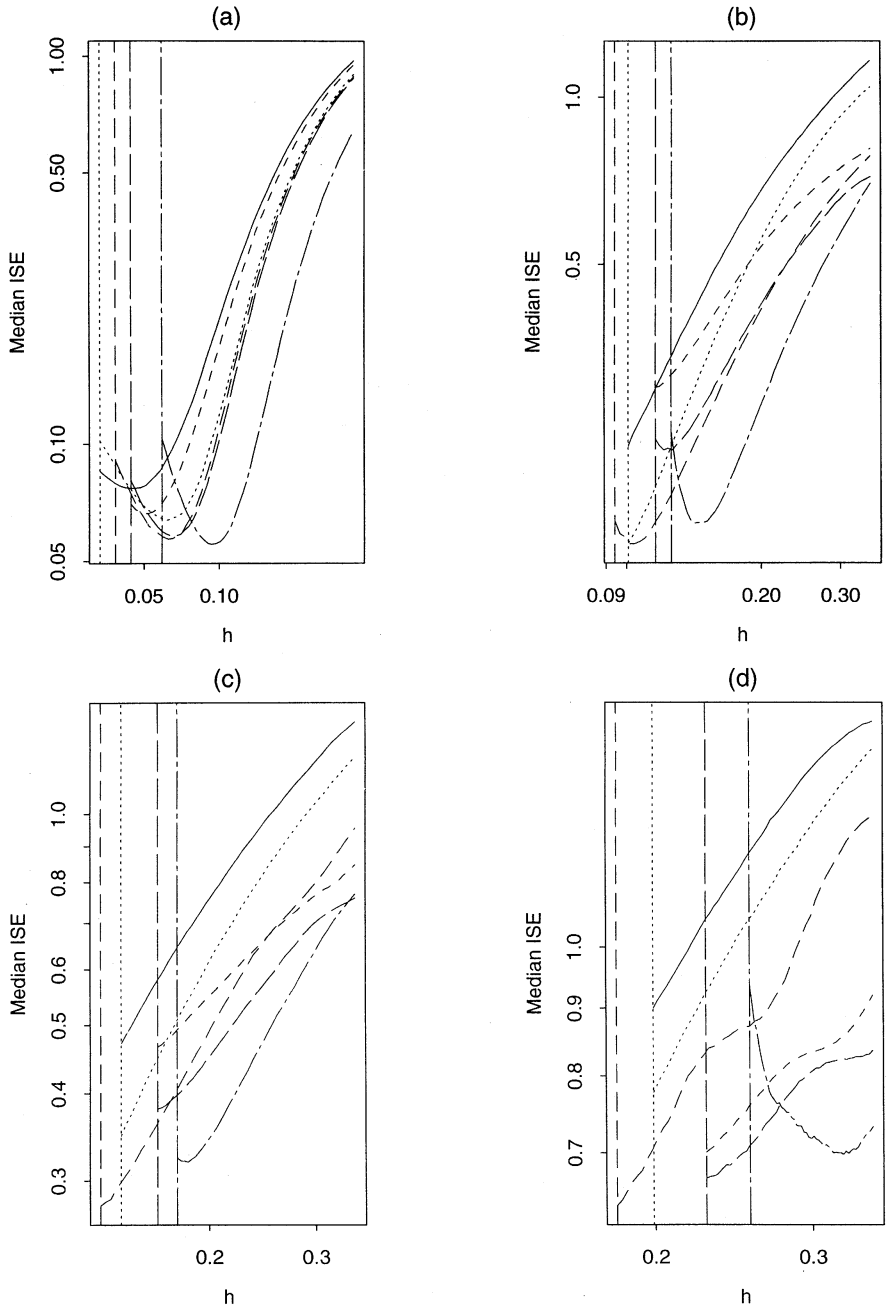

FIG. 2. *Median integrated squared error comparison of sharpened and local polynomial estimators with* $n = 100$ *and target function* g_2. *Panels* (a), (b), (c) *and* (d) *depict four different design density settings, with* $\alpha = 0$ *and* $(\alpha, q) = (1, 4), (1, 2)$ *and* (1, 1), *respectively. The solid, dotted, short-dashed, medium-dashed and long-dashed lines represent median integrated squared error curves for* $\hat{g}_{\mathrm{NW}}$, $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{LL}}$, $\hat{g}_{\mathrm{LL},1}$ *and* $\hat{g}_{\mathrm{LL},2}$, *respectively, while the dotted–dashed lines refer to* $\hat{g}_{\mathrm{quad}}$ *in panels* (b) *and* (c) *and to* $\hat{g}_{\mathrm{cub}}$ *in panels* (a) *and* (d).

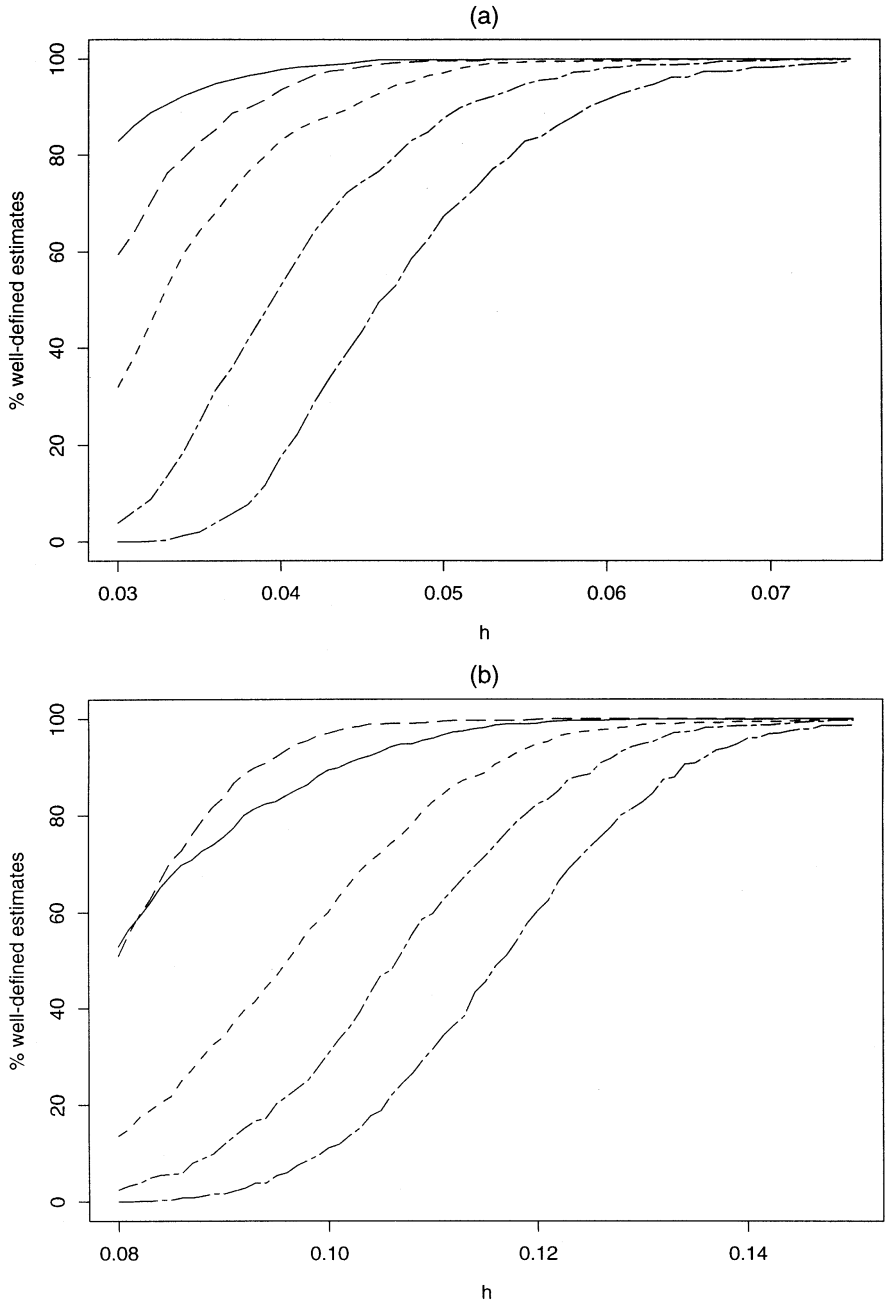

FIG. 3. *Percentage of times an estimate is well defined as function of bandwidth. Panels* (a) *and* (b) *correspond to the design density settings* $\alpha = 0$ *and* $(\alpha, q) = (1, 4)$, *respectively, in the context of Figure* 2. *Sample size was* $n = 100$. *Graphs for* $\hat{g}_{\mathrm{NW},2}$ *and* $\hat{g}_{\mathrm{LL},2}$ *are identical to those for* $\hat{g}_{\mathrm{NW}}$ *and* $\hat{g}_{\mathrm{LL}}$, *respectively, and so are not shown. Line types for estimators* $\hat{g}_{\mathrm{NW}}$, $\hat{g}_{\mathrm{LL}}$ *and* $\hat{g}_{\mathrm{LL},1}$ *are the same as in Figure* 2, *while dotted–short–dashed and dotted–long–dashed lines refer to* $\hat{g}_{\mathrm{quad}}$ *and* $\hat{g}_{\mathrm{cub}}$, *respectively.*

function is not co-symmetric with the design density, $\hat{g}_{\mathrm{LL},1}$ clearly outperforms $\hat{g}_{\mathrm{NW},2}$.

The case $(\alpha, 1)$ is more difficult, since the design is particularly sparse at the left end of $\mathscr{I}$. Here also $\hat{g}_{\mathrm{LL},1}$ attains least median ISE over $[h_{\min}, 0.35]$, thanks to its greater robustness against sparse design. This enables $\hat{g}_{\mathrm{LL},1}$ to have a significantly smaller value of $h_{\min}$ than the other estimators. Broadly similar conclusions may be drawn if we take $k = 3$ or $k = 5$ in (3.1).

Performance at the left boundary was also investigated in the case $k = 2$, along with $\alpha = 0$ and $(\alpha, q) = (1, 4)$. For each method, we considered the median of 1000 ISEs in [0,0.1] calculated at their respective median ISE optimal bandwidth (as found in the last paragraph). In this part of the study we did not generate design points outside $\mathscr{I}$. While all methods share the same order of performance in the uniform setting, with median ISE ranging from 0.003 (for $\hat{g}_{\mathrm{NW},2}$) to 0.007 (for $\hat{g}_{\mathrm{cub}}$), we observed huge differences for the sparse design setting $(\alpha, q) = (1, 4)$. Here median ISE ranged from 0.006 (for $\hat{g}_{\mathrm{NW},2}$), to 4.6 (for $\hat{g}_{\mathrm{quad}}$) and to 333.9 (for $\hat{g}_{\mathrm{cub}}$), the estimates $\hat{g}_{\mathrm{LL},1}$ and $\hat{g}_{\mathrm{LL},2}$ achieving 0.11 and 0.28, respectively. We conclude that high-degree-local-polynomial methods suffer particularly from sparse design at the edges.

We then explored the effects of l-fold data sharpening, of the type employed in $\hat{g}_{\mathrm{LL},2}$. The following results were observed for sample sizes $n = 50$, 100 and 200 and with $k = 2$ in the target function g_k at (3.1). There was generally an initial decrease in the minimum value of median ISE curves, the overall minimum occurring at two-fold sharpening (i.e., after one more sharpening step than was used to construct $\hat{g}_{\mathrm{LL},2}$). Beyond that point, however, the minimum slowly increased. Exceptions were the cases $(\alpha, q) = (1, 2)$ and (1,1), along with sample size $n = 50$ for which the minimum occurred after a single sharpening step. For all three sample sizes, the value of bandwidth at which the minimum occurred gradually increased with l, confirming suspicions that for a fixed sample size, iterating data sharpening beyond a certain point inflates variance to such an extent that deleterious effects are not outweighed by reductions in bias.

Finally we illustrate our methodology in the case $p = 2$ using a bivariate real-data example taken from Brinkman (1981). The data set consists of 88 measurements from an experiment in which ethanol was burned in a single-cylinder automobile test engine. The first component $X^{(1)}$ of the explanatory variable X represents the engine's compression ratio, the second component $X^{(2)}$ denotes the richness of the air–ethanol mixture fed to the engine and the response variable Y represents nitrous oxide concentration in the engine exhaust. Increasing $X^{(1)}$ tends to monotonically increase the response, even over a wide range, whereas increasing $X^{(2)}$ at first increases and then decreases nitrous oxide emissions. Therefore, a graph of $g(x) = E(Y|X = x)$ has a ridge virtually parallel to the $x^{(1)}$ axis, gradually increasing in the $x^{(1)}$ direction. Conventional low-order curve estimators can be expected to underestimate the true height of the ridge. Indeed, an application of $\hat{g}_{\mathrm{LL},1}$ increases estimated ridge height by 17%, relative to that for the conventional estimator $\hat{g}_{\mathrm{LL}}$, with negligible adverse effects caused by increased stochastic

error. See Figure 4. The bivariate bandwidth was $(h_1, h_2) = (1, 0.1)$, and the kernel was bivariate Normal.

4. Theoretical properties. Write $x = (x^{(1)}, \ldots, x^{(p)})$, and assume $K(x) = \prod_i L(x^{(i)})$, where L is a compactly supported, symmetric, univariate, Hölder-continuous probability density. (Of course, kernels that are not of the product type could be used instead; they include, for example, a p-variate, non-product form of the biweight kernel.) Suppose, too, that $Y - E(Y|X = x)$ may be written as $\epsilon\sigma(x)$, where the distribution of ϵ has unit variance and does not depend on x, and $\sigma(x)$ is a bounded, continuous function and that $h = h(n)$

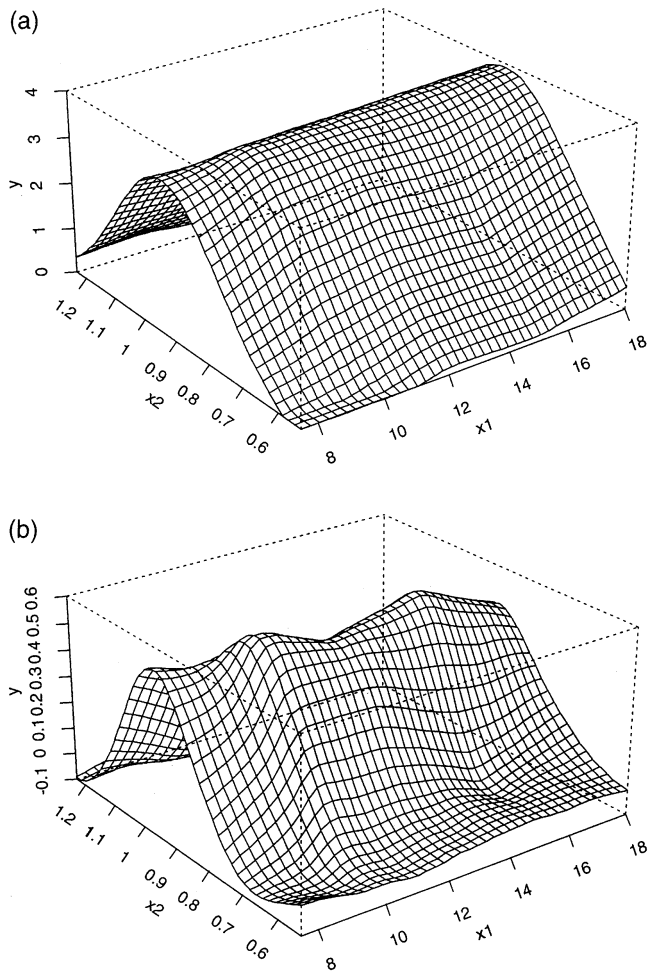

FIG. 4. *Perspective plots of regression surface estimates obtained using* $\hat{g}_{\mathrm{LL}}$ *and* $\hat{g}_{\mathrm{LL},1}$. *Panel* (a) *shows the estimate* $\hat{g}_{\mathrm{LL},1}$, *and panel* (b) *plots* $\hat{g}_{\mathrm{LL},1} - \hat{g}_{\mathrm{LL}}$. *The same bandwidths were used for both estimates. The effect of using* $\hat{g}_{\mathrm{LL},1}$ *rather than* $\hat{g}_{\mathrm{LL}}$ *is to increase ridge height by* 17%, *on average, while keeping features, in particular the slope of the ridge, unchanged.*

converges to 0 as $n \to \infty$, in such a manner that for some $\eta > 0$, $n^\eta h \to 0$ and $n^{1-\eta} h^p \to \infty$. Call these conditions (C_1).

To address performance in the interior of the support of the design distribution, let f denote the density of X, and assume that $f(x) > 0$ and both f and g have four bounded derivatives in a neighborhood of x. Call these conditions (C_2). To treat performance on the boundary, assume that the function representing the boundary of the support $\mathscr{S}$ of f has three bounded derivatives. [The boundary is a $(p-1)$-dimensional figure in p-dimensional space. For example, when $p = 2$ the boundary of the support of f is a curve—a one-dimensional figure—in the plane.] Suppose, too, that f has three bounded derivatives in the set $\mathscr{T}$ formed by the intersection of $\mathscr{S}$ with an open neighborhood of a point x_0 on the boundary, that g has three bounded derivatives in $\mathscr{T}$ and that $f(x_0) > 0$. Call these conditions (C_3).

Put $\nabla f(x) = (\partial f(x)/\partial x^{(1)}, \ldots, \partial f(x)/\partial x^{(p)})^T$, $\nabla^2 f(x) = (\partial^2 f(x)/\partial x^{(j)} \partial x^{(k)})$ (the former a p-vector and the latter a $p \times p$ matrix), $\kappa_1 = \int K^2$, $\kappa_2 = \int u^2 L(u)\, du$ and

$$\kappa_3 = \int \Big\{ 2K(u) - \int K(u+v) K(v)\, dv \Big\}^2 du.$$

If $\lambda(t) = \int L(s+t) L(s)\, ds$ then $\kappa_1 = \lambda(0)^p$ and

$$\kappa_3 = 4\lambda(0)^p + \Big\{ \int \lambda(t)^2\, dt \Big\}^p - 4 \Big\{ \int \lambda(t) L(t)\, dt \Big\}^p.$$

Values of κ_1, κ_2 and κ_3 are given in Table 1 when $p = 1$ and the kernel is uniform, Epanechnikov, biweight, triweight or standard Normal. In the first four cases, $L(t) = \frac{1}{2}$, $\frac{3}{4}(1-t^2)$, $\frac{15}{16}(1-t^2)^2$ or $\frac{35}{32}(1-t^2)^3$, respectively, on the interval $(-1, 1)$, and equals zero elsewhere.

Let $\mathrm{tr}(M)$ denote the trace of a $p \times p$ matrix M.

TABLE 1
Values of κ_1, κ_2 and κ_3 when $p = 1$.

Kernel	κ_1	κ_2	κ_3
Uniform	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{5}{6} \approx 0.833$
Epanechnikov	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{8387}{9856} \approx 0.851$
Biweight	$\frac{5}{7}$	$\frac{1}{7}$	$\frac{4665929295}{4635158528} \approx 1.007$
Triweight	$\frac{350}{429}$	$\frac{1}{9}$	$\frac{6000946648025}{5205581365248} \approx 1.153$
Normal	$\frac{1}{2\sqrt{\pi}}$	1	$\frac{4\sqrt{6} + \sqrt{3} - 8}{2\sqrt{6\pi}} \approx 0.407$

THEOREM 4.1. (i) *Under conditions* (C_1) *and* (C_2), $\hat{g}_{\mathrm{NW},1}$ *is asymptotically Normally distributed with mean* $g + \frac{1}{2}h^2 \operatorname{tr}(\nabla^2 g)\kappa_2 + o(h^2) + O\{(nh^p)^{-1}\}$ *and variance given by* $(nh^p)^{-1}\kappa_1 f^{-1}\sigma^2 + o\{(nh^p)^{-1}\}$ (*both the same as in the case of* $\hat{g}_{\mathrm{LL}}$), *and* $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{NW},3}$, $\hat{g}_{\mathrm{LL},1}$ *and* $\hat{g}_{\mathrm{LL},2}$ *are asymptotically Normally distributed with mean* $g + O\{h^4 + (nh^p)^{-1}\}$ *and variance* $(nh^p)^{-1}\kappa_3 f^{-1}\sigma^2 + o\{(nh^p)^{-1}\}$. (ii) *Under conditions* (C_1) *and* (C_3), *the asymptotic variances of all five estimators equal* $O\{(nh^p)^{-1}\}$ *at each point in* $\mathscr{T}$ (*including* x_0). *Furthermore, the asymptotic biases of* $\hat{g}_{\mathrm{NW},1}$, $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{NW},3}$ *and* $\hat{g}_{\mathrm{LL},1}$ *are* $O\{h^2 + (nh^p)^{-1}\}$ *on* $\mathscr{T}$, *while the asymptotic bias of* $\hat{g}_{\mathrm{LL},2}$ *equals* $O\{h^3 + (nh^p)^{-1}\}$ *there, and all five estimators are asymptotically Normally distributed at each point of* $\mathscr{T}$.

The conventional, uncorrected Nadaraya–Watson estimator has bias of order $h + O\{(nh^p)^{-1}\}$ at the boundary, and so part (ii) of the theorem demonstrates that all three types of data sharpening applied to $\hat{g}_{\mathrm{NW}}$ reduce boundary bias. Perhaps of greater interest, however, is the reduction of boundary bias in the case of $\hat{g}_{\mathrm{LL},2}$, effectively from $O(h^2)$ to $O(h^3)$.

It may be proved that $\hat{g}_{\mathrm{NW},1} = \hat{g}_{\mathrm{LL}} + o_p\{h^2 + (nh^p)^{-1/2}\}$. This is another way of stating that $\hat{g}_{\mathrm{NW},1}$ achieves, to first order, the same asymptotic bias and variance as $\hat{g}_{\mathrm{LL}}$. Therefore, in adjusting $\hat{g}_{\mathrm{NW}}$ to $\hat{g}_{\mathrm{NW},1}$ we have remedied inefficiency properties of the former. But that adjustment clearly involves increasing the susceptibility of $\hat{g}_{\mathrm{NW}}$ to design sparsity, and so we may conclude that the enhanced theoretical, asymptotic performance of $\hat{g}_{\mathrm{LL}}$, relative to $\hat{g}_{\mathrm{NW}}$, is achieved at the expense of design sparsity problems. This is readily observed in practice, too. It may also be explained through the fact that the "local constant" estimator $\hat{g}_{\mathrm{NW}}$ is obtained by fitting only one parameter, relative to the two parameters required for $\hat{g}_{\mathrm{LL}}$.

The potential for reducing mean squared error by data sharpening is clear from the theorem. In particular, the asymptotic mean squared errors of $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{NW},3}$, $\hat{g}_{\mathrm{LL},1}$ and $\hat{g}_{\mathrm{LL},2}$ all equal $O\{h^8 + (nh^p)^{-1}\}$, rather than $O\{h^4 + (nh^p)^{-1}\}$ as in the case of $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{NLL}}$. Therefore, for the first four estimators, by choosing h of size $n^{-1/(p+8)}$, the order of asymptotic mean squared error can be reduced to $n^{-8/(p+8)}$, which should be compared with the smallest order, $n^{-4/(p+4)}$, which is achievable for the estimators $\hat{g}_{\mathrm{NW}}$ and $\hat{g}_{\mathrm{LL}}$.

Note that, with $\|\cdot\|$ denoting the L^2 metric for functions $\kappa_3^{1/2} = \|2K - K * K\| > 2\|K\| - \|K\| = \kappa_1^{1/2}$, where "$*$" denotes convolution. Therefore, data sharpening inflates asymptotic variance by a constant factor, except in the case of $\hat{g}_{\mathrm{NW},1}$.

The smoothness assumptions in conditions (C_2) and (C_3) are clearly over restrictive in some instances. In particularly, to obtain the bias properties of $\hat{g}_{\mathrm{NW},1}$ in parts (i) and (ii) of the theorem we require only two derivatives of f and g, and to obtain the biases of $\hat{g}_{\mathrm{NW},2}$, $\hat{g}_{\mathrm{NW},3}$ and $\hat{g}_{\mathrm{LL},1}$ in part (ii) we need only two derivatives. Furthermore, the orders of asymptotic bias stated in the theorem for the interior of $\mathscr{S}$ are available uniformly in points in $\mathscr{S}$ that are distant at least Ch from the boundary, where $C > 0$ depends on the

support of K and on the first four derivatives of the boundary (assumed finite). Likewise, the asymptotic bias results stated for points that are within $O(h)$ of the boundary are valid uniformly there.

5. Outline of theoretical arguments. Let "as. bias $\hat{g}$" denote the asymptotic bias of an estimator $\hat{g}$, put $\epsilon_i = Y_i - E(Y_i|X_i)$, define $x + \hat{\psi}(x)$ to equal the ratio at (2.2) when X_j there is replaced by x, and put $\xi = (nh^p)^{-1}$. Standard Taylor-series methods may be used to prove that as. bias $\hat{\psi} = h^2 a_1 + o(h^2)$, as. bias $\hat{g}_{\mathrm{NW}} = g + h^2 a_2 + o(h^2) + O(\xi)$ and as. bias $\hat{g}_{\mathrm{LL}} = g + h^2 a_3 + o(h^2) + O(\xi)$, where $a_1 = \kappa_2(\nabla f) f^{-1}$ (a p-vector), $a_2 = \frac{1}{2}\kappa_2 \operatorname{tr}\{\nabla^2(fg) - g\nabla^2 f\} f^{-1}$ and $a_3 = \frac{1}{2}\kappa_2 \operatorname{tr}(\nabla^2 g)$ (both scalars). Therefore, $\widehat{X}_j = X_j + h^2 a_1(X_j) + Z_j^{(1)} + R_1$, $\widetilde{X}_j = X_j - h^2 a_1(X_j) + Z_j^{(2)} + R_2$, $\widetilde{Y}_j = g(X_j) - h^2 a_2(X_j) + Z_j^{(3)} + 2\epsilon_j + R_3$ and $\breve{Y}_j = g(X_j) - h^2 a_3(X_j) + Z_j^{(4)} + 2\epsilon_j + R_4$, where the variables $Z_j^{(k)}$ are measurable in the sigma-field generated by the X_i's and have zero mean, and R_k denotes a random variable that equals $o_p(h^2) + O_p\{(nh^p)^{-1/2}\}$. These formulae, while not in themselves adequate for the purpose of calculating bias and variance, guide our arguments below.

For each estimator the asymptotic bias may be calculated, up to terms $O(h^4) + O(\xi)$, as the ratio of the expected values of numerator and denominator. We shall give outline arguments in the cases of $\hat{g}_{\mathrm{NW},1}$ and $\hat{g}_{\mathrm{LL},1}$. For $\hat{g}_{\mathrm{NW},1}$, using the approximations developed in the previous paragraph, we see that

$$
\begin{aligned}
&\text{as. bias } \hat{g}_{\mathrm{NW},1}(x) + O(\xi) \\
&\quad = f(x)^{-1} h^{-p} \int \{g(y) - g(x)\} f(y) \\
&\qquad\qquad \times K\big[\{x - y - h^2 a_1(y)\}/h\big]\, dy + o(h^2) \\
&\quad = f(x)^{-1} \int \big[g\{x - hy - h^2 a_1(y)\} - g(x)\big] \qquad (5.1) \\
&\qquad\qquad \times f\{x - hy - h^2 a_1(y)\} K(y)\, dy + o(h^2) \\
&\quad = h^2 \{(\nabla f)^T (\nabla g)\kappa_2 + \tfrac{1}{2} f \operatorname{tr}(\nabla^2 g)\kappa_2 - f(\nabla g)^T a_1\} f^{-1} + o(h^2) \\
&\quad = \tfrac{1}{2} h^2 \operatorname{tr}(\nabla^2 g)\kappa_2 + o(h^2).
\end{aligned}
$$

Calculations in the case of $\hat{g}_{\mathrm{LL},1}$ are similar but more complex; we outline the argument. First, writing $U_i(x) = g(X_i) - g(x) - h^2 a_2(x)$ and $U = (U_1, \ldots, U_n)^T$, observe that with $\beta(x) =$ as. bias $\hat{g}_{\mathrm{LL},1}(x)$ we have

$$
\beta(x) + O(\xi) = e_1^T \Big[E\Big\{ \widetilde{X}(x)^T \widetilde{W}(x) \widetilde{X}(x) \Big\} \Big]^{-1} E\Big\{ \widetilde{X}(x)^T \widetilde{W}(x) U(x) \Big\} + o(h^2).
$$

In the formula for $U_i(x)$, Taylor-expand the term $g(X_i)$ around $\widetilde{X}_i$, obtaining $U_i(x) = V_i(x) + \{a_1(x)^T \nabla g(x) - a_2(x)\} h^2 + o_p(h^2)$ where $V_i(x) = g(\widetilde{X}_i) - g(x)$.

Therefore,

$$\begin{aligned}\beta(x) + O(\xi) &= e_1^T \Big[E\Big\{ \widetilde{X}(x)^T \widetilde{W}(x) \widetilde{X}(x) \Big\} \Big]^{-1} E\Big\{ \widetilde{X}(x)^T \widetilde{W}(x) V(x) \Big\} \\ &\quad + \Big\{ a_1(x)^T \nabla g(x) - a_2(x) \Big\} h^2 + o(h^2) \\ &= \text{as. bias } \tilde{g}_{\mathrm{LL}}(x) + \Big\{ a_1(x)^T \nabla g(x) - a_2(x) \Big\} h^2 + o(h^2),\end{aligned} \tag{5.2}$$

where $V = (V_1, \ldots, V_n)^T$ and $\tilde{g}_{\mathrm{LL}}$ is the estimator defined by a local-linear fit to the "pseudodata" $(\widetilde{X}_i, g(\widetilde{X}_i))$. To work out as. bias $\tilde{g}_{\mathrm{LL}}$ we use an argument familiar from calculating the asymptotic bias of $\hat{g}_{\mathrm{LL}}$: first, expand $g(\widetilde{X}_i)$ as

$$\begin{aligned}g(\widetilde{X}_i) &= g(x) + (\widetilde{X}_i - x)^T \nabla g(x) \\ &\quad + \tfrac{1}{2} (\widetilde{X}_i - x)^T \nabla^2 g(x) (\widetilde{X}_i - x) + o_p(\| \widetilde{X}_i - x \|^2);\end{aligned} \tag{5.3}$$

next, observe that the contribution to $\tilde{g}_{\mathrm{LL}}$ from $(\widetilde{X}_i - x)^T \nabla g(x)$ vanishes identically (not just asymptotically) when the expansion at (5.3) is substituted into the definition of $\tilde{g}_{\mathrm{LL}}$ and finally, note that the contribution to asymptotic bias from $\frac{1}{2}(\widetilde{X}_i - x)^T \nabla^2 g(x)(\widetilde{X}_i - x)$ is, up to terms of smaller order than h^2, identical to the contribution from $\frac{1}{2}(X_i - x)^T \nabla^2 g(x)(X_i - x)$ to the conventional local-linear estimator $\hat{g}_{\mathrm{LL}}$, and so equals $\frac{1}{2} h^2 \operatorname{tr}(\nabla^2 g) \kappa_2$. Combining the results from (5.2) down we conclude that

$$\begin{aligned}&\text{as. bias } \tilde{g}_{\mathrm{LL},\,1} + O(\xi) \\ &\quad = \tfrac{1}{2} h^2 \operatorname{tr}(\nabla^2 g) \kappa_2 + (a_1^T \nabla g - a_2) h^2 + o(h^2) = o(h^2).\end{aligned} \tag{5.4}$$

More extensive analysis shows that, when f and g have four bounded derivatives, the "$o(h^2)$" terms at (5.1) and (5.4) are actually $O(h^4)$.

Calculation of asymptotic variance is routine, and asymptotic Normality follows via Lindeberg's theorem. Bias computations at the boundary are similar.

Acknowledgments. We are grateful to Hans-Georg Müller for helpful discussions, and to the Associate Editor and two referees for constructive comments.

REFERENCES

BOSWELL, S. B. (1983). Nonparametric mode estimation for higher dimensional densities. Ph.D. dissertation, Dept. Statistics, Rice University.

BRINKMAN, N. D. (1981). Ethanol fuel: a single-cylinder engine study of efficiency and exhaust emissions. *SAE Transactions* **90** 1410–1424.

CHOI, E and HALL, P. (1999). Data sharpening as a prelude to density estimation *Biometrika.* **86** 941–947.

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

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

FWU, C., TAPIA, R. A. and THOMPSON, J. R. (1981). The nonparametric estimation of probability densities in ballistics research. *Proceedings Twenty-Sixth Conference Design of Experiments in Army Research Development and Testing* 309–326. Springfield, Virginia.

HALL, P. and PRESNELL, B. (1999). Intentionally biased bootstrap methods. *J. Roy. Statist. Soc. Ser. B* **61** 143–158.

HOUGAARD, P. (1988). A boundary modification of kernel function smoothing, with application to insulin absorption kinetics. In *Compstat Lectures* 31–36. Physica, Vienna.

HOUGAARD, P. PLUM, A. and RIBEL, U. (1989). Kernel function smoothing of insulin absorption kinetics. *Biometrics* **45** 1041–1052.

JONES, M. C., LINTON, O. and NIELSEN, J. P. (1995). A simple bias reduction method for density estimation *Biometrika* **82** 327–338.

JONES, R. H. and STEWART, R. C. (1997). A method for determining significant structures in a cloud of earthquakes. *J. Geophysical Res.* **102** 8245—8254.

LINTON, O. and NIELSEN, J. P. (1994). A multiplicative bias reduction method for nonparametric regression. *Statist. Probab. Lett.* **19** 181–187.

MAMMEN, E. and MARRON, J. S. (1997). Mass centred kernel smoothers. *Biometrika* **84** 765–777.

MÜLLER, H.-G. (1997). Density adjusted kernel smoothers for random design nonparametric regression. *Statist. Probab. Lett.* **36** 161–172.

MÜLLER, H.-G. and SONG, K.-S (1993). Identity reproducing multivariate nonparametric regression. *J. Multivariate Anal.* **46** 237–253.

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

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA
E-MAIL: halpstat@fac.anu.edu.au