Nonparametric Frontier Estimation: a Robust Approach*

Catherine Cazals
IDEI/GREMAQ
Université de Toulouse

Jean-Pierre Florens
IDEI/GREMAQ
Université de Toulouse

Léopold Simar†
Institut de Statistique
Université Catholique de Louvain

January 3, 2001

Abstract

A large amount of literature has been developed on how to specify and to estimate production frontiers or cost functions. Two different approaches have been mainly developed: the deterministic frontier model which relies on the assumption that all the observations are on a unique side of the frontier, and the stochastic frontier models where observational errors or random noise allows some observations to be outside of the frontier. In a deterministic frontier framework, nonparametric methods are based on envelopment techniques known as FDH (Free Disposal Hull) and DEA (Data Envelopment Analysis). Today, statistical inference based on DEA/FDH type of estimators is available but, by construction, they are very sensitive to extreme values and to outliers.

In this paper, we build an original nonparametric estimator of the "efficient frontier" which is more robust to extreme values, noise or outliers than the standard DEA/FDH nonparametric estimators. It is based on a concept of expected minimum input function (or expected maximal output function). We show how these functions are related to the efficient frontier itself. The resulting estimator is also related to the FDH estimator but our estimator will not envelop all the data. The asymptotic theory is also provided. Our approach includes the multiple inputs and multiple outputs cases.

As an illustration, the methodology is applied to estimate the expected minimum cost function for french post offices.

Keywords: production function, cost function, expected maximum production function, expected minimum cost function, frontier, nonparametric estimation.

JEL Classification: C13, C14, D20.

---

*This paper is a revised version of Cazals and Florens (1997).

†Visiting IDEI, Toulouse with the support of the Ministère de l'Education nationale , de la recherche et de la technologie, France. Research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) from the Belgian Government is also acknowledged.

1 Introduction

Since the basic work of Koopmans (1951) and Debreu (1951) on activity analysis, a large amount of literature has been developed on how to specify and to estimate production frontiers or cost functions and on how to measure technical efficiency of production units. See Shephard (1970) for a modern economic formulation of the problem. Consider a production technology where the activity of production units is characterized by a set of inputs $x \in I\!R_+^p$ used to produce a set of outputs $y \in I\!R_+^q$.

The production set is defined as the set

$$\Psi = \{(x, y) \in I\!R_+^{p+q} \mid x \text{ can produce } y\}. \tag{1.1}$$

This set can be described mathematically by its sections. For example, in the input space we have the input requirement sets defined for all $y \in \Psi$ as $C(y) = \{x \in I\!R_+^p \mid (x, y) \in \Psi\}$. The radial (input-oriented) efficiency boundary ("efficient frontier") is then defined by:

$$\partial C(y) = \{x \mid x \in C(y), \theta x \notin C(y) \;\forall\; 0 < \theta < 1\}. \tag{1.2}$$

The Farrell input measure of efficiency of a production unit working at level (x_0, y_0) is then defined as:

$$\theta(x_0, y_0) = \inf\{\theta \mid \theta x_0 \in C(y_0)\} = \inf\{\theta \mid (\theta x_0, y_0) \in \Psi\}. \tag{1.3}$$

Note that $\partial C(y) = \{x \mid \theta(x, y) = 1\}$.

The same could be done in the output space where the output requirement set is defined for all $x \in \Psi$ as $P(x) = \{y \in I\!R_+^q \mid (x, y) \in \Psi\}$. Its radial efficient boundary is then:

$$\partial P(x) = \{y \mid y \in P(x), \lambda y \notin P(x) \;\forall\; \lambda > 1\}. \tag{1.4}$$

Then the Farrell output measure of efficiency for a production unit working at level (x_0, y_0) is defined as

$$\lambda(x_0, y_0) = \sup\{\lambda \mid \lambda y_0 \in P(x_0)\} = \sup\{\lambda \mid (x_0, \lambda y_0) \in \Psi\}. \tag{1.5}$$

Here, $\partial P(x) = \{y \mid \lambda(x, y) = 1\}$.

Note that the frontier of Ψ is unique and $\partial C(y)$ and $\partial P(x)$ are two different ways of describing it. Different assumptions can be assumed on Ψ like free disposability[1] or convexity,... (see, e.g., Shephard, 1970 for details).

[1]Free disposability in inputs and outputs of Ψ means that if $(x, y) \in \Psi$ then $(x', y') \in \Psi$ for any $x' \geq x$ and $y' \leq y$, where inequality between vectors has to be understood element by element.

The econometric problem is thus how to estimate Ψ from a random sample of production units $\{(X_i, Y_i) \mid i = 1, \ldots, n\}$. Two different approaches have been mainly developed: the deterministic frontier model which relies on the assumption that the DGP (Data Generating Process) is such that $\text{Prob}((X_i, Y_i) \in \Psi) = 1$, and the stochastic frontier models where observational errors or random noise allows some observations to be outside of Ψ.

In a deterministic frontier framework, nonparametric methods have known an increasing success since the pioneering work of Farrell (1957). The methods are based on envelopment techniques known as FDH (Free Disposal Hull) estimators initiated by Deprins, Simar and Tulkens (1984) which rely only on free disposability assumptions for Ψ and DEA (Data Envelopment Analysis) estimators, which assumes free disposability and convexity of Ψ, initiated by Farrell and operationalized as linear programming estimators by Charnes, Cooper and Rhodes (1978). They can be defined as follows: the FDH estimator of Ψ is the free disposal hull of the set of observations:

$$\widehat{\Psi}_{FDH} = \left\{(x, y) \in \mathbb{R}_+^{p+q} | y \leq Y_i,\ x \geq X_i, \quad i = 1, \ldots, n\right\}.$$

Then the convex hull of $\widehat{\Psi}_{FDH}$ provides the DEA estimator of Ψ:

$$\begin{aligned} \widehat{\Psi}_{DEA} \quad = \quad & \{(x, y) \in \mathbb{R}_+^{p+q} | y \leq \sum_{i=1}^{n} \gamma_i Y_i\,;\ x \geq \sum_{i=1}^{n} \gamma_i X_i \text{ for } (\gamma_1, \ldots, \gamma_n) \\ & \text{such that } \sum_{i=1}^{n} \gamma_i = 1\,;\ \gamma_i \geq 0, i = 1, \ldots, n\}. \end{aligned}$$

It is the smallest free disposal convex set covering all the data.

Nonparametric envelopment estimators have been extensively used for estimating efficiency of firms (see Seiford, 1996, for a nice survey). Today, statistical inference based on DEA/FDH type of estimators is available either by using asymptotic results (Kneip, Park and Simar, 1998 and Park, Simar and Weiner, 2000) or by using the bootstrap, see Simar and Wilson (2000) for a recent survey of the available results. Nonparametric deterministic frontier models are very appealing because they rely on very few assumptions but it is known that by construction, they are very sensitive to extreme values and to outliers.

In a stochastic frontier framework, where noise is allowed, only parametric restrictions on the shape of the frontier and on the DGP allow identification of noise from efficiency and estimation of the frontier. Most of the available techniques are based on the maximum likelihood principle, in the spirit of the work of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broek (1977).

Our work is a part of the literature of nonparametric frontier estimation in that we build an original nonparametric estimator of the "efficient frontier" which is more robust to extreme values, noise or outliers than the standard DEA/FDH nonparametric estimators.

To the best of our knowledge, very few methods are proposed in the literature to address this important issue. Wilson (1993) and (1995) has proposed descriptive methods to detect outliers in this framework. This paper proposes robust estimators of frontiers with their full statistical treatment.

For sake of simplicity, we will make our presentation in the input-oriented framework, where we have one input x ($p = 1$) and q outputs y. The input efficient frontier can then be interpreted as a "minimum input function" or as a "minimum cost function". We will indicate, in Appendix A, how to make the changes for the output-oriented case where we have one output y ($q = 1$) and p inputs x: in this case, the efficient frontier is a "maximum production function". A shown later, a complete multivariate extension (multi-input and multi-output cases) is also possible.

We will define a concept of expected minimum input function (or expected minimum cost function) and present the methodology for a nonparametric estimation of it. The output oriented case provides the concept of expected maximum production function. We show how these functions are related to the efficient frontier defined above under the hypothesis of free disposability. The resulting estimator is also related to the FDH estimator but our estimator will not envelop all the data. The method can also be adapted if, in addition, the assumption of convexity of Ψ is made and convex estimators of Ψ are wanted. In this case, our estimator will be related to the DEA estimator.

The paper is organized as follows. Section 2 presents the basic concepts of expected minimum input function and its relation to frontiers. Section 3 proposes a nonparametric estimator and analyses its asymptotic properties and its relations with other nonparametric estimators. In Section 4, a numerical illustration is proposed: the methodology is applied to estimate the expected minimum cost function for french post offices. We use a data set on labor (as input) and mail volumes (as output) on around 10.000 post offices. Section 5 suggests some two useful extensions: how to introduce exogenous explanatory variables in the model and how to generalize the approach to the multivariate case (multi-input and multi-output). Section 6 concludes.

2 The Expected Minimum Input Function

Let us consider a random vector (X, Y) on $I\!R_+ \times I\!R_+^q$. The first element X is the input and the q-dimensional vector Y represents the outputs. The joint distribution on (X, Y) defines the production process. Such probability measure is usually decomposed into a marginal distribution on Y and a conditional distribution on X given $Y = y$.

In this paper, we will rather concentrate on an other characterization of the joint probability measure on (X, Y): the joint survivor function. Let us denote by $Y \geq y$ the property that $Y_j \geq y_j$ for $j = 1, \ldots, q$. We will consider the conditional probability measure on X given $Y \geq y$. If the joint probability measure is characterized by the joint survivor function:

$$S(x, y) = \text{Prob}(X \geq x, Y \geq y), \tag{2.1}$$

the conditional distribution on X given $Y \geq y$ may be described by its survivor function:

$$\begin{aligned} S_c(x \mid y) &= \text{Prob}(X \geq x \mid Y \geq y) && (2.2) \\ &= \frac{S(x, y)}{S_Y(y)}, && (2.3) \end{aligned}$$

where $S_Y(y) = S(0, y) = \text{Prob}(Y \geq y)$ denotes the marginal survivor function of Y. It is supposed here and below that $S_Y(y) \neq 0$ or that y is an interior point of the support of the marginal distribution of Y.

The lower boundary of the support of the conditional distribution whose survivor function is $S_c(x \mid y)$ is given by the function:

$$\varphi(y) = \inf\{x \mid S_c(x \mid y) < 1\}. \tag{2.4}$$

This function is monotone non decreasing in y as shown by the following theorem:

Theorem 2.1 *The frontier function $\varphi(y)$ is monotone non decreasing in y:*

$$\textit{For all } y' \geq y \textit{ we have } \varphi(y') \geq \varphi(y). \tag{2.5}$$

Proof:

$$\begin{aligned} \varphi(y) &= \inf\{x \mid S_c(x \mid y) < 1\} \\ &= \sup\{x \mid S_c(x \mid y) = 1\}. \end{aligned}$$

Denote A_y the set $\{x \mid S_c(x \mid y) = 1\}$. We have

$$\begin{aligned} A_y &= \{x \mid \text{Prob}(X \geq x \mid Y \geq y) = 1\} \\ &= \{x \mid \text{Prob}(X < x \mid Y \geq y) = 0\} \\ &= \{x \mid \text{Prob}(X < x, Y \geq y) = 0\}. \end{aligned}$$

If $y' \geq y$, $A_y \subseteq A_{y'}$, then $\sup\{x \mid x \in A_{y'}\} \geq \sup\{x \mid x \in A_y\}$ which completes the proof.■

Note that this minimum input (or cost) frontier function $\varphi(y)$ is always defined and monotone non decreasing: no particular assumption on Ψ are needed. By construction and from the preceding theorem, $\varphi(y)$ is the largest monotone function which is smaller than

$\partial C(y)$ the input-efficient frontier of Ψ (remember that here $p = 1$). It is clear that if the attainable set Ψ is free disposal, $\partial C(y)$ is monotone and coincides with $\varphi(y)$.

Consider now an integer $m \geq 1$ and let $(X^1, \ldots, X^m)$ be m independent identically distributed random variables generated by the distribution of X given $Y \geq y$.

Definition 2.1 *The expected minimum input function of order m denoted by $\varphi_m(y)$ is the real function defined on $I\!R^q_+$ as*

$$\varphi_m(y) = E(\min(X^1, \ldots, X^m) \mid Y \geq y), \tag{2.6}$$

where we assume the existence of this expectation.

The function $\varphi_m(y)$ can be computed as follows.

Theorem 2.2 *If $\varphi_m(y)$ exists, it is given by*

$$\varphi_m(y) = \int_0^\infty [S_c(u \mid y)]^m \, du. \tag{2.7}$$

Proof: This result is an elementary consequence of the rules of integration by parts, since if $X_{\min} = \min(X^1, \ldots, X^m)$, we have:

$$\text{Prob}(X_{\min} \geq u \mid Y \geq y) = [S_c(u \mid y)]^m, \tag{2.8}$$

from which the result derives.■

From its definition, it is clear that for any y fixed, $\varphi_m(y)$ is a decreasing function of m. The limiting case when $m \to \infty$ is of particular interest. It achieves the efficient frontier:

Theorem 2.3 *For any fixed value of y we have*

$$\lim_{m \to \infty} \varphi_m(y) = \varphi(y). \tag{2.9}$$

Proof:

$$\begin{aligned} \varphi_m(y) &= \int_0^\infty [S_c(u \mid y)]^m \, du \\ &= \int_0^{\varphi(y)} [S_c(u \mid y)]^m \, du + \int_{\varphi(y)}^\infty [S_c(u \mid y)]^m \, du. \end{aligned}$$

For all $u \leq \varphi(y),\ S_c(u \mid y) = 1$. So that

$$\varphi_m(y) = \varphi(y) + \int_{\varphi(y)}^\infty [S_c(u \mid y)]^m \, du. \tag{2.10}$$

For $u > \varphi(y)$, $S_c(u \mid y) < 1$, so $[S_c(u \mid y)]^m$ tends to zero when $m \to \infty$. Using the Lebesgue convergence theorem, the integral on the right hand side of (2.10) converges to zero when $m \to \infty$ giving the result. ■

The function $\varphi_m(y)$ converges to a monotone non decreasing function $\varphi(y)$ as $m \to \infty$, but it is not monotone non decreasing itself unless we add the following assumption.

Assumption 2.1 *The conditional distribution of X given $Y \geq y$ has the following property*

$$\text{For all } y' \geq y,\ S_c(x \mid y') \geq S_c(x \mid y). \tag{2.11}$$

This assumption is not needed for all the results of this paper except the next theorem, but it appears to be quite reasonable: it says that the chance of spending more than an input (or cost) x does not decrease if a firm produces more. So, if we want a joint survival function $S(x, y)$ to represent a production process, Assumption 2.1 is quite natural. It also implies the monotonicity of $\varphi_m(y)$:

Theorem 2.4 *Under Assumption 2.1, $\varphi_m(y)$ is monotone non decreasing in y.*

Proof: This is immediate by the expression of $\varphi_m(y)$ given in Theorem 2.2 and from properties of integrals. ■

From an economic point of view, the expected minimum input (cost) function of order m, $\varphi_m(y)$ has its own interest: it is not the efficient frontier of the production set but it might be useful in term of practical efficiency analysis. Suppose a production unit produces a quantity of output y_0 using the quantity x_0 of input, $\varphi_m(y_0)$ gives the expected minimum cost among a fixed number of m potential firms producing more than y_0. For this particular unit, working at level (x_0, y_0), it is certainly worth to know this value because it gives a clear indication of how efficient he is compared with these m potential units. This is achieved by comparing its own level x_0 with the "benchmarked" value $\varphi_m(y_0)$. At this stage, m could be any number from 1 to ∞. In practice, a few values of m could be used to guide the manager of the production unit to evaluate its own performance.

But the most attractive property of this function is that it can be easily non parametrically estimated without the drawbacks of the methods trying to estimate the frontier itself: it will be less sensitive to noise, extreme values or outliers. This is developed in the next section.

In Appendix A, we indicate how the concepts and properties can be adapted to the output oriented case with one output y and p inputs x.

3 Nonparametric Estimation

Consider, for simplicity an i.i.d. sample (x_i, y_i), $i = 1\dots, n$ of the random vector (X, Y). The empirical survivor function is defined by:

$$\widehat{S}_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!I(x_i \geq x, y_i \geq y). \tag{3.1}$$

The empirical version of $S_c(x \mid y)$ is then given by:

$$\widehat{S}_{c,n}(x \mid y) = \frac{\widehat{S}_n(x, y)}{\widehat{S}_{Y,n}(y)}, \tag{3.2}$$

where $\widehat{S}_{Y,n}(y) = (1/n)\sum_{i=1}^{n} 1\!\!I(y_i \geq y)$. Note that this estimator does not require any smoothing procedure as required when the conditional distribution of X given $Y = y$ is required.

All the properties of $\varphi(y)$ and $\varphi_m(y)$ of the preceding section remain valid when the function $S_c(x \mid y)$ is replaced by $\widehat{S}_{c,n}(x \mid y)$. In particular we have the lower boundary of the support of the empirical conditional distribution characterizing the estimated efficient frontier of the production set. It is given by the function:

$$\hat{\varphi}_n(y) = \inf\{x \mid \widehat{S}_{c,n}(x \mid y) < 1\}. \tag{3.3}$$

This function is monotone non decreasing in y. It is the input oriented efficient frontier obtained by the FDH estimator. The estimator of the expected minimum input function of order m is defined by:

$$\hat{\varphi}_{m,n}(y) = \widehat{E}(\min(X^1, \dots, X^m) \mid Y \geq y), \tag{3.4}$$

where $X^1, \dots, X^m$ are m i.i.d. random variables generated by the empirical distribution of X given $Y \geq y$ whose survivor function is $\widehat{S}_{c,n}(x \mid y)$. It is computed through

$$\hat{\varphi}_{m,n}(y) = \int_0^\infty [\widehat{S}_{c,n}(u \mid y)]^m \, du. \tag{3.5}$$

The relation (2.10) between $\varphi_m(y)$ and $\varphi(y)$ remains valid with their empirical versions:

$$\hat{\varphi}_{m,n}(y) = \hat{\varphi}_n(y) + \int_{\hat{\varphi}_n(y)}^\infty [\widehat{S}_{c,n}(u \mid y)]^m \, du, \tag{3.6}$$

from which we obtain again that for all y,

$$\lim_{m\to\infty} \hat{\varphi}_{m,n}(y) = \hat{\varphi}_n(y). \tag{3.7}$$

So, our estimator of the expected minimum input function of order m converges to the FDH input efficient frontier when m increases. In particular, in finite samples, it should be noticed that, even when $m = n$, our estimator is different from the FDH estimator:

$$\hat{\varphi}_{n,n}(y) \neq \hat{\varphi}_n(y).$$

Even for large finite values of m, the estimator $\hat{\varphi}_{m,n}(y)$ is less sensitive to extremes values than the FDH estimator $\hat{\varphi}_n(y)$ which by construction, envelopes all the observations. The asymptotic theory is discussed below. Note also that $\hat{\varphi}_{m,n}(y)$ is not necessarily monotone non decreasing. Indeed, even if Assumption 2.1 is assumed for the true conditional survivor function, it could not be verified by its empirical counterpart. Of course we know that for large sample size n, it will mostly be the case.

The integral in (3.5) defining our estimator may be easily computed in practice. Let $n(y)$ be the number of observations of y_i greater or equal to y, i.e. $n(y) = \sum_{i=1}^{n} 1\!\!I(y_i \geq y)$, and, for $j = 1, \ldots, n(y)$, denote by $x^y_{(j)}$ the j-th order statistic[2] of the observations x_i such that $y_i \geq y$: $x^y_{(1)} < x^y_{(2)} < \ldots < x^y_{(n(y))}$.

The function $\widehat{S}_{c,n}(u \mid y)$ is a step function such that:

$$\begin{aligned} \widehat{S}_{c,n}(u \mid y) &= 1 \quad \text{if } u \leq x^y_{(1)} \\ &= \frac{n(y) - j}{n(y)} \quad \text{if } x^y_{(j)} < u \leq x^y_{(j+1)} \\ &= 0 \quad \text{if } u > x^y_{(n(y))}. \end{aligned}$$

Then we have:

$$\hat{\varphi}_{m,n}(y) = x^y_{(1)} + \sum_{j=1}^{n(y)-1} \left[\frac{n(y) - j}{n(y)} \right]^m (x^y_{(j+1)} - x^y_{(j)}). \tag{3.8}$$

The following theorem summarizes the asymptotic properties of our estimator for any fixed value of m.

Theorem 3.1 *Assume that Ψ, the support of the random vector (X, Y) is compact, then for any interior point y in the support of the Y distribution, and for any $m \geq 1$:*

(i) $\hat{\varphi}_{m,n}(y) \to \varphi_m(y)$ *a.s. as* $n \to \infty$ *;*

(ii) $\mathcal{L}\left(\sqrt{n}(\hat{\varphi}_{m,n}(y) - \varphi_m(y))\right) \to N(0, \sigma^2(y))$ *as* $n \to \infty$*, where*

$$\sigma^2(y) = E\left[\frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1} 1\!\!I(X \geq u, Y \geq y)\, du - \frac{m\varphi_m(y)}{S_Y(y)} 1\!\!I(Y \geq y) \right]^2.$$

[2]We suppose here that there are no ties among the $x^y_{(j)}$: this allow the simple formulation of $\widehat{S}_{c,n}(u \mid y)$. In case of ties, all the theory remains valid but the explicit expression of $\hat{\varphi}_{m,n}(y)$ in (3.8) is no more valid. The general expression (3.5) has to be used.

Proof:

(i) This result follows from a strong law of large numbers which implies the almost sure convergence of $\widehat{S}_{c,n}(u \mid y)$ to $S_c(u \mid y)$ and from the Lebesgue dominated convergence theorem which warrants the convergence of the integrals defining $\hat{\varphi}_{m,n}(y)$ and $\varphi_m(y)$.

(ii) The argument will follow the standard Delta method (see Serfling, 1980, Chapter 6, Theorem A). Let us denote by

$$T(S) = \int_0^\infty [S_c(u \mid y)]^m \, du.$$

$T(S)$ is an operator which associates a real value to any survivor function S. This operator is differentiable at the Frechet sense w.r.t. the sup norm, that is:

$$T(R) - T(S) = DT_S(R - S) + \varepsilon(R - S)||R - S||, \tag{3.9}$$

for any two survivor functions S and R, where the sup norm is used:

$$||V(x, y)|| = \sup_{(x,y)\in\Psi} |V(x, y)|,$$

and where $\varepsilon(V) \to 0$ when $||V|| \to 0$. The Frechet derivative is obtained by standard calculus, noting that $S_c(u \mid y) = S(u, y)/S_Y(y)$:

$$DT_S(V) = \frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1} V(u, y) \, du - m\frac{\varphi_m(y)}{S_Y(y)} V(0, y). \tag{3.10}$$

Now, applying (3.9) and noting that $DT_S(\widehat{S}_n - S) = DT_S(\widehat{S}_n)$ we have:

$$\begin{aligned}
&\sqrt{n}[T(\widehat{S}) - T(S)] = \\
&\frac{\sqrt{n}}{n} \sum_{i=1}^{n} \left[\frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1} \, 1\!\!I(x_i \geq u, y_i \geq y) \, du - \frac{m\varphi_m(y)}{S_Y(y)} \, 1\!\!I(y_i \geq y) \right] \\
&+\varepsilon(\widehat{S}_n - S)(\sqrt{n}||\widehat{S}_n - S||).
\end{aligned} \tag{3.11}$$

As $\sqrt{n}||\widehat{S}_n - S|| = O_p(1)$ by the Dvoretzky, Kiefer and Wolfowitz inequality (see Serfling, 1980, Chapter 2, Theorem A) and $\varepsilon(\widehat{S}_n - S) \to 0$ in probability (because $\widehat{S}_n$ is uniformly convergent), the second term of the r.h.s. of (3.11) converges to 0. The theorem comes then from a central limit theorem applied to the first term of the r.h.s. of (3.11). In particular, it is easy to verify that the term between brackets has zero mean. Indeed:

$$\begin{aligned}
&E \left[\frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1} \, 1\!\!I(X \geq u, Y \geq y) \, du - \frac{m\varphi_m(y)}{S_Y(y)} \, 1\!\!I(Y \geq y) \right] \\
&= n \left[\frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1} \, S(u, y) \, du - \frac{m\varphi_m(y)}{S_Y(y)} \, S_Y(y) \right] = 0.\blacksquare
\end{aligned}$$

Note the $\sqrt{n}$ rate of convergence of $\hat{\varphi}_{m,n}(y)$ to $\varphi_m(y)$ which is rather unusual in nonparametric statistics. The expression of the variance can be used to derive asymptotic confidence intervals for $\varphi_m(y)$: by plugging estimators for the unknown quantities and taking the empirical mean for the expectation provides $\hat{\sigma}^2(y)$, a consistent estimator of the variance. Observe that for a given sample size, $\hat{\sigma}^2(y)$ will increase with y.

Note that these convergence results may be improved by a functional limit theorem which is given in Appendix B. With this functional theorem the asymptotic can be derived for transformations of φ_m.

The result can also be extended to the analysis of the asymptotic properties of a vector $(\hat{\varphi}_{m,n}(y^1), \ldots, \hat{\varphi}_{m,n}(y^r))$. We still have the asymptotic r-variate normal distribution with asymptotic covariances given by

$$\Sigma_{k,\ell} = \mathrm{Cov}(\hat{\varphi}_{m,n}(y^k), \hat{\varphi}_{m,n}(y^\ell)) = E\left[\Gamma(y^k, X, Y)\,\Gamma(y^\ell, X, Y)\right], \tag{3.12}$$

where

$$\Gamma(y, X, Y) = \frac{m}{S_Y(y)^m} \int_0^\infty S(u, y)^{m-1}\, 1\!\!1(X \geq u, Y \geq y)\, du - \frac{m\varphi_m(y)}{S_Y(y)}\, 1\!\!1(Y \geq y).$$

The issue of how to choose m in practice has been discussed above in Section 2. We know that the estimator $\hat{\varphi}_{m,n}(y)$ converges to the FDH estimator $\hat{\varphi}_n(y)$ defined in (3.3) as $m \to \infty$. But we know also from Park, Simar and Weiner (2000), that under regularity conditions, as $n \to \infty$, the FDH estimator $\hat{\varphi}_n(y)$ converges to the true unknown frontier $\varphi(y)$ defined in (2.4).

The value of m can thus be viewed as a "trimming" or "smoothing" parameter and the natural question then arises: how to define m as a function of n such that $\hat{\varphi}_{m,n}(y)$ converges to $\varphi(y)$, as $n \to \infty$. This could also give some insights on how to choose m in practice in order to obtain a consistent estimator of the true frontier, if wanted. The result follows from the next theorem.

Theorem 3.2 *Assume that the joint probability measure of (X, Y) on the compact support Ψ provides a strictly positive density on the frontier $\varphi(y)$ and that the function $\varphi(y)$ is continuously differentiable in y. Then, for any y interior to the support of Y we have:*

$$\mathcal{L}\left(n^{1/(1+q)}(\hat{\varphi}_{m_y(n),n}(y) - \varphi(y))\right) \to \mathrm{Weibull}(\mu_y^{1+q}, 1+q) \quad \text{as } n \to \infty, \tag{3.13}$$

where $m_y(n) = O(\beta\, n \log(n)\, S_Y(y))$, with $\beta > 1/(1+q)$ and μ_y is a constant.

Proof: From Park, Simar and Weiner (2000) we know that

$$\mathcal{L}\left(n^{1/(1+q)}(\hat{\varphi}_n(y) - \varphi(y))\right) \to \text{Weibull}(\mu_y^{1+q}, 1+q) \quad \text{as } n \to \infty,$$

where the parameter μ_y of the Weibull depends on local properties of the DGP near the frontier point $(\varphi(y), y)$. Now using (3.6) we obtain:

$$n^{1/(1+q)}(\hat{\varphi}_{m,n}(y) - \varphi(y)) = n^{1/(1+q)}(\hat{\varphi}_n(y) - \varphi(y)) + n^{1/(1+q)} \int_{\hat{\varphi}_n(y)}^{\infty} [\widehat{S}_{c,n}(u \mid y)]^m \, du.$$

So the question is to find the value of $m = m_y(n)$ such that the last term of the preceding expression is $o_p(1)$ as $n \to \infty$. Using a mean value theorem, we can write the integral as $(x^y_{(n(y))} - x^y_{(1)})[\widehat{S}_{c,n}(\tilde{u} \mid y)]^m$ where $\tilde{u} \in]x^y_{(1)}, x^y_{(n(y))}[$. Since the support of (X, Y) is compact, the range of X is bounded, in addition, for $u > \hat{\varphi}_n(y)$, $\widehat{S}_{c,n}(\tilde{u} \mid y)$ is bounded by $(n(y) - 1)/n(y)$. So to achieve our goal, it is sufficient that $m_y(n)$ is such that

$$[(n(y) - 1)/n(y)]^{m_y(n)} = O_p(n^{-\beta}),$$

where $\beta > 1/(1+q)$. Now, since $\log(1 - 1/n(y)) \asymp -1/n(y)$ and $n(y) \asymp n S_Y(y)$ this is equivalent to

$$m_y(n) = O(\beta \, n \, \log(n) \, S_Y(y)),$$

with $\beta > 1/(1+q)$. ■

In practice, if a consistent estimator of the frontier itself is wanted, we might plug the value of $\widehat{S}_Y(y)$ in the formula to get an idea of the order of $m_y(n)$, but of course the result is only an asymptotic one.

Note that we loose the $\sqrt{n}$-consistency because here we use $\hat{\varphi}_{m,n}(y)$ to estimate the frontier $\varphi(y)$ itself and not $\varphi_m(y)$, the "order-m" frontier which can be viewed as a "trimmed" frontier.

Remark 3.1 Convexifying the estimator: Robust DEA estimator
The above results do not rely on convexity assumption regarding the attainable set Ψ. If Ψ is convex, our estimator remains consistent with the same asymptotics but a convex estimates of Ψ is obtained by convexifying the set above the obtained frontier $\hat{\varphi}_{m,n}(y)$. This could be achieved by running a input-oriented DEA linear program on the set of points $\{(\hat{\varphi}_{m,n}(y_i), y_i),\ i = 1, \ldots, n\}$. Again the obtained estimator will converge to the DEA frontier estimator if $m \to \infty$, but, for finite m, it will not envelop all the data points.

4 Empirical Illustration

To illustrate our methodology, we analyze the production of the postal services in France. More precisely, we focus on the cost of the delivery activity.

We use a cross-section data set of around 10.000 post offices, observed in 1994. We have information about labor used and mail volumes for the delivery activity of each post office. So, in this example, we have one input X and one output Y.

For each post office i, the variable X_i is the labor cost, which represents more than 80% of the total cost of the delivery activity. It is measured by the quantity of labor. The output Y_i is defined as volume of the delivered mail (in number of objects). The data and the results are shown in Figures 1–3.

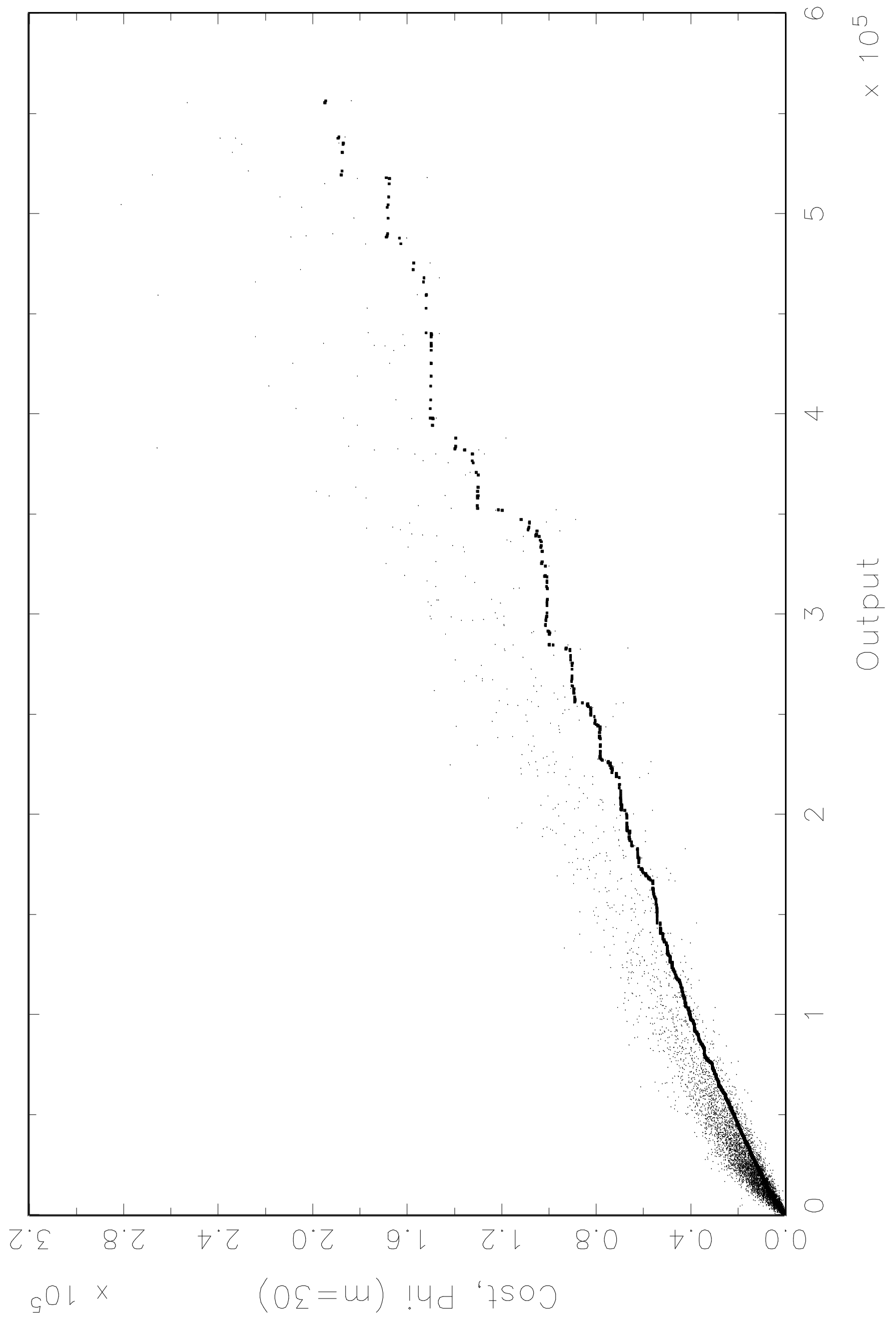
Figure 1 plots the observed data, the cost x_i (vertical axis) against the output y_i (horizontal axis), along with the nonparametric estimation of the expected minimum cost function of order m: $\hat{\varphi}_{m,n}(y)$. Here the value of m was fixed to 30. Figure 2 zooms in Figure 1 for the 3.000 first observations with the smallest output levels. It appears more clearly, in the zoom, that the estimates is typically monotone and that many points stay outside (below) the frontier of order $m = 30$.

We have also estimated the variance function $\sigma^2(y)$ given in Theorem 3.1. This allows to determine for some given points y confidence intervals. Figure 3 plots the pointwise confidence intervals for a selected grid of points y. As expected, the lengths of the confidence intervals increases when y is larger.

The FDH cost efficient frontier would envelop all the data points and is, of course, below our estimate. Our obtained expected minimum cost of order m can thus be viewed as a mark of "good practice" for producing units when studying their performance. However, this benchmark is less "severe" than the FDH frontier because it is less sensitive to extreme points.

With $m = 30$, 73% of the observations where used to determine the expected minimum cost estimate of order m and so 27% of points were left out. Figure 4 indicates how the percentage of points below the expected minimum cost estimate of order m decreases with m. We notice that, in our example, this percentage is very stable from $m = 50$ where roughly 10% of the observations are left out. These observations should be analyzed in details because they are really extreme and could be outlying or perturbed by noise.

Fig. 1 : Expected minimum cost function



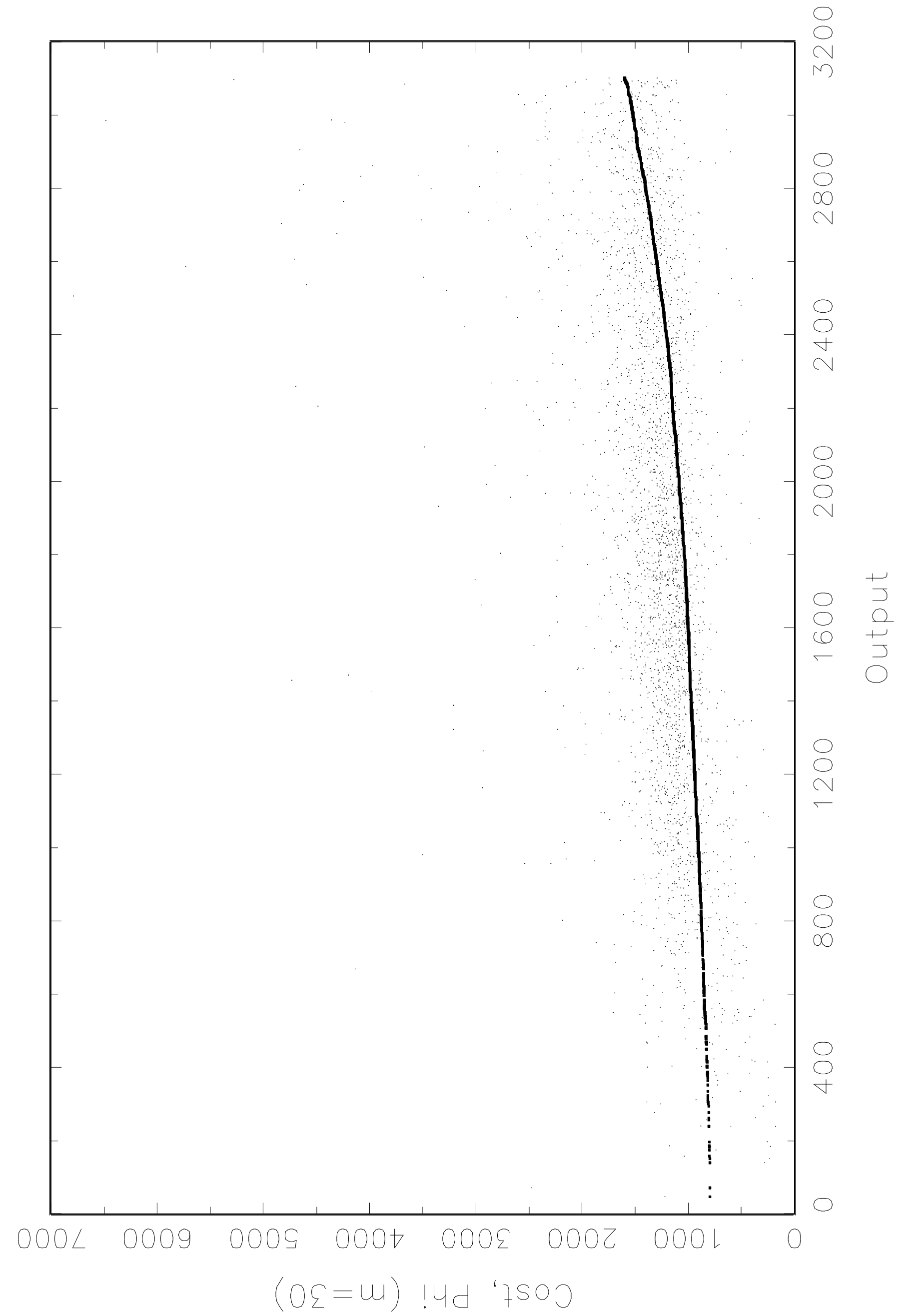


Fig. 2 : 3000 first observations

Fig. 3 : Confidence intervals

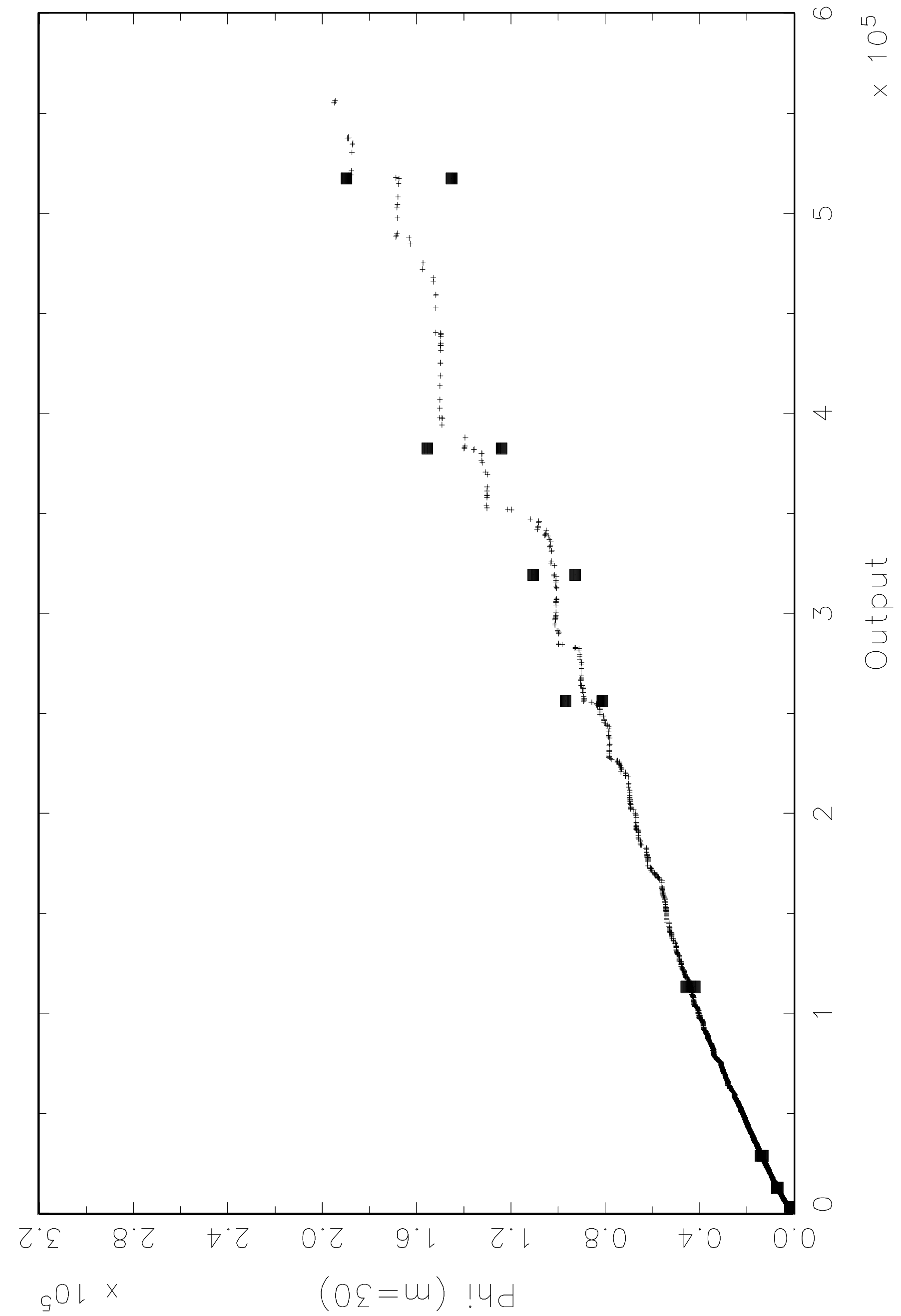
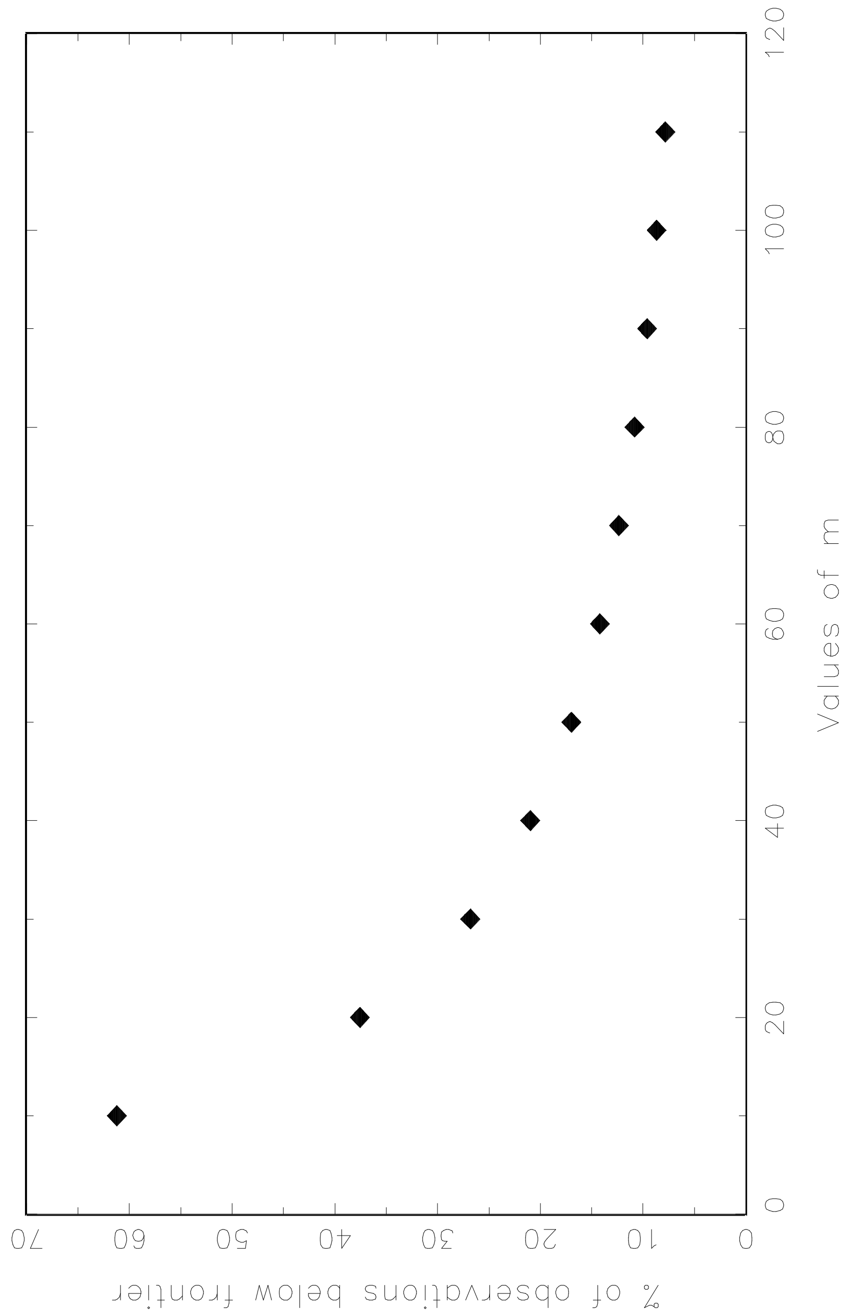


Fig. 4 : Evolution of the percentage of observations below frontier with m



5 Extensions

5.1 Introducing environmental factors

The analysis of the preceding section can easily be extended to the case where additional information is provided by other variables $Z \in \mathbb{R}^k$, exogenous to the production process itself, but which may explain part of it. It could be environmental variables, not under the control of the manager. For instance, in the case of our empirical illustration above, we could consider a variable Z representing the geographical density of the distribution area of a post office (the number of delivery points by unit of length of the distribution route). This variable, at least in the short term, is not under the control of the managers of the post office but might influence the cost of the post office.

One way for introducing in the model this additional information is to condition the production process to a given value of Z. Then, in the empirical example of the post offices, we could study the expected minimum cost function for a post office delivering a mail volume greater than y, with a geographical density equal to z.

Here, the joint survival function is written as $S(x, y, z) = \text{Prob}(X \geq x, Y \geq y, Z \geq z)$. Then the methodology can be adapted by replacing in Section 2, $S_c(x \mid y)$ by $S_c(x \mid y; z)$ where

$$S_c(x \mid y; z) = \text{Prob}(X \geq x \mid Y \geq y, Z = z) = \frac{\partial_z S(x, y, z)}{\partial_z S(0, y, z)}, \tag{5.1}$$

∂_z denoting the operator of derivative of order k with respect to all the components of z:

$$\partial_z = \frac{\partial^k}{\partial z_1 \ldots \partial z_k}.$$

The estimation of this conditional survivor function will require a smoothing technique in z. For example we will estimate $S(x, y \mid Z = z)$ by:

$$\widehat{S}_n(x, y \mid Z = z) = \frac{\sum_{i=1}^{n} \mathbb{1}(x_i \geq x, y_i \geq y) K\left(\frac{z - z_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{z - z_i}{h_n}\right)}, \tag{5.2}$$

where $K(\cdot)$ is a kernel and h_n is the smoothing bandwidth which has the appropriate size for getting the asymptotic theory. It can be shown that the resulting estimator of the mth order expected minimum cost will achieve the rate of convergence $\sqrt{nh_n^k}$ where k is the dimension of Z. The theory can be developed in a similar way as in the preceding sections. We sketch below the main points of the argument.

We want to estimate the conditional mth order expected minimum cost function defined as

$$\varphi_m(y,z) = \int_0^\infty [S_c(u \mid y; z)]^m du, \tag{5.3}$$

where

$$S_c(u \mid y; z) = \text{Prob}(X \geq u \mid Y \geq y, Z = z).$$

The estimator is given by:

$$\hat{\varphi}_{m,n}(y,z) = \int_0^\infty \left[\frac{\sum_{i=1}^n 1\!\!I(x_i \geq u, y_i \geq y) K\left(\frac{z - z_i}{h_n}\right)}{\sum_{i=1}^n 1\!\!I(y_i \geq y) K\left(\frac{z - z_i}{h_n}\right)} \right]^m du. \tag{5.4}$$

The asymptotic distribution of the estimator, with y and z fixed may be derived from the general results of Aït-Sahalia (1995). Under regularity conditions on the distribution of Z (continuous differentiable density function), if the kernel $K(\cdot)$ is symmetric and positive and with the usual conditions on the bandwidth implying asymptotic normality and unbiasedness of the estimator of the density of Z (i.e. $nh_n^k \to \infty$ and $nh_n^{k+4} \to 0$ as $n \to \infty$), we obtain that

$$\mathcal{L}\Big(\sqrt{nh_n^k}(\hat{\varphi}_{m,n}(y,z) - \varphi_m(y,z)\Big) \to N(0, \sigma^2(y,z) f_Z(z) \int K^2), \tag{5.5}$$

with

$$\sigma^2(y,z) = \text{Var}(A(X,Y;y,z) \mid Z = z),$$

where

$$\begin{aligned} A(X,Y;y,z) = \quad & \frac{m}{[\partial_z(S(0,y,z))]^m} \int_0^\infty [\partial_z(S(u,y,z))]^{m-1} 1\!\!I(X \geq u, Y \geq y) du \\ & - \frac{m\varphi_m(y,z)}{\partial_z(S(0,y,z))} 1\!\!I(Y \geq y). \end{aligned}$$

Note again that $E(A(X,Y;y,z) \mid Z = z) = 0$.

The main argument of the proof is similar to the developments in Theorem 3.1. In the linearization of $\sqrt{nh_n^k}(\hat{\varphi}_{m,n}(y,z) - \varphi_m(y,z))$, the analog of the leading term of (3.11) becomes

$$\begin{aligned} & \frac{\sqrt{nh_n^k}}{n} \sum_{i=1}^n \left[\frac{m}{[\partial_z(S(0,y,z))]^m} \int_0^\infty [\partial_z(S(u,y,z))]^{m-1} 1\!\!I(x_i \geq u, y_i \geq y) \frac{1}{h_n^k} K\left(\frac{z - z_i}{h_n}\right) du \right. \\ & \quad \left. - \frac{m\varphi_m(y,z)}{\partial_z(S(0,y,z))} 1\!\!I(y_i \geq y) \frac{1}{h_n^k} K\left(\frac{z - z_i}{h_n}\right) \right] \\ = \quad & \frac{\sqrt{nh_n^k}}{n} \sum_{i=1}^n A(x_i, y_i; y, z) \frac{1}{h_n^k} K\left(\frac{z - z_i}{h_n}\right), \end{aligned}$$

then, under the appropriate regularity conditions on the kernel, the standard theory applies.

5.2 Multivariate extensions

We consider here the setting of Section 1 where we have p inputs and q outputs. We continue the presentation in the input-oriented case . The modifications for the formulation in the output-oriented case are straightforward (see also Appendix A). The production process is here described by the joint probability measure of (X, Y) on $I\!R_+^p \times I\!R_+^q$. The support of (X, Y) is the attainable set Ψ.

There are several ways for describing the frontier $\partial C(y)$ when $p \geq 1$. For any $(x, y) \in \Psi$, we can indeed define, as in (1.3), the input efficiency measure of the point (x, y):

$$\theta(x, y) = \inf\{\theta \mid \theta x \in C(y)\} = \inf\{\theta \mid (\theta x, y) \in \Psi\}.$$

For any output level y in the interior of the support of Y, we want to describe, as in Section 2, the efficient frontier. In the multi-inputs case, the input-efficient frontier can then either be described through the efficiency measures, since , $\partial C(y) = \{x \mid \theta(x, y) = 1\}$ or through the efficient level of the inputs which, for any $x \in I\!R_+^p$ is given by:

$$x^{\partial}(y) = \theta(x, y)\, x. \in \partial C(y). \tag{5.6}$$

So the frontier, for any (x, y), can be characterized either by $\theta(x, y)$ or by $x^{\partial}(y)$. It is clear that if Ψ is free disposal, $y' \geq y \Rightarrow C(y') \subseteq C(y)$, then, for any x, $\theta(x, y)$ and $x^{\partial}(y)$ are non decreasing in y and when $p = 1$, $x^{\partial}(y)$ determines $\varphi(y)$ as defined in Section 2.

For a given level of outputs y_0 in the interior of the support of Y, consider now the m i.i.d. random variables $X_i, i = 1, \ldots, m$ generated by the conditional p-variate distribution function $F_X(x \mid y_0) = \text{Prob}(X \leq x \mid Y \geq y_0)$ and define the set:

$$\Psi_m(y_0) = \{(x, y) \in I\!R_+^{p+q} \mid x \geq X_i, y \geq y_0\}. \tag{5.7}$$

Then, for any x, we may define

$$\tilde{\theta}_m(x, y_0) = \inf\{\theta \mid (\theta x, y_0) \in \Psi_m(y_0)\}. \tag{5.8}$$

Note that $\tilde{\theta}_m(x, y_0)$ may be computed by the following formula:

$$\tilde{\theta}_m(x, y_0) = \min_{i=1,\ldots,m} \left\{ \max_{j=1,\ldots,p} \left(\frac{X_i^j}{x^j} \right) \right\} \tag{5.9}$$

where a^j denotes the jth component of a vector a.

The following definition is the multivariate extension of our minimum input function of order m as defined in Section 2 for $p = 1$.

Definition 5.1 *For any $x \in I\!R^p_+$, the expected minimum input level of order m denoted by $x^\partial_m(y)$ is defined for all y in the interior of the support of Y as:*

$$x^\partial_m(y) = x\, E(\tilde{\theta}_m(x,y) \mid Y \geq y), \tag{5.10}$$

where we assume the existence of the expectation.

The expected minimum level of inputs of order m may be computed as follows.

Theorem 5.1 *If $x^\partial_m(y)$ exists, it is given by:*

$$x^\partial_m(y) = x \int_0^\infty (1 - F_X(ux \mid y))^m du. \tag{5.11}$$

Proof: The conditional distribution of $\tilde{\theta}_m(x,y)$ is given by

$$\begin{aligned} P(\tilde{\theta}_m(x,y) \leq u \mid Y \geq y) &= P\Big(\min_{i=1,\ldots,m} \Big\{ \max_{j=1,\ldots,p} (\frac{X_i^j}{x^j}) \Big\} \leq u \mid Y \geq y \Big) \\ &= 1 - P\Big(\min_{i=1,\ldots,m} \Big\{ \max_{j=1,\ldots,p} (\frac{X_i^j}{x^j}) \Big\} > u \mid Y \geq y \Big) \\ &= 1 - \Big[P\Big(\max_{j=1,\ldots,p} (\frac{X_i^j}{x^j}) > u \mid Y \geq y \Big) \Big]^m \\ &= 1 - \Big[1 - P\Big(\max_{j=1,\ldots,p} (\frac{X_i^j}{x^j}) \leq u \mid Y \geq y \Big) \Big]^m \\ &= 1 - \Big[1 - P(X \leq ux \mid Y \geq y) \Big]^m . \end{aligned}$$

Now, we obtain

$$E(\tilde{\theta}_m(x,y) \mid Y \geq y) = \int_0^\infty (1 - F_X(ux \mid y))^m du\,, \tag{5.12}$$

from which the theorem follows.■

Here again, in this multivariate case, when $m \to \infty$, the expected minimum input level of order m converges to the efficient input level defining the frontier:

Theorem 5.2 *For any value of $x \in I\!R^p_+$ and for any y in the interior of the support of Y we have:*

$$\lim_{m\to\infty} x^\partial_m(y) = x^\partial(y). \tag{5.13}$$

Proof: This comes immediately from (5.12) where the integral is computed on two intervals of values for u:

$$\begin{aligned} E(\tilde{\theta}_m(x,y) \mid Y \geq y) &= \int_0^{\theta(x,y)} (1 - F_X(ux \mid y))^m du + \int_{\theta(x,y)}^{\infty} (1 - F_X(ux \mid y))^m du \\ &= \theta(x,y) + \int_{\theta(x,y)}^{\infty} (1 - F_X(ux \mid y))^m du, \end{aligned}$$

where the integral converges to zero as $m \to \infty$.■

The nonparametric estimation of $x_m^{\partial}(y)$ is straightforward: we replace the true $F_X(\cdot \mid y)$ by its empirical version, $\widehat{F}_{X,n}(\cdot \mid y)$. We have

$$\begin{aligned} \hat{\theta}_{m,n}(x,y) &= \widehat{E}(\tilde{\theta}_m(x,y) \mid Y \geq y) && (5.14) \\ &= \int_0^{\infty} (1 - \widehat{F}_{X,n}(ux \mid y))^m du. && (5.15) \end{aligned}$$

where

$$\widehat{F}_{X,n}(x \mid y) = \frac{\sum_{i=1}^n 1\!\!I(x_i \leq x, y_i \geq y)}{\sum_{i=1}^n 1\!\!I(y_i \geq y)}.$$

Then, the estimator of the expected minimum input level of order m is given by:

$$\hat{x}_{m,n}^{\partial}(y) = x\,\hat{\theta}_{m,n}(x,y). \quad (5.16)$$

Note that here, due to the multivariate nature of $\widehat{F}_{X,n}(x \mid y)$, there is no simple explicit expression of $\hat{\theta}_{m,n}(x,y)$. The easiest way to compute it is by Monte-Carlo simulations which can be performed as follows.

For a given y, draw a sample of size m with replacement among these x_i such that $y_i \geq y$ and denote this sample by $(X_{1,b}, \ldots, X_{m,b})$. Then compute $\tilde{\theta}_m^b(x,y)$ as:

$$\tilde{\theta}_m^b(x,y) = \min_{i=1,\ldots,m} \left\{ \max_{j=1,\ldots,p} \left(\frac{X_{i,b}^j}{x^j} \right) \right\}.$$

Redo this for $b = 1, \ldots, B$, where B is large. Then,

$$\hat{\theta}_{m,n}(x,y) \approx \frac{1}{B} \sum_{b=1}^{B} \tilde{\theta}_m^b(x,y). \quad (5.17)$$

The empirical frontier, which envelopes all the data points, is given by the standard FDH solution. For instance, the FDH input efficiency measure of any (x,y) is given by:

$$\hat{\theta}_n(x,y) = \inf\{\theta \mid (\theta x, y) \in \widehat{\Psi}_{FDH}\}, \quad (5.18)$$

where $\widehat{\Psi}_{FDH}$ is defined in (1.6). It is computed through the following formula:

$$\hat{\theta}_n(x,y) = \min_{i \mid y_i \geq y} \left\{ \max_{j=1,\ldots,p} \left(\frac{x_i^j}{x^j} \right) \right\}.$$

The corresponding estimated input-efficient level of inputs is given by:

$$\hat{x}^{\partial}_{n}(y) = x\,\hat{\theta}_n(x,y). \tag{5.19}$$

The asymptotic developed in Section 3 for $p = 1$ remains valid, in particular, by Theorem 3.1, we still achieve the $\sqrt{n}$-consistency of $\hat{x}^{\partial}_{m,n}(y)$ to $x^{\partial}_{m}(y)$ for m fixed as $n \to \infty$.

Note also, by (5.15), that

$$\hat{\theta}_{m,n}(x,y) = \hat{\theta}_n(x,y) + \int_{\hat{\theta}_n(x,y)}^{\infty} (1 - \widehat{F}_{X,n}(ux \mid y))^m du\,,$$

so that $\hat{\theta}_{m,n}(x,y) \to \hat{\theta}_n(x,y)$ as $m \to \infty$ for n fixed, or equivalently, the estimated minimum input level of order m, $\hat{x}^{\partial}_{m,n}(y)$ converges to the FDH efficient input level $\hat{x}^{\partial}_{n}(y)$ when $m \to \infty$. However, for a finite m our estimator does not envelop all the data points and is more robust to extreme values, noise or outliers.

From Park, Simar and Weiner (2000) we know that the rate of convergence of the FDH efficiency measures $\hat{\theta}_n(x,y)$ to $\theta(x,y)$ is $n^{1/(p+q)}$, so Theorem 3.2 as to be adapted accordingly.

6 Conclusions

In this paper, we define a statistical concept of a production frontier and we propose a nonparametric estimation of it. The concept is the expected minimum input level (or output level) of order m. It can be applied in very general settings with multiple inputs and multiple outputs. It is related to the usual FDH/DEA nonparametric envelopment estimators but is more robust to extreme values, noise or outliers, in the sense that it does not envelop all the data points. The estimator is easy to implement and the asymptotic properties have been developed. In particular our estimator converges at a rate of $\sqrt{n}$ to its population counterparts. By choosing m_n appropriately as a function of the sample size n, our estimator, as an estimator of the frontier itself, recovers the asymptotic properties of the FDH estimator.

Appendix

A The Expected Maximal Production Function

We follow here, without any proofs, the development of Section 2 in the case we have one output y and p inputs x. Here the production process is defined by the joint distribution of the random vector (X, Y) on $I\!R_+^p \times I\!R_+$. We will concentrate here on the conditional distribution of Y given $X \leq x$. Let the joint distribution be

$$F(x, y) = \text{Prob}(X \leq x, Y \leq y), \tag{A.1}$$

the conditional distribution on Y given $X \leq y$ is described by

$$F_c(y \mid x) = \text{Prob}(Y \leq x \mid X \leq x) \tag{A.2}$$

$$= \frac{F(x, y)}{F_X(x)}, \tag{A.3}$$

where $F_X(x) = \text{Prob}(X \leq x)$.

The upper boundary of the support of $F_c(y \mid x)$ is given by the function:

$$\psi(x) = \sup\{y \mid F_c(y \mid x) < 1\}. \tag{A.4}$$

This function is monotone nondecreasing in x. It is the smallest monotone nondecreasing function which is greater or equal to the output-efficient frontier $\partial P(x)$ as defined in Section 1. It is clear that if the attainable set Ψ is free disposal, the two functions coincide.

Consider now an integer $m \geq 1$ and let $(Y^1, \ldots, Y^m)$ be m independent identically distributed random variables generated by the distribution of Y given $X \leq x$.

Definition A.1 *The expected maximum production function of order m denoted by $\psi_m(x)$ is the real function defined on $I\!R_+^p$ as*

$$\psi_m(x) = E(\max(Y^1, \ldots, Y^m) \mid X \leq x), \tag{A.5}$$

where we assume the existence of this expectation.

The function $\psi_m(x)$ can be computed as follows.

Theorem A.1 *If $\psi_m(x)$ exists, it is given by*

$$\psi_m(y) = \int_0^\infty (1 - [F_c(u \mid x)]^m)\, du. \tag{A.6}$$

From its definition, it is clear that for any y fixed, the function is a increasing function of m. The limiting case when $m \to \infty$ is of particular interest. It achieves the output efficient frontier:

Theorem A.2 *For any fixed value of x we have*

$$\lim_{m\to\infty} \psi_m(x) = \psi(x). \tag{A.7}$$

In particular we have:

$$\psi_m(x) = \psi(x) - \int_0^{\psi(x)} [F_c(u \mid x)]^m \, du. \tag{A.8}$$

Assumption A.1 *The conditional distribution of Y given $X \le x$ has the following property*

$$\text{For all } x' \ge x,\ F_c(y \mid x') \le F_c(y \mid x). \tag{A.9}$$

This assumption says that the chance of producing less than a value y decreases if a firm utilizes more inputs. If $F(x, y)$ represent a production process, this hypothesis is natural.

Theorem A.3 *Under Assumption A.1, $\psi_m(x)$ is monotone nondecreasing in x.*

From an economic point of view, the expected maximal production function of order m, $\psi_m(x)$ has its own interest: it is not the efficient frontier of the production set but it might be useful in term of practical efficiency analysis. Suppose a production unit uses a quantity of input x_0, $\psi_m(x_0)$ gives the expected maximum production among a fixed number of m firms using less than x_0. For this particular unit, it is certainly worth to know this value because it gives a clear indication of how efficient it is compared with these m units. This is achieved by comparing its level y_0 with the value of $\psi_m(x_0)$.

The nonparametric estimator of $\psi_m(x)$ is given by replacing the conditional distribution $F_c(y \mid x)$ by its empirical version:

$$\widehat{F}_{c,n}(y \mid x) = \frac{\widehat{F}_n(x, y)}{\widehat{F}_{X,n}(x)}, \tag{A.10}$$

where $\widehat{F}_n(x, y) = \frac{1}{n} \sum_{i=1}^n 1\!\!I(x_i \le x, y_i \le y)$ and $\widehat{F}_{X,n}(y) = (1/n) \sum_{i=1}^n 1\!\!I(x_i \le x)$.

The estimated FDH output-efficient frontier of the production set is given by:

$$\hat{\psi}_n(x) = \sup\{y \mid \widehat{F}_{c,n}(y \mid x) < 1\}. \tag{A.11}$$

The estimator of the expected maximum output function of order m is defined by:

$$\hat{\psi}_{m,n}(x) = \widehat{E}(\max(Y^1, \ldots, Y^m) \mid X \leq x). \tag{A.12}$$

It is computed through

$$\hat{\psi}_{m,n}(x) = \int_0^\infty (1 - [\widehat{F}_{c,n}(u \mid x)]^m)\, du. \tag{A.13}$$

We have:

$$\hat{\psi}_{m,n}(x) = \hat{\psi}_n(x) - \int_0^{\hat{\psi}_n(x)} [\widehat{F}_{c,n}(u \mid x)]^m\, du, \tag{A.14}$$

from which we obtain again that for all x,

$$\lim_{m\to\infty} \hat{\psi}_{m,n}(x) = \hat{\psi}_n(x). \tag{A.15}$$

The asymptotic theory given by Theorems 3.1 and 3.2 can easily be adapted.

B A Functional Convergence Theorem

Coming back to Theorem 3.1, the asymptotic is developped for $n \to \infty$ for a fixed value of m and for a fixed value of y. It is a pointwise convergence result. In fact we can obtain a more general result by using convergence properties of functionals.

We know that the process $\sqrt{n}(\widehat{S}_n - S)$ indexed by elements $(x, y) \in I\!R_+^{1+q}$ converges in distribution to G, a $1+q$ dimensional S-brownian bridge. G is a gaussian process with zero mean and covariance function $\text{Cov}(f_1, f_2) = E(f_1, f_2) - E(f_1)E(f_2)$ where $f_1(u) = 1\!\!I(u \geq t_1)$ and $f_2(u) = 1\!\!I(u \geq t_2)$, $u, t_1, t_2 \in I\!R^{1+q}$ and the expectation is relative to the distribution characterized by the survivor function S (see van der Vaard and Wellner, 1996, Chapter 2, section 1).

Then the continuous mapping theorem implies that $\sqrt{n}(\hat{\varphi}_{m,n} - \varphi_m)$ converges in distribution to $DT_S(G)$ where DT_S is the continuous linear operator defined in (3.10). This process $DT_S(G)$ is then a q dimensional zero mean gaussian process indexed by y where covariance function is given by (3.12).

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] Aït-Sahalia, Y. (1995), The delta method for nonlinear kernel functionals, working paper, University of Chicago.

[3] Cazals, C. and J.P. Florens (1997), The expected minimum cost function: a nonparametric approach, working paper, IDEI, University of Toulouse.

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

[5] Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19(3), 273–292.

[6] 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.

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

[8] Kneip, A., B.U. Park, and L. Simar (1998), A note on the convergence of nonparametric DEA estimators for production efficiency scores, *Econometric Theory*, 14, 783–793.

[9] Koopmans, T.C. (1951), An Analysis of Production as an Efficient Combination of Activities, in *Activity Analysis of Production and Allocation*, ed. by T.C. Koopmans, Cowles Commission for Research in Economics, Monograph 13. New York: John-Wiley and Sons, Inc.

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

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

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

[13] Serfling, R.T. (1980), *Approximation of Mathematical Statistics*, Wiley, New-York.

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

[15] Simar, L., and P.W. Wilson (2000), Statistical inference in nonparametric frontier models: The state of the art, *Journal of Productivity Analysis* 13, 49–78.

[16] van der Vaard, A. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer, New-York.

[17] Wilson, P. W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319–323.

[18] Wilson, P. W. (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis* 6, 27–45.