SORT 43 (2) July-December 2019, 337-354 DOI: 10.2436/20.8080.02.91

Data envelopment analysis efficiency of public services: bootstrap simultaneous confidence region

Jesús A. Tapia[1], Bonifacio Salvador[1] and Jesús M. Rodríguez[2]

Abstract

Public services, such as higher education, medical services, libraries or public administration offices, provide services to their customers. To obtain opinion-satisfaction indices of customers, it would be necessary to survey all the customers of the service (census), which is impossible. What is possible is to estimate the indices by surveying a random customer sample. The efficiency obtained with the classic data envelopment analysis models, considering the opinion indices of the customers of the public service as output data estimated with a user sample, will be an estimation of the obtained efficiency if the census is available. This paper proposes a bootstrap methodology to build a confidence region to simultaneously estimate the population data envelopment analysis efficiency score vector of a set of public service-producing units, with a fixed confidence level and using deterministic input data and estimated customer opinion indices as output data. The usefulness of the result is illustrated by describing a case study comparing the efficiency of libraries.

MSC: 62D05, 90C99.

Keywords: Data envelopment analysis, sampling survey research, public sector, bootstrap, simultaneous confidence region.

1 Introduction

Data envelopment analysis (DEA) is clearly of enormous potential in measuring public sector efficiency, particularly in areas where there exists a large number of agencies to compare (see Smith and Mayston, 1987 or chapter 15 of Cooper, Seiford and Zhu, 2011). In this context, DEA efficiency is usually evaluated using determinist input/output data. However, the quality of the service delivered by a provider can therefore have important implications and available results. Bayraktar et al. (2012), Witte and Geys (2013), Mayston (2015, 2017), Santín and Sicilia (2017) and Førsund (2017) analyse the efficiency of any individual public services producer, where output

[1] Department of Statistic and Operative Research, University of Valladolid, Campus Miguel Delibes, Paseo Belén nº 7,47011 Valladolid, Spain.

[2] Statistical & Budgetary General Management of the Regional Government of Castile & Leon, Regional Tax Office, José Cantalapiedra nº 2, 47014 Valladolid, Spain

Received: July 2018
Accepted: October 2019

variables can include the quality of the service produced and the said quality is measured by a consumer satisfaction survey. Tapia, Salvador and Rodríguez (2018) study the relationship between customer sample size and accuracy in estimating the efficiency of public services. Of course, a public service is more efficient when, with its resources, it is able to achieve the highest opinion-satisfaction of its users. Customer opinion-satisfaction surveys are widely used tools to measure the perception of the quality of the service (Parasuraman, Berry and Zeithaml, 1993), and to obtain outputs and the DEA efficiency scores (Lee and Kim, 2014). In this paper, using confidence regions, we estimate the DEA efficiency of a fixed (not random) set of public service-producing units, i.e., our decision-making units (DMUs). We do so using indices of the service quality obtained as the mean of the answers given by the sample of surveyed people in the opinion-satisfaction survey as the data output, and the resources of the services measured in a deterministic way as the data input. For example, when comparing the DEA efficiency of all the cinemas in a city, user opinion on the quality of each cinema is measurable with opinion indices estimated using a survey of the cinemagoers to know their satisfaction with the location, the staff, the state of the cinema, etc. The number of seats, screens in the cinema, daily movies on show, or monthly premiering movies are the resources of the service.

The studies where the DEA efficiency is evaluated in the presence of sampling information have had two approaches until now. In the first one, the set of DMUs from which input/output information is known is considered as a sample of a population of DMUs and the randomness comes from the DMU sample. In this approach, Banker (1993), Simar and Wilson (1998, 2000, 2007, 2011, 2013, 2015), Kneip, Park and Simar (1998) and Kneip, Simar and Wilson (2008), have proved statistical properties of the nonparametric estimators used to estimate the productivity efficiency of DMUs, derived the asymptotic distribution of DEA estimators and tested hypotheses about the structure of the underlying nonparametric model.

In the second approach, samples are used to estimate input and/or output data. The efficiency is evaluated using linear programming (LP) problems subject to constraints defined in terms of probability, or chance-constrained problems. A great number of papers have reported a wide range of uses of chance-constrained programming, including: Charnes and Cooper (1959, 1963), Land, Lovell and Thore (1993), Olesen and Petersen (1995), Cooper, Huang and Li (1996), Cooper et al. (2002), Huang and Li (2001), Wu and Olson (2008), Khodabakhshi and Asgharian (2009), Khodabakhshi (2010), Wu and Lee (2010), Wu (2010) and Tavana, Shiraz and Hatami-Marbini (2014). Charles and Kumar (2014) introduced a chance-constrained model to measure the stochastic efficiency of the service quality.

In this paper, we assume a fixed set of homogeneous DMUs in terms of the nature of the operations they perform, the measures of their efficiency, and the conditions under which they operate, as in the classic DEA models (Charnes, Cooper and Rhodes, 1978). The randomness comes solely from the customer sample in each DMU with which we estimate the output data. However, to evaluate the DEA efficiency with a bootstrap confidence region, we do not use chance-constrained programming, only the classic DEA models with constant (CCR) and variable returns-to-scale (BCC), i.e., LP problems subject to deterministic constraints.

Using estimated output data with a sample of customers instead of population output data causes an estimation error to be transferred to the evaluation of the DEA efficiency (Ceyhan and Benneyan, 2014). The vector of DEA efficiency scores obtained is, therefore, an estimation of the vector of the population DEA efficiency scores that would be obtained if we had the customer population data, i.e., if the output data were obtained with a customer census in each DMU. In our study, we solve the problem of determining how many customers need to be surveyed in order to estimate the output data in each DMU, with a previously fixed estimation error, when estimating the vector of population DEA efficiency scores with a bootstrap simultaneous confidence region. With the same assumptions as in this study, Tapia et al. (2018) obtained the customer sample size needed in each DMU to estimate the population DEA efficiency with a fixed accuracy in each DMU; while in this paper, the customer sample size necessary in each public service-producing unit is determined so that the maximum efficiency estimation error in the service-producing units will be smaller than a previously fixed value.

Using bootstrap, smooth bootstrap or double-smooth bootstrap methodologies to evaluate the efficiency of the public sector with confidence intervals is not new (Simar and Wilson, 1998, 2000, Simar and Zelenyuk, 2006, Kneip, Simar and Wilson, 2011). For instance, these methodologies have been used to measure the efficiency in health care (Tsekouras et al., 2010, Chowdhury and Zelenyuk, 2016), universities and research institutes (Barra and Zotti, 2016), government (Benito, Solana and Moreno, 2014), public libraries (Liu and Chuang, 2009), schools (Essid, Ouellette and Vigeant, 2014, Alexander, Haug and Jaforullah, 2010), tourism (Assaf and Agbola, 2011), banks (Casu and Molyneux, 2003) or public transport services (Assaf, 2010, Gil, Turias and Cerbán, 2019). In all these references, the different bootstrap resampling techniques are used considering the observed DMUs to be a sample taken from a population of DMUs and the resampling is done over the estimated efficiencies. In our study, we consider a fixed (not random) set of services, a customer sample in each service to estimate the client opinion indices (outputs) and a bootstrap resampling on the customer sample. As far as we know, the bootstrap efficiency simultaneous confidence region introduced in this paper has not been attempted in the literature. Our confidence region is the product of intervals and these intervals allow efficiency rankings, dominance relations and efficiency bounds to be determined as in Salo and Punkka (2011).

The rest of the paper is organized as follows. The problem is introduced in Section 2. Section 3 studies the determination of the customer sample size in each public service, in order to achieve a fixed accuracy in the simultaneous DEA efficiency estimation. In Section 4, a bootstrap simultaneous confidence region to estimate the population DEA efficiency in a fixed set of public services is determined. Section 5 contains an application of the proposed approach using real inputs and opinion indices estimated with a user sample (output data) of 15 libraries. Finally, the main conclusions are given.

2 Preliminaries

Consider a fixed set of M service-producing units, our DMUs, m resources of the services as known inputs $\mathbf{X}_j = (x_{1j},\ldots,x_{mj})$; $j = 1,\ldots,M$ and s customer opinion indices as unknown outputs. We distinguish between the population and sampling contexts. As for the population context, we consider $\mathbf{U}_j = \left(\mathbf{U}_{1j},\ldots,\mathbf{U}_{N_j j}\right); j = 1,\ldots,M$ the opinion of all of the N_j customers of the jth DMU (DMU$_j$ for short). Each $\mathbf{U}_{kj} = (U_{k1j},\ldots,U_{ksj})$ is the quantitative answer of the kth customer $(k = 1,\ldots,N_j)$ of the DMU$_j$ $(j = 1,\ldots,M)$ to the s opinion items. The output data $\mathbf{Y}_j$ in the DMU$_j$ is $g(\mathbf{U}_j)$ where $g : \Re^{N_j\times s} \longrightarrow \Re^s$. In this paper we consider g as the sample mean $\mathbf{Y}_j = \left(\frac{\sum_{k=1}^{N_j} U_{k1j}}{N_j},\ldots,\frac{\sum_{k=1}^{N_j} U_{ksj}}{N_j}\right)$. The LP model CCR or BCC with the output orientation of Table 1 (CCR-O or BCC-O), taking data $\{(\mathbf{X}_j,\mathbf{Y}_j)\}_{j=1,\ldots,M}$, determines the population DEA efficiency scores $\{\varphi_j\}_{j=1,\ldots,M}$. The output orientation is selected because the interest is to know which services, with their resources, can improve the opinion indices of their customers. Keeping in mind the impossibility of getting the opinion of all the population of N_j customers of the DMU$_j$, the outputs $\mathbf{Y}_j$ and the efficiencies φ_j , $j = 1,\ldots,M$, are unknown.

Table 1: *DEA models with constant (CCR) and variable (BCC) returns-to-scale; output orientation.*

	$max \quad \varphi+\varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$ s.t. $\sum_{j=1}^{M} \lambda_j y_{rj} - s_r^+ = \varphi y_{ro},\ \ r = 1,\ldots,s$ $\sum_{j=1}^{M} \lambda_j x_{ij} + s_i^- = x_{io},\ \ i = 1,\ldots,m$	
CCR-O	$\lambda_j \geq 0, s_i^- \geq 0, s_r^+ \geq 0;\ \ j = 1,\ \ldots,\ M;\ i = 1,\ \ldots,\ m;\ \ r = 1,\ldots,s$	(1)
BCC-O	$\sum_{j=1}^{M} \lambda_j = 1$	(2)
	where s_i^- and s_r^+ are slack variables and $\varepsilon > 0$ is a non-Archimedian element.	

In the sampling context, in the DMU$_j$, we take a random customer sample $\left(\mathbf{U}_{1j},\ldots,\mathbf{U}_{n_j j}\right) \subset \mathbf{U}_j$ of size n_j and we estimate the output data $\mathbf{Y}_j$ by $\widehat{\mathbf{Y}}_j = \left(\frac{\sum_{k=1}^{n_j} U_{k1j}}{n_j},\ldots,\frac{\sum_{k=1}^{n_j} U_{ksj}}{n_j}\right)$. We denote by $\widehat{\mathbf{y}}_j$ the observed value of this estimator. The LP model (1) or (2), taking $\left\{\left(\mathbf{X}_j,\widehat{\mathbf{Y}}_j\right)\right\}_{j=1,\ldots,M}$ as input-output data, determines the estimators $\{\widehat{\varphi}_j\}_{j=1,\ldots,M}$ of the population efficiency scores $\{\varphi_j\}_{j=1,\ldots,M}$, understanding that the model is maximized with the data $\{(\mathbf{X}_j,\widehat{\mathbf{y}}_j)\}_{j=1,\ldots,M}$ to obtain the estimates $\{\widehat{\omega}_j\}_{j=1,\ldots,M}$. Tapia et al. (2018)

prove that the estimator $\widehat{\varphi}_j$ is statistically consistent in the particular case of the CCR-O model with one known input and one estimated output.

Therefore, our statistical model (Ω, P) corresponds to independent, random samples in each DMU, that is, the sample space is $\Omega = \prod_{j=1}^{M} \Omega_j$, where $\Omega_j = \{$samples u_j of sample size n_j in the DMU$_j\}$, and the probability P depends on the sample design used.

The problem in this paper is to estimate the population efficiency scores vector $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_M)$ with a simultaneous confidence region. Formally, for any $\delta \in (0,1)$ and $\alpha \in (0,1)$, we then calculate the customer sample size n_j in the DMU$_j$, $j = 1, \ldots, M$, to guarantee

$$P\left(max_{j=1,\ldots,M} |\widehat{\varphi}_j - \varphi_j| \leq \delta\right) \geq 1 - \alpha, \tag{3}$$

that is,

$$\prod_{j=1}^{M} [\widehat{\varphi}_j \pm \delta] \tag{4}$$

defines a simultaneous region of confidence $1 - \alpha$ for $\boldsymbol{\varphi}$.

3 How many customers to interview?

We analytically solve the problem to determine the customer sample size proposed in (3) of Section 2, proving Theorem 2 under these assumptions:

C1 Fixed M DMUs
C2 One known input $\{X_j\}_{j=1,\ldots,M}$ and one unknown opinion index (output) $\{Y_j\}_{j=1,\ldots,M}$
C3 CCR-O model

Lemma 1 is the result used to prove Theorem 2, establishing the relation between sample size and accuracy, in order to simultaneously estimate the vector of DEA efficiencies.

Lemma 1 *Under assumptions C1, C2 and C3, for any* $0 < p < 1$, *we consider the sets of* Ω*:*

$$A_j = \left\{ u = (u_1 \times \cdots \times u_M) \in \Omega \ / \ \left| \widehat{Y}_j(u_j) - Y_j \right| \leq pY_j \right\}; \ j = 1, \ldots, M \tag{5}$$

$$B_j = \left\{ u \in \Omega \ / \ \left| \widehat{\varphi}_j(u) - \varphi_j \right| \leq \frac{2p}{1+p} \right\}; \ j = 1, \ldots, M \tag{6}$$

then

$$\bigcap_{j=1}^{M} A_j \subset \bigcap_{j=1}^{M} B_j.$$

Theorem 2 *Under assumptions C1, C2 and C3, for any* $0 < \delta < 1$ *and any* $0 < \alpha < 1$*, for every* $j = 1, \ldots, M$*, let* n_j *be the sample size in the* DMU_j *such that*

$$P\left(\left|\widehat{Y}_j - Y_j\right| \leq pY_j\right) \geq \sqrt[M]{1-\alpha} \tag{7}$$

with $p = \frac{\delta}{2-\delta} \in (0,\ 1)$*. Then*

$$\prod_{j=1}^{M} [\widehat{\varphi}_j \pm \delta] \tag{8}$$

defines a simultaneous region of confidence $1-\alpha$ *for the population efficiency scores vector.*

Remark 3 gives the explicit formulas to obtain the sample size under the usual simple random sample without replacement in a finite population.

Remark 3 If the customer sampling in each DMU is a simple random sample without replacement and the output is a population mean then

$$Y_j = \frac{\sum_{k=1}^{N_j} u_{kj}}{N_j};\ j = 1, \ldots, M$$

where u_{kj} is the answer (opinion) of the customer k in the DMU$_j$ and N_j its population size; then the sample size n_j that it verifies

$$P\left(\left|\widehat{Y}_j - Y_j\right| \leq pY_j\right) \geq \alpha_1$$

is (Särndal, Swensson and Wretman, 2003)

$$n_j \geq \frac{n_{oj}}{\left(\frac{n_{oj}}{N_j} + 1\right)} \tag{9}$$

with $n_{oj} = \frac{\tau^2_{1-\left(\frac{1-\alpha_1}{2}\right)}}{\left(pY_j\right)^2} {\sigma_j}^2$ and $\tau_{1-\left(\frac{1-\alpha_1}{2}\right)} = \phi^{-1}\left(1 - \left(\frac{1-\alpha_1}{2}\right)\right)$, where ${\sigma_j}^2$ is the population variance and ϕ the normal standard distribution function.

4 Bootstrap efficiency simultaneous confidence region

We carried out a simulation to check the confidence of the simultaneous region (8) in the case of two known inputs, two outputs estimated with a simple random sample without replacement of customers of size (9) and BCC-O model (2). This confidence is approximately one, so the region (8) is very conservative. We propose, as an alternative, an algorithm to construct a simultaneous confidence region for the population DEA efficiency scores vector, using Theorem 2 to determine the sample size, and bootstrap resampling of the samples of the customers' answers to the opinion items to estimate the population efficiency.

The algorithm is, considering M DMUs, each one using $m \geq 1$ known inputs $\mathbf{X}_j = (X_{1j},\ \ldots,\ X_{mj})$ and $s \geq 2$ unknown outputs $\mathbf{Y}_j = (Y_{1j}, \ldots, Y_{sj})$:

i. Taking $0 < \delta < 1$ and $0 < \alpha < 1$, we calculate the customer sample size n_j in the DMU_j as

$$n_j = max\{n_{1j}, \ldots, n_{sj}\} \tag{10}$$

where n_{rj} is the sample size to estimate the rth output in the DMU_j such that

$$P\left(\left|\widehat{Y}_{rj} - Y_{rj}\right| \leq pY_{rj}\right) \geq \sqrt[M]{1-\alpha};\ r = 1, \ldots, s \tag{11}$$

with $p = \frac{\delta}{2-\delta} \in (0,\ 1)$.

ii. In the DMU_j, $j = 1, \ldots, M$, we take the simple random sample without replacement of customers $\mathbf{u}_{kj} = (u_{k1j}, \ldots, u_{ksj})$ of size n_j, $k = 1, \ldots, n_j$, and we estimate the outputs $\widehat{\mathbf{y}}_j = (\widehat{y}_{1j}, \ldots, \widehat{y}_{sj})$ with the sample mean:

$$\widehat{y}_{rj} = \frac{\sum_{k=1}^{n_j} u_{krj}}{n_j};\ r = 1, \ldots, s,\ j = 1, \ldots, M. \tag{12}$$

iii. We take a bootstrap sample with replacement $\widehat{\mathbf{u}}^*_{kj}$ from $\mathbf{u}_{kj}$ of size n_j, $j = 1, \ldots, M$, with which we obtain the bootstrap version of the s output estimations $\widehat{\mathbf{y}}^*_j = \left(\widehat{y}^*_{1j}, \ldots, \widehat{y}^*_{sj}\right);\ j = 1, \ldots, M$.

With the data $\left\{\left(X_{1j}, \ldots,,\ X_{mj}, \widehat{y}^*_{1j}, \ldots, \widehat{y}^*_{sj}\right)\right\}_{j=1,\ldots,M}$, using a Table 1 model, we obtain the bootstrap version of the estimated efficiency scores, $\left\{\widehat{\omega}^*_j\right\}_{j=1,\ldots,M}$.

iv. We repeat step iii B times and the B bootstrap versions of the estimated efficiency scores for the DMU_j, $j = 1, \ldots, M$, should be $\left\{\widehat{\omega}_j^{*(b)}\right\}_{b=1,\ldots,\ B}$.

For any $0 < \alpha' < 1$, let $1-\alpha'$ be the level of coverage intention; then the observed bootstrap simultaneous confidence region of the population efficiency vector is

$$RC^* = \prod_{j=1}^{M} \left(\widehat{\omega}_j^{*\left(\frac{\alpha'}{2n}\right)}, \widehat{\omega}_j^{*\left(1-\frac{\alpha'}{2n}\right)}\right) \tag{13}$$

where $\widehat{\omega}_j^{*(\alpha)}$ is the α-percentile of the B values $\left\{\widehat{\omega}_j^{*(b)}\right\}_{b=1,\ldots,\ B}$.

This algorithm is ad hoc. No theory is given to suggest if it permits an estimate of the confidence region with asymptotically correct coverages, only the simulation study check the estimate quality.

4.1 Simulation study

We generate a simulated population model, as in Tapia et al. (2018), using the health centre data of Cooper, Seiford, and Tone (2006) (Table 2): in the j-th health centre, $j = 1,\dots,12$, a finite patient population $\mathbf{U}_j = \{\mathbf{u}_{1j},\dots,\mathbf{u}_{N_j j}\}$, of size N_j, $j = 1,\dots,12$, is generated where N_j are independent random variables with uniform distribution in $[10000, 50000]$, and

$$\mathbf{u}_{kj} = (u_{k1j}, u_{k2j}) \to N_2\left(\left(\begin{array}{c} z_{1j} \\ z_{2j} \end{array}\right), \left(\begin{array}{cc} z_{1j}^2/4 & 0 \\ 0 & z_{2j}^2/4 \end{array}\right)\right);\ k = 1,\dots,N_j;\ j = 1,\dots,12$$

and (z_{1j}, z_{2j}) are the original value outputs of the jth health centre, columns 4 and 5 of Table 2.

Table 2: *Number of doctors, nurses, outpatients and inpatients in 12 health centres.*

DMU	Doctor X1	Nurse X2	Outpatient Z1	Inpatient Z2	CCR-O efficiency score	BCC-O efficiency score
1	2.0	15.1	10	9	1	1
2	1.9	13.1	15	5	1	1
3	2.5	16	16	5.5	0.883	0.925
4	2.7	16.8	18	7.2	1	1
5	2.2	15.8	9.4	6.6	0.763	0.767
6	5.5	25.5	23	9	0.835	0.955
7	3.3	23.5	22	8.8	0.902	1
8	3.1	20.6	15.2	8	0.796	0.826
9	3	24.4	19	10	0.960	0.990
10	5	26.8	25	10	0.871	1
11	5.3	30.6	26	14.7	0.955	1
12	3.8	28.4	25	12	0.958	1

Source: Table 1.5 Cooper et al. (2006)

Table 3 shows the simulated population model: the patient population size for each health centre (column 2), the known inputs (columns 3 and 4) and the simulated values of the two outputs (columns 5 and 6) obtained with the population means

$$(Y_{1j}, Y_{2j}) = \left(\frac{\sum_{k=1}^{N_j} u_{k1j}}{N_j}, \frac{\sum_{k=1}^{N_j} u_{k2j}}{N_j}\right);\quad j = 1,\dots,M \tag{14}$$

where u_{krj} is the answer of the kth patient of the jth health centre to the rth opinion question. Columns 6 and 7 show the population DEA efficiency CCR-O and BCC-O, respectively.

Table 3: *Simulated population model.*

DMU	Population size N_j	Doctor X1	Nurse X2	Y1	Y2	Population efficiency score φ_j	
						CCR-O	**BCC-O**
1	43341	2.0	15.1	9.98	9.01	1	1
2	24438	1.9	13.1	14.98	5.01	1	1
3	45606	2.5	16	15.99	5.50	0.883	0.926
4	12578	2.7	16.8	17.96	7.18	1	1
5	19314	2.2	15.8	9.40	6.58	0.763	0.766
6	21782	5.5	25.5	22.96	8.97	0.835	0.957
7	19024	3.3	23.5	21.99	8.77	0.901	0.998
8	36271	3.1	20.6	15.17	8.01	0.797	0.826
9	30691	3	24.4	19.02	10.04	0.963	0.991
10	28385	5	26.8	24.89	10.01	0.871	1
11	28005	5.3	30.6	26.11	14.69	0.958	1
12	49077	3.8	28.4	25.09	11.98	0.960	1

Supposing a simple random sample without replacement of patients in each health centre, having fixed an efficiency estimation error $\delta = 0.1$ and a probability $1-\alpha = 0.95$, Table 4 shows the customer sample size n_j, $j = 1,\ldots,12$, calculated using (10) and (11).

Table 4: *Patient sample size obtained for each DMU, fixed* $\delta = 0.1$ *and* $\alpha = 0.05$.

DMU	n_j
1	985
2	674
3	779
4	1078
5	900
6	1415
7	795
8	814
9	765
10	617
11	1194
12	682

With this sample size: we first repeat the bootstrapping methodology steps ii. - iv. 1000 times, obtaining 1000 observed bootstrap efficiency simultaneous confidence regions, $\left\{RC^{*(k)}\right\}_{k=1,\ldots,1000}$, as in (13). The confidence of the bootstrap efficiency simultaneous confidence region is approximated through

$$C^* = \frac{1}{1000}\sum_{k=1}^{1000} I_{(\varphi_1,\ldots,\varphi_{12})\in RC^{*(k)}}. \tag{15}$$

Having fixed a percentile bootstrap confidence $1-\alpha' = 0.9$, Table 5 shows the approximated confidence (15) of the bootstrap confidence region to simultaneously estimate the population efficiency score vector. This result confirms that, by determining the sample size in each DMU using (10), we get the fixed accuracy with the percentile bootstrap efficiency simultaneous confidence region.

Table 5: *Confidence approximation of the bootstrap efficiency simultaneous confidence region taking* $1-\alpha' = 0.9$ *and fixed* $\delta = 0.1$ *and* $\alpha = 0.05$. *CCR and BCC model with output orientation.*

	CCR-O	**BCC-O**
C^*	0.975	0.983

In order to justify the basic percentile method to obtain bootstrap confidence intervals, we analyse the bias of the bootstrap process. Using the 1000 data $\left\{\widehat{\omega}_j - \widehat{\omega}_j^{*(b)}\right\}_{b=1,\ldots,1000}$ for each DMU$_j$, we represent graphically the nonparametric density estimates with kernel $N(0,1)$ and smoothing parameter selected with rule-of-thumb (Silverman (1986)). The smooth density estimates obtained are approximately symmetrical with respect to 0, therefore the bias is negligible.

5 A Case Study

This section provides an empirical DEA efficiency analysis of libraries using real data input. Table 7 corresponds to the database from 15 libraries used in Tapia et al. (2018). The data input are the number of book loans, (X_1), the library's seating capacity and the number of computers for users, (X_2), and the data are scaled 0-10. Column 2 shows the distribution of the user population in each of the 15 libraries. In each library, we use only the output given by the users' mean, monthly time use of the library, in hours, measured for each user, on a scale of 0-10. In order to determine the user sample size n_j in the jth library, $j = 1,\ldots,15$, using Remark 3 in each library, we take a previous user simple random sample without replacement of 0.1% of the user population size and we estimate the output and the population variance, shown in columns 2, 3 and 4 of Table 7, respectively. Having fixed the efficiency estimation error $\delta = 0.1$ and a probability $1-\alpha = 0.9$, we determine the sample size n_j; $j = 1,\ldots,15$ with (10) and (11). We then take the user sample in each library and estimate the mean monthly time of permanence in the library $\{\widehat{y}_j\}_{j=1,\ldots,15}$, column 6 of Table 7.

Table 6: *Results of the previous simple random sample without replacement: User sample size, Estimation of monthly time use mean and of population variance.*

DMU	Sample size $n_j^{(0)}$	Estimation of the monthly time use library mean	Estimation of the population variance
1	89	6.61	14.34
2	79	5.78	13.76
3	64	3.87	12.93
4	59	5.06	13.26
5	57	6.12	14.48
6	51	2.88	8.67
7	42	5.41	17.08
8	37	4.31	13.79
9	35	6.06	12.79
10	32	4.10	10.24
11	24	4.57	15.16
12	21	5.27	16.97
13	19	3.09	13.10
14	17	4.37	14.55
15	13	7.81	10.63

Table 7: *Database for 15 libraries: User population size, book loans and user posts (inputs), user sample size to estimate the mean monthly library time use (output) with $\delta = 0.1$ and $\alpha = 0.1$.*

DMU	User population size N_j	Book loans X1	User posts X2	User sample size n_j	Estimated monthly time use library mean $\widehat{y}_j$
1	89300	7.04	7.82	855	6.03
2	78500	7.81	6.87	1065	5.33
3	64000	5.41	5.60	2190	3.26
4	59100	2.66	5.18	1326	5.33
5	56500	3.96	4.95	998	6.13
6	50700	3.28	4.44	2601	3.37
7	41600	4.36	3.64	1476	5.79
8	37000	6.29	3.24	1850	3.44
9	34600	5.82	3.03	891	5.14
10	32000	7.69	2.80	1523	3.48
11	23600	2.61	2.07	1761	5.28
12	21200	3.61	1.86	1492	5.87
13	18900	4.73	1.65	3028	5.33
14	17200	2.12	1.51	1791	3.40
15	12900	2.16	1.13	442	5.77

The estimated efficiency scores, $\{\widehat{\omega}_j\}_{j=1,\ldots,15}$ and the intervals whose product determines the bootstrap efficiency simultaneous confidence region (columns 2 and 3 in Table 8, respectively) are obtained using the data $\{(X_{1j}, X_{2j}, \widehat{y}_j)\}_{j=1,\ldots,15}$ and the BCC model with orientation output. The libraries $\{1, 5, 7, 14, 15\}$ can be considered efficient because the corresponding intervals of the bootstrap efficiency simultaneous confidence region contains the value 1.

Table 8: *Estimated efficiency scores and intervals whose product determines the bootstrap efficiency simultaneous confidence region, taking* $1-\alpha' = 0.9$. *BCC model with output orientation.*

DMU	**Estimated efficiency score**	**Intervals**
1	0.983	[0.904, 1]
2	0.870	[0.797, 0.942]
3	0.531	[0.493, 0.567]
4	0.908	[0.841, 0.982]
5	1	[0.959, 1]
6	0.562	[0.524, 0.596]
7	0.962	[0.902, 1]
8	0.574	[0.529, 0.611]
9	0.862	[0.791, 0.925]
10	0.585	[0.539, 0.625]
11	0.900	[0.826, 0.981]
12	1	[0.913, 1]
13	0.914	[0.839, 0.959]
14	1	[1, 1]
15	1	[1, 1]

6 Conclusions

Over the last decade, the use of opinion-satisfaction surveys on customers of public services has been an essential tool in measuring the quality of the service given. Without a doubt, a public service will be more efficient when, with its resources, it is able to have the highest opinion-satisfaction of its customers. The questionnaire is a common tool to find out customer opinion-satisfaction with the service received. The mean of the opinion-satisfaction answers of the sample of customers are indices, indicators of the service quality, that can be considered as output data. If we add the deterministic information of the resources of the public service-producing unit as input data, we will have the necessary input and output data to calculate the DEA efficiency in the set of services.

We focus on this DEA efficiency problem as a statistical one, considering an unknown population efficiency vector that would be obtained if we had the opinion of the entire service customer population (census). We estimate this parametric vector with a confidence region using the outputs estimated with the opinion of the user sample, the known inputs and the classical DEA models (LP models subject to deterministic constraints). To our knowledge, this statistical view of the DEA is totally novel and the use of a simultaneous confidence region is a statistical concept that has not been used in DEA efficiency analysis in the form that we propose. From a practical point of view, the application in library datasets shows the usefulness of the bootstrap region confidence efficiency methodology. This region, based on the product of confidence intervals in each library, detects whether the library is "efficient in any case" because the lower bound is equal to unity, or if the library is "efficient" because the upper bound is equal to unity and "inefficient" because the upper bound is less than unity. More public service examples where it may be interesting to apply the results of this paper are: leisure centres (where there are attractions for which it is necessary to maximize the demand, which can be evaluated by carrying out customer surveys), marketing or electoral polls (where the effect of the advertising or electoral campaign is evaluated with a survey), hospitals, airports, banks, universities, supermarkets, government services or schools (where existing resources can explain customer opinion-satisfaction).

In this paper, we obtain two types of confidence region for the population efficiency scores vector. Theorem 2 allows us to define the first simultaneous region, finding a relation between the accuracy of the simultaneous confidence region and the customer sample size needed to guarantee an output estimation error in each public service-producing unit. By simulation, we check that this confidence region is very conservative. As an alternative, we propose to determine the sample size of customers necessary using Theorem 2 and Remark 3 and to obtain an efficiency confidence region based on the basic percentile bootstrap method. The simulation shows that we are able to reach the confidence of the bootstrap efficiency simultaneous confidence region close to the desired level.

Other possible extensions currently under investigation by the authors also include considering stochastic inputs estimated with a provider sample or using other bootstrap methods, such as the adjusted percentile method or the ABC method, and comparing them with the method used in this paper.

7 Appendix section

7.1 Proof of Lemma 1

Let $u = (u_1 \times \cdots \times u_M) \in \bigcap_{j=1}^{M} A_j$.

Let us consider the DMU_r. Then, $\varphi_r = 1$ or $\varphi_r < 1$:

- If $\varphi_r = 1$, it is because $\frac{Y_r}{X_r} = max_j \frac{Y_j}{X_j}$.

  The most unfavourable situation, where B_r is verified, is that the output of the DMU_r is as small as possible and the rest of the DMUs are as big as possible, that is to say
  $$Y_j^* = (1-p)Y_j I_{(j=r)} + (1+p)Y_j I_{(j \neq r)}.$$

Let $\varphi_j^*,\ j=1,\ldots,M$ be the efficiencies of the DMUs obtained with the data $\left\{(X_j,Y_j^*)\right\}_{j=1,\ldots,M}$, therefore

$$\varphi_r^* = \begin{cases} 1 & \text{if } \frac{(1-p)Y_r}{X_r} \geq max_{j\neq r}\ \frac{(1+p)Y_j}{X_j} & (a) \\ \frac{(1-p)Y_r}{(1+p)\frac{Y_k}{X_k}X_r} & \text{if } (1+p)\frac{Y_k}{X_k} = max_{j\neq r}\ \frac{(1+p)Y_j}{X_j} > \frac{(1-p)Y_r}{X_r} & (b) \end{cases}$$

and

$$|\varphi_r - \widehat{\varphi}_r(u)| = 1-\widehat{\varphi}_r(u) \leq 1-\varphi_r^*$$

where the first equality is obtained because $\varphi_r = 1$ and the second inequality is verified because, as the efficiency of a DMU decreases when the output of this unit decreases, while the outputs of the rest of the DMUs also increase, $\widehat{\varphi}_r(u) \geq \varphi_r^*$, then

$$|\varphi_r - \widehat{\varphi}_r(u)| \leq \begin{cases} 0 & \text{if } (a) \\ 1-\frac{(1-p)}{(1+p)} = \frac{2p}{1+p} & \text{if } (b) \end{cases}$$

and it is verified that $u \in B_r\ \forall r\ /\ \varphi_r = 1$.

- If $\varphi_r < 1$

  Let $k \neq r\ \ /\ \ \frac{Y_k}{X_k} = max_j \frac{Y_j}{X_j}$ and $\varphi_r = \frac{\frac{Y_r}{X_r}}{\frac{Y_k}{X_k}} < 1$.

  There are two more favourable situations for B_r to be verified

  Case (I) $Y_j^* = (1-p)Y_jI_{(j=r)} + (1+p)Y_jI_{(j\neq r)}$.

  Case (II) $Y_j^{**} = (1+p)Y_jI_{(j=r)} + (1-p)Y_jI_{(j\neq r)}$.

  Case (I): Let $\varphi_j^*,\ j=1,\ldots,M,$ be the DMU efficiencies obtained with the data $\left\{(X_j,Y_j^*)\right\}_{j=1,\ldots,M}$, then

$$\varphi_r^* = \frac{(1-p)\frac{Y_r}{X_r}}{(1+p)\frac{Y_k}{X_k}} = \frac{(1-p)}{(1+p)}\varphi_r < \varphi_r.$$

  As the efficiency of a DMU decreases when the output of this unit decreases, while the outputs of the rest of the DMUs also increase,

$$|\varphi_r - \widehat{\varphi}_r(u)| \leq |\varphi_r - \varphi_r^*|$$

  and therefore

$$|\varphi_r - \widehat{\varphi}_r(u)| \leq \varphi_r - \varphi_r^* = \varphi_r\left(\frac{2p}{1+p}\right) < \frac{2p}{1+p}.$$

Case (II): Let $\varphi_j^{**},\ j = 1,\dots,M,$ be the DMU efficiencies obtained with the data $\left\{(X_j,Y_j^{**})\right\}_{j=1,\dots,M}$, then

– If $max\left(\frac{(1+p)Y_r}{X_r}, max_{j\neq r}\ \frac{(1-p)Y_j}{X_j}\right) = \frac{(1+p)Y_r}{X_r}$ then $\varphi_r^{**} = 1$ and $\varphi_r \geq \frac{(1-p)}{(1+p)}$ and therefore

$$|\varphi_r - \widehat{\varphi}_r(u)| \leq |\varphi_r - \varphi_r^{**}| = 1 - \varphi_r \leq \frac{2p}{1+p}.$$

– If $max\left(\frac{(1+p)Y_r}{X_r}, max_{j\neq r}\ \frac{(1-p)Y_j}{X_j}\right) = \frac{(1-p)Y_k}{X_k}$ for $k \neq r$ then $\varphi_r < \frac{(1-p)}{(1+p)}$ and $\varphi_r^{**} = \frac{(1+p)\frac{Y_r}{Y_r}}{(1-p)\frac{Y_k}{X_k}} = \frac{(1+p)}{(1-p)}\varphi_r > \varphi_r$ and therefore

$$|\varphi_r - \widehat{\varphi}_r(u)| \leq \varphi_r^{**} - \varphi_r \leq \frac{2p}{1-p}\varphi_r < \frac{2p}{1+p}.$$

In consequence $u \in B_r\ \forall r\ /\ \varphi_r < 1$.

7.2 Proof of Theorem 2

It is enough to prove that

$$P\left(max_{j=1,\dots,M}\left|\widehat{\varphi}_j - \varphi_j\right| \leq \delta\right) \geq 1-\alpha. \tag{16}$$

Using the notation of (5) and (6), by Lemma 1, we know that

$$\bigcap_{j=1}^{M} A_j \subset \bigcap_{j=1}^{n} B_j$$

and, as the events $\{A_j\}_{j=1,\dots,M}$ are independent, then

$$P\left(\bigcap_{j=1}^{M} B_j\right) \geq P\left(\bigcap_{j=1}^{M} A_j\right) = \prod_{j=1}^{M} P(A_j) \geq \left(\sqrt[M]{1-\alpha}\right)^M = 1-\alpha.$$

Acknowledgements

The authors thank the Editor and two anonymous referees for their constructive comments and suggestions that have helped to improve the quality of the paper.

References

Alexander, W.R.J., Haug, A.A. and Jaforullah, M. (2010). A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. *Journal of Productivity Analysis*, 34, 99–110.

Assaf, A. (2010). Bootstrapped scale efficiency measures of UK airports. *Journal of Air Transport Management*, 16, 42–44.

Assaf, A.G. and Agbola, F.W. (2011). Modelling the performance of Australian hotels: A DEA double bootstrap approach. *Tourism Economics*, 17, 73–89.

Banker, R.D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, 39, 1265–1273.

Barra, C. and Zotti, R. (2016). Measuring efficiency in higher education: An empirical study using a bootstrapped data envelopment analysis. *International Advances in Economic Research*, 22, 11–33.

Bayraktar, E., Tatoglu, E., Turkyilmaz, A., Delen, D. and Zaim, S. (2012). Measuring the efficiency of customer satisfaction and loyalty for mobile phone brands with DEA. *Expert Systems with Applications*, 39, 99–106.

Benito, B., Solana, J. and Moreno, M.R. (2014). Explaining efficiency in municipal services providers. *Journal of Productivity Analysis*, 42, 225–239.

Casu, B. and Molyneux, P. (2003). A comparative study of efficiency in European banking. *Applied Economics*, 35, 1865–1876.

Ceyhan, M.E. and Benneyan, J.C. (2014). Handling estimated proportions in public sector data envelopment analysis. *Annals of Operations Research*, 221, 107–132.

Charles, V. and Kumar, M. (2014). Satisficing data envelopment analysis: An application to servqual efficiency. *Measurement*, 51, 71–80.

Charnes, A. and Cooper, W.W. (1959). Chance-constrained programming. *Management Science*, 6, 73–79.

Charnes, A. and Cooper, W.W. (1963). Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*, 11, 18–39.

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

Chowdhury, H. and Zelenyuk, V. (2016). Performance of hospital services in Ontario: DEA with truncated regression approach. *Omega*, 63, 111–122.

Cooper, W., Huang, Z. and Li, S.X. (1996). Satisficing DEA models under chance constraints. *Annals of Operations Research*, 66, 279–295.

Cooper, W.W., Deng, H., Huang, Z., Li, S.X. et al. (2002). Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. *Journal of the Operational Research Society*, 53, 1347–1356.

Cooper, W.W., Seiford, L.M. and Tone, K. (2006). Introduction to data envelopment analysis and its uses: with DEA-solver software and references. *Springer Science & Business Media. New York, NY 10013, USA.*

Cooper, W.W., Seiford, L.M. and Zhu, J. (2011). Handbook on data envelopment analysis, volume 164. *Springer Science & Business Media. New York, NY 10013, USA.*

Essid, H., Ouellette, P. and Vigeant, S. (2014). Productivity, efficiency and technical change of Tunisian schools: a bootstrapped Malmquist approach with quasi-fixed inputs. *Omega*, 42, 88–97.

Førsund, F.R. (2017). Measuring effectiveness of production in the public sector. *Omega*, https://doi.org/10.1016/j.omega.2016.12.007.

Gil Ropero, A., Turias Dominguez, I. and Cerbán Jiménez, M.M. (2019). Bootstrapped operating efficiency in container ports: a case study in Spain and Portugal. *Industrial Management & Data Systems*, 119, 924–948.

Huang, Z. and Li, S.X. (2001). Stochastic DEA models with different types of input-output disturbances. *Journal of Productivity Analysis*, 15, 95–113.

Khodabakhshi, M. (2010). An output oriented super-efficiency measure in stochastic data envelopment analysis: Considering iranian electricity distribution companies. *Computers & Industrial Engineering*, 58, 663–671.

Khodabakhshi, M. and Asgharian, M. (2009). An input relaxation measure of efficiency in stochastic data envelopment analysis. *Applied Mathematical Modelling*, 33, 2010–2023.

Kneip, A., Park, B.U. and 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. and Wilson, P.W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24, 1663–1697.

Kneip, A., Simar, L. and Wilson, P.W. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics*, 38, 483–515.

Land, K.C., Lovell, C. and Thore, S. (1993). Chance-constrained data envelopment analysis. *Managerial and Decision Economics*, 14, 541–554.

Lee, H. and Kim, C. (2014). Benchmarking of service quality with data envelopment analysis. *Expert Systems with Applications*, 41, 3761–3768.

Liu, S.T. and Chuang, M. (2009). Fuzzy efficiency measures in fuzzy DEA/AR with application to university libraries. *Expert Systems with Applications*, 36, 1105–1113.

Mayston, D.J. (2015). Analysing the effectiveness of public service producers with endogenous resourcing. *Journal of Productivity Analysis*, 44, 115-126.

Mayston, D.J. (2017). Data envelopment analysis, endogeneity and the quality frontier for public services. *Annals of Operations Research*, 250, 185–203.

Olesen, O.B. and Petersen, N. (1995). Chance constrained efficiency evaluation. *Management Science*, 41, 442–457.

Parasuraman, A., Berry, L.L. and Zeithaml, V.A. (1993). More on improving service quality measurement. Journal of Retailing, 69, 140–147.

Salo, A. and Punkka, A. (2011). Ranking intervals and dominance relations for ratio-based efficiency analysis. *Management Science*, 57, 200–214.

Santín, D. and Sicilia, G. (2017). Dealing with endogeneity in data envelopment analysis applications. *Expert Systems with Applications*, 68, 173–184.

Särndal, C.-E., Swensson, B. and Wretman, J. (2003). Model assisted survey sampling. *Springer Science & Business Media. New York, NY 10010, USA.*

Silverman, B.W. (1986). Density estimation for statistics and data analysis. *Chapman and Hall. London; New York.*

Simar, L. and Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44, 49–61.

Simar, L. and Wilson, P.W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27, 779–802.

Simar, L. and Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136, 31–64.

Simar, L. and Wilson, P.W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36, 33–53.

Simar, L. and Wilson, P.W. (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives. *Foundations and Trends in Econometrics*, 5, 183–337.

Simar, L. and Wilson, P.W. (2015). Statistical approaches for non-parametric frontier models: a guided tour. *International Statistical Review*, 83, 77–110.

Simar, L. and Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews*, 25, 497–522.

Smith, P. and Mayston, D. (1987). Measuring efficiency in the public sector. *Omega*, 15, 181–189.

Tapia, J.A., Salvador, B. and RodrÃguez, J.M. (2018). Data envelopment analysis in satisfaction survey research: sample size problem. *Journal of the Operational Research Society*, 69, 1096-1104.

Tavana, M., Shiraz, R.K. and Hatami-Marbini, A. (2014). A new chance-constrained DEA model with birandom input and output data. *Journal of the Operational Research Society*, 65, 1824–1839.

Tsekouras, K., Papathanassopoulos, F., Kounetas, K. and Pappous, G. (2010). Does the adoption of new technology boost productive efficiency in the public sector? the case of ICUs system. *International Journal of Production Economics*, 128, 427–433.

Witte, K.D. and Geys, B. (2013). Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. *European Journal of Operational Research*, 224, 592–602.

Wu, D. and Olson, D.L. (2008). A comparison of stochastic dominance and stochastic DEA for vendor evaluationand. *International Journal of Production Research*, 46, 2313–2327.

Wu, D.D. (2010). A systematic stochastic efficiency analysis model and application to international supplier performance evaluation. *Expert Systems with Applications*, 37, 6257–6264.

Wu, D.D. and Lee, C.-G. (2010). Stochastic DEA with ordinal data applied to a multi-attribute pricing problem. *European Journal of Operational Research*, 207, 1679–1688.