On Estimation of Monotone and Concave Frontier Functions

by

I. Gijbels[1], E. Mammen[2], B.U. Park[1,3] and L. Simar[1]

August 1997

Abstract

When analyzing the productivity of firms, one may want to compare how the firms transform a set of inputs x (typically labor, energy or capital) into an output y (typically a quantity of goods produced). The economic efficiency of a firm is then defined in terms of its ability of operating close to or on the production frontier which is the boundary of the production set. The frontier function gives the maximal level of output attainable by a firm for a given combination of its inputs. The efficiency of a firm may then be estimated via the distance between the attained production level and the optimal level given by the frontier function. From a statistical point of view, the frontier function may be viewed as the upper boundary of the support of the population of firms density in the input and output space. It is often reasonable to assume that the production frontier is a concave monotone function. Then, a famous estimator, in the univariate input and output case, is the data envelopment analysis (DEA) estimator which is the lowest concave monotone increasing function covering all sample points. This estimator is biased downwards since it never exceeds the true production frontier. In this paper we derive the asymptotic distribution of the DEA estimator, which enables us to assess the asymptotic bias and hence to propose an improved bias corrected estimator. This bias corrected estimator involves consistent estimation of the density function as well as of the second derivative of the production frontier. We also discuss briefly the construction of asymptotic confidence intervals. The finite sample performance of the bias corrected estimator is investigated via a simulation study and the procedure is illustrated for a real data example.

AMS 1991 subject classification. Primary 62G20; secondary 60E20, 60Gxx.

Key words. Asymptotic distribution, bias correction, confidence interval, data envelopment analysis, density support, frontier function.

1 Introduction

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. with a density f in $I\!R^2$. The support of f is assumed to be of the following form:

$$\Psi = \{(x, y) \mid f(x, y) \geq 0\} = \{(x, y) \mid y \leq g(x)\}$$

where g is concave, and monotone increasing. The function g is the upper boundary of the support of the density f. We are interested in estimating g based on the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$.

The problem of estimating a concave and monotone boundary g appears naturally in the context of productivity analysis. When analyzing the productivity of firms, one may compare how the firms transform a set of inputs x (e.g. labor, energy or capital) into an output y (e.g. a quantity of goods produced). In this paper, we restrict to the case where inputs are characterized by a scalar measure x. In this context, Ψ is the attainable production set and $g(\cdot)$ is the production frontier function: it is the geometric locus of the optimal production. For a firm operating with input x_0, $g(x_0)$ is the maximal level of output attainable. The economic efficiency of a firm is then defined in terms of its ability of operating close to this optimal level $g(x_0)$: if its production level is y_0, its efficiency may be calculated via $g(x_0) - y_0$. With this measure, firms can be compared to detect the most efficient (or inefficient) ones. In practice, Ψ and its frontier g are unknown. So our prior interest is the estimation of this frontier from a set of observed firms $(x_i, y_i), i = 1, \ldots, n$. From a statistical point of view, the frontier function g may be viewed as the upper boundary of Ψ, the support of the population of firms density in the input and output space. Generally, the attainable set Ψ is supposed to be convex which implies the concavity of the production frontier g. The monotonicity of g is justified by the free disposability of inputs and outputs. See for example Shephard (1970) and Färe, Grosskopf and Lovell (1985).

In the econometrics literature, a lot of efforts have been devoted to using a parametric model for the frontier function which generally belongs to the class of linear models, such as Cobb-Douglas or translog models (see Berndt and Christensen (1973)). See also Greene (1992), for a nice survey of parametric approaches. For instance, in Section 3.2.2. Figure 5 shows the estimation of two parametric models for a data set of 123 American electric utility companies, described in Section 3.2. Both parametric fits appear as not very appropriate. In situations like this one prefers to avoid the specification of a particular parametric form for

g, and has to find a nonparametric estimator of the frontier function which has to be concave and monotone. This estimator is presented in the next section. Other approaches could be proposed: for instance the upper confidence band of the support of Ψ could be worthwhile to investigate. In this paper we concentrate on a popular nonparametric estimator of a frontier function used in the econometrics literature.

Farrell (1957) introduced the so called *data envelopment analysis* (DEA) estimator $\widehat{\Psi}$ of Ψ, which is the convex hull of $(X_1, Y_1), \ldots, (X_n, Y_n)$. It is the set under the "lowest" concave monotone increasing function covering all the sample points (X_i, Y_i). The DEA estimator of g at x_0 is then defined by the maximum of y such that (x_0, y) belongs to $\widehat{\Psi}$. The DEA estimator has been extensively used since Charnes, Cooper and Rhodes (1978) popularized it by introducing linear programming techniques. Today the estimator is used as a standard for ranking firms according to their relative performance with respect to the obtained frontier. See Charnes, Cooper, Lewin and Seiford (1995) for an exhaustive description of the DEA technique. This technique has been used in many fields of application: analysis of the performance of public services, banks, hospitals, etc. See Seiford (1996) for a recent survey of the DEA estimator, including a lot of references. Until recently however, no attention was devoted to the statistical properties of the DEA estimator. For recent work on statistical aspects see Grosskopf (1996) and Simar (1996). Consistency issues of the estimation procedure were addressed by Banker (1993), Korostelev, Simar and Tsybakov (1995a, b) and Kneip, Park and Simar (1996). A bootstrap technique for simulating sampling variation of the estimator has recently been proposed by Simar and Wilson (1996).

In this paper, we derive the asymptotic distribution of the DEA estimator of g. This motivates us to propose a "blown-up" version. Obviously, the DEA estimator is downward biased since it never exceeds g. The asymptotic distribution quantifies the downward bias, and by correcting this one can improve the DEA estimator. The bias correction involves estimation of the density at the boundary point $(x, g(x))$ and also estimation of the second derivative of g. In this paper, we propose simple and easy to implement estimation procedures for those quantities. In particular, our estimation procedure of g'' preserves the concavity of g, i.e., it guarantees that the estimator of g'' be always negative. We show that the estimator of the bias is consistent. The bias corrected estimator is therefore asymptotically unbiased and it has the same first order asymptotic variance as the DEA estimator. Also, we investigate the effect of the bias correction for finite samples through a simulation study.

This paper is related to Kneip, Park and Simar (1996) where the convergence rate of the DEA estimator is derived. From that paper it can be seen that the rate of convergence is $n^{-2/3}$ if the frontier g is twice continuously differentiable. There are some other recent related works, especially Härdle, Park and Tsybakov (1995), Korostelev, Simar and Tsybakov (1995a, b), and Mammen and Tsybakov (1995). Those papers focus on estimation of the density support rather than the boundary function, and consider only the rates of convergence. (Except for the paper of Korostelev, Simar and Tsybakov (1995a) which provides the exact constant too for the minimax risk). Other related work is concerned with the convex hull of i.i.d. samples in higher dimensions. Groeneboom (1988) and Cabo and Groeneboom (1994) derive the asymptotics for the number of vertices, the boundary length and the area of the convex hull of a uniform sample from the interior of a convex polygon. For other related probabilistic work in this direction see also the references cited in these both papers and Hueter (1992). We are not aware of any previous results on the asymptotic distributions of estimators of the boundary function. Parametric approaches restricting the function g to be linear have been considered by Park and Simar (1994), and Park, Sickles and Simar (1997).

Note that the DEA structure is also well-defined in a more general multi input – multi output setup. Although consistency and rate of convergence have been derived in this general setup in Kneip, Park and Simar (1996), it should be mentioned that the results in the present paper are only valid for a single input – single output variable case. It is not clear yet how to derive an asymptotic distribution result for the general multi input – multi output case, if even possible at all.

The main results of this paper, the asymptotic distribution of the DEA estimator, the proposed bias corrected estimation and the construction of approximate confidence intervals, are given in Section 2. The simulation study investigating the finite sample performance of the DEA estimator is presented in Section 3. We illustrate the bias corrected estimation procedure using the American electric utility data published in Christensen and Greene (1976) and discussed further in Greene (1990), among others. Section 4 contains the proofs of the theoretical results.

2 The Main Results

2.1 Asymptotic distribution of the DEA estimator

Based on the i.i.d. random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$, the DEA estimator of Ψ is formally defined by

$$\widehat{\Psi} = \{(x,y) \mid \quad y \leq \textstyle\sum_{i=1}^{n} \gamma_i Y_i; \; x \geq \sum_{i=1}^{n} \gamma_i X_i \text{ for some } (\gamma_1, \ldots, \gamma_n)$$
$$\text{such that } \textstyle\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, \ldots, n\}.$$

Let x_0 be an interior point in the support of the marginal density of X. Then, the DEA estimator of g at x_0 is defined by

$$\hat{g}(x_0) = \sup\{y \mid (x_0, y) \in \widehat{\Psi}\}.$$

The estimator $\hat{g}$, as a function, is piecewise linear with knots depending on the sample, and is concave and monotone increasing. Note that $\hat{g}(x_0)$ is well-defined whenever there exists an X_i such that $X_i \leq x_0$ and this happens with probability tending to one.

We assume that the frontier function g is twice continuously differentiable at x_0 and $g''(x_0) < 0$. We call this (A1). To obtain a proper asymptotic distribution we ask that near the point $(x_0, g(x_0))$ there be enough data and the density f be smooth. Specifically, we assume that the density function f is bounded away from zero and is continuous on the set

$$\mathcal{U} = \Psi \cap \{(x,y) \mid \quad \| (x,y) - (x_0, g(x_0)) \| < \varepsilon\}$$

for a positive number ε. We call this (A2).

From Kneip, Park and Simar (1996), it can be seen that $\hat{g}(x_0)$ converges to $g(x_0)$ at the rate $n^{-2/3}$ under the assumptions (A1) and (A2). This convergence rate generalizes to the rate $n^{-2/(2+p)}$ in the case of p-variate X, as was also shown in the same paper. Although the rate $n^{-2/3}$ is very good its optimality properties have not been studied so far. We describe below the limiting distribution of $n^{2/3}(\hat{g}(x_0) - g(x_0))$. For this, let b_0, b_1 and b_2 denote respectively $f(x_0, g(x_0))$, $g'(x_0)$ and $-g''(x_0)/2$.

Theorem 1 *Assume (A1) and (A2). Then for all $z < 0$ we have*

$$P\{n^{2/3}(\hat{g}(x_0) - g(x_0)) \leq z\} = \int_0^\infty h(v,z)dv + o(1) \tag{2.1}$$

where $h(v,z) = (1/2)b_0(b_2 v^2 - z) \exp\{-(1/6)b_0 b_2^{-2} v^{-3}(b_2 v^2 - z)^3\}$.

4

By transforming $v = \sqrt{-z/b_2} \cdot u$ for the integral of (2.1) and by a simple algebraic manipulation we get the following corollary to Theorem 1.

Corollary 1 *Under the assumptions of Theorem 1 and for $z < 0$ we have*

$$P\{n^{2/3}(b_0^2/b_2)^{1/3}(\hat{g}(x_0) - g(x_0)) \leq z\} = \int_0^\infty \varphi(u,z)du + o(1) \tag{2.2}$$

where $\varphi(u,z) = (1/2)(-z)^{3/2}(1 + u^2)\exp\{-(1/6)(-z)^{3/2}(u + u^{-1})^3\}$.

The limiting distribution in Corollary 1 is depicted as a thin solid line in Figure 1. The integral has been calculated via numerical integration. We also illustrate the appropriateness of the limiting distribution as an approximation to the finite sample distribution via a simulation study. For a given model (Model 1 of Section 3.1) we simulated 500 samples of size $n = 100$ and calculated for each sample $n^{2/3}(b_0/b_2)^{1/3}(\hat{g}(0.5) - g(0.5))$. Based on the resulting 500 values we obtained a histogram estimator from which an estimator for the cumulative distribution function on the right-hand side of (2.2) was derived. This estimated finite sample distribution is presented as a thick solid line in Figure 1. Note that even for a small sample of size 100 the finite sample distribution and the limiting distibution are pretty close. The closeness between the two improves with sample size (simulations have been done for other sample sizes but are for brevity not presented here).

If we denote the integral of (2.2) by $G(z)$, then the first and second asymptotic moments of $n^{2/3}(b_0^2/b_2)^{1/3}(\hat{g}(x_0) - g(x_0))$ are given respectively by $-\int_0^\infty G(-z)dz$ and $2\int_0^\infty zG(-z)dz$. Hence, the asymptotic bias and variance of $\hat{g}(x_0)$ are given by

$$\text{asymp.bias of } \hat{g}(x_0) = -n^{-2/3}(b_2/b_0^2)^{1/3}c_1, \quad \text{asymp.var. of } \hat{g}(x_0) = n^{-4/3}(b_2/b_0^2)^{2/3}c_2$$

where $c_1 = \int_0^\infty \int_0^\infty \varphi(u,-z)dudz$ and $c_2 = 2\int_0^\infty \int_0^\infty z\varphi(u,-z)dudz - c_1^2$. This gives the asymptotic mean squared error of $\hat{g}(x_0)$ as follows:

$$\text{asymp. MSE of } \hat{g}(x_0) = n^{-4/3}(b_2/b_0^2)^{2/3}(c_1^2 + c_2).$$

Straightforward calculations show that the constants c_1 and c_2 are given by $c_1 = 2 \cdot 6^{2/3}\Gamma(2/3)/9 \approx 0.99359976$ and $c_2 = 4 \cdot 6^{1/3}\Gamma(1/3)/15 - c_1^2 \approx 0.310880687$ respectively.

If we recall that $b_0 = f(x_0, g(x_0))$ and $b_2 = -g''(x_0)/2$, we can see that the asymptotic mean squared error of the DEA estimator is larger, as is expected, when the density at $(x_0, g(x_0))$ is lower or the frontier function g has more curvature at x_0.
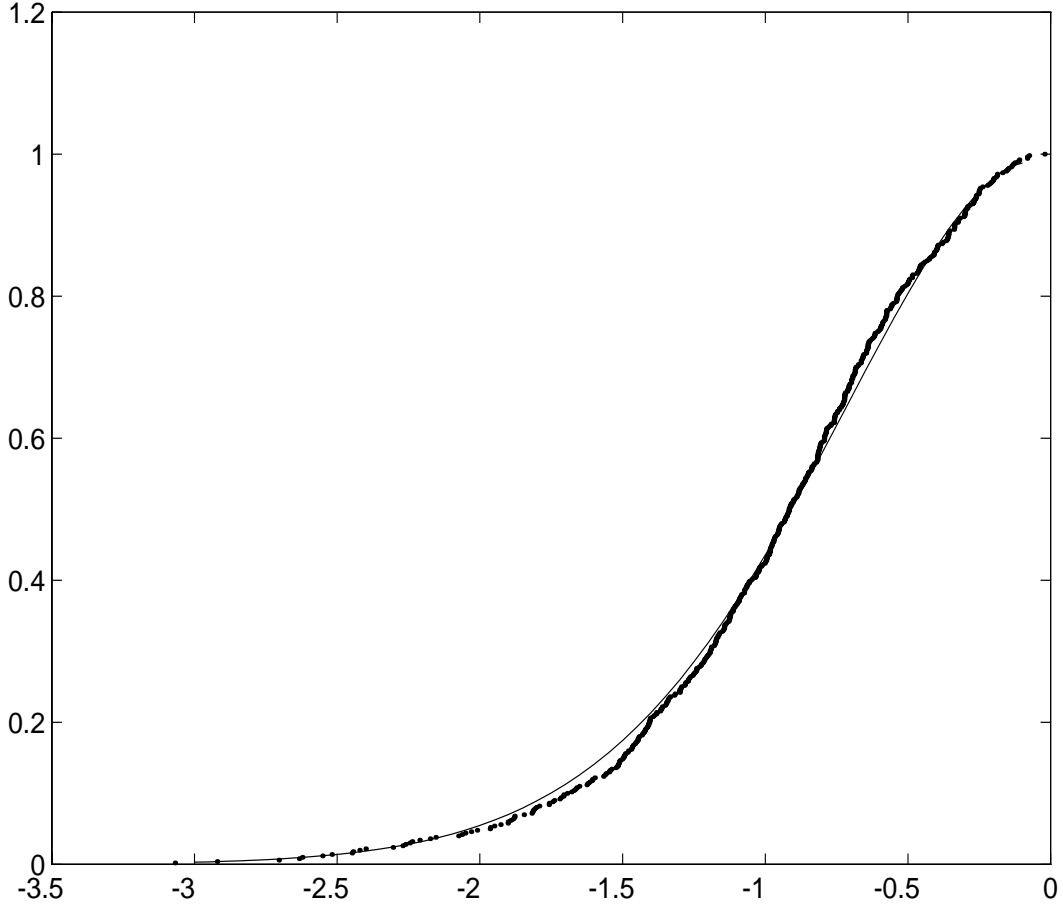
5

Figure 1: *The asymptotic distribution of $n^{2/3}(b_0^2/b_2)^{1/3}(\hat{g}(x_0) - g(x_0))$ (thin solid line) given in (2.2) together with an estimate for the finite sample distribution based on 500 samples of size $n = 100$ from Model 1 of Section 3.1 (thick solid line).*

Remark 1 In some instances, we may be interested in estimating $g^{-1}(y_0) := \inf\{x \mid (x, y_0) \in \Psi\}$ for a given y_0. This is particularly the case when one wants to measure the efficiency of a production unit with output level y_0 in input-oriented way. In this case, for a firm working at the level (x, y_0), the feasible reduction of input for being efficient is given by $x - g^{-1}(y_0)$. If we define $\hat{g}^{-1}(y_0) = \inf\{x \mid (x, y_0) \in \widehat{\Psi}\}$, then by parallel arguments leading to Theorem 1 we get, for $t > 0$, $P\{n^{2/3}(\hat{g}^{-1}(y_0) - g^{-1}(y_0)) > t\} = \int_0^\infty h(v, -b_1 t)dv + o(1)$, where we take $x_0 = g^{-1}(y_0)$ in the definitions of b_0, b_1 and b_2. This is intuitively clear if we observe that $b_1(\hat{g}^{-1}(y_0) - x_0) \simeq y_0 - \hat{g}(x_0)$. For, then $n^{2/3}(\hat{g}^{-1}(y_0) - x_0) > t$ is asymptotically equivalent to $n^{2/3}(\hat{g}(x_0) - y_0) < -b_1 t$.

Remark 2 Estimates for the distribution of $\hat{g}(x_0) - g(x_0)$ can be constructed using bootstrap

6

methods. However, the naive bootstrap method, based on drawing with replacement from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, does not work here. This follows e.g. from the fact that with positive probability the naive bootstrap estimate $\hat{g}^*(x_0)$ coincides with $\hat{g}(x_0)$. This can be seen as follows. Note first that $\hat{g}(x_0)$ is a broken line. With probability $1 - 2(1 - \frac{1}{n})^n + (1 - \frac{2}{n})^n \approx 1 - 2e^{-1} + e^{-2}$ the bootstrap resample contains both break points of $\hat{g}(x_0)$ that are neighbored to x_0. Then $\hat{g}^*(x_0) = \hat{g}(x_0)$. Therefore a more refined bootstrap approach is required. We propose to draw i.i.d. bootstrap resamples from a density $\tilde{f}$ that has support with concave monotone boundary function $\tilde{g}$. If the functions $\tilde{g}$ and $\tilde{f}$ are such that $\tilde{f}(x_0, \tilde{g}(x_0)) \to f(x_0, g(x_0))$ and $\tilde{g}''(x_0) \to g''(x_0)$, in probability, then it can be shown that the bootstrap estimate, i. e. the (conditional) distribution of $\hat{g}^*(x_0) - \tilde{g}(x_0)$, is a consistent estimate of the distribution of $\hat{g}(x_0) - g(x_0)$. Such bootstrap approaches serve asymptotic valid bias corrections of $\hat{g}(x_0)$ and confidence intervals for $g(x_0)$. The next two subsections present another approach for bias correction and construction of confidence intervals. This is based on the asymptotic formula of Theorem 1 using estimates of $g''(x_0)$ and $f(x_0, g(x_0))$. These estimates could also be used for constructions of $\tilde{f}$ and $\tilde{g}$ that let the (refined) bootstrap work.

2.2 A bias corrected estimator

Let $B(x)$ denote $\{-g''(x)/(2f^2(x, g(x)))\}^{1/3}$. Then, from the discussion following Corollary 1, the asymptotic bias of $\hat{g}(x_0)$ is $-n^{-2/3}c_1 B(x_0)$. A modification of the DEA estimator now has the following form:

$$\tilde{g}(x_0) = \hat{g}(x_0) + n^{-2/3}c_1 \widehat{B}(x_0)$$

where $\widehat{B}(x_0)$ denotes a proper estimator of $B(x_0)$. Below we propose a simple estimator of $B(x_0)$.

First, consider the interval $\mathcal{K}(\delta) = (x_0 - \delta/2, x_0 + \delta/2)$ for some positive δ. Let $\mathcal{S}(\delta)(\subset I\!\!R^2)$ denote $\mathcal{K}(\delta) \times I\!\!R$. For the estimation of b_0, the density at $(x_0, g(x_0))$, let $\mathcal{D} = \mathcal{S}(\delta) \cap \hat{\Psi} \cap \{(x, y) \mid y \geq \hat{g}(x_0) - \delta/2\}$. A simple estimator of b_0 is

$$\hat{b}_0 = \{\# \text{ of } (X_i, Y_i) \in \mathcal{D}\}/\{n\lambda(\mathcal{D})\}$$

where λ denotes the Lebesgue measure.

Now, for the estimation of b_2, take $h > 0$. One may think of fitting the second order polynomial passing through $(x_0 - h/2, \hat{g}(x_0 - h/2))$, $(x_0, \hat{g}(x_0))$ and $(x_0 + h/2, \hat{g}(x_0 + h/2))$,

and then take its second derivative to estimate b_2. However, this would yield a zero estimate when the three points sit on the same line, and this may happen frequently in practice. Instead, we propose the following estimation procedure. For a given w, let ℓ_w^L denote the line segment joining $(x_0, \hat{g}(x_0))$ and $(x_0 - h/2, w)$. Likewise, ℓ_w^R denotes the line segment joining $(x_0, \hat{g}(x_0))$ and $(x_0 + h/2, w)$. Define

$$Z^- = \max\{w \mid \text{ there exists } (X_i, Y_i) \in \ell_w^L \text{ for some } 1 \leq i \leq n\}$$

whenever it can be defined. This means that, whenever it is defined, there are one or more sample points sitting on $\ell_{Z^-}^L$ but no points beyond that in the stripe $(x_0 - h/2, x_0) \times I\!R$. Define Z^+ likewise with ℓ_w^R. We take $Z_1^- = \min\{Z^-, \hat{g}(x_0 - h/2)\}$ and $Z_1^+ = \min\{Z^+, \hat{g}(x_0 + h/2)\}$. When the stripe $(x_0 - h/2, x_0) \times I\!R$ is empty, in which case Z^- is not defined, we take $Z_1^- = \hat{g}(x_0 - h/2)$. We do the same thing for Z_1^+. Now, fit the second order polynomial passing through the three points, $(x_0 - h/2, Z_1^-)$, $(x_0, \hat{g}(x_0))$ and $(x_0 + h/2, Z_1^+)$. Call it $\breve{g}$. We take

$$\hat{b}_2 = -\breve{g}''(x_0)/2.$$

Note that the points $(x_0 - h/2, Z_1^-)$ and $(x_0 + h/2, Z_1^+)$ are always below $\hat{g}$. Hence $\breve{g}$ is concave by concavity of the estimator $\hat{g}$. The estimator $\hat{b}_2$ is therefore always positive as it should be, except the case when $(x_0 - h/2, Z_1^-)$, $(x_0 + h/2, Z_1^+)$ and $(x_0, \hat{g}(x_0))$ lie on the same line, but this can happen only with probability tending to zero.

The estimator of $B(x_0)$ is given by $\widehat{B}(x_0) = (\hat{b}_2/\hat{b}_0^2)^{1/3}$.

Theorem 2 *Under the assumptions of Theorem 1, and if $h \sim n^{-1/3+\epsilon_1}$ and $\delta \sim n^{-1/2+\epsilon_2}$ for some $0 < \epsilon_1 < 1/3$ and $0 < \epsilon_2 < 1/2$, then $\hat{b}_0$ and $\hat{b}_2$ are consistent estimators of b_0 and b_2.*

According to the above theorem, the modified estimator is asymptotically unbiased but with the same first order asymptotic variance as the DEA estimator. In Section 3, we investigate the effect of this modification for finite samples.

It is important to note that in our estimation procedure we only require consistency of the estimators of b_0 and b_2, and hence of $B(x_0)$. Further, it should be mentioned that the proposed estimators $\hat{b}_0$ and $\hat{b}_2$ are quite sensitive to the choices of the respective smoothing parameters δ and h. The effect of choosing the smoothing parameters is however less noticeable in the bias corrected estimator. Further research is needed to develop automatic choices of the local smoothing parameters δ and h.

8

Improvement of the estimators for b_0 and b_2, and hence of the bias correction $n^{-2/3}c_1B(x_0)$ would of course result in an improvement of the bias corrected estimator. It is outside the scope of this paper however to elaborate further on this point.

2.3   Construction of confidence intervals

Theorem 1 and its corollary can be used also to construct asymptotic confidence intervals for $g(x_0)$ from the DEA estimator or from the bias corrected estimator. Suppose we want to construct a $100 \times (1 - \alpha)\%$ confidence interval for $g(x_0)$. Then we will search for quantiles $z_{\alpha/2}$ and $z_{1-\alpha/2}$ such that

$$\lim_{n\to\infty} P\{n^{2/3}(b_0^2/b_2)^{1/3}(\hat{g}(x_0) - g(x_0)) < z_{\alpha/2}\} = \frac{\alpha}{2}$$
$$\lim_{n\to\infty} P\{n^{2/3}(b_0^2/b_2)^{1/3}(\hat{g}(x_0) - g(x_0)) \le z_{1-\alpha/2}\} = 1 - \frac{\alpha}{2},$$

or equivalently

$$\lim_{n\to\infty} P\{\hat{g}(x_0) - n^{-2/3}(b_2/b_0^2)^{1/3}z_{1-\alpha/2} \le g(x_0) \le \hat{g}(x_0) - n^{-2/3}(b_2/b_0^2)^{1/3}z_{\alpha/2}\} = 1 - \alpha.$$

Using Theorem 2, i.e. the consistency of the estimated bias correction, this then implies

$$\lim_{n\to\infty} P\{\hat{g}(x_0) - n^{-2/3}\hat{B}(x_0)z_{1-\alpha/2} \le g(x_0) \le \hat{g}(x_0) - n^{-2/3}\hat{B}(x_0)z_{\alpha/2}\} = 1 - \alpha, \qquad (2.3)$$

and also

$$\lim_{n\to\infty} P\{\tilde{g}(x_0) - n^{-2/3}\hat{B}(x_0)(z_{1-\alpha/2} + c_1) \le g(x_0) \le \tilde{g}(x_0) - n^{-2/3}\hat{B}(x_0)(z_{\alpha/2} + c_1)\} = 1 - \alpha.$$
$$(2.4)$$

Hence, for fixed sample size n, an approximate $100 \times (1 - \alpha)\%$ confidence interval for $g(x_0)$ is given by

$$\left[\tilde{g}(x_0) - n^{-2/3}\hat{B}(x_0)(z_{1-\alpha/2} + c_1), \tilde{g}(x_0) - n^{-2/3}\hat{B}(x_0)(z_{\alpha/2} + c_1)\right], \qquad (2.5)$$

with $z_{\alpha/2}$ and $z_{1-\alpha/2}$ determined from

$$G(z_{\alpha/2}) = \frac{\alpha}{2} \qquad \text{and} \qquad G(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2} \qquad (2.6)$$

respectively, using for example numerical integration techniques. Note that the quantiles $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are independent of x_0.

For $\alpha = 0.05$ we found the following approximated values for the 0.025^{th} quantile and the 0.975^{th} quantile respectively

$$z_{0.025} = -2.2962 \qquad \text{with precision} \quad G(-2.2962) = \int_0^\infty \varphi(u, -2.2962)du \approx 0.024996$$

$$z_{0.975} = -0.1727 \qquad \text{with precision} \quad G(-0.1727) = \int_0^\infty \varphi(u, -0.1727)du \approx 0.974992.$$

$$(2.7)$$

The above calculated approximate critical values could also be used in testing procedures concerning the production frontier. Testing problems are so far an unexploited area, and this would be a very interesting direction for future research.

As pointed out in the introduction, in econometric applications one wants to analyze for a particular observed firm whether it is economically efficient or not. For a production situation (x_0, y_0), with x_0 representing the input and y_0 the output, this efficiency can be measured via the distance $g(x_0) - y_0$. Note that a pointwise confidence interval for $g(x_0)$ can serve to build a confidence interval for this efficiency.

It would be of interest to derive a confidence band for the entire production frontier $g(\cdot)$. This could be an issue in testing nonparametric versus parametric models (see for example Section 3.2.2). The construction of a confidence band for $g(\cdot)$ would first of all involve a study of the limiting distribution of $\sup_z |\hat{g}(z) - g(z)|$, suitably normalized. But even with such a result it is not always clear how to construct good confidence bands. See for example Eubank and Speckman (1993). Note further that we can consider working in a regression type setup, with a model of the form $Y = g(X) - U$ where the random variable $U \geq 0$. It is not clear to us how to proceed in case of such a regression model, where we have a constraint involving the response variable Y and the predictor variable X. This situation is quite different from the one considered in for example Naiman (1987) who proposed a method for constructing simultaneous confidence bands in multiple regression in case of constraints on the predictor variables. The construction of a confidence band for the entire function $g(\cdot)$ is an interesting open problem.

3 Finite Sample Performance

In this section we investigate the finite sample performance of the bias corrected estimation procedure proposed in Section 2 via some simulation studies and the analysis of the American

electric utility data given in Christensen and Greene (1976).

3.1 Simulation Study

We investigate the finite sample performance of the DEA estimator and its bias corrected version for two simulation models.

Model 1:

$$X \sim \text{U}[0,1] \qquad Y = g(X)\exp\{-V\}, \qquad g(x) = x^{1/2},$$

$$V \sim \text{Exp}(3), \qquad \text{independent of } X.$$

Note that $E(\exp\{-V\}) = 3/4$. The stochastic scenario adapted here, i.e. an exponential distribution for the logarithm of the inefficiencies and a global average of inefficiency of 0.75, is reasonable with respect to many applications found in the econometrics literature.

Model 2:

Similar to Model 1 but with $V \sim \text{Exp}(1)$, so that $\exp\{-V\}$ is uniformly distributed on $[0,1]$, and $E(\exp\{-V\}) = 1/2$. This situation is clearly less favorable than the one described in Model 1, since it is expected to observe less points near the true frontier function. Hence the estimation task here is a bit more difficult.

For each simulation model we consider samples of size 100, 500 and 1000, and estimate the production function $g(\cdot)$ in three different points in the interior of the support of the marginal density of X, namely $x_0 = 0.25, 0.50$ and 0.75.

Recall that estimation of b_0, respectively b_2, involves a smoothing parameter δ, respectively h. In all simulations we took $\delta = 2n^{-1/2}$ and $h = 2n^{-1/3}$, keeping in mind the conditions on these smoothing parameters imposed in Theorem 2.

The number of simulations in each estimation situation is $N = 500$, and a table summarizes, for each sample size, on a first line the average (centered) values $\bar{\hat{g}}(x_0) - g(x_0)$ and $\bar{\tilde{g}}(x_0) - g(x_0)$ with

$$\bar{\hat{g}}(x_0) = \frac{1}{N}\sum_{s=1}^{N} \hat{g}^{(s)}(x_0) \qquad \text{and} \qquad \bar{\tilde{g}}(x_0) = \frac{1}{N}\sum_{s=1}^{N} \tilde{g}^{(s)}(x_0), \tag{3.1}$$
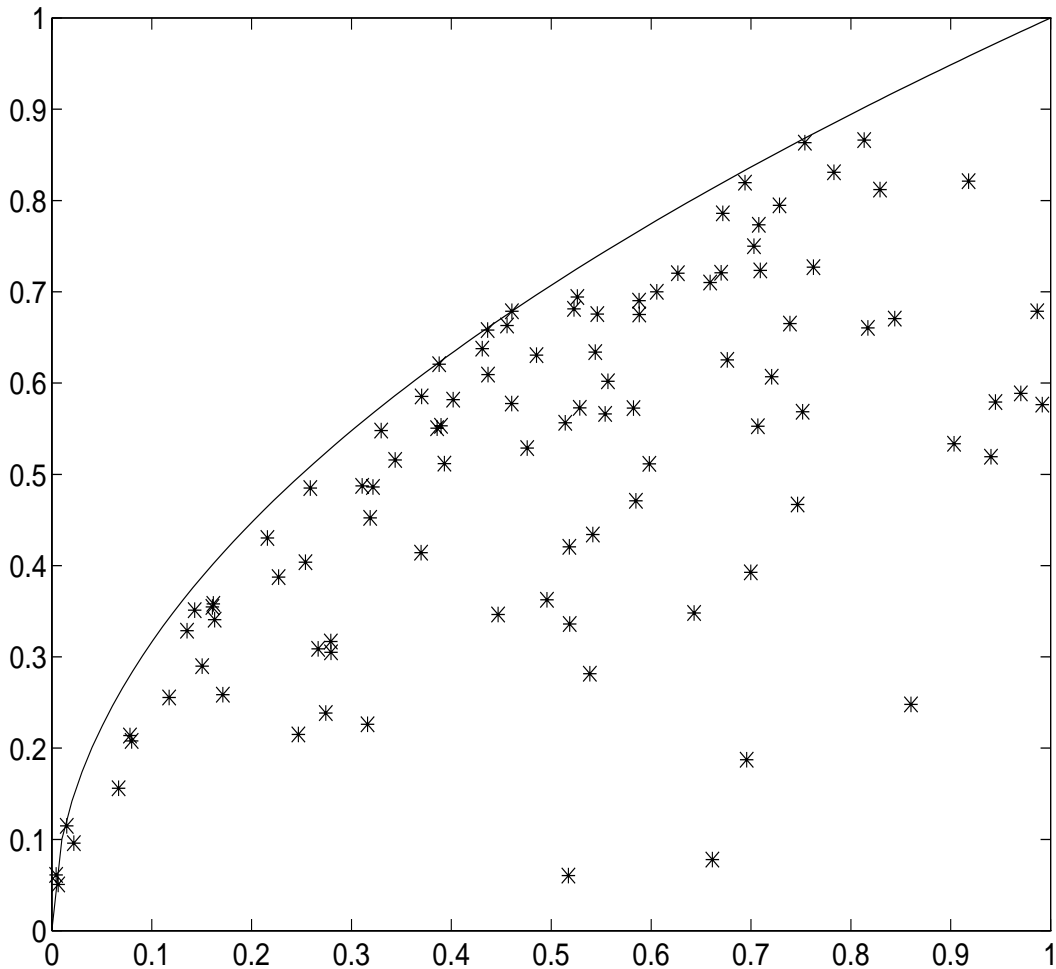
Figure 2: *A typical simulated data set of sample size 100 from Model 1. The solid line represents the true frontier function g.*

where $\hat{g}^{(s)}(x_0)$ and $\tilde{g}^{(s)}(x_0)$ denote the DEA and the bias corrected estimator for a particular simulation. In the tables we also present on the first line (between brackets) the corresponding standard errors of the bias estimator $\bar{\hat{g}}(x_0) - g(x_0)$, among the 500 simulations, i.e.

$$\frac{1}{\sqrt{N}}\sqrt{\frac{1}{N-1}\sum_{s=1}^{N}\left(\hat{g}^{(s)}(x_0) - \bar{\hat{g}}(x_0)\right)^2}, \tag{3.2}$$

and similarly for the bias corrected estimator $\tilde{g}(x_0)$. For each sample size we list, on a second line, the estimated mean squared error (MSE) calculated from the 500 simulations.

Figure 2 depicts a typical simulated data set for a sample of size $n = 100$ from Model 1, together with the true frontier function g. The simulation results for this model are presented in Table 1, showing clearly that the bias corrected estimator $\tilde{g}(x_0)$ performs much

12

better than the DEA estimator. The standard errors between brackets indicate that the bias corrected estimator is more variable, which is intuitively clear since this estimator involves extra estimation tasks, namely estimation of b_0 and b_2.

Table 1: *Simulation results for Model 1. Estimated Bias $(\times 10^{-2})$ with the corresponding standard error $(\times 10^{-4})$ between brackets and, on a second line, the estimated MSE $(\times 10^{-4})$.*

n	$x_0 = 0.25$		$x_0 = 0.50$		$x_0 = 0.75$	
	DEA	Bias Corrected	DEA	Bias Corrected	DEA	Bias Corrected
100	-1.4561 (3.8617)	0.4667 (4.6711)	-1.2319 (2.9936)	0.2726 (3.9720)	-1.2485 (3.2919)	0.5541 (4.6288)
	2.8657	1.3088	1.9656	0.8631	2.1007	1.3783
500	-0.4881 (1.1749)	-0.0520 (1.3872)	-0.4199 (1.0476)	0.0623 (1.3715)	-0.4057 (1.0528)	0.1141 (1.4230)
	0.3073	0.0989	0.2312	0.0979	0.2200	0.1143
1000	-0.2956 (0.7572)	0.0253 (0.8597)	-0.2582 (0.6096)	0.0345 (0.8005)	-0.2485 (0.6120)	0.0598 (0.8204)
	0.1160	0.0376	0.0853	0.0332	0.0805	0.0372

To get an idea about the finite sample distribution of the DEA and the bias corrected estimator we present in Figure 3 kernel density estimates for $\hat{g}(0.5) - g(0.5)$ and $\tilde{g}(0.5) - g(0.5)$, based on the 500 simulations. The kernel density estimates were calculated using the Gaussian kernel and the normal reference bandwidth $h = 1.06\hat{\sigma}n^{-1/5}$, with $\hat{\sigma} = \min\{\text{sample standard deviation}, (\text{sample interquartile range})/1.349\}$. See for example Silverman (1986).

In comparing the distributions of $\hat{g}(0.5) - g(0.5)$ and $\tilde{g}(0.5) - g(0.5)$ we see that the latter one is shifted to the right illustrating clearly the improvement of the bias corrected estimator. Note further that the distribution of the bias corrected estimator is slightly more variable (due to the extra estimation tasks involved). By looking at Figure 3 and focusing on the sample sizes we nicely see the asymptotic effect getting into action: the density estimates become spikier when n increases.

For each simulation we also calculated, from (2.5), the 95% confidence interval for $g(0.5)$, using the values $z_{0.025} = -2.2962$ and $z_{0.975} = -0.1727$ as calculated in Section 2.3. Based on the 500 simulations this led, for sample size $n = 100$, to the empirical coverage probability of 92.4% which should be compared with the theoretical 95%. For other points of the production frontier and other sample sizes we got similar results. Given the fact that the bias correction has been estimated in a quite elementary way (only consistency of the
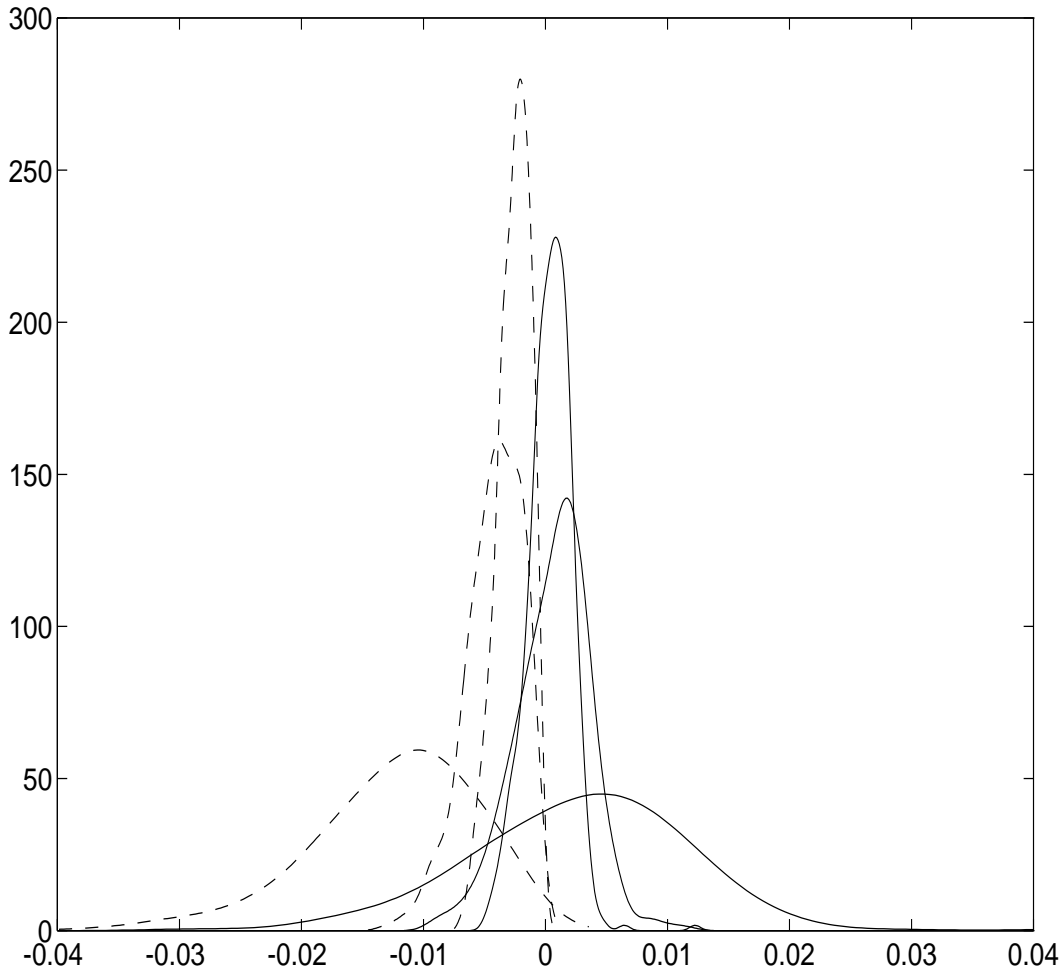
Figure 3: *Kernel density estimates for the DEA estimator and the bias corrected estimator evaluated at the point 0.5 and centered around the true value $g(0.5)$ of the frontier function. The kernel density estimates are based on 500 simulations of samples of sizes $n = 100$, 500 and 1000 (less spiky to spikier curves). Solid lines: kernel density estimates for the bias corrected estimator; Dashed lines: kernel density estimates for the DEA estimator.*

estimator was required) the obtained empirical coverage probabilities are quite satisfactory. It is worthwhile mentioning here that improvement of the estimators for b_0 and b_2 would very likely lead to an improvement of the achieved coverage probability.

As mentioned above, estimation of the frontier function is slightly more difficult in Model 2 since fewer points will be observed along the true frontier. See the simulation results for this model presented in Table 2. Again there is a considerable improvement obtained with the bias corrected estimator. The improvement is most remarkable for the point $x_0 = 0.25$ since the frontier function is more curved there, leading to a bigger bias correction term.

Table 2: *Simulation results for Model 2. Estimated Bias $(\times 10^{-2})$ with the corresponding standard error $(\times 10^{-4})$ between brackets and, on a second line, the estimated MSE $(\times 10^{-4})$.*

n	$x_0 = 0.25$ DEA	$x_0 = 0.25$ Bias Corrected	$x_0 = 0.50$ DEA	$x_0 = 0.50$ Bias Corrected	$x_0 = 0.75$ DEA	$x_0 = 0.75$ Bias Corrected
100	-3.0010 (7.3742)	0.2812 (9.9568)	-2.5319 (7.0128)	0.6570 (9.7465)	-3.0421 (9.5324)	0.7528 (12.463)
	11.725	5.0360	8.8692	5.1814	13.798	8.3330
500	-0.9702 (2.5701)	0.0703 (3.3205)	-0.8628 (2.1784)	0.2294 (3.1977)	-0.8352 (2.1956)	0.5075 (3.6078)
	1.2716	0.5562	0.9816	0.5639	0.9386	0.9083
1000	-0.6259 (1.5708)	0.0314 (2.0209)	-0.5380 (1.3341)	0.1687 (2.0727)	-0.5550 (1.4360)	0.2923 (2.3674)
	0.5152	0.2052	0.3785	0.2433	0.4111	0.3657

Simulations for models with other frontier functions $g(\cdot)$ have also been carried out, leading to similar conclusions. For example, for the above models with $g(x) = 1 - e^{-4x}$ we found an even better improvement of the bias corrected estimator $\tilde{g}(\cdot)$. This is as to be expected since this function g is more concave than the function $g(x) = x^{1/2}$ considered in Tables 1 and 2, and hence the bias is larger here. For $V \sim \text{Exp}(3)$, sample size $n = 100$ and at $x_0 = 0.5$ we find for $\hat{g}(0.5)$ an estimated bias of -2.1147 $\times 10^{-2}$ with corresponding standard error 5.3779 $\times 10^{-4}$ and estimated MSE 5.9182 $\times 10^{-4}$. For the bias corrected estimator $\tilde{g}(0.5)$ the estimated bias was 0.0354 $\times 10^{-2}$ with corresponding standard error 6.3165 $\times 10^{-4}$ and estimated MSE 1.9962 $\times 10^{-4}$. We do not present all simulation results for this example since they only confirm the messages obtained from the above described examples.

3.2 American electric utility data

Data on 123 American electric utility companies were collected. These consists of measurements on several variables. For our illustration we only use the measurements on the variable $Y = \log(Q)$, with Q the production output of a firm, and $X = \log(C)$, with C the total cost involved in the production. For detailed descriptions and analysis of these data see e.g. Christensen and Greene (1976) and Greene (1990).

3.2.1 Nonparametric estimation of the production frontier

Figure 4 shows the 123 observations, together with the DEA estimate $\hat{g}(\cdot)$ and the bias corrected estimate $\tilde{g}(\cdot)$ for $g(\cdot)$. For illustration purpose we also provide the pointwise 95% confidence intervals for $g(\cdot)$, calculated from (2.5), and depicted as a confidence band in Fig-
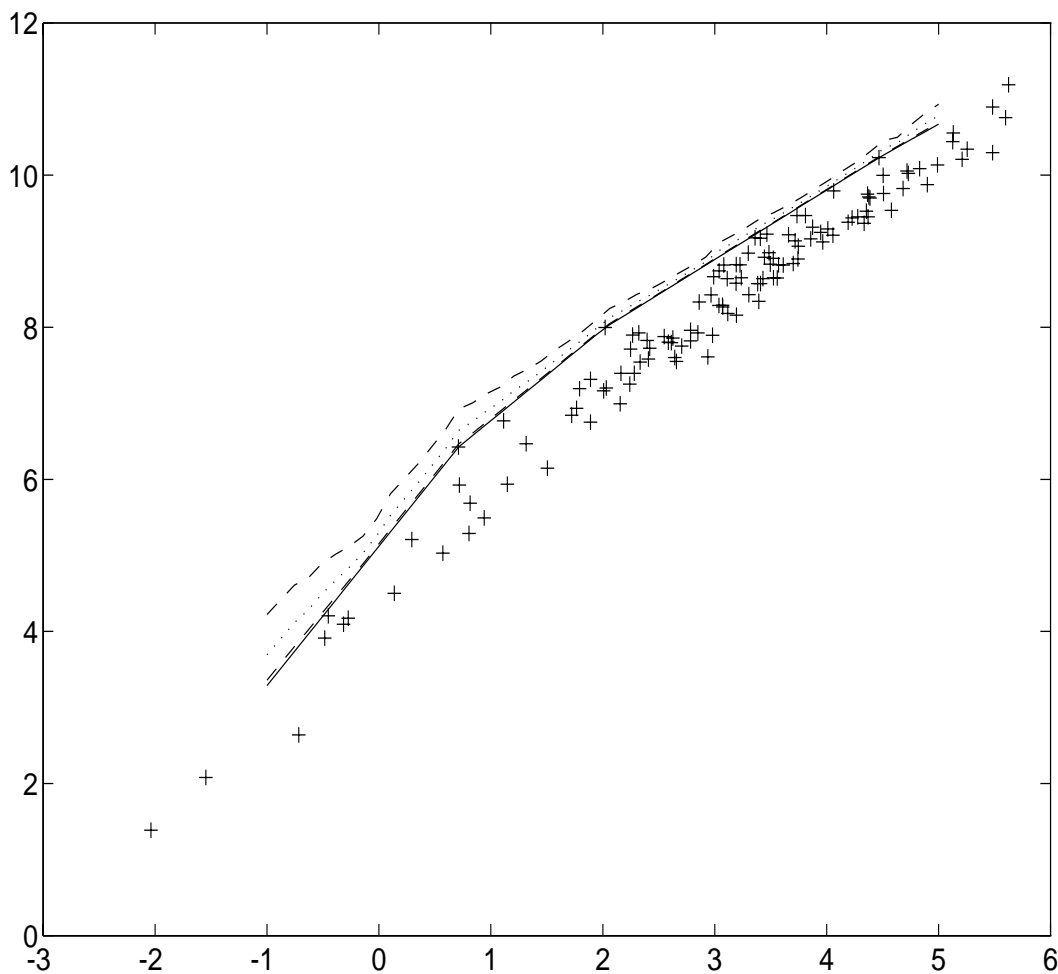
Figure 4: *Scatter plot of the American electric utility data. Solid line: the DEA estimator; Dotted line: the bias corrected estimator. The dashed lines represent the 95% confidence intervals for $g(\cdot)$.*

ure 4. We restricted the region of estimation to $x \in [-1, 5]$ to avoid dealing with boundary effects. For calculation of the bias corrected estimator we subdivided the x-region into two regions: $x \in [-1, 2]$ and $x \in [2, 5]$. The smoothing parameters δ and h were taken equal to 2 and 4 for the region $[-1, 2]$, and equal to 1 and 2 for the estimation region $[2, 5]$. Note that the confidence band is wider at places where the difference between the DEA estimator and the bias corrected estimator is bigger. This is also obvious from the definition of the confidence intervals in (2.5). Indeed, the width of the confidence interval for $g(x_0)$ depends on $\hat{B}(x_0)$, which involves estimation of the density f at the point $(x_0, g(x_0))$ as well as estimation of the curvature of g at the point x_0. In regions where there are more points the density estimate will be larger, and hence it is not surprising that the confidence intervals are

smaller there, as can be seen from Figure 4. On the other hand, sparsity of data will result in smaller estimated values of the joint density, and in wider confidence intervals. Note also that at some places the DEA estimate falls outside the confidence band.

3.2.2 Fitting some parametric models

In this paper we focus on nonparametric estimation of the production frontier g. It might be of interest to compare the performance of our nonparametric estimation procedure with some parametric estimators. The literature on parametric procedures for estimating g is rather limited. An example of a parametric model which has been studied is the linear model

$$Y_i = \alpha + \beta^T X_i - U_i, \qquad i = 1, \cdots, n \tag{3.3}$$

where $U_i \geq 0$ with $E(U_i) \geq 0$, $\beta \in I\!\!R^p$ and X_i is a p-variate random variable. Here the superscript T denotes the transpose of a vector. Note that under this model the production frontier equals $g(x) = \alpha + \beta^T x$, i.e. is a linear function. For estimating the parameters α and β descriptive programming estimators, least-squares methods and maximum likelihood estimators have been proposed. We focus here on the least-squares method since consistency results are available for these estimators. See for example Greene (1980). Least-squares estimation under the parametric model (3.3) is done as follows. One first centers the error term U_i, by subtracting its mean, leading to

$$Y_i = \alpha^* + \beta^T X_i + \varepsilon_i, \qquad i = 1, \cdots, n \tag{3.4}$$

where $\alpha^* = \alpha - E(U_i)$ and $\varepsilon_i = -U_i + E(U_i)$, $i = 1, \cdots, n$. Ordinary least-squares then provides estimators $\hat{\alpha}^*$ and $\hat{\beta}$ of α^* and β respectively. Moreover, the standard least-squares theory tells us that $\hat{\alpha}^*$ and $\hat{\beta}$ are consistent estimators of α^* and β respectively. A particularity here is that one needs some extra efforts in order to get an estimator for α, the intercept parameter. Consider the least-squares residuals $\hat{\varepsilon}_i = Y_i - (\hat{\alpha}^* - \hat{\beta}^T X_i)$. Then, an estimator for α is given by $\hat{\alpha} = \hat{\alpha}^* + \max_{1 \leq i \leq n} \hat{\varepsilon}_i$. The resulting parametric estimator for the production frontier is $\hat{g}(x) = \hat{\alpha} + \hat{\beta}^T x$. So far only the consistency of $\hat{g}$ has been proved (see Greene (1980)). To our knowledge there are no results available on the asymptotic distribution and on the convergence rate of $\hat{g}$. However, for some particular case (panel data and stochastic frontier function) the convergence rate of $\hat{g}$ has been established by Park and Simar (1994).
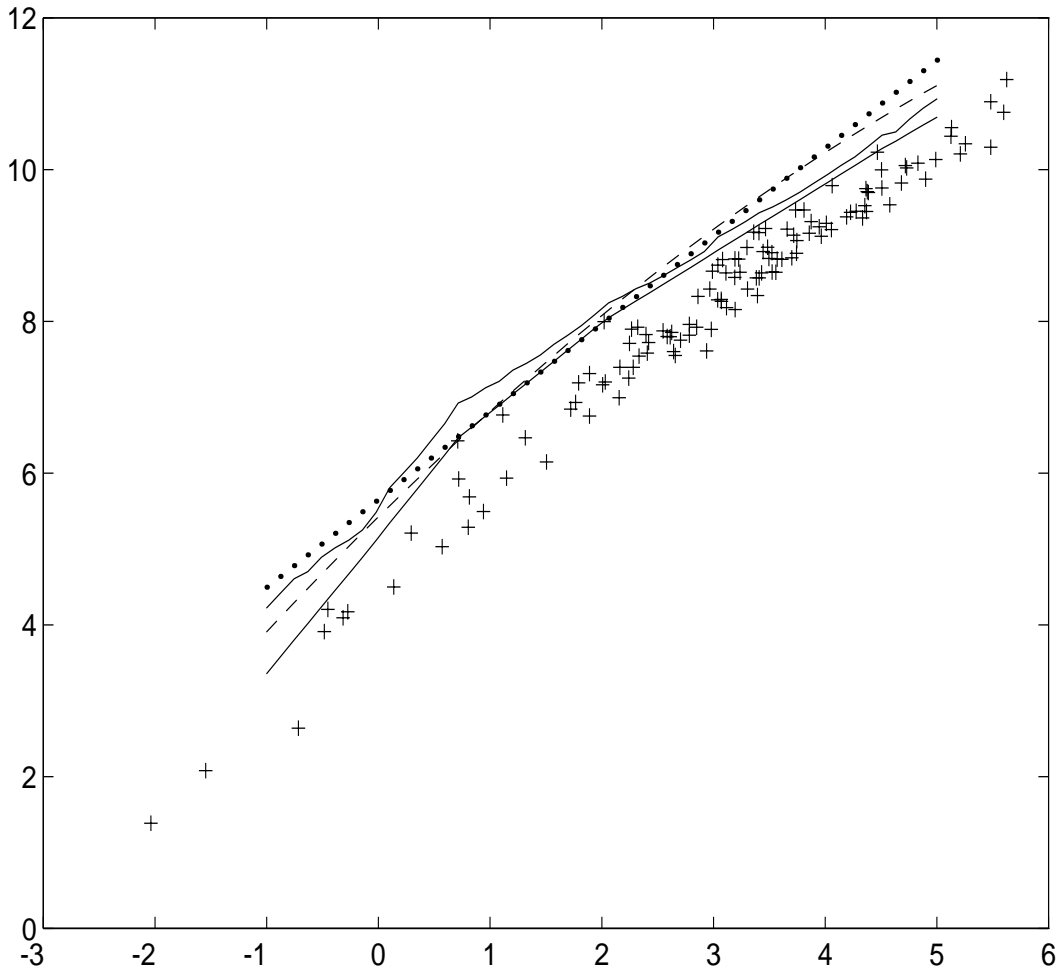
Figure 5: *Parametric fits of Models 1 and 2 (dotted and dashed line respectively) for the American electric utility data, together with the nonparametric confidence intervals for $g(\cdot)$ (the solid lines).*

For comparison purpose we fit two parametric models to the American electric utility data: a Cobb-Douglas model and a translog model (see Berndt and Christense (1973)). Such models are commonly-used in econometrics.

Parametric model 1: $\quad \log(Q) = \alpha + \beta \log(C) - U$

Parametric model 2: $\quad \log(Q) = \alpha + \beta_0 \log(C) + \frac{1}{2}\beta_1 \log^2(C) - U.$

Note that both models are of the form given in (3.3). We estimate the parameters as indicated above using the least-squares method. The resulting estimators for the production frontier are $\hat{\alpha} + \hat{\beta} \log(c)$ for model 1 and $\hat{\alpha} + \hat{\beta}_0 \log(c) + \frac{1}{2}\hat{\beta}_1 \log^2(c)$ for model 2.

The results of these parametric fits are presented in Figure 5 together with the pointwise

18

confidence intervals based on the nonparametric estimate for g. Note that both the linear and quadratic fit exceed the confidence intervals at certain regions, which indicates that these parametric models are not so appropriate.

4 Proofs

4.1 Proof of Theorem 1

Without loss of generality we assume that $x_0 = 0$ and $g(x_0) = 0$. Fix $z < 0$. Put $z' = n^{-2/3}z$. We denote, for $\omega \geq 0$, by $\mathcal{I}_\omega$ the line segment joining $(0, z')$ and $(\omega, g(\omega))$. Specifically,

$$\mathcal{I}_\omega = \{(x, y) \mid x = \alpha\omega \text{ and } y = \alpha g(\omega) + (1 - \alpha)z' \text{ for some } 0 \leq \alpha \leq 1\}.$$

Moreover, we denote by $\mathcal{J}_\omega$ the set surrounded by the function g and the straight line passing through the two points, $(0, z')$ and $(\omega, g(\omega))$. More precisely, we define

$$\mathcal{J}_\omega = \{(x, y) \mid x = \alpha\omega \text{ and } g(x) \geq y \geq \alpha g(\omega) + (1 - \alpha)z' \text{ for some } \alpha \leq 1\}.$$

For $0 \leq \omega_1 \leq \omega_2$, we define $\mathcal{I}_{\omega_1,\omega_2} = \mathcal{J}_{\omega_1}^c \cap \mathcal{J}_{\omega_2}$. It is the set surrounded by the function g, the line passing through $(0, z')$ and $(\omega_1, g(\omega_1))$, and the other line passing through $(0, z')$ and $(\omega_2, g(\omega_2))$. Note that with probability tending to one there exists $\omega \geq 0$ for which $\mathcal{I}_\omega$ includes a sample point. Hence, we put

$$W = \min\{\omega \geq 0 \mid \text{ there exists } (X_i, Y_i) \in \mathcal{I}_\omega \text{ for some } 1 \leq i \leq n\}.$$

The essential idea of the proof is to note that when $\hat{g}(x_0) \leq z'$ there is no second sample point in the random set $\mathcal{J}_W$.

We show that for an infinitesimal change dv

$$P(n^{1/3}W \in (v, v + dv), \hat{g}(x_0) \leq z') = h(v, z)dv(1 + o(1)) \qquad (4.1)$$

uniformly for v in compact subsets of $[0, \infty)$, and

$$n^{1/3}W = O_P(1). \qquad (4.2)$$

Theorem 1 directly follows from (4.1) and (4.2).

We prove (4.1) first. We note that, for an infinitesimal change $d\omega$,

$$P\{W \in (\omega, \omega + d\omega) \text{ and } \hat{g}(x_0) \leq z'\}$$

$$= nP\{(X_1, Y_1) \in \mathcal{I}_{\omega,\omega+d\omega}\}P\{(X_i, Y_i) \notin \mathcal{J}_\omega \text{ for any } 2 \leq i \leq n\} \qquad (4.3)$$

$$= n\int_{\mathcal{I}_{\omega,\omega+d\omega}} f(x, y)dxdy\{1 - \int_{\mathcal{J}_\omega} f(x, y)dxdy\}^{n-1}.$$

For $\omega = vn^{-1/3}$ with fixed $v > 0$, we get

$$R := \int_{\mathcal{I}_{\omega,\omega+d\omega}} f(x,y)dxdy = b_0 \lambda(\mathcal{I}_{\omega,\omega+d\omega})(1+o(1)) \qquad (4.4)$$

where λ denotes the Lebesgue measure. Now, ignoring $(d\omega)^2$, the area of $\lambda(\mathcal{I}_{\omega,\omega+d\omega})$ is the same as that of the triangle connecting $(0, z')$, $(\omega, g(\omega))$ and the intersection of the line segment $\mathcal{I}_{\omega+d\omega}$ with the line parallel to x-axis that passes through $(\omega, g(\omega))$. Specifically, $\lambda(\mathcal{I}_{\omega,\omega+d\omega}) = (1/2)\kappa(g(\omega) - z')$ where

$$\kappa = d\omega - (\omega + d\omega)\{g(\omega + d\omega) - g(\omega)\}/\{g(\omega + d\omega) - z'\}.$$

Since $g(\omega) = b_1\omega - b_2\omega^2 + o(n^{-2/3})$, by ignoring $(d\omega)^2$ again it follows that $\kappa = (b_2\omega^2 - z')d\omega/\{g(\omega) - z'\} + o(n^{-1/3})d\omega$. This gives

$$R = (1/2)b_0(b_2\omega^2 - z')d\omega + o(n^{-2/3})d\omega. \qquad (4.5)$$

We now compute $S := \int_{\mathcal{J}_\omega} f(x,y)dxdy$. As in the case of R we can write $S = b_0\lambda(\mathcal{J}_\omega)(1 + o(1))$. Denote by ω^- the x-coordinate of the point at which the extension of $\mathcal{I}_\omega$ to the left intersects with g. It has the property that

$$\{g(\omega) - z'\}/\omega = -\{z' - g(\omega^-)\}/\omega^-. \qquad (4.6)$$

Let $\gamma = b_1 - b_2\omega - z'/\omega$. Then we can write the common value of (4.6) as $\gamma + o(n^{-1/3})$. Also, we can write

$$\lambda(\mathcal{J}_\omega) = \int_{\omega^-}^{\omega} (b_1 x - b_2 x^2 - \gamma x - z')dx + o(n^{-1}). \qquad (4.7)$$

The integrand function in (4.7) has the values $o(n^{-2/3})$ at ω and ω^-. Let $\vartheta = (\omega - \omega^-)/2$. Without altering the value of the integral in (4.7), we can shift the integrand function so that it is expressed as an integral over $(-\vartheta, \vartheta)$. It is still the integral of a quadratic function of which the coefficient of x^2 is $-b_2$. But, this quadratic function has the values $o(n^{-2/3})$ now at $-\vartheta$ and ϑ. Thus, the quadratic function must have the form $-b_2 x^2 + b_2 \vartheta^2 + o(n^{-2/3})$, and this gives

$$\lambda(\mathcal{J}_\omega) = \int_{-\vartheta}^{\vartheta}(b_2\vartheta^2 - b_2 x^2)dx + o(n^{-1})$$
$$= (1/6)b_2(\omega - \omega^-)^3 + o(n^{-1}). \qquad (4.8)$$

Also, it is easily seen from (4.6) and an expansion for $g(\omega) - g(\omega^-)$ that

$$\omega - \omega^- = \omega - b_2^{-1}\omega^{-1}z' + o(n^{-1/3}). \qquad (4.9)$$

20

Putting (4.8) and (4.9) together, we get

$$(1 - S)^{n-1} = \exp\{-(1/6)nb_0 b_2^{-2}\omega^{-3}(b_2\omega^2 - z')^3\}(1 + o(1)). \tag{4.10}$$

Combining the results (4.3), (4.5) and (4.10) we deduce (4.1).

For the proof of (4.2) note first that

$$P(W > n^{-1/3}C) = P\{ \text{ there exists no } (X_i, Y_i) \text{ in } \mathcal{I}_{0,n^{-1/3}C} \text{ for } 1 \leq i \leq n\}.$$

Claim (4.2) now follows from the fact that for $\varrho > 0$ we have $\lambda(\mathcal{I}_{0,n^{-1/3}C}) \geq \varrho/n$ for sufficiently large C. This implies that for $\epsilon > 0$ we have that $P(W > n^{-1/3}C) \leq \epsilon$ for sufficiently large C. This concludes the proof.

4.2 Proof of Theorem 2

For the proof of the consistency of $\hat{b}_0$, let $\mathcal{D}_1$ denote $\mathcal{S}(\delta) \cap \Psi \cap \{(x,y) \mid y \geq g(x_0) - \delta/2\}$. Then, it can be shown that $\lambda(\mathcal{D} \triangle \mathcal{D}_1) = O_P(n^{-2/3}\delta)$ where $\triangle$ denotes the symmetric difference. From this we can deduce that $\hat{b}_0 = \tilde{b}_0 + o_P(1)$ where $\tilde{b}_0$ is constructed as $\hat{b}_0$ but with $\mathcal{D}_1$ replacing $\mathcal{D}$. The consistency of $\hat{b}_0$ now follows.

For the proof of the consistency of $\hat{b}_2$, we assume $x_0 = 0$ and $g(x_0) = 0$ without loss of generality. We will show first

$$g(-h/2) - Z_1^- = O_P(n^{-2/3}) = g(h/2) - Z_1^+. \tag{4.11}$$

For this, it is important to note that Z_1^- never exceeds $g(-h/2)$ (but Z^- may!). Hence for the first equality of (4.11) we only need to show that $\max\{g(-h/2) - Z^-, 0\} = O_P(n^{-2/3})$ since $g(-h/2) - \hat{g}(-h/2)$ is already within $O_P(n^{-2/3})$. Now, if $\mathcal{L}$ denotes the line segment connecting $(0,0)$ and $(-h/2, g(-h/2) - n^{-1}h^{-1}C)$ for $C > 0$, then

$$P\{g(-h/2) - Z^- > n^{-1}h^{-1}C\} \leq P\{ \text{ there exists no } (X_i, Y_i) \text{ in } (-h/2, 0) \times I\!R \text{ beyond } \mathcal{L}\}$$

The above probability can be bounded by $c' \exp(-c''C)$ for some positive c' and c''. Hence $\max\{g(-h/2) - Z^-, 0\} = O_P(n^{-1}h^{-1})$, and this proves the first equality of (4.11). The second equality of (4.11) can be shown in the same way. Next, we denote by g_1 the second order polynomial passing through $(-h/2, g(-h/2))$, $(0,0)$ and $(h/2, g(h/2))$. Then by (4.11) it can be shown that

$$\breve{g}''(0) = g_1''(0) + O_P(n^{-2/3}h^{-2}). \tag{4.12}$$

Also, it follows that

$$g_1''(0) = g''(0) + o(1). \tag{4.13}$$

The consistency of $\hat{b}_2$ now follows from (4.12) and (4.13). This completes the proof of Theorem 2.

References

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

[2] Berndt, E.R. and Christensen, L.R. (1973). The translog function and the substitution of equipment, structures and labor in U.S. manufacturing 1929–1968. *Journal of Econometrics*, 1, 81-114.

[3] Cabo, A. J. and Groeneboom, P. (1994). Limit theorems for functionals of convex hulls, *Probability Theory and Related Fields*, 100, 31-55.

[4] Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. eds. (1995). *Data Envelopment Analysis : Theory, Methodology and Applications*, Kluwer Academic Publishers, Boston.

[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). Economics of scale in U.S. electric power generation, *Journal of Political Economy*, 84, 653-667.

[7] Eubank, R.L. and Speckman, P.L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, 88, 1287-1301.

[8] Färe, R., Grosskopf, S. and Lovell, C.A.K. (1985). *The Measurement of Efficiency of production*. Kluwer-Nijhoff Publishing, Boston.

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

[10] Greene, W.H. (1980). Maximum likelihood estimation of econometric frontier, *Journal of Econometrics*, 13, 27-56.

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

[12] Greene, W.H. (1992). The econometric approach to efficiency analysis, in H. Fried, K. Lovell and S. Schmidt (eds) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press.

[13] Groeneboom, P. (1988). Limit theorems for convex hulls, *Probability Theory and Related Fields*, 79, 327-368.

[14] Grosskopf, S. (1996). Statistical inference and nonparametric efficiency: a selective survey, manuscript to appear in *Journal of Productivity Analysis*.

[15] Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries, *Journal of Multivariate Analysis*, 55, 205-218.

[16] Hueter, I. (1992). *The convex hull of n random points and its vertex process*, Doctoral dissertation, Universität Bern, Switzerland.

[17] Kneip, A., Park, B.U. and Simar, L. (1996). A note on the convergence of nonparametric DEA efficiency measures, Discussion Paper #9603, Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

[18] Korostelev, A., Simar, L. and Tsybakov, A.B. (1995a). Efficient estimation of monotone boundaries, *The 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. Paries*, XXXIX, 1, 3-18.

[20] Mammen, E. and Tsybakov, A.B. (1995). Asymptotical minimax recovery of sets with smooth boundaries, *The Annals of Statistics*, 23, 502-524.

[21] Naiman, D.Q. (1987). Simultaneous confidence bounds in multiple regression using predictor variable constraints. *Journal of the American Statistical Association*, 82, 214-219.

[22] Park, B.U., Sickles, R.C. and Simar, L. (1997). Stochastic panel frontiers: A semiparametric approach, manuscript to appear in *The Journal of Econometrics*.

[23] Park, B.U. and Simar, L. (1994). Efficient semiparametric estimation in a stochastic frontier model, *Journal of the American Statistical Association*, 89, 929-936.

[24] Seiford, L.M. (1996), Data Envelopment Analysis : The Evolution of the State-of-the-Art (1978–1995), manuscript to appear in *Journal of Productivity Analysis* .

[25] Shephard, R.W. (1970). *Theory of Cost and Production Function*. Princeton University Press, Princeton, New-Jersey.

[26] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[27] Simar, L. (1996). Aspects of statistical analysis in DEA-type frontier models, *Journal of Productivity Analysis*, 7, 177-185.

[28] Simar, L. and Wilson, P. (1996). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models, manuscript to appear in *Management Science*.