INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0222

PARAMETRIC APPROXIMATIONS OF NONPARAMETRIC FRONTIERS

J.P. FLORENS and L. SIMAR

http://www.stat.ucl.ac.be

Parametric Approximations of Nonparametric Frontiers

Jean-Pierre Florens
IDEI/GREMAQ
Université de Toulouse
Toulouse, France

Léopold Simar*
Institut de Statistique
Université Catholique de Louvain
Louvain-la-Neuve, Belgium

July 15, 2002

Abstract

A large amount of literature has been developed on how to estimate frontier functions. The idea is to analyze how firms combine their inputs to produce in an efficient way, the output. The maximal achievable level of output for a given level of inputs defines the production frontier. The efficiency of a particular firm is then characterized by the distance between its level of output and this optimal level it should obtain if it were efficient. From a nonparametric perspective, envelopment estimators have been mostly used, like the Free Disposal Hull (FDH) or the Data Envelopment Analysis (DEA). The statistical theory of these estimators is now available. Nonparametric estimators are very appealing because they rely on very few assumptions, on the other hand, a parametric form for the production function allows for a richer economic interpretation of the production process under analysis. Here, in a deterministic frontiers framework, most of the approaches rely on "ad hoc" procedures based on standard regression methods (shifted OLS, corrected OLS, and MLE) and are based on strong distributional assumptions on the production process. Also they characterizes rather properties of the center of the cloud of points rather than its boundary. In this paper, we investigated a new approach, which tries to capture the shape of the cloud points near its boundary. It combines the nonparametric and the parametric approaches, by offering parametric approximations of nonparametric frontiers. For the nonparametric part, we use the FDH estimator or expected frontier of order-m, introduced by Cazals, Florens and Simar (2002). We provide the statistical theory for the obtained estimators (consistency and asymptotic distribution). We illustrate with some simulated examples, showing the advantages of our method compared with the regression-type estimators.

Keywords: production function, frontier, nonparametric estimation.
JEL Classification: C13, C14, D20.

*Visiting IDEI, Toulouse with the support of the Ministère de l'Education nationale , de la recherche et de la technologie, France. Financial research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) from the Belgian Government, and from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) are also acknowledged.

1 Introduction

The estimation of technical efficiencies of production units from frontier models has been extensively used in the literature since the pioneering work of Farrell (1957), for a nonparametric approach and of Aigner and Chu (1968) for a parametric approach. The idea is to analyze how firms combine their inputs to produce in an efficient way, the output. The maximal achievable level of output for a given level of inputs defines the production frontier[1]. This production function is the boundary of the so-called attainable set $\Psi = \{(x, y)|x \text{ can produce } y\}$, where $x \in I\!R^p_+$ is a vector of inputs and $y \in I\!R_+$ is the output. The efficiency of a particular firm is then characterized by the distance between its level of output and this optimal level it should obtain if it were efficient.

Since the production function is unknown, it is estimated from a set of observations $\mathcal{X} = \{(x_i, y_i)|i = 1, \ldots, n\}$. From a nonparametric perspective, envelopment estimators have been mostly used, like the Free Disposal Hull (FDH, initiated by Deprins, Simar and Tulkens, 1984) or the Data Envelopment Analysis (DEA initiated by Farrell, 1957, and popularized as programming estimator by Charnes, Cooper and Rhodes, 1978). The statistical theory of these estimators is now available (see Simar and Wilson, 2000, for a recent survey). Recently, robust nonparametric envelopment estimators have been suggested by Cazals, Florens and Simar (2002). They introduce the concept of expected frontier of order-m. For instance, in the output-oriented case, it is defined as the expected maximum achievable level of output among m firms drawn in the population of firms using less than a given level of input. A simple nonparametric estimator is proposed, which does not envelop all the data points and so, is more robust to outliers and/or extreme values. All its statistical properties are known.

Nonparametric estimators are very appealing because they rely on very few assumptions: no particular shape for the attainable set and its frontier (only free disposability[2] for the FDH and in addition, convexity of the attainable set, for the DEA) and no particular distributional assumptions for the distribution of (x, y) on Ψ. The drawback of nonparametric approaches is that the results are more difficult to interpret in term of the sensitivity of the production of output to particular inputs (shape of production function, elasticities, . . .), and inference for the measures of interest (confidence intervals, test of hypothesis) is not easy (see Simar and Wilson, 2000). Also, the curse of dimensionality (with FDH and DEA methods) implies that large sample sizes are to be considered to get sensible results.

[1]In this paper, hereafter, we will refer to this output-oriented notion of technical efficiency. The same could be done for the input-oriented case, where for a given set of outputs, firms try to achieve a minimal level of the input

[2]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.

On the other hand, a parametric form for the production function allows for a richer economic interpretation of the production process under analysis: here the parameters are usually much easier to interpret and to estimate, but at a cost of a reasonable parametric specification of the model. For parametric models[3], we can identify two categories of approaches. the first one is due to Aigner and Chu (1968) where they try to envelop the data in a parametric way by solving appropriate mathematical programs; although the method seems to be appealing, it suffers from some drawbacks discussed below. The second family of parametric approaches are in the spirit of Greene (1980): they are based on "ad hoc" procedures based on standard regression methods (shifted OLS, corrected OLS, and MLE), but, as shown below, they are based on strong distributional assumptions on the production process, and characterizes properties of the center of the cloud of points rather than the shape of its boundary.

In this paper, we investigated a new approach, which is, in a certain sense, in the spirit of Aigner and Chu but avoiding its drawbacks, and tries to capture the shape of the cloud of points near its boundary. It combines the nonparametric and the parametric approaches. A parametric frontier model will be estimated using a two-step procedure: first, identify, by using a nonparametric method, where is located the production frontier, then, in a second step, adjust a parametric model to the obtained nonparametric frontier. The argument is that the production frontier is the locus of optimal production situations, so we might hope to get substantial improvements (bias, variance reduction, etc...), if we use only efficient observations to estimate it.

This technique was used, in a pure descriptive and pragmatic way, by Thiry and Tulkens (1992): they fitted the FDH-efficient units in the sample by a linear model using standard OLS. The standard OLS inference used there (p-values, confidence intervals,...) is of course incorrect. Simar (1992) developed the same idea, in a context of a panel of data and proposed a bootstrap algorithm to provide the inference on the parameters of the model. But, as pointed in Simar (1992), no theoretical results were so far obtained about the statistical properties of the obtained estimators (consistency, asymptotic distributions, ...). In our paper here, we provide the complete statistical analysis of a similar two-step procedure: we first project all the observations on a nonparametric estimator of the frontier and then we fit a parametric model to the obtained points. We also improve the method by using, in the first step of the procedure, the more robust order-m frontier estimator.

The paper is organized as follows. Section 2 introduces the basic concepts and notations

[3]In this paper, we only consider deterministic models where $\text{Prob}((x, y) \in \Psi) = 1$. We don't consider stochastic models where noise is introduced in the model. Here only parametric approaches have been developed so far in the spirit of Aigner, Lovell and Schmidt (1997) and of Meeusen and van den Broek (1977). In a nonparametric model, noise cannot be identified from inefficiencies, see Hall and Simar (2002).

and motivates our approach. Section 3 analyzes the statistical properties whereas Section 4 illustrates with some numerical examples. Section 5 concludes.

2 Basic Concepts and Notations

The production process is defined by the joint distribution of the random vector (X, Y) on $\mathbb{R}_+^p \times \mathbb{R}_+$: it characterizes the Data Generating Process (DGP). Such distribution is usually decomposed into a marginal distribution on X and a conditional distribution on Y, given $X = x$. We will rather concentrate on the conditional distribution of Y given $X \leq x$. Let the joint distribution be

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

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

$$\begin{aligned} F_c(y \mid x) &= \text{Prob}(Y \leq y \mid X \leq x) & (2.2) \\ &= \frac{F(x, y)}{F_X(x)}, & (2.3) \end{aligned}$$

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:

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

As shown in Cazals, Florens and Simar (2002), this function is monotone nondecreasing in x. It is the smallest monotone nondecreasing function which is greater or equal to the output-efficient frontier of Ψ defined for all x as $\phi(x) = \{y|(x, \lambda y) \notin \Psi, \forall \lambda > 1\}$ and if the attainable set Ψ is free disposal (a quite minimal assumption we will maintain hereafter), the two functions coincide.

2.1 Standard parametric approaches

Suppose we introduce a parametric model[4] for the frontier function $\varphi(x)$: we suppose the production frontier can be written as a specified analytical function depending on a finite number of parameters $\theta \in \mathbb{R}^k$. We denote this parametric model by $\varphi(x; \theta)$. Defining for all (x, y), $u = \varphi(x; \theta) - y$, we can write our parametric model as

$$y = \varphi(x; \theta) - u, \tag{2.5}$$

[4]From now on, unless necessary, we will not distinguish in our notation a random variable Z and a particular value z. So hereafter, by using the notation z, the context will tell if it represents the random variable or a particular value of it.

where $u \geq 0$, with probability one, and the random variables (x, u) have some joint distribution involved by $F(x, y)$. We want to estimate this model from a random sample of observation $\mathcal{X} = \{(x_i, y_i)|i = 1, \ldots, n\}$.

2.1.1 Mathematical programming estimators

Aigner and Chu (1968) propose to estimate the model, in the linear case $\varphi(x; \theta) = \alpha + \beta x$, by solving the following optimization problems, either a constrained linear program:

$$\begin{aligned} &\min_{\alpha,\beta} \sum_{i=1}^{n} |y_i - \alpha - \beta' x_i| \\ &\text{s.t.} \quad y_i \leq \alpha + \beta' x_i, \; i = 1, \ldots, n \end{aligned} \tag{2.6}$$

or a constrained quadratic program:

$$\begin{aligned} &\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta' x_i)^2 \\ &\text{s.t.} \quad y_i \leq \alpha + \beta' x_i, \; i = 1, \ldots, n \end{aligned} \tag{2.7}$$

The method is interesting since it does not imply any distributional assumptions, but, to the best of our knowledge, no statistical properties of the estimators of θ have been developped so far and so, no inference (even asymptotically) is available. Without consistency, even any bootstrap procedure should be questionable. In addition, the method has the main drawback of giving in the objective function (in particular in the quadratic case) too much weights to points which are far from the frontier.

Note that this approach can also be viewed as a maximum likelihood procedure under very restrictive assumptions: if the random term u is independent of x in (2.5) and if the law of u, $f(u)$, is an exponential (half-normal), the obtained estimators from the linear (quadratic) program are the maximum likelihood estimators of θ. However, this does not help to derive any statistical properties of the estimators, because we are in a non-standard problem: the support of y depends on θ and $f(0) > 0$.

2.1.2 Regression-type estimators

Under some restricting hypothesis, the model can be approached in terms of (shifted) regression function. This is the main idea of Greene (1980). Suppose indeed that u is independent[5] of x, then from (2.5), we have:

$$E(y|x) = \varphi(x; \theta) - E(u). \tag{2.8}$$

[5]Note that this hypothesis is much more strong than the usual $E(u|x) = \mu$, where μ is a constant, but the argument following (2.8) is still valid with this weaker assumption.

So, at a shift, $\varphi(x;\theta)$ is the regression function of y on x. Greene (1980) proposes a shifted OLS procedures to estimate (2.5) in the linear case where $\theta = (\alpha, \beta)$. The idea is to estimate the slope β by standard OLS and then shift the obtained hyperplane upward, such that all the residuals are negative. This define the estimator of α which is proven to be consistent. By specifying a parametric family of distributions for u, Greene proposes also a corrected OLS and a maximum likelihood estimator of (α, β): the corrected OLS, keep the standard OLS estimator of the slope and corrects the OLS estimator of α by taking the particular conditions on the first moments of u into account, derived from the chosen family of d.f., to estimate the shift $E(u)$. He shows also that, for a Gamma distribution of u, the MLE shares its traditional sampling properties, if the shape parameter of the gamma is large enough, to avoid unknown boundary problems ($f(u)$ has to converge smoothly to zero when $u \to 0^+$). See also Deprins and Simar (1985) for the asymptotic relative efficiency of the MLE with respect to OLS techniques in the estimation of the slope.

The usefulness of this approach is limited by the strong restriction on the law of (u, x): independence between u and x and restriction on the shape of $f(u)$ for the corrected OLS and the MLE approaches. But the basic drawback of the model is the consequence (2.8) of the independence: the frontier function is by construction, at a shift, the regression function of y on x. So any estimation procedure will basically capture the shape of the "middle" of the cloud of data points. This is not a natural approach in the context of frontier estimation in deterministic models where we would prefer to capture the shape of the *boundary* of the observed cloud of points. The following two simple examples show how dramatic could these approaches be when the independence assumption between u and x is violated.

Consider first, the simplest case $p = 1$ with $0 \leq x \leq 1$, and the frontier is defined by the equation $y = x$. Now suppose the DGP generates (x, y) uniformly on the triangle under the frontier. It is clear that any OLS-type technique (shifted and/or corrected) will behave dramatically: it provides biased, inconsistent estimators of all the parameters, the slope and the intercept. Consider now the situation where $y = x - u$, the distribution of $(u|x)$ being exponential with parameter $\lambda(x)$. Here again, regression type methods would provide inconsistent and biased estimators. We will come back to these examples in Section 4.

Of course, one could argue that this concerns specification problems, but the idea is to provide an other way to estimate model (2.5), which relies on a more natural approach and does not need too restrictive assumptions.

2.2 Nonparametric estimators

As pointed above, the idea is to estimate the frontier from points which are identified as being efficient by a first step nonparametric estimator. This first step can thus be viewed as a kind of “filtering” for eliminating from the sample clearly inefficient units which certainly does not provide substantial information to analyze how to transform efficiently inputs into an output. Two kinds of nonparametric estimator can be used in this first step filtering, depending on the frontier model we want estimate. They rely on a minimal set of assumptions on the observed technology. The two next subsections resume the basic definitions for these two estimators.

2.2.1 The FDH estimator

The FDH estimator was proposed by Deprins, Simar and Tulkens (1984). In this approach, the attainable set is estimated as the smallest free disposal set containing all the data points. In our context , this set is given by

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

The FDH estimator of the frontier function is the boundary of $\widehat{\Psi}_{FDH}$ in the output direction. It can be defined as:

$$\hat{\varphi}_{FDH,n}(x) = \hat{\varphi}_n(x) = \max_{i \,|\, x_i \leq x} (y_i). \tag{2.9}$$

As pointed in Cazals, Florens and Simar (2002), the FDH estimator can be viewed as a plug-in estimator of $\varphi(x)$, where the unknown $F_c(y|x)$ in (2.4) has been replaced by its empirical analog, $\widehat{F}_{c,n}(y \mid x)$, provided by $\mathcal{X}$.

This estimator relies on the minimal assumption of free disposability of Ψ, when convexity is also assumed, the estimator can be improved by convexifying $\widehat{\Psi}_{FDH}$. This provides the DEA estimator mentionned above. We will not pursue this idea below, preferring to rely on a minimal set of assumptions on Ψ.

The asymptotic of the FDH was first derived by Korostelev, Simar, and Tsybakov (1995) for the consitency and by Park, Simar and Weiner (2000) for asymptotic sampling distributions. In summary, it is shown in the latter, that for all x, as $n \to \infty$, $n^{1/(p+1)}\left(\widehat{\varphi}_n(x) - \varphi(x)\right)$ converges to a Weibull distribution whose parameters depends on the density of (x, y) near the frontier point. Notice that we have the “curse of dimensionality” which is often the price to pay for a flexible nonparametric approach.

2.2.2 The expected frontier of order-m

The FDH estimator provides an estimated production function which envelops all the data points. It is thus very sensitve to outliers and/or extreme values. In place of estimating the full frontier, Cazals, Florens and Simar (2002) propose rather to estimate an expected maximal output frontier of order-m. This order-m output frontier is defined as follows. Consider a fixed integer m, we can define, for a given level of the inputs x, the expected value of the maximum of m random variable $Y^1, \ldots, Y^m$, drawn from the conditional distribution of the output Y, given $X \leq x$. Formally:

$$\varphi_m(x) = E\left[\max(Y^1, \ldots, Y^m) \mid X \leq x\right] = \int_0^\infty \left(1 - [F_c(y \mid x)]^m\right) dy, \tag{2.10}$$

where the integrand is identically zero for $y \geq \varphi(x)$. From an economic point of view, this expected maximal production function of order m, has its own interest: it is not the efficient frontier of the production set but it gives the expected maximum production among a fixed number of m firms using less than x as inputs. This can be viewed as a reasonable benchmark for firms using a level x of inputs.

A natural estimator is provided by a plug-in argument:

$$\widehat{\varphi}_{m,n}(x) = \widehat{E}\left[\max(Y^1, \ldots, Y^m) \mid X \leq x\right], \tag{2.11}$$

it may be computed through

$$\widehat{\varphi}_{m,n}(x) = \int_0^\infty \left(1 - [\widehat{F}_{c,n}(y \mid x)]^m\right) dy, \tag{2.12}$$

where the integrand is identically zero for $y \geq \widehat{\varphi}(x)$. An exact fromula is available in order to compute $\widehat{\varphi}_{m,n}(x)$, but in practice, it is more easy to approximate (2.11), by the following Monte-Carlo method. For a given x, draw a random sample of size m with replacement among these y_i such that $x_i \leq x$ and denote this sample by $(y_b^1, \ldots, y_b^m)$. Then compute $\tilde{\varphi}_{b,m}(x) = \max_{i=1,\ldots,m}(y_b^i)$. Redo this for $b = 1, \ldots, B$ where B is large. Finally, we have

$$\widehat{\varphi}_{m,n}(x) \approx \frac{1}{B}\sum_{b=1}^{B} \tilde{\varphi}_{b,m}(x),$$

where the quality of the approximation can be tuned by the choice of B.

To summarize the properties of these functions, we have, for all x:

$$\lim_{m\to\infty} \varphi_m(x) = \varphi(x) \tag{2.13}$$

$$\lim_{m\to\infty} \widehat{\varphi}_{m,n}(x) = \widehat{\varphi}_n(x) = \widehat{\varphi}_{FDH,n}(x) \tag{2.14}$$

$$\sqrt{n}\left(\widehat{\varphi}_{m,n}(x) - \varphi_m(x)\right) \to N(0, \sigma^2(x)) \text{ as } n \to \infty, \tag{2.15}$$

where an expression for $\sigma^2(x)$ is available. A stronger functional convergence theorem of $\widehat{\varphi}_{m,n}$ to φ_m is provided in Appendix B of Cazals, Florens and Simar (2002).

First we note that due to the expectation in (2.11), the estimator will not envelop all the data points, even for large m. So it more robust to extreme values and/or to outliers than the FDH estimator. It is also important to note that the nonparametric estimator of the order-m frontier achieves $\sqrt{n}$ consistency and that we don't have the curse of dimensionality. These two argument are in favour of using the order-m estimator in our two-step estimator below. As shown below, the asymptotic is also much easier.

3 Our Two-Steps Estimators

Consider a parametric family of functions defined on $\mathbb{R}_+^p$, $\{\varphi(\cdot;\theta) \,|\, \theta \in \Theta\}$. We want to estimate the value of $\theta \in \mathbb{R}^k$ which correspond to the best parametric approximation of the frontier function. We will consider two cases: (1) we want to estimate the best parametric approximation of the order-m frontier function $\varphi_m(x)$ and (2) we want to estimate the best parametric approximation of the full frontier function $\varphi(x)$.

According to the case, we define our estimators of θ as

$$\hat{\theta}_n^m = \arg\min_\theta \left[\sum_{i=1}^n (\widehat{\varphi}_{m,n}(x_i) - \varphi(x_i;\theta))^2\right] \tag{3.1}$$

$$\hat{\theta}_n = \arg\min_\theta \left[\sum_{i=1}^n (\widehat{\varphi}_{n}(x_i) - \varphi(x_i;\theta))^2\right] \tag{3.2}$$

In this section, we analyze the convergence of our estimators. For the order-m frontier case, we obtain the rates of convergence and the asymptotic normality. For the full frontier, we obtain the consitency. A more general presentation of the asymptotic of this kind of estimators is provided in Appendix A.

In order to simplify the presentation here, we restrict our parametric family, in this section, to the class of linear model, *i.e.*

$$\varphi(x;\theta) = g'(x)\,\theta, \tag{3.3}$$

where $g'(x) = (g_1(x) \dots g_k(x))$ with the k functions $g_j(\cdot)$ being known scalar functions of the vector x. For more general parametric models, see Appendix A.

3.1 Expected frontier of order m

In this case the estimator of θ given by (3.1), has the explicit expression:

$$\hat{\theta}_n^m = \left\{\widehat{E}[g(x)g'(x)]\right\}^{-1} \widehat{E}\left[g(x)\widehat{\varphi}_{m,n}(x)\right], \tag{3.4}$$

where $\widehat{E}$ stands for the empirical average: $\widehat{E}(h(x)) = (1/n)\sum_{i=1}^{n} h(x_i)$, for any function $h(x)$. We know also that a pseudo-true value[6] of θ can be defined as:

$$\theta^m = \arg\min_{\theta} \left[\sum_{i=1}^{n} (\varphi_m(x_i) - \varphi(x_i;\theta))^2 \right]. \tag{3.5}$$

In our case here, θ^m can be explicitely written as:

$$\theta^m = \left\{ \widehat{E}[g(x)g'(x)] \right\}^{-1} \widehat{E}\left[g(x)\varphi_m(x)\right]. \tag{3.6}$$

The error of estimation can be written as:

$$\hat{\theta}_n^m - \theta^m = \left\{ \widehat{E}[g(x)g'(x)] \right\}^{-1} \widehat{E}\left[g(x)(\hat{\varphi}_{m,n}(x) - \varphi_m(x))\right] \tag{3.7}$$

Consistency
Under general regularity conditions, we have, by the law of large numbers, that $\left\{ \widehat{E}[g(x)g'(x)] \right\}^{-1}$ converges to $E\left[g(x)g'(x)\right]^{-1}$, where the expectation E is with respect to f_X. So the problem is to analyze if we have the convergence of $\widehat{E}\left[g(x)(\hat{\varphi}_{m,n}(x) - \varphi_m(x))\right]$ to zero. First, we approximate, at the first order, the difference between $\hat{\varphi}_{m,n}(x)$ and $\varphi_m(x)$ by linearizing the difference between (2.12) and (2.10). We obtain:

$$\begin{aligned} \hat{\varphi}_{m,n}(x) - \varphi_m(x) &= -\int_0^\infty \frac{m(1-F_c(y|x))^{m-1}}{F_X(x)} \\ &\quad \left\{ [\widehat{F}_n(x,y) - F(x,y)] - F_c(y|x)[\widehat{F}_{X,n}(x) - F_X(x)] \right\} \, dy. \end{aligned}$$

Thus we have:

$$\begin{aligned} \widehat{E}\left[g(x)(\hat{\varphi}_{m,n}(x) - \varphi_m(x))\right] &= -\frac{1}{n}\sum_{i=1}^{n} g(x_i) \int_0^\infty \frac{m(1-F_c(y|x_i))^{m-1}}{F_X(x_i)} \\ \times \left\{ [\widehat{F}_n(x_i,y) - F(x_i,y)] - F_c(y|x_i)[\widehat{F}_{X,n}(x_i) - F_X(x_i)] \right\} \, dy & \qquad (3.8) \\ &= -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int_0^\infty g(x_i) \frac{m(1-F_c(y|x_i))^{m-1}}{F_X(x_i)} \\ \times \left\{ [1\!\!1(x_j \le x_i, y_j \le y) - F(x_i,y)] - F_c(y|x_i)[1\!\!1(x_j \le x_i) - F_X(x_i)] \right\} \, dy. & \qquad (3.9) \end{aligned}$$

Now, the expression can be symmetrized (see below) to obtain a U-statistics (by permuting all the indices i and j and taking the arithmetic mean) and clearly, each term in this double summation has an expectation equal to zero, where the expectation is over (x_i, y_i, x_j, y_j). So

[6] A pseudo-true value of θ, in our model, is the unknown value of the best parametric approximation of the true frontier of order-m. In (3.5), we use the empirical average over x_i for defining the objective function. We could use other weight function, as *e.g.*, the true expectation with respect to f_X. More details are provided in the Appendix A.

by Serfling (1980, Theorem A, page 190), the U-statistics converges almost surely to zero as $n \to \infty$. As a direct consequence, we obtain:

$$\hat{\theta}_n^m \to \theta^m, \ a.s. \text{ when } n \to \infty \tag{3.10}$$

Asymptotic Normality

We now analyze the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^m - \theta^m)$. By using (3.7) we have:

$$\sqrt{n}(\widehat{\theta}_n^m - \theta^m) = \left\{\widehat{E}[g(x)g'(x)]\right\}^{-1} \sqrt{n}\widehat{E}\left[g(x)(\hat{\varphi}_{m,n}(x) - \varphi_m(x))\right].$$

It is clear that, in the asymptotic distribution we can replace $\left\{\widehat{E}[g(x)g'(x)]\right\}^{-1}$ by its limiting value $\{E[g(x)g'(x)]\}^{-1}$. Now, we use again the properties of the symmetrized U-statistics, using (3.9):

$$\sqrt{n}\widehat{E}\left[g(x)(\hat{\varphi}_{m,n}(x) - \varphi_m(x))\right] = \frac{\sqrt{n}}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} U_{i,j} \ ,$$

where

$$\begin{aligned}
U_{i,j} &= U_{i,j}^{(1)} + U_{i,j}^{(2)} \\
&= -\int_0^\infty g(x_i)\frac{m(1-F_c(y|x_i))^{m-1}}{F_X(x_i)} \\
&\qquad \times \left\{[1\!\!1(x_j \le x_i, y_j \le y) - F(x_i,y)] - F_c(y|x_i)[1\!\!1(x_j \le x_i) - F_X(x_i)]\right\} \, dy. \\
&- \int_0^\infty g(x_j)\frac{m(1-F_c(y|x_j))^{m-1}}{F_X(x_j)} \\
&\qquad \times \left\{[1\!\!1(x_i \le x_j, y_i \le y) - F(x_j,y)] - F_c(y|x_j)[1\!\!1(x_i \le x_j) - F_X(x_j)]\right\} \, dy.
\end{aligned}$$

Now the expectation of the first term of $U_{i,j}$, with respect to (x_j, y_j) is equal to zero: $E(U_{i,j}^{(1)}|(x_i,y_i)) = 0$. By Serfling (1980, Theorem A, page 192), we have

$$\sqrt{n}\widehat{E}\left[g(x)(\widehat{\varphi}_{m,n}(x) - \varphi_m(x))\right] \to N(0,\Sigma),$$

where $\Sigma = Var\left[E(U_{i,j}^{(2)}|(x_i,y_i))\right]$.

The asymptotic normality for $\hat{\theta}_n^m$ now follows directly:

$$\sqrt{n}(\widehat{\theta}_n^m - \theta^m) \to N(0, A\,\Sigma\,A). \tag{3.11}$$

where $A = \{E[g(x)g'(x)]\}^{-1}$.

The above asymptotic results (consitency and asymptotic normality) are simplified particular cases of more general theorems (Theorem A.1 and Theorem A.2) provided in Appendix A. Here we only consider linear parametric models and we choose an empirical weighting function for defining the pseudo-true value. It should also be noted that in many cases (see Section 4), the true φ^m can be expressed as a linear function of x, as in (3.3). In these cases, the pseudo-true value θ^m is simply the true value of θ.

Practical computations

The evaluation of the covariance matrix in (3.11) is rather complicated. For practical purposes of inference on the value of θ^m, it will be easier to approximate this matrix by a bootstrap method. The idea is very simple, and as shown in Section 4, it is easy to implement in a reasonable computing time.

The aim is to approximate the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n^m - \theta^m)$ by its bootstrap analog. The algorithm could be described as follows

[1] Draw a random sample of size n with replacement from $\mathcal{X}$ to obtain the bootstrap sample $\mathcal{X}_b^* = \{(x_{b,i}^*, y_{b,i}^*)|i = 1, \ldots, n\}$.

[2] With this sample $\mathcal{X}_b^*$, compute $\sqrt{n}(\widehat{\theta}_{b,n}^{m,*} - \theta_b^{m,*})$, where

$$\widehat{\theta}_{b,n}^{m,*} = \left\{\widehat{E}_b^*[g(x)g'(x)]\right\}^{-1} \widehat{E}_b^* \left[g(x)\widehat{\varphi}_{m,n}^{*,b}(x)\right] \tag{3.12}$$

$$\theta_b^{m,*} = \left\{\widehat{E}_b^*[g(x)g'(x)]\right\}^{-1} \widehat{E}_b^* \left[g(x)\widehat{\varphi}_{m,n}(x)\right] \tag{3.13}$$

where $\widehat{E}_b^*$ stands for the empirical average $\widehat{E}_b^*(h(x)) = (1/n)\sum_{i=1}^n h(x_{b,i}^*)$, for any function $h(x)$ and where $\widehat{\varphi}_{m,n}^{*,b}(x)$ is the nonparametric estimator of the order-m frontier obtained from the bootstrap sample $\mathcal{X}^*$. In fact, $\widehat{\theta}_{b,n}^{m,*}$ and $\theta_b^{m,*}$ are the bootstrap analogs of $\widehat{\theta}_n^m$ and of θ^m, respectively.

[3] Redo steps [1] and [2], a large number of times, $b = 1, \ldots, B$.

The empirical distribution of $\sqrt{n}(\widehat{\theta}_{b,n}^{m,*} - \theta_b^{m,*}),\ b = 1, \ldots, B$ approximate the distribution of $\sqrt{n}(\widehat{\theta}_n^m - \theta^m)$ and can be used for doing inference (bootstrap confidence intervals, percentiles methods, ...). In particular, its variance matrix approximate $A\,\Sigma\,A$.

In Appendix A, variations of the bootstrap algorithm are proposed according to the pseudo-true value we are interested in.

3.2 Estimation of the efficient frontier

Now we analyze the asymptotic behavior of $\widehat{\theta}_n$, given in (3.2), considered as an estimator of the pseudo-true value of θ, defined as

$$\theta = \arg\min_{\theta} \left[\sum_{i=1}^{n}(\varphi(x_i) - \varphi(x_i;\theta))^2\right], \tag{3.14}$$

which provides the best parametric approximation of the efficient frontier function.

As pointed in Appendix A, the asymptotic behavior of $\widehat{\theta}_n$ is much more complex, because extreme values statistics are more difficult to handle. The main difficulty is about the asymptotic behavior of $(1/n)\sum_{i=1}^n g(x_i)\max(y_j)1\!\!I(x_j \leq x_i)$ which cannot be analyzed using the properties of U-statistics as above. Note also that the available asymptotic of the FDH first-step estimator $\widehat{\varphi}_n(x)$ is less powerful than for the order-m frontier (no functional convergence theorem available).

However, as shown in the appendix (Theorem A.3), we have the consistency of our estimator:

$$(\widehat{\theta}_n - \theta) \xrightarrow{p} 0. \tag{3.15}$$

In particular, if the parametric model is correct, θ is the true value of the parameter.

4 Numerical Illustrations

In this section, we presents 4 simulated examples to illustrate our procedure.

4.1 Example 1

We consider first the case where (x, y) is uniformly distributed over the region $D = \{(x, y)|0 \leq x \leq 1,\ 0 \leq y \leq x\}$. So the frontier function is $\varphi(x) = x$ and the order-m frontier can be computed as

$$\varphi_m(x) = x(1 - A_m),$$

where $A_m = \sum_{j=0}^m \binom{m}{j} \frac{(-1)^j 2^{m-j}}{m+j+1}$. So, in this particular case, the frontier and the order-m frontiers, are both linear in x but with different slopes. The results for a sample of size $n = 100$ are displayed on Figure 1.

In Table 1 we give the corresponding estimates of the intercept α and of the slope β of the corrsponding frontier. We remark in Figure 1, that the order-100 frontier achieves almost the FDH estimates. Of course, in this example, the shifted OLS is a catastrophy: the slope capture the slope of the regression of y on x (1/3), then the intercept is adjusted.

The covariance matrix $A\Sigma A$ given in (3.11) was estimated by our bootstrap procedure described above. For the case $m = 25$, we obtain with $B = 200$ bootstrap loops (computation time with our Matlab code, less than 15', on a Pentium III, 450 Mhz):

$$\widehat{A\Sigma A} = \begin{pmatrix} 0.0893 & -0.1403 \\ -0.1403 & 0.2987 \end{pmatrix},$$

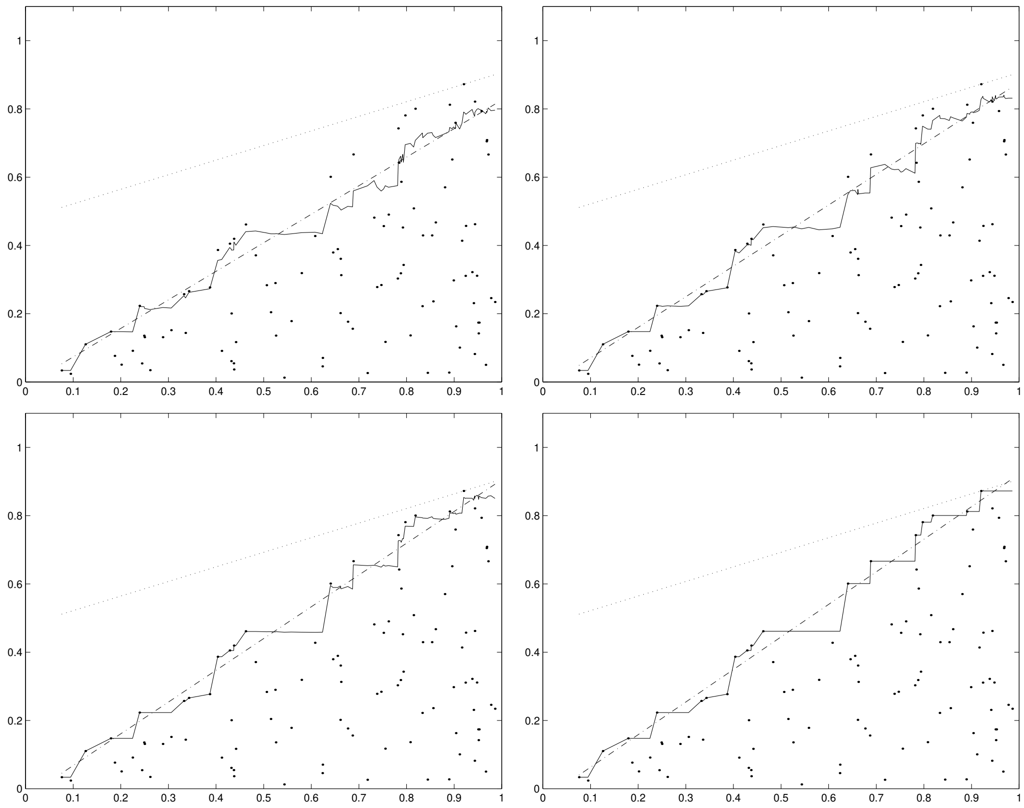

Figure 1: Results for the Example 1: $n = 100$. In solid line, the order-$m$frontier, in dash-dotted, our linear approximation of the order-m frontier. In dotted line, the shifted OLS fit. From left to right and from top to bottom: $m = 25, 50, 100$ and the FDH.

value of m	$\widehat{\alpha}^m$	$\widehat{\beta}^m$
$m = 25$	-0.0102	0.8360
$m = 50$	-0.0193	0.8963
$m = 100$	-0.0252	0.9305
FDH ($m = \infty$)	-0.0323	0.9539
Shifted-OLS	0.4789	0.4276
True full-frontier	0.0	1.0

Table 1: Results for the Example 1: $n = 100$.

providing the asymptotic normal, 95% confidence intervals:

$$\alpha^{25} \in [-0.0643, 0.0430]$$
$$\beta^{25} \in [0.7335, 0.9364].$$

The 95% percentile bootstrap confidence limits are:

$$\alpha^{25} \in [-0.0780, 0.0204]$$
$$\beta^{25} \in [0.7226, 0.9365].$$

This is quite similar than those obtained with the normal approximation. So, for $m = 25$, a sample of size $n = 100$ allow to use the normal approximation. For the case $m = 100$:

$$\widehat{A\Sigma A} = \begin{pmatrix} 0.0586 & -0.0765 \\ -0.0765 & 0.1421 \end{pmatrix},$$

providing the asymptotic normal, 95% confidence intervals:

$$\alpha^{100} \in [-0.0727, 0.0222]$$
$$\beta^{100} \in [0.8565, 1.0043].$$

The 95% percentile bootstrap confidence limits are:

$$\alpha^{100} \in [-0.0030, -0.0971]$$
$$\beta^{100} \in [0.8506, 0.9919],$$

showing that for the slope, the normal approximation is still valid for $n = 100$ and $m = 100$, but for α, the normal appriximation is not so good and needs larger sample sizes. This is not a surprise, we know that for $m = 100$, we are not far from the FDH estimator which has a Weibull limiting distribution.

4.2 Example 2

We now consider a case where the regression-type estimators are valid: we choose a Cobb-Douglas, log-linear frontier given by the model

$$y = x^{0.5}\, e^{-u},$$

where x is uniform on $[0, 1]$ and u, independent of x, is Exponential with parameter $\lambda = 3$. Here $E(u|x) = E(u) = 1/3$. This corresponds to an average output-efficiency of 0.75. So here, the true frontier function is $\varphi(x) = x^{0.5}$ and the order-m frontier can be computed as

$$\varphi_m(x) = x^{0.5}(1 - B_m),$$

where $B_m = 2^m \sum_{j=0}^{m} \binom{m}{j} \frac{(-1)^{m+j} 1.5^j}{3m-j+1}$. So, in this particular case, the frontier and the order-m frontier are both log-linear in x, but with a shift for the order-m frontier in the log scale. The results for a sample of size $n = 100$ are displayed on Figure 2.

In Table 2 we give the corresponding estimates of the intercept α and β of the corrsponding functions in the log-linear model. We remark again, in Figure 2, that the order-100 frontier achieves almost the FDH estimates. Of course, in this example, the shifted OLS estimates of the slope β is not so bad, because we are in a model where u is independent of x, but still, the estimation of α is is not so good. That's because it relies on one particular residual.

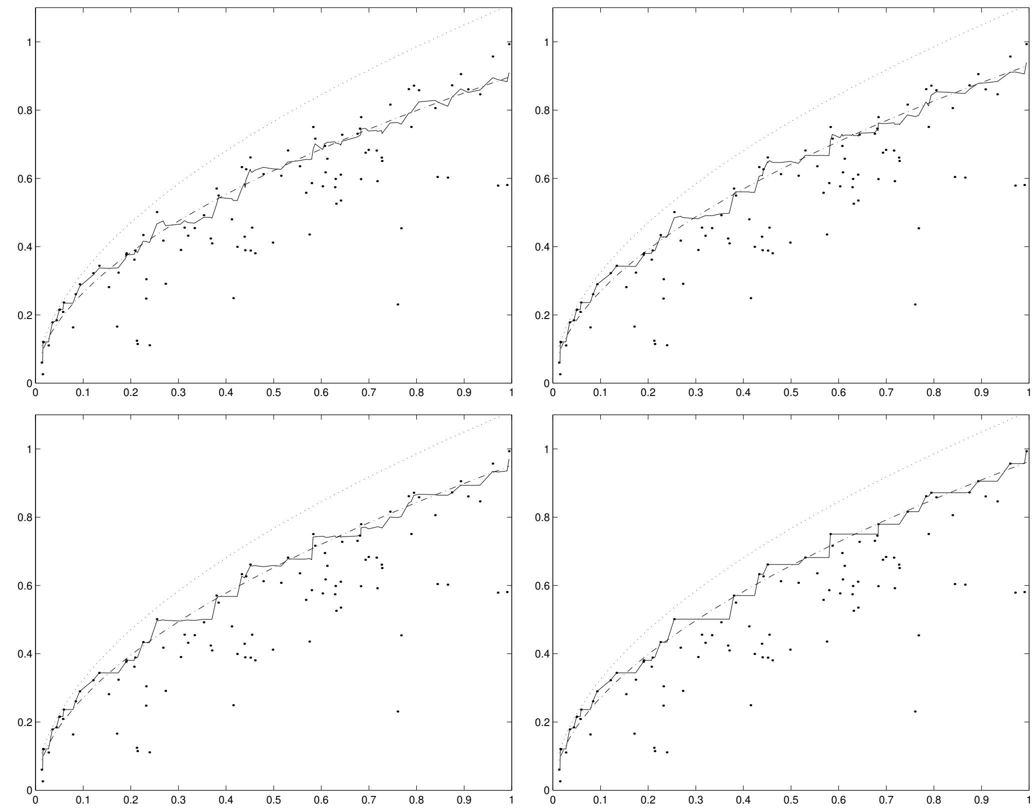

Figure 2: Results for the Example 2: $n = 100$. In solid line, the order-$m$frontier, in dash-dotted, our linear approximation of the order-m frontier. In dotted line, the shifted OLS fit. From left to right and from top to bottom: $m = 25, 50, 100$ and the FDH.

The bootstrap procedure provided the following results. For the case $m = 25$, we obtain with $B = 200$ bootstrap loops:

$$\widehat{A\Sigma A} = \begin{pmatrix} 0.0671 & 0.0670 \\ 0.0670 & 0.0870 \end{pmatrix},$$

value of m	$\widehat{\alpha}^m$	$\widehat{\beta}^m$
$m = 25$	-0.1063	0.5301
$m = 50$	-0.0702	0.5389
$m = 100$	-0.0493	0.5452
FDH ($m = \infty$)	-0.0373	0.5495
Shifted-OLS	0.1045	0.5337
True full-frontier	0.0	0.5

Table 2: Results for the Example 2: $n = 100$.

providing the asymptotic normal, 95% confidence intervals:

$$\alpha^{25} \in [-0.1455, -0.0673]$$
$$\beta^{25} \in [0.4926, 0.5674].$$

The 95% percentile bootstrap confidence limits are:

$$\alpha^{25} \in [-0.1395, -0.0670]$$
$$\beta^{25} \in [0.5195, 0.5894],$$

with the same conclusions as above for example 1. For the case $m = 100$:

$$\widehat{A\Sigma A} = \begin{pmatrix} 0.0709 & 0.0761 \\ 0.0761 & 0.0974 \end{pmatrix},$$

providing the asymptotic normal, 95% confidence intervals:

$$\alpha^{100} \in [-0.0857, -0.0118]$$
$$\beta^{100} \in [0.5047, 0.5861].$$

The 95% percentile bootstrap confidence limits are:

$$\alpha^{100} \in [-0.0856, -0.0098]$$
$$\beta^{100} \in [0.5375, 0.6094],$$

which are, in this case, very similar to the intervals obtained above.

4.3 Example 3

This is the same data set as in the preceding example but we add three outliers. The results for a sample of size $n = 100$ are displayed on Figure 3.

In Table 3 we give the corresponding estimates of the intercept α and β of the corrsponding functions in the log-linear model. The order-m frontiers estimators are more robust to the three outliers than the OLS-shifted frontier, in particular, the estimation of α is very influenced by the most extreme point. This is particularly bad when the quantity of interest is the efficiencies of the firms.

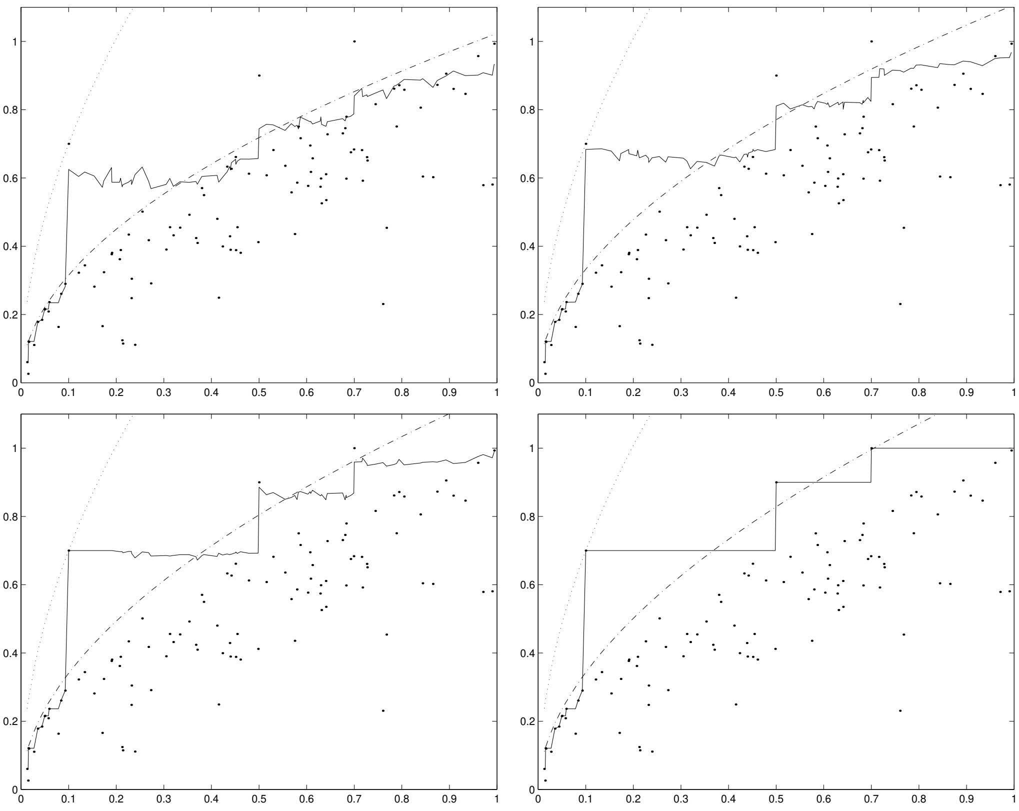

Figure 3: Results for the Example 3: $n = 103$, the 3 outliers included. In solid line, the order-$m$frontier, in dash-dotted, our linear approximation of the order-m frontier. In dotted line, the shifted OLS fit. From left to right and from top to bottom: $m = 25, 50, 100$ and the FDH.

4.4 Example 4

This is a similar scenario as in Example 2, but here, u is not independent of x. So we have the model:

$$y = x^{0.5}\, e^{-u},$$

where x is uniform on $[0, 1]$ and $u|x$ is Exponential with parameter $\lambda(x) = \frac{3}{2(1-x)}$. Here $E(u|x) = \frac{2}{3}(1 - x)$ and $E(u) = 1/3$, as above in Example 2. Here we don't have an explicit

value of m	$\widehat{\alpha}^m$	$\widehat{\beta}^m$
$m = 25$	0.0232	0.5111
$m = 50$	0.1016	0.5224
$m = 100$	0.1525	0.5341
FDH ($m = \infty$)	0.1912	0.5476
Shifted-OLS	0.8553	0.5264
True full-frontier	0.0	0.5

Table 3: Results for the Example 3: $n = 103$, including 3 outliers.

analytical expression for $\varphi_m(x)$. The results for a sample of size $n = 100$ are displayed on Figure 4.

In Table 4 we give the corresponding estimates of the intercept α and β of the corrsponding functions in the log-linear model. We remark again, in Figure 4, that the order-100 frontier achieves almost the FDH estimates. Of course, in this example, the shifted OLS is a catastrophy.

value of m	$\widehat{\alpha}^m$	$\widehat{\beta}^m$
$m = 25$	-0.0505	0.5815
$m = 50$	-0.0227	0.5901
$m = 100$	-0.0063	0.5964
FDH ($m = \infty$)	0.0016	0.5999
Shifted-OLS	0.8696	0.7204
True full-frontier	0.0	0.5

Table 4: Results for the Example 4: $n = 100$.

5 Conclusion

In this paper we have proposed a way for approximating nonparametric frontier models by some parametric models. The method is based on a two-step procedure: we approximate by the appropriate model the nonparametric estimation of the frontier. We investigate the statistical properties of the obtained estimators, we prove its consistency and provide asymptotic sampling distributions when available. For practical purposes, the bootstrap can help to estimate the sampling distributions of the desired estimator. The procedure is illustrated through numerical examples which show the advantages of our method compared with the traditional regression-type estimators.

In our approach, we have chosen to use the order-m frontier concept to derive nice

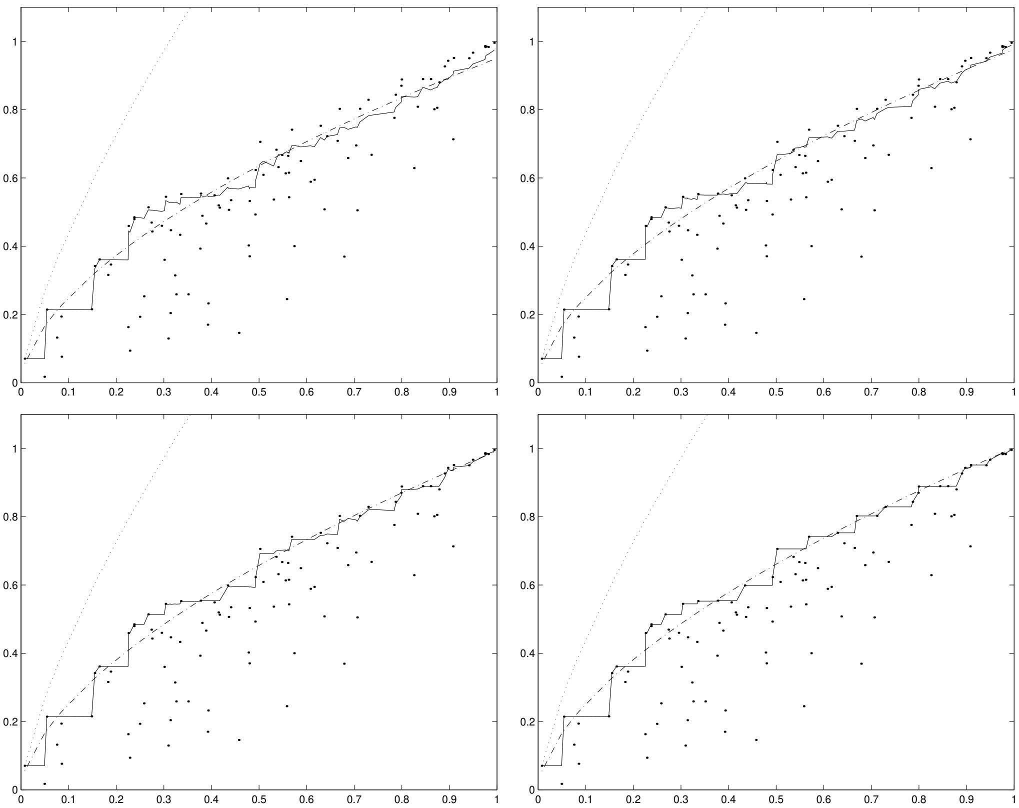

Figure 4: Results for the Example 4: $n = 100$. In solid line, the order-$m$frontier, in dash-dotted, our linear approximation of the order-m frontier. In dotted line, the shifted OLS fit. From left to right and from top to bottom: $m = 25, 50, 100$ and the FDH.

theoretical properties of our estimators. When using a full frontier estimate (FDH) as a first step, we only prove the consistency of our estimator. It is still an open issue on how to derive asymptotic properties of the estimator in the latter case: it is not yet clear what are the properties of the process $\widehat{\varphi}$, considered as a functional. So most of our results are related to the order-m frontier approach which has also the superiority of being more robust to outliers and/or extreme values than the FDH.

In this paper, we have not addressed the issue of testing a parametric model for the order-m frontier, like

$$H_0 \ : \ \varphi^m \in \{\varphi_\theta \,|\, \theta \in \Theta\}.$$

This problem could be approached in several ways:

- we could use a generalized test of Hansen, by computing $\int \left(\widehat{\varphi}^m(x) - \varphi_{\widehat{\theta}^m}\right)^2 w(x)\, dx$. The distribution of this statistics under the null hypothesis is rather complex and has been derived only under the optimal weighting function (see Carrasco and Florens, 2000).

- we could use a Hausmann-type test, based on the comparison of the two estimators obtained with two different weight functions. We can then verify, that the difference between the two estimators converges to a normal distribution, centered at the rate $\sqrt{n}$ with a variance equal to the difference of the two variances, if one of the two estimators is obtained with the optimal weighting function.

The testing issue is thus not so easy to solve and could be the topic of further research.

Appendix

A Some Asymptotic Theory

In this appendix, we describe some convergence properties of the estimators defined in this paper. The presentation is general and we tried, for sake of readability, to avoid unnecessary mathematical complexities. In particular, in this appendix, we slightly adapt the notation to improve the readability: here a parametric model will be denoted by φ_θ, whereas φ and φ^m will be the true frontier and the true expected frontier of order-m, respectively, as they were defined in Section 2. We will denote by $\widehat{\varphi}$ and $\widehat{\varphi}^m$ the corresponding plug-in estimators also defined in Section 2.

Consider now $\{\varphi_\theta \,|\, \theta \in \Theta\}$ a parametric family of functions defined on $I\!\!R_+^p$. We define the following "pseudo-true" values of θ:

$$\theta^m(w) = \arg\min_\theta \int \left(\varphi^m(x) - \varphi_\theta(x)\right)^2 w(x)\, dx \tag{A.1}$$

$$\theta(w) = \arg\min_\theta \int \left(\varphi(x) - \varphi_\theta(x)\right)^2 w(x)\, dx, \tag{A.2}$$

where $w(\cdot)$ is a given weight function. In a certain sense, the pseudo-true values of θ define the best parametric approximation of the true model (φ^m or φ respectively) in the parametric family $\{\varphi_\theta \,|\, \theta \in \Theta\}$, in the L_2 norm, according to the weight function $w(\cdot)$. The existence and unicity of these two pseudo-true values are based on technical conditions (integrability, identification, structure of the functional space $\{\varphi_\theta \,|\, \theta \in \Theta\}$). We will not expicit these technical hypothesis and assume that these pseudo-true values exist and are unique.

The weight function w can be viewed as a density on x weighting the error term. It might be natural to chose $w(\cdot) = f_X(\cdot)$, the density of the observations but since f_X is usually unknown, we can then define the "finite sample pseudo-true" values, by using the empirical discrete density $\hat{f}_n$, putting a mass $1/n$ at each observed $x_i, i = 1, \ldots, n$. So, we define:

$$\theta^m(\hat{f}_n) = \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \left(\varphi^m(x_i) - \varphi_\theta(x_i)\right)^2 \tag{A.3}$$

$$\theta(\hat{f}_n) = \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \left(\varphi(x_i) - \varphi_\theta(x_i)\right)^2 . \tag{A.4}$$

which can be viewed as conditional pseudo-true values, conditional on the observed values of the exogeneous variables x.

Now we can introduce the estimators of θ by plugging the nonparametric estimators of φ^m and of φ in the above formulae. We define:

$$\widehat{\theta}^m(w) = \arg\min_\theta \int \left(\widehat{\varphi}^m(x) - \varphi_\theta(x)\right)^2 w(x)\, dx \tag{A.5}$$

$$\widehat{\theta}^m(\hat{f}_n) = \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \left(\widehat{\varphi}^m(x_i) - \varphi_\theta(x_i)\right)^2 \tag{A.6}$$

$$\widehat{\theta}(w) = \arg\min_\theta \int \left(\widehat{\varphi}(x) - \varphi_\theta(x)\right)^2 w(x)\,dx \tag{A.7}$$

$$\widehat{\theta}(\hat{f}_n) = \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \left(\widehat{\varphi}(x_i) - \varphi_\theta(x_i)\right)^2. \tag{A.8}$$

Here again, we suppose the existence and unicity of these estimators.

In this appendix, we analyze the convergence properties of these estimators. We start with the properties related to the m-frontiers: Theorem A.1 concerns the consistency properties in the estimation of θ^m and Theorem A.2, asymptotic normality. The convergence of $\widehat{\theta}$, related to the true frontier will be addressed in Theorem A.3.

A.1 Estimation of θ^m

Theorem A.1 *For m fixed, we have, under regularity conditions, when $n \to \infty$,*

$$\widehat{\theta}^m(w) \xrightarrow{p} \theta^m(w) \tag{A.9}$$

$$\widehat{\theta}^m(\hat{f}_n) \xrightarrow{p} \theta^m(f_X) \tag{A.10}$$

$$\theta^m(\hat{f}_n) \xrightarrow{p} \theta^m(f_X). \tag{A.11}$$

Proof: From the functional convergence theorem of $\widehat{\varphi}^m$ to φ^m (Cazals, Florens and Simar, 2002, Appendix B), we obtain the convergence in probability of $\int \left(\widehat{\varphi}^m(x) - \varphi^m(x)\right)^2 w(x)\,dx$ to zero. Therefore $\int \left(\widehat{\varphi}^m(x) - \varphi_\theta(x) + \varphi_\theta(x) - \varphi^m(x)\right)^2 w(x)\,dx \xrightarrow{p} 0$, uniformly in θ. In other words, $\int \left(\widehat{\varphi}^m(x) - \varphi_\theta(x)\right)^2 w(x)\,dx \xrightarrow{p} \int \left(\varphi_\theta(x) - \varphi^m(x)\right)^2 w(x)\,dx$, uniformly in θ. which implies $\widehat{\theta}^m(w) \xrightarrow{p} \theta^m(w)$.

The convergence of $\widehat{\theta}^m(\hat{f}_n)$ to $\theta^m(f_X)$ can be proven by the same way, when we notice that $\frac{1}{n}\sum_{i=1}^n \left(\widehat{\varphi}^m(x_i) - \varphi^m(x_i)\right)^2 = \int \left(\widehat{\varphi}^m(x) - \varphi^m(x)\right)^2 f_X(x)\,dx + o_p(1)$.

Finally, $\theta^m(\hat{f}_n) \xrightarrow{p} \theta^m(f_X)$ is a direct consequence of the law of large numbers.■

The asymptotic normality of the estimators of the pseudo-true values of θ is obtained in the following theorem.

Theorem A.2 *For m fixed, we have, under regularity conditions, when $n \to \infty$,*

$$\sqrt{n}(\widehat{\theta}^m(w) - \theta^m(w)) \to N(0, \Sigma_1) \tag{A.12}$$

$$\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(\hat{f}_n)) \to N(0, \Sigma_2) \tag{A.13}$$

$$\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(f_X)) \to N(0, \Sigma_3), \tag{A.14}$$

where the structures of $\Sigma_i, i = 1, 3$ is described in the proof.

Proof: Consider the first order conditions which define $\widehat{\theta}^m(w)$, we have:

$$\int \left(\widehat{\varphi}^m(x) - \varphi_{\widehat{\theta}^m(w)}(x)\right) \frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\widehat{\theta}^m(w)} w(x)\, dx = 0.$$

Hence we obtain:

$$\begin{aligned}
&\sqrt{n}\int \frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\widehat{\theta}^m(w)} (\widehat{\varphi}^m(x) - \varphi^m(x))w(x)\, dx \\
&+\sqrt{n}\int \frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\widehat{\theta}^m(w)} (\varphi^m(x) - \varphi_{\theta^m(w)}(x))w(x)\, dx \\
&= \sqrt{n}\int \frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\widehat{\theta}^m(w)} (\varphi_{\widehat{\theta}^m(w)}(x) - \varphi_{\theta^m(w)}(x))w(x)\, dx.
\end{aligned}$$

In this equality, $\frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\widehat{\theta}^m(w)}$ can be replaced by $\frac{\partial \varphi_\theta}{\partial \theta}(x)\Big|_{\theta=\theta^m(w)}$ and we note that the second term on the left hand side is then equal to zero, by the definition of $\theta^m(w)$. Now by linearizing $(\varphi_{\widehat{\theta}^m(w)}(x) - \varphi_{\theta^m(w)}(x))$ (Taylor expansion of the first order) we obtain:

$$\sqrt{n}(\widehat{\theta}^m(w) - \theta^m(w)) = \left\{\int \frac{\partial \varphi_\theta}{\partial \theta}(x)\frac{\partial \varphi_\theta}{\partial \theta'}(x)w(x)\, dx\right\}^{-1} \int \frac{\partial \varphi_\theta}{\partial \theta}(x)\, \sqrt{n}(\widehat{\varphi}^m(x) - \varphi^m(x))w(x)\, dx$$

and so, using the properties of $\widehat{\varphi}^m$ (functional convergence theorem, Cazals, Florens and Simar, 2000) we obtain:

$$\sqrt{n}(\widehat{\theta}^m(w) - \theta^m(w)) \to N(0, A^{-1}BA^{-1}),$$

where

$$\begin{aligned}
A &= \int \frac{\partial \varphi_\theta}{\partial \theta}(x)\frac{\partial \varphi_\theta}{\partial \theta'}(x)w(x)\, dx \\
B &= \int \frac{\partial \varphi_\theta}{\partial \theta}(x)\frac{\partial \varphi_\theta}{\partial \theta'}(z)\sigma(x,z)w(x)w(z)\, dx\, dz,
\end{aligned}$$

where $\sigma(x,z)$ is the asymptotic covariance between $\sqrt{n}(\widehat{\varphi}^m(x) - \varphi^m(x))$ and $\sqrt{n}(\widehat{\varphi}^m(z) - \varphi^m(z))$.

For the second part of the theorem, the convergence of $\widehat{\theta}^m(\hat{f}_n)$ to a normal distribution can be analyzed from two points of view, according we center its distribution on $\theta^m(\hat{f}_n)$ or on $\theta^m(f_X)$. We can verify, by following the same argument as in the preceding theorem, that $\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(\hat{f}_n))$ has the same asymptotic distribution as $\sqrt{n}(\widehat{\theta}^m(f_X) - \theta^m(f_X))$. So, we obtain the same limiting covariance matrix as in the preceding case where w is replaced by f_X.

By following the same kind of arguments, we can prove that the limiting law of $\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(f_X))$ is also a normal but the structure of the limiting covariance matrix is much more complicated.■

Remark 1: Optimal approximation

We have chosen here to approximate the order-m frontier by a parametric family, in the least-squares sense, weighted by the function $w(x)$. A natural question could then concerns the existence of an optimal approximation and optimal choice of $w(x)$. Two elements can be said concerning this issue.

(1) If the parametric model is well-specified, *i.e.* if $\varphi^m \in \{\varphi_\theta \,|\, \theta \in \Theta\}$, the estimated parameter converges to the true value of θ and the question about the optimality concerns the minimazation of the variance. Note that some examples in Section 4 show that this hypothesis is not absurd and in these examples, φ^m is linear for all m. However, the derivation of the optimal weighting is complex even in this case. In finite dimension, it is well known that the optimal weights must converge to Σ^{-1}, where Σ is the asymptotic variance of $\widehat{\varphi}^m - \varphi^m$, but in infinite dimension, it is much more complicated: the methodology is described in Carrasco and Florens (2002).

We could simplify the problem, and not search for the absolute optimal solution, but rather correct the weighting system only for the heteroscedasticity. We could then chose $w(x)$ proportional to the inverse of $\sigma^2(x)$, the variance of $\widehat{\varphi}^m(x) - \varphi^m(x)$, and then scaling such that $\int w(x)\,dx = 1$.

(2) If the model is not well specified, $\varphi^m \notin \{\varphi_\theta \,|\, \theta \in \Theta\}$, then the pseudo-true value $\theta^m(w)$ depends on w, and so the comparison of the variances of the resulting estimators is useless.

Remark 2: The Bootstrap

The bootstrap algorithm proposed in Section 3, is one particular case of the algorithms we propose here. Indeed, in the bootstrap world, we must "mimic" the sampling distribution of the estimator of interest which estimates the pseudo-true value of interest. Therefore 3 versions of the bootstrap have to be carefully defined depending on the problem at hand.

case 1: $\sqrt{n}(\widehat{\theta}^m(w) - \theta^m(w)) \to N(0, \Sigma_1)$. It can be approximated by the distribution of $\sqrt{n}(\widehat{\theta}^{m,*}(w) - \widehat{\theta}^m(w)) \,|\, \mathcal{X}$.

case 2: $\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(\hat{f}_n)) \to N(0, \Sigma_2)$. It can be approximated by the distribution of $\sqrt{n}(\widehat{\theta}^{m,*}(\hat{f}_n^*) - \widehat{\theta}^m(\hat{f}_n^*)) \,|\, \mathcal{X}$.

case 3: $\sqrt{n}(\widehat{\theta}^m(\hat{f}_n) - \theta^m(f_X)) \to N(0, \Sigma_3)$. It can be approximated by the distribution of $\sqrt{n}(\widehat{\theta}^{m,*}(\hat{f}_n^*) - \widehat{\theta}^m(\hat{f}_n)) \,|\, \mathcal{X}$.

The notation is very explicit and does not involve any ambiguity. For instance, denoting $\mathcal{X}^* = \{(x_i^*, y_i^*) | i = 1, \ldots, n\}$ a bootstrap sample, resampled with replacement from $\mathcal{X}$, we have

$$\widehat{\theta}^{m,*}(w) = \arg\min_{\theta} \int \left(\widehat{\varphi}^{m,*}(x) - \varphi_\theta(x)\right)^2 w(x)\, dx \tag{A.15}$$

$$\widehat{\theta}^{m}(\hat{f}_n^*)) = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\varphi}^{m}(x_i^*) - \varphi_\theta(x_i^*)\right)^2 \tag{A.16}$$

$$\widehat{\theta}^{m,*}(\hat{f}_n^*) = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\varphi}^{m,*}(x_i^*) - \varphi_\theta(x_i^*)\right)^2 \tag{A.17}$$

where for any $z \in I\!R^p$, $\widehat{\varphi}^{m,*}(z)$ is the nonparametric estimation of $\varphi^m(z)$ obtained from the bootstrap sample $\mathcal{X}^*$. The appropriate bootstrap algorithm can be used, in particular, to get an estimation of the covariance matrices $\Sigma_i,\ i = 1, 2, 3$.

In each case, the bootstrap algorithm follows the ideas of the algorithm described in Section 3. There, we were interested in **case 2**, where the $\widehat{\theta}^{m,*}(\hat{f}_n^*)$ and $\widehat{\theta}^{m}(\hat{f}_n^*)$ had explicit expressions:

$$\begin{aligned}
\widehat{\theta}_n^{m,*} &= \widehat{\theta}^{m,*}(\hat{f}_n^*) = \left\{\widehat{E}^*[g(x)g'(x)]\right\}^{-1} \widehat{E}^* \left[g(x)\widehat{\varphi}^{m,*}(x)\right] \\
\theta^{m,*} &= \widehat{\theta}^{m}(\hat{f}_n^*) = \left\{\widehat{E}^*[g(x)g'(x)]\right\}^{-1} \widehat{E}^* \left[g(x)\widehat{\varphi}^{m}(x)\right],
\end{aligned}$$

where $\widehat{E}^*$ is the empirical average with respect to the bootstrap sample $\mathcal{X}^*$.

A.2 Estimation of θ

The analysis of the convergence of $\widehat{\theta}(w)$ to $\theta(w)$ and of $\widehat{\theta}(\hat{f}_n)$ to $\theta(f_X)$ is more complex. We start with a preliminary lemma, concerning uniform convergence of the FDH estimator $\widehat{\varphi}$ to φ.

Lemma A.1 *Under regularity conditions, $\widehat{\varphi}$ converges uniformly to φ in the following sense:*

$$\sup_x |\widehat{\varphi}(x) - \varphi(x)| \to 0, \ \textit{as } n \to \infty \tag{A.18}$$

Proof: This lemma results from different elements:

1. φ^m converges uniformly to φ as $m \to \infty$. Indeed the sequence $(\varphi^m)_m$ is an increasing sequence of continuous functions with, for all x, $\varphi^m(x) \to \varphi(x)$ continuous. By Dini's theorem (see *e.g.*, Schwartz, 1991, Chapter 13), we have the uniform convergence.

2. $\widehat{\varphi}^m$ converges uniformly to φ^m (see Appendix B of Cazals, Florens and Simar, 2002) and the FDH estimator, $\widehat{\varphi}$, satisfies

$$\widehat{\varphi}^m(x) \leq \widehat{\varphi}(x) \leq \varphi(x), \ \text{for all } x. \tag{A.19}$$

3. Hence we have,

$$\forall \varepsilon > 0, \exists m_\varepsilon \text{ such that } \sup_x |\varphi(x) - \varphi^{m_\varepsilon}(x)| < \frac{\varepsilon}{2}, \text{ and}$$
$$\exists N_\varepsilon \text{ such that for } n \geq N_\varepsilon, \sup_x |\widehat{\varphi}^{m_\varepsilon}(x) - \varphi^{m}(x)| < \frac{\varepsilon}{2}.$$

Therefore

$$\forall \varepsilon > 0, \exists m_\varepsilon, N_\varepsilon \text{ such that } \sup_x |\widehat{\varphi}^{m_\varepsilon}(x) - \varphi(x)| < \varepsilon,$$

which implies by (A.19), $\sup_x |\widehat{\varphi}(x) - \varphi(x)| < \varepsilon$.

This completes the proof of the lemma.■

This lemma implies:

$$1) \quad \int \left(\widehat{\varphi}(x) - \varphi(x)\right)^2 w(x)\, dx \xrightarrow{p} 0 \tag{A.20}$$

$$2) \quad \frac{1}{n}\left(\widehat{\varphi}(x_i) - \varphi(x_i)\right)^2 = \int \left(\widehat{\varphi}(x) - \varphi(x)\right)^2 f_X(x)\, dx + o_p(1) \xrightarrow{p} 0. \tag{A.21}$$

Now we can state our result:

Theorem A.3 $\widehat{\theta}(w) \xrightarrow{p} \theta(w)$ *and* $\widehat{\theta}(\hat{f}_n) \xrightarrow{p} \theta(f_X)$. *In particular, if the parametric model* $\{\varphi_\theta \,|\, \theta \in \Theta\}$ *is correctly specified for the frontier,* $\widehat{\theta}(w)$ *and* $\widehat{\theta}(\hat{f}_n)$ *converge in probability to the* θ *such that* $\varphi = \varphi_\theta$.

Proof: The proof is identical to that of Theorem A.1, using the properties (A.20) and (A.21).■

At the best of our knowledge, no functional convergence theorem exists for the FDH estimator $\widehat{\varphi}$. In Park, Simar and Weiner (2000), only pointwize distributional convergence is derived. So, the asymptotic distributional properties of $\widehat{\theta}(w)$ and/or of $\widehat{\theta}(\hat{f}_n)$ cannot be derived so easely. This could be the topic of forthcoming works.

References

[1] Aigner, D.J. and S.F. Chu (1968), On estimating the industry production function, *American Economic Review*, 58, 826–839.

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

[3] Carrasco, M. and .P. Florens (2000), Generalization of GMM to a continuum of moment conditions, *Econometric Theory*, Vol 16, 797–834.

[4] Cazals, C., Florens, J.P. and L. Simar (2002), Nonparametric frontier estimation: a robust approach, *Journal of Econometrics*, 106, 1–25.

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

[6] Deprins, D. and L. Simar (1985), A Note on the asymptotic relative efficiency of the M.L.E. in a linear model with Gamma disturbances, *Journal of Econometrics*, 27, 383-386.

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

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

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

[10] Hall, P., and L. Simar (2002), Estimating a change point, boundary or frontier in the presence of observation errors, *Journal of the American Statistical Association*, 97,523–534.

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

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

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

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

[15] Schwartz, L. (1991), *Analyse I*, Hermann, éditeurs des sciences et des arts, Paris.

[16] Simar, L. (1992), Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping, *Journal of Productivity Analysis*, 3, 167–203.

[17] 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.

[18] Thiry, B. and H. Tulkens (1992), Allowing for technical inefficiency in parametric estimation of production functions for urban transit firms, *Journal of Productivity Analysis*, 3, 45–65.