Data envelopment analysis efficiency in the public sector using provider and customer opinion: An application to the Spanish health system

Jesús A. Tapia[1] · Bonifacio Salvador[1]

Abstract

Measuring the relative efficiency of a finite fixed set of service-producing units (hospitals, state services, libraries, banks,...) is an important purpose of Data Envelopment Analysis (DEA). We illustrate an innovative way to measure this efficiency using stochastic indexes of the quality from these services. The indexes obtained from the opinion-satisfaction of the customers are estimators, from the statistical view point, of the quality of the service received (outputs); while, the quality of the offered service is estimated with opinion-satisfaction indexes of service providers (inputs). The estimation of these indicators is only possible by asking a customer and provider sample, in each service, through surveys. The technical efficiency score, obtained using the classic DEA models and estimated quality indicators, is an estimator of the unknown population efficiency that would be obtained if in each one of the services, interviews from all their customers and all their providers were available. With the object of achieving the best precision in the estimate, we propose results to determine the sample size of customers and providers needed so that with their answers can achieve a fixed accuracy in the estimation of the population efficiency of these service-producing units through the use of a novel one bootstrap confidence interval. Using this bootstrap methodology and quality opinion indexes obtained from two surveys, one of doctors and another of patients, we analyze the efficiency in the health care system of Spain.

Keywords Data envelopment analysis · Sampling survey research · Public sector · Sample size · Bootstrap · Efficiency confidence interval

1 Introduction

Data Envelopment Analysis (DEA) has become a widely used technique to compare the efficiency of service-producing units because it easily handles the multiple outputs characteristic of public sector production, is non-parametric and does not require input price data [34]. The popular application is with deterministic information and classic DEA models [3, 10], for instance, in health care

[1, 37] or [25]), universities and research institutes [26], government [17], public libraries [22, 23], schools [35], public transport services [15] or banks [28].

In recent years, some approaches consider the opinion of the customer to be crucial for measuring DEA efficiency in public services ([5, 16]). Consumer satisfaction-opinion surveys are a common tool for building opinion indexes, which measure the quality of the service, and to be used as output variables in DEA ([20, 27, 31, 32, 35, 40, 41] and [42]).

Besides customer opinion-satisfaction (output), provider opinion-satisfaction (input) is also a fundamental protagonist in public service-producing units (SPUs), our decision-making units. The provider's positive opinion-satisfaction concerning the service offered may result in a better customer opinion concerning the service received. For instance, a city's public bus drivers with a good opinion of their salary, timetable, partners, driven vehicles, etc., may influence a better satisfaction-opinion of the travellers using the

✉ Jesús A. Tapia
    jesus.tapia@uva.es

    Bonifacio Salvador
    bosal@eio.uva.es

[1] Department of Statistic and Operative Research, University
    of Valladolid, Campus Miguel, Delibes, Paseo Belén,
    nº7,47011 Valladolid, Spain

service. One way to measure the service quality offered is with satisfaction-opinion indexes obtained through survey samples carried out with the providers. In Tapia et al. [40–42] the protagonism of the opinion of the providers is not considered, that is, the inputs are deterministic and only the outputs are estimated using a customer sample. Introducing the opinion-satisfaction of the providers as estimated indexes increases the field of application of this work, where opinion-satisfaction indexes, estimated from a sample of providers (inputs) and a sample of customers (outputs), are used as information to measure the DEA efficiency of the public services. To do so, it is necessary to determine the sampling design, the sample size and the estimators of the satisfaction-opinion indexes of both the providers and the customers in each SPU.

In practice, this stochastic input and output information is available in many services, as it is increasingly common to conduct opinion-satisfaction surveys for both providers and customers. For instance, the sample data used in the Spanish health system application of Section 5.

The efficiency obtained with DEA models, using the opinion indexes estimated with the survey answer of samples of providers and consumers as data, will be an estimation of the population DEA efficiency. This population efficiency is an unknown non-evaluable parameter, since it would be necessary to use the indexes obtained with the opinion of all the customers and all the providers of all the services as data and this is a census of the entire population [40]. This statistical analysis of the DEA efficiency gives rise to the problem of determining the sample size of customers and providers needed to guarantee an a priori fixed accuracy in the estimation, with confidence interval, of the DEA efficiency of each public service, which is the object of our investigation. Liu et al. [30] considered the statistical analysis and sampling process as an important direction for handling the DEA. Ceyhan and Benneyan [7] investigated the impact of the sample size on the measures of efficiency when the DEA problem was carried out on values that include such estimated proportions as defect, satisfaction, mortality, or adverse event rates estimated from samples. Nevertheless, they do not propose a solution to the necessary sample size in order to control the error in the measures of efficiency. The problem of calculating the sample size of customers necessary when the objective is to estimate, with a fixed precision in the estimate error, the population efficiency in a finite set of public services using stochastic data output (customer opinion-satisfaction indexes) and known (non-stochastic) data input was resolved in Tapia et al. [40, 42].

In this paper, we propose a solution to determine the customer and provider sample sizes needed to estimate the DEA efficiency in public services with bootstrap confidence intervals and a fixed accuracy. These intervals capture the random variations introduced in the DEA analysis by using outputs and inputs estimated with a sample. So far, Simar and Wilson's [38, 39] methodology has been the most common for measuring efficiency with bootstrap confidence intervals in such public services as health care ([11]), universities and research institutes [4], government [6], public libraries [29], schools [19], tourism [2], banks [28] or public transport services [21]. The problem posed by Simar and Wilson is different from ours because the stochastic character of the inputs and outputs is different. In Simar and Wilson, the stochastic character of the input and output information comes from considering the set of available SPUs, S_n, as a sample from an infinite population and the sample observations in S_n are realizations of identically, independently distributed (iid) random variables with a probability density function with support over $P = \{(x, y)|x \text{ can produce } y\}$, [14]. In this work, the set of available SPUs are the only units whose efficiency we wish to evaluate. Therefore, we do not consider them a sample as in the classical DEA models, [9]. The stochastic character of our output and input information comes from the fact that these data are opinion-satisfaction indexes that it is necessary to estimate by taking independent random samples in each SPU, and the probability, in the statistical model, depends on the sample design used in each SPU.

In Shwartz [37], a random sample of patients arriving for health care services were bootstrap resampled to obtain data input and interval estimates of the DEA efficiency. These interval estimates are conservative (large) and the problem of the patient sample size necessary to obtain the desired accuracy in the estimation of the DEA efficiency is not resolved.

In other approaches, where samples in public services are used to estimate the data (Charles2014), the efficiency is estimated with stochastic DEA models. These models, which use LP problems subject to constraints defined in terms of probability, are also called chance-constrained problems. The deterministic characterization of "efficient" is then changed by the probabilistic characterization "probably efficient". A vast number of papers show a wide range of uses of chance-constrained programming (CCP), including [8, 9, 12, 33] or [43]. One of the main advantages of the technique presented in this study is its simplicity, as we use the original DEA models with constant (CCR) and variable returns-to-scale (BCC), that is, linear programming (LP) problems subject to deterministic constraints [3, 10].

In this paper, we therefore examine the implications of input and output data estimated with a provider and a customer sample, respectively, for the performance analysis of public services using DEA analysis techniques. In Section 2, we examine the nature of the problem.

Section 3 presents a theoretical method to determine the customer and provider sample size needed to estimate the population DEA efficiency with bootstrap confidence interval, which is examined in Section 4. Section 5 includes the empirical application in the Spanish health system we have undertaken. Section 6 contains our conclusions. All the software is our own elaboration using MATLAB and it is included in Appendix II.

2 Nature of the problem

We consider a finite fixed set of L SPUs, service-producing units, one provider interview to estimate m inputs and one customer interview to estimate s outputs. For example, in each hospital of a homogeneous set of L, it is possible to estimate the general satisfaction of the personnel, or their annual time of formation, or other personal opinion indexes (input data) by interviewing a sample of the personnel. It is also possible to estimate the satisfaction of the patients attention they have received, the human and material resources of the hospital, etc., by interviewing to a patient sample (output data). The main problem approached in this paper is to estimate the unknown parameter population DEA efficiency. The population efficiency, or census efficiency, is unknown because, in order to know it, it would be necessary to interview all the providers and all the customers (census) of each one of the L public services and, with this population data, to obtain the opinion indexes and use them as input/output data in the classic LP model CCR or BCC of Table 1; in real applications, these censuses are completely non-viable.

The error made in estimating the input/output data with samples is transferred to the DEA efficiency estimation. In this study, we propose a methodology that guarantees a reasonable quality of the DEA efficiency estimation.

Formally, in the jth service-producer unit (SPU$_j$ for short), we consider the finite provider population $\mathbf{U_j} = \left\{\mathbf{U_{1j}}, \ldots, \mathbf{U_{N_{x_j}j}}\right\}$ of size N_{x_j} and the customer population $\mathbf{W_j} = \left\{\mathbf{W_{1j}}, \ldots, \mathbf{W_{N_{y_j}j}}\right\}$ of size N_{y_j}. Each provider is a quantitative vector $\mathbf{U_{kj}} = \left(U_{k1j}, \ldots, U_{kmj}\right)$, where U_{kij} is the answer of the kth provider of the jth SPU to the ith opinion provider item, i.e., each customer $\mathbf{W_{hj}} = \left(W_{h1j}, \ldots, W_{hsj}\right)$ is a quantitative vector where W_{hrj} is the answer of the hth customer of the jth SPU to the rth customer opinion item. The population input and output data are opinion indexes obtained as a function of the provider and customer population answers, in general:

$$\mathbf{X_j} = \left(f(U_{11j}, \cdots, U_{N_{x_j}1j}), \ldots, f(U_{1mj}, \cdots, U_{N_{x_j}mj})\right);\ j=1,\ldots,L \quad (5)$$

$$\mathbf{Y_j} = \left(g(W_{11j}, \cdots, W_{N_{y_j}1j}), \ldots, g(W_{1sj}, \cdots, W_{N_{y_j}sj})\right);\ j=1,\ldots,L \quad (6)$$

These functions f and g can be of any type, with or without weights, whenever they admit an interpretation of m population opinion-satisfaction indexes as input and s population opinion-satisfaction indexes as output.

Using the data $\left\{(\mathbf{X_j}, \mathbf{Y_j})\right\}_{j=1,\ldots,L}$ in the Table 1 models, we obtain the population DEA efficiency scores $\left\{\varphi_j\right\}_{j=1,\ldots,L}$. If it were possible to carry out a census and to know these population indexes, the information would be fixed or deterministic and the character of the problem would not be stochastic.

A lack of knowledge concerning the population information input and output makes the taking of samples to estimate $\left\{\varphi_j\right\}_{j=1,\ldots,L}$ necessary. In the SPU$_j$, let $\left\{\mathbf{U_{1j}}, \ldots, \mathbf{U_{n_{x_j}j}}\right\} \subset \mathbf{U_j}$ and $\left\{\mathbf{W_{1j}}, \ldots, \mathbf{W_{n_{y_j}j}}\right\} \subset \mathbf{W_j}$ be random samples of size n_{x_j} and n_{y_j} of the provider and customer populations, respectively. To obtain the estimators of the input indexes Eq. 5 and the output indexes Eq. 6, the same functions f and g are used with the random samples:

$$\widehat{\mathbf{X}}_\mathbf{j} = \left(f(U_{11j}, \cdots, U_{n_{x_j}1j}), \ldots, f(U_{1mj}, \cdots, U_{n_{x_j}mj})\right);\ j=1,\ldots,L \quad (7)$$

$$\widehat{\mathbf{Y}}_\mathbf{j} = \left(g(W_{11j}, \cdots, W_{n_{y_j}1j}), \ldots, g(W_{1sj}, \cdots, W_{n_{y_j}sj})\right);\ j=1,\ldots,L \quad (8)$$

Having observed the provider and customer sample answers $\left\{\mathbf{u_{kj}} = \left(u_{k1j}, \ldots, u_{kmj}\right)\right\}_{k=1,\ldots,n_{x_j}}$ and $\left\{\mathbf{w_{hj}} = \left(w_{h1j}, \ldots, w_{hsj}\right)\right\}_{h=1,\ldots,n_{y_j}}$, respectively, Eqs. 9 and 10 are the estimates of the input and output indexes, respectively; that is, the values that the estimators Eqs. 7 and 8 take with the sample answers of the customers and providers:

$$\widehat{\mathbf{x}}_j = \left(f(u_{11j}, \cdots, u_{n_{x_j}1j}), \ldots, f(u_{1mj}, \cdots, u_{n_{x_j}mj})\right)$$
$$= \left(\widehat{x}_{1j}, \cdots, \widehat{x}_{mj}\right);\ j=1,\cdots,L \quad (9)$$

$$\widehat{\mathbf{y}}_j = \left(g(w_{11j}, \cdots, w_{n_{y_j}1j}), \ldots, g(w_{1sj}, \cdots, w_{n_{y_j}sj})\right)$$
$$= \left(\widehat{y}_{1j}, \cdots, \widehat{y}_{sj}\right);\ j=1,\cdots,L \quad (10)$$

Using $\left(\widehat{\mathbf{X}}_\mathbf{j}, \widehat{\mathbf{Y}}_\mathbf{j}\right)_{j=1,\ldots,L}$ in the Table 1 models, we obtain the estimators $\left(\widehat{\varphi}_j\right)_{j=1,\ldots,L}$ of the population efficiency scores $\left(\varphi_j\right)_{j=1,\ldots,L}$, on the understanding that the DEA model is maximized, or minimized, with the data $\left(\widehat{\mathbf{x}}_j, \widehat{\mathbf{y}}_j\right)_{j=1,\ldots,L}$ in order to obtain the estimation $\widehat{\omega}_j$ of the estimator $\widehat{\varphi}_j$.

Therefore, our statistical model (Ω, P) corresponds to independent, random samples in each SPU, that is, the sample space is $\Omega = \prod_{j=1}^{L} \Omega_j$, where $\Omega_j = \{$samples $\mathbf{u_{kj}}$ of size n_{x_j} and samples $\mathbf{w_{hj}}$ of size n_{y_j} in SPU$_j\}$, and the probability P depends on the sample design used.

The first objective of this paper, having fixed $\delta \in (0, 1)$ and $\alpha \in (0, 1)$, is to obtain the provider and customer sampling size, n_{x_j} and n_{y_j}, respectively, to estimate $\widehat{X}_j$ and

Table 1 DEA models with variable (BCC) and constant returns-to-scale (CCR)

Model	Orientation		
	Input		Output
	$min \ \theta - \epsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)$ $s.t.$ $\sum_{j=1}^{L} \lambda_j x_{ij} + s_j^- = \theta x_{io}; \ i = 1, \ldots, m$ $\sum_{j=1}^{L} \lambda_j y_{rj} - s_r^+ = y_{ro}; \ r = 1, \ldots, s$	(1)	$max \ \varphi + \epsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)$ $s.t.$ $\sum_{j=1}^{L} \lambda_j x_{ij} + s_i^- = x_{io}; \ i = 1, \ldots, m$ $\sum_{j=1}^{L} \lambda_j y_{rj} - s_r^+ = \varphi y_{ro}; \ r = 1, \ldots, s$ (2)
CCR	$\lambda_j \geq 0, s_i^- \geq 0, s_r^+ \geq 0; \ \ j = 1, \ldots, L; \ r = 1, \ldots, s \ ; i = 1, \ldots, m$		(3)
BCC	$\sum_{j=1}^{L} \lambda_j = 1$		(4)
	where $\mathbf{X_j} = (x_{1j}, .., x_{mj})$ and $\mathbf{Y_j} = (y_{1j}, .., y_{sj})$ are the m inputs and the s outputs of the SPU$_j$, s_i^- and s_r^+ are slack variables and $\varepsilon > 0$ is a non-Archimedian element.		

$\widehat{Y}_j$ and the estimator $\widehat{\varphi}_j$ such that:

$$P \left(\left| \widehat{\varphi}_j - \varphi_j \right| \leq \delta \right) \geq 1 - \alpha; \ j = 1, \ldots, L. \qquad (11)$$

The second objective is to determine a confidence interval for the populational efficiency, in each SPU, with a fixed accuracy.

3 How many providers and customers need to be interviewed?

The provider and customer sample size problem Eq. 11 can only be analytically resolved in the case of one input and one output and the CCR model. A rigorous proof of all the results are provided in the Appendix I.

3.1 CCR model with one provider and one customer opinion index

Let us consider these assumptions:

C1 Fixed L SPUs.
C2 One provider population opinion index $\{X_j\}_{j=1,...,L}$ as input and one customer population opinion index $\{Y_j\}_{j=1,...,L}$ as output. We consider $\{\widehat{X}_j\}_{j=1,...,L}$ and $\{\widehat{Y}_j\}_{j=1,...,L}$ to be the corresponding estimators.
C3 CCR model Eq. 3 with output orientation (CCR-O).

Let $(Z_j = \frac{Y_j}{X_j})_{j=1, \ldots, L}$ and $(\widehat{Z}_j = \frac{\widehat{Y}_j}{\widehat{X}_j})_{j=1,...,L}$. In this situation and with the given notation, the population efficiency obtained with the CCR-O model and its estimator

are:

$$\varphi_j = \frac{Z_j}{max_{j=1,...,L} \{Z_j\}} \qquad (12)$$

$$\widehat{\varphi}_j = \frac{\widehat{Z}_j}{max_{j=1,...,L} \{\widehat{Z}_j\}}. \qquad (13)$$

Having fixed $p \in (0, 1)$, we thus consider the sets of Ω:

$$A_j = \left(\left| \widehat{Y}_j - Y_j \right| \leq p Y_j \right) \cap \left(\left| \widehat{X}_j - X_j \right| \leq p X_j \right); \ j = 1, \ldots, L \quad (14)$$

The following Lemmas 1, 2 and 3 are used to prove Theorem 1 which, in this particular case, establishes the relation between the accuracy of the provider and customer opinion estimation indexes and that of the population CCR-O efficiency estimation.

Lemma 1 *Under assumptions C1, C2 and C3, with the given notation, fixed $p \in (0, 1)$ and m such that $Z_m = max_{j=1,...,L} \{Z_j\}$, then:*

$$A_j = \left(\frac{(1-p)}{(1+p)} \widehat{Z}_j \leq Z_j \leq \frac{(1+p)}{(1-p)} \widehat{Z}_j \right)$$

$$= \left(Z_j \frac{(1-p)}{(1+p)} \leq \widehat{Z}_j \leq Z_j \frac{(1+p)}{(1-p)} \right)$$

$$= \left(\frac{\widehat{Z}_j(1-p)}{Z_m(1+p)} \leq \varphi_j \leq \frac{\widehat{Z}_j(1+p)}{Z_m(1-p)} \right).$$

Lemma 2 *Under assumptions C1, C2 and C3, with the given notation, fixed $p \in (0, 1)$, m and Z_m defined as in Lemma 1 and l such that :*

$$\widehat{Z}_l = max_{j=1,...,L} \{\widehat{Z}_j\}$$

Consider the set of Ω

$$B = \left(Z_m \frac{(1-p)}{(1+p)} \leq \widehat{Z}_l \leq Z_m \frac{(1+p)}{(1-p)} \right), \qquad (15)$$

then:

$m = l$ implies $A_l = B$

$m \neq l$ implies $A_l \cap A_m \subset B$

Lemma 3 *Under assumptions C1, C2 and C3, with the given notation, fixed $p \in (0, 1)$ and for any $j = 1, \ldots, L$ we have that*

$$A_j \cap B \subset \left\{ \widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \leq \frac{\widehat{Z}_j(1-p)}{Z_m(1+p)} \leq \varphi_j \right\}$$
$$\cap \left\{ \varphi_j \leq \frac{\widehat{Z}_j(1+p)}{Z_m(1-p)} \leq \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right\}$$

Theorem 1 *Let us consider the assumptions C1, C2 and C3. Having fixed $p \in (0, 1)$ and $\alpha \in (0, 1)$, if $P(A_j) \geq 1 - \alpha \; \forall j = 1, \ldots, L$, then*

$$P\left(\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \leq \varphi_j \leq \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right) \geq (1-\alpha)^3. \quad (16)$$

Lemma 4 is an instrumental result, used to prove Theorem 2 and Corollary 1, which establishes how to determine the provider and customer sample size in each SPU, so the CCR-O efficiency estimator has the precision fixed in Eq. 11.

Lemma 4 *Under the hypotheses of Theorem 1, if*

$$\varphi_j \in \left[\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2}, \min \left\{ 1, \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right\} \right]$$

then

$$\varphi_j \in \left[\widehat{\varphi}_j - \frac{4p}{(1+p)^2}, \min \left\{ 1, \widehat{\varphi}_j + \frac{4p}{(1+p)^2} \right\} \right]$$

Theorem 2 *Let us consider the assumptions C1, C2 and C3. Having fixed $p \in (0, 1)$ and $\alpha \in (0, 1)$, for every $j = 1, \ldots, L$, let n_{x_j} be the sampling size in the SPU_j, such that*

$$P\left(\left| \widehat{X}_j - X_j \right| \leq pX_j \right) \geq \sqrt[6]{1-\alpha} \quad (17)$$

and n_{y_j} be the sampling size such that

$$P\left(\left| \widehat{Y}_j - Y_j \right| \leq pY_j \right) \geq \sqrt[6]{1-\alpha} \quad (18)$$

then

$$P\left(\left| \widehat{\varphi}_j - \varphi_j \right| \leq \frac{4p}{(1+p)^2} \right) \geq 1 - \alpha; \; \forall j = 1, \ldots, L. \quad (19)$$

Corollary 1 *Let us consider the assumptions C1, C2 and C3. Having fixed $\delta \in (0, 1)$ and $\alpha \in (0, 1)$, for every $j = 1, \ldots, L$, let n_{x_j} be the sampling size in the SPU_j, such that*

$$P\left(\left| \widehat{X}_j - X_j \right| \leq \left(\frac{2 - \delta - 2\sqrt{1-\delta}}{\delta} \right) X_j \right) \geq \sqrt[6]{1-\alpha} \quad (20)$$

and n_{y_j} be the sampling size such that

$$P\left(\left| \widehat{Y}_j - Y_j \right| \leq \left(\frac{2 - \delta - 2\sqrt{1-\delta}}{\delta} \right) Y_j \right) \geq \sqrt[6]{1-\alpha} \quad (21)$$

then

$$P\left(\left| \widehat{\varphi}_j - \varphi_j \right| \leq \delta \right) \geq 1 - \alpha; \; \forall j = 1, \ldots, L \quad (22)$$

Remark 1 gives the explicit formula to obtain the sample size under the usual simple random sample without replacement sample design.

Remark 1 If the design in each SPU is simple random sampling without replacement, and the output (i.e. input) is the mean of all the answers of the population to a survey item, then the sampling size n_{θ_j} that it verifies

$$P\left(\left| \widehat{\theta}_j - \theta_j \right| \leq p\theta_j \right) \geq 1 - \alpha_1; \; \theta = X \text{ or } Y; \; p, \alpha_1 \in (0, 1)$$

is ([36])

$$n_{\theta_j} \geq \frac{n_{oj}}{\left(\frac{n_{oj}}{N_j} + 1 \right)} \quad (23)$$

with $n_{oj} = \frac{\tau_{1-\alpha_1/2}^2}{(p\theta_j)^2} \sigma_{\theta_j}{}^2$ and $\tau_{1-\alpha_1/2} = \phi^{-1}(1 - \alpha_1/2)$, where $\sigma_{y_j}{}^2$ is the population variance and ϕ the normal standard distribution function.

3.2 CCR or BCC model with two or more provider and/or customer opinion indexes

In this section, we report on our simulation study to check that Theorem 2 and Corollary 1 also work in the BCC model with two or more estimated provider and/or customer opinion indexes.

If we consider m items (the same in all SPUs) to estimate the m provider opinion indexes (inputs) $(X_{1j}, \ldots, X_{mj})$ with $(\widehat{X}_{1j}, \ldots, \widehat{X}_{mj})$, Remark 1 calculates the sample size $n_{x_{ij}}$ necessary to achieve

$$P\left(\left| \widehat{X}_{ij} - X_{ij} \right| \leq pX_{ij} \right) \geq \sqrt[6]{1-\alpha}; \; i = 1, ..., m; \; j = 1, ..., L. \quad (24)$$

We propose to determine the provider sample size n_{x_j} in the SPU_j as

$$n_{x_j} = max_{i=1,...,m} \{n_{x_{ij}}\}, \qquad (25)$$

i.e., if we consider s items (the same in all SPUs) to estimate the s customer opinion indexes (outputs) the provider sample size n_{y_j} in the SPU_j is determined as

$$n_{y_j} = max_{r=1,...,s} \{n_{y_{rj}}\}, \qquad (26)$$

where $n_{y_{rj}}$ is the sample size necessary to achieve

$$P\left(\left|\widehat{Y}_{rj} - Y_{rj}\right| \leq pY_{rj}\right) \geq \sqrt[6]{1-\alpha};\ r=1,...,s;\ j=1,...,L. \qquad (27)$$

$$\begin{pmatrix} U_{k1j} \\ U_{k2j} \end{pmatrix} \hookrightarrow N_2\left(\begin{pmatrix} doct_j \\ nurs_j \end{pmatrix}, \begin{pmatrix} doct_j^2/4 & 0 \\ 0 & nurs_j^2/4 \end{pmatrix}\right); k=1,....,N_{x_j}; j=1,...,12$$

$$\begin{pmatrix} W_{h1j} \\ W_{h2j} \end{pmatrix} \hookrightarrow N_2\left(\begin{pmatrix} out_j \\ inp_j \end{pmatrix}, \begin{pmatrix} out_j^2/4 & 0 \\ 0 & inp_j^2/4 \end{pmatrix}\right); h=1,....,N_{y_j}; j=1,...,12$$

where doc_j, $nurs_j$ and out_j, inp_j are the original value doctor, nurse and outpatient, inpatient of the jth health center, columns 2, 3, 4 and 5 of Table 2, respectively. We consider the population mean to the simulated answers to the provider and customer items in the jth center, $j = 1, ..., 12$, to simulate the population inputs and outputs, columns 3, 4, 5 and 6 in Table 3:

$$(X_{1j}, X_{2j}) = \left(\frac{\sum_{k=1}^{N_{x_j}} U_{k1j}}{N_{x_j}}, \frac{\sum_{k=1}^{N_{x_j}} U_{k2j}}{N_{x_j}}\right)$$

$$(Y_{1j}, Y_{2j}) = \left(\frac{\sum_{h=1}^{N_{y_j}} W_{h1j}}{N_{y_j}}, \frac{\sum_{h=1}^{N_{y_j}} W_{h2j}}{N_{y_j}}\right)$$

The last two columns in Table 3 show the population efficiency scores CCR and BCC with output orientation.

To check the relation between sample size, estimation of the input/output indexes and estimation of the DEA efficiency, using Theorem 2 and Corollary 1 and these simulated population data, we follow the next steps:

i. In the jth health center, $j = 1, ..., 12$, a previous simple random sample without replacement of 25 providers $\left(n_{x_j}^{(0)} = 25\right)$ is taken to estimate the two inputs $\left(\widehat{x}_{1j}^{(0)}, \widehat{x}_{2j}^{(0)}\right)$, and their variances $\left(\widehat{\sigma}_{1jx}^{2(0)}, \widehat{\sigma}_{2jx}^{2(0)}\right)$ using the sample means and the sample quasi-variances, respectively, i.e., we estimate the two outputs $\left(\widehat{y}_{1j}^{(0)}, \widehat{y}_{2j}^{(0)}\right)$ and their variances, $\left(\widehat{\sigma}_{1jy}^{2(0)}, \widehat{\sigma}_{2jy}^{2(0)}\right)$ with a previous simple random sample without replacement of 25 customers $\left(n_{y_j}^{(0)} = 25\right)$.

3.2.1 Simulation study

We use the [13] health center data (Table 2) to simulate a population model: in the jth health center a population size of providers and customers, N_{x_j} and N_{y_j}, are generated from a random uniform distribution, according to the intervals [10000, 50000] and [30000, 80000], respectively. For each provider, we generate two item answers (U_{k1j}, U_{k2j}), i.e., for each customer (W_{h1j}, W_{h2j}), from a bivariate normal distribution as

ii. Fixed $\delta = 0.1$ or 0.2 and $1 - \alpha = 0.9$ as in Corollary 1, and with the estimates of step i., the sample sizes, n_{x_j} and n_{y_j}, are determined using Eqs. 23, 25 and 26.

iii. In the jth health center, the simple random samples without replacement of size n_{x_j} and n_{y_j} are taken and the inputs $(\widehat{x}_{1j}, \widehat{x}_{2j})$ and outputs $(\widehat{y}_{1j}, \widehat{y}_{2j})$ are estimated. With the data $\left\{(\widehat{x}_{1j}, \widehat{x}_{2j}, \widehat{y}_{1j}, \widehat{y}_{2j})\right\}_{j=1,...,12}$, the estimated efficiencies $\{\widehat{\omega}_j\}_{j=1,...,12}$ are obtained, maximizing the LP model Eqs. 3 or 4 with output orientation.

iv. One thousand iterations of step iii. are carried out obtaining, for the jth health center, 1000 estimated efficiency scores $\left\{\widehat{\omega}_j^{(k)}\right\}_{k=1,...,1000}$ and 1000 intervals

$$H_j^{(k)} = \left[\widehat{\omega}_j^{(k)} - \delta, min\left\{\widehat{\omega}_j^{(k)} + \delta, 1\right\}\right];\ k=1,...,1000 \qquad (28)$$

v. The probability $P\left(\left|\widehat{\varphi}_j - \varphi_j\right| \leq \delta\right)$ is approximated by calculating

$$C_j = \frac{1}{1000} \sum_{k=1}^{1000} I_{\left(\varphi_j \in H_j^{(k)}\right)} \qquad (29)$$

Table 4 shows the sampling sizes, n_{x_j} and n_{y_j}, obtained for the jth health center in the last iteration, for the two values of δ and $\alpha = 0.1$. In health center 3 or 6, the customer sample size increases up to 6 times when fixing a maximum $\delta = 0.2$ to 0.1.

The probabilities $P\left(\left|\widehat{\varphi}_j - \varphi_j\right| \leq \delta\right)$ approximated with Eq. 29 take the value one for all the health centers, the two values of δ, the CCR-O or BCC-O model and with $\alpha = 0.1$. Therefore, the confidence intervals for the population

Table 2 Number of doctors, nurses, outpatients and inpatients in 12 health centers

Health center	Doctor	Nurse	Outpatient	Inpatient	Efficiency score CCR-O	Efficiency score BCC-O
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

Font: Table 1.5 [13]

efficiency score φ_j, obtained with the samples of the size of Table 4 and Corollary 1, are very conservative. However, in the next section, we will see that these same sample sizes allow less conservative bootstrap DEA efficiency confidence intervals to be obtained.

4 Description of the bootstrap efficiency confidence interval technology

Bootstrap uses resampling to estimate the value of a parameter of a population ([18]). For the problem suggested in Section 2, we propose bootstrap resampling of the samples of the provider and customer answers to the opinion item to obtain confidence intervals for the population efficiencies, following these steps:

i. Having fixed δ and a probability $1 - \alpha$, we determine the sample sizes n_{x_j} and n_{y_j}, in the SPU_j, using Corollary 1, Remark 1 and Eqs. 25 and 26.

1. In the SPU_j, we take a provider and a customer simple random samples without replacement $\{\mathbf{u_{kj}} = (u_{k1j}, ..., u_{kmj})\}_{k=1,...,n_{x_j}}$ and $\{\mathbf{w_{hj}} = (w_{h1j}, ..., w_{hsj})\}_{h=1,...,n_{y_j}}$, respectively, to estimate the provider and customer opinion indexes $(\widehat{x}_{1j}, \ldots, \widehat{x}_{mj})$ and $(\widehat{y}_{1j}, \ldots, \widehat{y}_{sj})$, respectively, for example, with the sample means:

$$\widehat{x}_{ij} = \frac{\sum_{k=1}^{n_{x_j}} u_{kij}}{n_{x_j}}; \ i = 1, ..., m; \ j = 1, ...L \quad (30)$$

$$\widehat{y}_{rj} = \frac{\sum_{h=1}^{n_{y_j}} w_{hrj}}{n_{y_j}}; \ r = 1, ..., s; \ j = 1, ...L. \quad (31)$$

ii. In the SPU_j, we take a bootstrap sample with replacement $\{\mathbf{u_{kj}}^*\}_{k=1,...,n_{x_j}}$ from $\{\mathbf{u_{kj}}\}_{k=1,...,n_{x_j}}$, i.e., $\{\mathbf{w_{hj}}^*\}_{h=1,...,n_{y_j}}$ of size n_{y_j} from $\{\mathbf{w_{hj}}\}_{h=1,...,n_{y_j}}$, with which we obtain the bootstrap version of the m inputs, $\widehat{\mathbf{x}}_j^* = (\widehat{x}_{1j}^*, ..., \widehat{x}_{mj}^*)$, and s outputs, $\widehat{\mathbf{y}}_j^* = (\widehat{y}_{1j}^*, ..., \widehat{y}_{sj}^*)$. With the data $\left\{ (\widehat{x}_{1j}^*, ..., \widehat{x}_{mj}^*, \widehat{y}_{1j}^*, ..., \widehat{y}_{sj}^*) \right\}_{j=1,...,L}$ and the DEA model of Table 1, we obtain the bootstrap version of the estimated DEA efficiency, $(\widehat{\omega}_j^*)_{j=1,...,L}$.

iii. The step iii. is repeated B times and the B bootstrap versions of the estimated DEA efficiency for the SPU_j, $j = 1, ..., L$, are $\left\{ \widehat{\omega}_j^{*(b)} \right\}_{b=1,..., B}$.

2. In the SPU_j, the observed percentile bootstrap confidence interval for the population efficiency score φ_j is obtained, having fixed a coverage intention of level $1 - \alpha'$, as

$$I_j = \left[\widehat{\omega}_j^{*(\alpha'/2)}, \widehat{\omega}_j^{*(1-\alpha'/2)} \right] \quad (32)$$

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

4.1 Simulation study

To illustrate the bootstrap efficiency confidence interval methodology and to check the estimation quality of the DEA population efficiency obtained, a simulation was performed using the population model simulated from Table 3.

First, we fixed $\delta = 0.2$ and the confidence $1 - \alpha = 0.9$ to calculate the provider and customer sample size to

Table 3 Simulated population model

Health center	Provider population size N_{x_j}	X1	X2	Customer Population size N_{y_j}	Y1	Y2	Population efficiency score CCR-O	Population efficiency score BCC-O
1	16684	2.04	15.52	57128	10.29	9.25	1	1
2	26950	1.95	13.40	36283	15.36	5.15	1	1
3	26583	2.56	16.41	65876	16.41	5.66	0.882	0.926
4	18974	2.77	17.22	33569	18.44	7.37	1	1
5	43106	2.26	16.26	62522	9.63	6.79	0.761	0.764
6	21305	5.65	26.27	68349	23.68	9.24	0.834	0.952
7	10599	3.40	23.98	36417	22.61	9.07	0.901	1
8	27432	3.18	21.13	73596	15.60	8.22	0.798	0.828
9	11113	3.09	24.94	38962	19.54	10.29	0.954	0.987
10	31717	5.15	27.44	76479	25.70	10.28	0.875	1
11	39802	5.44	31.44	36356	26.63	15.12	0.955	1
12	32282	3.91	29.26	55847	25.67	12.33	0.955	1

estimate the input and output data; supposing a simple random sample without replacement, these sample sizes are columns 2 and 3 in Table 4.

The steps ii.-v. are iterated 1000 times, fixing the confidence of the bootstrap efficiency interval $1 - \alpha' = 0.9$ or 0.95 and 2000 resampling Bootstraps ($B = 2000$).

The confidence of the bootstrap interval for the DEA efficiency in the SPU_j is approximated with:

$$C_j = \frac{1}{1000} \sum_{k=1}^{1000} I_{(\varphi_j \in I_j^{(k)})}; \quad j = 1, ..., 12 \tag{33}$$

where $\left\{ I_j^{(k)} = [I_j^{(k)L}, \ I_j^{(k)U}] \right\}_{k=1,...,1000}$ are the 1000 bootstrap efficiency confidence intervals obtained in step v.

Table 5 shows the approximate confidence of the bootstrap intervals, output orientation DEA models. We

observe that the control of the coverage level $1 - \alpha'$ leads to the achievement of the confidence of the bootstrap efficiency interval required by the experimenter.

The amplitude of the bootstrap efficiency confidence intervals is analysed with the approximation of the expected value of the bounds

$$E\left(I_j^{L}\right) = \frac{\sum_{k=1}^{1000} I_j^{(k)L}}{1000} \text{ and } E\left(I_j^{U}\right) = \frac{\sum_{k=1}^{1000} I_j^{(k)U}}{1000}. \tag{34}$$

Table 6 shows the approximation of the expected values of the bounds of the bootstrap efficiency confidence interval, considering the BCC model with output orientation. If

Table 4 Customer and provider sample size, taking two values of δ, 0.2 or 0.1, and a confidence $1 - \alpha = 0.9$

Health center	$\delta = 0.2$		$\delta = 0.1$	
	Provider sample size	Customer sample size	Provider sample size	Customer sample size
1	463	560	1747	2654
2	681	358	1858	3627
3	398	380	2481	1268
4	544	487	1047	1749
5	495	404	2440	1807
6	441	403	1650	2397
7	399	551	1533	2131
8	492	418	2010	2519
9	542	371	1495	2045
10	394	515	1731	2900
11	457	612	2872	1456
12	464	257	1265	2015

Table 5 Simulated confidence of bootstrap efficiency intervals, having fixed $\delta = 0.2$

SPU	$1 - \alpha' = 0.90$		$1 - \alpha' = 0.95$	
	CCR-O	BCC-O	CCR-O	BCC-O
1	100	100	100	100
2	100	100	100	100
3	92.5	91.9	96.5	95.8
4	98.4	99.9	99.5	99.9
5	92.2	91	96.5	95.5
6	89.7	89.9	95.2	94.6
7	89.3	98.3	93.7	99.2
8	91.1	92.4	96.3	97.3
9	91.9	91.4	96.4	95.8
10	90.6	98.9	94.8	99.4
11	89.4	100	95	100
12	92	100	95	100

Output-oriented CCR and BCC model

we look at the SPUs in which the expected efficiency confidence interval contains the one, {1, 2, 4, 7, 10, 11, 12}, these SPUs coincide with the efficient population units (value 1 in column 9 from Table 3). As expected, the increase in the trust $1 - \alpha'$ of the interval Bootstrap leads to an increase in the amplitude.

In conclusion, the bootstrap efficiency confidence interval methodology has the advantage that, after determining the provider and customer sample size using Corollary 1, Remark 1 and Eqs. 25 and 26, the experimenter can achieve the confidence required, $1 - \alpha'$, to estimate the population efficiency.

5 Application to the Spanish health system

This section provides an empirical analysis of health production for Spain's 18 Autonomous Communities (CCAA).

Spain's Health Ministry has, for some time, been compiling the statistic "Health Barometer" (HB), where a group of individuals is selected in each CCAA, and a questionnaire is carried out to test the health system. One of the survey question blocks take the opinion of the individual concerning the attention provided by the doctors of primary attention and pediatrics in Spain with the following items:

Table 6 Approximation of the expected values of the bounds of the bootstrap efficiency confidence intervals, $\delta = 0.2$, $\alpha = 0.1$ and $(1 - \alpha') = 0.9$ or 0.95

SPU	$1 - \alpha' = 0.9$		$1 - \alpha' = 0.95$	
	$E\left(I_i^{L}\right)$	$E\left(I_i^{U}\right)$	$E\left(I_i^{L}\right)$	$E\left(I_i^{U}\right)$
1	1	1	1	1
2	1	1	1	1
3	0.880	0.966	0.872	0.974
4	0.979	1	0.972	1
5	0.737	0.821	0.731	0.835
6	0.894	0.985	0.885	0.990
7	0.956	1	0.948	1
8	0.796	0.854	0.791	0.860
9	0.941	0.998	0.932	0.999
10	0.954	1	0.943	1
11	1	1	1	1
12	1	1	1	1

Output-oriented BCC model

Either from your personal experience or in your own opinion, we would like you to evaluate the following aspects of the public health service, concerning the attention provided by the GP or the paediatrician. Do so using the scale of 1 to 10, where 1 means 'totally unsatisfactory' and 10 means 'totally satisfactory'.

- **(P-1)** The attention received from the healthcare personnel
- **(P-2)** The time dedicated by the doctor to each patient
- **(P-3)** The confidence and security that the doctor transmits
- **(P-4)** The information received concerning your health problem
- **(P-5)** The time between making the appointment and the visit to the doctor.

A principal components analysis (PCA) is carried out over these 5 items. The PCA is a statistical technique for reducing the variable dimensionality of the dataset, minimizing information loss. It does so by creating new uncorrelated variables that successively maximize the explained variance of the dataset ([24]). The first component (PCA1) obtained, interpretable as the size of satisfaction, explains 98.1% of the variability. The sample mean of the answers to this component in every CCAA is our estimated output index, interpreted as the patient mean satisfaction with the CCAA's healthcare system (column 9 in Table 7).

The input data is estimated using the results of the "Survey on the current situation of GPs in Spain", carried out by the Spanish Medical Colleges Organization (OMC) in 2015 on the population of Spanish GPs and paediatricians. We use the following items:

- **(C1)** Workload as number of patients attended per day, answering 1 if the workload is normal or low (inferior to 40 patients) and 0 if the workload is high.
- **(C2)** Occupation of the team of doctors, answering 1 if the occupation is normal or low and 0 if it is high.
- **(C3)** Time dedicated to ongoing training.

The sample means of the doctors, answers to these items in each CCAA estimate the three opinion indexes used as inputs: the proportion of doctors with a normal or low workload $\left(\overline{C1}\right)$, the proportion of teams of doctors with a normal or low occupation $\left(\overline{C2}\right)$, and the mean time doctors dedicated to ongoing training $\left(\overline{C3}\right)$. From our point of view, an increase in the value of these input indexes in the population of doctors of a CCAA would lead to a bigger satisfaction in the population of patients attended in the same CCAA. Columns 2 and 3 of Table 7 show the

Table 7 Spain's Autonomous Communities, population size and sample size of providers, estimation of the inputs, population size and sample size of customers and estimation of the output

| CCAA | N_{x_j} | n_{x_j} | Input data | | | N_{y_j} | n_{y_j} | Output data |
			$\overline{C1}$	$\overline{C2}$	$\overline{C3}$			$\overline{PCA1}$
1	5960	950	0.361	0.420	27.34	8399043	726	15.830
2	1132	76	0.947	0.092	31.91	1317847	319	18.109
3	762	133	0.699	0.263	27.73	1051229	306	16.632
4	666	107	0.907	0.224	31.66	1104479	286	16.634
5	1486	32	0.594	0.219	35.16	2100306	349	16.462
6	444	140	0.793	0.157	22.86	585179	243	17.610
7	1600	203	0.419	0.320	25.43	2059191	348	16.735
8	2630	383	0.705	0.272	29.21	2472052	382	16.999
9	5441	243	0.934	0.243	29.37	7508106	707	15.947
10	3553	290	0.645	0.283	27.67	4980689	540	16.250
11	950	178	0.562	0.124	25.00	1092997	315	16.202
12	2194	136	0.765	0.191	28.13	2732347	405	15.965
13	4383	431	0.594	0.179	30.08	6436996	607	16.379
14	1072	197	0.457	0.051	21.89	1467288	320	16.649
15	497	150	0.940	0.060	21.00	640476	269	17.657
16	1780	291	0.959	0.107	22.12	2189257	384	16.650
17	257	108	0.759	0.157	28.24	317053	241	17159
18	93	12	0.333	0.083	20.83	169847	464	14.707

Table 8 Spain's CCAAs, public health efficiency scores estimated point and by bootstrapping confidence interval, with $1 - \alpha' = 0.9$, using data from The Health Barometer of 2015 and the "Survey on the situation of Primary Care doctors in Spain, 2015"

| CCAA | $\widehat{\omega}_i$ | Bootstrap efficiency confidence interval | |
		Lower bound	Upper bound
1	1	0.941	1
2	1	1	1
3	0.956	0.923	0.977
4	0.925	0.903	0.950
5	0.961	0.924	0.994
6	1	0.983	1
7	1	0.988	1
8	0.977	0.949	0.989
9	0.886	0.868	0.905
10	0.941	0.915	0.955
11	0.954	0.928	0.984
12	0.910	0.883	0.926
13	0.958	0.933	0.972
14	1	1	1
15	1	1	1
16	0.941	0.918	1
17	0.979	0.949	0.998
18	1	0.876	1

BCC-O model

population and sample size of doctors (i.e., columns 7 and 8 of patients).

Table 8 shows the results of the estimation point ($\widehat{\omega}_\mathbf{i}$) and bootstrap interval of the population efficiency scores, with confidence $1 - \alpha' = 0.9$, in each of Spain's CCAAs, considering variable returns-to-scale and output orientation. The CCAAs in which the hypothesis of an efficient public health service (DEA efficiency equal one) is rejected are {3, 4, 5, 8, 9 10, 11, 12, 13, 17}. The CCAAs which are benchmark for the rest are {2 6, 7, 14 , 15}, because the upper and lower confidence interval bounds have value one. In general, the efficiency in all CCAAs is good, the inferior bound of the confidence interval is superior to 0.9 in all the cases, except in the CCAA {12}, according to our results, the CCAA with the least efficient health service.

6 Conclusions

The approach presented in this paper provides a step towards producing valid estimates of technical efficiency in public services, using provider and customer satisfaction-opinion indexes estimated with samples. These indexes measure the service quality from the perspective of both the provider and the customer.

We have developed statistical results for comparing the efficiencies of public services. These results are novel in the sense that: (i) We resolve the problem of determining the customer and provider sample size necessary to estimate the opinion indexes and the population efficiency with an accuracy fixed a priori; (ii) We build confidence intervals for the population DEA efficiency using bootstrap replicates of the providers sample and the customers sample in each public service; (iii) It is possible to achieve the level of confidence of the bootstrap efficiency confidence interval required by the experimenter (iv) The DEA models used are the original linear programming models; (iv) The new approach can be readily implemented. As far as we know, the approach of this paper has not been attempted in the literature.

While this study provides a useful methodology to measure public service efficiency, its limitations should also be acknowledged. First, the results can only be proven analytically for the CCR case, with one input and one output. Second, to obtain the input and output data, a provider and customer opinion survey is necessary. Finally, the presented methodology also has to allow deterministic inputs and/or output data to be considered.

This statistical efficiency methodology, used with opinion indexes from doctors and patients allows us to conclude

that, in Spain, there are ten autonomous communities that can improve their efficiency and five autonomous communities that act as benchmarks for the rest.

The results of this paper can have other important implications in practice. It can also be used to measure the efficiency in all services with users and providers, for instance, markets, health care, banks, casinos, schools, universities or public transport.

Appendix I

Proof of Lemma 1

The first equality is obtained by

$$
\left(\left| \widehat{Y}_j - Y_j \right| \le p Y_j \right) = \left(\frac{\widehat{Y}_j}{(1+p)} \le Y_j \le \frac{\widehat{Y}_j}{(1-p)} \right) \quad (35)
$$

$$
\left(\left| \widehat{X}_j - X_j \right| \le p X_j \right) = \left(\frac{\widehat{X}_j}{(1+p)} \le X_j \le \frac{\widehat{X}_j}{(1-p)} \right) (36)
$$

and, dividing the upper and lower extremes of Eq. 35 by the lower and upper extremes of Eq. 36, respectively, we obtain

$$
A_j = \left(\frac{(1-p)}{(1+p)} \widehat{Z}_j \le Z_j \le \frac{(1+p)}{(1-p)} \widehat{Z}_j \right). \quad (37)
$$

The second equality is obtained by clearing $\widehat{Z}_j$, and the third by dividing by Z_m.

Proof of Lemma 2

If $m = l$, from Lemma 1, it follows that $A_l = B$.

If $m \ne l$, then $\widehat{Z}_m < \widehat{Z}_l$ and $Z_l < Z_m$ and, if $A_l \cap A_m$ is verified, it follows that:

$$
\frac{(1-p)}{(1+p)} Z_m \le \widehat{Z}_m < \widehat{Z}_l \le \frac{(1+p)}{(1-p)} Z_l < \frac{(1+p)}{(1-p)} Z_m.
$$

Proof of Lemma 3

Dividing the first inequality that defines the set B by $(1+p)$ and multiplying by $(1-p)$, we obtain:

$$
\frac{(1-p)^2}{(1+p)^2} \le \frac{\widehat{Z}_l (1-p)}{Z_m (1+p)}
$$

Now, for $j = 1, \ldots, L$, as $\widehat{\varphi}_j = \frac{\widehat{Z}_j}{\widehat{Z}_l}$, we have that:

$$
\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \le \widehat{\varphi}_j \frac{\widehat{Z}_l (1-p)}{Z_m (1+p)} = \frac{\widehat{Z}_j (1-p)}{Z_m (1+p)};
$$

Dividing the second inequality that defines the set B by $(1-p)$ and multiplying by $(1+p)$, we have:

$$
\frac{\widehat{Z}_l (1+p)}{Z_m (1-p)} \le \frac{(1+p)^2}{(1-p)^2}.
$$

Given that $j = 1, \ldots, L$: we get $\widehat{\varphi}_j \frac{\widehat{Z}_l (1+p)}{Z_m (1-p)} \le \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2}$ and, as $\widehat{\varphi}_j = \frac{\widehat{Z}_j}{\widehat{Z}_l}$, we get $\frac{\widehat{Z}_j (1+p)}{Z_m (1-p)} \le \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2}$.

As from Lemma 1, $A_j = \left(\frac{\widehat{Z}_j (1-p)}{Z_m (1+p)} \le \varphi_j \le \frac{\widehat{Z}_j (1+p)}{z_m (1-p)} \right)$, then, whenever $A_j \cap B$ happens, we have

$$
\left(\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \le \frac{\widehat{Z}_j (1-p)}{Z_m (1+p)} \le \varphi_j \le \frac{\widehat{Z}_j (1+p)}{Z_m (1-p)} \le \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right).
$$

Proof of Theorem 1

$$
P \left(\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \le \varphi_j \le \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right) \ge P \left(A_j \cap B \right)
$$
$$
\ge P \left(A_j \cap A_l \cap A_m \right) = (1-\alpha)^3
$$

The first inequality follows from Lemma 3, the second from Lemma 2, and the last inequality follows because the events A_j are independent.

Proof of Lemma 4

If $\widehat{\varphi}_j \frac{(1-p)^2}{(1+p)^2} \le \varphi_j \implies \widehat{\varphi}_j \left(1 - \frac{4p}{(1+p)^2} \right) \le \varphi_j \Rightarrow \widehat{\varphi}_j - \varphi_j \le \frac{4p}{(1+p)^2} \widehat{\varphi}_j \le \frac{4p}{(1+p)^2}$; the last inequality follows from $0 \le \widehat{\varphi}_j \le 1$.

If $min \left\{ 1, \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right\} = 1 \implies \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \ge 1 \implies \widehat{\varphi}_j \ge \frac{(1-p)^2}{(1+p)^2} \implies \widehat{\varphi}_j + \frac{4p}{(1+p)^2} \ge \frac{(1-p)^2}{(1+p)^2} + \frac{4p}{(1+p)^2} = 1$ and $min \left\{ 1, \widehat{\varphi}_j + \frac{4p}{(1+p)^2} \right\} = 1$.

If $\varphi_j \le min \left\{ 1, \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \right\} = \widehat{\varphi}_j \frac{(1+p)^2}{(1-p)^2} \implies \varphi_j \le \widehat{\varphi}_j \left(\frac{1}{1 - \frac{4p}{(1+p)^2}} \right) \implies \varphi_j \left(1 - \frac{4p}{(1+p)^2} \right) \le \widehat{\varphi}_j \implies \varphi_j - \widehat{\varphi}_j \le \frac{4p}{(1+p)^2} \varphi_j \le \frac{4p}{(1+p)^2}$; the last inequality follows from $0 \le \varphi_j \le 1$.

Proof of Theorem 2

The result follows from Theorem 1, Lemma 4 and $P(A_j) = P \left(\left(\left| \widehat{X}_j - X_j \right| \le p X_j \right) \cap \left(\left| \widehat{Y}_j - Y_j \right| \le p Y_j \right) \right) = P \left(\left| \widehat{X}_j - X_j \right| \le p X_j \right) P \left(\left| \widehat{Y}_j - Y_j \right| \le p Y_j \right)$.

Appendix II

The software to reproduce the methods is available in the following link https://uvaes-my.sharepovint.com/:f:/g/personal/jesus_tapia_uva_es/EpnYVaKm1OZMpKUyyrI-bS0BNUj7aDumIquTR7Ipxt422A?e=GvrDtG

This file contains the following programs:

– Program to take survey sampling without replacement "mas.m"
– Program to take survey sampling with replacement "mascon.m"
– Programs to obtain the sample size: "masnp.m" "masn.m" "masnvar.m"
– Program to obtain efficiency bootstrap confidence intervals "percboot2outp2inp2pob.m"
– Program to obtain efficiency CCR "ccrout.m"
– Program to obtain efficiency BCC "bccout.m"
– Program to obtain the intervals and the results of the simulation "CIE15ybootalfab01.m"
– Simulation data "datcoop2i2ost2pob.mat"

Declarations

Conflict of Interests Author 1 declares that he has no conflict of interest. Author 2 declares that he has no conflict of interest.

References

1. Almeida Botega L, Andrade MV, Guedes GR (2020) Brazilian hospitals? performance: an assessment of the unified health system (SUS). Health Care Manag Sci 23:443–452
2. Assaf A (2010) Bootstrapped scale efficiency measures of UK airports. J Air Transport Manag 16(1):42–44
3. Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092
4. Barra C, Zotti R (2016) Measuring efficiency in higher education: an empirical study using a bootstrapped data envelopment analysis. Int Adv Econ Res 22(1):11–33
5. Bayraktar E, Tatoglu E, Turkyilmaz A, Delen D, Zaim S (2012) Measuring the efficiency of customer satisfaction and loyalty for mobile phone brands with DEA. Expert Syst Appl 39(1):99–106
6. Benito B, Solana J, Moreno MR (2014) Explaining efficiency in municipal services providers. J. Product. Anal. 42(3):225–239
7. Ceyhan ME, Benneyan JC (2014) Handling estimated proportions in public sector data envelopment analysis. Ann Oper Res 221(1):107–132
8. Charles V, Kumar M (2014) Satisficing data envelopment analysis: an application to SERVQUAL efficiency. Measurement 51:71–80
9. Charnes A, Cooper WW (1959) Chance-constrained programming. Manag Sci 6(1):73–79
10. Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444
11. Chowdhury H, Zelenyuk V (2016) Performance of hospital services in Ontario: DEA with truncated regression approach. Omega 63:111–122
12. Cooper WW, Deng H, Huang Z, Li SX (2002) Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. J Oper Res Soc 53(12):1347–1356
13. Cooper W. W., Seiford L. M., Tone K (2006) Introduction to data envelopment analysis and its uses: with DEA-solver software and references. Springer Science & Business Media
14. Cooper W. W., Seiford L. M., Zhu J (2011) Handbook on data envelopment analysis. Springer Science & Business Media
15. Daraio C, Diana M, Di Costa F, Leporelli C, Matteucci G, Nastasi A (2016) Efficiency and effectiveness in the urban public transport sector: a critical review with directions for future research. Eur J Oper Res 248(1):1–20
16. De Witte K, Geys B (2013) Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. Eur J Oper Res 224(3):592–602
17. Diewert WE (2011) Measuring productivity in the public sector: some conceptual problems. J Prod Anal 36(2):177–191
18. Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. CRC press
19. Essid H, Ouellette P, Vigeant S (2014) Productivity, efficiency and technical change of Tunisian schools: a bootstrapped Malmquist approach with quasi-fixed inputs. Omega 42(1):88–97
20. Førsund FR (2017) Measuring effectiveness of production in the public sector. Omega 73:93–103
21. Gil Ropero A, Turias Dominguez I, Cerbán Jiménez MM (2019) Bootstrapped operating efficiency in container ports: a case study in Spain and Portugal. Ind Manag Data Syst 119(4):924–948
22. Hammond CJ (2002) Efficiency in the provision of public services: a data envelopment analysis of uk public library systems. Appl Econ 34(5):649–657
23. Hemmeter JA (2006) Estimating public library efficiency using stochastic frontiers. Public Finance Rev 34(3):328–348
24. Jobson JD (1992) Applied multivariate data analysis. Volume II: Categorical and Multivariate Methods. Springer
25. Kohl S, Schoenfelder J, Fügener A (2019) The use of Data Envelopment Analysis (DEA) in healthcare with a focus on hospitals. Health Care Manag Sci 22:245–286
26. Korhonen P, Tainio R, Wallenius J (2001) Value efficiency analysis of academic research. Eur J Oper Res 130(1):121–132
27. Lee H, Kim C (2014) Benchmarking of service quality with data envelopment analysis. Expert Syst Appl 41(8):3761–3768
28. Li Y, Chiu Y, Lin T, Huang Y. Y. u. (2019) Market share and performance in Taiwanese banks: min/max SBM DEA. TOP 2:233–252
29. Liu ST, Chuang M (2009) Fuzzy efficiency measures in fuzzy DEA/AR with application to university libraries. Expert Syst Appl **36**(2):1105–1113
30. Liu W, Wang Y-M, Lyu S (2017) The upper and lower bound evaluation based on the quantile efficiency in stochastic data envelopment analysis. Expert Syst Appl 85:14–24

31. Mayston DJ (2015) Analysing the effectiveness of public service producers with endogenous resourcing. J Prod Anal 44(1):115–126

32. Mayston DJ (2017) Data envelopment analysis, endogeneity and the quality frontier for public services. Ann Oper Res 250(1):185–203

33. Olesen OB, Petersen N (1995) Chance constrained efficiency evaluation. Manag Sci 41(3):442–457

34. Ruggiero J (1996) On the measurement of technical efficiency in the public sector. Eur J Oper Res 90(3):553–565

35. Santín D, Sicilia G (2017) Dealing with endogeneity in data envelopment analysis applications. Expert Syst Appl 68:173–184

36. Särndal C-E, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Science & Business Media

37. Shwartz M, Burgess JF, Zhu J (2016) A DES based composite measure of quality and its associated data uncertainty interval for health care provider profiling and pay-for-performance. Eur J Oper Res 253(2):489–502

38. Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Manag Sci 44(1):49–61

39. Simar L, Wilson PW (2000) A general methodology for bootstrapping in non-parametric frontier models. J Appl Stat 27(6):779–802

40. Tapia JA, Salvador B, Rodríguez JM (2018) Data envelopment analysis in satisfaction survey research: sample size problem. J Oper Res Soc 69:7,1096–1104

41. Tapia JA, Salvador B, Rodríguez JM (2019) Data envelopment analysis efficiency of public services: bootstrap simultaneous confidence region. SORT-stat Oper Res Trans:337–354

42. Tapia JA, Salvador B, Rodríguez JM (2020) Data envelopment analysis with estimated output data: Confidence intervals efficiency. Measurement 152(107364)

43. Zandkarimkhani S, Mina H, Biuki M (2020) A chance constrained fuzzy goal programming approach for perishable pharmaceutical supply chain network design. Ann Oper Res 295:425–452