I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0324

STATISTICAL INFERENCE FOR AGGREGATES OF FARRELL-TYPE EFFICIENCIES

SIMAR, L. and V. ZELENYUK

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

Statistical Inference for Aggregates of Farrell-type Efficiencies

Léopold Simar* **and Valentin Zelenyuk****

Abstract

In this study, we merge results of two recent directions in efficiency analysis research—the Aggregation and the Bootstrap—applied, as an example, to one of the most popular point-estimators of individual efficiency: the Data Envelopment Analysis (DEA) estimator. A natural context of the methodology developed here is a study of efficiency of a particular economic system (e.g., an industry) as a whole, or a comparison of efficiencies of distinct groups within such a system (e.g., private vs. public or regulated vs. non-regulated firms, etc). Our methodology is justified by the (neo-classical) economic theory and is supported by carefully adapted statistical methods.

15 October, 2003

* Institut de Statistique, Université Catholique de Louvain, Belgium

** EERC (Economics Education and Research Consortium) and National University "Kyiv-Mohyla Academy", Kyiv (Kiev), Ukraine

Acknowledgement:
We would like to thank Tom Coupé and Natalia Konstandina, as well as participants of seminars/workshops at the Institut de Statistique, Université Catholique de Louvain, EERC (Ukraine), IAMO (Germany), and 8th European Workshop of Productivity and Efficiency Analysis (Oviedo, Spain) for their fruitful comments and interest that have stimulated our work. Research support from "Projet d'Actions de Recherche Concertées" (No.98/03-217) and from the "Interuniversity Attraction Pole", Phase V (No.P5/24) from the Belgian Government are also acknowledged.

1. Introduction

Many applied economists devote their attention to analyzing economic *efficiency* of various economic systems: firms, industries, countries, regions, etc. After performing various techniques for estimating efficiencies of individual units (say, firms) of a system, most researchers inevitably come to a question like 'What is the efficiency of the *entire* system (the industry)?' or 'What are the efficiencies of distinct groups within this system? Which one is more efficient?' For example, many researchers have recently analyzed and compared efficiencies of *groups* of firms operating under different regulatory regimes, or different ownership structures (private vs. public, domestic vs. foreign, etc.), or operating in different regions, or efficiencies of countries at different stages of economic development or transition. The answers to such questions are of high importance not only to researchers but also, perhaps even more important, for policy makers, for voters, for educators in related areas, for many others. Sometimes, economic theory cannot give precise general answer on which group of firms must be more efficient in a particular environment, or this may depend on various unobserved conditions, or different models may suggest different conclusions. All this makes the *empirical* studies on efficiency of various groups and sub-groups interesting and important, demanding reliable methods of estimation and inference.

For comparison of efficiencies of different groups—the context of our study—there are at least two critical issues about the appropriateness of methodology: (i) reliable *point*-estimators of group (or sub-group) efficiencies, and (ii) reliable *interval*-estimators of group efficiencies. The first issue can be viewed as an aggregation question—a question of obtaining an (appropriate) *aggregate* efficiency score from (appropriate) *individual* efficiency scores. This question has been recently explored in a number of studies.[1] One of the most important issues here is the choice of *weights* in the aggregation. For example, the answer for such an important question as: 'What is the more efficient way of regulation used in practice?' may merely depend on the (researchers) *choice* of weights that are distributed among the estimated efficiencies of each economic unit in the system. Consider for example, a hypothetical industry consisting of two types of firms, two firms in each type, whose efficiency and an economic weight is given in Table 1.

According to this example, if a researcher uses the simple average (as many studies have been doing in practice) then the conclusion would be that, on average, group A is as efficient as the group Z. Noting that the efficiency scores per se are 'standardized' to be between zero and 1 and

[1] See Blackorby and Russell (1999), Färe and Zelenyuk (2003), Färe, Grosskopf and Zelenyuk (2002), Li and Ng (1995), and Ylvinger S., (2000). Also see the latter source for a review of earlier studies.

thus ignore the relative effort or (economic) importance of firm that earned this score, another researcher may want to use the *weighted* average. In this case, a dramatically different conclusion would be reached: group A is more efficient than Z, but the industry average is still very low because the type Z firms dominate. The policy implications would differ dramatically.

Table 1. A hypothetical example

Firms in A (10%)	Weight in sub-group	Efficiency (%)	Firms in Z (90%)	Weight in sub-group	Efficiency (%)	Efficiency of Entire Group
A1	90%	100%	*Z1*	10%	100%	--
A2	10%	50%	*Z2*	90%	50%	--
Simple Average		**75%**	**Simple Average**		**75%**	**75%**
Weighted Average		**95%**	**Weighted Average**		**55%**	**59%**

Of course, the main question here is like that of A. Griboyedov's play, *Woe from Wit*: "And who are the judges…?"—or in our case "And what are the weights?" Clearly, a strong justification for choice of weights is needed. One of such justifications for Farrell-type efficiencies was recently proposed by Färe and Zelenyuk (2003) and is based on economic optimization. The resulting group efficiency measure that they derived is the average of the efficiencies of individual units *weighted* by their realized shares (cost or revenue, depending on optimization assumed) in the industry. This result gives to applied researchers a *point-estimator* of group efficiency with meaningful weights derived from economic principles. Here, we extend their result to aggregation *within* and *between* sub-groups in a given group.

The main goal of this paper is to propose a way of constructing reliable confidence intervals and bias corrections for the DEA-estimated *aggregate* efficiencies of a group and also its sub-groups, as well as to propose an appropriate test for comparison of such *aggregate* efficiencies.

The choice of DEA is not necessary but is motivated by its increasing popularity, especially since some good statistical properties of DEA have been recently unveiled. This includes most recent discovery of the limiting distribution by Kneip et al. (2003a), who also prove the consistency of the sub-sampling bootstrap for DEA estimator of individual efficiency scores.

The goal of this paper is to merge the existing works on bootstrap for DEA with the works on *aggregation* of efficiency scores—to provide researchers with a theoretically appropriate and reliable *practical* tool of statistical inference on the size of *aggregated* efficiency scores and for their comparison between each other.

2. Efficiency Measurement

2.1. Measurement of Individual efficiency

The methodology developed below can be used to analyze virtually any economic system that can be viewed as a composition of several units. To facilitate our discussion we consider an example of an industry with *n* firms. For each firm *k* (*k = 1, 2, … , n*) we will use vector $x^k = (x_1^k,..., x_N^k)' \in \Re_+^N$ to denote *N inputs* that the firm *k* uses to produce a vector of *M outputs*, denoted by $y^k = (y_1^k,..., y_M^k)' \in \Re_+^M$. We assume the technology of firm *k* can be characterized by the set T^k,

$$T^k \equiv \{(x^k, y^k) \ : \ x^k \ can \ produce \ y^k\}. \qquad (2.1.1)$$

Equivalently, the technology can be characterized by the *output sets*

$$P^k(x^k) \equiv \{y^k : \ x^k \ can \ produce \ y^k\}, \qquad x^k \in \Re_+^N. \qquad (2.1.2)$$

Throughout, we assume the technology satisfies the usual regularity axioms of production theory, under which we can use the *output orientated* Shephard (1970) distance function $D_o^k : \Re_+^N \times \Re_+^M \to \Re_+^1 \cup \{\infty\}$, defined as

$$D_o^k(x^k, y^k) \equiv \inf\{q : \ y^k / q \in P^k(x^k)\} \qquad (2.1.3)$$

to have a complete (*primal*) characterization of the technology of firm *k*, in the sense that

$$D_o^k(x^k, y^k) \le 1 \quad \Leftrightarrow \quad y^k \in P^k(x^k) \qquad (2.1.4)$$

This function is particularly convenient as a criterion for technical efficiency of a firm *k* since, roughly speaking, it gives a 'measure' (valued between 0 and 1) of a distance from a point y^k in $P^k(x^k)$ to the 'upper' boundary of $P^k(x^k)$. Such efficiency criterion often appears in another form, as the *Farrell* output oriented measure of *technical efficiency*, defined for all $y^k \in P^k(x^k)$ as[2]

$$TE^k(x^k, y^k) \equiv \max\{q : \ q \ y^k \in P^k(x^k)\} = 1 / D_o^k(x^k, y^k). \qquad (2.1.5)$$

Formally, if we let the technological frontier to be the 'upper' boundary of $P^k(x^k)$ defined as $\partial P^k(x^k) = \{ y \in \Re_+^M \ : \ y \in P^k(x^k), \ \ ly \notin P^k(x^k), \ \ \forall l \in (1,\infty)\}$ then, whenever we have $0 < D_o^k(x^k, y^k) < 1 \ \Leftrightarrow \ y^k \in P^k(x^k), y^k \notin \partial P^k(x^k), y^k \neq 0$, we would call (x^k, y^k) as *technically* inefficient, with inefficiency score given by (2.1.5) (or its reciprocal). Alternatively, we

[2] Farrell (1957) originally used *input* orientation, conceptually the same idea. Similar idea also appeared in Debreu (1951), as a capacity utilization measure. See Russell (1990) for properties of this 'measure'.

call (x^k, y^k) *technically* efficient if and only if $D_o^k(x^k, y^k) = 1 \Leftrightarrow y^k \in \partial P^k(x^k)$. Finally, $D_o^k(x^k, y^k) = 0$, if and only if $y^k = 0$ (which is usually not the case in practice).

An alternative, *dual*, characterization of $P^k(x^k)$ can be given via the *revenue* function,

$$R^k(x^k, p) \equiv \max_y \{py : y \in P^k(x^k)\}. \qquad (2.1.6)$$

where $p = (p_1, \ldots, p_M) \in \mathbb{R}^M_{++}$ denotes the vector of output prices.[3] As it is well known, the revenue function, $R^k(x^k, p)$, is the *dual* to the distance function $D_o^k(x^k, y^k)$, since[4]

$$D_o^k(x^k, y^k) = \sup_p \{py^k : R^k(x^k, p) \le 1\} \qquad (2.1.7)$$

A natural criterion of efficiency of a firm in the *dual* framework is what is often called the *revenue* (or overall output) efficiency and is defined as,

$$RE^k(x^k, y^k, p) \equiv R^k(x^k, p) / py^k. \qquad (2.1.8)$$

The Mahler's inequality (which can be obtained from *(2.1.7)*) tells us that

$$R^k(x^k, p) \ge py^k / D_o^k(x^k, y^k), \qquad (2.1.9)$$

and the multiplicative residual that closes the inequality (2.1.9) is often interpreted as the criterion or a measure of the *allocative* (in)efficiency of firm k, and is formally defined as

$$AE^k(x^k, y^k, p) \equiv RE^k(x^k, y^k, p) / TE^k(x^k, y^k). \qquad (2.1.10)$$

The decomposition (2.1.10) goes back at least to Farrell (1957) and will prove very useful in deriving the results for aggregating the technical efficiencies into (sub)group measures.

2.2. Group Efficiency Measures

Let us focus first on a *sub-group*, call it sub-group l, of n_l firms taken from the original group of n firms (e.g., the selection can be based on exogenous economic criterion such as ownership structure, regulation regimes, etc). We will denote the input allocation among firms *within* the group l by $X^l = (x^{l,1}, \ldots, x^{l,n_l})$ and the *sum* of output vectors over all firms in l^{th} group with $\overline{Y}^l = \sum_{k=1}^{n_l} y^k$.

[3] For the purpose of obtaining the desired aggregation results we have make a *necessary* assumption that all firms face the same *output* prices.

[4] To achieve this result, convexity of the output sets is needed, in addition to other regularity axioms mentioned above; see Färe and Primont (1995) for details.

A crucial step now is to define a *group technology*, that is, the *aggregate* technology of *all* firms within a (sub)group. In the context we have chosen—the *output orientation*, i.e., consideration of output changes *given fixed* levels of inputs—a natural way to define a (sub)group technology is to assume a linear structure of aggregation of the output sets (Färe and Zelenyuk, 2003), i.e.,

$$\overline{P}^{l}(X^{l}) \equiv \sum_{k=1}^{n_l} P^{l,k}(x^{l,k}) \qquad (2.2.1)$$

Thus, the *output set* of a sub-group of firms, $\overline{P}^{l}(X^{l})$, is the sum of the individual *output sets* of all firms in this sub-group. Clearly, the properties of the group technology depend on the properties of technologies of each firm in the group. In particular, $\overline{P}^{l}(X^{l})$ inherits the regularity conditions imposed above and is convex if the individual output sets are convex.

Given the l^{th} sub-group technology (2.2.1), the *sub-group revenue* function can be defined as

$$\overline{R}^{l}(X^{l}, p) \equiv \max_{y}\{py \;:\; y \in \overline{P}^{l}(X^{l})\} \qquad (2.2.2)$$

and the l^{th} sub-group revenue *efficiency*, analogue of (2.1.9) is therefore defined as

$$\overline{RE}^{l}(X^{l}, \overline{Y}^{l}, p) \equiv \overline{R}^{l}(X^{l}, p) / p\overline{Y}^{l}. \qquad (2.2.3)$$

The next theorem and its corollaries give the aggregation results needed for our study.

Theorem.[5] The maximal revenue of the sub-group of firms is equal to the sum of the maximal revenues of all its member firms, i.e.,

$$\overline{R}^{l}(X^{l}, p) = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p). \qquad (2.2.4)$$

This theorem (as well as the first two corollaries below) is from Färe and Zelenyuk (2003), adapted to our context, and for the sake of completeness, the proof is provided in the appendix. The economic intuition of this theorem is straightforward. It says that the sum of the revenues of individual (independent) revenue-maximizing firms in a given subgroup would be the same as the revenue obtained by *one* revenue-maximizing firm (e.g., a revenue-maximizing social planner) whose technology is defined in (2.2.1), given that the output price vector is the same for all firms. In the next corollary, we will use this theorem to obtain some results for aggregating efficiencies.

[5] This theorem is a revenue analog to the Koopmans (1957) theorem of aggregation of the profit functions. The cost analog is proven in Färe, Grosskopf and Zelenyuk (2002).

Corollary 1. The revenue efficiency of the l^{th} sub-group of firms is equal to the *weighted sum* of revenue efficiencies of all its member firms, where the weights are the *actual* (observed) revenue shares of these firms in the sub-group, i.e.,

$$\overline{RE}^{l}(X^{l}, \overline{Y}^{l}, p) \;=\; \sum_{k=1}^{n_l} RE^{l,k}(x^{l,k}, y^{l,k}, p)\cdot S^{l,k}, \qquad (2.2.5)$$

where

$$S^{l,k} = py^{l,k} / p\overline{Y}^{l}, \qquad k = 1, \ldots, n_l. \qquad (2.2.6)$$

Corollary 2. The aggregate revenue efficiency can be decomposed into the weighted sum of the *technical* efficiencies (where the weights are the actual revenue shares) and the weighted sum of the *allocative* efficiencies (where the weights are the revenue shares corrected for inefficiency). I.e.,

$$\overline{RE}^{l}(X^{l}, \overline{Y}^{l}, \; p) \;=\; \overline{TE}^{l} \;\times\; \overline{AE}^{l} \qquad (2.2.7)$$

where

$$\overline{TE}^{l} \equiv \sum_{k=1}^{n_l} TE^{l,k}(x^{l,k}, y^{l,k})\cdot S^{l,k} \qquad (2.2.8)$$

and

$$\overline{AE}^{l} \equiv \sum_{k=1}^{n_l} AE^{l,k}(x^{l,k}, y^{l,k}, p)\cdot S_{ae}^{l,k}, \qquad (2.2.9)$$

where

$$S^{l,k} \equiv \frac{py^{l,k}}{p\overline{Y}^{l}}, \quad S_{ae}^{lk} \equiv \frac{p\left(y^{l,k} TE^{l,k}(x^{l,k}, y^{l,k})\right)}{p\sum_{k=1}^{n_l}\left(y^{l,k} TE^{l,k}(x^{l,k}, y^{l,k})\right)}, \quad k = 1, \ldots, n_l. \qquad (2.2.10)$$

Remark 1. Note that if $L = 1$, then the aggregate measures in (*2.2.7*)-(*2.2.*9) are measures for the entire group. Moreover, the measure (*2.2.*8) is a *multi-output* generalization of what Farrell called the "*structural efficiency of an industry*," (Farrell, 1957, p. 261-262).

Remark 2 Note that the weights of aggregation for obtaining the (sub)group *technical* efficiency derived above *depend* on prices. This may seem somewhat undesirable for at least two reasons. First of all, the *technical* efficiency, at least in principle, is often thought of as a price independent measure of efficiency, e.g., as in our disaggregate case or single-output aggregate case. Note however that these weights were *not* chosen arbitrarily or in an ad hoc way, but came out as a result of imposing a standard economic criterion—optimization behavior—which researchers, at least implicitly, consider when making their choice of orientation (input, output, etc) in measuring efficiency. Moreover, if the goal is to account for an *economic* weight of each firm, its relative *economic* effort in earning the particular 'standardized' efficiency score then, since prices contain some *economic* information, it must not be surprising to derive the price-dependent weights from imposing the *economic* optimization principle. The second consideration is more practical: price

information may be unavailable (or unreliable) in a given study. One way around this is to use the shadow prices instead. Another way is to impose some additional standardization that will make the weights derived above being price-independent (as we will do below).

Remark 3. It is not the first time that *positive* aggregation results in economics requires some additional, often strong and perhaps sometimes undesirable assumptions (e.g., the reader can recall assumptions needed for aggregation of demands over consumers or over goods). In fact, in a more general context of aggregating efficiencies (without optimization criterion as in our case) Blackorby and Russell (1999) have shown an *impossibility* result for their general case and the need of quite strong assumptions on the technology in special cases.

Let us now consider a case when (given some exogenous economic criterion) a researcher is interested in comparing aggregate efficiencies of certain sub-groups *within* the entire group and relative to the entire group. In particular, consider a case of partitioning the entire group into L non-intersecting and exhaustive sub-groups, indexed by $l = 1, \ldots, L$, and let $\overline{Y} \equiv \sum_{k=1}^{n} y^k$. An immediate consequence of the previous theorem would be the following result.

Corollary 3. The maximal revenue of the entire group of firms is equal to the sum of maximal revenues of *all* its (non-intersecting) sub-groups of firms, i.e.,

$$\overline{R}(X, p) = \sum_{l=1}^{L} \overline{R}^l(X^l, p). \qquad (2.2.11)$$

This is an analog of the theorem 1 (extension to aggregation *between* the sub-groups) and the next corollary is the corresponding analog of corollary 1.

Corollary 4. The revenue efficiency of the entire group of firms is equal to the *weighted sum* of revenue efficiencies of all its (non-intersecting) sub-groups of firms, where the weights are the *actual* revenue shares of these sub-groups in the entire group, i.e.,

$$\overline{RE}(X, \overline{Y}, p) = \sum_{l=1}^{L} \overline{RE}^l(X^l, \overline{Y}^l, p) \cdot S^l, \qquad (2.2.12)$$

where

$$S^l = p\overline{Y}^l / p\sum_{l=1}^{L} \overline{Y}^l, \qquad l = 1, \ldots, L. \qquad (2.2.13)$$

Finally, the 'between-the-group aggregation' analog of the corollary 2 is given in the next corollary.

Corollary 5. The revenue efficiency of the entire group can be decomposed into the weighted sum of the *sub-group* technical efficiencies of all non-intersecting sub-groups (where the weights are the actual revenue shares of these sub-groups) and the weighted sum of the *sub-group* allocative efficiencies of all non-intersecting sub-groups (where the weights are the actual revenue shares corrected for the sub-groups technical inefficiency). Formally,

$$\overline{RE}(X,\overline{Y},p) \;=\; \overline{TE} \;\times\; \overline{AE} \qquad (2.2.14)$$

where

$$\overline{TE} = \sum_{l=1}^{L} \overline{TE}^{\,l} \cdot S^{l}, \qquad \overline{AE} = \sum_{l=1}^{L} \overline{AE}^{\,l} \cdot S^{l}_{ae}, \qquad (2.2.15)$$

and

$$S^{l} = p\overline{Y}^{\,l} \,/\, p\sum\nolimits_{l=1}^{L} \overline{Y}^{\,l}, \quad \text{and} \quad S^{l}_{ae} = p\overline{Y}^{\,l}\,\overline{TE}^{\,l} \,/\, p\sum\nolimits_{l=1}^{L} \overline{Y}^{\,l}\,\overline{TE}^{\,l}, \; l = 1, \ldots, L. \qquad (2.2.16)$$

That is, the efficiencies of the sub-groups of firms are aggregated into efficiencies of the entire group much like the efficiencies of individual members of the sub-group are aggregated into the sub-group efficiencies.

Price Independent Weights

Here, we adopt the standardization proposed by Färe and Zelenyuk (2003) for making the weights derived above being price independent, while still preserving the aggregation structure based on the economic optimization criterion that we have used above. Let us first illustrate this for the case of aggregating efficiencies of the *entire* group. The trick is based on the following 'standardization':

$$p_m \overline{Y}_m = A, \qquad m = 1, \ldots, M \qquad (2.2.17)$$

where $\overline{Y}_m \equiv \sum_{k=1}^{n} y_m^k$ and A is a positive constant. Intuitively, this standardization means that, in cases when we do not want the price information to impact the aggregate measure of *technical* efficiency (or simply when we do not have such information) we choose to regard that each output is valued, on industry level, as any other output among $m = 1, \ldots, M$. If instead we were interested in valuing each output differently—the price-dependent weights we have obtained above would be a natural candidate. Also note that, incidentally, expression (2.2.17) is the output analog to what Cornes (1992, p.42) uses to illustrate the duality theory in economics (for $A=1$).

Now, denote the firm's *k* share in the group in terms of m^{th}-output as $v_m^k = y_m^k / \bar{Y}_m$, then the standardization imposed by (2.2.17) onto the weights for aggregation of technical and revenue efficiencies derived above will be the *average* of these shares, i.e.,

$$S^k = \frac{1}{M}\sum_{m=1}^{M} v_m^k, \qquad k = 1, \ldots, n. \qquad (2.2.18)$$

Analogously, the price-independent weights for aggregating allocative efficiencies are similar to the weights in aggregating technical efficiencies but, as before, the actual outputs are replaced with their *technically-efficient* prototypes, i.e.,

$$S_{ae}^k = \frac{1}{M}\sum_{m=1}^{M} \frac{y_m^k TE^k(x^k, y^k)}{\sum_{k=1}^{n} y_m^k TE^k(x^k, y^k)}, \qquad k = 1, \ldots, n. \qquad (2.2.19)$$

We can now also use the standardization (2.2.17) and combine with (2.2.16) to obtain the '*between* the sub-groups' weights as

$$S^l = \frac{1}{M}\sum_{m=1}^{M} W_m^l, \qquad S_{ae}^l = \frac{1}{M}\sum_{m=1}^{M} \frac{\bar{Y}_m^l \overline{TE}^l}{\sum_{l=1}^{L} \bar{Y}_m^l \overline{TE}^l}, \qquad l = 1, \ldots, L. \qquad (2.2.20)$$

where, $W_m^l = \bar{Y}_m^l / \bar{Y}_m$ is the *l's sub-group* share in the entire group in terms of m^{th}-output (analogous to what we had for the individual firms). This in turn helps getting the weight '*within* a sub-group *l*' for an individual efficiency of firm *k* to be

$$S^{l,k} = \frac{1}{M}\sum_{m=1}^{M} \frac{y_m^{l,k}}{\bar{Y}_m^l \cdot S^l}, \qquad k = 1, \ldots, n_l \; ; \; l = 1, \ldots, L.$$

Intuitively, it is exactly what we had in (2.2.18) except that we now account for the weight of the particular group in the entire group. The analog of this for the allocative inefficiency is

$$S_{ae}^{l,k} = \frac{1}{M}\sum_{m=1}^{M} \frac{y_m^{l,k} TE^{l,k}(x^{l,k}, y^{l,k})}{\sum_{k=1}^{n_l} y_m^{l,k} TE^{l,k}(x^{l,k}, y^{l,k}) \cdot S^l}, \qquad k = 1, \ldots, n_l \; ; \; l = 1, \ldots, L$$

In the next section, we discuss the means of estimation of the above-presented measures.

2.3. The DEA Point-Estimator

In the previous section we have outlined the *theoretical* measures of individual and aggregate efficiencies. All these measures require the knowledge of $TE^k(\cdot)$ or/and $R^k(\cdot)$, or at least their

values at (x^k, y^k) for all firms $k = 1, \ldots, n$, whose efficiencies are of interest. In turn, obtaining such information requires knowledge of the technology characterization, for example in terms of $P^k(x^k)$, or in terms of its frontier. In practice, such information is unlikely to be available and an appropriate estimation method is needed. In this study we will focus on the class of estimators known under the general name—Data Envelopment Analysis (DEA). There are many variations in this class, all intending to estimate the technological *frontier* of some set-wise characterization of the technology and then compute a point estimate of efficiency scores for each observation, relative to this estimated frontier. Here, for the sake of brevity, we will focus only on the most common DEA model (that uses output orientation, assumes variable returns to scale, and free disposability of inputs and outputs) and only on the estimation (and bootstrap) of the *technical* efficiency. The methodology, however, can be extended to other cases.

One fundamental assumption in most DEA estimations is that all firms have access to the *same* technology, which we will denote as $P(x)$ or T.[6] This is needed to justify the estimation of *one frontier* from the entire data, often called the (observed) *best practice frontier*. Another fundamental assumption of the DEA estimator is that all observed input-output combinations (x^k, y^k), $k = 1, \ldots, n$ are *feasible* under T, i.e., $y^k \in P(x^k)$, $k = 1, \ldots, n$. This assumption implicitly assumes no errors and all deviations from the frontier are assumed to be due to technical inefficiency (however, the data is allowed to be random; see below for statistical assumptions).

With these assumptions and allowing for variable returns to scale and free disposability of inputs and outputs, the (observed) *best practice frontier* under DEA is defined as

$$\partial\hat{P}(x) = \{y \in \Re_+^M \;:\; y \in \hat{P}(x), \quad ly \notin \hat{P}(x), \quad l \in (1, \infty)\}, \qquad (2.3.1)$$

where

$$\hat{P}(x) = \{y \in \Re_+^M \;:\; \sum_{k=1}^{n} z_k y_m^k \geq y_m, \;\; m = 1, ..., M, \;\; \sum_{k=1}^{n} z_k x_n^k \leq x, \;\; n = 1, ..., N,$$
$$z_k \geq 0, \;\; k = 1, \ldots, n, \;\; \sum_{k=1}^{n} z_k = 1 \;\}. \qquad (2.3.2)$$

Thus, $\hat{P}(x)$ is the smallest convex free-disposal hull that fits the observed data, and $\partial\hat{P}(x)$ is its 'upper' boundary and is a *piece-wise linear* estimate of the true best-practice frontier of $P(x)$.

The DEA estimator of individual *technical* efficiency at a fixed point (x, y), is computed relative to this estimated frontier—as a solution to the following linear programming problem (LPP)

$$\hat{TE}(x,y) = \max_{q,z_1,\ldots,z_n} \{q \quad : \; qy \in \hat{P}(x)\} \qquad (2.3.3)$$

Finally, the DEA estimator of the technically efficient level ('à la Farrell') of output at a particular level of input x, is defined as

$$\hat{y}^{\partial}(x) \equiv y \cdot \hat{TE}(x,y) \qquad (2.3.4)$$

While these DEA estimators can be applied to any point in the estimated output set, i.e., $\forall\, y \in \hat{P}(x)$, researchers usually are interested in the *observed* points, and thus apply the corresponding liner programming problem (2.3.3) for each (x^k, y^k), $k = 1, \ldots, n$. In the next section we discuss the statistical issues of this basic DEA estimator.

3. Known Statistical Results for the DEA Estimator

First of all, it must be clear that $\hat{P}(x) \subseteq P(x)$, and therefore $\partial\hat{P}(x)$ is a *downward biased* estimator of $\partial P(x)$. As a result, $\hat{TE}(x,y)$ is a downward biased estimator of $TE(x,y)$, i.e.,

$$1 \le \hat{TE}(x,y) \le TE(x,y), \qquad \forall\, y \in \hat{P}(x) \qquad (3.1)$$

The asymptotic statistical properties have recently been discovered for the DEA estimator presented above. In particular, (2.3.2) is consistent and is the maximum-likelihood estimator of the frontier of *P(x)*, as shown by Korostelev et al. (1995) and generalized by Kneip et al. (1998), who also derived the rates of convergence. Gijbels et al. (1999) provided the limiting distribution of DEA in the 1-input-1-output case and most recently, Kneip et al., (2003a) have unveiled it for the multi-output-multi-input case.

These statistical results require additional assumptions that help defining the data generating process (DGP) and converting our economic model of production into a statistical model. Before listing these axioms, we represent the problem by using the *polar* coordinates of $y \in \Re_+^M$ defined by the modulus $w = w(y) \in \Re_+^1$, where $w(y) \equiv \sqrt{y'y}$, and the angle $h \equiv h(y) \in [0, p/2]^{M-1}$, where $h_m \equiv \arctan(y_{m+1} / y_1)$ *if* $y_1 > 0$ or $h_m \equiv p/2$, if $y_1 = 0$ for $m = 1, \ldots, M$. The following assumptions, adapted to our context from Kneip et al. (1998), define the DGP we will work with.

---

[6] Standard regularity conditions must also be imposed, see Färe and Primont (1995) for details.

A1. $\{(x^k, y^k) : k = 1,...,n\}$ are *independent* random variables on the *convex* technology set *T*. All observations $\{(x^k, y^k) : k = 1,...,n\}$ can be partitioned into *L sub-samples* (by some exogenous criterion) such that each sub-sample *l* (*l = 1, …, L*) represents a distinct *sub-group l* (*l = 1, …, L*) of interest that exists in the population (e.g., public vs. private firms in an industry).

A2. For all *l* (*l = 1,…,L*), the inputs $x \in \Re_+^N$ has density $f_{x,l}(x)$, with compact support $\aleph \subseteq \Re_+^N$.

A3. For all *l* (*l = 1,…,L*) and all $x \in \aleph$, the vector $h \equiv (h_1,...,h_{M-1})$ has a conditional p.d.f. $f_{(h|x),l}(h \mid x)$ on $[0, \pi/2]^{M-1}$ and the modulus w has a conditional p.d.f. $f_{(w|h,x),l}(w \mid h, x)$.

A4. For all *l* (*l = 1,…,L*), all $x \in \aleph$, and all $h \in [0, \pi/2]^{M-1}$ there exist constants $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ such that $\forall w \in [w(y^\partial(x)),\ w(y^\partial(x)) + \varepsilon_2]$, $f_{(w|h,x),l}(w \mid h, x) \geq \varepsilon_1$, *l = 1, …, L*.

A5. The technical efficiency measure $TE(x, y)$ is differentiable in both its vectors.

Note that for all *l* and given (h, x), the efficient output level $y^\partial(x)$ has the modulus equal to

$$w_l(y^\partial(x)) = \sup\{w \in \Re_+^1 \ : \ f_{(w|h,x),l}(w \mid h, x) > 0\}, \qquad l = 1, \ldots, L, \qquad (3.2)$$

so that the relation between $w_l(y)$ and the technical efficiency measure at (*x, y*), $TE^l(x, y)$, is now

$$TE^l(x, y) = w_l(y^\partial(x)) / w(y), \qquad l = 1, \ldots, L. \qquad (3.3)$$

Thus, A3 along with (3.3) implies the existence of a conditional (on (h,x)) density for $TE^l(x, y)$, *l = 1, …, L* (with the support $[1,\infty)$), which we will denote with $f_l(TE \mid h, x)$. Moreover, A4 along with (3.3), implies that $f_l(TE \mid h, x) \geq \varepsilon_1, \forall TE \in [1, 1+\varepsilon_2]$, *l = 1, …, L*. Finally, with assumptions A1-5, the DGP, denoted with $\wp = \wp(P(x), g_l(TE, h, x), l = 1,...,L)$ is completely defined through the *joint* densities of (TE, h, x), for all sub-groups *l = 1, …, L*.

$$g_l(TE, h, x) = f_l(TE \mid h, x) f_{(h|x),l}(h \mid x) f_{x,l}(x), \qquad l = 1, \ldots, L, \qquad (3.4)$$

each with the support $\Omega \equiv [1,\infty) \times [0, \pi/2]^{M-1} \times \aleph$. It is this DGP that we assume has generated our sample $\Xi_n = \{(x^k, y^k) \ : \ k = 1,...,n\}$ of independent observations that are identically distributed *within* each sub-group *l* (*l = 1, …, L*) but not necessarily *across* them.

With these assumptions, following Kneip et al., (1998), the DEA estimator is consistent, and

$$\hat{TE}^{l}(x, y) - TE^{l}(x, y) = O_p(n^{-2/(M+N+1)}). \qquad (3.5)$$

Note that such DGP still assumes that *all* firms have *access* to the *same* technology, but the conditions of this access ("how easy it is to get to the frontier") might be different for different sub-groups. In economic terms, such DGP can be well justified. Different sub-groups may have considerably different regulation regimes, ownership structures, environments, etc—some exogenous factors that may cause *systematic* differences in economic *incentives* or just 'physical' capabilities of firms in different sub-groups to reach the *frontier* of the *same* technology. For example, private firms may have different incentives for being closer to the frontier than state-owned firms; firms under average-cost pricing regulation may have, theoretically, different incentives than firms under rate-of-return regulation, which are in turn different from unregulated firms. In all these cases the marginal densities that generate technical (in)efficiency (as well as densities generating inputs and outputs) for these firms might be different across sub-groups, while the technology is still the same. All of this provides an intuitive justification of the *group-wise heterogeneous bootstrap* for the DEA estimates of a common technology frontier.

4. Bootstrap for Aggregate Efficiency Scores from the DEA Estimator

Statistical bootstrap is a method of estimation of unknown sampling distribution of an estimator by means of re-sampling from original data. The theory of statistical bootstrap was originated by Efron (1979) and developed in many studies since then.[7] Perhaps the most encouraging result from the general bootstrap theory is that under fairly moderate assumptions on the DGP, the bootstrap provides approximation to the *unknown* sampling distribution that is at least as good as the approximation given by the first-order asymptotic theory. It can give even better approximation if the estimator is asymptotically pivotal (i.e., if the asymptotic distribution of the estimator of interest is independent from the unknown population parameters). For the case where the limiting distribution is unknown, as ours, the bootstrap is the only appropriate alternative. To the efficiency analysis, the bootstrap was introduced by Simar (1992) and later developed by Simar and Wilson (1998, 2000a) and most recently by Kneip et al. (2003a), whom we follow here.

To outline the basic principle of the bootstrap in our context, at this point we focus on bootstrapping the aggregate efficiency of the *entire* group, assuming we have a data set

[7] For a recent survey of main results (and references to them), see Horowitz (2002).

$\Xi_n = \{(x^k, y^k) \ : \ k = 1, ..., n\}$ generated from a DGP $\wp = \wp(P(x), g(TE, \boldsymbol{h}, x))$ that satisfies the assumptions imposed above. Of course, no one of $P(x), g(TE, \boldsymbol{h}, x), TE$ is observed, but we can obtain *consistent* estimates of them, $\hat{P}(x), T\hat{E}$, using the data Ξ_n from this DGP and the DEA estimator (2.3.3). We then can aggregate the estimates of $T\hat{E}$ over all firm k in the sample using the aggregation procedures presented above to obtain $\overline{T\hat{E}}$ as an estimator of $\overline{TE}$. What we are interested now is in the sampling distribution of $\overline{T\hat{E}} - \overline{TE} \mid \wp$. The idea of the bootstrap is to approximate this distribution by treating Ξ_n as the population, whose properties then can be inferred by operation with *pseudo*-samples, $\Xi_n^* = \{(x^{*k}, y^{*k}) \ : \ k = 1, ..., n\}$, drawn randomly (with replacement) from this population, Ξ_n. Since we have all the population, Ξ_n, we thus can learn everything about the distribution of Ξ_n^*. In particular, we can use the same formula for estimation of technical efficiency as the one applied to the original sample, (2.3.3), but applied to the *pseudo*-sample Ξ_n^*, obtaining $T\hat{E}^{*k}$—the bootstrap estimate of $T\hat{E}^k$, for all k. We then can aggregate $T\hat{E}^{*k}$ over k using (using the same formulas as for the original DEA estimates) now with weights based on the pseudo-sample, to obtain $\overline{T\hat{E}}_b^*$—a bootstrap estimate of $\overline{T\hat{E}}$. If the bootstrap is consistent then the relationship between the bootstrap (pseudo) estimate and the original estimate will mimic the relationship between the original estimate and the true unobserved value of what we want to estimate. In our case, if the bootstrap is consistent, then

$$\overline{T\hat{E}^*} - \overline{T\hat{E}} \mid \hat{\wp} \quad \overset{asy.}{\sim} \quad \overline{T\hat{E}} - \overline{TE} \mid \wp \tag{4.1}$$

Since we know 'everything' about the distribution of Ξ_n^*, at least in principle, the sampling distribution of $\overline{T\hat{E}}^*$ is also completely known and, although its analytical form is most likely unknown, it can be approximated with arbitrary degree of accuracy by Monte-Carlo simulations.

The key, of course, is to have the bootstrap that is *consistent*. Indeed, the context of technology frontier estimation turns out to be a case where the consistency of variants of basic, or naïve, bootstrap is questionable. In fact, a few such approaches of bootstrapping DEA efficiency scores have appeared in the literature, and were shown to offer inconsistent estimates. Briefly put (and see Simar and Wilson (2000 b) for details) the naïve bootstrap does not account for specifics of

the problem—estimation of the (upper) *boundary* of the *unknown* (technology) set. In this case, the assumptions of the theorem of consistency for the naïve bootstrap are violated and thus the naïve bootstrap (sampling distribution) does not correctly mimic the true distribution of the DEA. One remedy was offered by Simar and Wilson (1998, 2000a) who showed how to use the "smooth bootstrap" in the DEA context. The idea of the smooth bootstrap is based on re-sampling *not* from the original input-output data but from the (kernel-estimated) density of technical efficiency scores. Simar and Wilson (1998, 2000a) suggested two variants of the smooth bootstrap: (i) the *homogeneous* bootstrap, and (ii) the heterogeneous bootstrap. Under the homogeneous case, it is assumed that $f(TE \mid h, x) = f(TE)$, i.e., the same probability law is dictating how far any firm is from the frontier, regardless of the input-output mix. The heterogeneous bootstrap does not make this assumption, and thus requires consideration of the joint density $g(TE, h, x)$. Both variants can be extended to cover the group-wise heterogeneous case as ours.

Most recently, Kneip, Simar and Wilson (2003a) offered another alternative—the sub-sampling (with replacement) bootstrap—and, most importantly, showed that it is *consistent* (for *any* sub-sample that has a smaller size than the original sample). They also showed Monte-Carlo evidence that the smooth bootstrap (in the homogeneous case) is a good approximation of this consistent bootstrap. The sub-sampling however has important advantages over the smooth bootstrap. The main advantage is that it accounts for heterogeneity but does not require density estimation as the smooth bootstrap does. Another advantage is that it is much simpler and faster to compute. The main disadvantage of the sub-sampling bootstrap in the DEA context is that the choice of the sub-sample size is not clear at this point. Although consistency of the sub-sampling bootstrap for DEA is proven for *any* sub-sample size smaller than the original, precision of bootstrap estimates in finite samples may be different for different sub-sample sizes (see Kneip et al., 2003). This problem is intuitively similar to the problem of the bandwidth choice in the density estimation. However, many methods exist for the latter, while little is developed for the former in the DEA context. Recent work of Kneip, Simar and Wilson, (2003b), investigates the use of iterated bootstrap to select the appropriate sub-sample size.

Below, we present the algorithm of the sub-sampling bootstrap for DEA estimated aggregate efficiencies, adapted to our context of the *group-wise heterogeneous* case.

4.1. Algorithm of the Group-Wise Heterogeneous Sub-Sampling Bootstrap of Aggregates of DEA Efficiency Scores

1. For each observation in the sample $\Xi_n = \{(x^k, y^k) \ : \ k = 1, ..., n\}$ compute $\hat{TE}(x, y)$, from (2.3.3), obtaining $\{\hat{TE}(x^k, y^k)$: $k=1,\ldots, n\}$.

2. A*ggregate* the estimates of individual efficiencies from step 1 into the *L* sub-group *estimated* aggregate measures of technical efficiency using formulas outlined in section 2.2.

3. Obtain the bootstrap sequence $\Xi^*_{s_l,b} = \{(x^{*k}_b, y^{*k}_b) \ : \ k = 1, ..., s_l\}$ (*b* denotes the bootstrap iteration, *b=1,…,B*), by sub-sampling with replacement independently from data on each sub-group *l* of the original sample, $\Xi_{n_l} = \{(x^k, y^k) \ : \ k = 1, ..., n_l\}$, where $s_l \equiv (n_l)^k, k < 1$, *l=1,…,L*.

4. Compute the bootstrap estimates of $\hat{TE}(x, y)$ via (2.3.3), using *the bootstrapped sample* $\Xi^*_{n,b}$ obtained *from step 3*, call them $\hat{TE}^{*l,k}_b$, for $k = 1, \ldots, s_l < n_l$, all $l = 1, \ldots, L$.

5. Compute the *bootstrap* estimates of the *aggregate* efficiency scores, using

$$\overline{\hat{TE}^{*l}_b} = \sum\nolimits_{k=1}^{s_l} \hat{TE}^{*l,k}_b \cdot S^{*l,k}_b, \quad \text{where} \quad S^{*l,k}_b = \frac{py^{*l,k}_b}{p\sum_{k=1}^{s_l} y^{*l,k}_b}, \ k = 1, \ldots, s_l < n_l; \qquad (4.2a)$$

and

$$\overline{\hat{TE}^{*}_b} = \sum_{l=1}^{L} \overline{TE^{*l}_b} \cdot S^{*l}_b, \quad \text{where} \quad S^{*l}_b = \frac{p\sum_{k=1}^{s_l} y^{*l,k}_b}{p\sum_{l=1}^{L}\sum_{k=1}^{s_l} y^{*l,k}_b}, \quad l = 1, \ldots, L. \qquad (4.2b)$$

or using the price independent weights (e.g., if the price information is unavailable), using

$$S^{*l}_b = \frac{1}{M}\sum_{m=1}^{M} \frac{\sum_{k=1}^{s_l} y^{*l,k}_{m,b}}{\sum_{l=1}^{L}\sum_{k=1}^{s_l} y^{*l,k}_{m,b}}, \qquad l = 1, \ldots, L.$$

and

$$S^{*l,k}_b = \frac{1}{M}\sum_{m=1}^{M} \frac{y^{l,k}_m}{\sum_{k=1}^{s_l} y^{*l,k}_{m,b} \cdot S^{*l}_b}, \qquad k = 1, \ldots, s_l < n_l, \quad l = 1, \ldots, L \qquad (4.3)$$

6. Repeat the steps 3-5, *B* times (obtain the above bootstrap estimates for each $b = 1, \ldots, B$).

At the end, the bootstrap will provide *B* bootstrap-estimates of *estimated* aggregate efficiencies $\{\overline{\hat{TE}}^{*l}_b\}_{b=1}^B$ for each sub-group *l* (*l*=1,…,*L*) and of the entire group $\{\overline{\hat{TE}}^{*}_b\}_{b=1}^B$. These estimates can be used to obtain the bootstrap *confidence intervals*, bias corrected estimates and standard errors of the

estimates. Simar and Wilson (1998, 2000a) have shown how to do this for the *individual* technical efficiency scores. In the next section, we adapt this procedure to obtain the bootstrap confidence intervals and bias correction for the *aggregate* efficiencies.

4.2. Bootstrap Confidence Intervals and Bias Correction for Aggregate Efficiency

Here, we use the most recent approach of Simar and Wilson (2000a), where the bootstrap constructed confidence intervals automatically account for the bias. The key is the expression

$$\overline{\hat{TE}^{*l}} - \overline{\hat{TE}^{l}} \mid \hat{\wp} \quad \overset{asy}{\sim} \quad \overline{\hat{TE}^{l}} - \overline{TE^{l}} \mid \wp \tag{4.4}$$

which is satisfied if the bootstrap is consistent. Given this, the true confidence interval given by $\Pr\left(-b_a \le \overline{\hat{TE}^{l}} - \overline{TE^{l}} \le -a_a \mid \wp\right) = 1-a$ can be approximated with its bootstrap analog

$$\Pr\left(-\hat{b}_a \le \overline{\hat{TE}^{*l}} - \overline{\hat{TE}^{l}} \le -\hat{a}_a \mid \hat{\wp}\right) = 1-a \tag{4.5}$$

where a is the significance level (size of the test) chosen by the researcher, and $\hat{b}_a$ and $\hat{a}_a$ are obtained as the endpoints of the truncated sorted (in ascending order) list of $\left(\overline{\hat{TE}_b^{*l}} - \overline{\hat{TE}^{l}}\right)$, $b = 1, \ldots, B$, where the truncation is done by deleting $(a/2)\times 100$ percent of the elements at each end of the sorted list. The resulting bootstrap confidence interval around the *unknown* aggregate efficiency, $\overline{TE}$, with significance level a, is therefore,

$$\overline{\hat{TE}^{l}} + \hat{a}_a \;\le\; \overline{TE^{l}} \;\le\; \overline{\hat{TE}^{l}} + \hat{b}_a \tag{4.6}$$

To obtain a bias corrected estimate of aggregate efficiency we, again relying on (4.4), note that the true bias $Bias(\overline{\hat{TE}^{l}} \mid \wp) = E\left(\overline{\hat{TE}^{l}}\right) - \overline{TE^{l}}$ can be approximated with its bootstrap analog

$$Bias(\overline{\hat{TE}^{*l}} \mid \hat{\wp}) = E\left(\overline{\hat{TE}^{*l}}\right) - \overline{\hat{TE}^{l}} \tag{4.7}$$

where, $E\left(\overline{\hat{TE}^{*l}}\right)$ can be approximated with its Monte-Carlo analog (from the original bootstrap procedure) $\overline{\overline{\hat{TE}^{*l}}} \equiv \frac{1}{B}\sum_{b=1}^{B} \overline{\hat{TE}_b^{*l}}$, therefore the *estimated* bias is $\hat{Bias}(\overline{\hat{TE}^{*l}} \mid \hat{\wp}) = \overline{\overline{\hat{TE}^{*l}}} - \overline{\hat{TE}^{l}}$ and the resulting bias corrected estimate of the aggregate efficiency is

$$\overline{\breve{TE}^{I}} \;=\; \overline{\hat{TE}^{I}} - \hat{Bias}(\overline{\hat{TE}^{*I}} \mid \hat{\wp}) \;=\; 2\overline{\hat{TE}^{I}} - \overline{\overline{\hat{TE}^{*I}}} \qquad (4.8)$$

Finally, the standard error of $\overline{\hat{TE}^{I}}$ can be computed as

$$se(\overline{\hat{TE}_b^{*I}}) \equiv \frac{1}{B-1}\left[\sum_{b=1}^{B}\left(\overline{\hat{TE}_b^{*I}} - \overline{\overline{\hat{TE}^{*I}}}\right)^2\right]^{1/2} \qquad (4.9)$$

The next section discusses how to make statistical inference on the difference between the aggregate efficiencies of any two distinct sub-groups in a given group.

4.3. The Test for Equality of Aggregate Efficiencies of Two Sub-Groups

In empirical literature, when making judgement on efficiency of certain groups in the industry after using DEA, researchers often resort to such popular non-parametric test as, for example, the Kruskal-Wallis test. However, a direct application of this test to analysis of DEA estimates does not take into account the fact that the *estimates* are used instead of the true efficiencies, thus ignoring the corresponding issues of finite-sample bias and dependency. Most importantly for our context, such tests use *equal* weights, ignoring the economic weights associated with each 'standardized' (to be between 0 and 1 or 1 and ∞) efficiency score.

The goal of this section is to propose a bootstrap-based test of equality of aggregate efficiencies, say, of two sub-groups in an industry. The test can be based on a pair-wise comparison of the aggregate efficiencies of sub-groups. For example, for group A and Z we can postulate

$$H_0 : \overline{TE^A} = \overline{TE^Z} \qquad \text{against} \qquad H_1 : \overline{TE^A} \neq \overline{TE^Z}\,.$$

We are then interested in how far, in a statistical sense, the quantity $RD_{A,Z} = \overline{TE^A} / \overline{TE^Z}$ is different from unity, and we can infer about it by considering its DEA estimator $\hat{RD}_{A,Z} = \overline{\hat{TE}^A} / \overline{\hat{TE}^Z}$, whose behavior can be mimicked by its bootstrap analog, $\hat{RD}^*_{A,Z,b} = \overline{\hat{TE}^{*A}_b} / \overline{\hat{TE}^{*Z}_b}$, *b = 1, …, B.* The bootstrap-based bias correction and the confidence interval for this statistic can be constructed in the same fashion as we described for the aggregate efficiencies in the previous sub-section. The decision rule would then be: "Reject the null if the bootstrap confidence interval does not cover unity". The next section presents a few illustrations of the methods described above for simulated data.

5. Simulated Examples

The goal of this section is to illustrate the methods described above for some simulated examples where we know the 'truth' and thus can get a feeling of the performance of the proposed techniques. In all examples, we assume the prices are not available so that we have to construct the price-independent weights as proposed above.

Playing with many different scenarios we have noticed that the precision of estimation results is sensitive to the choice of sub-sample size $s_l \equiv (n_l)^k, k < 1,\ \ l=1,\ldots,L.$ In any case, reasonable precision was reached for values of k in between 0.5 and 0.7. Monte-Carlo evidence from Kneip et al (2003a) also indicates that, depending on original sample size and the dimension of the technology set, good precision is reached for these values of k. To illustrate, how our methods work we try 3 values of k = {0.5, 0.6, 0.7} and present only the result that gave the best performance among these choices.

5.1. Example 1: Single-output-single-input

We decided to present this example since it allows us to visualize the plot of true technology as well as the spread of the observed realizations of input-output combinations for each firm. We assume that the entire population (e.g., industry) has two types of firms (sub-groups)—A and Z—and we observe 100 firms of each type. We assume that the *true* technology frontier is characterized by the Shephard output distance function of the following simple form:

$$D_o(x, y) = y/(x)^{0.5}$$

For sub-groups A and Z, the only input is assumed to come from *Uniform(0,1).*[8] We assume that $TE^{l,k} = 1 + u^{l,k}$ where $(u^{l,k} \mid x) \sim N^+(m_l, s_l^2(x))$, l = *A, Z,* which we call the '*true*' inefficiencies. We choose $m_A > 0$, intuitively representing some pathological tendency for inefficiency existing for this group, say, of the state-owned firms (such tendency could be justified with one of the economic theories of incentives, etc). For simplicity, we assume $s_A(x) = s_A$.[9] On the other hand, we assume that the other type of firms has a tendency to be technically efficient by having the mode of *TE* at 1, i.e., $m_Z = 0$, but also has $s_Z(x) = s_Z(1-x)^2$. Such heteroskedasticity here can be motivated by vulnerability of this type of firms at low levels of operations (this can be

[8] We have tried other distributions (e.g., Beta, with various parameters) and the results did not change qualitatively.
[9] We have tried it heteroskedastic too and the results did not change qualitatively.

supported by empirical evidence of, say, private firms having higher risk of being unstable when they are small but becoming more stable as they grow). For purpose of supporting the message conveyed by Table 1 in the Introduction section, we set $m_A = 0.25, s_A = 0.2, s_Z = 1.4$.

Figure 1 visualizes the technology set and the input-output realizations for each firm in the sample, and clearly shows the tendency of decreasing inefficiency with increase of scale for Z-type firms. Four panels of Figure 2 give the plots of *estimated* densities of efficiencies for each group. Specifically, (i) compares densities estimated from the 'true realizations' of inefficiency for the two sub-groups (generated from their distributions), and (ii) does the same for the densities estimated using the DEA-*estimates* of efficiencies for the two sub-groups. The two panels look very similar, supporting the fact that in 1-input-1-output case, the rate of convergence of the DEA estimator is better than the usual parametric one. This is also supported by panels (iii) and (iv) that illustrate the difference between the densities estimated with the 'true' and with the DEA-estimates.[10] The rugged shape observed for the sub-group Z is a consequence of our heteroskedasticity assumption. The results of bootstrapping the aggregate efficiencies are given in Table 2.

<Insert Figure 1 Here> **<Insert Figure 2 Here>**

Table 2. Estimation results for Example 1.

	DEA Estim.	True Estim.	Bias Corr. Estim.	Estim. Bias	Estim. St. Dev.	Est. lower CI bound	Est. upper CI bound
AgEf. A	1.2528	1.2785	1.2768	-0.0239	0.0358	1.2065	1.3459
AgEf. Z	1.1237	1.1424	1.1394	-0.0157	0.0311	1.0695	1.1895
AgEf.	1.1817	1.2036	1.2017	-0.0200	0.0245	1.1488	1.2483
MeEf. A	1.2592	1.3007	1.2973	-0.0381	0.0345	1.2290	1.3621
MeEf. Z	1.2811	1.3236	1.3279	-0.0468	0.0789	1.1510	1.4596
MeEf.	1.2701	1.3122	1.3126	-0.0424	0.0431	1.2174	1.3897
$RD_{A,Z;Ag}$	1.1149	1.1191	1.1199	-0.0050	0.0447	1.0340	1.2061
$RD_{A,Z;Mean}$	0.9776	0.9827	0.9621	0.0156	0.0678	0.8321	1.0994

Notes: CI = Confidence Intervals are all at the 0.95 level; AgEf. = aggregate efficiency, MeEf = mean efficiency; $RD_{A,Z;Ag}$ and $RD_{A,Z;Mean}$ are $RD_{A,Z;}$ for aggregate (weighted mean) and (non-weighted) mean efficiencies, respectively; 'True' estimates are obtained by aggregating (with appropriate weights) the 'true' efficiencies that were drawn from the specified above densities when constructing the example. (k_A, k_Z)=(0.7, 0.7); estimation time = 1017.8 (sec.)

[10] The boundary issue is dealt with Silverman (1996) 'reflection' method; bandwidth is by Sheather and Jones (1991).

Perhaps the first thing that catches the eye in Table 2 is that the estimated aggregate efficiencies are, as expected, biased downward for the 'true' ones and that the bias correction makes them closer to the truth, with a little noise. The confidence intervals cover the true quantities. Overall, despite the fact that our estimator knew almost nothing about the true DGP, it still produced good estimates of the aggregate efficiencies, after applying our bootstrap procedure.

The most important observation is that the point-estimates of the *non-weighted* means tell us that the sub-group A is more efficient relative to the sub-group Z, while the point-estimates of the *aggregate* efficiencies (weighted means) tell us the opposite story. The bootstrap confidence intervals (CIs) for these efficiencies suggest that the difference between the non-weighted means is *not* statistically significant, while it is significant for the aggregate efficiencies. This last argument is also supported by the bootstrap CIs for our RD-measure—applied for comparison of sub-group efficiencies using the weighted means ($RD_{A,Z;Ag}$) and using the non-weighted means ($RD_{A,Z;Mean}$).

All this supports the corner-stone argument of our research: Tests on sample means of estimated DEA efficiencies may lead to quite different conclusion than the tests based on aggregate efficiencies, whose weights account for economic importance of each firm in the sample.

5.2. Example 2: Two-outputs-two-inputs

Here we modify the example from Park et al., (2000). We assume the technology is characterized by the Shephard's output distance function of the following simple form:

$$D_o(x, y) = (y_2 + y_1)/(x_1)^{0.2}(x_2)^{0.3}$$

For both sub-groups A and Z, the two inputs are drawn from *Uniform(0,1)*. The outputs are generated by first drawing the pseudo-outputs, $\tilde{y}_1^{l,k}$ and $\tilde{y}_2^{l,k}$, from *Uniform(0.2,1)* for both sub-groups, which are then used to generate random rays in the output space characterized by the slopes $s^{l,k} = \tilde{y}_1^{l,k} / \tilde{y}_2^{l,k}$ for each k in each l (l=*A,Z*), which are in turn used to generate the *efficient* outputs (i.e., when $D_o(x, y) = 1$) as: $y_{1,eff}^{l,k} = (x_1)^{0.2}(x_2)^{0.3}/(s^{l,k} + 1)$, and $y_{2,eff}^{l,k} = (x_1)^{0.2}(x_2)^{0.3} - y_{1,eff}^{l,k}$. Finally, the 'realized' (or observed) outputs are constructed as $y_1^{l,k} = y_{1,eff}^{l,k} / TE^{l,k}$, and $y_2^k = y_{2,eff}^k / TE^{l,k}$ where, $TE^{l,k} = 1 + u^{l,k}$, $(u^{l,k} \mid x) \sim N^+(m_l, s_l^2(x))$, for all k, l = *A, Z*.

Here we assume that $m_A = -0.1$, $m_Z = 0$, and now inefficiency is heteroskedastic for *both* type of firms, however in a very different manner. For the Z-type firms, $s_Z(x) = s_Z(1 - x_1)^{g_Z}$,

while for the A-type firms: $s_A(x) = s_A \cdot (x_1)^{g_A}$, that is, the A-type firms here tend to be more inefficient as they use more of the input x_1 (e.g., increase in number of employees may increase the asymmetric information problem between the employees and management and thus lead to decrease in firm's efficiency). Again, for the purpose of supporting the message conveyed by Table 1 above, we set $s_A = 0.5$, $s_Z = 1.7$ and $g_A = 3$, $g_Z = 1/5$. Figure 3 visualizes this scenario with the kernel-estimated densities of true and DEA-estimated efficiencies for each group.

<Insert Figure 3 here>

Table 3. Estimation results for Example 2.

	DEA Estim.	True Estim.	Bias Corr. Estim.	Estim. Bias	Estim. St. Dev.	Est. lower CI bound	Est. upper CI bound
AgEf. A	1.2083	1.2718	1.2787	-0.0704	0.0312	1.2118	1.3360
AgEf. Z	1.1081	1.1591	1.1561	-0.0480	0.0268	1.0965	1.2000
AgEf.	1.1573	1.2144	1.2168	-0.0595	0.0202	1.1731	1.2540
MeEf. A	1.2700	1.3138	1.3154	-0.0834	0.0345	1.2435	1.3779
MeEf. Z	1.2990	1.3141	1.2852	-0.0904	0.0516	1.1688	1.3673
MeEf.	1.2134	1.3139	1.3003	-0.0869	0.0309	1.2331	1.3543
$RD_{A,Z;Ag}$	1.0904	1.0972	1.1068	-0.0164	0.0408	1.0252	1.1885
$RD_{A,Z;Mean}$	0.9733	0.9998	0.9045	0.0688	0.0570	0.7968	1.0209

Notes: CI = Confidence Intervals are all at the 0.95 level; AgEf. = aggregate efficiency, MeEf = mean efficiency; $RD_{A,Z;Ag}$ and $RD_{A,Z;Mean}$ are $RD_{A,Z;}$ for aggregate (weighted mean) and (non-weighted) mean efficiencies, respectively; 'True' estimates are obtained by aggregating (with appropriate weights) the 'true' efficiencies that were drawn from the specified above densities when constructing the example. (k_A, k_Z)=(0.7, 0.7); Time of estimation = 1827.1 (sec.)

Table 3 presents the results of estimation and the first thing that must catch the eye here is that again the estimated aggregate efficiency scores are quite *far* from the 'true' ones and, remarkably, the bias correction definitely improves our estimate. The confidence intervals also cover the true quantities. Overall, despite all the complications we have imposed in our scenario we still recover the truth very well with our bootstrap application to DEA estimates.

Note that the point-estimates tell us that the efficiency scores of the two sub-groups, A and Z, are very similar when the non-weighted means are used, but quite different when the *aggregate* efficiencies are used. In turn, the bootstrap CIs for these efficiencies and for the RD-measures suggest that the non-weighted means are not significantly different from each other, while the aggregate efficiencies of the two sub-groups do differ significantly.

All this gives another support for the argument given in Table 1, and illustrates that the method we propose in this paper helps making inferences in a quite complex environment, but with a fully non-parametric approach, where the knowledge of that complexity is not required.

5.3 Example 3: Two-outputs-two-inputs

Here we modify the example above under assumption that the two groups share the same distribution. In particular, we assume that $m_A = m_Z = 0$ and $s_A(x) = s_Z(x) = 0.35$, and the rest is the same as in the example 2. The goal is to see if our method generates any 'spurious' difference between the groups. Although both groups have the same underlying DGP, the particular draws we got look a bit different even in terms of estimated densities based on true efficiencies, as revealed by Figure 4 (panel i), but this difference is *not* exaggerated (but mimicked) when the DEA-estimates are used instead (panel ii). It is still possible, in principle, that the weights happen to be distributed such that the aggregate efficiencies look different in estimation and in a bootstrap iteration. The results of bootstrapping aggregate DEA scores are given in Table 4.

Table 4. Estimation results for Example 3.

	DEA Estim.	True Estim.	Bias Corr. Estim.	Estim. Bias	Estim. St. Dev.	Est. lower CI bound	Est. upper CI bound
AgEf. A	1.1306	1.2229	1.2082	-0.0776	0.0327	1.1339	1.2586
AgEf. Z	1.1591	1.2405	1.2520	-0.0929	0.0286	1.1895	1.3006
AgEf.	1.1445	1.2315	1.2285	-0.0841	0.0203	1.1846	1.2631
MeEf. A	1.1487	1.2577	1.2384	-0.0897	0.0365	1.1513	1.2948
MeEf. Z	1.1746	1.2757	1.2781	-0.1035	0.0331	1.2048	1.3318
MeEf.	1.1617	1.2667	1.2571	-0.0954	0.0237	1.2056	1.2967
$RD_{A,Z;Ag}$	0.9754	0.9858	0.9625	0.0130	0.0424	0.8753	1.0404
$RD_{A,Z;Mean}$	0.9816	0.9859	0.9734	0.0082	0.0472	0.8778	1.0596

Notes: CI = Confidence Intervals are all at the 0.95 level; AgEf. = aggregate efficiency, MeEf = mean efficiency; $RD_{A,Z;Ag}$ and $RD_{A,Z;Mean}$ are $RD_{A,Z;}$ for aggregate (weighted mean) and (non-weighted) mean efficiencies, respectively; 'True' estimates are obtained by aggregating (with appropriate weights) the 'true' efficiencies that were drawn from the specified above densities when constructing the example. (k_A, k_Z)=(0.6, 0.6); Time of estimation = 971 (sec.)

As in the example above, the point-estimates of the aggregate DEA-efficiency scores are quite *far* from the 'true' ones and the bias correction does a good job improving them. The

bootstrap CI's cover the true quantities and overlap for the two sub-groups. The tests based on the RD-measure also suggest that the two sub-groups are not statistically different from each other both in terms of the aggregate efficiencies and in terms of the sample means. Thus, when the groups have identical DGP, the group-wise heterogeneous sub-sampling bootstrap (which does not take into account information that the DGP is the same for both sub-groups) does a good job estimating the true model.

6. Conclusion

In this paper we have merged two streams of the current literature in efficiency analysis—bootstrap and aggregation—and thus proposed a new way of making statistical inference on the relative efficiency among the distinct sub-groups of (e.g., public vs. private) firms within a population. Simulations that we have tried, a few of which we presented here, suggest that the proposed methodology have a good potential to be very useful for practitioners and we are ready to share our code (for Matlab) to facilitate the use of the proposed method in empirical research.

A natural extension of this work would be to provide extensive Monte-Carlo (MC) evidence of performance of the proposed methodology (including analysis of empirical size and power of the proposed test) for various scenarios. Although it is quite a time-expensive exercise in itself, given that just one MC replication takes about 1000-2000 seconds (for 1133MHz, 253MB machine) for a 2x2 case with 200 firms, it can shed light on the precision of the method we proposed under various circumstances. Another useful extension for this work (and any other sub-sampling bootstrap application) is to develop a data driven method of choosing the sub-sample size, especially the one that maximizes the power of the proposed test for the comparison of aggregate efficiencies among different sub-groups. This could be done along the lines of the iterated bootstrap procedure proposed by Kneip et al. (2003b).

References

Blackorby, C., and R. Russell (1999), "Aggregation of Efficiency Indices", *Journal of Productivity Analysis* 12:1, 5-20.

Cornes, R. (1992), *Duality and Modern Economics*, Cambridge University Press, Cambridge.

Debreu, G. (1951), "The coefficient of resource utilization," *Econometrica*, 19, 273-292.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife," *Annals of Statistics* 7, 1-26.

Färe, R., S. Grosskopf and V. Zelenyuk (2002), "Aggregation of Cost Efficiency Indicators and Indexes Across Firms", *Scientific Report 2002:*1, R.R. Institute of Applied Economics, Sweden.

Färe, R. and D. Primont (1995), Multi-Output Production and Duality: Theory and Applications, Kluwer Academic Publishers, Boston.

Färe, R. and V. Zelenyuk (2003), "On Aggregate Farrell Efficiency Scores," *European Journal of Operations Research* 146:3, 615-620.

Farrell, M.J. (1957), "The Measurement of Productive Efficiency," *Journal of Royal Statistical Society*, Series A, General, 120, part 3, 253-281.

Gijbels, I., E. Mammen, B.U. Park and L. Simar (1999), "On Estimation of Monotone and Concave Frontier Functions", *Journal of the American Statistical Association* 94, 220-228.

Horowitz, J.L. (2002), "The Bootstrap" in J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, Amsterdam.

Li, S.K., and Y.C. Ng (1995), "Measuring the Productive Efficiency of a Group of Firms," *International Advances in Economic Research* 1(4), 377-90.

Kneip, A., L. Simar and P. Wilson (2003a), "Asymptotics for DEA Estimators in Non-parametric Frontier Models", Discussion Paper #0317, Institut de Statistique, Université Catholique de Louvain, Belgium .

Kneip, A., L. Simar and P. Wilson (2003b), "A simplified bootstrap for DEA estimators", paper presented at the 8th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, September 2003.

Kneip, A., B. Park and L. Simar (1998), "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores", *Econometric Theory* 14, 783-793.

Korostelev, A., L. Simar, and A. Tsybakov (1995), "Efficient Estimation of Monotone Boundaries", *The Annals of Statistics* 23(2), 476-489.

Koopmans, T.C. (1957), *Three Essays on the State of Economic Analysis*, New York: McGraw-Hill.

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

Russell, R. (1990), "Continuity of Measures of Technical Efficiency," *Journal of Economic Theory* 51, 255-267.

Shephard, R. (1970), *Theory of Cost and Production Function*s, Princeton: Princeton University Press.

Sheather, S.J. and M.C. Jones (1991), "Areliable Data Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of Royal Statistical Society*, Series B, 53, 683-690.

Silverman B.W. (1996), *Density Estimation for Statistics and Data Analysis*, New York, Chapman and Hall.

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.

Simar, L. and P. Wilson (1998), "Sensitivity of efficiency scores : How to bootstrap in Nonparametric frontier models", *Management Sciences* 44(1), 49-61.

Simar, L. and P. Wilson (2000 a), "A General Methodology for Bootstrapping in Nonparametric Frontier Models", *Journal of Applied Statistics* 27, 779-802.

Simar, L. and P. Wilson (2000 b), "Statistical Inference in Nonparametric Frontier Models: The State of the Art," *Journal of Productivity Analysis* 13, 49-78.

Ylvinger S., (2000) "Industry performance and structural efficiency measures: Solutions to problems in firm models," *European Journal Of Operational Research* 121-1, 164-174.

APPENDIX

Proof of (2.2.4).

Recall that $x^k = (x_1^k,..., x_N^k)' \in \Re_+^N$ and $y^k = (y_1^k,..., y_M^k)' \in \Re_+^M$ are input and output vectors, respectively, of a particular firm k ($k = 1, \ldots, n$). Input allocation over the entire group of n firms is denoted with a (N by n) matrix $X = (x^1,...,x^n)$. Now suppose the entire group of firms must be partitioned into (non-intersecting) *subgroups* by some exogenous criterion. Suppose there are L subgroups (indexed as: $l = 1, \ldots, L$) with number of firms in each group l equal to a positive integer n_l. The input allocation among firms *within* a group l will be denoted by a (N by n_l) matrix $X^l = (x^{l,1},...,x^{l,n_l})$. In general, technology of a particular firm k ($k = 1,..., n_l$) within a group l is assumed to be characterized by the output sets

$$P^{l,k}(x^{l,k}) \equiv \{ y^{l,k} : x^{l,k} \text{ can produce } y^{l,k} \}, \quad x^{l,k} \in \Re_+^N \qquad (A1)$$

Technology of a particular *sub-group l* is *assumed* to be related to technologies of its firms as

$$\overline{P}^l(X^l) \equiv \sum_{k=1}^{n_l} P^{l,k}(x^{l,k}) \qquad (A2)$$

which yields one of the main aggregation results we stated in the text as (2.2.4):

$$\overline{R}^l(X^l, p) = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \qquad (A3)$$

where

$$R^{l,k}(x^{l,k}, p) \equiv \max_y \{ py : y \in P^{l,k}(x^{l,k}) \} \qquad (A4)$$

and

$$\overline{R}^l(X^l, p) \equiv \max_y \{ py : y \in \overline{P}^l(X^l) \} \qquad (A5)$$

To prove this, for each $k = 1,..., n_l$, take $y^{l,k}$ to be an arbitrary vector in $P^{l,k}(x^{l,k})$ and use them to define $\overline{Y}^l = \sum_{k=1}^{n_l} y^{l,k}$. Because of (A2) we have $\overline{Y}^l \in \overline{P}^l(X^l)$, and due to (A5), we obtain

$$p\overline{Y}^l \le \overline{R}^l(X^l, p) \qquad (A6)$$

Since $y^{l,k}$ is an *arbitrary* vector in $P^{l,k}(x^{l,k})$, it implies that (A6) also holds for those $y^{l,k}$ that *solve* (A4), call them $\tilde{y}^{l,k}$, in which case we would have

$$p\tilde{\overline{Y}}^l \equiv \sum_{k=1}^{n_l} p\tilde{y}^{l,k} = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \le \overline{R}^l(X^l, p) \qquad (A7)$$

On the other hand, let $\overline{Y}^l$ be an arbitrary vector in $\overline{P}^l(X^l)$, then due to (A2) there exist $y^{l,k} \in P^{l,k}(x^{l,k})$, for each $k = 1,..., n_l$, such that $\overline{Y}^l = \sum_{k=1}^{n_l} y^{l,k}$. Therefore, due to (A4) we have

$$p\overline{Y}^{l} \equiv \sum_{k=1}^{n_l} py^{l,k} \le \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p), \qquad \textit{(A8)}$$

and since $\overline{Y}^{l}$ is an *arbitrary* vector in $\overline{P}^{l}(X^{l})$, expression (A8) is also true for those $\overline{Y}^{l}$ that solve (A5), call them $\tilde{\overline{Y}}^{l}$, in which case we get

$$p\tilde{\overline{Y}}^{l} \equiv \overline{R}^{l}(X^{l}, p) \le \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \qquad \textit{(A9)}$$

Clearly, expressions (A7) and (A9) can simultaneously hold if and only if

$$\overline{R}^{l}(X^{l}, p) = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p). \qquad \textit{Q.E.D.}$$

One immediate implication from this conclusion is that if $L = 1$, i.e., the subgroup is the entire group (thus indexing with l can be dropped, e.g., $n_l = n$), then

$$\overline{R}(X, p) = \sum_{k=1}^{n} R^{k}(x^{k}, p), \qquad \textit{(A10)}$$

where

$$\overline{R}(X, \ p) \equiv \max_{y} \{py \ : \ y \in \overline{P}(X)\} \qquad \textit{(A11)}$$

and

$$\overline{P}(X) = \sum_{k=1}^{n} P^{k}(x^{k}) \qquad \textit{(A12)}$$

Another important implication is about the relationship between the maximal revenues of the sub-groups to the maximal revenue of the entire group. In particular, since

$$\sum_{l=1}^{L} \overline{R}^{l}(X^{l}, p) = \sum_{l=1}^{L} \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) = \sum_{k=1}^{n} R^{k}(x^{k}, p) \qquad \textit{(A13)}$$

thus along with (A10) we get

$$\overline{R}(X, p) = \sum_{l=1}^{L} \overline{R}^{l}(X^{l}, p). \qquad \textit{(A14)}$$

This insures 'internally consistent' aggregation *within* and *between* the subgroups in the sense that

$$\overline{RE} = \sum_{l=1}^{L} \overline{RE}^{l} \cdot S^{l} = \sum_{k=1}^{n} RE^{k} \cdot S^{k}, \quad \overline{TE} = \sum_{l=1}^{L} \overline{TE}^{l} \cdot S^{l} = \sum_{k=1}^{n} TE^{k} \cdot S^{k}, \quad \text{where } S^{k} \equiv \frac{py^{k}}{p\overline{Y}}, \text{ and}$$

$$\overline{AE} = \sum_{l=1}^{L} \overline{AE}^{l} \cdot S_{ae}^{l} = \sum_{k=1}^{n} AE^{k} \cdot S_{ae}^{k}, \text{ where } S_{ae}^{k} \equiv \frac{p\left(y^{k} TE^{k}(x^{k}, y^{k})\right)}{p\sum_{k=1}^{n}\left(y^{k} TE^{k}(x^{k}, y^{k})\right)}, \qquad k = 1, \ldots, n.$$

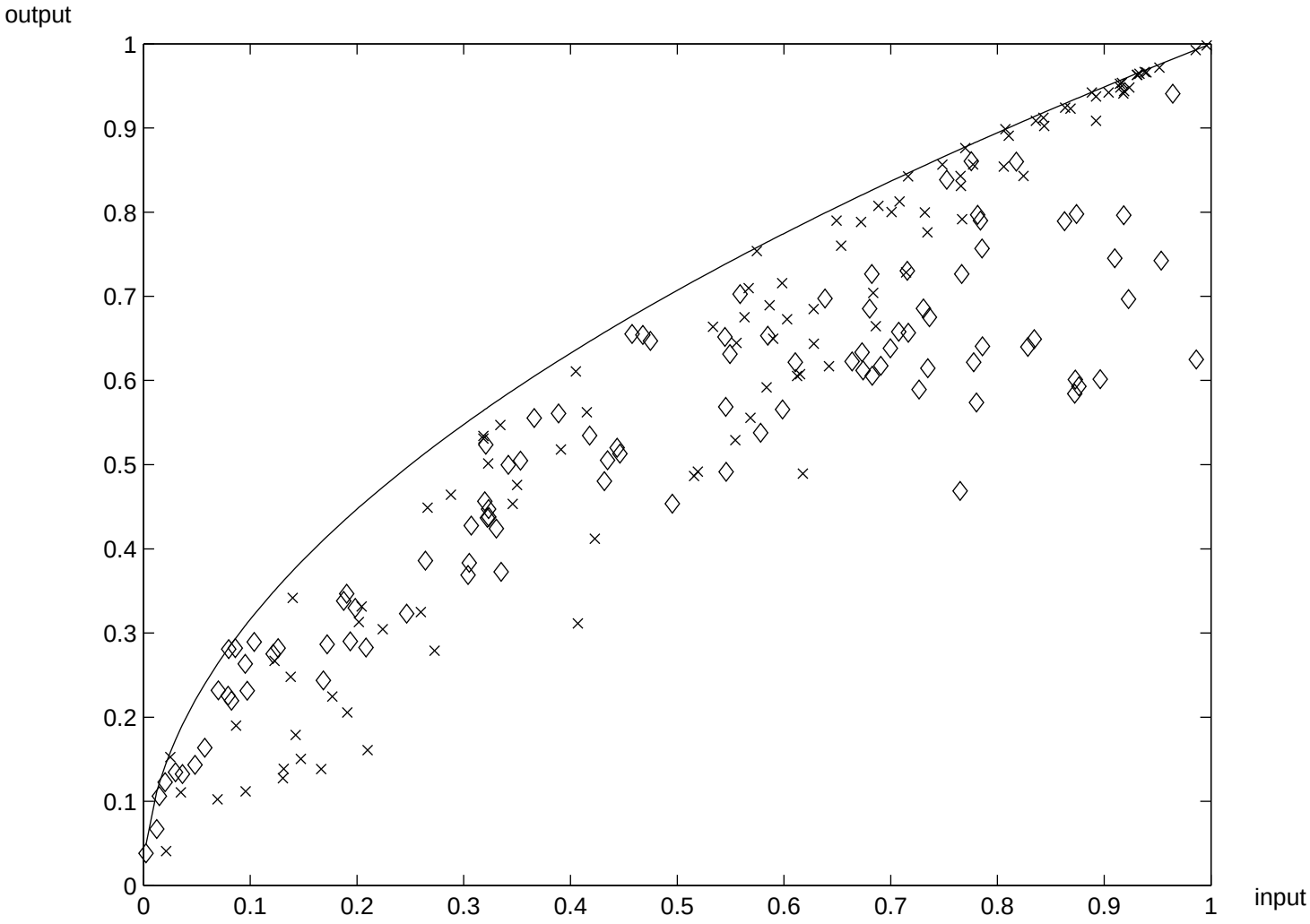
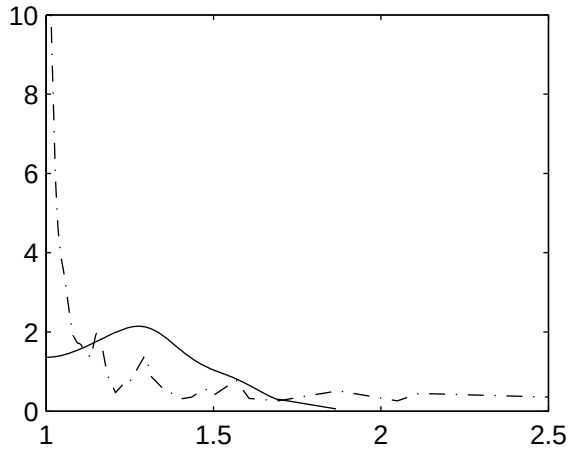
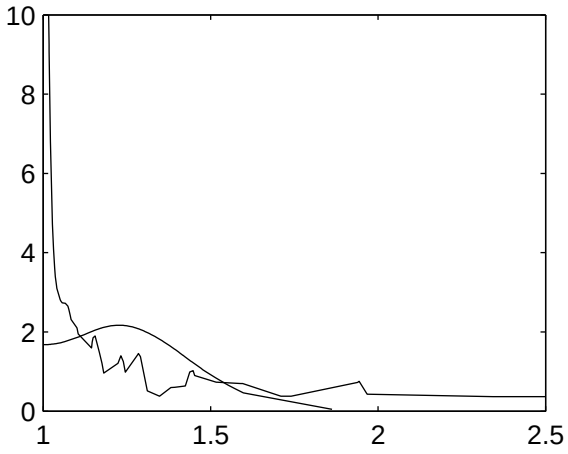


Figure 1. True technology set and observed firms (sub−group Z is indicated by 'x')

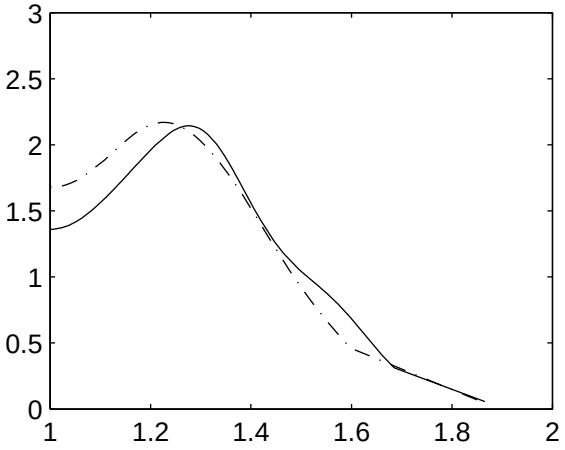
i. Estimated from true efficiencies for A and Z



ii. Est. from DEA−est.efficiencies for A and Z



iii. Sub−group A: True vs DEA−estimated efficiencies



iv. Sub−group Z: True vs DEA−estimated efficiencies

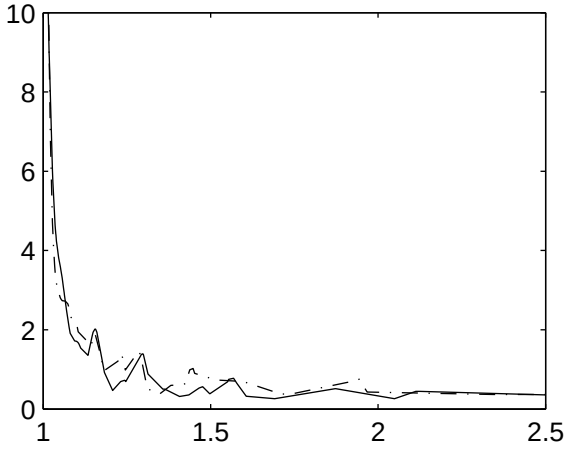
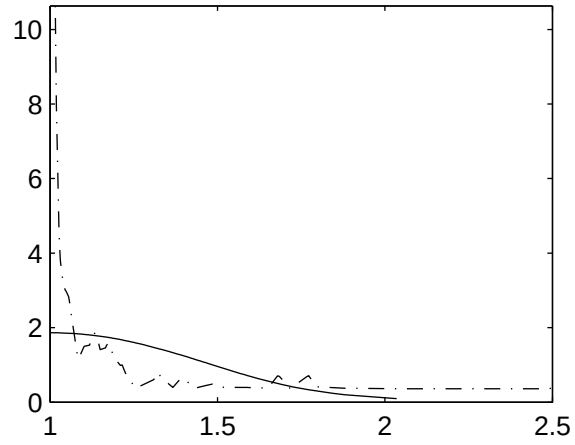
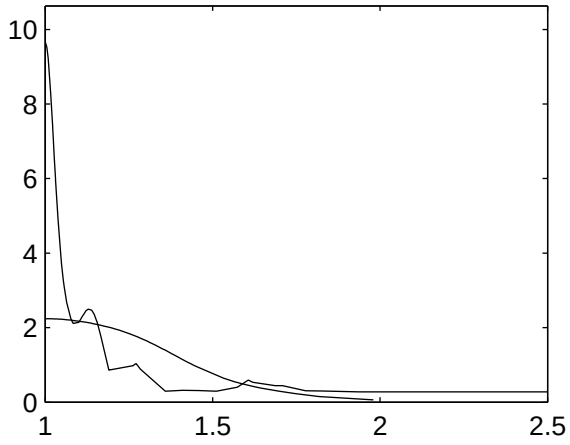


Figure 2. Kernel−estimated densities of efficiencies for sub−group A and Z,
estimated from true or DEA−estimates of efficiencies: Example 1

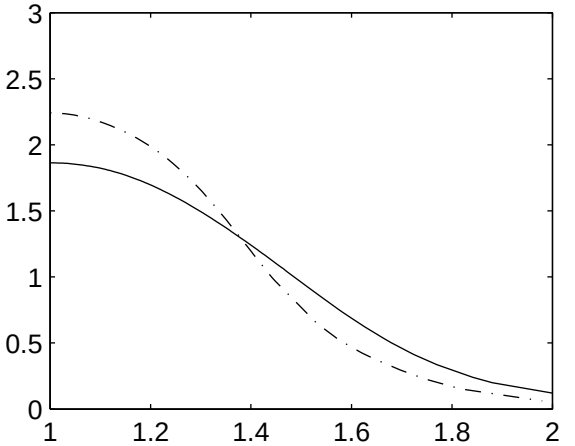
i. Estimated from true efficiencies for A and Z



ii. Est. from DEA−est.efficiencies for A and Z



iii. Sub−group A: True vs DEA−estimated efficiencies



iv. Sub−group Z: True vs DEA−estimated efficiencies

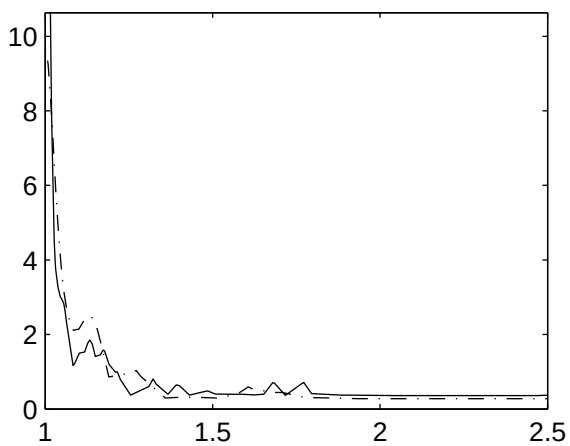
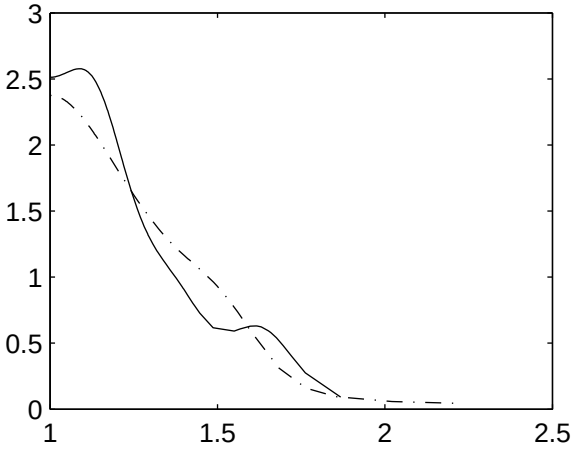
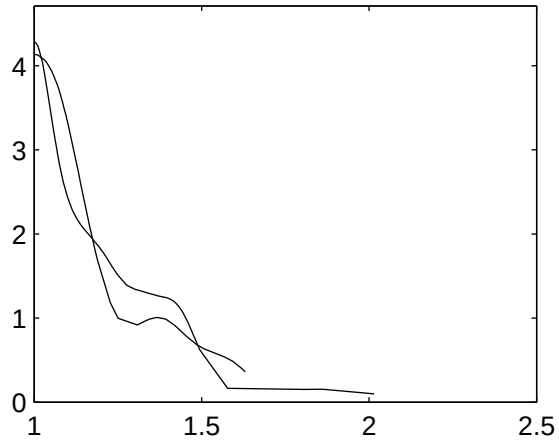


Figure 3. Kernel−estimated densities of efficiencies for sub−group A and Z,
estimated from true or DEA−estimates of efficiencies: Example 2

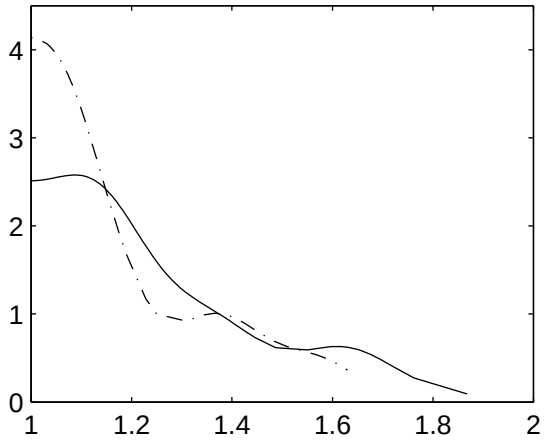
i. Estimated from true efficiencies for A and Z



ii. Est. from DEA−est.efficiencies for A and Z



iii. Sub−group A: True vs DEA−estimated efficiencies



iv. Sub−group Z: True vs DEA−estimated efficiencies

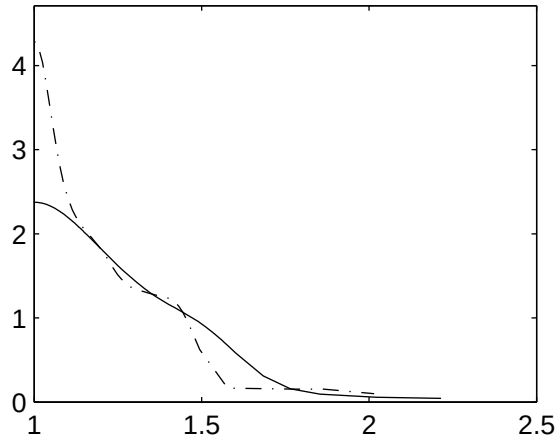


Figure 4. Kernel−estimated densities of efficiencies for sub−group A and Z,
estimated from true or DEA−estimates of efficiencies: Example 3