ESTIMATING A CHANGEPOINT, BOUNDARY OR FRONTIER IN THE PRESENCE OF OBSERVATION ERROR

Peter Hall [1] and Léopold Simar[1,2]

ABSTRACT. When stochastic errors are added to data from a distribution with a sharp boundary, such as a changepoint or a frontier, nonparametric estimation of the boundary can be interpreted as a problem of deconvolution. However, that view is not necessarily the most appropriate one, since convergence rates in deconvolution problems are often pathologically slow, particularly when the error distribution is Normal. For this reason we argue that, rather than attempting to estimate the distribution of the uncorrupted data, and thereby approximate the boundary, one might focus more directly on the boundary estimation problem. We suggest estimating the boundary by locating the steepest part of an estimate of the density of the corrupted data. It is shown that the bias intrinsic to this approach is of the same order as the variance, not simply the standard deviation, of the error distribution. This relatively fast rate of convergence in low-noise settings motivates our development of methodology in both univariate and bivariate cases. For example, we show that local linear approximations to the boundary in the univariate case can be used to improve accuracy, by reducing intrinsic bias from the square to the third power of standard deviation. Numerical and theoretical properties of estimators are described.

KEYWORDS. Changepoint analysis, data envelopment analysis, deconvolution, frontier estimation, kernel methods, nonparametric estimation, panel data, production frontier.

SHORT TITLE. Boundary estimation from noisy data.

[1]Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

[2]Institut de Statistique, Université Catholique de Louvain, B 1348 Louvain-la-Neuve, Belgium. Research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) from the Belgian Government is acknowledged.

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 endpoint of the distribution of Y_i is of particular interest in the context of frontier analysis, and we shall focus solely on that setting, although more general problems may be treated using similar methods.

Naturally, the problem of uniquely defining θ, as well as estimating it, requires restrictions on the distributions of Y and Z. The challenge is to find assumptions that are at once realistic, constrain the problem sufficiently narrowly to permit a unique characterisation of θ, and allow practicable estimators with reasonable rates of convergence. We shall suggest several related approaches to solving this problem.

Applications in frontier analysis motivate both univariate and bivariate contexts, each of which is treated in this paper. We shall shortly give detail of these applications and related methodologies. Several different approaches to inference are possible in bivariate contexts, their respective suitabilities depending on matters such as whether the distributions of X or Y may be regarded as genuinely "spatial" in character, and whether one of the components of X is replicated conditionally on the other.

The problem of estimating properties of the distribution of X from those of the distribution of Y, when X is obtained from Y by adding an independent error, is of course one of deconvolution. However, a direct assault on the problem from that viewpoint, in a nonparametric setting, is not necessarily productive. In particular, optimal nonparametric estimators of the distribution of Y from data on X, even when the distribution of Z is known, can have pathologically slow convergence rates. For example, the rates are only logarithmic in the sample size, n, if Z is Normally distributed; see Carroll and Hall (1988) and Fan (1991, 1992, 1993).

The appropriate approach is arguably not to attempt estimation of the full distribution of X, but instead to identify a feature of the distribution that is of particular practical interest, and confine attention to that. The feature considered here is the boundary, and we focus on the setting where error variance is relatively small; this is often the case in frontier estimation, since the amount of error in output data (e.g. company profits, or measures of customer service) is generally low relative to the true values of output. In this context we develop boundary-estimation methods based on nonparametric density estimation. In contrast to traditional deconvolution approaches, our methods are not adversely affected by assumptions such as Normal errors.

One of our contributions is the result that, if inference is conducted as though the density of Y were perfectly flat in the neighbourhood of its jump discontinuity (the "flat-density hypothesis"), then even when this condition is violated, the bias of the boundary estimator is of the same order as the variance, σ^2, not simply the standard deviation, of the error distribution.

The flat-density approximation suggests estimating the boundary as though it were a place where the density of X had maximum gradient, and this is the approach we take. The fast convergence rate, $O(\sigma^2)$, is available in both univariate and spatial-bivariate settings. In the latter case, 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 biases of estimators can be further reduced by estimating and correcting for asymmetries that are 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.

Next we discuss applications of our methods in the area of economics, where boundaries may be interpreted as production frontiers. There, one wishes to analyse how different firms transform a set of inputs (typically labour, energy or capital) into an output (typically a quantity of goods produced). The boundary represents the upper envelope of attainable production; it is the geometric locus of optimal production. The economic efficiency of a firm is defined in terms of the firm's ability to operate close to the boundary. In this context, firms can be compared in order to determine those that are more efficient.

There is a large econometric literature on boundary estimation. Parametric approaches have been suggested by Aigner, Lovell and Schmidt (1977), Meeusen and van den Broek (1977) and Greene (1990). In the context of panel data they have semiparametric generalisations; see Simar (1994) and Park, Sickles and Simar (1998). Nonparametric techniques, including those suggested in the present paper, 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, introduced by Charnes, Cooper and Rhodes (1978) for computing the DEA estimator, have proved particularly popular. Deprins, Simar and Tulkens (1984) extended the idea to non-convex sets, and suggested the "free disposal hull" (FDH) estimator, equal to the smallest free disposal set containing all 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) and Park, Simar and Weiner (1999).

Frontier estimation methods have been applied in many settings, including to public services, banks, hospitals and other institutions; see Seiford (1996) for an extensive survey. They may be extended to multivariate settings; see for example 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, as opposed to *stochastic* frontier models where the data are perturbed by noise. In the presence of noise, the envelopment techniques described above 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. By way of contrast, the conditions required in the present paper are very mild.

2 Methodology in 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-hand 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 zero for all $y > \theta$. We assume too, 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 neighbourhood of θ). The unimodality property assumed of f_Z therefore implies that θ can be characterised as that point, ω say, in $[a+c, \infty)$ at which the slope of the density of X achieves its greatest absolute value.

Of course, this characterisation 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. Section 2.1 will 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, which is not an 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 maximised, is attractive. For example, we may construct a nonparametric estimator $\hat{f}_X$ of f_X, and estimate θ by $\hat{\theta} = \operatorname{argmax} |\hat{f}'_X|$, where the "argmax" is taken in the neighbourhood of the true value of θ. See section 2.4 for details.

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, the approximation should nevertheless become increasingly accurate as the standard deviation, σ say, of Z decreases. The assumption that the "signal to noise ratio" — that is, the ratio $\mathrm{var}(Y)/\mathrm{var}(Z)$ — is high is often appropriate in practice. We shall explore these issues in detail in our numerical work in section 4.

We shall show below 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 might have been expected in this problem. The $O(\sigma^2)$ convergence rate will be verified 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 neighbourhood of 0, and (e) $g''(0) \neq 0$. Call these conditions (C_1). The assumption that g is compactly supported is made only so as 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. In contrast, a direct approach to deconvolution is impractical in the Normal setting (Carroll and Hall, 1988).

Assume initially that

$$f_Y(\theta - y) = A_0 + A_1 y + A_2 y^2 \,, \tag{2.1}$$

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

$$\begin{aligned} -f_X'(x) = A_0\, f_Z(x-\theta) \quad &+ \quad A_1 \{1 - F_Z(x-\theta)\} \\ &+ \quad 2A_2 \int_0^\infty \{1 - F_Z(x - \theta + u)\}\, du \,, \end{aligned} \tag{2.2}$$

where F_Z denotes the distribution function corresponding to f_Z. For the sake of simplicity we have assumed here that f_Z is supported in a sufficiently small interval about its mode, 0. However, our 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-hand side of (2.2) in the context of fixed g, and noting that (C_1) implies $f_Z'(0) = 0$, it may be shown that as $\sigma \to 0$,

$$\begin{aligned} -f_X'(x) = A_0 \{f_Z(0) \quad &+ \quad \tfrac{1}{2}\,(x-\theta)^2\, f_Z''(0)\} \\ &+ \quad A_1 \{\tfrac{1}{2} - (x-\theta)\, f_Z(0)\} + A_2 \gamma \sigma + o(\sigma) \,, \end{aligned} \tag{2.3}$$

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

$$\omega = \theta + \frac{A_1 f_Z(0)}{A_0 f_Z''(0)} + o(\sigma^2) = \theta + O(\sigma^2) \tag{2.4}$$

as $\sigma \to 0$.

If instead of (2.1) 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 (2.3) is still valid, with a slightly more complex derivation. See Appendix (i) for an illustration of the argument.

To better appreciate the orders of magnitude claimed at (2.4), we should comment on the orders at (2.3). Since f_Z is modelled as $\sigma^{-1} g(\cdot/\sigma)$ for a fixed density g, then it follows from (C_1) that $f_Z(0) = \sigma^{-1}\, g(0)$ and 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 cubic terms at (2.3) are both of size σ, and the remainder $o(\sigma)$ is negligible. Likewise, the approximation error $\omega - \theta$ at (2.4) is of size σ^2. Indeed, (2.4) is equivalent to

$$\omega = \theta + \frac{f_Y'(\theta-)\ g(0)}{f_Y(\theta-)\ |g''(0)|}\,\sigma^2 + o(\sigma^2)\,. \tag{2.5}$$

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 neighbourhood 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 so as to improve its accuracy as an approximation to θ.

At the level of (2.3) 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" only becomes discernible if we progress to cubic and higher-order terms in the expansion. As we shall 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-hand side of (2.4), 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, although we shall not consider them 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 (2.3) 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 neighbourhood of 0, (e) $g^{(j)}(0) = 0$ for $j = 1, 3$, and (f) $g''(0) \neq 0$. Call these conditions (C_2). Property (e) is a little weaker than asking that the distribution of Z be symmetric.

We shall prove in Appendix (i) that under (C_2),

$$-f_X'(\omega + u) = C_0 - C_2 u^2 + C_3 u^3 + C_4 u^4 + O(\sigma^4)\,, \tag{2.6}$$

where

$$
\begin{aligned}
C_0 &= f_Y(\theta-)\, f_Z(0) + O(1), \\
C_2 &= \tfrac{1}{2}\, f_Y(\theta-)\, |f_Z''(0)| + O(\sigma^{-2}), \\
C_3 &= -\tfrac{1}{6}\, f_Y'(\theta-)\ |f_Z''(0)| \left(1 - \frac{f_Z(0)\, f_Z^{(4)}(0)}{f_Z''(0)^2}\right) + O(\sigma^{-2})\,, \\
C_4 &= \tfrac{1}{24}\, f_Y(\theta-)\, f_Z^{(4)}(0) + O(\sigma^{-4})\,.
\end{aligned}
$$

It may be deduced from (2.6) that, assuming (C_2), the remainder $o(\sigma^2)$ at (2.5) is actually $O(\sigma^3)$:

$$
\omega = \theta + \frac{f_Y'(\theta-)\ g(0)}{f_Y(\theta-)\ |g''(0)|}\, \sigma^2 + O(\sigma^3)\,. \tag{2.7}
$$

More importantly, (2.6) gives us the leverage we need to correct the approximation ω (of θ) for the term of order σ^2. In particular, (2.6) includes a cubic term, which would vanish if f_Y were perfectly flat in a neighbourhood immediately to the left of θ (i.e. if $A_1 = 0$), but which is in general nonzero.

Property (2.7), and the formulae for $C_0, \ldots, C_4$, imply that

$$
\frac{3C_0C_3}{2(C_2^2 - 6C_0C_4)} = (\theta - \omega)\, \{1 + O(\sigma)\}
$$

as $\sigma \to 0$. Therefore, since $\omega = \theta + O(\sigma^2)$ then

$$
\theta = \omega + \frac{3C_0C_3}{2(C_2^2 - 6C_0C_4)} + O(\sigma^3)\,. \tag{2.8}
$$

Formula (2.8) suggests the following approach to bias-correcting a flat-density approximation estimator of θ. Estimate f_X' by $\hat f_X'$, say, and put $\widehat\omega = \operatorname{argmax} |\hat f_X'|$. Next, estimate constants $\widehat C_0, \ldots, \widehat C_4$ by fitting a quartic to $-\hat f_X'$ in a neighbourhood of $\widehat\omega$, obtaining

$$
-\hat f_X'(\widehat\omega + u) \approx \widehat C_0 - \widehat C_2 u^2 + \widehat C_3 u^3 + \widehat C_4 u^4\,, \tag{2.9}
$$

say. Take

$$
\hat\theta = \widehat\omega + \frac{3\widehat C_0 \widehat C_3}{2(\widehat C_2^2 - 6\widehat C_0 \widehat C_4)} \tag{2.10}
$$

to be an estimator of θ.

It is implicit in this argument that $C_2^2 - 6C_0C_4$ does not vanish. Now,

$$
C_2^{-2}\, (C_2^2 - 6C_0C_4)\, \{1 + O(\sigma)\} = 1 - \frac{g(0)\, g^{(4)}(0)}{g''(0)^2}\,,
$$

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 that are present in formulae such as (2.6), we note that C_0, C_2, C_3, C_4 are asymptotic to constant multiples of $\sigma^{-1}, \sigma^{-3}, \sigma^{-3}, \sigma^{-5}$ respectively, as $\sigma \to 0$. Therefore, if u equals a constant multiple of σ^2 (the same size as $\omega - \theta$) then the terms $C_0, C_2u^2, C_3u^3, C_4u^4$ at (2.6) are asymptotic to constant multiples of $\sigma^{-1}, \sigma, \sigma^3, \sigma^3$ respectively. In particular, the cubic and quartic terms at (2.6) 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 at (2.6). These properties justify the claim, implicit in our motivation of the estimator $\hat{\theta}$, that the polynomial at (2.9) 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

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

where K is a kernel function and h is a bandwidth. We shall 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, means they are not so attractive when the task at hand is to find turning points of derivatives.

Define $\widehat{\omega} = \operatorname{argmax} |\hat{f}_X'(x)|$, where the "argmax" is taken over a closed interval $\mathcal{I}$ in the right-hand tail of the distribution of X. In describing properties of this estimator we treat first 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 minimising asymptotic mean squared error of $\widehat{\omega}$ as an estimator of ω, is $h \asymp n^{-1/9}$. For this choice of h the stochastic error of $\widehat{\omega}$, and the asymptotic bias, are both of size $n^{-2/9}$. Result (5.1) 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 $\widehat{\omega}$, 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 $X = Y + Z$, where again Y (the variable of intrinsic interest) and Z (representing error) are independent, but this time X, Y, 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, $\mathcal{C}$ say, but the sharpness of the change along $\mathcal{C}$ is blunted in the observed data X by the addition of errors Z.

There are at least three ways of generalising our methods for the univariate problem so as to approximate or estimate $\mathcal{C}$. The generalisations 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 vanishing second coordinate, but their first coordinate would have a continuous distribution with its mode at the origin.

Alternatively, in either this setting or that where there are 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 analysed 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 $\mathcal{X}_b(x^{(1)}) = \{X_i^{(2)} : x^{(1)} - b \le X_i^{(1)} \le x^{(1)} + b\}$, where $b > 0$ was a bandwidth, and estimate $\psi(x^{(1)})$ from $\mathcal{X}_b(x^{(1)})$. In practice one would typically take b 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 described above are distinctly non-spatial, 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. the number of employees of a company) and the second is an output (e.g. the annual profit of the company).

In spatial settings, however, both the model and the method would typically 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

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

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 are 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 (ii) we shall 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 $\widehat{S}$, defined as the locus of points (v, x) determined by $v = |\hat{f}_X'(x)|^2$, where

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

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

4 Numerical Properties

In this section we report the results of simulation studies in the univariate case, and 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, respectively. 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 $\operatorname{var} Y = 1/12$.

The first two models reflect assumptions that are 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 as $\mu < 0$ or $\mu > 0$, respectively. 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 $\widehat{\omega}$ would not be attempted. It is therefore 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, $E(Y) \approx \frac{1}{2}$. Various values of n and σ_z were chosen, the latter allowing us to tune the noise to signal ratio, $\rho_{\mathrm{nts}} = \sigma_Z/\sigma_Y$.

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 $\widehat{\omega}$ 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 $\widehat{\omega}$ by calculating the zero of $\hat{f}''_X(x)$ in a neighborhood of the previous value of $\widehat{\omega}$; and so on. We stopped iteration when the absolute value of the difference in the values of $\widehat{\omega}$ between two iterations increased. Thus we obtained both a value for h and the corresponding $\widehat{\omega}$.

The correction term, $\mathrm{ct} = 3\widehat{C}_0\widehat{C}_3/\{2(\widehat{C}_2^2 - 6\widehat{C}_0\widehat{C}_4)\}$, that is added to $\widehat{\omega}$ at (2.10) can be unstable since 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 $\widehat{\omega}$ from the original sample $\mathcal{X}$ as discussed above, we drew a bootstrap resample $\mathcal{X}^*$ from $\mathcal{X}$ and calculated the bootstrap versions $\widehat{\omega}^*$ and ct^* of $\widehat{\omega}$ and ct, respectively, for the given bandwidth h. The final correction term, $\widehat{\mathrm{ct}}$ say, may be taken to be any robust average of the values ct^* drawn by repeated resampling; we took $\widehat{\mathrm{ct}}$ to be the mean of those out of $B = 100$ values of ct^* that satisfied $|\mathrm{ct}^*| \le h$.

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 employed 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.

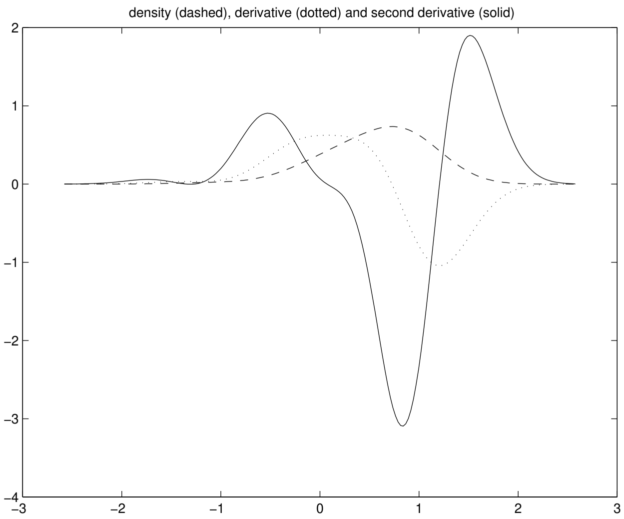

Figure 1: Density and its derivatives for typical sample of size $n = 100$ generated from Model 1 with $\theta = 1$, $\lambda = 2$, $\sigma_Z = 0.2$, $\mu_Y = \sigma_Y = 0.5$ and $\rho_{\mathrm{nts}} = 0.40$. For these data, $\max(X_i) = 1.2145$, $\widehat{\omega} = 1.2072$, $\hat{\theta} = 1.0767$ and $h = 0.3435$.

4.1.3 Results

Tables 1–9 give numerical approximations to mean squared error (MSE), bias, variance (var), and the standard deviation of bias (std(bias)), obtained by averaging over 200 simulations. It will be seen that $\hat{\theta} = \widehat{\omega} + \widehat{\text{ct}}$, the corrected version of $\widehat{\omega}$, almost always has less bias and less mean squared error than $\widehat{\omega}$. Indeed, $\hat{\theta}$ improves substantially on $\widehat{\omega}$ 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 commonly 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 $\widehat{\omega}$ 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 that is 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 that 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, since the "flat-density approximation" is valid, and it might be expected that attempting a correction would significantly reduce performance. This turns out not to be the case.

4.2 Real-data example in the bivariate case

4.2.1 Semiparametric fit of Cobb-Douglas model

Data on 123 American electric utility companies presented by Christensen and Greene (1976) have become a popular example for the study of frontier models. In Greene (1990), several fully parametric models are fitted and compared. The data are there described in detail: for each utility its 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:

$$y_i = \alpha + \beta' x - u_i + v_i \,, \tag{4.1}$$

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 or Gamma) for the probability density function of the inefficiency term u_i convolved with Normal for the noise v_i. The procedure is maximum likelihood in the spirit of the

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0267	0.1046	(0.0089)	0.0158
	$\hat{\theta}$	0.0052	0.0127	(0.0050)	0.0051
$n = 100$	$\hat{\omega}$	0.0040	0.0468	(0.0030)	0.0018
	$\hat{\theta}$	0.0008	0.0058	(0.0020)	0.0008
$n = 500$	$\hat{\omega}$	0.0013	0.0277	(0.0016)	0.0005
	$\hat{\theta}$	0.0003	0.0051	(0.0011)	0.0003
$n = 1000$	$\hat{\omega}$	0.0005	0.0178	(0.0009)	0.0002
	$\hat{\theta}$	0.0002	0.0053	(0.0009)	0.0002

Table 1: Monte-Carlo simulations Model 1: Exponential ($\lambda = 2$), $E(Y) = 0.5$, $\sigma_Y = 0.5$, $\sigma_Z = 0.05$ and $\rho_{\text{nts}} = 0.10$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0306	0.1253	(0.0087)	0.0150
	$\hat{\theta}$	0.0065	0.0215	(0.0055)	0.0060
$n = 100$	$\hat{\omega}$	0.0071	0.0689	(0.0035)	0.0024
	$\hat{\theta}$	0.0021	0.0120	(0.0032)	0.0020
$n = 500$	$\hat{\omega}$	0.0021	0.0422	(0.0013)	0.0004
	$\hat{\theta}$	0.0012	0.0141	(0.0022)	0.0010
$n = 1000$	$\hat{\omega}$	0.0016	0.0379	(0.0010)	0.0002
	$\hat{\theta}$	0.0009	0.0158	(0.0017)	0.0006

Table 2: Monte-Carlo simulations Model 1: Exponential ($\lambda = 2$), $E(Y) = 0.5$, $\sigma_Y = 0.5$, $\sigma_Z = 0.1$ and $\rho_{\text{nts}} = 0.20$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0099	0.0358	(0.0066)	0.0087
	$\hat{\theta}$	0.0059	-0.0294	(0.0050)	0.0050
$n = 100$	$\hat{\omega}$	0.0017	0.0207	(0.0026)	0.0013
	$\hat{\theta}$	0.0013	-0.0064	(0.0025)	0.0012
$n = 500$	$\hat{\omega}$	0.0004	0.0064	(0.0013)	0.0003
	$\hat{\theta}$	0.0003	-0.0027	(0.0011)	0.0002
$n = 1000$	$\hat{\omega}$	0.0002	0.0048	(0.0008)	0.0001
	$\hat{\theta}$	0.0001	-0.0020	(0.0007)	0.0001

Table 3: Monte-Carlo simulations Model 2: Half-Normal ($\sigma = 0.6267$), $E(Y) = 0.5$, $\sigma_Y = 0.3778$, $\sigma_Z = 0.0378$ and $\rho_{\text{nts}} = 0.10$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0113	0.0417	(0.0069)	0.0096
	$\hat{\theta}$	0.0066	-0.0174	(0.0056)	0.0064
$n = 100$	$\hat{\omega}$	0.0025	0.0238	(0.0031)	0.0020
	$\hat{\theta}$	0.0025	-0.0119	(0.0034)	0.0023
$n = 500$	$\hat{\omega}$	0.0005	0.0136	(0.0013)	0.0004
	$\hat{\theta}$	0.0005	-0.0013	(0.0016)	0.0005
$n = 1000$	$\hat{\omega}$	0.0003	0.0102	(0.0009)	0.0002
	$\hat{\theta}$	0.0003	-0.0027	(0.0012)	0.0003

Table 4: Monte-Carlo simulations Model 2: Half-Normal ($\sigma = 0.6267$), $E(Y) = 0.5$, $\sigma_Y = 0.3778$, $\sigma_Z = 0.0756$ and $\rho_{\text{nts}} = 0.20$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0151	0.0664	(0.0073)	0.0107
	$\hat{\theta}$	0.0079	-0.0064	(0.0063)	0.0079
$n = 100$	$\hat{\omega}$	0.0032	0.0378	(0.0030)	0.0018
	$\hat{\theta}$	0.0014	0.0028	(0.0026)	0.0014
$n = 500$	$\hat{\omega}$	0.0010	0.0249	(0.0015)	0.0004
	$\hat{\theta}$	0.0006	0.0043	(0.0017)	0.0005
$n = 1000$	$\hat{\omega}$	0.0005	0.0177	(0.0010)	0.0004
	$\hat{\theta}$	0.0003	0.0034	(0.0013)	0.0003

Table 5: Monte-Carlo simulations Model 3: Truncated-Normal ($\mu = -0.40$, $\sigma = 0.7834$), $E(Y) = 0.5$, $\sigma_Y = 0.4046$, $\sigma_Z = 0.0809$ and $\rho_{\text{nts}} = 0.20$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 100$	$\hat{\omega}$	0.0046	-0.0506	(0.0032)	0.0021
	$\hat{\theta}$	0.0060	-0.0612	(0.0033)	0.0022
$n = 500$	$\hat{\omega}$	0.0024	-0.0452	(0.0014)	0.0004
	$\hat{\theta}$	0.0022	-0.0395	(0.0018)	0.0006
$n = 1000$	$\hat{\omega}$	0.0019	-0.0416	(0.0010)	0.0002
	$\hat{\theta}$	0.0012	-0.0295	(0.0013)	0.0003

Table 6: Monte-Carlo simulations Model 3: Truncated-Normal ($\mu = 0.40$, $\sigma = 0.3768$), $E(Y) = 0.5$, $\sigma_Y = 0.3033$, $\sigma_Z = 0.0303$ and $\rho_{\text{nts}} = 0.10$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 100$	$\hat{\omega}$	0.0039	-0.0441	(0.0031)	0.0020
	$\hat{\theta}$	0.0064	-0.0611	(0.0037)	0.0027
$n = 500$	$\hat{\omega}$	0.0022	-0.0430	(0.0014)	0.0004
	$\hat{\theta}$	0.0024	-0.0419	(0.0018)	0.0007
$n = 1000$	$\hat{\omega}$	0.0020	-0.0418	(0.0011)	0.0002
	$\hat{\theta}$	0.0015	-0.0324	(0.0016)	0.0005

Table 7: Monte-Carlo simulations Model 3: Truncated-Normal ($\mu = 0.40$, $\sigma = 0.3768$), $E(Y) = 0.5$, $\sigma_Y = 0.3033$, $\sigma_Z = 0.0607$ and $\rho_{\text{nts}} = 0.20$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0393	-0.1109	(0.0117)	0.0272
	$\hat{\theta}$	0.0417	-0.1296	(0.0112)	0.0250
$n = 100$	$\hat{\omega}$	0.0164	-0.0635	(0.0079)	0.0125
	$\hat{\theta}$	0.0163	-0.0631	(0.0079)	0.0123
$n = 500$	$\hat{\omega}$	0.0009	-0.0129	(0.0019)	0.0007
	$\hat{\theta}$	0.0008	-0.0085	(0.0019)	0.0007
$n = 1000$	$\hat{\omega}$	0.0004	-0.0085	(0.0014)	0.0004
	$\hat{\theta}$	0.0004	-0.0054	(0.0014)	0.0004

Table 8: Monte-Carlo simulations Model 4: Decreasing density for Y with ($a = 1$, $b = 2$), $E(Y) = 0.4375$, $\sigma_Y = 0.2891$, $\sigma_Z = 0.0578$ and $\rho_{\text{nts}} = 0.20$.

sample size	estimate	MSE	bias	std(bias)	var
$n = 20$	$\hat{\omega}$	0.0159	-0.0313	(0.0086)	0.0150
	$\hat{\theta}$	0.0176	-0.0641	(0.0082)	0.0136
$n = 100$	$\hat{\omega}$	0.0019	-0.0016	(0.0031)	0.0019
	$\hat{\theta}$	0.0025	-0.0131	(0.0034)	0.0024
$n = 500$	$\hat{\omega}$	0.0004	-0.0016	(0.0014)	0.0004
	$\hat{\theta}$	0.0005	-0.0037	(0.0016)	0.0005
$n = 1000$	$\hat{\omega}$	0.0002	-0.0005	(0.0010)	0.0002
	$\hat{\theta}$	0.0003	-0.0022	(0.0011)	0.0002

Table 9: Monte-Carlo simulations Model 5: Uniform (0,1), $E(Y) = 0.5$, $\sigma_Y = 0.2887$, $\sigma_Z = 0.0577$ and $\rho_{\text{nts}} = 0.20$.

papers of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broek (1977) on parametric stochastic frontiers.

The method described in Section 2 can be used to motivate a semiparametric approach. The idea is straightforward: the parametric part is the Cobb-Douglas model as above, 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, respectively.

Arguing in this manner, we suggest the following estimation procedure. Rewrite (4.1) 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 $\widehat{\alpha}^\circ$ and $\widehat{\beta}$. Define the shifted OLS residuals,

$$\hat{\epsilon}_i = y_i - \widehat{\beta}' x_i \, , \quad i = 1, \ldots, n \, ,$$

which are estimators of the true residuals,

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

Using the method of Section 2, compute $\widehat{\alpha}$ from the sample $\hat{\epsilon}_i, i = 1, \ldots, n$. Note too that an estimator of $\mu = E(u)$ is given by

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

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, respectively.

4.2.2 Application of method of section 3

As noted in Section 3, the univariate methodology can be extended to multivariate cases. We shall consider a bivariate setting where the discontinuity is along a smooth boundary $\mathcal{C}$, and we shall illustrate the case where $\mathcal{C}$ is determined by the formula $x^{(2)} = \psi(x^{(1)})$. The framework will be non-spatial: $\psi(x^{(1)})$ could represent, for instance, a production frontier where $x^{(2)}$ is the maximum output 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 below we chose the bandwidth to be appropriate for estimating the marginal density of $X^{(1)}$. The projected data were analyzed using the algorithm described above, producing $\widehat{\psi}(x^{(1)})$. We shall briefly illustrate the application of this method in a Monte-Carlo experiment, and then we 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:

$$\begin{aligned}
&X^{(1)} \sim U[0,1], \text{ and } X^{(2)} = \psi(X^{(1)}) \exp(-V) \exp(W), \text{ where } \psi(x) = x^{1/2}, \\
&\text{with } V \sim \exp(3) \text{ and } W \sim N(0, (0.0667)^2)\,,
\end{aligned}$$

with the random variables $X^{(1)}, V, W$ being independent. Note that $E\{\exp(-V)\} = 3/4$. The noise to signal ratio is $\rho_{\text{nts}} = \sigma_W/\sigma_V = 0.20$. Figure 2 depicts a typical simulated dataset for a sample of size $n = 100$.

Table 10 summarizes simulation results for two sample sizes at three different points, 0.25, 0.50 and 0.75, 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 zero.

	$x^{(1)} = 0.25$		$x^{(1)} = 0.50$		$x^{(1)} = 0.75$	
n	bias (std(bias))	MSE	bias (std(bias))	MSE	bias (std(bias))	MSE
100	0.0048 (0.0017)	0.0014	0.0017 (0.0022)	0.0025	-0.0003 (0.0024)	0.0028
500	0.0009 (0.0010)	0.0005	0.0041 (0.0012)	0.0007	0.0041 (0.0014)	0.0011

Table 10: Estimated bias and MSE at three different values of $X^{(1)}$, for two sample sizes. Standard errors of the bias estimator are given in parentheses.

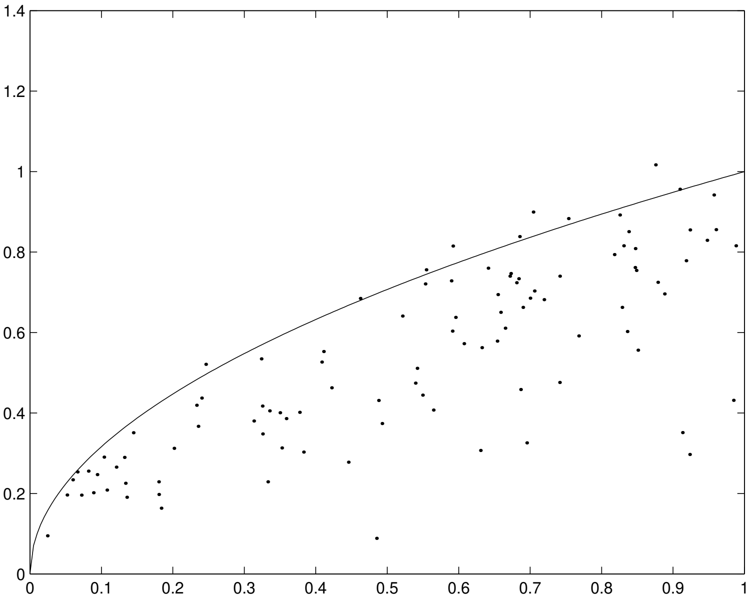

Figure 2: A typical simulated sample of size 100. The solid line represents the true frontier function.

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, together with pointwise estimates of $\psi(x^{(1)})$ over a selected grid of 50 values for $x^{(1)}$. Bandwidth was taken equal to 0.51. The figure also shows a smooth estimator of the frontier obtained by running a kernel smoother (quartic kernel and same bandwidth 0.51) through the estimated boundary points. In order to avoid edge effects we restricted the region of estimation to $[1.5, 5]$.

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 $\omega = \text{argmax}\,|f'_X(x)|$ is an interior point of $\mathcal{I}$, and the density f_X has four continuous derivatives in an open interval containing $\mathcal{I}$, with $f_X^{(3)}(\omega) \neq 0$, $f_X^{(2)}(\omega) = 0$ and $f_X^{(2)}(x) \neq 0$ for all points $x \in \mathcal{I}$ with $x \neq \omega$; and suppose K is a compactly supported, symmetric probability density with three bounded

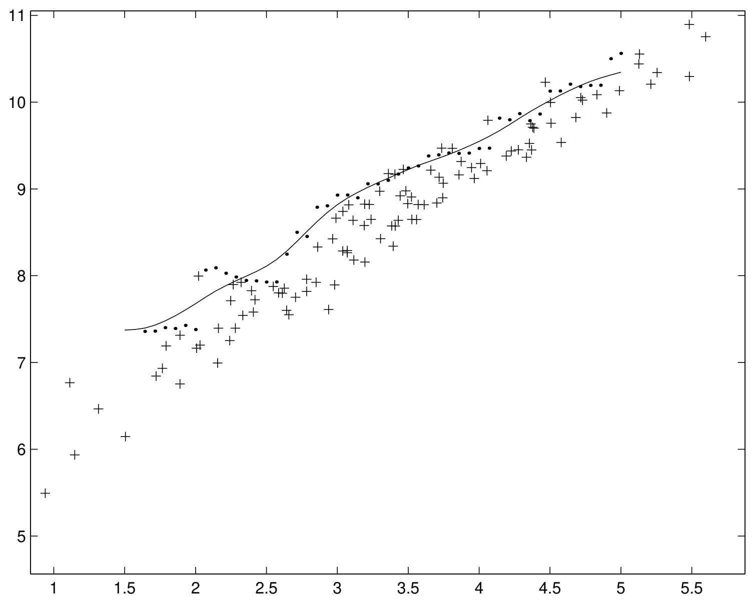

Figure 3: Scatterplot of the American Electric Utility data. The +'s represent the observations, the ·'s are the pointwise estimates of the boundary over a selected grid of 50 values for $x^{(1)}$, and the solid line is a smooth version of these boundary points. The bandwith for $X^{(1)}$ is 0.5128.

derivatives, and $h \asymp n^{-1/9}$. Then,

$$\widehat{\omega} - \omega = \{f_X^{(3)}(\omega)\}^{-1} \Big[(nh^5)^{-1/2} \{\kappa\, f_X(x)\}^{1/2}\, N_n - \tfrac{1}{2}\, h^2 \kappa_2\, f_X^{(4)}(\omega)\Big] + o_p(h^2)\,, \tag{5.1}$$

where $\kappa = \int (K'')^2$, $\kappa_2 = \int u^2\, K(u)\, du$, and the random variable N_n has a standard Normal distribution. Note that, since h is of size $n^{-1/9}$, the terms in $(nh^5)^{-1/2}$ and h^2 on the right-hand side of (5.1) are both of size $n^{-2/9}$. An outline proof of (5.1) is given in Appendix (iii).

Next we suppose 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 $\widehat{\omega}$ 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.

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

$$h = o\{\sigma^{6/5} + (n^{-1}\sigma^{14})^{1/15}\}\,, \tag{5.2}$$

$$\sigma = O\{(nh^5)^{1/4} n^{-\epsilon}\}\,, \tag{5.3}$$

all holding for some $\epsilon > 0$. (Note that we do not require $nh^5 \to \infty$.) Suppose too that K satisfies the conditions imposed prior to (5.1). Then,

$$\begin{aligned} \widehat{\omega} - \theta \;=\; & \frac{f_Y'(\theta-)\, g(0)}{f_Y(\theta-)\, |g''(0)|}\, \sigma^2 \\ & + \sigma^3\, (nh^5)^{-1/2}\, \{\kappa^{1/2} f_Y(\theta-)^{3/2}\, |g''(0)|\}^{-1}\, M_n + o_p(\sigma^2)\,, \end{aligned} \tag{5.4}$$

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

Note that in (5.4), in contrast to (5.1), we effectively treat $\widehat{\omega}$ as an estimator of θ rather than of ω. Of course, this is possible since we now allow the error variance σ^2 to converge to 0. Taking this view, the principal contribution to the bias of $\widehat{\omega}$, given by the first term on the right-hand side of (5.4), 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 at (5.1).

Observe too that the stochastic error term, given by the second component on the right-hand side of (5.4), is now of size $\sigma^3\,(nh^5)^{-1/2}$ rather than $(nh^5)^{-1/2}$, as at (5.1). 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, since the peak in f_X' becomes more pronounced.

Property (5.4) implies that the asymptotic mean squared error of $\widehat{\omega}$ is of size $\sigma^4 + \sigma^6 (nh^5)^{-1}$, and is minimised by taking h to be of size $(\sigma^2/n)^{1/5}$. This is of course a very different size of bandwidth from that in the problem of estimating ω for a fixed error distribution, since we are treating a different problem — in the present setting we are considering $\widehat{\omega}$ 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 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 asymptotic mean squared error of $\widehat{\omega}$ as an estimator of θ is $O(\sigma^4)$. Note that this is the same order as the mean squared error of the non-stochastic approximation ω to θ; see section 2.2 for that result.

Appendix: Proofs of Main Theoretical Results

A.1 Derivation of (2.6)

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

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

Assuming conditions (C_2), the expansion at (2.3) 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,

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

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

$$\begin{aligned} &B_0 = A_0 f_Z(0) + A_1\, F_Z(0) + A_{20}\sigma + A_{30}\sigma^2 + A_{40}\sigma^3\,, \\ &B_1 = -A_1 f_Z(0) + A_{21} + A_{31}\sigma\,, \quad B_2 = \tfrac{1}{2}\, A_0\, |f_Z''(0)| - A_{22}\sigma^{-1}\,, \\ &B_3 = -\tfrac{1}{6}\, A_1\, f_Z''(0)\,, \quad B_4 = \tfrac{1}{24}\, A_0 f_Z^{(4)}(0)\,. \end{aligned}$$

Note that A_{23}, which would otherwise 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, since their contributions, when multiplied by the appropriate power of $x - \theta$, are of the same order as the $O(\sigma^4)$ remainder at (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 (2.6) and the claimed formulae for C_0, C_2, C_3, 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, and which does not touch the boundary of $\mathcal{R}$ except at the two points on opposite sides of $\mathcal{R}$. Assume that, in a neighbourhood 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 neighbourhood 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 that are bounded uniformly in P. Suppose too that $f_Y(y)$ may 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 $\mathcal{R}$ that lies 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 $z \to 0$, Q being 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, rotate the axes slightly. If f_W denotes the density of a random 2-vector W, define

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

Then

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

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

$$f_X^{(j_1,\ldots,j_r)}(\theta) = \int_{\mathcal{T}(\theta)} \Big\{ A_0(\theta) + \sum_{i=1}^{2} A_1^{(i)}(\theta)\, v^{(i)} + \sum_{i_1,i_2} A_2^{(i_1,i_2)}(\theta)\, v^{(i_1)} v^{(i_2)}$$

$$+\sum_{i_1,i_2,i_3} A_3^{(i_1,i_2,i_3)}(\theta)\, v^{(i_1)} v^{(i_2)} v^{(i_3)} + \ldots \Big\} f_Z^{(j_1,\ldots,j_r)}(v)\, dv. \qquad \text{(A.2)}$$

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 results ($\mathrm{R}_1$). 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),

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

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

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

$$\begin{aligned} \gamma_{ij}(u) &\equiv \frac{\partial}{\partial u^{(j)}}\, \beta_i(u) \\ &= f_X^{(i,j,1)}(\theta+u)\, f_X^{(1)}(\theta+u) + f_X^{(i,j,2)}(\theta+u)\, f_X^{(2)}(\theta+u) \\ &\qquad + f_X^{(i,1)}(\theta+u)\, f_X^{(j,1)}(\theta+u) + f_X^{(i,2)}(\theta+u)\, f_X^{(j,2)}(\theta+u)\,. \end{aligned} \qquad \text{(A.5)}$$

Results (R_1) imply that $\beta_i(0) = O(\sigma^{-6})$; call this result ($\mathrm{R}_2$). Substituting (A.3) and (A.4) into (A.5), as well as using (R_1), we deduce that

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

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 zeroth and first derivatives, respectively, 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 (5.1)

By Taylor expansion, and since $f''_X(\omega) = 0$, then $f''_X(\omega + x) = x\, f'''_X(\omega) + O(x^2)$ 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 $D(x) = (nh^5)^{1/2}\{\hat{f}''_X(x) - E\hat{f}''_X(x)\}$, we find that $D(x)$ is asymptotically Normally distributed with zero mean 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, results of Komlós, Major and Tusnády (1975) may be used to prove that

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

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

Combining these results we deduce first that with probability 1,

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

uniformly in u such that $\omega + hu \in \mathcal{I}$, and secondly 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 $\widehat{\omega}$ of x that maximises $|\hat{f}'_X(x)|$ satisfies $(\widehat{\omega} - \omega)/h \to 0$ with probability 1. Therefore,

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

(It is straightforward to derive (A.7) when the term $(nh^5)^{-1/2} D(\omega)$ on the right-hand side is replaced by $(nh^5)^{-1/2} D(\widehat{\omega})$. However, the weak convergence of ξ to ξ_0, and the fact that $\widehat{\omega} = \omega + o(h)$ with probability 1, imply that $(nh^5)^{-1/2} D(\widehat{\omega}) = (nh^5)^{-1/2} D(\omega) + o_p\{(nh^5)^{-1/2}\}$. And since $h \asymp n^{-1/9}$ then the remainder here equals $o_p(h^2)$.) Solving (A.7) for $\widehat{\omega} - \omega$ we deduce (5.1).

A.4 Derivation of (5.4)

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:

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

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

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

Note that $B_2 \asymp \sigma^{-3}$, $B_3 = O(\sigma^{-3})$ and $B_4 = O(\sigma^{-5})$. Therefore, since $h = o(\sigma)$, then $B_4 h^2 = o(B_2)$. Put $\delta_\epsilon = \sigma^2 + \sigma^3 (nh^5)^{-1/2} n^\epsilon$, and note that (5.3) and the property $\sigma \to 0$ imply that $\delta_\epsilon \to 0$ for some $\epsilon > 0$. It follows from (5.2) that $\sigma^{-7} h^5 = o(\sigma^{-3}\delta_0)$, where δ_0 denotes δ_ϵ with $\epsilon = 0$. More simply, the fact that $h = o(\sigma)$ implies that $\sigma^{-5} h^4 = o(\sigma^{-3}\delta_0)$. Using (5.3) and the property $h = o(\sigma)$ we may deduce that

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

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-hand 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 $\epsilon > 0$,

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

From this property, and borrowing from the argument used in Appendix (iii), we deduce that the solution $u = \hat{u} = \widehat{\omega} - \theta$ of the equation $\hat{f}_X''(\theta + u) = 0$ satisfies

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

Solving for $\hat{u}$ we deduce (5.4). (The asymptotic Normality of $D(\omega)$ requires the assumption that $nh \to \infty$, but does not need $nh^5 \to \infty$.)

References

[1] Aigner, D.J., Lovell, C.A.K. and P. Schmidt (1977). Formulation and estimation of stochastic frontier models. *Journal of Econometrics*, 6, 21-37.

[2] Banker, R.D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, 39, 10, 1265–1273.

[3] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 26, 123–140.

[4] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83, 1184–1186.

[5] Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the inefficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.

[6] Christensen, L. and Greene, R. (1976). Economies of scale in U.S. electric power generation. *Journal of Political Economy*, 84, 653–667.

[7] Deprins, D., Simar, L. and H. Tulkens (1984). Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

[8] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics,* 19, 1257–1272.

[9] Fan, J. (1992). Deconvolution with supersmooth distributions. *Canadian Journal of Statistics*, 20, 155–169.

[10] Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Annals of Statistics*, 21, 600–610.

[11] Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120, 253–281.

[12] 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, 445, 220–228.

[13] Greene, W.H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–163.

[14] Hall, P., Qian, W. and Titterington, D.M. (1992). Ridge finding from noisy data. *Journal of Computational and Graphical Statistics*, 1, 197–211.

[15] 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.

[16] Kneip, A. and L. Simar (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis*, 7, 187–212.

[17] 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.

[18] Korostelev, A., Simar, L. and Tsybakov, A.B. (1995a). Efficient estimation of monotone boundaries. *Annals of Statistics*, 23, 476–489.

[19] Korostelev, A., Simar, L. and Tsybakov, A.B. (1995b). On estimation of monotone and convex boundaries. *Pub. Inst. Stat. Univ. Paris*, XXXIX, 1, 3–18.

[20] Meeusen, W. and J. van den Broek (1977). Efficiency estimation from Cobb-Douglas production function with composed error. *International Economic Review*, 8, 435–444.

[21] Park, B.U., Sickles, R.C. and Simar, L (1998). Stochastic frontiers: a semiparametric approach. *Journal of Econometrics*, 84, 273–301.

[22] Park, B.U. and Simar, L. (1995). Efficient semiparametric estimation in stochastic frontier models. *Journal of the American Statistical Association*, 89, 929–36.

[23] Park, B.U., Simar, L. and Ch. Weiner (1999). FDH efficiency scores from a stochastic point of view. *Econometric Theory*, to appear.

[24] Seiford, L.M. (1996). Data envelopment analysis: the evolution of the state-of-the-art (1978–1995). *Journal of Productivity Analysis*, 7, 99–137.