Estimating a Changepoint, Boundary, or Frontier in the Presence of Observation Error

Peter HALL and Léopold SIMAR

A range of problems in economics and statistics involve calculation of the boundary, or frontier, of the support of a distribution. Several practical and attractive solutions exist if the sampled distribution has a sharp discontinuity at the frontier, but accuracy can be greatly diminished if the data are observed with error. Indeed, if the error is additive and has variance σ^2 then inaccuracies are usually of order σ, for small σ. In this article we suggest an elementary method for reducing the effect of error to $O(\sigma^2)$, and show that refinements can improve accuracy still further, to $O(\sigma^3)$ or less. The problem is inherently ill-posed, however, to such an extent that the frontier is generally not even identifiable unless the error distribution is known. The latter assumption is unreasonable in most practical settings, not in the least because the error is often asymmetrically distributed. For example, in the context of productivity analysis the error distribution tends to have a longer tail in the direction of underestimation of production. Nevertheless, even when the error distribution is unknown, it is often true that error variance is relatively low, and so methods for reducing systematic error in that case are useful in practice.

KEY WORDS: Changepoint analysis; Data envelopment analysis; Deconvolution; Frontier estimation; Kernel methods; Nonparametric estimation; Panel data; Production frontier.

1. INTRODUCTION

We observe independent data $X_1, \ldots, X_n$ generated by the model $X_i = Y_i + Z_i$, where the density of the distribution of Y_i has a jump discontinuity at a point θ and Y_i and Z_i are independent. The Z_i's may be thought of as observational "errors," added to the variables Y_i, which are really the focus of attention. We wish to estimate θ. The case where θ is an upper endpoint of the distribution of Y_i is of particular interest in the context of frontier analysis, and here we focus solely on that setting, although more general problems may be treated using similar methods.

In many cases this problem does not admit a unique solution, and so θ is not identifiable. Trivially, the roles of Y and Z can be interchanged; if the distributions of Y and Z end in points θ_Y and θ_Z, then the problems of estimating θ_Y and θ_Z are confounded. Even if it is assumed that the density of Z is symmetric and unimodal (with mode at 0) and infinitely differentiable, and that the distribution of Y has a jump discontinuity at its right endpoint θ, there may be an infinite number of possible values of θ for a given distribution of X. For instance, suppose that $X = V + \sum_{i \ge 1} U_i$ where each U_i is uniformly distributed on $[-2^{-i}, 2^{-i}]$, V is normally distributed with mean 0; and $V, U_1, U_2, \ldots$ are independent. Then for each i we may write $X = Y + Z$, where $Y = U_i$ and $Z = V + \sum_{j \ne i} U_j$. In this representation the distribution of Y has as its right endpoint $\theta = 2^{-i}$, and the distribution of Z is symmetric about 0, unimodal, and infinitely differentiable.

Even if the distribution of Z were perfectly known, estimating the distribution of X would not necessarily be straightforward. For example, if Z is known to be normal $N(0, \sigma^2)$ and σ^2 is known, then the minimax-optimal rate of convergence for estimating the density of Y from data on X is only $(\log n)^{-k/2}$, where n denotes sample size and k represents the number of derivatives of the density of Y. (See Carroll and Hall 1988 and Fan 1991, 1992, 1993 for discussion of this and related results.) The problem that we treat here is, in a sense, an order of magnitude more difficult, in that we make only minimal assumptions about the distribution of Z; for example, that $\text{var}(Z)$ is relatively small.

Despite these pessimistic views of the problem, there are grounds for believing that it can be posed in a manner that leads to useful and practicable solutions. For example, if the density of Y is perfectly flat in a short interval to the left of θ, and if the density of Z is unimodal and sufficiently short-tailed to the right, then it may be shown that θ can be characterized as a local maximum of the gradient of the density of X. This result fails if the density of Y approaches its endpoint at an angle, but nevertheless it still holds to a good approximation, provided that the error variance, $\sigma^2 = \text{var}(Z)$, is small. Indeed, we show that the L_1 error in the flat-density approximation is only $O(\sigma^2)$ as $\sigma \to 0$. These results should be compared with the rate of $O(\sigma)$ that obtains if standard frontier estimation methods are applied without allowing for the effects of noise. The biases of estimators can be further reduced by estimating, and correcting for, asymmetries introduced by departures from the flat-density hypothesis. We show how to reduce bias from order σ^2 to σ^3 in this way; higher-order corrections are also feasible.

Therefore, although the problem that we are endeavoring to solve is intrinsically ill-posed and can have infinitely many solutions, a useful and well-defined practical solution can be obtained in the low-noise case. This solution generally will not be consistent unless the density of Y is completely flat, but the asymptotic bias will be proportional to the error variance, not to the error standard deviation.

The case where error variance is small is commonly encountered in practical settings, and so there is substantial motivation for developing a solution along the lines just described, despite the admittedly irregular nature of the problem. A major application is in economics, where boundaries may be interpreted as production frontiers. There, one wishes to analyze how different firms transform a set of inputs (typically labor, energy, or capital) into an output (typically a quantity of goods produced). The boundary represents the upper

Peter Hall is Professor, Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia. Léopold Simar is Professor, Institut de Statistique, Université Catholique de Louvain, B 1348 Louvain-la-Neuve, Belgium (E-mail: simar@stat.ucl.ac.be). This research was supported by the Projet d'Actions de Recherche Concertées (No. 98/03-217) from the Belgian Government.

Journal of the American Statistical Association
June 2002, Vol. 97, No. 458, Theory and Methods

envelope of attainable production; it is the geometric locus of optimal production. A firm's economic efficiency of a firm is defined in terms of its ability to operate close to the boundary. In this context, firms can be compared to determine the more efficient. Although production is usually measured with error it is typically small relative to the level of production itself. Therefore, the condition of high signal-to-noise ratio is generally valid, and a low-noise approximation can be effective.

Importantly, our method is readily extended to spatial-bivariate settings, where the convergence rate $O(\sigma^2)$ continues to be valid. In such problems the boundary is a curve in a plane rather than a point on a line. It can be estimated as a ridge line of the surface defined by the maximum directional derivative of a bivariate kernel density estimator, or by estimating the maximum gradient of the appropriate conditional density.

The econometric literature on boundary estimation includes parametric approaches, suggested by Aigner, Lovell, and Schmidt (1977), Meeusen and van den Broek (1977), and Greene (1990). In the context of panel data, these approaches have semiparametric generalizations (see Park and Simar 1994; Park, Sickles, and Simar 1998). Nonparametric techniques, including those suggested in this article, are based on the idea of "enveloping" the data. In particular, Farrell (1957) introduced data envelopment analysis (DEA) estimators of the boundary, based on the convex hull of the data. Linear programming techniques for computing the DEA estimator introduced by Charnes, Cooper, and Rhodes (1978) have proven particularly popular. Deprins, Simar, and Tulkens (1984) extended the idea to nonconvex sets and suggested the free disposal hull (FDH) estimator, equal to the smallest free disposal set containing all of the data. Statistical properties of DEA and FDH estimators are well known (see Banker 1993; Korostelev, Simar, and Tsybakov 1995a,b; Kneip, Park, and Simar 1998; Gijbels, Mammen, Park, and Simar 1999; Park, Simar, and Weiner 2000).

Frontier estimation methods have been applied in many settings, including public services, banks, hospitals, and other institutions (see Seiford 1996 for an extensive survey). They may be extended to multivariate settings: (see, e.g., Shephard 1970). However, existing methods rely on the unrealistic assumption that the data Y_i are observed without noise. In the literature these approaches are referred to as *deterministic* frontier models, in contrast to *stochastic* frontier models where the data are perturbed by noise. In the presence of noise, the envelopment techniques described earlier will be significantly biased, to such an extent that they will generally not be consistent. Kneip and Simar (1996) proposed a method that was arguably the first attempt to estimate stochastic frontiers, but it is limited to panel data and requires particularly restrictive assumptions; for example, inefficiency must be assumed constant over long time periods. In contrast, the conditions required in this article are very mild.

2. METHODOLOGY IN THE UNIVARIATE CASE

2.1 A "Toy" Problem: The Case Where the Y-density Is Flat

Let (X, Y, Z) denote a generic version of (X_i, Y_i, Z_i). Recall from Section 1 that $X = Y + Z$, where X is observed, the density f_Y of Y has its right endpoint at θ, and Z represents a random error. In this section we treat the "toy" problem, where $f_Y(y)$ is positive for y in an interval (a, θ) to the left of θ and 0 for all $y > \theta$. We also assume, more realistically than the assumption that f_Y is flat, that the density f_Z of Z is unimodal with a unique mode at 0 and is supported on a compact interval $[b, c]$, with $\theta > a + c$. If f_Y is constant on (a, θ), then $f'_X(x)$, denoting the derivative at x of the density f_X of X, is proportional to $-f_Z(x-\theta)$ for $x > a + c$ (and in particular, for x in a neighborhood of θ). Thus the unimodality property assumed of f_Z implies that θ can be characterized as that point, say ω, in $[a + c, \infty)$ at which the slope of the density of X achieves its greatest absolute value.

Of course, this characterization rests critically on the assumption that f_Y is constant in an interval to the left of θ. If that condition were not satisfied, and instead f_Y had a jump discontinuity at θ without, in particular, $f'_Y(\theta-)$ vanishing, then ω would be only an approximation to θ. In Section 2.2 we explore the accuracy of the approximation when the assumption of flatness of f_Y is violated. Nevertheless, the condition that f_Z is unimodal is an attractively mild assumption about the error distribution, and if f_Z is also symmetric about the origin (a not-uncommon condition on an error density), then the assertion that f_Z has a unique mode at 0 is also mild.

Most importantly, the fact that we can estimate θ by simply estimating the point at which the value of $|f'_X|$ is maximized is attractive. For example, we may construct a nonparametric estimator $\hat{f}_X$ of f_X and estimate θ by $\operatorname{argmax} |\hat{f}'_X|$, where the argmax is taken in the neighborhood of the true value of θ; see Section 2.4 for details.

One aspect of the fact that the distributions of Y and Z are not identifiable, noted in Section 1, is that any translation of one can be accommodated by translating the other by an equal but opposite amount. Our insistence that the mode of the error distribution be unimodal *with mode at* 0 eliminates this indeterminism. Clearly, a centering convention of some type is necessary, just as it is in simpler problems of regression; in the current setting, centering at the mode is more natural than centering at the mean.

2.2 Accuracy of ω as an Approximation to θ

The flat-density approximation, $\theta \approx \omega \equiv \operatorname{argmax} |f'_X|$, is derived under the assumption that f_Y is perfectly flat to the left of θ. It is intuitively clear that if the latter condition is violated, then the approximation should nevertheless become increasingly accurate as the standard deviation, say σ, of Z decreases. The assumption that the signal-to-noise ratio [i.e., the ratio $\operatorname{var}(Y)/\operatorname{var}(Z)$] is high is often appropriate in practice. We explore these issues in detail in our numerical work in Section 4.

We next show that the flat-density approximation is in error by only $O(\sigma^2)$ as $\sigma \to 0$. The $O(\sigma^2)$ convergence rate is an order of magnitude faster than the rate $O(\sigma)$ that would arise if we were to ignore the presence of noise and apply a standard frontier estimation method to the noisy data X as though they were really data on Y. We verify the $O(\sigma^2)$ rate in the case of error distributions where the density can be represented as $f_Z(z) = \sigma^{-1} g(z/\sigma)$, with g denoting a fixed density and σ (in our asymptotic model) converging to 0.

Specifically, we assume that (a) f_Y has two continuous derivatives to the left of θ, (b) $f_Y(\theta-) > 0$, (c) g is compactly supported and unimodal with its mode at 0, (d) g has two continuous derivatives in a neighborhood of 0, and (e) $g''(0) \neq 0$. Call these collective conditions (C_1). The assumption that g is compactly supported is made only to avoid having to discuss specific moment conditions, which depend on the number of terms to which our Taylor expansion approximations are taken, and because it is in keeping with conditions imposed in Section 2.1. Nevertheless, our results remain valid if the error distribution has sufficiently many finite moments, and in particular if it is normal.

Assume initially that

$$f_Y(\theta - y) = A_0 + A_1 y + A_2 y^2 \tag{1}$$

for y in a small interval $(0, \epsilon)$, where $A_0 > 0$ and $-\infty < A_1 < \infty$. Then, for x sufficiently close to θ,

$$-f_X'(x) = A_0 f_Z(x-\theta) + A_1\{1 - F_Z(x-\theta)\} + 2A_2 \int_0^\infty \{1 - F_Z(x - \theta + u)\}\,du, \tag{2}$$

where F_Z denotes the distribution function corresponding to f_Z. (Our techniques do not require that f_Z or F_Z be known; these functions are introduced here only to develop methodology.) For the sake of simplicity, we assume here that f_Z is supported in a sufficiently small interval about its mode, 0. However, the main result, that $\theta - \omega = O(\sigma^2)$, is available much more generally, for example, in the case of normal errors.

Put $\gamma = \int |u| g(u)\,du$. Taylor expanding the right side of (2) in the context of fixed g, and noting that (C_1) implies $f_Z'(0) = 0$, it may be shown that as $\sigma \to 0$,

$$-f_X'(x) = A_0\{f_Z(0) + \tfrac{1}{2}(x-\theta)^2 f_Z''(0)\} + A_1\{\tfrac{1}{2} - (x-\theta) f_Z(0)\} + A_2\gamma\sigma + o(\sigma) \tag{3}$$

uniformly in $|x - \theta| \leq c\sigma^2$ for any fixed $c > 0$. Deriving the turning point of the quadratic on the right side, we conclude that $\omega = \operatorname{argmax}|f_X'|$ satisfies

$$\omega = \theta + \frac{A_1 f_Z(0)}{A_0 f_Z''(0)} + o(\sigma^2) = \theta + O(\sigma^2) \tag{4}$$

as $\sigma \to 0$.

Analogously, suppose that we estimate θ by

$$\hat{\omega} = \operatorname{argmax}|\hat{f}_X'|, \tag{5}$$

where $\hat{f}_X'$ denotes the derivative of a kernel density estimator based on data on X, and the argmax is taken over a closed interval $\mathcal{J}$ in the right tail of the distribution of X, including the point θ. Then $\hat{\omega}$ estimates θ with systematic error $O(\sigma^2)$, for large n,

$$\hat{\omega} = \theta + O(\sigma^2) + o_p(1) \tag{6}$$

as $n \to \infty$. In this formula the term $O(\sigma^2)$ denotes the order of systematic error that occurs even in large samples. This level of bias remains uncorrected, even in the limit as $n \to \infty$. The quantity $o_p(1)$ denotes a stochastic term that converges to 0 as $n \to \infty$. Sufficient regularity conditions for (6) are (a) f_Y has a bounded derivative in a neighborhood to the left of θ, (b) g is compactly supported and unimodal with its mode at 0 and has two continuous derivatives in a neighborhood of 0, (c) $f_Y(\theta-) g''(0) \neq 0$, and (d) $\hat{f}_X'$ is a kernel density estimator based on a bounded, symmetric, compactly supported, continuously differentiable kernel and a bandwidth $h = h(n)$ that satisfies $h \to 0$ and $nh^3 \to \infty$ as $n \to \infty$. A proof of (6) is similar to the proofs given in the Appendix.

If instead of (1) we assume that f_Y has two continuous derivatives in $(\theta - \epsilon, \theta]$ for some $\epsilon > 0$, and that $A_0 \equiv f_Y(\theta-) > 0$, and if we define $A_1 \equiv -f_Y'(\theta-)$, then (3) is still valid, with a slightly more complex derivation. See Appendix A.1 for an illustration of the argument.

To better appreciate the orders of magnitude claimed in (4), consider the orders of the terms appearing in (3). Because f_Z is modeled as $\sigma^{-1} g(\cdot/\sigma)$ for a fixed density g, it follows from (C_1) that $f_Z(0) = \sigma^{-1} g(0)$ and $f_Z''(0) = -\sigma^{-3}|g''(0)|$, where $g(0), |g''(0)|$ are both strictly positive. Therefore, if $x - \theta \asymp \sigma^2$, then the linear and quadratic terms in (3) are both of size σ, and the remainder $o(\sigma)$ is negligible. Likewise, the approximation error $\omega - \theta$ at (4) is of size σ^2. Indeed, (4) is equivalent to

$$\omega = \theta + \frac{f_Y'(\theta-)\, g(0)}{f_Y(\theta-)\, |g''(0)|}\sigma^2 + o(\sigma^2). \tag{7}$$

2.3 Enhancing the Accuracy of ω by Local Linear Approximation to f_Y

If the error density f_Z is compactly supported and symmetric about 0, and f_Y is perfectly flat in a sufficiently large interval to the left of θ, then the density derivative f_X' will be symmetric about θ. That is, it will satisfy $f_X'(\theta + x) \equiv f_X'(\theta - x)$ for x in a neighborhood of 0. Conversely, if f_Y is not flat to the left of θ but f_Z is nevertheless symmetric, then f_X' will not be symmetric about θ. This lack of symmetry can be used as a diagnostic for lack of flatness, but more practically it can be exploited as a means for correcting ω to improve its accuracy as an approximation to θ.

At the level of (3), the main effects of lack of flatness are not really apparent. In particular, even if f_Y were perfectly flat to the left of θ, f_X' would still equal a quadratic plus higher-order terms. Departure from the flat-density hypothesis becomes discernible only if we progress to cubic and higher-order terms in the expansion. As we show, the cubic and quartic terms can be used to approximate the component $A_1 f_Z(0)/A_0 f_Z''(0)$, on the right side of (4), of the error in the flat-density approximation to θ. Arguing in this way, the error in the approximation can be reduced from $O(\sigma^2)$ to $O(\sigma^3)$. Higher-order corrections can reduce the error still further, but we do not consider these corrections here.

We continue to interpret f_Z as $\sigma^{-1} g(\cdot/\sigma)$, where g is a fixed density and $\sigma \to 0$. As a prelude to developing lengthier versions of (3), we assume that (a) f_Y has five bounded derivatives to the left of θ, (b) $f_Y(\theta-) > 0$, (c) g is compactly supported and unimodal with its mode at 0, (d) g has five bounded derivatives in a neighborhood of 0, (e) $g^{(j)}(0) = 0$

for $j = 1, 3$, and (f) $g''(0) \neq 0$. Call these collective conditions (C_2). Property (e) is a little weaker than asking that the distribution of Z be symmetric.

We prove in Appendix A.1 that under (C_2),

$$-f_X'(\omega + u) = C_0 - C_2 u^2 + C_3 u^3 + C_4 u^4 + O(\sigma^4), \quad (8)$$

where

$$C_0 = f_Y(\theta-) f_Z(0) + O(1),$$

$$C_2 = \tfrac{1}{2} f_Y(\theta-) |f_Z''(0)| + O(\sigma^{-2}),$$

$$C_3 = -\tfrac{1}{6} f_Y'(\theta-) |f_Z''(0)| \left(1 - \tfrac{f_Z(0) f_Z^{(4)}(0)}{f_Z''(0)^2}\right) + O(\sigma^{-2}),$$

and

$$C_4 = \tfrac{1}{24} f_Y(\theta-) f_Z^{(4)}(0) + O(\sigma^{-4}).$$

It may be deduced from (8) that, assuming (C_2), the remainder $o(\sigma^2)$ in (7) is actually $O(\sigma^3)$,

$$\omega = \theta + \frac{f_Y'(\theta-) g(0)}{f_Y(\theta-) |g''(0)|} \sigma^2 + O(\sigma^3). \quad (9)$$

More importantly, (8) gives us the leverage we need to correct the approximation ω (of θ) for the term of order σ^2. In particular, (8) includes a cubic term that would vanish if f_Y were perfectly flat in a neighborhood immediately to the left of θ (i.e., if $A_1 = 0$) but is in general nonzero.

Property (9) and the formulas for $C_0, \ldots, C_4$ imply that

$$\frac{3C_0 C_3}{2(C_2^2 - 6C_0 C_4)} = (\theta - \omega)\{1 + O(\sigma)\}$$

as $\sigma \to 0$. Thus, because $\omega = \theta + O(\sigma^2)$,

$$\theta = \omega + \frac{3C_0 C_3}{2(C_2^2 - 6C_0 C_4)} + O(\sigma^3). \quad (10)$$

Formula (10) suggests the following approach to bias-correcting a flat-density approximation estimator of θ. Estimate f_X' by $\hat{f}_X'$, say, and define $\hat{\omega}$ as in (5). Next, estimate constants $\widehat{C}_0, \ldots, \widehat{C}_4$ by fitting a quartic to $-\hat{f}_X'$ in a neighborhood of $\hat{\omega}$, obtaining

$$-\hat{f}_X'(\hat{\omega} + u) \approx \widehat{C}_0 - \widehat{C}_2 u^2 + \widehat{C}_3 u^3 + \widehat{C}_4 u^4, \quad (11)$$

say. Take

$$\hat{\theta} = \hat{\omega} + \frac{3\widehat{C}_0 \widehat{C}_3}{2(\widehat{C}_2^2 - 6\widehat{C}_0 \widehat{C}_4)} \quad (12)$$

to be an estimator of θ. Analogously to (6), it may be proved by Taylor expansion that the correction in (12) achieves the level of bias suggested by (10),

$$\hat{\theta} = \theta + O(\sigma^3) + o_p(1), \quad (13)$$

where $O(\sigma^3)$ denotes the order of bias that remains even in the limit as $n \to \infty$ and $o_p(1)$ denotes a stochastic term that converges to 0 as $n \to \infty$. Sufficient regularity conditions for (13), in addition to (C_2), are that $\hat{f}_X'$ is a kernel density estimator based on a bounded, symmetric, compactly supported kernel with five continuous derivatives, and a bandwidth $h = h(n)$ that satisfies $h \to 0$ and $nh^9 \to \infty$ as $n \to \infty$.

It is implicit in this argument that $C_2^2 - 6C_0 C_4$ does not vanish. Now,

$$C_2^{-2}(C_2^2 - 6C_0 C_4)\{1 + O(\sigma)\} = 1 - \frac{g(0) g^{(4)}(0)}{g''(0)^2},$$

which quantity is nonzero for common error distributions. In particular, it equals -2 when the error distribution is normal.

To better appreciate the orders of magnitude present in formulas such as (8), note that C_0, C_2, C_3, and C_4 are asymptotic to constant multiples of $\sigma^{-1}, \sigma^{-3}, \sigma^{-3}$, and σ^{-5} as $\sigma \to 0$. Therefore, if u equals a constant multiple of σ^2 (the same size as $\omega - \theta$), then the terms C_0, $C_2 u^2$, $C_3 u^3$, and $C_4 u^4$ in (8) are asymptotic to constant multiples of $\sigma^{-1}, \sigma, \sigma^3$, and σ^3. In particular, the cubic and quartic terms in (8) are of the same size, σ^3, and so the quartic term in particular cannot be dropped, and the remainder, $O(\sigma^4)$, is an order of magnitude smaller than the terms explained by the quartic polynomial in (8). These properties justify the claim, implicit in our motivation of the estimator $\hat{\theta}$, that the polynomial given in (11) represents an accurate approximation to $-f_X'$.

2.4 Empirical Methods

We may estimate ω by finding the point at which the gradient of a kernel estimator of f_X achieves its maximum. To this end, define

$$\hat{f}_X(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is a kernel function and h is a bandwidth. We take K to be a compactly supported, symmetric probability density. Higher-order kernels are of course feasible, but their greater tendency to produce oscillatory density estimates, owing to the so-called "side lobes" of K, makes them not so attractive when the task at hand is to find turning points of derivatives.

Define $\hat{\omega}$ as in (5). In describing properties of this estimator, we first treat the case where the error distribution is held fixed as sample size increases. It may be shown that in this setting, the optimal order of bandwidth (in the sense of minimizing asymptotic mean squared error of $\hat{\omega}$ as an estimator of ω) is $h \asymp n^{-1/9}$. For this choice of h, the stochastic error of $\hat{\omega}$ and the asymptotic bias are both of size $n^{-2/9}$. Result (15) in Section 5 is a more detailed statement of this property. Further results in that section show that the $O(\sigma^2)$ order of error of the flat-density approximation ω of θ can be achieved empirically by $\hat{\omega}$, but there $h \asymp (\sigma^2/n)^{1/5}$.

3. EXTENSION TO MULTIVARIATE DATA

We consider only bivariate cases, noting that the same principles apply to multivariate data. Suppose that $X = Y + Z$, where again Y (the variable of intrinsic interest) and Z (representing error) are independent, but this time X, Y, and Z are vectors of length 2. We represent their coordinates by bracketed superscripts, for example, $X = (X^{(1)}, X^{(2)})$. The bivariate density of the distribution of Y is assumed to have a sharp discontinuity along a smooth, curved boundary, say $\mathcal{C}$, but the sharpness of the change along $\mathcal{C}$ in the observed data X is blunted by the addition of errors Z.

Our methods for the univariate problem can be generalized so as to approximate or estimate $\mathcal{C}$ in at least three ways. The generalizations depend to some extent on the type of data available. In the first approach, repeated values of $X^{(2)}$ are available for each of a sequence of values of $X^{(1)}$, and the replications are in sufficient quantity for us to simply apply the methods suggested in Section 2 to each $X^{(1)}$. In this setting, the errors Z would generally have a vanishing second coordinate, but their first coordinate would have a continuous distribution with its mode at the origin.

Alternatively, in either this setting or a setting with no replications of the $X^{(2)}$'s, we could project onto a straight line passing through a point $x^{(1)}$ on the first axis and parallel to the second axis those data values X whose first coordinate lay within a given bandwidth of $x^{(1)}$. The projected data, being on a line rather than in the plane, would be analyzed as though they arose in the context discussed in Section 2. For example, if $\mathcal{C}$ were determined by the formula $x^{(2)} = \psi(x^{(1)})$, then to estimate $\psi(x^{(1)})$ from the dataset $\mathcal{X} = \{X_1, \ldots, X_n\}$, we would first form the univariate dataset $\mathcal{X}_h(x^{(1)}) = \{X_i^{(2)} : x^{(1)} - h \le X_i^{(1)} \le x^{(1)} + h\}$, where $h > 0$ was a bandwidth, and estimate $\psi(x^{(1)})$ from $\mathcal{X}_h(x^{(1)})$. In practice, one would typically take h equal to the bandwidth for the nonparametric density estimator computed from the projected data on the line. Thus the effective kernel for the latter estimator would be bivariate and supported on a rectangle, rather than univariate and supported on a line.

The two methods just described are distinctly nonspatial and are appropriate when the first and second data coordinates are so different in character that geometric operations, such as rotations, in the $(x^{(1)}, x^{(2)})$ plane are not really justified. This is the case in many practical settings, for example, when the first coordinate represents an input to a commercial process (e.g., number of employees of a company) and the second is an output (e.g., annual profit of the company).

In spatial settings, however, both the model and the method typically would be different. In particular, the error distribution would be genuinely bivariate and continuous, rather than concentrated on just one coordinate. Our third approach to approximating and estimating $\mathcal{C}$ is appropriate in this setting. As before, define f_X to be the density of X and put $f_X^{(i)}(x) = (\partial/\partial x^{(i)}) f_X(x)$. Let

$$|f'_X(x)| \equiv \sup\{u^{(1)} f_X^{(1)}(x) + u^{(2)} f_X^{(2)}(x) : (u^{(1)})^2 + (u^{(2)})^2 = 1\}$$
$$= \left[\{f_X^{(1)}(x)\}^2 + \{f_X^{(2)}(x)\}^2\right]^{1/2}$$

denote the maximum of the directional derivative of f_X over all directions. Using this definition of $|f'_X|$, we consider the surface $\mathcal{S}$ represented by $v = |f'_X(x)|^2$ and take a ridge line of the surface to be an approximation to $\mathcal{C}$.

A ridge line of $\mathcal{S}$ is the projection into the x-plane, Π_x, of a ridge of $\mathcal{S}$. Several different definitions of a ridge have been given by Hall, Qian, and Titterington (1992). For any of these definitions, the ridge line approximation to $\mathcal{C}$ is in error by only $O(\sigma^2)$, not simply $O(\sigma)$, under mild regularity conditions. In Appendix A.2 we verify this claim, which we call result (R), in the case of the following specific definition of a ridge:

> A ridge (analogous to the popular definition if $\mathcal{S}$ represents a mountain range) or an antiridge (corresponding to a line along a valley floor) is a locus of points where successive "contour lines," in planes perpendicular to Π_x, have local maxima in the absolute values of their curvatures, taken over all planes passing through the point that are perpendicular to Π_x.

The true surface $\mathcal{S}$ is estimated by the empirical surface $\widehat{\mathcal{S}}$, defined as the locus of points (v, x) determined by $v = |\hat{f}'_X(x)|^2$, where

$$\hat{f}_X(x) = (nh^2)^{-1} \sum_i K\left(\frac{x - X_i}{h}\right)$$

is a kernel estimator of f, computed using a bivariate kernel K. The analogous estimator of $\mathcal{C}$ is a ridge line of $\widehat{\mathcal{S}}$. As in the univariate case, here it can be shown that the $O(\sigma^2)$ convergence rate may be attained empirically, although for the sake of brevity we do not pursue that matter further here.

4. NUMERICAL PROPERTIES

In this section we report the results of simulation studies in the univariate case and give a real data example in a bivariate setting.

4.1 Simulations in the Univariate Case

4.1.1 Models. We analyzed five models for (Y, Z) in terms of which X was defined by $X = Y + Z$:

- Model 1: $\theta - Y$ is exponential with mean $1/\lambda$ and Z is normal $N(0, \sigma_z^2)$. Thus $E(Y) = \theta - (1/\lambda)$ and $\operatorname{var} Y = \lambda^{-2}$.
- Model 2: $\theta - Y$ is distributed as $|N(0, \sigma^2)|$ and Z is normal $N(0, \sigma_z^2)$. Thus $E(Y) = \theta - (2/\pi)^{1/2}\sigma$ and $\operatorname{var} Y = (1 - 2\pi^{-1})\sigma^2$.
- Model 3: $\theta - Y$ is distributed as $N^+(\mu, \sigma^2)$ [i.e., $N(\mu, \sigma^2)$ conditioned on $N(\mu, \sigma^2) > 0$] and Z is $N(0, \sigma_z^2)$. Thus $E(Y) = \theta - (\mu + c\sigma)$ and $\operatorname{var} Y = \sigma^2\{1 - c(\mu\sigma^{-1} + c)\}$, where $c = \phi(\mu/\sigma)/\Phi(\mu/\sigma)$ and ϕ and Φ are the standard normal density and distribution functions. Note that $\mu = 0$ corresponds to model 2.
- Model 4: $\theta = 1$, Y has density $f_Y(y) = \{1 + a(1 - y)^b\}/\{1 + a(b+1)^{-1}\}$ for $0 \le y \le 1$, where $a, b > 0$, and Z is $N(0, \sigma_z^2)$.
- Model 5: $\theta - Y$ is uniformly distributed on the interval $(0, 1)$ and Z is $N(0, \sigma_z^2)$. Thus $E(Y) = \frac{1}{2}$ and $\operatorname{var} Y = 1/12$.

Models 1 and 2 reflect assumptions commonly made in parametric frontier models. In particular, they imply that the density f_Y of Y is increasing immediately before the boundary. In Model 3 the density of Y is either increasing or decreasing immediately before the boundary according to whether $\mu < 0$ or $\mu > 0$. In Model 4 the density is always decreasing at the boundary. Model 5 represents a case where $\omega = \theta$, and if this were known to the experimenter, then correction of $\hat{\omega}$ would not be attempted. It is thus of interest to ascertain the relative performance of the corrected estimator.

In each model we took $\theta = 1$ and selected the parameter values so that $E(Y) = \frac{1}{2}$ or, in the case of model 4, $E(Y) \approx \frac{1}{2}$. We chose various values of n and σ_z, the latter allowing us to tune the noise to signal ratio, $\rho_{\text{nts}} = \sigma_Z/\sigma_Y$.

4.1.2 Practical Computations. In our simulations we used the following iterative procedure to determine the bandwidth, h. Starting with an overly large h, we computed an initial value of $\hat{\omega}$ by calculating the zero of $\hat{f}''_X(x)$ in a neighborhood of $\max(X_i)$. Then, over a decreasing set of values of h on a fine grid, we computed a new value of $\hat{\omega}$ by calculating the zero of $\hat{f}''_X(x)$ in a neighborhood of the previous value of $\hat{\omega}$, and so on. We stopped iteration when the absolute value of the difference in the values of $\hat{\omega}$ between two iterations increased. Thus we obtained both a value for h and the corresponding $\hat{\omega}$.

The correction term, $\text{ct} = 3\widehat{C}_0\widehat{C}_3/\{2(\widehat{C}_2^2 - 6\widehat{C}_0\widehat{C}_4)\}$, added to $\hat{\omega}$ in (12) can be unstable because it is a ratio of two random variables. We dampened the fluctuations by using a variant of Breiman's (1996) bagging method, as follows. Having computed h and $\hat{\omega}$ from the original sample $\mathcal{X}$ as discussed earlier, we drew a bootstrap resample $\mathcal{X}^*$ from $\mathcal{X}$ and calculated the bootstrap versions $\hat{\omega}^*$ and ct^* of $\hat{\omega}$ and ct, for the given bandwidth h. The final correction term, say $\widehat{\text{ct}}$, may be taken to be any robust average of the values ct^* drawn by repeated resampling; we took $\widehat{\text{ct}}$ to be the mean of those values of ct^* out of $B = 100$ that satisfied $|\text{ct}^*| \le h$.

As suggested in Section 2, we used quartic interpolation to calculate the values of $\widehat{C}_j^*$ (the bootstrap variant of $\widehat{C}_j$) needed to compute ct^*. To implement the interpolation, we used a grid of 10 points in a neighborhood of width $h/2$. As an illustration, Figure 1 depicts the density and its derivatives for a typical sample generated from model 1.

4.1.3 Results. Tables 1–9 give numerical approximations to mean squared error (MSE), bias, variance, and the stan-

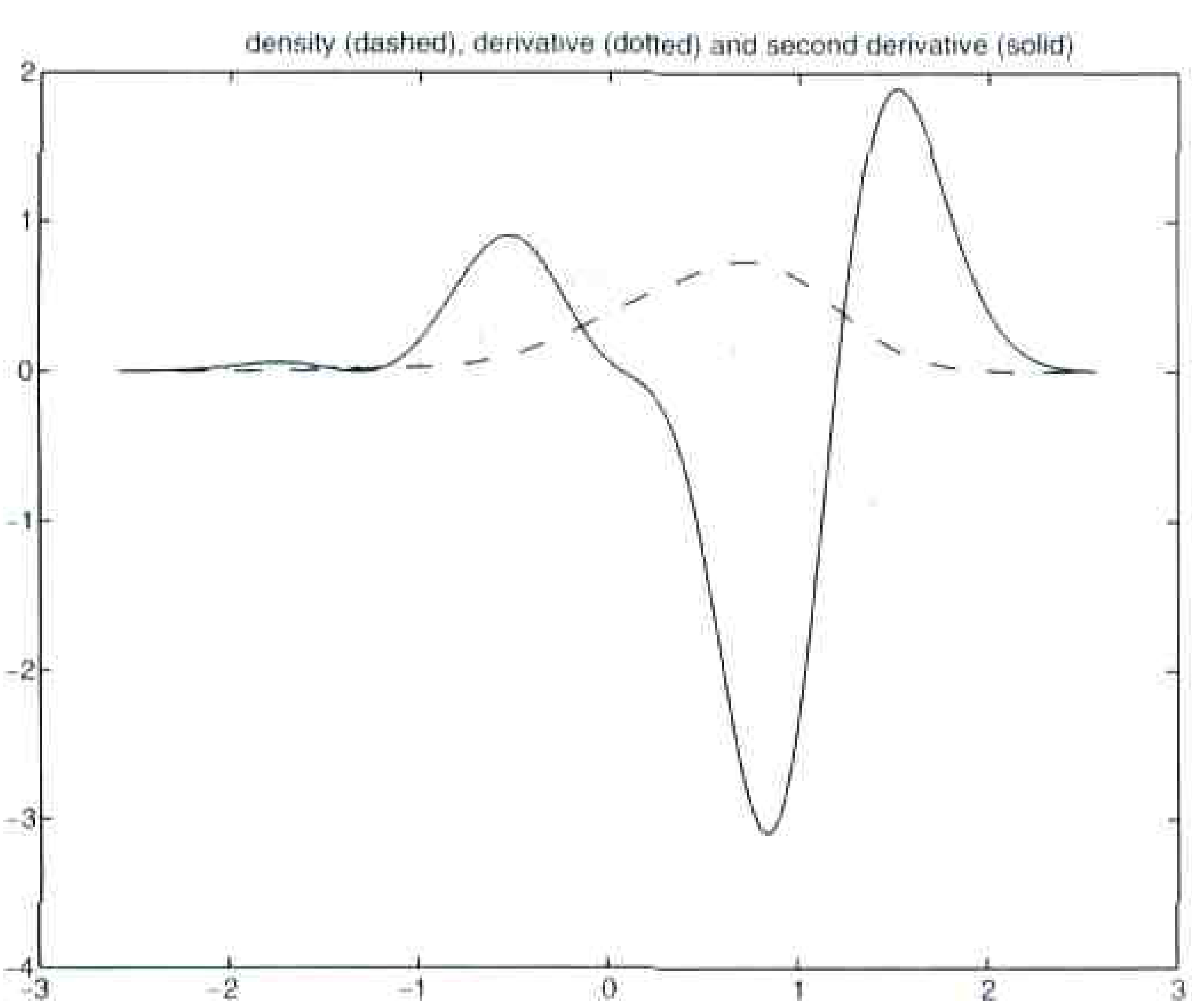

Figure 1. Density and Its Derivatives for a Typical Sample of Size $n = 100$ Generated From Model 1 With $\theta = 1$, $\lambda = 2$, $\sigma_Z = .2$, $\mu_Y = \sigma_Y = .5$, and $\rho_{nts} = .40$. For these data, $\max(X_i) = 1.2145$, $\hat{\omega} = 1.2072$, $\hat{\theta} = 1.0767$, and $h = .3435$.

Table 1. Monte Carlo Simulations for Model 1: Exponential ($\lambda = 2$), $E(Y) = .5$, $\sigma_Y = .5$, $\sigma_Z = .05$, and $\rho_{nts} = .10$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat{\omega}$	.0267	.1046	(.0089)	.0158
	$\hat{\theta}$	.0052	.0127	(.0050)	.0051
$n = 100$	$\hat{\omega}$	.0040	.0468	(.0030)	.0018
	$\hat{\theta}$	.0008	.0058	(.0020)	.0008
$n = 500$	$\hat{\omega}$	.0013	.0277	(.0016)	.0005
	$\hat{\theta}$	.0003	.0051	(.0011)	.0003
$n = 1000$	$\hat{\omega}$	.0005	.0178	(.0009)	.0002
	$\hat{\theta}$	.0002	.0053	(.0009)	.0002

Table 2. Monte Carlo Simulations for Model 1: Exponential ($\lambda = 2$), $E(Y) = .5$, $\sigma_Y = .5$, $\sigma_Z = .1$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat{\omega}$	.0306	.1253	(.0087)	.0150
	$\hat{\theta}$	.0065	.0215	(.0055)	.0060
$n = 100$	$\hat{\omega}$	.0071	.0689	(.0035)	.0024
	$\hat{\theta}$	.0021	.0120	(.0032)	.0020
$n = 500$	$\hat{\omega}$	.0021	.0422	(.0013)	.0004
	$\hat{\theta}$	.0012	.0141	(.0022)	.0010
$n = 1000$	$\hat{\omega}$	.0016	.0379	(.0010)	.0002
	$\hat{\theta}$	.0009	.0158	(.0017)	.0006

Table 3. Monte Carlo Simulations for Model 2: Half-Normal ($\sigma = .6267$), $E(Y) = .5$, $\sigma_Y = .3778$, $\sigma_Z = .0378$, and $\rho_{nts} = .10$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat{\omega}$	.0099	.0358	(.0066)	.0087
	$\hat{\theta}$	.0059	−.0294	(.0050)	.0050
$n = 100$	$\hat{\omega}$	.0017	.0207	(.0026)	.0013
	$\hat{\theta}$	.0013	−.0064	(.0025)	.0012
$n = 500$	$\hat{\omega}$	.0004	.0064	(.0013)	.0003
	$\hat{\theta}$	.0003	−.0027	(.0011)	.0002
$n = 1000$	$\hat{\omega}$	.0002	.0048	(.0008)	.0001
	$\hat{\theta}$	.0001	−.0020	(.0007)	.0001

Table 4. Monte Carlo Simulations for Model 2: Half-Normal ($\sigma = .6267$), $E(Y) = .5$, $\sigma_Y = .3778$, $\sigma_Z = .0756$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat{\omega}$	.0113	.0417	(.0069)	.0096
	$\hat{\theta}$	.0066	−.0174	(.0056)	.0064
$n = 100$	$\hat{\omega}$	.0025	.0238	(.0031)	.0020
	$\hat{\theta}$	.0025	−.0119	(.0034)	.0023
$n = 500$	$\hat{\omega}$	.0005	.0136	(.0013)	.0004
	$\hat{\theta}$	.0005	−.0013	(.0016)	.0005
$n = 1000$	$\hat{\omega}$	.0003	.0102	(.0009)	.0002
	$\hat{\theta}$	.0003	−.0027	(.0012)	.0003

Table 5. Monte Carlo Simulations for Model 3: Truncated Normal ($\mu = -.40$, $\sigma = .7834$), $E(Y) = .5$, $\sigma_Y = .4046$, $\sigma_Z = .0809$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat{\omega}$	.0151	.0664	(.0073)	.0107
	$\hat{\theta}$	.0079	−.0064	(.0063)	.0079
$n = 100$	$\hat{\omega}$	.0032	.0378	(.0030)	.0018
	$\hat{\theta}$	.0014	.0028	(.0026)	.0014
$n = 500$	$\hat{\omega}$	.0010	.0249	(.0015)	.0004
	$\hat{\theta}$	.0006	.0043	(.0017)	.0005
$n = 1000$	$\hat{\omega}$	.0005	.0177	(.0010)	.0004
	$\hat{\theta}$	.0003	.0034	(.0013)	.0003

Table 6. Monte Carlo Simulations for Model 3: Truncated Normal ($\mu = .40$, $\sigma = .3768$), $E(Y) = .5$, $\sigma_Y = .3033$, $\sigma_Z = .0303$, and $\rho_{nts} = .10$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 100$	$\hat\omega$	.0046	−.0506	(.0032)	.0021
	$\hat\theta$	.0060	−.0612	(.0033)	.0022
$n = 500$	$\hat\omega$	.0024	−.0452	(.0014)	.0004
	$\hat\theta$	.0022	−.0395	(.0018)	.0006
$n = 1000$	$\hat\omega$	.0019	−.0416	(.0010)	.0002
	$\hat\theta$	.0012	−.0295	(.0013)	.0003

Table 8. Monte Carlo Simulations for Model 4: Decreasing Density for Y With ($a = 1$, $b = 2$), $E(Y) = .4375$, $\sigma_Y = .2891$, $\sigma_Z = .0578$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat\omega$	.0393	−.1109	(.0117)	.0272
	$\hat\theta$	.0417	−.1296	(.0112)	.0250
$n = 100$	$\hat\omega$	.0164	−.0635	(.0079)	.0125
	$\hat\theta$	.0163	−.0631	(.0079)	.0123
$n = 500$	$\hat\omega$	.0009	−.0129	(.0019)	.0007
	$\hat\theta$	.0008	−.0085	(.0019)	.0007
$n = 1000$	$\hat\omega$	.0004	−.0085	(.0014)	.0004
	$\hat\theta$	.0004	−.0054	(.0014)	.0004

dard deviation of bias (std(bias)), obtained by averaging over 200 simulations. It will be seen that $\hat\theta = \hat\omega + \widehat{ct}$, the corrected version of $\hat\omega$, almost always has less bias and less MSE than $\hat\omega$. Indeed, $\hat\theta$ improves substantially on $\hat\omega$ when the density is monotonically increasing up to the boundary, for example, in the cases of models 1 and 2. In practical terms, this is often the case for data that have an economic origin, where the boundary represents a theoretical limit to performance and firms tend to be clustered at or near the boundary.

The improvement offered by $\hat\theta$ over $\hat\omega$ is still positive, but not so large, when the mode of f_Y occurs near the boundary, for example, in the case of model 3 with $\mu > 0$. There the quartic fit used to compute the correction term is sometimes confused by the presence of the mode. This difficulty could be remedied by fitting the quartic by hand in difficult cases, although we found this prohibitively time-consuming to do in a simulation study. Performance of the method in model 5 is surprisingly good. In theory, this case requires no correction, because the flat-density approximation is valid, and attempting a correction might be expected to significantly reduce performance. This turns out to not be the case.

4.2 Real Data Example in the Bivariate Case

4.2.1 Semiparametric Fit of Cobb–Douglas Model. A dataset on 123 American electric utility companies presented by Christensen and Greene (1976) has become a popular example for the study of frontier models. Greene (1990) fitted and compared several fully parametric models and described the data in detail: for each utility, cost, C, output, Q, price of labor, P_l, price of fuel, P_f, and price of capital, P_k, were measured. The parametric stochastic frontier is a cost-efficient frontier described by a Cobb–Douglas model,

$$y_i = \alpha + \beta' x - u_i + v_i, \tag{14}$$

Table 7. Monte Carlo Simulations for Model 3: Truncated Normal ($\mu = .40$, $\sigma = .3768$), $E(Y) = .5$, $\sigma_Y = .3033$, $\sigma_Z = .0607$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 100$	$\hat\omega$	.0039	−.0441	(.0031)	.0020
	$\hat\theta$	.0064	−.0611	(.0037)	.0027
$n = 500$	$\hat\omega$	.0022	−.0430	(.0014)	.0004
	$\hat\theta$	.0024	−.0419	(.0018)	.0007
$n = 1000$	$\hat\omega$	.0020	−.0418	(.0011)	.0002
	$\hat\theta$	.0015	−.0324	(.0016)	.0005

where $y = -\log(C/P_f)$, $x_1 = \log Q$, $x_2 = (\log Q)^2$, $x_3 = \log(P_l/P_f)$, and $x_4 = \log(P_k/P_f)$.

Greene (1990) proposed parametric models (exponential, half-normal, truncated normal, and gamma) for the probability density function of the inefficiency term u_i convolved with normal for the noise v_i. (This noise variable is assumed to be part of the recorded data and is being modeled here; it is, of course, not being added as part of the analysis.) The procedure is maximum likelihood in the spirit of the work of Aigner et al. (1977) and Meeusen and van den Broek (1977) on parametric stochastic frontiers.

We implement a semiparametric approach using the method described in Section 2, based on the estimator $\hat\omega$ defined by (5). The idea is straightforward: The parametric part is the Cobb–Douglas model as before, but now $u_i > 0$ and the noise variables v_i are independent and identically distributed with unspecified probability densities. We make the same assumptions on the probability densities of u_i and v_i as in Section 2 for the densities of Y_i and Z_i.

Arguing in this manner, we suggest the following estimation procedure. Rewrite (14) as

$$y_i = \alpha^\circ + \beta' x_i - u_i^\circ + v_i,$$

where $\mu = E(u_i)$, $u_i^\circ = u_i - \mu$, and $\alpha^\circ = \alpha - \mu$. An ordinary least squares (OLS) procedure produces consistent estimators $\hat\alpha^\circ$ and $\hat\beta$. Define the shifted OLS residuals,

$$\hat\epsilon_i = y_i - \hat\beta' x_i, \qquad i = 1, \ldots, n,$$

which are estimators of the true residuals,

$$\epsilon_i = y_i - \beta' x_i = \alpha - u_i + v_i.$$

Table 9. Monte Carlo Simulations for Model 5: Uniform (0,1), $E(Y) = .5$, $\sigma_Y = .2887$, $\sigma_Z = .0577$, and $\rho_{nts} = .20$

Sample size	Estimate	MSE	Bias	Std(bias)	Variance
$n = 20$	$\hat\omega$	.0159	−.0313	(.0086)	.0150
	$\hat\theta$	.0176	−.0641	(.0082)	.0136
$n = 100$	$\hat\omega$	.0019	−.0016	(.0031)	.0019
	$\hat\theta$	.0025	−.0131	(.0034)	.0024
$n = 500$	$\hat\omega$	.0004	−.0016	(.0014)	.0004
	$\hat\theta$	.0005	−.0037	(.0016)	.0005
$n = 1000$	$\hat\omega$	.0002	−.0005	(.0010)	.0002
	$\hat\theta$	.0003	−.0022	(.0011)	.0002

Using the method of Section 2, compute $\hat{\alpha}$ from the sample $\hat{\epsilon}_i$, $i = 1, \ldots, n$. Specifically, we used $\hat{\omega}$ defined by (5), taking $\hat{f}_X$ there to be a kernel density estimator using a bandwidth h computed as described in Section 4.1.2. Note too that an estimator of $\mu = E(u)$ is given by

$$\hat{\mu} = -n^{-1} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}' x_i) \equiv \hat{\alpha} - \hat{\alpha}^{\circ}.$$

The resulting estimator of α in the case of the electric utility company data equals 7.37 and lies between the value 7.21 of the OLS estimator and the values 7.41, 7.50, and 7.53 of maximum likelihood estimators in the cases of $|N(0, \sigma^2)|$, $N^+(\mu, \sigma^2)$, and exponential models. Thus the new nonparametric estimator lies approximately in the middle of the range of values obtained using alternative parametric approaches and so gives a meaningful result.

4.2.2 Application of the Method of Section 3. As noted in Section 3, the univariate methodology can be extended to multivariate cases. We consider a bivariate setting where the discontinuity is along a smooth boundary $\mathcal{C}$, and illustrate the case where $\mathcal{C}$ is determined by the formula $x^{(2)} = \psi(x^{(1)})$. The framework is nonspatial: $\psi(x^{(1)})$ could represent, for instance, a production frontier where $x^{(2)}$ is the maximum output that a firm can produce for a given input $x^{(1)}$. Due to production inefficiency, the "true" value of an observation lies below the frontier, but the observation is contaminated by noise. This is exactly the setting of Gijbels et al. (1999), who proposed an estimator of ψ (a bias-corrected version of the convex hull of the cloud of points) when there is no noise and the frontier function is known to be monotone and concave.

In this setting, to estimate $\psi(x^{(1)})$ for a given $x^{(1)}$, we project onto a straight line passing through $x^{(1)}$ and parallel to the second axis those data whose first coordinate lies within a given bandwidth of $x^{(1)}$. In the illustrations that follow we chose the bandwidth to be appropriate for estimating the marginal density of $X^{(1)}$. The projected data were analyzed using the algorithm described earlier, producing $\hat{\psi}(x^{(1)})$. We briefly illustrate the application of this method in a Monte Carlo experiment, and then apply the method to the electric utility company data.

4.2.3 Simulated Data. Here we use model 1 of Gijbels et al. (1999) and perturb the data by noise:

$$X^{(1)} \sim U[0, 1] \text{ and } X^{(2)} = \psi(X^{(1)}) \exp(-V) \exp(W),$$

where

$$\psi(x) = x^{1/2}, \text{ with } V \sim \exp(3)$$

and

$$W \sim N(0, (.0667)^2),$$

with the random variables $X^{(1)}$, V, and W independent. Note that $E\{\exp(-V)\} = 3/4$. The noise to signal ratio is $\rho_{\text{nts}} = \sigma_W / \sigma_V = .20$. Figure 2 depicts a typical simulated dataset for a sample of size $n = 100$.

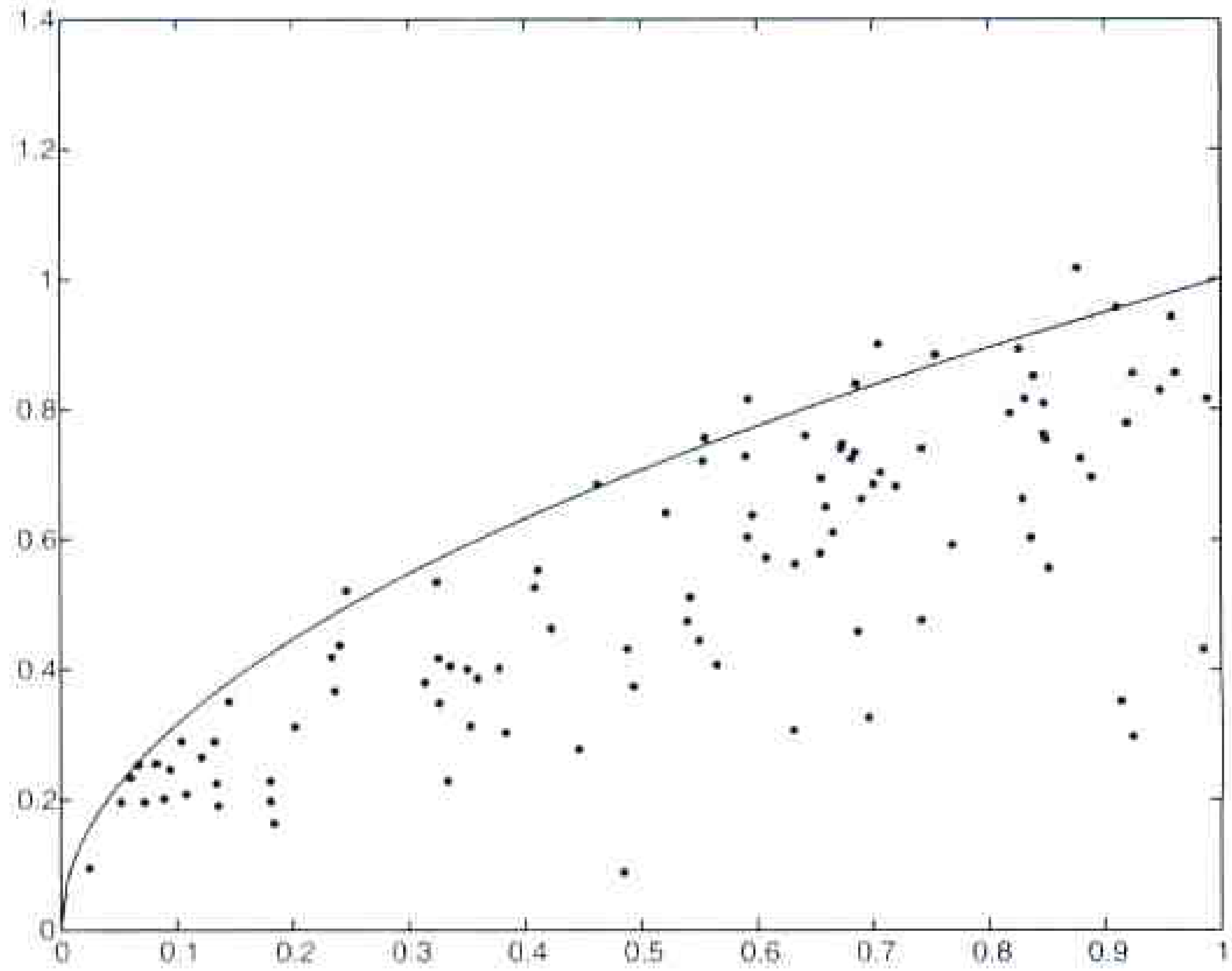

Figure 2. A Typical Simulated Sample of Size 100. The solid line represents the true frontier function.

Table 10 summarizes simulation results for two sample sizes at three different points, .25, .50, and .75, interior to the support of the marginal density of $X^{(1)}$. The results represent the averages over 500 simulations. It is evident from the table that the estimator performs well, even for n as small as 100. Indeed, the bias is often not significantly different from 0.

4.2.4 Application to Electric Utility Data. Returning to the electric utility data of Christensen and Greene (1976), we use only measurements of the variables $X^{(2)} = \log Q$ and $X^{(1)} = \log C$, where Q is the production output of a firm and C is the total cost involved in production. Figure 3 depicts the data along with pointwise estimates of $\psi(x^{(1)})$ over a selected grid of 50 values for $x^{(1)}$. The bandwidth was taken equal to .51. The figure also shows a smooth estimator of the frontier obtained by running a kernel smoother (quartic kernel and same bandwidth .51) through the estimated boundary points. To avoid edge effects, we restricted the region of estimation to [1.5, 5].

Here, of course, we focused on estimation of the upper boundary, because we are interested in estimating the "best practice" frontier that maximizes the production Q for a given level of the cost C. If for some reason one wanted to consider the lower boundary, then it could be dealt with in the same way as the upper boundary.

Table 10. Estimated Bias and MSE at Three Different Values of $X^{(1)}$, for Two Sample Sizes

	$x^{(1)} = .25$		$x^{(1)} = .50$		$x^{(1)} = .75$	
n	Bias (std(bias))	MSE	Bias (std(bias))	MSE	Bias (std(bias))	MSE
100	.0048 (.0017)	.0014	.0017 (.0022)	.0025	−.0003 (.0024)	.0028
500	.0009 (.0010)	.0005	.0041 (.0012)	.0007	.0041 (.0014)	.0011

NOTE: Standard errors of the bias estimator are given in parentheses.

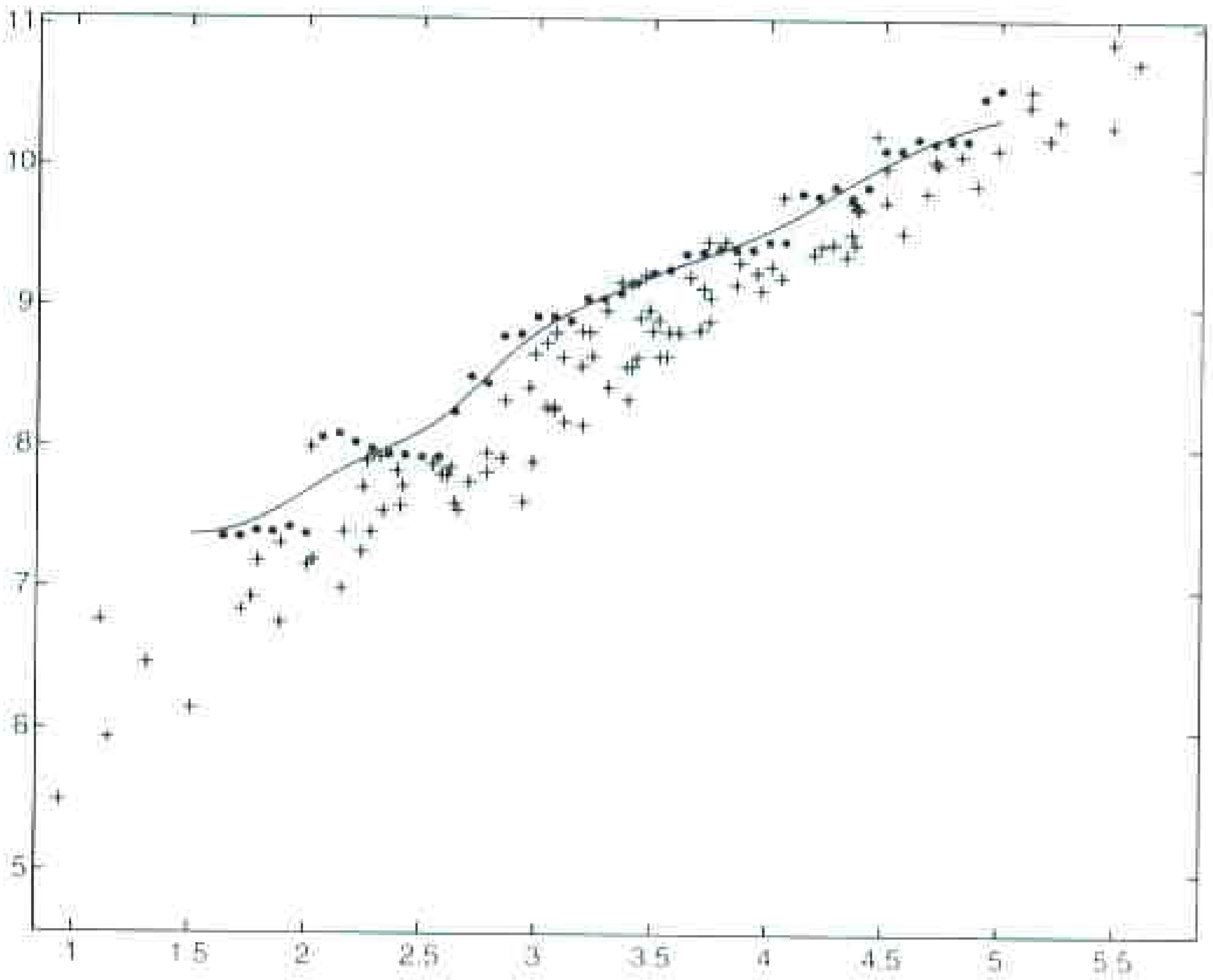

Figure 3. Scatterplot of the American Electric Utility Data. The +'s represent the observations, the ·'s are the pointwise estimates of the boundary over a selected grid of 50 values for $x^{(1)}$, and the solid line is a smooth version of these boundary points. The bandwidth for $X^{(1)}$ is .5128.

5. THEORETICAL STATISTICAL PROPERTIES

First, we treat the case where the distribution of the errors, Z_i, is held fixed as sample size increases, and introduce regularity conditions. Assume that $\omega = \operatorname{argmax}|f_X'(x)|$ is an interior point of a closed interval $\mathcal{J}$ and the density f_X has four continuous derivatives in an open interval containing $\mathcal{J}$, with $f_X^{(3)}(\omega) \neq 0$, $f_X^{(2)}(\omega) = 0$ and $f_X^{(2)}(x) \neq 0$ for all points $x \in \mathcal{J}$ with $x \neq \omega$. Also suppose that K is a compactly supported, symmetric probability density with three bounded derivatives and $h \asymp n^{-1/9}$. Then

$$\hat{\omega} - \omega = \left\{f_X^{(3)}(\omega)\right\}^{-1}\left[(nh^5)^{-1/2}\{\kappa f_X(x)\}^{1/2} N_n - \tfrac{1}{2} h^2 \kappa_2 f_X^{(4)}(\omega)\right] + o_p(h^2), \quad (15)$$

where $\kappa = \int (K'')^2$, $\kappa_2 = \int u^2 K(u)\,du$, and the random variable N_n has a standard normal distribution. [The departure of the distribution of N_n from standard normal may be absorbed into the $o_p(h^2)$ term.] Note that, because h is of size $n^{-1/9}$, the terms in $(nh^5)^{-1/2}$ and h^2 on the right side of (15) are both of size $n^{-2/9}$. An outline proof of (15) is given in Appendix A.3.

Next, we suppose that the density f_Z of Z may be represented in the form $\sigma^{-1} g(\cdot/\sigma)$, where we allow $\sigma = \sigma(n)$ to converge to 0 as $n \to \infty$. This asymptotic regime will give us insight into properties of the estimator $\hat{\omega}$ in cases where simultaneously sample size is moderately large and error variance is moderately small. In such contexts it is reasonable to ask that the bandwidth, h, be of smaller order than σ; otherwise, the kernel smoothing step will blur the endpoint θ of the density f_Y to at least the same extent as adding the errors Z_i, and so can be expected to hinder rather than help estimation of θ from some points of view.

To further elaborate on this technique, we note that in asymptotic analysis one usually keeps the model fixed as n increases. There the parameter of interest is identifiable, and the parameter estimator generally converges to the parameter. In the present setting, however, the parameter is not identifiable, and we rely on low noise to achieve relatively low systematic error. We have suggested methods for reducing the order of magnitude of this bias, but because the model is not identifiable, we are not able to reduce bias to 0, even in the limit as $n \to \infty$, unless we allow σ to decrease to 0 as n increases. [The dichotomy between nonvanishing systematic error if σ does not converge to 0 and asymptotically negligible stochastic error as $n \to \infty$ is evident in the two remainder terms on the right sides of (6) and (13).] Taking $\sigma = \sigma(n)$ to converge to 0 is essential if we are to achieve a limit theorem of conventional type that demonstrates consistency in the standard way.

Assume six bounded derivatives of f_Y and g, and that $g''(0) \neq 0$, $h = o(\sigma)$, $\sigma = O(n^{-\epsilon})$, $n^{1-\epsilon} h \to \infty$,

$$h = o\{\sigma^{6/5} + (n^{-1}\sigma^{14})^{1/15}\}, \quad (16)$$

and

$$\sigma = O\{(nh^5)^{1/4} n^{-\epsilon}\}, \quad (17)$$

all holding for some $\epsilon > 0$. (Note that we do not require that $nh^5 \to \infty$.) Suppose also that K satisfies the conditions imposed before (15). Then

$$\hat{\omega} - \theta = \frac{f_Y'(\theta-)g(0)}{f_Y(\theta-)|g''(0)|}\sigma^2 + \sigma^3 (nh^5)^{-1/2} \times \{\kappa^{1/2} f_Y(\theta-)^{3/2} |g''(0)|\}^{-1} M_n + o_p(\sigma^2), \quad (18)$$

where M_n denotes a random variable that is asymptotically normal $N(0, 1)$. An outline proof is given in Appendix A.4.

Note that in (18), in contrast to (15), we effectively treat $\hat{\omega}$ as an estimator of θ rather than of ω. Of course, this is possible because we now allow the error variance σ^2 to converge to 0. Taking this view, the principal contribution to the bias of $\hat{\omega}$, given by the first term on the right side of (18), is of order σ^2. This is one of the consequences of taking h to be of smaller order than σ; without that assumption, there would be an extra term in h^2 in the bias contribution, much as in (15).

Observe also that the stochastic error term, given by the second component on the right side of (18), is now of size $\sigma^3 (nh^5)^{-1/2}$ rather than $(nh^5)^{-1/2}$, as in (15). This is a consequence of the fact that the value of $\omega = \operatorname{argmax}|f_X'|$ becomes more easy to estimate when σ is small, because the peak in f_X' becomes more pronounced.

Property (18) implies that the asymptotic MSE (AMSE) of $\hat{\omega}$ is of size $\sigma^4 + \sigma^6 (nh^5)^{-1}$ and is minimized by taking h to be of size $(\sigma^2/n)^{1/5}$. This is of course a very different-sized bandwidth than that in the problem of estimating ω for a fixed error distribution, because we are treating a different problem here, considering $\hat{\omega}$ to be an estimator of θ rather than ω. The constraint $h \asymp (\sigma^2/n)^{1/5}$ is allowed by our regularity conditions, provided that σ does not converge to 0 too quickly; specifically, we need $n^{-1/4} = O(\sigma n^{-\epsilon})$, as well as $\sigma = O(n^{-\epsilon})$, for some $\epsilon > 0$.

For the optimal choice of bandwidth, the minimum AMSE of $\hat{\omega}$ as an estimator of θ is $O(\sigma^4)$. Note that this is the same order as the MSE of the nonstochastic approximation ω to θ; see Section 2.2 for that result.

APPENDIX: PROOFS OF MAIN THEORETICAL RESULTS

A.1 Derivation of (8)

Put $A_j = (-1)^j (j!)^{-1} f_Y^{(j)}(\theta-)$ for $j = 0, \ldots, 5$, and

$$A_{ji} = A_j \begin{cases} 2(-1)^{i+1} j! \{(j-i-1)!\}^{-1} \int |y|^{j-i-1} g(y)\, dy & \text{if } i \le j-1 \\ (-1)^{i+1} j!\, g^{(i-j)}(0) & \text{if } i \ge j. \end{cases}$$

Assuming condition (C_2), the expansion in (3) may be extended to

$$\begin{aligned} -f_X'(x) &= A_0\{f_Z(0) + \tfrac{1}{2}(x-\theta)^2 f_Z''(0) + \tfrac{1}{24}(x-\theta)^4 f_Z^{(4)}(0)\} \\ &\quad + A_1\{F_Z(0) - (x-\theta) f_Z(0) - \tfrac{1}{6}(x-\theta)^3 f_Z''(0)\} \\ &\quad + \sum_{j=2}^{4} \sum_{i=0}^{4-j} A_{ji} \sigma^{j-i-1} (x-\theta)^i + O(\sigma^4). \end{aligned}$$

[Condition (C_2) allows us to drop terms in f_Z' and $f_Z^{(3)}$.] Equivalently,

$$\begin{aligned} -f_X'(x) = B_0 + B_1(x-\theta) - B_2(x-\theta)^2 + B_3(x-\theta)^3 \\ + B_4(x-\theta)^4 + O(\sigma^4) \qquad \text{(A.1)} \end{aligned}$$

uniformly in $|x-\theta| \le C\sigma^2$ for any $C > 0$, where

$$\begin{aligned} B_0 &= A_0 f_Z(0) + A_1 F_Z(0) + A_{20}\sigma + A_{30}\sigma^2 + A_{40}\sigma^3, \\ B_1 &= -A_1 f_Z(0) + A_{21} + A_{31}\sigma, \\ B_2 &= \tfrac{1}{2} A_0 |f_Z''(0)| - A_{22}\sigma^{-1}, \\ B_3 &= -\tfrac{1}{6} A_1 f_Z''(0), \end{aligned}$$

and

$$B_4 = \tfrac{1}{24} A_0 f_Z^{(4)}(0).$$

Note that A_{23}, which otherwise would appear in the expansion of B_3, equals 0, and that higher-order terms in expansions of $B_0, \ldots, B_4$ may be neglected because their contributions, when multiplied by the appropriate power of $x - \theta$, are of the same order as the $O(\sigma^4)$ remainder in (A.1).

Therefore, defining $\alpha_j = f_Z^{(j)}(0)$ and $\omega = \operatorname{argmax}|f_X'(x)|$, we see that

$$\omega = \theta + \frac{A_1\alpha_0}{A_0\alpha_2} + O(\sigma^3) = \theta + O(\sigma^2)$$

and, uniformly in $|u| \le c\sigma^2$ for any $c > 0$,

$$-f_X'(\omega + u) = C_0 - C_2 u^2 + C_3 u^3 + C_4 u^4 + O(\sigma^4),$$

where $C_0 = A_0\alpha_0 + O(1)$, $C_2 = \frac{1}{2}A_0|\alpha_2| + O(\sigma^{-2})$, $C_4 = \frac{1}{24}A_0\alpha_4 + O(\sigma^{-4})$, and

$$C_3 = \tfrac{1}{6} A_1 |\alpha_2| \left(1 - \frac{\alpha_0\alpha_4}{\alpha_2^2}\right) + O(\sigma^{-2}).$$

This proves both (8) and the claimed formulas for C_0, C_2, C_3, and C_4.

A.2 Derivation of Result (R) of Section 3

For simplicity, we take $\mathcal{C}$ to be a non–self-intersecting curve passing from one side of a closed rectangle $\mathcal{R}$ to the other that does not touch the boundary of $\mathcal{R}$ except at the two points on opposite sides of $\mathcal{R}$. Assume that in a neighborhood of each point on $\mathcal{C}$, the curve has a Cartesian representation in terms of functions with three uniformly bounded derivatives. That is, if P is a point on $\mathcal{C}$, and if the axes are translated and rotated so that the $x^{(1)}$ axis passes through P and is tangential to $\mathcal{C}$ at P, then in the neighborhood of P, $\mathcal{C}$ is the locus of points $(x^{(1)}, x^{(2)})$, where $x^{(2)} = \psi_P(x^{(1)})$ and the function ψ_P has three derivatives bounded uniformly in P. Suppose also that $f_Y(y)$ can be written as $f(y)I(y \in \mathcal{Y})$, where the function f has two continuous derivatives in $\mathcal{R}$, $\mathcal{Y}$ denotes that part of $\mathcal{R}$ lying on a specific side of $\mathcal{C}$, and $f(\theta) > 0$ for $\theta \in \mathcal{C}$. Take $f_Z = \sigma^{-2} g(\cdot/\sigma)$, where the fixed bivariate density g is compactly supported and unimodal with its mode at 0, has two continuous derivatives, and satisfies $g(z) = g(0) - z^{\mathrm{T}} Q z + o(\|z\|^2)$ as $z \to 0$, with Q a strictly positive-definite 2×2 matrix.

Let $\theta = (\theta^{(1)}, \theta^{(2)})$ denote a point on $\mathcal{C}$, and let $\mathcal{T}(\theta)$ represent the set of points v such that $f_Y(\theta + v) > 0$. We may assume without loss of generality that the tangent to $\mathcal{C}$ at θ is not parallel to either of the coordinate axes; if it is, then rotate the axes slightly. If f_W denotes the density of a random two-vector W, then define

$$f_W^{(i_1, \ldots, i_r)}(u) = \frac{\partial^{i_1 + \cdots + i_r}}{\partial u^{(i_1)} \cdots \partial u^{(i_r)}} f_W(u)$$

and

$$A^{(i_1, \ldots, i_r)}(\theta) = (r!)^{-1} \lim_{v \to 0,\, v \in \mathcal{T}(\theta)} f_Y^{(i_1, \ldots, i_r)}(\theta + v).$$

Then

$$\begin{aligned} &f_X(\theta - u) \\ &\quad = \int f_Y(\theta + v) f_Z(-u - v)\, dv \\ &\quad = \int_{\mathcal{T}(\theta)} \Bigg\{ A_0(\theta) + \sum_{i=1}^{2} A_1^{(i)}(\theta) v^{(i)} + \sum_{i_1, i_2} A_2^{(i_1, i_2)}(\theta) v^{(i_1)} v^{(i_2)} \\ &\qquad + \sum_{i_1, i_2, i_3} A_3^{(i_1, i_2, i_3)}(\theta) v^{(i_1)} v^{(i_2)} v^{(i_3)} + \cdots \Bigg\} f_Z(-u - v)\, dv. \end{aligned}$$

(Here and later we represent Taylor expansions in a formal sense, without concise estimates of remainder terms, to clarify the nature of our arguments.) Formulas for derivatives of $f_X(\theta - u)$ with respect to components of u may be developed similarly. Thus

$$\begin{aligned} &f_X^{(j_1, \ldots, j_r)}(\theta) \\ &\quad = \int_{\mathcal{T}(\theta)} \Bigg\{ A_0(\theta) + \sum_{i=1}^{2} A_1^{(i)}(\theta) v^{(i)} + \sum_{i_1, i_2} A_2^{(i_1, i_2)}(\theta) v^{(i_1)} v^{(i_2)} \\ &\qquad + \sum_{i_1, i_2, i_3} A_3^{(i_1, i_2, i_3)}(\theta) v^{(i_1)} v^{(i_2)} v^{(i_3)} + \cdots \Bigg\} f_Z^{(j_1, \ldots, j_r)}(v)\, dv. \end{aligned} \qquad \text{(A.2)}$$

It follows from (A.2), the unimodality and smoothness of g, and the smoothness of the boundary of $\mathcal{T}(\theta)$ that $f_X^{(i)}(\theta) = O(\sigma^{-2})$ and $f_X^{(i_1, i_2)}(\theta) = O(\sigma^{-2})$. Call these collective results (R_1). To derive a more concise formula for $f_X^{(i)}(\theta)$ and a formula for $f_X^{(i_1, i_2)}(\theta)$, we assume without loss of generality that (a) θ lies at the origin of the $(x^{(1)}, x^{(2)})$ coordinate system, (b) the $x^{(1)}$ axis is parallel to the tangent to the boundary of $\mathcal{T}(\theta)$ at θ, and (c) the positive half of the $x^{(2)}$ axis is on the side of the $x^{(1)}$ axis away from $\mathcal{T}(\theta)$. Then, by (A.2),

$$\begin{aligned} f_X^{(i)}(\theta) &= A_0 \int_{\mathcal{T}(\theta)} f_Z^{(i)}(v)\, dv + o(\sigma^{-2}) \\ &= \begin{cases} A_0(\theta)\sigma^{-2} g(0) + o(\sigma^{-2}) & \text{if } i = 2 \\ o(\sigma^{-2}) & \text{if } i = 1 \end{cases} \end{aligned} \qquad \text{(A.3)}$$

and

$$f_X^{(i,j,k)}(\theta) = A_0 \int_{\mathcal{Z}(\theta)} f_Z^{(i,j,k)}(v)\,dv + o(\sigma^{-6})$$
$$= \begin{cases} A_0(\theta)\sigma^{-6} g^{(2,2)}(0) + o(\sigma^{-6}) & \text{if } i=j=2 \\ o(\sigma^{-6}) & \text{otherwise.} \end{cases} \quad (A.4)$$

Put $\beta_i(u) \equiv f_X^{(i,1)}(\theta+u) f_X^{(1)}(\theta+u) + f_X^{(i,2)}(\theta+u) f_X^{(2)}(\theta+u)$, and note that

$$\gamma_{ij}(u) \equiv \frac{\partial}{\partial u^{(j)}} \beta_i(u)$$
$$= f_X^{(i,j,1)}(\theta+u) f_X^{(1)}(\theta+u) + f_X^{(i,j,2)}(\theta+u) f_X^{(2)}(\theta+u)$$
$$+ f_X^{(i,1)}(\theta+u) f_X^{(j,1)}(\theta+u)$$
$$+ f_X^{(i,2)}(\theta+u) f_X^{(j,2)}(\theta+u). \quad (A.5)$$

Result (R_1) implies that $\beta_i(0) = O(\sigma^{-6})$; call this result (R_2). Substituting (A.3) and (A.4) into (A.5) and using (R_1), we deduce that

$$\gamma_{ij}(0) = \begin{cases} A_0(\theta)^2 \sigma^{-8} g(0) g^{(2,2)}(0) + o(\sigma^{-8}) & \text{if } i=j=2 \\ o(\sigma^{-8}) & \text{otherwise.} \end{cases} \quad (A.6)$$

Note that $\beta_i(u) = \frac{1}{2}(\partial/\partial u^{(i)})|f_X'(\theta+u)|^2$. Result (A.6) shows that a ridge line of the surface $v = |f_X'(\theta+u)|^2$, as defined at the end of Section 3, is asymptotically, as $\sigma \to 0$, in the direction parallel to the $x^{(2)}$ axis. By constructing a two-term Taylor expansion of $\beta_2(u)$ around $u=0$, using (R_2) and (A.6) to obtain the zero and first derivatives of $\beta_2(u)$ at $u=0$, and thereby obtaining an asymptotic solution of the equation $\beta_2(u)=0$, we deduce that the ridge line passes a distance $O(\sigma^{-6}/\sigma^{-8}) = O(\sigma^2)$ from the point on $\mathcal{C}$ with coordinates θ.

A.3 Derivation of (15)

By Taylor expansion, and because $f_X''(\omega) = 0$, $f_X''(\omega + x) = x f_X'''(\omega) + O(x^2)$ as $x \to 0$. Likewise,

$$E\{\hat f_X''(x)\} = f_X''(x) + \tfrac{1}{2} h^2 \kappa_2 f_X^{(4)}(x) + o(h^2)$$

uniformly in $x \in \mathcal{I}$, as $h \to 0$. Also, defining $D(x) = (nh^5)^{1/2} \times \{\hat f_X''(x) - E\hat f_X''(x)\}$, we find that $D(x)$ is asymptotically normally distributed with mean 0 and variance $(nh^5)^{-1}\kappa f_X(x)$, and, more generally, $\xi(u) \equiv D(\omega + hu)$ converges weakly, as a stochastic process defined on the interval $\mathcal{J} = [-c, c]$ for any fixed $c > 0$, to a Gaussian process $\xi_0(u)$ whose covariance function is

$$\gamma(u,v) = f_X(x) \int K''(u+w) K''(v+w)\,dw.$$

Furthermore, theorem 3 of Komlós, Major, and Tusnády (1975) may be used to prove that

$$|\xi(u) - \xi(0)| = O\Big[\{|u| \log(1+|u|)\}^{1/2}\Big]$$

with probability 1, uniformly in values u such that $\omega + hu \in \mathcal{I}$.

Combining these results, we deduce first that with probability 1,

$$\hat f_X''(\omega+hu) = \hat f_X''(\omega) + hu\, f_X^{(3)}(\omega) + O\Big(h^2\Big[\{|u|\log(1+|u|)\}^{1/2} + u^2\Big]\Big),$$

uniformly in u such that $\omega + hu \in \mathcal{I}$, and second that $\hat f_X''(x) - f_X''(x) \to 0$ uniformly in $x \in \mathcal{I}$. It follows from these two properties and the properties assumed of f_X'' on $\mathcal{I}$ that the value $\hat\omega$ of x that maximizes $|\hat f_X'(x)|$ satisfies $(\hat\omega - \omega)/h \to 0$ with probability 1. Therefore,

$$0 = \hat f_X''(\hat\omega) = (\hat\omega - \omega) f_X^{(3)}(\omega) + \tfrac{1}{2} h^2 \kappa_2 f_X^{(4)}(\omega)$$
$$+ (nh^5)^{-1/2} D(\omega) + o_p(h^2). \quad (A.7)$$

It is straightforward to derive (A.7) when the term $(nh^5)^{-1/2}D(\omega)$ on the right side is replaced by $(nh^5)^{-1/2}D(\hat\omega)$. However, the weak convergence of ξ to ξ_0 and the fact that $\hat\omega = \omega + o(h)$ with probability 1 imply that $(nh^5)^{-1/2}D(\hat\omega) = (nh^5)^{-1/2}D(\omega) + o_p\{(nh^5)^{-1/2}\}$. And because $h \asymp n^{-1/9}$, the remainder here equals $o_p(h^2)$. Solving (A.7) for $\hat\omega - \omega$, we deduce (15).

A.4 Derivation of (18)

Under the assumed conditions, the following version of (A.1) is valid:

$$-f_X'(\theta+u) = B_0 + B_1 u - B_2 u^2 + B_3 u^3 + B_4 u^4 + O(\sigma^{-5}|u|^5 + \sigma^{-7}|u|^6),$$

uniformly in $|u| \le C$ for any $C > 0$. Likewise, the derivative of this formula is valid:

$$-f_X''(\theta+u) = B_1 - 2B_2 u + 3B_3 u^2 + 4B_4 u^3 + O(\sigma^{-5}|u|^4 + \sigma^{-7}|u|^5),$$

uniformly in $|u| \le C$. Therefore,

$$E\{\hat f_X''(\theta+u)\} = \int K(v) f_X''(\theta+u-hv)\,dv$$
$$= B_1 - 2B_2 u + 3B_3(u^2 + h^2\kappa_2) + 4B_4(u^3 + 3\kappa_2 h^2 u)$$
$$+ O\{\sigma^{-5}(u^4+h^4) + \sigma^{-7}(|u|^5 + h^5)\}. \quad (A.8)$$

Note that $B_2 \asymp \sigma^{-3}$, $B_3 = O(\sigma^{-3})$, and $B_4 = O(\sigma^{-5})$. Therefore, because $h = o(\sigma)$, $B_4 h^2 = o(B_2)$. Put $\delta_\epsilon = \sigma^2 + \sigma^3(nh^5)^{-1/2} n^\epsilon$ and note that (17) and the property $\sigma \to 0$ imply that $\delta_\epsilon \to 0$ for some $\epsilon > 0$. It follows from (16) that $\sigma^{-7}h^5 = o(\sigma^{-3}\delta_0)$, where δ_0 denotes δ_ϵ with $\epsilon = 0$. Stated more simply, the fact that $h = o(\sigma)$ implies that $\sigma^{-5}h^4 = o(\sigma^{-3}\delta_0)$. Using (17) and the property $h = o(\sigma)$, we may deduce that

$$|u| + \sigma^{-2}u^2 + \sigma^{-2}|u|^3 + \sigma^{-4}u^4 \to 0, \quad (A.9)$$

uniformly in values of u such that $|u| \le C(h^2 + \delta_\epsilon)$ for any $C > 0$ and some $\epsilon > 0$. Now a constant multiple of the left side of (A.9) dominates $(|B_3 u| + |B_4| u^2 + \sigma^{-5}|u|^3 + \sigma^{-7}u^4)/B_2$. Combining these results and (A.8), we deduce that, uniformly in u such that $|u| \le C(h^2 + \delta_\epsilon)$, for some $\epsilon > 0$,

$$E\{\hat f_X''(\theta+u)\} = \{B_1 + o(\sigma^{-1})\} - 2B_2\{1 + o(1)\}u$$
$$+ o\{\sigma^{-1} + (nh^5)^{-1/2}\}.$$

From this property, and borrowing from the argument used in Appendix A.3, we deduce that the solution $u = \hat u = \hat\omega - \theta$ of the equation $\hat f_X''(\theta+u) = 0$ satisfies

$$0 = \{B_1 + o_p(\sigma^{-1})\} - 2B_2\{1 + o_p(1)\}\hat u$$
$$+ (nh^5)^{-1/2} D(\omega) + o_p\{(nh^5)^{-1/2}\}.$$

Solving for $\hat u$, we deduce (18). The asymptotic normality of $D(\omega)$ requires the assumption that $nh \to \infty$, but does not need the assumption that $nh^5 \to \infty$.

[Received April 2000. Revised March 2001.]

REFERENCES

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977), "Formulation and Estimation of Stochastic Frontier Models," *Journal of Econometrics*, 6, 21–37.

Banker, R. D. (1993), "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation," *Management Science*, 39, 1265–1273.

Breiman, L. (1996), "Bagging Predictors," *Machine Learning*, 26, 123–140.

Carroll, R. J., and Hall, P. (1988), "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184–1186.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978), "Measuring the Inefficiency of Decision Making Units," *European Journal of Operational Research*, 2, 429–444.

Christensen, L., and Greene, R. (1976), "Economies of Scale in U.S. Electric Power Generation," *Journal of Political Economy*, 84, 653–667.

Deprins, D., Simar, L., and Tulkens, H. (1984), "Measuring Labor Inefficiency in Post Offices," in *The Performance of Public Enterprises: Concepts and Measurements*, eds. M. Marchand, P. Pestieau and H. Tulkens, North-Holland: Amsterdam, pp. 243–267.

Fan, J. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

——— (1992), "Deconvolution With Supersmooth Distributions," *Canadian Journal of Statistics*, 20, 155–169.

——— (1993), "Adaptively Local One-Dimensional Subproblems With Application to a Deconvolution Problem," *The Annals of Statistics*, 21, 600–610.

Farrell, M. J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Ser. A, 120, 253–281.

Gijbels, I., Mammen, E., Park, B. U., and Simar, L. (1996), "On Estimation of Monotone and Concave Frontier Functions," *Journal of the American Statistical Association*, 94, 220–228.

Greene, W. H. (1990), "A Gamma-Distributed Stochastic Frontier Model," *Journal of Econometrics*, 46, 141–163.

Hall, P., Qian, W., and Titterington, D. M. (1992), "Ridge Finding From Noisy Data," *Journal of Computational and Graphical Statistics*, 1, 197–211.

Kneip, A., Park, B. U., and Simar, L. (1998), "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores," *Econometric Theory*, 14, 783–793.

Kneip, A., and Simar, L. (1996), "A General Framework for Frontier Estimation With Panel Data," *Journal of Productivity Analysis*, 7, 187–212.

Komlós, J., Major, P., and Tusnády, G. (1975), "An Approximation of Partial Sums of Independent RV's, and the Sample DF. I," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 32, 111–131.

Korostelev, A., Simar, L., and Tsybakov, A. B. (1995a), "Efficient Estimation of Monotone Boundaries," *The Annals of Statistics*, 23, 476–489.

——— (1995b), "On Estimation of Monotone and Convex Boundaries," *Publications des Instituts de Statistique des Universités de Paris*, XXXIX, 1, 3–18.

Meeusen, W., and van den Broek, J. (1977), "Efficiency Estimation From Cobb–Douglas Production Function With Composed Error," *International Economic Review*, 8, 435–444.

Park, B. U., Sickles, R. C., and Simar, L. (1998), "Stochastic Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273–301.

Park, B. U., and Simar, L. (1994), "Efficient Semiparametric Estimation in Stochastic Frontier Models," *Journal of the American Statistical Association*, 89, 929–36.

Park, B., Simar, L., and Weiner, Ch. (2000), "The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties," *Econometric Theory*, 16, 855–877.

Seiford, L. M. (1996), "Data Envelopment Analysis: The Evolution of the State-of-the-Art (1978–1995)," *Journal of Productivity Analysis*, 7, 99–137.

Shephard, R. W. (1970), *Theory of Cost and Production Function*, Princeton, NJ: Princeton University Press.