JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **22**: 1367–1394 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.991

STATISTICAL INFERENCE FOR AGGREGATES OF FARRELL-TYPE EFFICIENCIES

LÉOPOLD SIMAR[a] AND VALENTIN ZELENYUK[b]*
[a] *Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*
[b] *Kyiv Economics Institute, KEI and UPEG/EERC Kiev, Ukraine*

SUMMARY
In this study, we merge results of two recent directions in efficiency analysis research—aggregation and 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., regulated vs. non-regulated firms or private vs. public firms). Our methodology is justified by the (neoclassical) economic theory and is supported by carefully adapted statistical methods. Copyright © 2007 John Wiley & Sons, Ltd.

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, and for many others. Sometimes, economic theory cannot give a 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 subgroups 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 concerning the appropriateness of methodology: (i) reliable *point* estimators of group (or subgroup) 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

* Correspondence to: Valentin Zelenyuk, Kyiv Economics Institute, and Kyiv School of Economics (EERC), vul. Dekhtyarivska 51, Kyiv, Ukraine, 03113. E-mail: vzelenyuk@eerc.kiev.ua

Table I. A hypothetical example

Firms in A (10%)	Weight in subgroup	Efficiency (%)	Firms in Z (90%)	Weight in subgroup	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%

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 to such an important question as: 'What is the more efficient way of regulation used in practice?' may merely depend on the (researcher's) 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 I.

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 group Z. Noting that the efficiency scores per se are 'standardized' to be between zero and one and 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.

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 subgroups 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 subgroups, 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 the most recent discovery of the limiting distribution by Kneip *et al.* (2003a), who also prove the consistency of the subsampling bootstrap for the DEA estimator of individual efficiency scores.

The goal of this paper is to merge the existing works on bootstrap for DEA with those 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.

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

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 $(k = 1, 2, \ldots, n)$ we will use vector $x^k = (x_1^k, \ldots, x_N^k)' \in \Re_+^N$ to denote N inputs that firm k uses to produce a vector of M outputs, denoted by $y^k = (y_1^k, \ldots, y_M^k)' \in \Re_+^M$. We assume the technology of firm k can be characterized by the set T^k:[2]

$$T^k \equiv \{(x^k, y^k) : x^k \text{ can produce } y^k\} \tag{1}$$

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

$$P^k(x^k) \equiv \{y^k : x^k \text{ can produce } y^k\}, x^k \in \Re_+^N \tag{2}$$

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

$$D_o^k(x^k, y^k) \equiv \inf\{\theta : y^k/\theta \in P^k(x^k)\} \tag{3}$$

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

$$D_o^k(x^k, y^k) \leq 1 \Leftrightarrow y^k \in P^k(x^k) \tag{4}$$

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 an 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[3]

$$TE^k(x^k, y^k) \equiv \max\{\theta : \theta y^k \in P^k(x^k)\} = 1/D_o^k(x^k, y^k) \tag{5}$$

Formally, if we let the technological frontier be the 'upper' boundary of $P^k(x^k)$, defined as $\partial P^k(x^k) = \{y \in \Re_+^M : y \in P^k(x^k), \lambda y \notin P^k(x^k), \forall \lambda \in (1, \infty)\}$ 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 equation (5) (or its reciprocal). Alternatively, we 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:

$$R^k(x^k, p) \equiv \max_y\{py : y \in P^k(x^k)\} \tag{6}$$

[2] Note that to develop theory as general as possible we allow for technology to be different over firms. For estimation purposes we will then assume that all firms have access to the same ('best-practice') technology.

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

DOI: 10.1002/jae

where $p = (p_1, \ldots, p_M) \in \Re^{M}{}_{++}$ denotes the vector of output prices.[4] As 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[5]

$$D_o^k(x^k, y^k) = \sup_p\{py^k : R^k(x^k, p) \leq 1\} \tag{7}$$

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

$$RE^k(x^k, y^k, p) \equiv R^k(x^k, p)/py^k \tag{8}$$

Mahler's inequality (which can be obtained from (7)) tells us that

$$R^k(x^k, p) \geq py^k/D_o^k(x^k, y^k) \tag{9}$$

and the multiplicative residual that closes the inequality (9) is often interpreted as the criterion or a measure of the *allocative* (in)efficiency of firm k, and is formally defined as

$$AE^k(x^k, y^k, p) \equiv RE^k(x^k, y^k, p)/TE^k(x^k, y^k) \tag{10}$$

The decomposition (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 subgroup, call it subgroup l, of n_l firms taken from the original group of n firms (e.g., the selection can be based on exogenous economic criteria such as ownership structure or regulation regimes). 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 the lth group with $\overline{Y}^l = \sum_{k=1}^{n_l} y^k$.

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.:

$$\overline{P}^l(X^l) \equiv \sum_{k=1}^{n_l} P^{l,k}(x^{l,k}) \tag{11}$$

Thus, the *output set* of a subgroup of firms, $\overline{P}^l(X^l)$, is the sum of the individual output sets of all firms in this subgroup. 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.

---

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

[5] 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.

DOI: 10.1002/jae

Given the lth subgroup technology (11), the *subgroup revenue* function can be defined as

$$\overline{R}^l(X^l, p) \equiv \max_y\{py : y \in \overline{P}^l(X^l)\} \tag{12}$$

and the lth subgroup revenue efficiency, the analogue of (9) is therefore defined as

$$\overline{RE}^l(X^l, \overline{Y}^l, p) \equiv \overline{R}^l(X^l, p)/p\overline{Y}^l \tag{13}$$

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

Theorem[6] *The maximal revenue of the subgroup of firms is equal to the sum of the maximal revenues of all its member firms, i.e.:*

$$\overline{R}^l(X^l, p) = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \tag{14}$$

This theorem (as well as the first two corollaries below) is adapted from Färe and Zelenyuk (2003) to the context of subgroups, 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 (11), 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.

Corollary 1 The revenue efficiency of the lth subgroup 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 subgroup, i.e.:

$$\overline{RE}^l(X^l, \overline{Y}^l, p) = \sum_{k=1}^{n_l} RE^{l,k}(x^{l,k}, y^{l,k}, p) \cdot S^{l,k} \tag{15}$$

where

$$S^{l,k} = py^{l,k}/p\overline{Y}^l, k = 1, \ldots, n_l \tag{16}$$

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.:

$$\overline{RE}^l(X^l, \overline{Y}^l, p) = \overline{TE}^l \times \overline{AE}^l \tag{17}$$

[6] This theorem is a revenue analogue to the Koopmans (1957) theorem of aggregation of the profit functions. The cost analogue is proven in Färe *et al*. (2002).

DOI: 10.1002/jae

so that

$$\overline{TE}^{l} \equiv \sum_{k=1}^{n_l} TE^{l,k}(x^{l,k}, y^{l,k}) \cdot S^{l,k} \tag{18}$$

and

$$\overline{AE}^{l} \equiv \sum_{k=1}^{n_l} AE^{l,k}(x^{l,k}, y^{l,k}, p) \cdot S_{ae}^{l,k} \tag{19}$$

with

$$S_{ae}^{lk} \equiv \frac{p y_*^{l,k}}{p \sum_{k=1}^{n_l} y_*^{l,k}}, k = 1, \ldots, n_l \tag{20}$$

where $y_*^{l,k} \equiv y^{l,k} \cdot TE^{l,k}(x^{l,k}, y^{l,k})$ is the output vector corrected for technical inefficiency.

Remark 1 Note that if $L = 1$, then the aggregate measures in (17)–(19) are measures for the entire group. Moreover, the measure (18) is a *multi-output* generalization of what Farrell called the '*structural efficiency of an industry*' (Farrell, 1957, pp. 261–262). Also note that the resulting weighting scheme is quite intuitive: technical efficiencies are measured starting from actual output levels—so, naturally, these efficiencies are weighted at the actual output levels. On the other hand, allocative efficiencies are measured starting from technically efficient output levels—so, naturally, these efficiencies are weighted at the technically efficient output levels. Importantly, these weights were not made up ad hoc to be intuitive, but are derived from economic optimization behaviour.

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, 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 behaviour—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, that is, account for its relative economic effort in earning the particular 'standardized' efficiency score, then, since prices contain important economic information, it will not be surprising that we obtained price-dependent weights from imposing the economic optimization principle. The second consideration is more practical: price information may be unavailable 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 price-independent (as we shall do below).

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

DOI: 10.1002/jae

Let us now consider a case when (given some exogenous economic criterion) a researcher is interested in comparing aggregate efficiencies of certain subgroups *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) subgroups of firms, i.e.:

$$\overline{R}(X, p) = \sum_{l=1}^{L} \overline{R}^l(X^l, p) \tag{21}$$

This is an analogue of Theorem 1 (extension to aggregation *between* subgroups) and the next corollary is the corresponding analogue 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) subgroups of firms, where the weights are the actual revenue shares of these subgroups in the entire group, i.e.:

$$\overline{RE}(X, \overline{Y}, p) = \sum_{l=1}^{L} \overline{RE}^l(X^l, \overline{Y}^l, p) \cdot S^l \tag{22}$$

where

$$S^l = p\overline{Y}^l / p \sum_{l=1}^{L} \overline{Y}^l, \; l = 1, \ldots, L \tag{23}$$

Finally, the 'between-group aggregation' analogue of 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 subgroup technical efficiencies of all non-intersecting subgroups (where the weights are the actual revenue shares of these subgroups) and the weighted sum of the subgroup allocative efficiencies of all non-intersecting sub-groups (where the weights are the actual revenue shares corrected for the subgroups technical inefficiency). Formally:

$$\overline{RE}(X, \overline{Y}, p) = \overline{TE} \times \overline{AE} \tag{24}$$

where

$$\overline{TE} = \sum_{l=1}^{L} \overline{TE}^l \cdot S^l, \;\; \overline{AE} = \sum_{l=1}^{L} \overline{AE}^l \cdot S^l_{ae} \tag{25}$$

and

$$S^l = p\overline{Y}^l / p \sum_{l=1}^{L} \overline{Y}^l, \text{ and } S^l_{ae} = p\overline{Y}^l \overline{TE}^l / p \sum_{l=1}^{L} \overline{Y}^l \overline{TE}^l, \; l = 1, \ldots, L \tag{26}$$

DOI: 10.1002/jae

That is, the efficiencies of the subgroups of firms are aggregated into efficiencies of the entire group in a similar manner as the efficiencies of individual members of the subgroup are aggregated into the subgroup efficiencies.

Price-Independent Weights

Here, we elaborate on 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 used above. Let us illustrate this for the case of aggregating technical efficiencies of the entire group. The trick requires the following additional assumption:

$$p_m \overline{Y}_m / \sum\nolimits_{m=1}^{M} p_m \overline{Y}_m = a_m, m = 1, \ldots, M \tag{27}$$

where $\overline{Y}_m \equiv \sum_{k=1}^{n} y_m^k$, while a_m $(m = 1, \ldots, M)$ are known (estimated or assumed) constants between zero and unity. Intuitively, (27) says that the share of the industry revenue from output m in the industry total revenue equals a_m, so $\sum_{m=1}^{M} a_m = 1$. Such aggregate information is often available from industry or government reports. Now, letting $\varpi_m^k = y_m^k / \overline{Y}_m$ be the share of k's firm in the group in terms of mth output and imposing (27) on the weights for aggregation of technical and revenue efficiencies derived above would yield

$$S^k = \sum_{m=1}^{M} a_m \varpi_m^k, k = 1, \ldots, n \tag{28}$$

Intuitively, (28) says that a firm's weight is the weighted average over all output shares of this firm in the group, where the weights are the shares of the industry revenue from output m in the industry (observed) total revenue. When a_m is unavailable, then one might assume it is the same constant for all m, which would turn (28) into an equally weighted arithmetic average, as in Färe and Zelenyuk (2003).

We can now also use the standardization (27) and combine it with (26) to obtain the '*between* subgroups' weights as

$$S^l = \sum\nolimits_{m=1}^{M} a_m W_m^l, l = 1, \ldots, L \tag{29}$$

where, $W_m^l = \overline{Y}_m^l / \overline{Y}_m$ is l's subgroup share in the entire group in terms of mth output (analogous to what we had for the individual firms). This in turn helps in obtaining the weight for an individual efficiency of firm k '*within* subgroup l', by simply noting that

$$S^{l,k} = S^k / S^l, k = 1, \ldots, n_l, l = 1, \ldots, L \tag{30}$$

Intuitively, it is exactly what we had in (28) except that we now account for the weight of the particular group in the entire group. 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 knowledge of $TE^k(\cdot)$ or/and $R^k(\cdot)$, or at least their values

DOI: 10.1002/jae

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 of 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 (which uses output orientation, and assumes variable returns to scale and free disposability of inputs and outputs) and only on the estimation (and bootstrap) of *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.[7] 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 are 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

$$\partial\hat{P}(x) = \{y \in \Re^M_+ : y \in \hat{P}(x), \lambda y \notin \hat{P}(x), \lambda \in (1, \infty)\} \tag{31}$$

where

$$\hat{P}(x) = \{y \in \Re^M_+ : \sum_{k=1}^{n} z_k y_m^k \geq y_m, m = 1, \ldots, M,$$
$$\times \sum_{k=1}^{n} z_k x_n^k \leq x, n = 1, \ldots, \mathrm{N}, z_k \geq 0, k = 1, \ldots, n, \sum_{k=1}^{n} z_k = 1\} \tag{32}$$

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):

$$T\hat{E}(x, y) = \max_{\theta, z_1, \ldots, z_n} \{\theta : \theta y \in \hat{P}(x)\} \tag{33}$$

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

$$\hat{y}^{\partial}(x) \equiv y \cdot T\hat{E}(x, y) \tag{34}$$

While these DEA estimators can be applied to any point in the estimated output set, i.e., $\forall y \in \hat{P}(x)$, researchers are usually interested in the *observed* points, and thus apply the corresponding liner

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

DOI: 10.1002/jae

programming problem (33) 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, 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, $T\hat{E}(x, y)$ is a downward-biased estimator of $TE(x, y)$, i.e.:

$$1 \leq T\hat{E}(x, y) \leq TE(x, y), \forall y \in \hat{P}(x) \tag{35}$$

The asymptotic statistical properties have recently been discovered for the DEA estimator presented above. In particular, (32) 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 in defining the data-generating process (DGP) and converting our economic model of production into a statistical model. Before listing assumptions, we represent the problem by using the *polar* coordinates of $y \in \Re_+^M$ defined by the modulus $\omega = \omega(y) \in \Re_+^1$, where $\omega(y) \equiv \sqrt{y'y}$, and the angle $\eta \equiv \eta(y) \in [0, \pi/2]^{M-1}$, where $\eta_m \equiv \arctan(y_{m+1}/y_1)$ *if* $y_1 > 0$ or $\eta_m \equiv \pi/2$, if $y_1 = 0$ for $m = 1, \ldots, M-1$.[8] The following assumptions, adapted to our context from Kneip *et al.* (1998), define the DGP we will work with.

A1. $\{(x^k, y^k) : k = 1, \ldots, n\}$ are independent random variables on the convex technology set T. All observations $\{(x^k, y^k) : k = 1, \ldots, n\}$ can be partitioned into L subsamples (by some exogenous criterion) such that each subsample l $(l = 1, \ldots, L)$ represents a distinct subgroup l $(l = 1, \ldots, L)$ of interest that exists in the population (e.g., public vs. private firms in an industry).
A2. For all l $(l = 1, \ldots, L)$, the inputs $x \in \Re_+^N$ has density $f_{x,l}(x)$, with compact support $\aleph \subseteq \Re_+^N$.
A3. For all l $(l = 1, \ldots, L)$ and all $x \in \aleph$, the vector $\eta \equiv (\eta_1, \ldots, \eta_{M-1})$ has a conditional p.d.f. $f_{(\eta|x),l}(\eta|x)$ on $[0, \pi/2]^{M-1}$ and the modulus ω has a conditional p.d.f. $f_{(\omega|\eta,x),l}(\omega|\eta, x)$.
A4. For all l $(l = 1, \ldots, L)$, all $x \in \aleph$, and all $\eta \in [0, \pi/2]^{M-1}$ there exist constants $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ such that $\forall\omega \in [\omega(y^\partial(x)), \omega(y^\partial(x)) - \varepsilon_2]$, $f_{(\omega|\eta,x),l}(\omega|\eta, x) \geq \varepsilon_1, l = 1, \ldots, L$, where $y^\partial(x)$ was defined in (34).
A5. The technical efficiency measure $TE(x, y)$ is differentiable in both its vectors.

Since all the groups have access to the same technology T, the lower boundary of the support $f_{(\omega|\eta,x),l}(\omega|\eta, x)$ is the same for each group, $l = 1, \ldots, L$, so that

$$\omega(y^\partial(x)) = \sup\{\omega \in \Re_+^1 : f_{(\omega|\eta,x),l}(\omega|\eta, x) > 0\} \tag{36}$$

[8] Our way of defining the polar coordinates is the one that has been chosen in all previous works on analyzing the statistical properties of DEA estimators. It is not the most standard one and any other definition could be used. This would not change the argument that follows. For instance, as pointed out by an anonymous referee, one might choose the polar parameterization defined in Anderson (1984, p. 279).

and the relation between $\omega(y)$ and technical efficiency at (x, y), $TE(x, y)$, is now

$$TE(x, y) = \omega(y^{\partial}(x))/\omega(y) \tag{37}$$

Although the same efficiency measure (37) is applied for all firms in the population, the efficiency score it yields for particular points may have different densities $f_{(\omega|\eta,x),l}(\omega|\eta, x)$ for different subpopulations l $(l = 1, \ldots, L)$. From now on, superscripts k and l on the measure (37) will imply that this measure has been applied to a firm k belonging to group l $(l = 1, \ldots, L)$.

Thus, A3 along with (37) is implied the existence of a conditional (on (η, x)) density for $TE^l(x, y), l = 1, \ldots, L$ (with support $[1, \infty]$), which we will denote by $f_l(TE|\eta, x)$. Moreover, A4 along with (37), implies that $f_l(TE|\eta, x) \geq \varepsilon_1$, $\forall TE \in [1, 1 + \varepsilon_2], l = 1, \ldots, L$. Finally, with assumptions A1–A5, DGP, denoted by $\wp = \wp(P(x), g_l(TE, \eta, x), l = 1, \ldots, L)$ is completely defined through the *joint* densities of (TE, η, x), for all subgroups $l = 1, \ldots, L$:

$$g_l(TE, \eta, x) = f_l(TE|\eta, x) f_{(\eta|x),l}(\eta|x) f_{x;l}(x),\ l = 1, \ldots, L \tag{38}$$

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

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

$$T\hat{E}^l(x, y) - TE^l(x, y) = O_P(n^{-2/(M+N+1)}) \tag{39}$$

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 subgroups. In economic terms, such DGP can be well justified. Different subgroups 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 subgroups 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 from 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 subgroups, 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 for estimation of unknown sampling distribution of an estimator by means of resampling from original data. The theory of statistical bootstrap was originated by Efron (1979) and has been developed in many studies since then.[9] Perhaps the most encouraging

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

DOI: 10.1002/jae

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. For 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 dataset $\Xi_n = \{(x^k, y^k) : k = 1, \ldots, n\}$ generated from a DGP $\wp = \wp(P(x), g(TE, \eta, x))$ that satisfies the assumptions imposed above. Of course, no one of $P(x)$, $g(TE, \eta, x)$, 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 (33). We can then aggregate the estimates of $T\hat{E}$ over all firms 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 in now is the sampling distribution of $\overline{T\hat{E}} - \overline{TE}|\wp$. 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, $\Xi_n^* = \{(x^{*k}, y^{*k}) : k = 1, \ldots, n\}$, drawn randomly (with replacement) from this population, Ξ_n. Since we have all the population, Ξ_n, we can learn everything about the distribution of Ξ_n^*. In particular, we can use the same formula for estimation of technical efficiency as that applied to the original sample (33), 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 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

$$\overline{T\hat{E}^*} - \overline{T\hat{E}}|\hat{\wp} \overset{asy.}{\sim} \overline{T\hat{E}} - \overline{TE}|\wp \tag{40}$$

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 (see Simar and Wilson, 2000b, 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 resampling *not* from the original input–output data but from the (kernel-estimated) density

DOI: 10.1002/jae

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 $f(TE|\eta, x) = f(TE)$, 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 $g(TE, \eta, x)$. Both variants can be extended to cover the group-wise heterogeneous case as ours.

Most recently, Kneip *et al.* (2003a) offered another alternative—the subsampling (with replacement) bootstrap—and, most importantly, showed that it is *consistent* (for any subsample 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 subsampling, 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 subsampling bootstrap in the DEA context is that the choice of subsample size is not clear at this point. Although consistency of the subsampling bootstrap for DEA is proven for any subsample size smaller than the original, precision of bootstrap estimates in finite samples may be different for different subsample sizes (see Kneip *et al.*, 2003a). This problem is intuitively similar to the problem of 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 by Kneip *et al.* (2003b) investigates the use of iterated bootstrap to select the appropriate subsample size.

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

4.1. Algorithm of Group-Wise Heterogeneous Subsampling Bootstrap of Aggregates of DEA Efficiency Scores

1. For each observation in the sample $\Xi_n^{=}\{(x^k, y^k) : k = 1, \ldots, n\}$ compute $T\hat{E}(x, y)$ from (33), obtaining $\{T\hat{E}(x^k, y^k) : k = 1, \ldots, n\}$.
2. Aggregate the estimates of individual efficiencies from step 1 into the L subgroup estimated aggregate measures of technical efficiency using formulas outlined in Section 2.2.
3. Obtain the bootstrap sequence $\Xi^*_{s_l,b} = \{(x_b^{*k}, y_b^{*k}) : k = 1, \ldots, s_l\}$ (b denotes the bootstrap iteration, $b = 1, \ldots, B$), by subsampling with replacement independently from data on each subgroup l of the original sample, $\Xi_{n_l}^{=}\{(x^k, y^k) : k = 1, \ldots, n_l\}$, where $s_l \equiv (n_l)^\kappa, \kappa < 1, l = 1, \ldots, L$.
4. Compute the bootstrap estimates of $T\hat{E}(x, y)$ via (33), using the bootstrapped sample $\Xi^*_{n,b}$ obtained from step 3; call them $T\hat{E}_b^{*l,k}$, for $k = 1, \ldots, s_l < n_l$, for all $l = 1, \ldots, L$.
5. Compute the bootstrap estimates of the aggregate efficiency scores, using

$$\overline{T\hat{E}_b^{*l}} = \sum\nolimits_{k=1}^{s_l} T\hat{E}_b^{*l,k} \cdot S_b^{*l,k},\ \text{where} S_b^{*l,k} = \frac{p y_b^{*l,k}}{p \sum_{k=1}^{s_l} y_b^{*l,k}}, k = 1, \ldots, s_l < n_l \tag{41a}$$

DOI: 10.1002/jae

and

$$\overline{T\hat{E}_b^*} = \sum_{l=1}^{L} \overline{TE_b^{*l}} \cdot S_b^{*l}, \text{ where } S_b^{*l} = \frac{p \sum_{k=1}^{s_l} y_b^{*l,k}}{p \sum_{l=1}^{L} \sum_{k=1}^{s_l} y_b^{*l,k}}, l = 1, \ldots, L \tag{41b}$$

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

$$S_b^{*l} = \frac{1}{M} \sum_{m=1}^{M} \frac{\sum_{k=1}^{s_l} y_{m,b}^{*l,k}}{\sum_{l=1}^{L} \sum_{k=1}^{s_l} y_{m,b}^{*l,k}}, l = 1, \ldots, L$$

and

$$S_b^{*l,k} = \frac{1}{M} \sum_{m=1}^{M} \frac{y_m^{l,k}}{\sum_{k=1}^{s_l} y_{m,b}^{*l,k} \cdot S_b^{*l}}, k = 1, \ldots, s_l < n_l, l = 1, \ldots, L \tag{42}$$

6. Repeat 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 $\{\overline{T\hat{E}_b^{*l}}\}_{b=1}^{B}$ for each subgroup l ($l = 1, \ldots, L$) and of the entire group $\{\overline{T\hat{E}_b^{*}}\}_{b=1}^{B}$. 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

$$\overline{T\hat{E}^{*l}} - \overline{T\hat{E}^{l}} | \hat{\wp} \overset{asy.}{\sim} \overline{T\hat{E}^{l}} - \overline{TE^{l}} | \wp \tag{43}$$

which is satisfied if the bootstrap is consistent. Given this, the true confidence interval given by $\Pr\left(-b_\alpha \le \overline{T\hat{E}^l} - \overline{TE^l} \le -a_\alpha | \wp\right) = 1 - \alpha$ can be approximated by its bootstrap analogue:

$$\Pr\left(-\hat{b}_\alpha \le \overline{T\hat{E}^{*l}} - \overline{T\hat{E}^{l}} \le -\hat{a}_\alpha | \hat{\wp}\right) = 1 - \alpha \tag{44}$$

where α is the significance level (size of the test) chosen by the researcher, and $\hat{b}_\alpha$ and $\hat{a}_\alpha$ are obtained as the endpoints of the truncated sorted (in ascending order) list of $\left(\overline{T\hat{E}_b^{*l}} - \overline{T\hat{E}^{l}}\right)$,

DOI: 10.1002/jae

$b = 1, \ldots, B$, where the truncation is done by deleting $(\alpha/2) \times 100\%$ 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 α, is therefore[10]

$$\overline{T\hat{E}^l} + \hat{a}_\alpha \leq \overline{TE^l} \leq \overline{T\hat{E}^l} + \hat{b}_\alpha \tag{45}$$

To obtain a bias-corrected estimate of aggregate efficiency, again relying on (43), we note that the true bias $\text{Bias}(\overline{T\hat{E}^l}|\wp) = E\left(\overline{T\hat{E}^l}\right) - \overline{TE^l}$ can be approximated by its bootstrap analogue:

$$\text{Bias}(\overline{T\hat{E}^{*l}}|\hat{\wp}) = E\left(\overline{T\hat{E}^{*l}}\right) - \overline{T\hat{E}^l} \tag{46}$$

where $E\left(\overline{T\hat{E}^{*l}}\right)$ can be approximated by its Monte Carlo analogue (from the original bootstrap procedure) $\overline{\overline{T\hat{E}^{*l}}} \equiv \frac{1}{B}\sum_{b=1}^{B} \overline{T\hat{E}_b^{*l}}$; therefore the estimated bias is $\hat{\text{B}}\text{ias}(\overline{T\hat{E}^{*l}}|\hat{\wp}) = \overline{\overline{T\hat{E}^{*l}}} - \overline{T\hat{E}^l}$ and the resulting bias-corrected estimate of the aggregate efficiency is

$$\overline{T\check{E}^l} = \overline{T\hat{E}^l} - \hat{\text{B}}\text{ias}(\overline{T\hat{E}^{*l}}|\hat{\wp}) = 2\overline{T\hat{E}^l} - \overline{\overline{T\hat{E}^{*l}}} \tag{47}$$

Finally, the standard error of $\overline{T\hat{E}^l}$ can be computed as

$$\text{SE}(\overline{T\hat{E}_b^{*l}}) \equiv \frac{1}{B-1}\left[\sum_{b=1}^{B}\left(\overline{T\hat{E}_b^{*l}} - \overline{\overline{T\hat{E}^{*l}}}\right)^2\right]^{1/2} \tag{48}$$

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

4.3. Test for Equality of Aggregate Efficiencies of Two Subgroups

In empirical literature, when making a judgement on the 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 *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' (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 subgroups in an industry. The test can be based on a pair-wise comparison of the aggregate efficiencies of subgroups. For example, for groups A and Z we can postulate

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

[10] We have chosen for the presentation an equal-tailed confidence interval, but the procedure could be adapted to produce symmetric intervals around $\overline{T\hat{E}^l}$ or intervals constructed for being the shortest for a given confidence level.

DOI: 10.1002/jae

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 $\hat{RD}_{A,Z} = \overline{T\hat{E}^A}/\overline{T\hat{E}^Z}$, whose behaviour can be mimicked by its bootstrap analogue, $\hat{RD}^*_{A,Z,b} = \overline{T\hat{E}^{*A}_b}/\overline{T\hat{E}^{*Z}_b}$, $b = 1, \ldots, B$. 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 subsection. 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 subsample size $s_l \equiv (n_l)^\kappa$, $\kappa < 1$, $l = 1, \ldots, L$. In any case, reasonable precision was reached for values of κbetween 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 κ. To illustrate how our methods work we try three values of $\kappa = \{0.5, 0.6, 0.7\}$ and present only the result that gave the best performance among these choices.

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 (subgroups)—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 subgroups A and Z, the only input is assumed to come from Uniform(0,1).[11] We assume that $TE^{l,k} = 1 + u^{l,k}$, where $(u^{l,k}|x) \sim N^+(\mu_l, \sigma_l^2(x))$, $l = A, Z$, which we call the '*true*' inefficiencies. We choose $\mu_A > 0$, intuitively representing some pathological tendency for inefficiency existing for this group, say, of the state-owned firms (such a tendency could be justified with one of the economic theories of incentives, etc.). For simplicity, we assume $\sigma_A(x) = \sigma_A$. [12] 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., $\mu_Z = 0$, but also has $\sigma_Z(x) = \sigma_Z(1 - x)^2$. Such heteroskedasticity here can be motivated by vulnerability of this type of firms at low levels of operations (this can be supported by empirical evidence of, say, private firms having higher risk of being unstable when

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

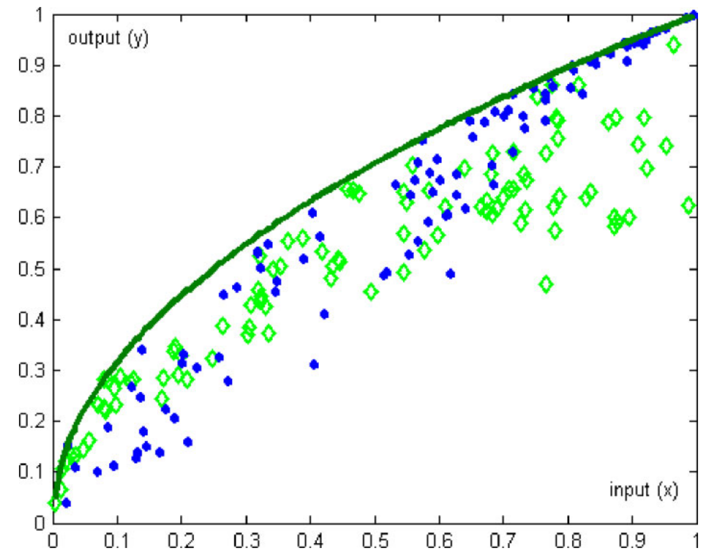

Figure 1. True technology set and observed firms for simulation example 1. *Note*: Groups A and Z are indicated by diamonds and dots, respectively. This figure is available in color online at www.interscience.wiley.com/journal/jae

they are small but becoming more stable as they grow). For the purpose of supporting the message conveyed by Table I in the Introduction, we set $\mu_A = 0.25, \sigma_A = 0.2, \sigma_Z = 1.4$.

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. Figure 2 gives the plots of estimated densities of efficiencies for each group. Specifically, part (i) compares densities estimated from the 'true realizations' of inefficiency for the two subgroups (generated from their distributions), and part (ii) does the same for densities estimated using the DEA estimates of efficiencies for the two subgroups. The two panels look very similar, supporting the fact that in the one-input–one-output case the rate of convergence of the DEA estimator is better than the usual parametric one. This is also supported by parts (iii) and (iv), which illustrate the difference between the densities estimated with the 'true' and with the DEA estimates.[13] The rugged shape observed for the subgroup Z is a consequence of our heteroskedasticity assumption. The results of bootstrapping (with 2000 replications) the aggregate efficiencies are given in Table II.

Perhaps the first thing that catches the eye in Table II 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 subgroup A is more efficient relative to subgroup Z, while the point estimates of the aggregate efficiencies (weighted means) tell us the opposite story. The bootstrap confidence intervals (CIs)

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

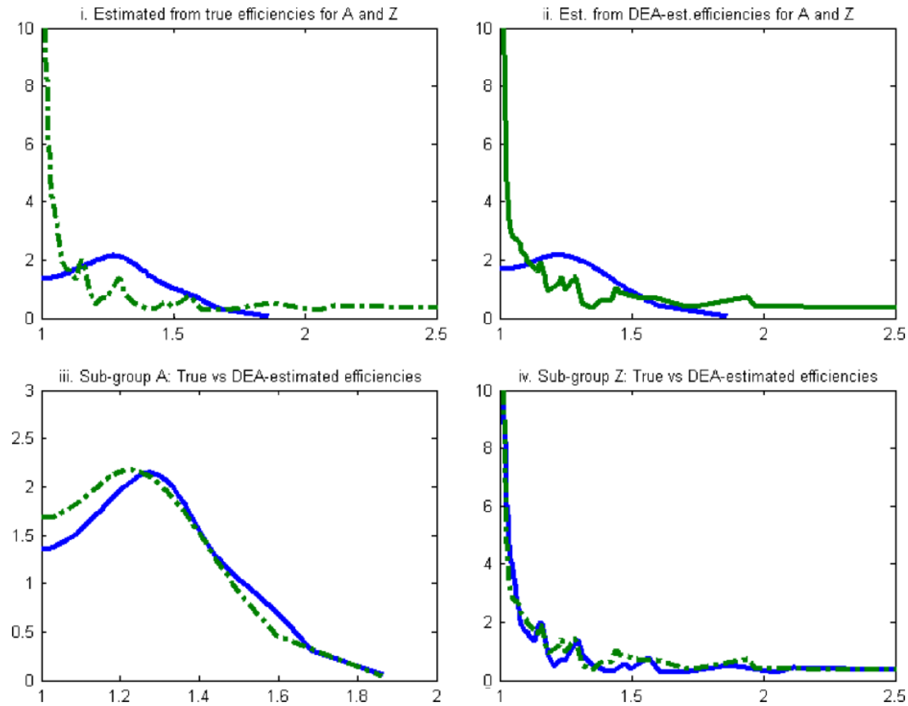

Figure 2. Kernel-estimated densities of efficiency scores for subgroups A and Z (simulation Example 1). This figure is available in color online at www.interscience.wiley.com/journal/jae

Table II. Estimation results for Example 1

	DEA Estim.	True Estim.	Bias corr. estim.	Estim. bias	Estim. SD	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
$\mathrm{RD}_{A,Z;\mathrm{Ag}}$	1.1149	1.1191	1.1199	−0.0050	0.0447	1.0340	1.2061
$\mathrm{RD}_{A,Z;\mathrm{Mean}}$	0.9776	0.9827	0.9621	0.0156	0.0678	0.8321	1.0994

Note: CI, confidence intervals are all at the 0.95 level; AgEf., aggregate efficiency; MeEf, mean efficiency; $\mathrm{RD}_{A,Z;\mathrm{Ag}}$ and $\mathrm{RD}_{A,Z;\mathrm{Mean}}$ are $\mathrm{RD}_{A,Z;}$ 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. $(\kappa_A, \kappa_Z) = (0.7, 0.7)$; 2000 bootstrap replications; estimated time = 1017.8 s.

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 subgroup efficiencies using the weighted means ($\mathrm{RD}_{A,Z;\mathrm{Ag}}$) and using the non-weighted means ($\mathrm{RD}_{A,Z;\mathrm{Mean}}$).

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

Example 2: Two Outputs–Two Inputs

Here we modify the example from Park *et al.* (2000). We assume the technology is characterized by 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 subgroups 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 subgroups, 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 $TE^{l,k} = 1 + u^{l,k}$, $(u^{l,k}|x) \sim N^+(\mu_l, \sigma_l^2(x))$, for all $k, l = A, Z$.

Here we assume that $\mu_A = -0.1$, $\mu_Z = 0$, and now inefficiency is heteroskedastic for *both* types of firm, but in a very different manner. For the Z-type firms $\sigma_Z(x) = \sigma_Z(1 - x_1)^{\gamma_Z}$, while for the A-type firms $\sigma_A(x) = \sigma_A \cdot (x_1)^{\gamma_A}$; 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 a decrease in a firm's efficiency). Again, for the purpose of supporting the message conveyed by Table I above, we set $\sigma_A = 0.5$, $\sigma_Z = 1.7$ and $\gamma_A = 3$, $\gamma_Z = 1/5$. Figure 3 visualizes this scenario with the kernel-estimated densities of true and DEA-estimated efficiencies for each group.

Table III presents the results of bootstrap (with 2000 replications) estimation and the first thing that must catch the eye here is that again the estimated aggregate efficiency scores are quite far

Table III. Estimation results for Example 2

	DEA estim.	True estim.	Bias corr. estim.	Estim. bias	Estim. SD	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

Note: 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. $(\kappa_A, \kappa_Z) = (0.7, 0.7)$; 2000 bootstrap replications; estimated time = 1827.1 s.

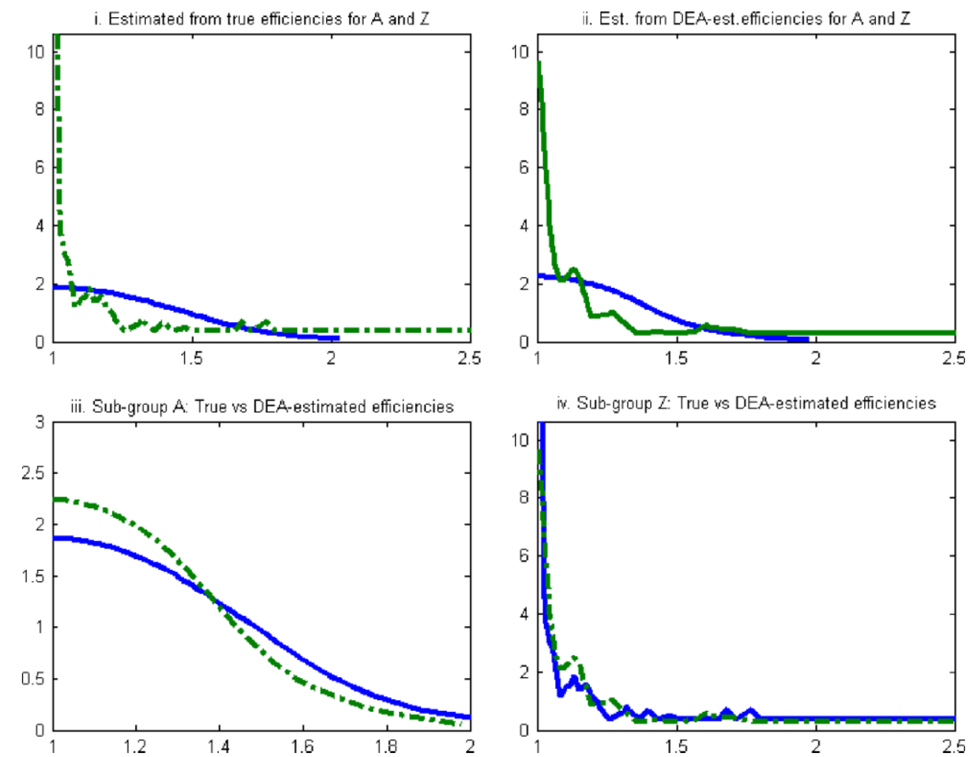

Figure 3. Kernel-estimated densities of efficiency scores for subgroups A and Z (simulation Example 2). This figure is available in color online at www.interscience.wiley.com/journal/jae

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 subgroups, 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 subgroups do differ significantly.

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

Example 3: Two Outputs–Two Inputs

Here we modify the example above under the assumption that the two groups share the same distribution. In particular, we assume that $\mu_A = \mu_Z = 0$ and $\sigma_A(x) = \sigma_Z(x) = 0.35$, and the rest is the same as in 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

DOI: 10.1002/jae

we obtained 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 (with 2000 replications) aggregate DEA scores are given in Table IV.

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

6. EMPIRICAL ILLUSTRATION

The goal of this section is to provide an empirical illustration using the methods developed above. The details of data sources, description and economic context can be found in a recent paper of

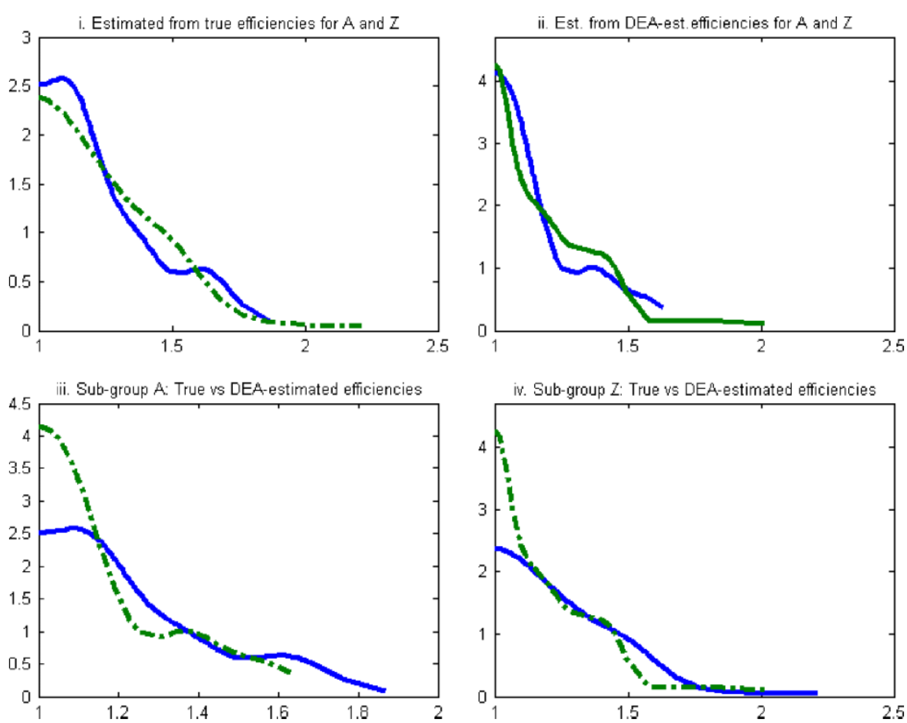

Figure 4. Kernel-estimated densities of efficiency scores for subgroups *A* and *Z* (simulation Example 3). This figure is available in color online at www.interscience.wiley.com/journal/jae

Table IV. Estimation results for Example 3

	DEA estim.	True estim.	Bias corr. estim.	Estim. bias	Estim. SD	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
$\text{RD}_{A,Z;\text{Ag}}$	0.9754	0.9858	0.9625	0.0130	0.0424	0.8753	1.0404
$\text{RD}_{A,Z;\text{Mean}}$	0.9816	0.9859	0.9734	0.0082	0.0472	0.8778	1.0596

Note: CI, confidence Intervals are all at the 0.95 level; AgEf., aggregate efficiency; MeEf, mean efficiency; $\text{RD}_{A,Z;\text{Ag}}$ and $\text{RD}_{A,Z;\text{Mean}}$ are $\text{RD}_{A,Z;}$ 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. $(\kappa_A, \kappa_Z) = (0.7, 0.7)$; 2000 bootstrap replications; estimated time = 971 s.

Zelenyuk and Zheka (2004).[14] In the dataset we use, there are 152 firms from various business sectors in Ukraine: Chemical, Construction, Engineering, Metallurgy, Services, Transport and Food Industries. For each firm we have information on (proxies of) three inputs (labour, capital and materials) and one proxy of all outputs produced (total revenue). The fact that all inputs and outputs are measured in (the same) monetary units gives us some justification for pooling the data over industries to measure the efficiency of each firm relative to one 'best-practice frontier'. In this sense, the technology set characterization (equation (1)) bears not an engineering, but a broader, economic meaning: this is a set characterizing all the possibilities to make 'revenue' out of investments into major inputs. In essence, we are modelling a technology of 'using money to make more money', regardless of particular engineering, managerial and other features of the business. In the light of the curse of dimensionality problem (see equation (39)), it should be clear that such pooling of the data over industries is also a practical necessity—to have reasonably large number of observations estimating one frontier. We are also willing to accept that each enterprise is there to strive to have as much of *revenue* from given inputs as possible, regardless of the specifics of the business, so that we choose the output-oriented measurement.

Many applied econometric studies have investigated the relationship between foreign ownership and performance of firms (e.g., see Yudaeva *et al*., 2000; Blomstrom and Kokko, 2003; Akimova and Schwödiauer, 2004; Javorcik, 2004; and earlier works cited therein). The debate is far from over and we decided also to try our methodology to tackle this interesting issue. In our dataset, 50 firms have some foreign share of ownership and the rest do not have any. Our goal would be to estimate (for the two groups and the entire sample) the original DEA estimates, their bootstrap-based bias-corrected analogues, standard errors and confidence intervals. Finally, we would like to test whether the group of firms with a foreign share of ownership experienced higher efficiency than the group of purely local firms. Application of the methods developed above yield the results summarized in Tables V and VI. In Table V we present results of DEA and bootstrap under the assumption that technology exhibits *constant* returns to scale (CRS); i.e., the last constraint in (32) is removed. Perhaps the first thing that catches the eye is a sizable difference between the original DEA scores and their bias-corrected analogues for both types of group measures.

[14] We thank Vitaliy Zheka for generously providing us with the data set for this illustration.

DOI: 10.1002/jae

Table V. Results for empirical application (CRS case)

	DEA estim.	SE	Bias corr. estim.	95% CI. bounds		99% CI bounds		90% CI bounds	
				Lower	Upper	Lower	Upper	Lower	Upper
AgEf. A	1.50	0.14	1.79	1.46	1.98	1.32	2.00	1.52	1.97
AgEf. Z	1.71	0.22	2.08	1.52	2.36	1.19	2.39	1.67	2.33
AgEf.	1.58	0.13	1.90	1.60	2.10	1.39	2.12	1.66	2.08
MeEf. A	1.78	0.19	2.22	1.73	2.44	1.57	2.47	1.81	2.41
MeEf. Z	1.78	0.17	2.22	1.78	2.41	1.70	2.43	1.85	2.39
MeEf.	1.78	0.17	2.22	1.81	2.39	1.72	2.42	1.85	2.38
$RD_{A,Z;Ag}$	0.88	0.14	0.84	0.53	1.11	0.40	1.19	0.60	1.06
$RD_{A,Z;Mean}$	1.00	0.08	0.99	0.82	1.15	0.74	1.20	0.85	1.13

Note: CI, confidence intervals are all at the 0.95 level; 2000 bootstrap replications; 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; $(\kappa_A, \kappa_Z) = (0.7, 0.7)$; group A = group of firms with foreign share of ownership, group Z is otherwise.

Table VI. Results for empirical application (VRS case)

	DEA estim.	SE	Bias corr. estim.	95% CI bounds		99% CI bounds		90% CI bounds	
				Lower	Upper	Lower	Upper	Lower	Upper
AgEf. A	1.11	0.04	1.16	1.06	1.22	1.00	1.22	1.08	1.21
AgEf. Z	1.17	0.14	1.19	0.80	1.33	0.67	1.50	0.88	1.33
AgEf.	1.13	0.04	1.19	1.09	1.25	1.02	1.29	1.11	1.25
MeEf. A	1.39	0.09	1.57	1.34	1.70	1.20	1.73	1.39	1.69
MeEf. Z	1.53	0.09	1.84	1.59	1.96	1.51	1.99	1.64	1.95
MeEf.	1.48	0.08	1.74	1.54	1.85	1.49	1.86	1.58	1.84
$RD_{A,Z;Ag}$	0.95	0.12	0.97	0.78	1.22	0.62	1.27	0.81	1.19
$RD_{A,Z;Mean}$	0.91	0.08	0.82	0.66	0.99	0.60	1.04	0.70	0.96

Note: 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; $(\kappa_A, \kappa_Z) = (0.7, 0.7)$; group A = group of firms with foreign share of ownership, group Z is otherwise.

An interesting coincidence is that the group efficiency scores (original and bias corrected) of the two groups are equal (up to the second digit) when the equally weighted aggregation is used, so, not surprisingly, the confidence interval for the RD statistic overlaps with unity, suggesting that the two groups are not statistically different with respect to their group efficiency scores. However, when the weighted average is applied to obtain the group efficiency scores, we get a different picture. The group of firms with foreign ownership has a group efficiency score that is lower (i.e., closer to perfect efficiency) than the group of purely local firms. Despite a sizable difference, the confidence interval for the RD statistic in the weighted aggregation case also overlaps with unity. Thus we cannot reject the hypothesis (even at the 10% level) that the efficiency scores of the two groups (foreign and local) are the same. A similar conclusion is reached by looking at the (overlap of) estimated confidence intervals of the group efficiencies.

In Table VI we present results of DEA and bootstrap estimation under the assumption that technology exhibits *variable* returns to scale (VRS), i.e., as stated in (32). One thing that catches the eye is that the efficiency scores are smaller than those we observed for the CRS case. This is not surprising since VRS imposes more constraints in the maximization problem (33) than CRS does, and now some firms that were inefficient under CRS become less inefficient or even have unity efficiency score.

For both types of aggregation, we now observe the group of firms with foreign ownership having less inefficiency. However, only for the case of non-weighted aggregation, such a difference is statistically significant (at the 5% level). Also note that the group weighted efficiency scores are substantially smaller than the equally weighted group efficiency scores (especially after bias correction). This was also the case under CRS, but here it is more pronounced. It suggests a tendency of the largest firms in the sample (that have higher weight) to be more efficient, under CRS and especially under VRS.

It should be noted that we presented results only for $(\kappa_A, \kappa_Z) = (0.7, 0.7)$, but we also tried other levels from 0.6 to 0.8 and the conclusions were very robust. Also, one of the reasons for not rejecting the hypothesis of equality of group efficiency scores might be the relatively small size of the entire sample (and, in particular, of the sample of group of firms with foreign ownership). However, it is also not clear what the averages would be if we had more observations.

Further analysis is desirable to understand the reasons of our evidence, but our goal is only an illustration of our methods, while more analysis can be found in Zelenyuk and Zheka (2004).

7. 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 subgroups of (e.g., public vs. private, foreign vs. domestic) firms within a population. Simulations that we have tried, a few of which we presented here, suggest that the proposed methodology have 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. For the sake of space we have presented only a small application in this paper, but the interested reader is referred to more applications of these (along with other) methods in Henderson and Zelenyuk (2004) and Zelenyuk and Zheka (2004).

A natural extension of this work would be to provide extensive Monte Carlo evidence of performance of the proposed methodology for various scenarios. Although it is quite a time-expensive exercise in itself, given that just one Monte Carlo replication takes about 1000–2000 seconds (for an 1133 MHz, 253 MB machine) for a 2×2 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 subsampling bootstrap application) is to develop a data-driven method of choosing the subsample size, especially that which maximizes the power of the proposed test for comparison of aggregate efficiencies among different subgroups. This could be done along the lines of the iterated bootstrap procedure proposed by Kneip *et al*. (2003b). Yet another natural extension of our work would be an intertemporal extension: developing a reliable bootstrap procedure for statistical inference on aggregate Malmquist quantity as well as Malmquist and Hicks–Moorsteen productivity indexes (for related aggregation results, see Zelenyuk, 2005).

Overall, we believe that the methods developed above will prove to be very useful techniques for applied econometric studies of efficiency of various economic systems and their distinct groups.

DOI: 10.1002/jae

APPENDIX: PROOF OF EQUATION (14)

Recall that $x^k = (x_1^k, \ldots, x_N^k)' \in \Re_+^N$ and $y^k = (y_1^k, \ldots, y_M^k)' \in \Re_+^M$ 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 by an $(N \times n)$ matrix $X = (x^1, \ldots, x^n)$. 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 the 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 an $(N \times n_l)$ matrix $X^l = (x^{l,1}, \ldots, x^{l,n_l})$. In general, the technology of a particular firm k ($k = 1, \ldots, n_l$) within a group l is assumed to be characterized by the output sets:

$$P^{l,k}(x^{l,k}) \equiv \{y^{l,k} : x^{l,k} \text{ can produce } y^{l,k}\}, x^{l,k} \in \Re_+^N \tag{A.1}$$

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

$$\overline{P}^l(X^l) \equiv \sum_{k=1}^{n_l} P^{l,k}(x^{l,k}) \tag{A.2}$$

which yields one of the main aggregation results we stated in the text as (14):

$$\overline{R}^l(X^l, p) = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \tag{A.3}$$

where

$$R^{l,k}(x^{l,k}, p) \equiv \max_y\{py : y \in P^{l,k}(x^{l,k})\} \tag{A.4}$$

and

$$\overline{R}^l(X^l, p) \equiv \max_y\{py : y \in \overline{P}^l(X^l)\} \tag{A.5}$$

To prove this, for each $k = 1, \ldots, n_l$, 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 (A.2) we have $\overline{Y}^l \in \overline{P}^l(X^l)$, and due to (A.5) we obtain

$$p\overline{Y}^l \le \overline{R}^l(X^l, p) \tag{A.6}$$

Since $y^{l,k}$ is an arbitrary vector in $P^{l,k}(x^{l,k})$, it implies that (A.6) also holds for those $y^{l,k}$ that solve (A.4)—call them $\tilde{y}^{l,k}$, in which case we would have

$$p\overline{\tilde{Y}}^l \equiv \sum_{k=1}^{n_l} p\tilde{y}^{l,k} = \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \le \overline{R}^l(X^l, p) \tag{A.7}$$

On the other hand, let $\overline{Y}^l$ be an arbitrary vector in $\overline{P}^l(X^l)$, then due to (A.2) there exist $y^{l,k} \in P^{l,k}(x^{l,k})$, for each $k = 1, \ldots, n_l$, such that $\overline{Y}^l = \sum_{k=1}^{n_l} y^{l,k}$. Therefore, due to (A.4) we have

$$p\overline{Y}^l \equiv \sum_{k=1}^{n_l} py^{l,k} \le \sum_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \tag{A.8}$$

and since $\overline{Y}^l$ is an *arbitrary* vector in $\overline{P}^l(X^l)$, expression (A.8) is also true for those $\overline{Y}^l$ that solve (A.5)—call them $\tilde{\overline{Y}}^l$—in which case we get

$$p\tilde{\overline{Y}}^l \equiv \overline{R}^l(X^l, p) \leq \sum\nolimits_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) \tag{A.9}$$

Clearly, expressions (A.7) and (A.9) can simultaneously hold if and only if

$$\overline{R}^l(X^l, p) = \sum\nolimits_{k=1}^{n_l} R^{l,k}(x^{l,k}, p)$$

QED

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

$$\overline{R}(X, p) = \sum\nolimits_{k=1}^{n} R^k(x^k, p) \tag{A.10}$$

where

$$\overline{R}(X, p) \equiv \max\nolimits_y\{py : y \in \overline{P}(X)\} \tag{A.11}$$

and

$$\overline{P}(X) = \sum_{k=1}^{n} P^k(x^k) \tag{A.12}$$

Another important implication concerns the relationship between the maximal revenues of the subgroups to the maximal revenue of the entire group. In particular, since

$$\sum\nolimits_{l=1}^{L} \overline{R}^l(X^l, p) = \sum\nolimits_{\mathrm{l}=1}^{\mathrm{L}} \sum\nolimits_{k=1}^{n_l} R^{l,k}(x^{l,k}, p) = \sum\nolimits_{k=1}^{n} R^k(x^k, p) \tag{A.13}$$

thus along with (A.10) we get

$$\overline{R}(X, p) = \sum\nolimits_{l=1}^{L} \overline{R}^l(X^l, p) \tag{A.14}$$

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

$$\overline{RE} = \sum_{l=1}^{L} \overline{RE}^l \cdot S^l = \sum_{k=1}^{n} RE^k \cdot S^k, \overline{TE} = \sum_{l=1}^{L} \overline{TE}^l \cdot S^l = \sum_{k=1}^{n} TE^k \cdot S^k$$

where $S^k \equiv \dfrac{py^k}{p\overline{Y}}$, and

$$\overline{AE} = \sum_{l=1}^{L} \overline{AE}^l \cdot S^l_{ae} = \sum_{k=1}^{n} AE^k \cdot S^k_{ae}$$

DOI: 10.1002/jae

where $S_{ae}^{k} \equiv \dfrac{p\left(y^k TE^k(x^k, y^k)\right)}{p\sum_{k=1}^{n}\left(y^k TE^k(x^k, y^k)\right)}$, $k = 1, \ldots, n$.

ACKNOWLEDGEMENTS

We thank Tom Coupé, Roy Gardner, Alexander Romanov, two anonymous referees, as well as participants of seminars/workshops at the Institut de Statistique, Université Catholique de Louvain, EERC (Ukraine), IAMO (Germany), and the 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 is also acknowledged. We also thank Vitaliy Zheka for generously providing us with data for the empirical illustration. We remain responsible for the views expressed and mistakes made.

REFERENCES

Akimova I, Schwödiauer G. 2004. Ownership structure, corporate governance and enterprise performance: empirical results for Ukraine. *International Advances in Economic Research* **10**(1): 28–42.

Anderson TW. 1984. *An Introduction to Multivariate Statistical Analysis* (2nd edn). Wiley: New York.

Blackorby C, Russell R. 1999. Aggregation of efficiency indices. *Journal of Productivity Analysis* **12**(1): 5–20.

Blomstrom M, Kokko A. 2003. The economics of foreign direct investment incentives. NBER working paper no. W9489. http://ssrn.com/abstract=379960 [14 October 2005].

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, Primont D. 1995. *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic: Boston, MA.

Färe R, Zelenyuk V. 2003. On aggregate Farrell efficiency scores. *European Journal of Operational Research* **146**(3): 615–620.

Färe R, Grosskopf S, Zelenyuk V. 2002. Aggregation of cost efficiency indicators and indexes across firms. *Academia Economic Papers* **32**(2): 395–411.

Farrell MJ. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A, General* **120**(3): 253–281.

Gijbels I, Mammen E, Park BU, Simar L. 1999. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* **94**: 220–228.

Henderson D, Zelenyuk V. 2004. Testing for catching-up: statistical analysis of DEA efficiency estimates. Disc. Paper #0431, Institut de Statistique, Université Catholique de Louvain, Belgium.

Horowitz JL. 2002. The bootstrap. In *Handbook of Econometrics*, Vol. 5, Heckman J, Leamer E (eds). Elsevier: Amsterdam; 3159–3223.

Javorcik BS. 2004. Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review* **94**: 605–627.

Li SK, Ng YC. 1995. Measuring the productive efficiency of a group of firms. *International Advances in Economic Research* **1**(4): 377–390.

Kneip A, Park B, Simar L. 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* **14**: 783–793.

Kneip A, Simar L, Wilson P. 2003a. Asymptotics for DEA estimators in non-parametric frontier models. Discussion Paper #0317, Institut de Statistique, Université Catholique de Louvain, Belgium. *Econometric Theory*.

DOI: 10.1002/jae

Kneip A, Simar L, Wilson P. 2003b. A simplified bootstrap for DEA estimators. Paper presented at the *8th European Workshop on Efficiency and Productivity Analysis*, Oviedo, Spain, September 2003.
Korostelev A, Simar L, Tsybakov A. 1995. Efficient estimation of monotone boundaries. *Annals of Statistics* **23**(2): 476–489.
Koopmans TC. 1957. *Three Essays on the State of Economic Analysis*. McGraw-Hill: New York.
Park B, Simar L, Weiner Ch. 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 Functions*. Princeton University Press: Princeton, NJ.
Sheather SJ, Jones MC. 1991. A reliable data based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* **53**: 683–690.
Silverman BW. 1996. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall: New York.
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, Wilson P. 1998. Sensitivity of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Sciences* **44**(1): 49–61.
Simar L, Wilson P. 2000a. A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics* **27**: 779–802.
Simar L, Wilson P. 2000b. 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.
Yudaeva K, Kozlov K, Melentieva N, Ponomareva N. 2003. Does foreign ownership matter? Russian experience. *Economics of Transition* **11**(3): 383–409.
Zelenyuk V. 2005. On aggregation of Malmquist productivity indexes. Discussion Paper 0505 of Institute of Statistics, University Catholique de Louvain, Belgium. *European Journal of Operational Research* **174**: 1076–1086.
Zelenyuk V, Zheka V. 2004. Corporate governance and firm's efficiency: the case of a transitional country, Ukraine. Disc. Paper #0429, Institut de Statistique, Université Catholique de Louvain, Belgium, *Journal of Productivity Analysis* **25**: 143–168.

DOI: 10.1002/jae