ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 82 – 2006

Efficiency Measurement: a Nonparametric Approach

Yves ARAGON, Abdelaati DAOUIA, Christine THOMAS-AGNAN*

ABSTRACT. – The aim of the paper is to present a statistical methodology allowing a meaningful comparison of the production performance of firms without resorting to the usual concept of production frontier. We introduce an efficiency measure based on a nonstandard conditional distribution and propose a two-stage estimation procedure with a smoothing step followed by an isotonization step. We illustrate the approach through a simulated example and an analysis of the performance of Spanish electricity distributors.

Mesure d'efficacité : une approche non paramétrique

RÉSUMÉ. – L'objet de cet article est de présenter une méthodologie statistique permettant de mesurer la performance de production des firmes sans recourir au concept usuel de la frontière de production. Notre mesure d'efficacité est basée sur une loi de probabilité conditionnelle non-standard. La construction d'un estimateur non-paramétrique de cette mesure résulte d'une étape de lissage suivie d'une étape de monotonisation. On fournit une illustration par des données simulées et des données réelles sur la distribution de l'électricité en Espagne.

* Y. ARAGON, A. DAOUIA, C. THOMAS-AGNAN : GREMAQ, Manufacture des Tabacs, 21 allée de Brienne, 31042 Toulouse Cedex, France. Tel. : +33(0)5-61-12-85-48 ; Fax : +33(0)5-61-22-55-63 ; e-mail : aragon@cict.fr, daouia@cict.fr, cthomas@cict.fr Key words: Efficiency Measurement Using a Conditional Distribution, Integrated Kernel Estimators; Isotonization.

1 Introduction

The performance of production units in terms of the conventional efficiencies has received widespread attention in the economics, statistics and related literatures. KOOPMANS [1951] and DEBREU [1951] introduced the basic concept of a production set $\Psi = \{(x,y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}$, a set of technically possible pairs of input $x \in \mathbb{R}_+^p$ and output $y \in \mathbb{R}_+^q$. Associated with this set are the production frontier, *i.e.*, the set of the most efficient production units, and the efficiency of a unit, *i.e.*, the distance of its output to the production frontier in the output-orientation (resp.: the distance of its input to the production frontier in the input-orientation). FARRELL [1957] is the first empirical work where the problem of measuring efficiencies for a set of observed production units is analyzed. In the output-orientation, the Farrell technical efficiency of a production unit working at a level (x,y) is defined as $\lambda(x,y) = \sup\{\lambda \mid (x,\lambda y) \in \Psi\}$. $\lambda(x,y) \geq 1$ gives the radial feasible proportionate increase in outputs the unit (x,y) should perform to be considered as "output technically efficient". Likewise, in the input-orientation, the Farrell efficiency score is defined as $\theta(x,y) = \inf\{\theta \mid (\theta x,y) \in \Psi\}$. Here $\theta(x,y) \leq 1$ measures the radial feasible proportionate reduction of inputs the unit (x,y) should perform to be considered as "input technically efficient".

Both Farrell efficiencies are radial measures of the distance between the point of interest (x,y) and the production frontier (*i.e.*, the topological upper surface of Ψ). See, *e.g.*, SHEPHARD [1970] for a modern economic formulation of the problem. The conventional Shephard input and output distance functions are merely the reciprocals of the corresponding Farrell measures. The econometric problem is thus the estimation of Ψ or equivalently its frontier from a random sample of production units $\mathcal{X}_n = \{(X_i,Y_i) \mid i = 1,\cdots,n\}$. Two different methods have been mainly developed: the stochastic frontier approach where random noise allows some observations to be outside of the attainable set Ψ, and the deterministic frontier approach which supposes that Ψ contains all the observations with probability one.

In the stochastic frontier models, where noise is allowed, the identification of the frontier from the sample is only possible with additional parametric restrictions on the shape of the frontier function. Most of the available techniques used to estimate the parameters of the frontier function are based on the maximum likelihood methods, in the spirit of the work of AIGNER, LOVELL and SCHMIDT [1977] and MEEUSEN and VAN DEN BROEK [1977]. These methods may lack robustness if the assumed distributional form of the error term does not hold. In particular, outliers in the data may unduly affect the estimate of the production frontier. Furthermore, this estimate may be biased if the error structure is not correctly specified, or, if heteroscedasticity in this structure is not properly accounted for. Biased or non-robust estimation of the input or output-oriented production frontier may result in poor estimation of the corresponding technical efficiency. For more discussions about drawbacks of the stochastic frontier approach, we refer to KOKIC, CHAMBERS, BRECKLING and BEARE [1997]. See also KUMBHAKAR and LOVELL [2000] for a nice summary of the theory of stochastic frontier models.

The deterministic frontier approach, by contrast, is popular because it avoids having to specify a particular functional relationship to be estimated; the data are allowed to speak for themselves. It involves the nonparametric estimation of Ψ, or equivalently its frontier which is referred to as production frontier, by either the free disposal hull (FDH) of the sample $\mathcal{X}_n$ or the data envelopment analysis (DEA). The FDH estimator of Ψ was introduced by Deprins, Simar and Tulkens [1984]. It only relies on the free disposability assumption on Ψ. The DEA estimator is defined as the convex hull of the FDH estimator. It requires stronger assumptions: in addition to the free disposability, it relies on a convexity assumption for Ψ. The method is due to Farrell [1957] and has been made operational in a linear programming setting by Charnes, Cooper and Rhodes [1978]. The FDH and DEA estimators of the Farrell-Shephard technical efficiencies are obtained by plugging the empirical analog of Ψ into the appropriate formulae. These nonparametric estimators are very appealing because they rely on few assumptions and their statistical theory is mostly available (see Simar and Wilson, [2000], for a nice survey). However, in the presence of outliers, they could behave dramatically since, by construction, they are very sensitive to extreme values. A more serious problem is that these envelopment estimators suffer from the curse of dimensionality.

Recently, robust nonparametric estimators have been suggested by Cazals, Florens and Simar [2002] and Aragon, Daouia and Thomas-Agnan [2005], when either inputs or outputs are univariate. In this case, Farrell-Shephard efficiencies estimates can be obtained through direct estimation of the production frontier. But, in place of estimating the full frontier of Ψ, Cazals *et al.* [2002] propose rather to estimate an expected partial frontier of order *m*, a positive integer. Similarly, Aragon *et al.* [2005] build a new estimator of the efficient frontier, replacing the concept of "discrete" order-*m* partial frontier by a "continuous" order-α partial frontier where $\alpha \in [0, 1]$ corresponds to the level of a nonstandard conditional quantile function. Both nonparametric order-*m* and order-α frontiers of Ψ do not suffer from the curse of dimensionality. However, as demonstrated in Daouia and Ruiz-Gazen [2004], the quantile-based frontiers have better robustness properties than the standard (FDH, DEA) estimators and than the order-*m* estimators of Cazals *et al.* [2002]. They also possess remarkable asymptotic properties as established in Daouia [2005]. In our present work, we show under quite reasonable conditions from an economic point of view, that the nonstandard conditional distribution function (df) associated to the quintile-type frontiers, defines itself a good efficiency measure even in the full multivariate case with multiple inputs and multiple outputs.

The paper is organized as follows. In section 2 we motivate our approach and in section 3 we propose a two-stage estimation procedure with a smoothing step followed by an isotonization step. The computation algorithm and the bandwidth selection rules are discussed in Section 4. In section 5, a numerical illustration is proposed with simulated and real data. Section 6 concludes the paper. The lemmas and proofs of all theoretical results in the sequel can be found in the Appendix.

2 The Production Performance Measure

Let us consider a random vector $(X, Y) \in \mathbb{R}_+^p \times \mathbb{R}_+^q$ defined on the probability space (Ω, $\mathcal{A}$, *P*). In this approach, we define the production set Ψ to be the support of the joint distribution of (X,Y), where X represents the inputs and Y represents the outputs. Such a joint distribution is usually decomposed into a marginal distribution of X and a conditional distribution of Y given $X = x$. In this paper, we will rather concentrate on another characterization of this joint probability measure by considering the conditional probability measure of Y given $X \leq x$. If the joint probability measure is characterized by the joint df $F(x,y) = P(X \leq x, Y \leq y)$, the conditional distribution of Y given $X \leq x$ may be described by its df $F(y \mid x) = F(x,y)/F_X(x)$, where F_X denotes the marginal df of X. Note that if $(x,y) \in \Psi$ and $F_X(x) = 0$, then (x,y) may not be viewed as an effective production unit. In this work, we only focus on the production set $\Psi^* = \{(x,y) \in \Psi \mid F_X(x) > 0\}$.

In the univariate-output case where $q = 1$ and $p \geq 1$, the nonstandard conditional df $F(y \mid x)$ can be characterized by the family of quantile curves $q_\alpha(x) = \inf\{y \geq 0 \mid F(y \mid x) \geq \alpha\}$, where α ranges in [0, 1]. As pointed out in ARAGON *et al.* [2005], $q_\alpha(x)$ is the production threshold exceeded by $100(1-\alpha)\%$ of firms that use less than the level x of inputs. If the quantile curve of order α passes through a production unit $(x,y) \in \Psi^*$, *i.e.*, $y = q_\alpha(x)$, then this unit produces more than $100\alpha\%$ of all production units using inputs smaller than or equal to x, and produces less than the $100(1-\alpha)\%$ remaining units. Thus the quantile function $q_\alpha(x)$, or equivalently its related conditional df $F(y \mid x)$, quantifies the production efficiency of the unit working at (x,y) by comparing it with all units which use the same level of inputs x as well as with those which use strictly less than x. It is then natural to set the performance measure for a production unit $(x,y) \in \Psi^*$, to be the order $\alpha(x,y)$ of the quantile curve which passes through this unit, or equivalently:

$$\alpha(x,y) = F(y \mid x). \tag{1}$$

The efficiency measure (1) is defined as the conditional df itself. This simple formulation allows us to take into account the full multivariate case (multi-inputs and multi-outputs). It does not involve quantiles and therefore avoids the problem of multiple definitions of multivariate quantiles. The natural question which is then raised is: what are the properties that a good efficiency measure $\alpha(x,y)$ should possess? First of all, each unit $(x,y) \in \Psi^*$ must have a unique measure $\alpha(x,y)$, which is satisfied by (1). The second natural property from an economic point of view would be the strict monotonicity[1] of $\alpha(x,y)$ with respect to both inputs and outputs. It is clear that (1) is increasing with respect to outputs $y \in Y(x)$ if and only if the df $F(\cdot \mid x)$ is increasing on its support, for any x such that $F_X(x) > 0$. It is also decreasing with respect to inputs $x \in X(y)$ if and only if the function $x \mapsto F(y \mid x)$ is decreasing on $\{x \mid 0 < F_X(x) < 1\}$ for every y such that $0 < F_Y(y) < 1$. This neces-

1. A real valued function G on $\mathbb{R}^k$ is said to be nondecreasing (resp.: nonincreasing) with respect to the partial order induced on $\mathbb{R}^k$ by "$\leq$" (understood componentwise) if $u \leq v$ implies $G(u) \leq G(v)$ (resp.: $G(u) \geq G(v)$). We will say that G is decreasing (resp.: increasing), if $u \leq v$ and $u \neq v$ implies $G(u) > G(v)$ (resp.: $G(u) < G(v)$).

sary and sufficient condition is quite reasonable in view of the fact that the chance of producing less than a value *y* decreases if a firm uses more inputs.

Additional requirements for an efficiency measure to be acceptable in the case of one output and multiple inputs, are argued by Kokic *et al.* [1997]: the measure should lie between 0 and 1; the poorest performing firms should have measures close to 0; the best performing firms should have measures close to 1; the distribution of the measure should not depend on the level of inputs of the firm.

Since $\alpha(x,y)$ represents the percentage of units which produce less than the value *y* among all units using inputs smaller than or equal to the level *x*, if $\alpha(x,y)$ is close to 1, then the unit (x,y) can be seen to be performing relatively efficiently, and likewise, if $\alpha(x,y)$ is close to 0, then the unit (x,y) would be performing relatively inefficiently. Thus, the first three criteria of Kokic *et al.* [1997] are satisfied by the measure (1). To check the last one, we need in addition to the strict monotonicity of $F(\cdot\mid x)$ that the joint df $F(\cdot,\cdot)$ possesses a density. Indeed, it can be easily seen under these conditions that:

$$P\big(\alpha\,(X,\ Y)\le u\big)=u,\quad \text{for any } u\in[0,1].$$

As a corollary, our efficiency measure is uniform on the interval [0, 1], which is quite reasonable in practice. Indeed, if the measure would for instance happen to concentrate close to 1, no firm would appear to be relatively inefficient.

The statistical problem is now to find a consistent order-preserving estimator of the nonstandard conditional df $\alpha(x,y)$ from the random sample $\mathcal{X}_n$, which will be monotone with respect to both inputs and outputs.

3 Nonparametric Estimation

Smoothing step: Since the distribution of *Y* is conditioned by $X\le x$, the estimate of $\alpha(x,y)=F(y\mid x)$ does not require a smoothing procedure which would be the case if the distribution of *Y* were conditioned by $X=x$. To estimate $F(y\mid x)$ we could simply use its empirical version $\alpha_n(x,y)=\sum_{i=1}^{n}1\!\!1(X_i\le x, Y_i\le y)/\sum_{i=1}^{n}1\!\!1(X_i\le x)$, with $\alpha_n(x,y)=0$ if the denominator is 0. However, as a function of *y* (resp.: of *x*), $\alpha_n(x,y)$ is piecewise constant. From an economic viewpoint it does not make sense to consider that two firms using the same level of inputs, for example, have the same efficiency when they do not produce the same quantity of outputs. To overcome this deficiency, we propose in a first step to smooth $\alpha_n(x,y)$ with respect to *x* and *y*, replacing the indicator functions by integrated kernels. This technique produces increasing estimators of the conditional df $F(\cdot\mid x)$ that always lie between 0 and 1.

Several recent order-preserving methods for the nonparametric estimation of the usual conditional df of Y given $X = x$, can be adapted to our nonstandard conditioning framework. They include the identity-reproducing regression or mass-centered smoothing techniques, discussed by MULLER and SONG [1993] and MAMMEN and MARRON [1997]; the biased bootstrap form of the Nadaraya-Watson estimator (HALL and PERSNELL [1999]); and some data-sharpening techniques (CHOI, HALL and ROUSSON [2000]). However, all these methods suffer from excessive bias at the boundaries (see HALL and MULLER [2003], for more details). An alternative technique which is more attractive from the viewpoint of mathematical efficiency, would be to use local linear methods (see, *e.g.*, FAN [1992]; FAN and GIJBELS [1992]; YU and JONES [1998]). But those methods, much as discussed, *e.g.*, by HALL, WOLFF and YAO [1999] and PERACCHI [2002] in the context of standard conditional df estimation, have the disadvantage of producing df estimators that are not constrained either to lie between 0 and 1 or to be monotone. HALL and MULLER [2003] showed that with probability tending to 1, local linear methods fail to be order-preserving.

For our smoothing step we consider product kernels $K_0(x) = \prod_{i=1}^{p} k_0^i\,(x^i)$ and $K_1(y) = \prod_{j=1}^{q} k_1^j\,(y^j)$ for $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$, with k_0^i and k_1^j being probability densities on $\mathbb{R}$. Given bandwidths h_0 and h_1 and integrated kernels $H_0(x) = \int_{-\infty}^{x} K_0(u)du$ and $H_1(y) = \int_{-\infty}^{y} K_1(u)du$, we define the following smoothed version of $\alpha_n(x,y)$:

$$(2) \qquad \widehat{\alpha}_n(x,y) = \frac{(1/n)\sum_{i=1}^{n} H_0\left(\frac{x-X_i}{h_0}\right) H_1\left(\frac{y-Y_i}{h_1}\right)}{(1/n)\sum_{i=1}^{n} H_0\left(\frac{x-X_j}{h_0}\right)} := \frac{\widehat{F}(x,y)}{\widehat{F}_X(x)}$$

with $\widehat{\alpha}_n(x,y) = 0$ if $\widehat{F}_X(x) = 0$. Using standard theory, we easily show that if the marginals of X and Y are continuous and $h_i \to 0$ as $n \to \infty$, for $i = 0, 1$, then $\widehat{\alpha}_n(x, y)$ converges completely to $\alpha(x, y)$. Under more stringent conditions on the joint df F, on the univariate kernels k_0^i, k_1^j and the bandwidths h_0, h_1, we show:

$$\sqrt{n}\left(\widehat{\alpha}_n(x, y) - \alpha(x, y)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\,(0, \sigma^2(x,y)), \quad n \to \infty,$$

for any $(x, y) \in \Psi$ such that $0 < F(x, y) < F_X(x)$, where $\sigma^2(x, y) = [F_X(x) - F(x, y)]F(x, y)/F_X^3(x)$. Replacing $F_X(x)$ and $F(x,y)$ by their empirical versions in $\sigma^2(x,y)$, we obtain a strongly consistent estimator for this unknown asymptotic variance that can be used to construct confidence intervals for $\alpha(x,y)$. In the next theorem, we derive the limiting process of $\sqrt{n}\,(\widehat{\alpha}_n - \alpha)$ when viewed as a process indexed by (x,y). To this end, we need the following assumptions:

- (H1) The joint df F is continuously differentiable on $\mathbb{R}^{p+q}$ up to order $p+q+r$. Its derivatives are zero on the boundary of the compact support Ψ.

For simplicity, we consider the same bandwidth $h = h_0 = h_1$ and the same univariate kernel $k(\cdot) = k_0^i(\cdot) = k_1^j(\cdot)$, and we assume that h and the product kernel $K(z) = \prod_{i=1}^{p+q} k(z^i) = K_0(x)\, K_1(y)$, for $z = (x, y) \in \mathbb{R}^{p+q}$, satisfy:

• (H2) $n^{1/2}h^r \to 0$ as $n \to \infty$; $K(-z) = K(z)$ for all $z \in \mathbb{R}^{p+q}$; $\int_{-\infty}^{\infty} K(z)\, dz = 1$; K is a kernel function of order r, is differentiable up to order $p+q+r$ on $\mathbb{R}^{p+q}$, and its derivatives of order up to r are $L^2(\mathbb{R}^{p+q})$.

THEOREM 3.1. *Let* $\Psi_* \subset \Psi$ *such that* $\inf_{(x,y)\in\Psi_*} F_X(x) > 0$. *Given* (*H*1)-(*H*2), *the process* $\widehat{\mathbb{A}}_n = \sqrt{n}(\widehat{\alpha}_n - \alpha)$ *converges in distribution in the space* $L^\infty(\Psi_*)$ *of bounded functions on* Ψ_*, *to the centered Gaussian process* $\mathbb{A}$ *defined, for any* $(x,y) \in \Psi_*$, *by* $\mathbb{A}\,(x,y) = [\mathbb{F}\,(x,y) - F\,(y\,|\,x)\,\mathbb{F}\,(x,\infty)]\,/\,F_X(x)$, *where* $\mathbb{F}$ *is a* $p+q$ *dimensional F-Brownian bridge, i.e., a Gaussian process with zero mean and covariance function* $E\,(\mathbb{F}\,(t_1)\,\mathbb{F}\,(t_2)) = F\,(\min\,(t_1,\ t_2)) - F(t_1)\,F(t_2)$, *for all* $t_1, t_2 \in \overline{\mathbb{R}}^{p+q}$.

When the data $\{(X_i, Y_i)\}$ are not independent, the statements of Theorem 3.1 hold under the additional condition:

• (H3) The sequence $\{(X_i, Y_i)\}$ is a strictly stationary β-mixing sequence satisfying: $\lim_{k\to\infty} k^\delta \beta_k = 0$ for some fixed $\delta > 1$. The β-mixing coefficients $\{\beta_k\}$ are defined as:

$$\beta_k = \sup_{I,J,l} \left\{ \sum_{i=1}^{I}\sum_{j=1}^{J} |\, P(A_i \cap B_j) - P(A_i)\, P(B_j)\,|:\ \{A_i\} \text{ is a partition of } \sigma_1^l,\right.$$
$$\left. \{B_j\} \text{ is a partition of } \sigma_{k+l}^\infty \right\}$$

where σ_1^l and σ_{k+l}^∞ are the σ-fields respectively generated by $\{(X_i, Y_i)\}_{i=1,\cdots,l}$ and $\{(X_i, Y_i)\}_{i \geq k+l}$.

Indeed, following AIT-SAHALIA [1993, Theorem 1], the conditions (H1)-(H3) ensure the functional convergence in distribution of $\sqrt{n}\left(\widehat{F} - F\right)$ to $\mathbb{F}$ in $L^\infty(\Psi)$. By applying the same technique of proof as in Theorem 3.1 in conjunction with this result, we obtain the functional convergence result of Theorem 3.1. From now on, we only focus on the independent identically distributed data framework.

Monotonicity step: The quotient estimator $\widehat{\alpha}_n(x, y)$ is increasing with respect to *y* but is not necessarily nonincreasing with respect to *x*. We will construct an isotonic version of $\widehat{\alpha}_n(x, y)$, which will still be increasing with respect to the output and will only be nonincreasing with respect to the input. This isotone estimator will also be shown to converge completely and uniformly to the efficiency measure

$\alpha(x, y)$ on some given subset $D \times S$ of Ψ^*. We use the symbol $\|\cdot\|_T$ to indicate the sup-norm of a real valued function over a set *T*. For the sake of simplicity we assume throughout this paper that $D = \prod_{i=1}^{p} [a_i, b_i]$ is a subset interior to the support of *X* and that $S = \prod_{j=1}^{q} [c_j, d_j]$ is a subset of the support of *Y*. The monotone estimator is obtained from the following isotonization procedure: for a real valued function *r* defined on *D*, let us define the following three functions:

$$
\begin{aligned}
r^u(x) &= \sup_{x' \in D | x' \geq x} r(x'), \\
r^l(x) &= \inf_{x' \in D | x' \leq x} r(x'), \qquad (3)\\
r^{\#}(x) &= (r^u(x) + r^l(x)) / 2.
\end{aligned}
$$

It is clear that $r^u(x)$, $r^l(x)$ and $r^{\#}(x)$ are nonincreasing and that they satisfy the following inequality for all *x* in their domain of definition *D*,

$$r^l(x) \leq r(x) \leq r^u(x).$$

It is then natural to define our isotonic estimator of $\alpha(x, y)$, for any $(x, y) \in D \times S$, with a slight abuse of notation, by:

$$\hat{\alpha}_n^{\#}(x, y) = [\hat{\alpha}_n^l(x, y) + \hat{\alpha}_n^u(x, y)] / 2, \qquad (4)$$

where

$$\hat{\alpha}_n^l(x, y) = \inf_{x' \in D | x' \leq x} \hat{\alpha}_n(x', y), \quad \hat{\alpha}_n^u(x, y) = \sup_{x' \in D | x' \geq x} \hat{\alpha}_n(x', y). \qquad (5)$$

The basic idea of the present isotonization is not new. MUKERJEE and STERN [1994] use a similar principle to isotonize a Nadaraya-Watson kernel estimator of the regression function, and with a slight difference, which is in fact a computational artifact: the sup and inf in (3) are taken over a discrete grid instead of the whole domain *D*.

The argument that will be used to show the uniform complete convergence of our isotonic estimator $(x, y) \mapsto \hat{\alpha}_n^{\#}(x, y)$ on $D \times S$ is first based on the fact that the # operator is sup-norm contracting (Lemmas A.1-A.2), and on the uniform complete convergence of the estimators $\widehat{F}$ and $\widehat{F}_X$ (Lemmas A.3-A.4).

THEOREM 3.2. *Given the conditions of Lemmas* A.2-A.4, *we have for any* $\theta < 1/2$,

$$n^{\theta} \left\| \hat{\alpha}_n^{\#} - \alpha \right\|_{D \times S} \leq n^{\theta} \left\| \hat{\alpha}_n - \alpha \right\|_{D \times S} \xrightarrow{co.} 0 \text{ as } n \to \infty.$$

This theorem also shows that the isotonized estimator $\hat{\alpha}_n^{\#}(\cdot,\cdot)$ becomes globally more stable and closer to the true efficiency measure $\alpha(\cdot,\cdot)$ than the original estimator $\hat{\alpha}_n(\cdot,\cdot)$ as n goes to ∞. The standardizing factor, n^{θ} with $\theta < 1/2$, indicates a lower rate of convergence than $\sqrt{n}$, which is natural and optimal in a certain sense since $\sqrt{n}\left\|\hat{\alpha}_n - \alpha\right\|_{D\times S} \xrightarrow{P} \left\|\mathbb{A}\right\|_{D\times S}$ following Theorem 3.1.

Let us now turn to the analysis of the quantitative robustness of the estimators of our efficiency measure. Let $(x,y) \in \Psi^*$ be fixed and consider the statistical functional $\Phi_{x,y}$ that maps a df $G(\cdot,\cdot)$ on $\mathbb{R}_+^{p+q}$ such that $G(x,\infty) > 0$ to the real value $\Phi_{x,y}(G) = G(x,y)\,/\,G(x,\infty)$. The efficiency measure $\alpha(x,y)$ being representable as $\Phi_{x,y}(F)$, its influence function can be defined according to HAMPEL *et al.* [1985], Definition 1, p. 84, as $(x_0,y_0) \in \mathbb{R}_+^{p+q} \mapsto I\,F((x_0,y_0)\,;\Phi_{x,y},\, F) = (\partial/\partial\delta)\,\Phi_{x,y}\,(F + \delta(\Delta_{(x_0,y_0)} - F))\,|_{\delta=0+}$, where $\Delta_{(x_0,y_0)}(u,v) = 1\mathrm{I}\,(x_0 \le u,\ y_0 \le v)$ for any $(u,v) \in \mathbb{R}_+^{p+q}$. We have:

$$IF\,((x_0,y_0)\,;\Phi_{x,y},\ F) = 1\mathrm{I}\,(x_0 \le x)\,[1\mathrm{I}\,(y_0 \le y) - F(y\,|\,x)]\,/\,F_X(x).$$

The *IF* measures the asymptotic bias caused by contamination in the observations (X_i,Y_i). Indeed, it can be easily seen that $IF\,((X_i,Y_i)\,;\,\Phi_{x,y},F)$ represents the approximate influence of the observation (X_i,Y_i) on the estimation errors $\alpha_n(x,y) - \alpha\,(x,y)$, $\hat{\alpha}_n(x,y) - \alpha(x,y)$ and $\hat{\alpha}_n^{\#}(x,y) - \alpha(x,y)$, since:

$$\alpha_n(x,y) - \alpha(x,y) = \frac{1}{n}\sum_{i=1}^{n} IF\,((X_i,Y_i)\,;\Phi_{x,y},\ F) + o_p(1) \quad \text{as} \quad n \to \infty.$$

The same asymptotic representation holds for $\hat{\alpha}_n(x,y)$ and $\hat{\alpha}_n^{\#}(x,y)$ in view of their consistency. The fact that $I\,F\,((X_i,Y_i)\,;\Phi_{x,y},F)$ is zero whenever $X_i \not\le x$, ensures that outlying points (X_i,Y_i) with $X_i \not\le x$ have no effect on $\alpha(x,y)$ estimates. Even in the presence of outliers with $X_i \le x$, $\alpha_n(x,y)$, $\hat{\alpha}_n(x,y)$ and $\hat{\alpha}_n^{\#}(x,y)$ can be viewed as bias-robust since the gross-error sensitivity γ^*, *i.e.*, the sup-norm of *IF*, is finite. γ^* is interpreted as the worst possible influence which a fixed amount of contamination can have upon the estimators. However, these estimators may suffer excessive bias at the boundaries of the sample in the input-orientation where $F_X(x)$ is negligible. To correct this failure, an attractive technique which attains the desirable boundary bias rate, would be to adapt the data imputation method of HALL and MULLER [2003], which involves reflection in points on the boundaries. Instead, we will propose in Subsection 5.3 a simple practical method combining FDH efficiency scores and an approximative empirical choice of bandwidths.

4 Practical Aspects

Computational algorithm: To compute in practice the monotone estimator $\widehat{\alpha}_n^{\#}$ we use a discrete grid instead of the whole domain D. For instance, we could consider the minimal rectangular set with edges parallel to the coordinate axes that covers all of the observations X_i, and then choose a discrete grid $D_n = \{x_{n,1}, \cdots, x_{n,k}\}$ in this rectangular set containing the unique minimal and maximal (with respect to the partial order "$\leq$") points of this set. Such a choice makes it easier to compute $\widehat{\alpha}_n^{\#}(x,y)$ for x in the rectangular set by using the following algorithm. First compute $\widehat{\alpha}_n^{u}$ successively along D_n starting from its maximal point, using the fact that:

$$\widehat{\alpha}_n^{u}(x_{n,i},y) = \widehat{\alpha}_n(x_{n,i},y) \vee \max\{\widehat{\alpha}_n^{u}(x_{n,j},y) : x_{n,j} \text{ is an immediate successor of } x_{n,i}\},$$

for all $x_{n,i} \in D_n$. Compute also $\widehat{\alpha}_n^{l}$ successively along D_n starting this time from its minimal point, using the fact that:

$$\widehat{\alpha}_n^{l}(x_{n,i},y) = \widehat{\alpha}_n(x_{n,i},y) \wedge \min\{\widehat{\alpha}_n^{l}(x_{n,j},y) : x_{n,j} \text{ is an immediate predecessor of } x_{n,i}\}.$$

The monotone estimator $\widehat{\alpha}_n^{\#}(x,y)$ can be therefore easily computed, for any x in the rectangular set, as the mean of $\widehat{\alpha}_n^{u}(x,y)$ and $\widehat{\alpha}_n^{l}(x,y)$ with:

$$\widehat{\alpha}_n^{l}(x,y) = \min_{x_{n,i} \in D_n | x_{n,i} \leq x} \widehat{\alpha}_n(x_{n,i},y) \text{ and } \widehat{\alpha}_n^{u}(x,y) = \max_{x_{n,i} \in D_n | x_{n,i} \geq x} \widehat{\alpha}_n(x_{n,i},y).$$

This algorithm is inspired from MUKERJEE and STERN [1994]. By an appropriate choice of D_n, they prove that when their initial estimator of the regression function is strongly uniformly consistent and the regression function is nondecreasing, then their isotonized estimator is also strongly uniformly consistent. We can easily adapt their proof to show the uniform complete convergence of our isotonic estimator $(x,y) \mapsto \widehat{\alpha}^{\#}(x,y)$ on $D \times S$ under some further regularity conditions on $x \mapsto \alpha(x,y)$.

For $\delta > 0$, let $D_\delta \supset D$ be the closed δ-neighborhood of D which we assume to be interior to the support of X, and let $\{b_n\}$ be a positive sequence tending to 0. We then define D_n as the set of vectors in D_δ with components that are integral multiples of b_n. As pointed out by Mukerjee and Stern, if D is rectangular with edges parallel to the coordinate axes, as is often the case, then we could consider only the minimal subset of D_n that covers D by convex combinations. The minimal and maximal points of this subset being unique, we can then isotonize $\widehat{\alpha}_n(x,y)$ over $D \times S$ by applying the algorithm described above.

It is important to note that in this case, since D_n is not contained in D, the proof of Lemma A.1 is no more valid, and then this lemma cannot be applied to derive the uniform complete convergence of $\widehat{\alpha}_n^{\#}$ to α on $D \times S$. But if we use the same arguments as Mukerjee and Stern ([1994], see the paragraph below Equation (4), p. 78), we can easily show that $\left\|\widehat{\alpha}_n^{\#} - \alpha\right\|_{D\times S} \leq \left\|\widehat{\alpha}_n - \alpha\right\|_{D_\delta \times S} + R_n$, where:

$$R_n = \sup_{(x,y)\in D\times S} \{\alpha(x,y) - \alpha(\overline{x}_n, y)\} \vee \sup_{(x,y)\in D\times S} \{\alpha(\overline{\overline{x}}_n, y) - \alpha(x,y)\},$$

with $\overline{x}_n \geq x$ and $\overline{\overline{x}}_n \leq x$ denote the unique nearest neighbors of x in D_n satisfying the given order restrictions. If we assume that $(x,y) \mapsto \alpha(x,y)$ is uniformly continuous on $D_\delta \times S$, then by using the fact that $\left\|x - \overline{x}_n\right\|$ and $\left\|x - \overline{\overline{x}}_n\right\|$ are bounded by $\sqrt{p}\, b_n$ for $x \in D$ and $b_n \to 0$, we get $R_n = o(1)$.

Now let $D = \prod_{i=1}^{p} [a_i, b_i]$, $S = \prod_{j=1}^{q} [c_j, d_j]$ and let $D_\delta = \prod_{i=1}^{p} [a_i - \delta, b_i + \delta]$ be interior to the support of X. Then, by using the following decomposition

$$\left\|\widehat{\alpha}_n - \alpha\right\|_{D_\delta \times S} \leq \left[\left\|\widehat{F} - F\right\|_{D_\delta \times S} + \left\|\widehat{F}_X - F_X\right\|_{D_\delta}\right] / \inf_{x\in D_\delta} \widehat{F}_X(x),$$

and by following step by step the proof of Theorem 3.2, we get $\left\|\widehat{\alpha}_n - \alpha\right\|_{D_\delta \times S} \xrightarrow{co.} 0$ under the conditions of Lemmas A.3 and A.4, which finally gives the uniform complete convergence of $\widehat{\alpha}_n^{\#}$ to α on $D \times S$.

Bandwidths selection: Deriving asymptotically optimal bandwidths is a tedious matter. Using cross-validation demands selection of the amount of data that are left out and using the bootstrap calls for selection of subsidiary smoothing parameters (using plug-in methods in the time series case requires explicit estimation of complex functions using dependent data). Such complexity is arguably not justified (see Hall *et al.* [1999], p. 156). Instead, we suggest an approximate empirical method as follows. In place of using a unique bandwidth, as in formula (3), we adapt a bandwidth to each component, *i.e.*, we use in (3) the vectors $\left(\frac{x^1 - X_i^1}{h_0^1}, \cdots, \frac{x^p - X_i^p}{h_0^p}\right)$ and $\left(\frac{y^1 - Y_i^1}{h_1^1}, \cdots, \frac{y^q - Y_i^q}{h_1^q}\right)$ rather than $\frac{x - X_i}{h_0}$ and $\frac{y - Y_i}{h_1}$, respectively, and we tune each bandwidth h_0^j so that approximately a reasonable percentage $\rho\%$ of the data points $X_1^j, \cdots, X_n^j$ fall into the support of $u \mapsto K_0^j(\frac{x^j - u}{h_0^j})$. Similarly we tune each bandwidth h_1^k so that $\rho\%$ of the data points $Y_1^k, \cdots, Y_n^k$ fall into the support of $u \mapsto K_0^k\left(\frac{y^k-u}{h_1^k}\right)$. In case of kernels with support [-1, 1], as Triweight and Epanechnikov kernels, we obtain explicit formulas of h_0^j and h_1^k :

$$h_0^j = \frac{\rho}{200}\left(\max_{i=1,\cdots,n} X_i^j - \min_{i=1,\cdots,n} X_i^j\right), \quad h_1^k = \frac{\rho}{200}\left(\max_{i=1,\cdots,n} Y_i^k - \min_{i=1,\cdots,n} Y_i^k\right).$$

Our method itself requires selection of a smoothing parameter ρ, but this parameter is dimensionless and the selection rule has the advantage to be very simple to interpret and to implement, particularly in the difficult context of adapting a bandwidth to each component. Moreover, we propose in Subsection 5.3 a simple algorithm for practical efficiency measurement, that starts with any initial value of ρ and leads to reasonable efficiency estimates at the end of the procedure by improving the value of ρ.

5 Numerical Illustration

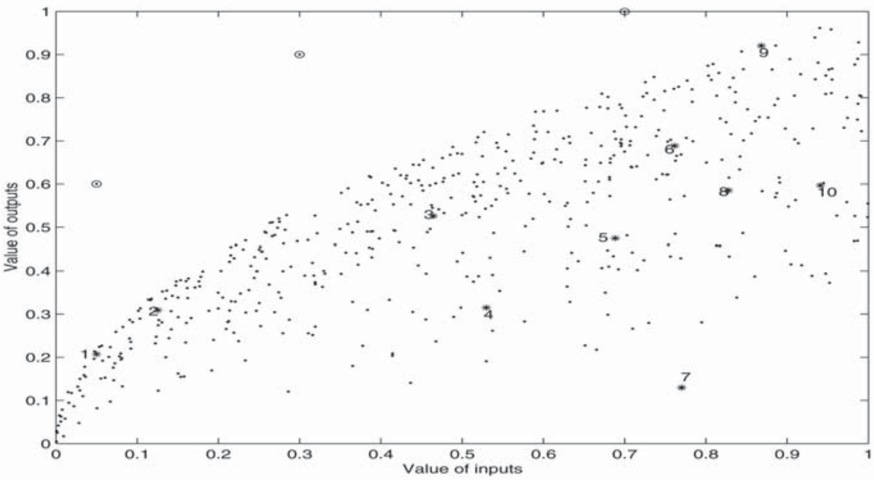
For our practical computations, the smooth measure $\widehat{\alpha}_n$ will be evaluated for a grid of values of ρ by using a Triweight kernel $k_0^i = k_1^j$ for any $i = 1,\cdots,p$ and any $j = 1,\cdots,q$. For the isotone measure $\widehat{\alpha}_n^{\#}$, we choose the discrete grid D_n to be the set of the observations $\{X_i\}$ and the minimal and maximal points X_0 and X_{n+1}, where $X_0^j = \min_{i=1,\cdots,n} X_i^j$ and $X_{n+1}^j = \max_{i=1,\cdots,n} X_i^j$, for $j = 1,\cdots,p$. In order to investigate the performance of $\widehat{\alpha}_n^{\#}$ (precision and robustness), we compare it with the FDH estimator of Shephard output distance function. As indicated in Section 1, Shephard output-efficiency of a unit $(x,y) \in \Psi$ is merely the reciprocal of the conventional Farrell efficiency score $\lambda(x,y)$ whose FDH estimator is given by $\widehat{\lambda}_n(x,y) = \max_{i|X_i \le x} \min_{j=1,\cdots,q} Y_i^j / y^j$. The asymptotic properties of this estimator are provided in Park, Simar and Weiner [2000].

5.1 Simulated Data: Cobb-Douglas Model

We simulate $n = 500$ data points (X_i, Y_i) according to the Cobb-Douglas model $Y = X^{0.5} * \exp(-U)$, where X is uniform on [0, 1] and U is exponential with mean 1/3. To save space, we show in Table 1 the efficiency estimates $\widehat{\lambda}_n^{-1}$ and $\widehat{\alpha}_n^{\#}$ of 10 units chosen at random among the simulated observations, which we represent in Figure 1 by '*'. Note that the three outliers in this figure are not generated by the same model as the 500 simulated data points, and that $\widehat{\alpha}_n^{\#}$ is computed with a Triweight kernel.

FIGURE 1

500 simulated observations with three outliers included: '*' indicate the 10 evaluation units showed in Table 1.



It is clear from Figure 1 that both $\hat{\lambda}_n^{-1}$ and $\hat{\alpha}_n^{\#}$ give reasonable efficiency estimates of the 10 evaluation points, and are very similar except for units 1 and 2 when ρ exceeds the value 10. This worse behavior of $\hat{\alpha}_n^{\#}$ for high values of ρ can be explained as follows: since units 1 and 2 are near the border in the input orientation, the choice of h_0 according a high percentage $\rho\%$ generates a very large bandwidth, which results in a poor estimate $\hat{\alpha}_n^{\#}$. It seems following Table 1 that we shall choose a small value of ρ in order to overcome such a deficiency. But, in fact, even for ρ small enough, $\hat{\alpha}_n^{\#}$ may underestimate extreme units at the border of the sample with very small inputs, as can be seen from the following result:

(6)

$$\lim_{\rho\searrow 0}\hat{\alpha}_n(x,y) = \frac{\sum_{i=1}^{n} 1\!\mathrm{I}(Y_i<y)[1\!\mathrm{I}(X_i<x)+\frac{1}{2}1\!\mathrm{I}(X_i=x)] + \frac{1}{2}\sum_{i=1}^{n} 1\!\mathrm{I}(Y_i=y)[1\!\mathrm{I}(X_i<x)+\frac{1}{2}1\!\mathrm{I}(X_i=x)]}{\sum_{i=1}^{n}[1\!\mathrm{I}(X_i<x)+\frac{1}{2}1\!\mathrm{I}(X_i=x)]}$$

for any observation $(x,y)\in\{(X_i,Y_i)\}_{i=1,\cdots,n}$. For instance, the observation (x,y) using the minimal input (if it is unique) is FDH efficient $\left(\hat{\lambda}_n^{-1}=1\right)$, whereas it is underestimated by our measure since $\lim_{\rho\searrow 0}\hat{\alpha}_n(x,y)=0.5$.

TABLE 1

***n* = 500, Efficiencies of the 10 units indicated by '*' in Figure 1.**

Unit	$\hat{\lambda}_n^{-1}$	$\hat{\alpha}_n^{\#}$ ($\rho = 1$)	$\hat{\alpha}_n^{\#}$ ($\rho = 5$)	$\hat{\alpha}_n^{\#}$ ($\rho = 10$)	$\hat{\alpha}_n^{\#}$ ($\rho = 20$)	$\hat{\alpha}_n^{\#}$ ($\rho = 30$)
1	0.97167	0.95139	0.94501	0.92597	0.85498	0.76822
2	0.92053	0.92242	0.92365	0.91624	0.88551	0.85141
3	0.80255	0.8641	0.86526	0.86255	0.8548	0.84488
4	0.43719	0.38054	0.37463	0.37452	0.37547	0.37759
5	0.58608	0.57983	0.57852	0.57946	0.5789	0.57696
6	0.81224	0.88289	0.88499	0.88667	0.8845	0.87852
7	0.153	0.053017	0.053141	0.053407	0.056026	0.059483
8	0.66668	0.67896	0.67686	0.67616	0.67798	0.68103
9	1	0.99944	0.99931	0.99879	0.99778	0.99554
10	0.62052	0.65206	0.65474	0.65662	0.65707	0.65916

To summarize, an efficient production unit at the border of the sample in the direction of small inputs, could be underestimated by our efficiency measure when ρ exceeds some high threshold (as it is the case for Units 1 and 2 in Table 1) and even when $\rho \searrow 0$ (as indicated by (6)). A simple methodology which, simultaneously, identifies the observations which suffer from severe border effects and generates an "optimal" value of ρ with satisfactory efficiency estimates, is proposed in Subsection 5.3 for the more general framework of multiple inputs and multiple outputs.

Now, to test the robustness of the estimator $\hat{\alpha}_n^{\#}$, we add in the data set three outliers as shown in Figure 1 and we compute $\hat{\lambda}_n^{-1}$ and $\hat{\alpha}_n^{\#}$. Table 2 shows the difference between these efficiency estimates and those measured in Table 1. As can be seen from this table, the estimates $\hat{\alpha}_n^{\#}$ are overall very robust ($0 \leq \Delta\hat{\alpha}_n^{\#} < 0.008$ for units 3, …, 10, and $0.01 < \Delta\hat{\alpha}_n^{\#} < 0.02$ for units 1,2), whereas $\hat{\lambda}_n^{-1}$ has dramatic changes for most units $(0.02 < \Delta\hat{\lambda}_n^{-1} \leq 0.62)$.

5.2 Real Data: Spanish Electricity Distributors

The data set used in GRIFELL and LOVELL [1999] contains information concerning the production of electricity by 61 firms in Spain. The output (y) is the amount of low, medium and high voltage electricity distributed (GWh) and the 3 inputs are the population density (x^1), the substation transformer capacity from medium voltage to low voltage (x^2) and the length in Km of voltage lines (x^3).

The results are shown in Table 3. We first remark that many of the firms are FDH-efficient ($\hat{\lambda}_n^{-1}$ often equal to one), but this is due to the high dimensionality of the space (3 + 1) and to the small sample size ($n = 61$). All of these units belong to the efficient surface (FDH frontier) of the smallest free disposal set containing all the data.

TABLE 2

n = 503, The operator Δ gives the difference between the efficiency estimates measured in Table ?? and those measured after including the three outliers.

Unit	$\Delta\hat{\lambda}_n^{-1}$	$\Delta\hat{\alpha}_n^{\#}$ ($\rho = 1$)	$\Delta\hat{\alpha}_n^{\#}$ ($\rho = 5$)	$\Delta\hat{\alpha}_n^{\#}$ ($\rho = 10$)	$\Delta\hat{\alpha}_n^{\#}$ ($\rho = 20$)	$\Delta\hat{\alpha}_n^{\#}$ ($\rho = 30$)
1	0.62609	0.01947	0.01843	0.01787	0.017221	0.01562
2	0.40552	0.015699	0.015433	0.015322	0.014708	0.013993
3	0.21812	0.0074232	0.0076263	0.00768	0.0079734	0.0077666
4	0.087276	0.0028808	0.0030007	0.002816	0.0027316	0.0026144
5	0.059044	0.0032924	0.0033455	0.0036382	0.0039355	0.0040526
6	0.12364	0.0041436	0.0040354	0.0041623	0.0045075	0.0047589
7	0.02329	0.00035328	0.00045438	0.00030477	0.00019667	0
8	0.081454	0.0047678	0.0047315	0.0042029	0.0037278	0.0035866
9	0.079688	0.0022348	0.0022588	0.002523	0.0029747	0.0031667
10	0.023728	0.0040973	0.0032257	0.0031422	0.0030559	0.0029503

We also remark that the two measures $\hat{\lambda}_n^{-1}$ and $\hat{\alpha}_n^{\#}$ agree overall except for very few units: 20, 21, 23, 33, 37, 58. A careful reading of the data confirms that the firms 20, 21 and 37 can be seen to be performing relatively efficiently as indicated by $\hat{\alpha}_n^{\#}$. Their efficiencies are underestimated by $\hat{\lambda}_n^{-1}$ due to the presence of some super-efficient firms which produce by far a much larger amount of electricity while using less input. More precisely, they are influenced, respectively, by the following sets of very extreme units: $\{44, 59\}$, $\{9, 24\}$ and $\{2, 3, 6\}$. Note that all of these influent units belong to the FDH frontier. Firms 59, 24 and 2 are particularly influent because they belong to the DEA frontier, *i.e.*, the upper surface of the smallest free disposal convex set covering all of the 61 firms. As expected, when deleting units $\{44, 59\}$ from the data points, we get the new measures $\hat{\lambda}_n^{-1} = 1$ and $\hat{\alpha}_n^{\#} = 0.99566$ for unit 20. Likewise, when deleting units $\{9, 24\}$ (resp.: $\{2, 3, 6\}$) from the data set, efficiencies of unit 21 (resp.: 37) become: $\hat{\lambda}_n^{-1} = 1$ and $\hat{\alpha}_n^{\#} = 0.9259$ (resp.: $\hat{\lambda}_n^{-1} = 1$ and $\hat{\alpha}_n^{\#} = 0.99805$).

The other units of interest 23, 33 and 58 are also underestimated, but this time by $\hat{\alpha}_n^{\#}$. A careful analysis of the small data set shows that these firms are very extreme in both output and input directions. This is the reason which explains the fact that their integrated kernel-based efficiencies are underestimated. We can overcome this border problem for units 23 and 33 by using small values of ρ (see Table 3). But the measure $\hat{\alpha}_n^{\#} = 0.5$ of unit 58 does not change due to the fact that this firm is very isolated with respect to the other firms. As a matter of fact, like in the univariate case (see (6)), it can be easily seen for the general multivariate case that $\lim_{\rho \searrow 0} \hat{\alpha}_n$ does not achieve a reasonable level, which should be close to 1, for isolated FDH points like unit 58 which is clearly atypical.

In order to deal with this limitation, we propose in the next subsection a simple procedure based on the fact that a FDH point (x, y) for which $\hat{\alpha}_n^{\#}$ does not exceed the threshold $t = 0.9$ for any ρ can be viewed as a potential outlier in the input-orientation. We could choose other values of t, say 0.95, 0.85 or 0.8: the more t is smaller than 1, the more (x, y) is suspicious: it can be a very extreme point at the border of the sample (very small w.r.t. both inputs and outputs) and/or an isolated one in the input-orientation in the sense that some of its components are outliers. This procedure allows in particular to perform efficiency estimates of observations situated at the border of the sample and to detect those which suffer from severe border effects, essentially atypical and/or outlying units with very small inputs.

5.3 **Practical Efficiency Measurement**

To perform any efficiency analysis, we can start our computation method by using any initial value of ρ, for instance 1, 5, 10 or 20.

[1] Compute $\hat{\lambda}_n^{-1}$ and $\hat{\alpha}_n^{\#}$ for a given value of ρ.

[2] Identify efficient FDH units $\left(\hat{\lambda}_n^{-1} = 1\right)$ with measure $\hat{\alpha}_n^{\#} \leq 0.9$ (the identification can be addressed to automatically):

[2.a] If there are no identified points, then the border problem does not have severe effects on the estimator $\hat{\alpha}_n^{\#}$. We can therefore stop the procedure and evaluate the production performance of the data points under analysis by using their efficiency estimates $\hat{\alpha}_n^{\#}$ for the given value of ρ.

The use of $\hat{\alpha}_n^{\#}$ rather than $\hat{\lambda}_n^{-1}$ is motivated essentially by the fact that it is very resistant to extreme values in the output-orientation as it is shown in Examples 1-2, where $\hat{\lambda}_n^{-1}$ behaves dramatically for the majority of evaluation points in Example 1, after including 3 outliers, and for units 20, 21 and 37 in Example 2.

[2b] If some points are identified, then proceed to step [3].

Each identified point may be either outlying in the input-orientation or just underestimated by $\hat{\alpha}_n^{\#}$ due to border effects when ρ is not small enough. For Example 2, if we start by $\rho = 1$ (see Table 3) then only units 23, 33, 58 are identified. If we rather start by $\rho = 5$ or 10, we obtain at the end the same results as will be seen below.

[3] For a smaller value of ρ, compute again $\hat{\alpha}_n^{\#}$ for the identified units in the preceding step. Then identify those of measure $\hat{\alpha}_n^{\#}$ still smaller than 0.9:

[3.a] If there are no identified points, then proceed to [2.a].

[3.b] If there is an identified point for which $\hat{\alpha}_n^{\#}$ increases, then repeat [3] for all identified units with $\hat{\alpha}_n^{\#} \leq 0.9$.

For Example 2, if we take $\rho = 0.1$ (see Table 3) then only unit 58 is identified, but its measure $\hat{\alpha}_n^{\#}$ does not change and so, we should proceed to step [3.c].

[3.c] If $\hat{\alpha}_n^{\#}$ does not change for all identified units, then these units are suspicious (potential outliers):

– compute $\hat{\alpha}_n^{\#}$ for all data points under analysis for the last value of ρ;

TABLE 3

The data set $\{(x_i, y_i), i = 1,\dots, 61\}$, and the efficiencies $\hat{\alpha}_n^{\#}$ and $\hat{\lambda}_n^{-1}$.

Unit	*y*	x^1	x^2	x^3	$\hat{\lambda}_n^{-1}$	$\hat{\alpha}_n^{\#}$ (ρ = 1)	$\hat{\alpha}_n^{\#}$ (ρ = 5)	$\hat{\alpha}_n^{\#}$ (ρ = 10)
1	1241	28.83076923	439	7007	1	0.98759	0.98244	0.97173
2	3334	50.27580645	1165	5577	1	0.99766	0.9976	0.99748
3	1871	23.12402875	865	5960	1	0.99747	0.99535	0.98689
4	1489	30.97287894	843	6840	0.8533	0.96126	0.94835	0.93222
5	1450	23.33291345	728	5586	1	0.9906	0.97972	0.96755
6	2724	91.25431152	1118	604	1	0.99805	0.99425	0.99026
7	4500	50.00441721	1935	17050	1	0.99851	0.99768	0.99328
8	684	7.47295423	314	4272	1	0.99111	0.98667	0.94334
9	504	7.085373364	237	3774	1	0.98989	0.97516	0.92999
10	2177	276.1740644	1012	3859	1	0.99635	0.99617	0.99575
11	968	9,927531182	407	5459	1	0.97214	0.92427	0.90828
12	316	51.87694145	142	4383	0.81443	0.89424	0.86947	0.78648
13	1227	11.94137353	404	5239	1	0.99584	0.99411	0.98362
14	1097	12.46605886	869	7692	0.89405	0.96277	0.91355	0.84461
15	297	4.96336056	147	4370	0.76546	0.82134	0.82351	0.75795
16	388	3.90584575	110	2169	1	0.97102	0.97439	0.92567
17	358	9.212554927	393	2961	0.92268	0.86467	0.88833	0.82569
18	1036	12.44840598	306	4869	1	0.99496	0.98704	0.97591
19	971	17.39387475	460	4102	1	0.99474	0.98905	0.97121
20	1267	1407.346153	654	9182	0.59567	0.9373	0.91498	0.90099
21	415	16.79987577	277	5871	0.44576	0.78817	0.75607	0.69546
22	1393	15.43905681	365	9829	1	0.99641	0.99537	0.99078
23	23	4.116877045	12	721	1	0.77838	0.55366	0.50552
24	931	9.428855657	250	4690	1	0.99337	0.99011	0.98073
25	705	16.69379752	675	8463	0.57457	0.67688	0.68086	0.6593
26	95	10.77475363	54	1406	1	0.9797	0.77067	0.60485
27	809	17.49376518	269	5685	0.86896	0.91915	0.93686	0.92241
28	501	15.6301784	306	3746	1	0.97848	0.95465	0.90473
29	212	4.603279324	63	1774	1	0.96675	0.94088	0.77896
30	87	4.888839285	23	1781	1	0.96387	0.70532	0.56864
31	1745	25.9270113	700	5192	1	0.99675	0.99482	0.98781
32	410	11.6827005	162	4711	1	0.9814	0.93311	0.85979
33	22	6.847996695	10	1272	1	0.68916	0.52584	0.4801
34	3476	1348.198148	1729	8594	1	0.99864	0.99244	0.98564
35	2844	184.3938193	840	7038	1	0.99823	0.99823	0.99821
36	1872	30.81301394	1080	8089	1	0.98267	0.97947	0.96485
37	1868	169.2686671	1344	6058	0.56029	0.93433	0.93483	0.92886
38	93	46.11206896	76	973	1	0.97059	0.74719	0.60679

TABLE 3, CONT.

Unit	*y*	x^1	x^2	x^3	$\hat{\lambda}_n^{-1}$	$\hat{\alpha}_n^{\#}$ ($\rho = 1$)	$\hat{\alpha}_n^{\#}$ ($\rho = 5$)	$\hat{\alpha}_n^{\#}$ ($\rho = 10$)
39	435	74.6214605	251	3745	1	0.9716	0.93656	0.87538
40	150	10.22512234	118	438	1	0.94444	0.88708	0.72749
41	913	10.28438	628	8071	0.94318	0.88638	0.84739	0.80467
42	3317	49.61331626	1309	9165	1	0.98991	0.99107	0.991
43	4397	154.5850094	1259	20925	1	0.99867	0.99867	0.99867
44	2127	46.01906334	581	6784	1	0.9977	0.99772	0.99721
45	7049	110.339922	1932	17353	1	0.99871	0.99871	0.9987
46	270	3.437563171	62	2410	1	0.96254	0.92054	0.80524
47	855	15.88256346	446	5427	0.69682	0.83094	0.84564	0.84332
48	750	9.157086772	322	4681	1	0.97502	0.94573	0.92361
49	4858	71.23629629	1494	14214	1	0.99859	0.99859	0.99856
50	212	148.0787878	97	1394	1	0.98615	0.93638	0.77915
51	339	15.30875576	143	6186	0.87371	0.88686	0.85633	0.77458
52	732	28.7513053	432	9602	0.59658	0.73957	0.73321	0.74075
53	2080	84.46452476	988	10075	0.9779	0.97417	0.98148	0.97862
54	957	32.76927651	327	3196	1	0.99315	0.99323	0.99253
55	10470	184.5294044	3266	22811	1	0.99886	0.99885	0.99884
56	6065	417.2896551	4610	16179	1	0.99885	0.99885	0.99885
57	3347	49.05046844	829	13977	1	0.99826	0.99826	0.99814
58	5	2.517326732	4	35	1	0.5	0.5	0.48899
59	1793	45.42970036	531	3208	1	0.99439	0.99429	0.99397
60	4992	164.1621212	1759	7426	1	0.9984	0.99842	0.99842
61	5362	243.5128552	3612	7621	1	0.99852	0.99852	0.99835

– set the efficiency measure to 1 for the identified suspicious units and to $\hat{\alpha}_n^{\#}$ for the remaining units;

– delete the identified suspicious units from the data and then proceed to step [4] in order to avoid the masking effect.

Indeed, a FDH potential outlier in the input-orientation could mask other suspicious points situated inside the FDH surface, near the first one. If we delete the FDH potential outlier from the data, then at least one of the masked points will become FDH efficient and can be therefore easily detected by comparing initial FDH points with those obtained after removing the FDH potential outlier.

[4] Compute $\hat{\lambda}_n^{-1}$ for the remaining data points after deleting the suspicious units and then, by comparing the resulting FDH points with those of the preceding step, identify the additional ones. In order to avoid the masking effect, the analysis should be repeated starting at step [1] and using, as data points, only the identified additional FDH points.

In our procedure, the computations are so fast that the whole process can be repeated several times, by deleting iteratively the identified suspicious observa-

TABLE 4

For $t = .95$ the extreme four isolated units 23, 33, 40 and 58 are identified as potential outliers in the input-orientation and for the smaller threshold $t = .9$ only the unit 58, which is clearly atypical, is detected by our efficiency analysis.

Unit	$\hat{\alpha}_n$ ($\rho = 0.001$)	$\hat{\alpha}_n$ ($\rho = 0.01$)	$\hat{\alpha}^{\#}_n$ ($\rho = 0.1$)	*Unit*	$\hat{\alpha}_n$ ($\rho = 0.001$)	$\hat{\alpha}_n$ ($\rho = 0.01$)	$\hat{\alpha}^{\#}_n$ ($\rho = 0.1$)
1	0.9973	0.9973	0.9973	32	0.9938	0.9938	0.9938
2	0.9976	0.9976	0.9976	33	0.9444	0.9444	0.9441
3	0.9974	0.9974	0.9974	34	0.9986	0.9986	0.9986
4	0.9608	0.9608	0.9608	35	0.9982	0.9982	0.9982
5	0.9972	0.9972	0.9972	36	0.9980	0.9980	0.9885
6	0.9981	0.9981	0.9981	37	0.9128	0.9128	0.9140
7	0.9985	0.9985	0.9985	38	0.9706	0.9706	0.9706
8	0.9923	0.9923	0.9923	39	0.9938	0.9938	0.9938
9	0.9912	0.9912	0.9910	40	0.9444	0.9444	0.9444
10	0.9964	0.9964	0.9964	41	0.8636	0.8636	0.8639
11	0.9952	0.9952	0.9953	42	0.9984	0.9984	0.9984
12	0.8951	0.8951	0.8951	43	0.9987	0.9987	0.9987
13	0.9961	0.9961	0.9961	44	0.9977	0.9977	0.9977
14	0.9444	0.9437	0.9434	45	0.9987	0.9987	0.9987
15	0.8265	0.8265	0.8128	46	0.9444	0.9444	0.9444
16	0.9444	0.9444	0.9528	47	0.8399	0.8399	0.8386
17	0.8509	0.8509	0.8509	48	0.9938	0.9938	0.9938
18	0.9952	0.9952	0.9952	49	0.9986	0.9986	0.9986
19	0.9948	0.9948	0.9948	50	0.9848	0.9848	0.9848
20	0.9396	0.9396	0.9396	51	0.8836	0.8836	0.8836
21	0.8429	0.8429	0.8429	52	0.7378	0.7378	0.7378
22	0.9964	0.9964	0.9963	53	0.9742	0.9742	0.9742
23	0.9444	0.9444	0.9444	54	0.9932	0.9932	0.9932
24	0.9932	0.9932	0.9932	55	0.9989	0.9989	0.9989
25	0.6946	0.6946	0.6990	56	0.9988	0.9988	0.9988
26	0.9800	0.9800	0.9800	57	0.9983	0.9983	0.9983
27	0.9190	0.9190	0.9190	58	0.5000	0.5000	0.5000
28	0.9932	0.9932	0.9932	59	0.9944	0.9944	0.9944
29	0.9706	0.9706	0.9705	60	0.9984	0.9984	0.9984
30	0.9706	0.9706	0.9706	61	0.9985	0.9985	0.9985
31	0.9967	0.9967	0.9967				

tions, in a reasonable computing time. Obviously, the process stops when we get either [2.a] or [3.a] or if there are no identified additional FDH units in step [4].

For Example 2, as pointed out above, for $\rho = 0.1$ we identify unit 58 for which [3.c] applies. After removing this atypical unit from the data, no unit shows suspicious behavior anymore. Indeed, there are no identified additional FDH units. Thus, we set the efficiency measure to 1 for the atypical unit 58 and to the estimates $\hat{\alpha}_n^{\#}$ ($\rho = 0.1$) for the remaining sixty units.

The same analysis was repeated starting the algorithm with $\rho = 5$ (resp.: $\rho = 10$) and using respectively $\{1, 0.1\}$ as smaller values of ρ for step [3] and we obtained the same results. This can be easily checked by looking carefully to Table 3.

Starting with $\rho = 1$ (resp.: 5 and 10) and using this time the larger threshold $t = 0.95$, besides the atypical unit 58 we also identified the additional suspicious units 23, 33 and 40, as it can be easily seen from Table 3. So, when t is closer to 1, we identify outliers in the input-orientation as well as very extreme units at the border of the sample which are potential outliers. For $t = 0.95$, we set the efficiency measure to 1 for the potential outliers 23, 33, 40 and 58 and to the estimates $\hat{\alpha}_n^{\#}$ ($\rho = 0.001$) for the remaining 57 units.

6 Concluding Remarks

If we assume, in productivity analysis, that observations $\{(X_i, Y_i)\}$ are independently drawn from a population where the support $Supp(X, Y)$ of the underlying distribution defines the production set Ψ, then the nonstandard conditional probability measure $\alpha(x, y) = \text{Prob}(Y \leq y \mid X \leq x)$ can be viewed as a good benchmark value in the output-orientation for producing units (x, y) interior to Ψ.

Order-preserving estimation: We argue that the production efficiency measure $\alpha(x, y)$ should be increasing in y and decreasing in x in order to equally take into account the waste of production resources in efficiency measurement. We propose a two-stage estimation procedure in which we incorporate this information. We first construct a smoothed kernel integrated-based estimator which is asymptotically normally distributed and only monotone with respect to outputs. Then we propose an isotonized version of the initial smoothed estimator with respect to inputs and we establish its uniform and complete convergence.

Bias-robustness: As indicated by the influence function, although our kernel integrated-based estimator enjoys the bias-robustness in the output-orientation, the desired ordering properties and always lies between 0 and 1, it may have worse behavior at the border of the sample in the input-orientation due to the conditioning by $X \leq x$. To correct this failure, we suggest a simple practical algorithm making use of the FDH efficiency scores in conjunction with an approximative empirical choice of bandwidths. This allows to improve efficiencies for production units which suffer from the border effects and eventually also to point out outliers in the input-orientation or at least very extreme observations in the direction of inputs. The results for the Spanish electricity distributors confirm that: our method gives overall reasonable efficiency estimates of the 61 electricity distributors and allows a meaningful comparison of the performance of their activities.

Environmental variables: In order to improve managerial performance by identifying economic conditions that lead to inefficiency, it is possible to introduce some external or environmental factors $Z \in \mathbb{R}^r$ which might influence the production process but that are not under the control of the firms. It would then be interesting to extend our approach by conditioning the production process to a given value of $Z = z$ to explore the influence of environmental factors on the efficiency of the firms. The estimation of the resulting conditional efficiency measure $\alpha(x, y \mid z) = \text{Prob}(Y \leq y \mid X \leq x, Z = z)$ could be addressed using standard theory.

Open issues: Important topics of interest for future research are the development of other monotonization techniques and an investigation of the asymptotic distribution of the isotone version of our smoothed estimator. On the other hand, the efficiency measure α(X, Y) is uniformly distributed on [0, 1] which seems to be a desirable property when examining returns to scale (*i.e.*, the effects of firm size on production performance) as motivated in the analysis of Kokic *et al*. [1997]. It would then be interesting to investigate how this efficiency measure could be used to make inference on returns to scale as it is the case for the measures of Farrell-Shephard and Kokic *et al.* [1997]. ■

References

Aigner D.J., Lovell C.A.K. and Schmidt P. (1977). – « Formulation and Estimation of Stochastic Frontier Production Function Models », Journal of Econometrics, 6: pp. 21-37.

Ait-Sahalia Y. (1993). – « Nonparametric Functional Estimation with Applications to Financial Models », *MIT PhD thesis*.

Aragon Y., Daouia A. and Thomas-Agnan C. (2005). – « Nonparametric Frontier Estimation: A Conditional Quantile-Based Approach », Econometric Theory, 21: pp. 358-389.

Cazals C., Florens J.P. and Simar L. (2002). – « Nonparametric Frontier Estimation: A Robust Approach », Journal of Econometrics, 106: pp. 1-25.

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

Choi E., Hall P. and Rousson V. (2000). – « Data Sharpening for Bias Reduction in Nonparametric Regression », *Annals of Statistics*, 28: pp. 1339-1355.

Daouia A. (2005). – « Asymptotic Representation Theory for Nonstandard Conditional Quantiles », Journal of Nonparametric Statistics, 17(2): pp. 253-268.

Daouia A. and Ruiz-Gazen A. (2004). – « Robust Nonparametric Frontier Estimators: Influence Function and Qualitative Robustness », to appear in Volume 16, Number 4, October 2006, Statistica Sinica.

Debreu G. (1951). – « The Coefficient of Resource Utilization », Econometrica*,* 19(3): pp. 273-292.

Deprins D., Simar L. and Tulkens H. (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, pp. 243-267.

Fan J.Q. (1992). – « Design-Adaptive Nonparametric Regression », Journal of the American Statistical Association, 87: pp. 998-1004.

Fan J.Q. and Gijbels I. (1992). – « Variable Bandwidth and Local Linear Regression Smoothers », *Annals of Statistics*, 20: pp. 2008-2036.

FARRELL M.J. (1957). – « The Measurement of Productive Efficiency », *Journal of the Royal Statistical Society, Series A*, 120: pp. 253-281.

GRIFFEL-TATJÉ E. and LOVELL C.A.K. (1999). – « The Managers Versus the Consultants, Sixth European Workshop on Efficiency and Productivity Analysis », 29 to 31 October, 1999, *University of Copenhagen*, Denmark.

HALL P. and MULLER H.G. (2003). – « Order-Preserving Nonparametric Regression with Applications to Conditional Distribution and Quantile Function Estimation », *Journal of the American Statistical Association*, 98: pp. 598-608.

HALL P. and PRESNELL B. (1999). – « Intentionally Biased Bootstrap Methods », *Journal of the Royal Statistical Society, Series B*, 61: pp. 143-158.

HALL P., WOLFF R.C.L. and YAO Q.W. (1999). – « Methods for Estimating a Conditional Distribution Function », *Journal of the American Statistical Association*, 94: pp. 154-163.

HAMPEL F.R., RONCHETTI E.M., ROUSSEEUW P.J. and STAHEL W.A. (1986). – « *Robust Statistics: The Approach Based on Influence Functions* », *Wiley*, New-York.

KNEIP A., and SIMAR L. (1996). – « A General Framework for Frontier Estimation with Panal Data », Journal of Productivity Analysis, 7: pp. 187-212.

KOKIC P., CHAMBERS R., BRECKLING J. and BEARE S. (1997). – « A Measure of Production Performance », Journal of Business & Economic Statistics, 15(4): pp. 445-451.

KOOPMANS T.C. (1951). – « An Analysis of Production as an Efficient Combination of Activities », in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph 13, New York: John Wiley.

KUMBHAKAR S.C. and LOVELL C.A. (2000). – « *Stochastic Frontier Analysis* », *Cambridge University Press*.

MAMMEN E. and MARRON J.S. (1997). – « Mass Centered Kernel Smoothers », Biometrika, 84: pp. 765-777.

MEEUSEN W. and VAN DEN BROEK (1977). – « Efficiency Estimation from Cobb-Douglas Production Function with Composed Error », International Economic Review, 8: pp. 435-444.

MUKERJEE H. and STERN S. (1994). – « Feasible Nonparametric Estimation of Multiargument Monotone Functions », *Journal of the American Statistical Association*, Vol. 89, 425: pp. 77-80.

MULLER H.G. and SONG K.S. (1993). – « Identity Reproducing Multivariate Nonparametric Regression », *Journal of Multivariate Analysis*, 46: pp. 237-253.

PARK B., SIMAR L. and WEINER CH. (2000). – « The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties », *Econometric Theory*, Vol. 16, pp. 855-877.

PERACCHI F. (2002). – « *On Estimation Conditional Quantiles and Distribution Functions* », Computational Statistics Data Analysis, 38: pp. 433-447.

SHEPHARD R.W. (1970). – « *Theory of Cost and Production Function* », Princeton, New Jersey: *Princeton University Press*.

SIMAR L. and WILSON P.W. (2000). – « Statistical Inference in Nonparametric Frontier Models: The State of the Art », *Journal of Productivity Analysis*, 13: pp. 49-78.

YU K. and JONES M.C. (1998). – « Local Linear Quantile Regression », *Journal of the American Statistical Association*, 93: pp. 228-237.

APPENDIX: LEMMAS AND PROOFS

PROOF OF THEOREM 3.1: As is well known (see, *e.g.*, Ait Sahalia 1993, Theorem 1), under the conditions (H1)-(H2), the process $\sqrt{n}(\widehat{F}-F)$ converges in distribution to $\mathbb{F}$ in the space $C(\Psi)$ of continuous functions on the compact support Ψ. Since $C(\Psi)\subset L^{\infty}(\Psi)\subset L^{\infty}(\Psi_*)$, the process $\sqrt{n}(\hat{F}-F)$ also converges in distribution to $\mathbb{F}$ in $L^{\infty}(\Psi_*)$. Likewise $\sqrt{n}(\hat{F}-F_X)$ converges in distribution to the p dimensional F_X-Brownian bridge $\mathbb{F}(\cdot,\infty)$ in the space of bounded functions on the support of X. On the other hand, we have $\sup_{(x,y)\in\Psi_*}|\hat{\mathbb{A}}_n(x,y)-\mathbb{A}(x,y)|\leq B_n+C_n$, where

$$
\begin{aligned}
B_n &= \sup_{(x,y)\in\Psi_*}\Big|\Big[\sqrt{n}(\hat{F}(x,y)-F(x,y))-\sqrt{n}F(y\mid x)\,(\hat{F}_{X,n}(x)-F_X(x))\Big]\\
&\qquad\times\Big\{(1/\hat{F}_{X,n}(x))-(1/F_X(x))\Big\}\Big|,\\
C_n &= \sup_{(x,y)\in\Psi_*}(1/F_X(x))\Big|\Big[\sqrt{n}(\hat{F}(x,y)-F(x,y))-\sqrt{n}F(y\mid x)(\hat{F}_{X,n}(x)-F_X(x))\Big]\\
&\qquad-\big[\mathbb{F}(x,y)-F(y\mid x)\,\mathbb{F}(x,\infty)\big]\Big|.
\end{aligned}
$$

By making use of the fact that $\sup_{(x,y)\in\Psi_*}|\sqrt{n}(\hat{F}(x,y)-F(x,y))-\mathbb{F}(x,y)|\overset{P}{\to}0$ and $\sup_{(x,y)\in\Psi_*}|\sqrt{n}(\hat{F}_{X,n}(x)-F_X(x))-\mathbb{F}(x,\infty)|\overset{P}{\to}0$, as $n\to\infty$, it can be easily seen that $\sup_{(x,y)\in\Psi_*}|\hat{\mathbb{A}}_n(x,y)-\mathbb{A}(x,y)|\overset{P}{\to}0$.

LEMMA A.1. *If r and s are two functions defined on D, then*

$$\|r^{\#}-s^{\#}\|_D\leq\|r-s\|_D .$$

PROOF: Let $M=\sup_{x\in D}|r(x)-s(x)|$. The lemma will follow from the following sets of inequalities: $r^u-M\leq s^u\leq r^u+M$ and $r^l-M\leq s^l\leq r^l+M$. The right two inequalities follow from taking the $\sup_{x'\geq x}$ (resp.: the $\inf_{x'\leq x}$) in the inequality $s(x')\leq r(x')+M$, and the left ones follow from taking the $\sup_{x'\geq x}$ (resp.: the $\inf_{x'\leq x}$) in the inequality $r(x')-M\leq s(x')$.

As an immediate consequence, we show in the next lemma that the asymptotic properties of $\hat{\alpha}_n^{\#}$ are driven by those of $\hat{\alpha}_n$.

LEMMA A.2. *If the true efficiency measure $x\mapsto F(y\mid x)$ is nonincreasing on D for any $y\in S$, then* $\|\hat{\alpha}_n^{\#}-\alpha\|_{D\times S}\leq\|\hat{\alpha}_n-\alpha\|_{D\times S}$.

PROOF: For any $y \in S$, we consider the functions $r_y(.)$ and $\hat{r}_y(.)$ defined on D by

$$r_y(x) = \alpha(x,y), \quad \hat{r}_y(x) = \hat{\alpha}_n(x,y).$$

From (3) and (4) we have $\hat{r}_y^{\#}(x) = \hat{\alpha}_n^{\#}(x,y)$, and since $r_y(.)$ is nonincreasing, we also have $r_y^{\#}(x) = r_y(x)$ on D. Therefore, we obtain by Lemma A.1, for any $(x, y) \in D \times S$

$$|\hat{r}_y^{\#}(x) - r_y(x)| \leq \sup_{x \in D} |\hat{r}_y^{\#}(x) - r_y^{\#}(x)| \leq \sup_{x \in D} |\hat{r}_y(x) - r_y(x)| \leq \sup_{(x,y)\in D\times S} |\hat{r}_y(x) - r_y(x)|.$$

Thus: $\sup_{(x,y)\in D\times S} |\hat{\alpha}_n^{\#}(x,y) - \alpha(x,y)| \leq \sup_{(x,y)\in D\times S} |\hat{\alpha}_n(x,y) - \alpha(x,y)|$.

LEMMA A.3. *Assume that the kernel function* $(x,y) \mapsto H(x,y) = H_0(x)H_1(y)$ *is Lipschitz of some order* $\beta > 0$ *with a constant* $L > 0$, *and that*

(i) F has a twice continuously differentiable density with integrable second partial derivatives on $\mathbb{R}^{p+q}$*;*

(ii) $\forall i = 1, ..., p,\ \int_{\mathbb{R}} zK_0^i(z)dz = 0,\ \int_{\mathbb{R}} |z|K_0^i(z)dz < \infty,\ \int_{\mathbb{R}} z^2 K_0^i(z)dz < \infty;$

(iii) $\forall j = 1, ..., q,\ \int_{\mathbb{R}} zK_1^j(z)dz = 0,\ \int_{\mathbb{R}} |z|K_1^j(z)dz < \infty,\ \int_{\mathbb{R}} z^2 K_1^j(z)dz < \infty;$

(iv) $h_i \to 0,\ \sqrt{n}h_i^2 \to 0,\ nh_i \to \infty,$ *for* $i = 0, 1,$ *as* $n \to \infty$.

Then, for every $\theta < 1/2$, $n^{\theta} \| \hat{F} - F \|_{D\times S} \xrightarrow{co.} 0$ *as* $n \to \infty$.

PROOF: Let $\theta < 1/2$, and write $\| \hat{F} - F \|_{D\times S} \leq \| \hat{F} - E\hat{F} \|_{D\times S} + \| E\hat{F} - F \|_{D\times S}$. By applying Fubini's Theorem and a change of variables, we obtain for any $(x, y) \in \Psi$,

$$E\hat{F}(x,y) - F(x,y)$$
$$= \int_{\mathbb{R}^{p+q}} \left(\int_{-\infty}^{x} \int_{-\infty}^{y} [f(u - h_0 z_1, v - h_1 z_2) - f(u,v)]dudv \right) K_0(z_1)K_1(z_2)dz_1 dz_2.$$

By applying a Taylor expansion to $f(u - h_0 z_1, v - h_1 z_2) - f)(u,v)$ and using the conditions (i)-(iv), we conclude that $\sqrt{n} \| E\hat{F} - F \|_{\Psi} \to 0$ as $n \to \infty$. It follows that $n^{\theta} \| E\hat{F} - F \|_{D\times S} \leq \sqrt{n} \| E\hat{F} - F \|_{D\times S} \to 0$ as $n \to \infty$. It is then enough to prove that $n^{\theta} \| \hat{F} - E\hat{F} \|_{D\times S} \xrightarrow{co.} 0$ as $n \to \infty$.

Let m be an integer greater than or equal to $\frac{\theta}{\beta} + 2$. For a fixed $n \geq 1$ and each $i = 1, \ldots, p$ (resp.: each $j = 1, \ldots, q$), consider a subdivision of $[a_i, b_i]$ (resp.: of $[c_j, d_j]$) to n^m subintervals with same length $(b_i - a_i)/n^m$ (resp.: $(d_j - c_j)/n^m$). Then $D \times S$ is covered by $n^{m(p+q)}$ hypercubes $B_k (D \times S = \bigcup_{k=1}^{n^{(p+q)m}} B_k)$ such that, for any $z = (x,y) = (x^1, ..., x^p, y^1, ..., y^q) \in B_k$, we have $\left| x^i - t_{z,x}^i \right| \leq (b_i - a_i)/n^m$ for every i, and $\left|y^j - t^j_{z,y}\right| \le (d_j - c_j)/n^m$ forevery j,where $(t_{z,x}, t_{z,y}) = (t^1_{z,x},..., t^p_{z,x}, t^1_{z,y},..., t^q_{z,y})$ denotes the center of B_k.

For every ε > 0, we have the following decomposition:

$$P(n^\theta \| \hat{F} - E\hat{F} \|_{D\times S} > \varepsilon)$$
$$= P\{ \sup_{z=(x,y)\in D\times S} | (\hat{F}(x,y) - E\hat{F}(x,y)) - (\hat{F}(t_{z,x}, t_{z,y}) - E\hat{F}(t_{z,x}, t_{z,y}))$$
$$+ (\hat{F}(t_{z,x}, t_{z,y}) - E\hat{F}(t_{z,x}, t_{z,y})) | > \varepsilon / n^\theta \}$$
$$(A.1)$$
$$\le P(\sup_{z=(x,y)\in D\times S} | (\hat{F}(x,y) - \hat{F}(t_{z,x}, t_{z,y})) - (E\hat{F}(x,y)) - E\hat{F}(t_{z,x}, t_{z,y})) | > \varepsilon / 2n^\theta)$$
$$+ P(\sup_{z=(x,y)\in D\times S} | \hat{F}(t_{z,x}, t_{z,y}) - E\hat{F}(t_{z,x}, t_{z,y}) | > \varepsilon / 2n^\theta).$$

Denoting by, $\| z \|_k = (\Sigma^k_{i=1} z_i^2)^{1/2}$, the usual Euclidean norm on $\mathbb{R}^k$, we have for any $z = (x, y) \in D \times S$,

$$\left|\hat{F}(x,y) - \hat{F}(t_{z,x}, t_{z,y})\right| \le L \| (\frac{x - t_{z,x}}{h_0}, \frac{y - t_{z,y}}{h_1}) \|^\beta_{p+q}$$
$$\le L\{h_1^2 \| b - a \|_p^2 + h_0^2 \| d - c \|_q^2\}^{\beta/2} / (n^m h_0 h_1)^\beta,$$

where a, b, c, d are, respectively, the vectors of components a_i, b_i, c_j, d_j. Therefore,

$$\sup_{z=(x,y)\in D\times S} \left|(\hat{F}(x,y) - \hat{F}(t_{z,x}, t_{z,y})) - (E\hat{F}(x,y)) - E\hat{F}(t_{z,x}, t_{z,y}))\right|$$
$$\le 2L\{h_1^2 \| b - a \|_p^2 + h_0^2 \| d - c \|_q^2\}^{\beta/2} / (n^m h_0 h_1)^\beta.$$

Since $h_i \to 0$ and $nh_i \to \infty$ as $n \to \infty$, for $i = 0, 1$, and since $m \ge \frac{\theta}{\beta} + 2$, we obtain $\{h_1^2 \| b - a \|_p^2 + h_0^2 \| d - c \|_q^2\}^{1/2} / n^{(m - \frac{\theta}{\beta})} h_0 h_1 \to 0$ as $n \to \infty$. Consequently the first term on the right-hand side of inequality (A.1) is 0 for n large enough. On the other hand, if we denote by t_k the center of B_k, $k = 1,..., n^{(p+q)m}$, we obtain by using Hoeffding's Inequality, for all $n \ge 1$,

$$P(\sup_{z=(x,y)\in D\times S} \left|\hat{F}(t_{z,x}, t_{z,y}) - E\hat{F}(t_{z,x}, t_{z,y})\right| \varepsilon / 2n^\theta)$$
$$\le \sum_{k=1}^{n^{(p+q)m}} P(\left|\hat{F}(t_k) - E\hat{F}(t_k)\right| > \varepsilon / 2n^\theta) \le 2n^{(p+q)m} e^{-n\varepsilon^2 / 2n^{2\theta}}.$$

Since $\theta < 1/2$, $\log(n)/n^{(1-2\theta)} \to 0$ as $n \to \infty$. Then there exist an integer N_* such that $\log n / n^{(1-2\theta)} \le \varepsilon^2 / 2[(p+q)m+2]$ for all $n > N_*$, which yields

$$\log(n^{(p+q)m}) - n^{(1-2\theta)}\varepsilon^2 / 2 \le -2\log n.$$

Consequently $n^{(p+q)m}e^{-n\varepsilon^2/2n^{2\theta}} \le n^{-2}$ for all $n > N_*$. Hence $\Sigma_{n=1}^{\infty} n^{(p+q)m}e^{-n\varepsilon^2/2n^{2\theta}}$ is finite, and thus $n^{\theta} \| \hat{F} - E\hat{F} \|_{D\times S}$ converges completely to 0 as $n \to \infty$.

LEMMA A.4. *Assume that H_0 is Lipschitz and F_X has a twice continuously differentiable density with integrable second partial derivatives on $\mathbb{R}^p$, and that the condition (ii) of Lemma A.3 holds with $\sqrt{n}h_0^2 \to 0$ and $nh_0 \to \infty$ as $n \to \infty$. Then, for every $\theta < 1/2$, $n^{\theta} \| \hat{F}_X - F_X \|_D \overset{co.}{\to} 0$ as $n \to \infty$.*

To prove Lemma A.4, it suffices to adapt the proof of Lemma A.3 by restricting ourself on the domain D instead of $D \times S$, and by considering respectively the functions $x \mapsto F_X(x)$ and $x \mapsto H_0(x)$ in place of $(x,y) \mapsto F(x,y)$ and $x \mapsto H(x,y)$.

Using Lemmas A.2-A.4, we are now in a position to prove the complete uniform convergence property of $\hat{\alpha}_n^{\#}$.

PROOF OF THEOREM 3.2: By Lemma A2, we have

$$n^{\theta} \| \hat{\alpha}_n^{\#} - \alpha \|_{D\times S} \le [n^{\theta} \| \hat{F} - F \|_{D\times S} + n^{\theta} \| \hat{F}_X - F_X \|_D] / \inf_{x\in D} \hat{F}_X(x).$$

Since the term between brackets converges completely to 0 from Lemmas A.3 and A.4, to complete the proof it suffices to show that there exists a constant $\delta > 0$ such that $\Sigma_{n=1}^{\infty} P(\inf_{x\in D} \hat{F}_X(x) \le \delta) < \infty$. By a simple computation, we get $\left|\inf_{x\in D} \hat{F}_X(x) - \inf_{x\in D} F_X(x)\right| \le \| \hat{F}_X - F_X \|_D$. Therefore, by Lemma A.4, $\inf_{x\in D} \hat{F}_X(x)$ converges completely to $\inf_{x\in D} F_X(x) > 0$. Hence, for every $\delta > 0$, $\Sigma_{n=1}^{\infty} P(\left|\inf_{x\in D} \hat{F}_X(x) - \inf_{x\in D} F_X(x)\right| \ge \delta) < \infty$. It follows that

$$\sum_{n=1}^{\infty} P(\inf_{x\in D} \hat{F}_X(x) \le \inf_{x\in D} F_X(x) - \delta) < \infty.$$

Thus, we conclude by putting $\delta = \inf_{x\in D} F_X(x)/2$.