



# Nonparametric efficiency measurement of undergraduate teaching by university size

Manuel Salas-Velasco<sup>1</sup>

Received: 29 November 2022 / Revised: 30 November 2023 / Accepted: 27 December 2023 /  
Published online: 19 February 2024  
© The Author(s) 2024

## Abstract

Conventional data envelopment analysis (DEA) models assume that all decision-making units (DMUs) are homogenous. While higher education institutions (HEIs) of very different sizes challenge the homogeneity of DMUs, DEA studies have paid relatively little attention to university size when assessing the performance of HEIs. This article proposes novel, effective methods for evaluating university performance and identifying useful benchmarks for improving the operations of inefficient performers. Specifically, DEA and cluster analysis (CA) are applied for the evaluation of the performance of traditional Spanish public universities. DEA is utilized to examine the relative performance of these universities in terms of undergraduate teaching output. CA is applied to find similar-in-scale universities prior to the DEA to facilitate peer-groupings. The advantage of this method is that when DMUs are clustered based on their size, one can obtain homogenous groups of units with comparable operating environments. Furthermore, using the meta-frontier framework, this research finds significant evidence that there is an efficiency advantage for medium- and large-sized universities over small ones in providing undergraduate teaching. A bootstrapped, non-parametric meta-frontier approach also verifies this latter result. Some of the factors that contribute to the differences in the relative efficiencies are identified as well.

## Highlights

- Nonparametric approach for performance evaluation of Spanish public universities.
- Technical efficiency of universities in the provision of undergraduate education.
- DEA and clustering methodology for efficiency measurement of HEIs of different size.
- Application of cluster analysis to identify homogeneous DMUs.
- Meta-frontier framework for comparing the efficiencies of heterogeneous groups.

**Keywords** Cluster analysis · Data envelopment analysis · meta-frontier frameworks · Nonparametric approach in efficiency analysis · Spanish public universities

## 1 Introduction

Performance evaluation of higher education institutions (HEIs) is vital for judging the degree to which resources made available to the higher education sector are managed in the process of achieving desired outcomes. Different techniques have been used in applied research to assess HEIs' learning and research results. These techniques range from the use of performance indicators (e.g., Johnes and Taylor 1990), benchmarking (e.g., OECD 2017), or university rankings (e.g., Rauhvargers 2011) to mathematical techniques such as data envelopment analysis (DEA) (e.g., Abbott and Doucouliagos 2003; Johnes 2006) or econometric analysis such as stochastic frontier analysis (SFA) (e.g., Titus and Eagan 2016). Both non-parametric and parametric frontier techniques allow efficiency measurement (e.g., Murillo-Zamorano 2004). However, we defend in this article the use of non-parametric efficiency analysis techniques such as DEA. On the one hand, DEA allows the simultaneous use of inputs and outputs, which is an important element in our study. In fact, the ease with which DEA can manage several inputs and outputs makes it a common choice of technique for computing the efficiency of universities (e.g., Nunamaker 1985). On the other hand, since our goal was also to find comparison groups for inefficient universities (i.e., efficient peers to emulate), DEA was a suitable benchmarking technique for our purpose (e.g., Rostamzadeh et al. 2021).

DEA was presented by Charnes et al. (1978) for evaluating the relative efficiencies of a set of decision-making units (DMUs), and it has been applied in numerous settings over the past 40 years, including higher education (Emrouznejad and Yang 2018; Liu et al. 2013). In the present study, the focus was on assessing the technical efficiency of traditional (campus-based) Spanish public universities in producing undergraduate teaching. This assessment exercise is relevant and desirable for several reasons.

Firstly, higher education in Spain consists almost exclusively of universities, and about 90% of students are enrolled in undergraduate education (Hernández-Armenteros and Pérez-García 2017). Research is now increasingly focused on student persistence and completion, both for equity and efficiency reasons (Salas-Velasco 2020a). Poor student academic performance is associated with two main inefficiencies. On the one hand, students will need a longer time than scheduled to finish university studies; that is, they will need more than four years to finish a four-year undergraduate degree program (e.g., Huntington-Klein and Gill 2021). On the other hand, poor student academic productivity is the prelude to dropping out of HEIs (e.g., Casanova et al. 2018).

Secondly, whether Spanish universities are using resources productively is a crucial research question in the context of university accountability. Debates in Spain about the performance of the HEIs have often focused on public universities because they are funded mostly by the public budget and enroll the vast majority of under-

graduate students. Spanish public universities are self-governing; however, funding is heavily dependent on regional governments (*comunidades autónomas*). Undergraduate studies in Spain are highly subsidized (tuition fees are relatively low). A good performance, including achieving excellent academic results, should therefore be the core mission of Spanish public universities.

Thirdly, Spanish universities have undergone profound changes over the past decade, primarily as a result of changes in undergraduate curricula. Spain signed the Bologna Declaration along with other European countries, which aimed to create the European Higher Education Area. In the 2010/2011 academic year, all Spanish universities started to offer new undergraduate degrees called *grados* (four-year bachelor's degree programs). However, we know little about the efficiency of undergraduate teaching provision at Spanish public universities after the implementation of the Bologna reform, and one of the objectives of this study was to fill that research gap. For the applied analysis of efficiency measurement, the 2015/2016 academic year was chosen, which is the first one with complete data once the Bologna reform is complete. Given that the primary mission of Spanish universities is undergraduate teaching, it is outside the scope of this paper to evaluate the efficiency of research and graduate education production. Furthermore, not all Spanish institutions have a postgraduate and research orientation.

This article innovates the existing literature in several ways, as we will discuss in the following sections. Essentially, the novelty of the current study lies in the fact that efficiency measurement was done in a setting where DMUs are non-homogeneous. Although traditional Spanish public HEIs (47 campus-based universities) are mainly fed by public money, share a common regulatory framework (Organic Law of the University System), and use similar educational production technology (distance/on-learning public universities and all private universities were excluded in this study), they range in size from “too small” to “too large”. The average size of the Spanish public university (on-campus degree programs) is around 19,200 enrolled undergraduate students and 1,600 full-time equivalent employed professors (Appendix Table A1). Nonetheless, there are noteworthy differences in both dimension indicators (students and faculty), with Universidad de La Rioja (just over 3,000 undergraduate students) and Universidad Complutense de Madrid (almost 53,000 undergraduate students) being at the extremes (Appendix Table A1).

Furthermore, this is not the only sort of heterogeneity that might be important. In general, “large” universities are “old” and are considered research-oriented institutions, while “small” universities are “young” and are essentially teaching institutions (Appendix Table A1). Carrington et al. (2005) already stated that we would expect that a large university would not be comparable to a small university and that an established research-intensive university should not be compared with a newly established university. More recently, Daraio et al. (2021) considered the foundation year (i.e., the year when the institution was established) among additional variables that are neither inputs nor outputs but which may affect the production process.

While HEIs of very different sizes challenge the homogeneity of DMUs, a basic assumption in DEA (Dyson et al. 2001), DEA studies have paid relatively little attention to university size when assessing the performance of HEIs. As we will show in this article, benchmarking, the practice of contrasting a DMU's performance to the

best practices formed by a set of DMUs, is debatable in the presence of HEIs' scale heterogeneity. Indeed, efficiency estimations have limited value if the heterogeneity among the observations is not taken into account (Fried et al. 2008).

Acknowledging the heterogeneity of Spanish public universities and aiming to compare them with appropriate peers might overcome that limitation. Potentially, the non-homogeneous DMU issue could be handled by breaking the set of DMUs into multiple groups and then doing a separate DEA analysis for each group. In this way, a DMU is evaluated against only "true" peers. In particular, in order to deliver realistic results, this article proposed splitting the sample of DMUs (i.e., Spanish public HEIs) into groups based on scale homogeneity using cluster analysis. The idea to cluster non-homogeneous units into homogeneous sets was already suggested in the DEA literature by Dyson et al. (2001). Once the homogeneous subsets were delimited, we assessed the efficiency of the HEIs in producing undergraduate teaching. Within each cluster, for each inefficient university, DEA showed its peers of similar scale to emulate, showing reasonable comparison results. Furthermore, this study aimed to test, using the meta-frontier framework, whether university size matters when explaining efficiency differences between groups. By measuring efficiency relative to a common meta-frontier, we are able to perform efficiency comparisons between groups where one set of DMUs is compared (benchmarked) against another.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces both linear programming models and empirical methods. Section 4 runs different DEA models on the pooled sample of HEIs. In Sect. 5, we divide the data set of HEIs into homogeneous subsets by university size using cluster analysis. In Sect. 6, we compute the teaching efficiency of institutions measured against their own group's frontier. The intra-group analysis provides much more realistic peer groups. Section 7 carries out an inter-group analysis using the meta-frontier framework. Section 8 discusses the results and policy implications. Finally, the summary of our conclusions is presented in Sect. 9.

## 2 Background

In microeconomics, production is the process of transforming inputs into outputs. In particular, production theory examines the relation between the quantity of inputs used within a system and the quantity of outputs that such a system is able to produce (Shephard 1971), where inputs and outputs are measured in physical or technical units. Technical efficiency can be defined as the ability of a system to optimize its physical resources in the production of output. Early studies, such as Koopmans (1951) and Farrell (1957), developed specific methods in order to analyze productive efficiency at a microeconomic level. Technical efficiency reflects the ability of the firm to produce the maximum amount of output from a set of resource inputs given the technology, and technical inefficiency arises when less than the maximum output is obtained from a given bundle of factors (Farrell 1957). The term "firm", sometimes replaced by the more encompassing term "DMU", may include non-profit or public organizations such as HEIs.

The performance evaluation carried out in this article stands on the microeconomic concept of technical efficiency. The microeconomic analysis applied to the study of efficiency in higher education should provide a measurement of how well (or efficiently) HEIs are using their physical inputs to produce physical outputs. As techniques such as DEA have developed to capture the performance of DMUs in a multi-output, multi-input production framework, DEA has also become a widely used technique for measuring the technical efficiency of HEIs. The results from the DEA should help university managers identify the possibilities of increasing teaching and research outputs while conserving resources or educational inputs. Universities are multi-product firms that produce multiple outputs from multiple inputs (Cohn et al. 1989), so productivity indicators based on simple ratios are unlikely to capture an accurate picture of performance (Johnes 2016). However, DEA will attach an efficiency score to each HEI as a summary measure of individual performance, and those institutions that get a score equal to one will be considered technically efficient. Technically efficient universities are not able to deliver more teaching and research output (without reducing quality) given their current capital, labor, and other inputs (Abbott and Doucouliagos 2003).

## 2.1 Production economics in the context of the higher education sector

While there is no agreement about the exact functioning of the universities' production processes (Worthington 2001), one needs to decide which input and output measures are associated with HEIs' activities in an exercise of efficiency measurement in higher education. Commonly accepted inputs to the university production process are the total number of full-time equivalent (FTE) academic staff, which is used to reflect labor input (e.g., Abbott and Doucouliagos 2003; Athanassopoulos and Shale 1997; Flegg et al. 2004; Johnes 2006; Salas-Velasco 2020a). Other separately designated inputs are the total number of undergraduate and graduate students (e.g., Athanassopoulos and Shale 1997; Flegg et al. 2004; Johnes 2006) and physical investments (e.g., Ahn et al. 1988). The outputs of the university production process most frequently used are the number of undergraduates and graduates as a proxy for teaching (e.g., Athanassopoulos and Shale 1997; Flegg et al. 2004; Johnes 2006). However, Avkiran (2001), Coelli (1996), and Abbott and Doucouliagos (2003) included the number of equivalent full-time students as a measure of teaching output. As a proxy for research activity, typical academic research output is the number of publications, such as articles in indexed journals (e.g., Gralka et al. 2019; Moreno-Gómez et al. 2020; Thanassoulis et al. 2016; Visbal-Cadavid et al. 2017). "Publication in high status refereed journals has become a major criterion of academic success in the competitive environment of global higher education" (Altbach 2015, p. 6). Other outputs that have been considered include the number of citations (e.g., Moncayo-Martínez et al., 2020). The performance assessment of the HEIs' activities has also been supplemented over the past few decades with a "third mission" involving the generation, use, application, and exploitation of university knowledge (Compagnucci and Spigarelli 2020). For the Spanish case, a better understanding of the role of the third mission in HEIs' performance was enabled by Berbegal-Mirabent et al. (2013) and De la Torre et al. (2017). In the former, knowledge transfer was proxied by the

number of spin-offs created in 2009, while in the latter, authors used intellectual or industrial property (IIP) agreements, income from R&D and consultancy contracts, and the number of university spin-offs established in the last 5 years.

However, in a strict sense of microeconomic theory, data envelopment analysis and its application to the measurement of technical efficiency in higher education should consider only physical or technical units of measurement. Farrell (1957) already proposed a measure for firm efficiency that was divided into two components. On the one hand, technical efficiency shows a firm's ability to maximize output levels depending on the input factor set. On the other hand, economic efficiency reflects the firm's capacity to use appropriate inputs for given technology and factor prices. Nevertheless, the majority of applied studies mixed technical units of measurement with monetary units when measuring educational inputs and outputs. For instance, Johnes (2006) used the value of interest payments and depreciation as a measure of capital stock. Also, as input measures, Athanassopoulos and Shale (1997) considered expenditure on library and computing services, while Flegg et al. (2004) used aggregate departmental expenditure. In relation to the measures of output, Abbott and Doucouliagos (2003) used the Research Quantum Allocation that each university received, while Johnes (2006) used the value of the recurrent grant for research awarded by the Higher Education Funding Council for England (HEFCE). Note also that this "research output" (income from applied research) might bias the results favoring institutions with strong fields in medical and experimental sciences (Kao and Hung 2008). In any case, all these works cannot be called proper studies that measure technical efficiency in the production of higher education.

In contrast to the previous studies, the approach followed in the current study has been to use non-monetary units for all inputs and outputs in the different DEA models. Furthermore, since the focus was on the measurement of technical efficiency in the provision of undergraduate teaching, the choice was guided by the selection made in other works that have assessed the efficiency performance of the Spanish university system before the university reform of 2010 (e.g., Salas-Velasco 2020a). In particular, in addition to the total number of professors (academic staff), we used the total number of students' academic loads as input. As an output variable of the instructional component of higher education, we used for each institution the total number of credits passed by its undergraduate students. The majority of university efficiency studies have considered enrolled undergraduate students as inputs and undergraduates as outputs, but they are individuals from different cohorts. "By using the credits earned over the total number of credits enrolled (registered), the efficiency in teaching can be assessed more accurately" (Salas-Velasco 2020b, p. 836) because we are considering the same time dimension of inputs and outputs. Besides, since our DEA models are fundamentally based on the "Farrellian" radial efficiency measurement, we avoided the use of variables measured in the form of ratios, percentages, and indices. All inputs and outputs were (total) volume measures (strictly positive continuous variables). In fact, the suitability of ratio data in DEA models is still a subject of academic debate (e.g., Gerami et al. 2022; Olesen et al. 2015).

## 2.2 DEA models for higher education performance evaluation

Performance evaluation of the education sector has received continuous attention in the DEA literature (e.g., Daraio 2020; Liu et al. 2013; Kounetas et al. 2023). Nigsch and Schenker-Wicki (2015) and De Witte and López-Torres (2017), among others, provided an overview of the strengths and limitations of the DEA approach for higher education performance evaluation over the past years. Many applications of DEA in higher education are aimed at just measuring HEIs' production efficiency and use mainly output-oriented radial DEA models. Some of them assumed constant returns to scale (CRS) (e.g., Flegg et al. 2004; Guironnet and Peypoch 2018), while others assumed variable returns to scale (VRS) (e.g., Avkiran 2001; Johnes 2006). The former measures the overall technical efficiency. The latter estimates the technical efficiency without considering inefficiencies due to the scale of operation of the productive unit. All these models are envelopment DEA models because the identified best-practice frontier envelops all the DMUs (Zhu 2014). These standard DEA models are also known as radial DEA models because they give the value of the radial efficiency score (maximal equiproportional reduction in all inputs or maximum equiproportional increase in all outputs).

Most studies on efficiency in higher education attempted to measure efficiency in the joint production of different university outputs (e.g., Abbot & Doucouliagos, 2003; Athanassopoulos and Shale 1997; Avkiran 2001; Breu and Raab 1994; Fandel 2007; García-Aracil 2013; Johnes 2006; McMillan and Datta 1998; Nazarko and Šaparauskas 2014). However, some research studies focused on the efficiency of producing either teaching only (e.g., Agasisti and Dal Bianco 2009) or research only (e.g., Johnes and Yu 2008; Ng and Li 2000). Cross-country efficiency assessments have also been presented (e.g., Agasisti and Pérez-Esparrells 2010; Sinuany-Stern and Hirsh 2021). Yet, results from studies adopting a cross-country perspective should be taken with caution because they mix institutions that face very different operating environments (legal, cultural, financial, etc.).

In addition to the standard envelopment DEA models, other DEA models such as directional distance function (DDF), closest target DEA, or slacks-based measure of efficiency (SBM) have been used in the literature (e.g., Johnes and Tone 2017; Ruiz et al. 2015; Villano and Tran 2018). DEA models have also been proposed to evaluate a DMU's performance in a two-stage production process, where the outputs of the first stage are taken as the inputs for the second stage (system inputs  $\diamond$  intermediate measures  $\diamond$  system outputs). Nevertheless, two-stage DEA models such as the network DEA approach, which are common in management (e.g., An et al. 2018; Sadri et al. 2022), are still rare in higher education performance evaluation (e.g., Tavares et al. 2021).

The determinants of efficiency are usually identified in a second-stage parametric regression of the DEA scores on variables in the environment of the DMUs that are expected to have an influence on efficiency (e.g., Salas-Velasco 2020b; Wolszczak-Derlacz and Parteka 2011). A rule of thumb that is commonly used is to use only environmental variables outside the control of the DMUs themselves in the second stage. However, this approach has limitations. On the one hand, the methodology is questionable in the presence of non-zero slacks, that is, input excesses and/or output

shortfalls (De Witte and Marques 2010). On the other hand, such studies are valid only if the separability condition between the input-output space of the first stage and the space of the external factors in the second stage holds. These factors are neither inputs nor outputs of the production process but may influence the performance of the units analyzed. In the situation where the separability condition does not hold, then a conditional DEA model might be an appropriate approach (Cazals et al. 2002; Daraio and Simar 2005). However, studies that investigate the issue of separability and apply a conditional nonparametric approach in higher education are relatively scarce to date (e.g., Agasisti et al. 2023; Mastromarco et al. 2022).

### 2.3 Concerns regarding the informative value of DEA

For educational policy purposes, DEA provides a wealth of managerial insights for the performance evaluation of institutions of higher education, such as universities. DEA benchmarks universities against the best performers in a given sample to provide efficiency scores, with a technical efficiency score equal to unity meaning that the institution is efficient. If the university's efficiency score is different than 1, the institution under evaluation is inefficient. Under output orientation, a score greater than 1 corresponds to the required production expansion to make the HEI efficient, keeping input levels fixed. Furthermore, DEA provides very valuable information for university managers by offering information on peers. A peer is a university against which the technically inefficient institution may be benchmarked to become efficient. When yielding efficiency scores, DEA has become a convenient tool for ranking universities (Bougnol and Dulá 2006). Assessing the performance of universities on efficiency-based rankings is a way to enhance the accountability and transparency of institutions. Several methods have also been developed to completely rank the efficient DMUs in DEA (e.g., Adler et al. 2002).

Nonetheless, the credibility of the DEA methodology as a benchmarking tool requires that there be a certain homogeneity in the sample of producers under evaluation. A fundamental assumption of DEA is that a set of DMUs should be homogenous in the sense that all DMUs are "alike" and therefore directly comparable (Dyson et al. 2001). In particular, they use similar resources to obtain similar products, have access to similar technology for converting inputs into outputs, face similar environmental conditions, and are similar in their scale of operations (Dyson et al. 2001). The non-homogeneity of DMUs significantly influences their efficiency scores, which also leads to an unfair ranking of DMUs (Kazemi et al. 2021). A major challenge in benchmarking and performance assessment of services is actually accounting for heterogeneity (Daraio et al. 2021).

Since the main objective of the current study was not to rank universities but to compare them with the aim of establishing best practices, we must ensure that homogeneity assumptions are met. By providing empirical results, our major contribution was to show that using a sample of producers or DMUs of widely varying scale or size can lead to inappropriate comparisons between DMUs. To identify suitable benchmarks that can be used by poorly performing HEIs as a basis for improvement, we should use "similar-in-scale" universities.



## 2.4 Performance evaluation in the presence of sample heterogeneity

The quantitative evaluations and comparisons should take into account the main features of the assessed units (Daraio et al. 2021). If there are differences between specific subsets of the full data set, it is more appropriate to apply DEA separately to each subset to get suitable peer groups for the inefficient DMUs (Charnes et al. 1981). For example, Johnes (2006) divided the sample of English HEIs, taking into account the Further and Higher Education Act 1992, which granted university status to former polytechnics. Applying DEA and the Charnes et al. (1981) method, Johnes (2006) tested for significant differences in the efficiency distributions of pre-1992 universities and post-1992 universities. However, in many real-life cases, the sample of DMUs may not consist of two or more naturally occurring subsets. In such situations, a useful statistical method to group the DMUs into homogenized subsets is cluster analysis. Although there are several clustering techniques, K-means clustering is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into  $k$  subsets (Steinley 2006). The K-means method produces exactly  $k$  different clusters or groups of the greatest possible distinction, implying a certain degree of homogeneity within clusters and heterogeneity between clusters (Steinley 2006). One of the first works in the DEA literature to propose K-means clustering in the presence of sample heterogeneity is due to Samoilenko and Osei-Bryson (2008). Other methods, such as fuzzy clustering algorithms, have been recently proposed in the DEA literature (e.g., Hajiagha et al. 2016; Kazemi et al. 2021). However, it is not the objective of this article to review all the available methods for partitioning data into a collection of clusters but to apply one of them, the extensively studied and implemented K-means partitioning clustering, to the objectives of our study.

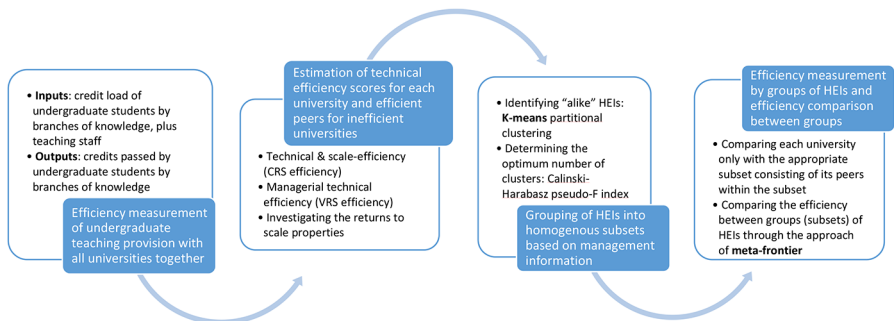
The current article takes the paper by Samoilenko and Osei-Bryson (2008) as a reference but presents some important differences. On the one hand, Samoilenko and Osei-Bryson (2008) used cluster analysis (K-means) to determine whether all DMUs in the sample (transition economies in their study) were similar in terms of the levels of the received inputs and the levels of the produced outputs. In our case, the inputs in the system are the credits enrolled and the outputs of the system are the credits passed, and the definition of “credit” is common not only to all Spanish universities but also across the European Higher Education Area, the so-called ECTS (European Credit Transfer System). Nonetheless, for realistic and reasonable benchmarking, we used cluster analysis to divide the sample of DMUs into groups based on scale homogeneity. The efficiency score of a DMU is not defined by the peer group but by an efficient subset of the same peer subset. Specifically, K-means analysis was performed in this research to “partition” Spain’s public university system into similar groups using university size indicators. On the other hand, Samoilenko and Osei-Bryson (2008) obtained several sets of clusters, but they finally chose the one with two subgroups of countries, the “leaders” cluster and the “majority” cluster, based on expert knowledge. Instead, this article used the so-called pseudo-F index of Calinski-Harabasz (1974) to determine the precise number of clusters. The advantage of our analysis is that it reduces subjectivity in grouping DMUs and concentrates on differences that are evident from the data rather than perceived differences.

### 3 Methods

This study contributed to the existing literature on performance evaluation and accountability in higher education by providing a DEA and clustering methodology for the measurement of university productive efficiency. The stages of our proposal for the technical efficiency measurement of universities in undergraduate teaching provision are presented in Fig. 1, but they can be extended to other areas of higher education (e.g., research evaluation, the evaluation of university departments, etc.). It is important to note that the focus of this approach is not to rank universities based on their relative efficiency but to provide a peer-to-peer comparison of inefficient and efficient HEIs. In the first stage of the evaluation, we must select a sample of producers and the inputs and outputs involved in the production process. Selecting inputs and outputs is a critical part of university performance analysis. In the next stage of the evaluation, we will apply the DEA methodology to obtain efficiency scores that allow an institution to be labeled as efficient or inefficient, as well as the efficient peers that inefficient institutions would have to emulate. Based on objective results and expert knowledge, we must decide whether or not the peer groups are effective. The improvement in efficiency may not be possible due to factors such as the significant difference in the university size of the compared institutions. In this latter case, we will proceed to the next stage of evaluation by partitioning the set of DMUs into homogeneous subsets. In the final stage, we will measure the efficiency of the HEIs in each cluster (“frontier separation” approach). Additionally, we may want to find out if there are differences in the relative efficiencies between groups and some of the factors that contribute to the differences in relative efficiencies.

#### 3.1 Technical efficiency measurement

In their seminal paper, Charnes et al. (1978) used the “Farrellian” efficiency concept in a linear programming model known as DEA CCR. The original CCR model aimed to build an empirical production frontier so that the DMUs that showed the best practices were considered efficient and were located on this empirical frontier, and the inefficient ones were situated below the border region (*efficiency is a rela-*



**Fig. 1** A multi-stage methodological framework for the study of the technical efficiency of universities in the provision of undergraduate teaching. *Source:* author’s elaboration

ive concept). The DEA CCR model assumed a production technology with constant returns to scale (CRS), meaning any proportional change in inputs should produce the same proportional change in output. However, to take into account that part of the inefficiency of DMUs might be due to their scale or size while still showing managerial efficiency, the BCC model by Banker et al. (1984) introduced a variation in the formulation of CCR to examine the variable returns to scale (VRS) nature of DEA. These traditional (radial) DEA models can be either input- or output-oriented. In the former, DEA defines the frontier by looking for the maximum possible proportional reduction in input usage, with production levels held constant. In the output-oriented case, DEA pursues the maximum proportional increase in production with input levels held fixed. The latter is more suitable for our objective of measuring university efficiency since the goal of Spanish public universities should be for undergraduate students to pass as many credits as possible out of those they enroll in, given their teaching staff. Also, labor is surely a quasi-fixed factor because a significant percentage of professors are permanent staff.

Let us consider a sample of  $n$  DMUs (HEIs in our analysis),  $DMU_j$  ( $j = 1, 2, \dots, n$ ), for which a common set of “ $m$ ” inputs,  $\{x_{ij}\}_{i=1}^m x \in \mathbb{R}_+^m$ , are converted into a common set of “ $s$ ” outputs,  $\{y_{rj}\}_{r=1}^s y \in \mathbb{R}_+^s$ .<sup>1</sup> The technology that models the conversion of inputs into outputs is represented by

$$T = \{(x, y) \mid x \text{ can produce } y\} \tag{1}$$

A constant returns-to-scale (CRS) DEA technology for output-oriented technical efficiency measurement involves the solution of the following linear programming (LP) problem, in which we label the DMU evaluated by the subscript  $k$  (Zhu 2014)

$$\begin{aligned} & \text{maximize } \varphi_k \\ & \text{subject to} \\ & \varphi_k y_{rk} - \sum_{j=1}^n \lambda_j y_{rj} \leq 0 \quad r = 1, \dots, s \\ & x_{ik} - \sum_{j=1}^n \lambda_j x_{ij} \geq 0 \quad i = 1, \dots, m \\ & \lambda_j \geq 0 \forall j = 1, \dots, n \end{aligned} \tag{2}$$

The above LP problem, the CCR model, should be solved  $n$  times – once for each DMU in the sample. The VRS efficiency model, which is usually referred to as the BCC model (Banker et al. 1984), is estimated by adding the convexity constraint  $\sum_{j=1}^n \lambda_j = 1$  in Eq. (2).

In Eq. (2),  $\varphi$  is the technical efficiency (TE) score. In the output-oriented formulation, efficiency scores,  $j=1, 2, \dots, n$ , are circumscribed between unity and infinity. If the optimal value of  $\varphi$  is equal to unity ( $\varphi^* = 1$ ), it implies that the DMU under assessment is on the frontier (i.e., it is efficient). A technically efficient DMU cannot produce any additional output from its existing input mix. Otherwise, if  $\varphi^* > 1$ , the DMU under evaluation is inefficient. In this latter case,  $[(\varphi^* - 1) 100]$  is the percent-

<sup>1</sup> For the units under evaluation, inputs and outputs must be greater than 0.

age increase in all outputs – the maximum possible radial expansion – that could be achieved by the technically inefficient DMU under analysis to become efficient with input quantities held constant.

### 3.2 Investigating the returns to scale properties

In the measurement of HEIs' performance, the DEA CCR model presented above provides a measure of the overall (global) performance of each university. An institution would be overall technically efficient if its CRS TE score was 1; otherwise, it would be overall technically inefficient (i.e., CRS TE score > 1). Furthermore, TE scores gotten from the CCR model can be decomposed into two components: one due to scale inefficiency and one due to managerial ("pure") technical inefficiency (i.e., VRS TE scores).<sup>2</sup> That is (Färe et al. 1985): technical & scale-efficiency (CRS efficiency) = scale-efficiency x technical (VRS) efficiency. This breakdown of the efficiency scores provides valuable managerial information. In the event that an institution is overall technically inefficient (i.e., CRS TE score > 1), the university manager may want to know if the inefficiency is solely due to the scale or size of the institution, and thus it is "too large" or "too small" to be considered scale efficient even when it presents managerial efficiency (i.e., VRS TE score = 1), or if the institution is both management inefficient (i.e., VRS TE score > 1) as a result of the institution's poor management practices and scale-inefficient as well.<sup>3</sup>

Scale efficiency states whether a DMU is operating at its "optimal size". Scale efficiency measures can be obtained by running both the CRS and VRS DEA models. For each DMU, the scale efficiency score is gotten by dividing the CRS TE score by the VRS TE score (Färe et al. 1985): scale efficiency (SE) = CRS TE score / VRS TE score. If SE = 1, then the DMU is scale-efficient. If there is a dissimilarity in the CRS and VRS TE scores for a particular DMU, then this shows that the DMU has scale inefficiency (Coelli et al. 2005). In our analysis (output-oriented TE scores), a public university would be scale-inefficient if SE > 1.<sup>4</sup>

Scale inefficiency arises due to either increasing returns to scale (IRS)<sup>5</sup> or decreasing returns to scale (DRS).<sup>6</sup> In order to test returns to scale (RTS) with CCR models, Banker et al. (2011) suggested: constant RTS if  $\sum \lambda_j^* = 1$ ; DRS if  $\sum \lambda_j^* > 1$ ; IRS if  $\sum \lambda_j^* < 1$ .<sup>7</sup> However, the method of testing a DMU's RTS nature using the sum of the CCR optimal lambda values may fail when the DEA model has alternate optima (Seiford and Zhu 1999). By adding in Eq. (2) the constraint  $\sum_{j=1}^n \lambda_j \leq 1$ , this leads to a non-increasing returns-to-scale (NIRS) technology. The nature of the scale inefficiencies for a specific DMU can be ascertained by examining whether the NIRS TE

<sup>2</sup> Since the technical efficiency scores gotten from the CCR model will be affected by scale effects, if they exist, the "pure" label refers to the fact that the variable returns-to-scale specification provides technical efficiency scores that are free of scale efficiency effects.

<sup>3</sup> If the CRS TE score is 1, the university is both scale-efficient and management-efficient.

<sup>4</sup> For (globally) inefficient DMUs, CRS TE scores are always greater than VRS TE scores.

<sup>5</sup> For example, if output more than doubles when all inputs are doubled.

<sup>6</sup> For instance, if the output is less than doubled when all inputs are doubled.

<sup>7</sup> Output-oriented returns to scale (RTS).

score is equal to the VRS TE score. If they are equal, then DRS applies; otherwise, scale inefficiency due to IRS exists for that DMU (Seiford and Zhu 1999).<sup>8</sup> If scale inefficiency does not exist, TE scores are coincident and equal to 1 under CRS, VRS, and NIRS technologies. In the higher education context, however, an institution does not have the flexibility to change its scale of operation as a private sector firm would. The adjustment of public universities to their optimal size could be gradual due to quasi-fixed inputs (e.g., teaching staff) and the regulatory environment in which they operate.

### 3.3 Cluster analysis for grouping homogenous DMUs

Once the efficiency scores have been obtained and we have investigated the RTS nature of the institutions, if both the results obtained and the expert knowledge point to inconsistent comparisons between HEIs, we must move on to the next stage suggested in Fig. 1. We need to recognize which DMUs belong to each group to achieve as much group homogeneity as possible. In particular, we should divide our sample into clusters within which HEIs are homogeneous in size. As we have advanced, cluster analysis is a statistical technique that identifies sets of observations that behave alike or exhibit comparable characteristics (look-alike groups). In particular, cluster analysis is convenient for merging observations into groups such that: (i) each group is homogeneous with respect to certain features (observations of each group are similar to each other), and (ii) each group should be different from the other groups with respect to these characteristics. The method developed by MacQueen (1967), recognized as “K-means”, offers partitions of observations into groups that are sensibly efficient in the sense of intra-class variance. Generally, this technique produces exactly  $k$  different clusters, maximizing the similarities within the clusters and minimizing the similarities between clusters. As the number of clusters is not known a priori, one needs a rule to “break” the initial sample of observations. Milligan and Cooper (1985) evaluated 30 stopping rules, singling out the Calinski–Harabasz index and the Duda–Hart index as two of the best. Stata® 17 statistical software provides both indices, but the latter is allowed only for hierarchical cluster analyses. Thus, for choosing  $k$ , we used in this study the Calinski–Harabasz pseudo-F index as the stopping rule. The optimal grouping is the one with the highest value for the index (Calinski and Harabasz 1974).

### 3.4 The meta-frontier framework for comparing the efficiency of subsets of DMUs

In the last stage of university performance evaluation, we should measure the technical efficiency of HEIs within each cluster by applying the methods suggested above. For each inefficient university, DEA will now show its efficient peers of similar scale to emulate. Next, with the main objective of providing comparable averaged technical efficiency scores across different clusters, a meta-frontier production function model is proposed and applied in the analysis of the technical efficiencies of HEIs. In fact, meta-frontier is a method tailored to deal with heterogeneity when entities

<sup>8</sup> In this latter case (i.e., IRS), the NIRS TE score would be equal to the CRS TE score.

are partitioned into different groups. Basically, a different production possibility set is reconstructed for each group, while an “overall” set is defined as the envelopment of the group-specific counterparts. In particular, efficiencies measured relative to a common meta-frontier can be decomposed into two components (O’Donnell et al. 2008): (a) a component that calculates the distance from an input-output point to the group frontier (the usual measure of technical efficiency); and (b) a component that calculates the distance between the group frontier and the meta-frontier. In relation to component a), the distance from an input-output point to its group frontier is a measure of the technical efficiency of that DMU, denoting how well a DMU is performing in comparison with the rest of the DMUs in its own group (i.e., individual efficiency scores in each subset). To obtain component b), we must run a DEA model on the pooled sample of DMUs. Then, the gap between the group frontier and the meta-frontier is measured by getting the so-called “meta-technology ratio” (MTR).

#### 4 DEA models performed on the pooled sample of HEIs

This section presents the results of our efficiency analysis using the full sample of traditional (face-to-face) Spanish public universities. The principal data source for the assessment comes from the Spanish Association of University Rectors (*Conferencia de Rectores de las Universidades Españolas*, CRUE). As we anticipated, in selecting input and output variables, our choice was guided both by the objectives pursued in this study and variables used in other works for the Spanish case before the reform of the curricula (e.g., Salas-Velasco 2020a). Specifically, input variables included the total number of credits enrolled by undergraduate students at each university in the 2015/2016 academic year (the total students’ academic load) plus the total number of FTE teaching staff (the academic labor input). In the Spanish education system, college students take courses over two semesters. Typically, each course has 6 credits (about 4 class hours per week). The average academic course load is 60 credits (five courses per semester). Since we focus on students’ academic performance, we did not include non-academic personnel since they do not participate in the process of converting inputs into outputs that we are describing. Nevertheless, we distinguished in the analysis between degrees in social sciences and humanities, on the one hand, and degrees in hard sciences, engineering, and health sciences, on the other hand. Both the difficulty of the studies and the type of students admitted to them differ between those two broad fields of knowledge.<sup>9</sup> Regarding the output variables of the instructional component of higher education, we used for each institution the total number of credits passed by its undergraduate students in the 2015/2016 academic year by broad fields of study. By using the credits passed over the credits enrolled, referring to the same academic year, we can more accurately assess the efficiency of teaching than using enrolled students as inputs and graduates as outputs. We also avoid the problem of having full-time and part-time students.

---

<sup>9</sup> The disciplinary field is also a dimension of heterogeneity, but given our degrees-of-freedom constraints, the inclusion in the DEA models of more disaggregation of fields of study would not be wise in any case.

Table A2 in the Appendix shows the descriptive statistics of input and output variables for the pooled sample. We finally worked with 46 HEIs, excluding the Universitat Politècnica de Catalunya because it only offers engineering degrees. One of the central assumptions of DEA is that all DMUs in the sample are functionally alike in the sense that all DMUs obtain the same number and type of inputs and outputs (Dyson et al. 2001). Cook et al. (2015) already argued that if a DMU has chosen not to produce a certain output, and thus does not put the resources in place to do so, then it would be inappropriate to artificially assign that DMU a zero value or some average value for the nonexistent factor. Full analysis results are offered in Table A3 in the Appendix. DEA models were performed using the DEA-Frontier© software (Zhu 2014). In addition to efficiency scores, Table A3 showed a list of the peers that each inefficient university should ideally emulate in order to become efficient (benchmarks whose practices should try to emulate inefficient DMUs). Indeed, the frequency with which an efficient DMU appears as a peer is of interest for benchmarking. “A low frequency suggests that it has an extreme characteristic (for example size) which makes it an unsuitable peer to emulate” (Johnes 2006, p. 280). As can be seen in Appendix Table A3, Universidad de León and Universidad Pablo de Olavide are both efficient (CRS TE score = 1) but do not appear as peers for any other university. Such universities may be deemed efficient because of their “extreme operating or size characteristics” (Athanassopoulos and Shale 1997, p. 129). In fact, Universidad de León is a young and small university with 10,000 undergraduate students, and Universidad Pablo de Olavide is also a young university with 8,800 undergraduate students (Appendix Table A1). For these institutions, we would not be able to determine whether the relative efficiency is indeed caused by their efficiency, or whether a relatively efficient status was awarded because they have a size for which no other comparison unit could be found.

The DEA results in Appendix Table A3 also show that some universities are inefficient under CRS and efficient under VRS (i.e., they are only scale-inefficient DMUs), but their comparison peers (benchmarks) may not be appropriate. For example, for the Universidad Politécnica de Cartagena, with around 4,500 undergraduate students, the analysis shows that it was operating under IRS (it is “too small”), and its main comparison DMU (efficient peer) was the Universitat Politècnica de València, which has around 20,000 students. When a DMU is found to be operating in the region of increasing returns to scale, an inferred conclusion is that it is smaller than its optimal size. Should the Universidad Politécnica de Cartagena grow that large to reduce its scale inefficiency? At the opposite pole, we find the Universidad de Granada, which has around 42,500 undergraduate students. The DEA analysis in Table A3 shows that it was operating under DRS (it is “too large”), and among its main comparison peers was the Universitat Pompeu Fabra, which has around 10,000 undergraduate students. A DMU functioning in the region of diminishing returns to scale is judged to be too large. Should the Universidad de Granada reduce its size that much to improve its efficiency? Ultimately, these results are telling us that the population of producers in the analysis is heterogeneous in the scale of their operations, and it is necessary to find “similar-in-scale” groups to obtain practical results. In short, the DMUs in the peer group should be similar enough in size for reasonable comparison results. We

propose to divide the sample of HEIs into groups by university size using cluster analysis.

## 5 Cluster analysis for identifying similar-in-scale HEIs

This section aimed to “partition” the set of DMUs (i.e., HEIs) into similar-in-scale producers using both indicators of institutional dimension: total enrollment of undergraduate students and total academic staff (full-time equivalent). K-means clustering is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of  $k$  groups (i.e.,  $k$  clusters). As a preliminary step, both indicators of university size were normalized to get scale-invariant results for the required distance measurement. Specifically, for each university, each indicator was subtracted from its minimum value (numerator), which was divided by the difference between the maximum value and minimum value (denominator) to obtain values on the interval  $[0, 1]$  (Clermont and Schaefer 2019; Milligan and Cooper 1985). The cluster analysis was performed using Stata® 17 (StataCorp 2021). To find a satisfactory clustering result, we run the clustering algorithm with different values of  $k$  ( $=2, 3, 4, 5$ ). We used the Calinski–Harabasz pseudo-F index as the stopping rule. The results showed that: with  $k=2$ ,  $F=82.14$ ; with  $k=3$ ,  $F=141.89$ ; with  $k=4$ ,  $F=119.62$ ; and with  $k=5$ ,  $F=116.42$ . Larger values indicate more distinct clustering. Thus, three clusters should be taken into account. However, one cluster had only six universities (the first large universities in Appendix Table A1). A rule of thumb establishes that the number of DMUs should be at least twice the number of inputs and outputs considered (Golany and Roll 1989). Hence, our 3-input, 2-output DEA models would recommend using at least 10 DMUs in each cluster. Therefore, we considered together the DMUs of the first two clusters that we labeled as “medium- and large-sized universities” (DMUs in Appendix Table A5), while DMUs in the third cluster were labeled as “small-sized universities” (DMUs in Appendix Table A6).

## 6 Teaching efficiency by university size: making peer groups effective

This section presents the empirical results. Efficiency scores were computed for each institution measured against its own group’s frontier, that is, medium- and large-sized public universities and small-sized public universities. Table A4 in the Appendix displays the summary statistics of input and output variables used in the different output-oriented DEA models. The statistics suggest heterogeneity across groups (clusters) in terms of their input and output compositions, justifying our discussion of the results according to the university size. Table A5 and Table A6 in the Appendix show full details of estimates of different efficiency measures, the nature of RTS, and a list of peers. Within each group, for each inefficient university, DEA shows its peers of similar size to emulate, getting reasonable comparison results according to our expert knowledge of the Spanish higher education system. For example, if we go back to the examples taken in Sect. 4, Universidad de Granada obtains a CRS TE score greater than 1 and exhibits scale disadvantages



in undergraduate education production (operates under DRS), but its peers are now two medium-sized institutions, in particular, Universidad Autónoma de Madrid and Universidad Rey Juan Carlos (Appendix Table A5). These last two universities are in fact the main benchmarks for the HEIs in this subset (Table 1). Nonetheless, we should note that the term “scale”, which is typical of the microeconomic analysis of the firm, is very elusive in the context of our performance evaluation as we are dealing with the production of an intangible (knowledge). In seeking sources of best operating practices for teaching efficiency, the Universidad de Granada, for instance, could eliminate certain bachelor’s degrees (or reduce the number of places) for which student academic performance is very low, taking its peers into account.

In relation to small universities, the most frequent peers are Universitat Pompeu Fabra (Barcelona), Universidad Carlos III de Madrid, and Universitat de Lleida. Going back to the discussion in Sect. 4, Universidad Politécnica de Cartagena is efficient when compared to other universities of a similar size (Appendix Table A6), and Universidad de León serves as a benchmark for 11 other small universities (in addition to itself) (Table 1). These results are much more realistic than those obtained by using the entire population of HEIs in the DEA analysis.

Table 2 presents the descriptive statistics, by university size, of the estimates of the efficiency scores shown in Table A5 and Table A6 in the Appendix. The average efficiency of the group of medium- and large-sized public universities is 1.0481 under CRS (about 96%) and 1.0171 under VRS (about 98%).<sup>10</sup> As expected, the VRS approach provides a higher number of technically efficient universities when compared to the CRS approach. Scale efficiency is also high, with an average of 1.0308 (about 97%). On the other hand,

<sup>10</sup> To obtain the efficiency as a percentage (shown in parentheses), we must calculate the inverse of the

**Table 1** Frequency of efficient universities being used as peers based on output-oriented CCR undergraduate teaching efficiency. Traditional Spanish public universities, 2015/2016

Medium- and large-sized universities		Small-sized universities	
Benchmarks	Frequency	Benchmarks	Frequency
Universidad Autónoma de Madrid	17	Universidad Carlos III de Madrid	13
Universidad Rey Juan Carlos	14	Universitat de Lleida	13
Universitat Politècnica de València	11	Universitat Pompeu Fabra	13
Universidad Politécnica de Madrid	9	Universidad de León	12
Universitat de Barcelona	1	Universidad Politécnica de Cartagena	7
		Universidad de Jaén	3
		Universidad Pablo de Olavide	2

Source: author’s elaboration

efficiency scores of each DMU, and then we average these scores.

**Table 2** The efficiency of undergraduate teaching provision. Traditional Spanish public universities. Separate analyses by universities' scale of operation, 2015/2016

	CCR efficiency (CRS TE scores) <sup>a)</sup>	BCC efficiency (VRS TE scores) <sup>b)</sup>	Scale efficiency <sup>c)</sup>
<b>Small-sized universities</b>			
Mean	1.0554	1.0279	1.0270
Standard deviation	0.0548	0.0476	0.0358
Minimum	1	1	1
Maximum	1.1594	1.1585	1.1196
Observations	25	25	25
Number of efficient units	7	14	7
<b>Medium- and large-sized universities</b>			
Mean	1.0481	1.0171	1.0308
Standard deviation	0.0425	0.0256	0.0407
Minimum	1	1	1
Maximum	1.1468	1.0810	1.1468
Observations	21	21	21
Number of efficient units	5	13	5

<sup>a)</sup> Output-oriented technical efficiency measures under constant returns to scale (CRS).

<sup>b)</sup> Output-oriented technical efficiency measures under variable returns to scale (VRS).

<sup>c)</sup> Scale efficiency is calculated as the ratio of the CRS efficiencies to the VRS efficiencies.

Source: author's calculations

the mean efficiency of the group of small-sized public universities is 1.0554 under CRS (about 95%) and 1.0279 under VRS (about 97%). Scale efficiency is also high, with an average of 1.0270 (about 97%).

Is student academic performance – what percentage of the enrolled credits a student is able to pass in one academic year – higher in medium- and large-sized Spanish public universities than in small ones? The mean teaching efficiency of medium- and large-sized institutions is greater than that of small ones, according to Table 2. However, as a common rule, efficiency levels measured relative to one frontier (e.g., the medium- and large-sized university frontier) cannot be compared with efficiency levels measured relative to another frontier (the small-sized university frontier). HEIs in each group may be facing different production opportunities attributed to the scale of operation, including the environment in which the production process takes place. We will use the non-parametric meta-frontier framework, which has been proposed in the literature to compare the technical efficiencies of firms that may be classified into different groups (e.g., O'Donnell et al. 2008), for making efficiency evaluations across groups (subsets) of Spanish public institutions.

## 7 Non-parametric meta-frontier framework for efficiency evaluation

### 7.1 Traditional meta-frontier approach

The meta-frontier framework was described in Sect. 3.4. The results of the traditional non-parametric meta-frontier approach to account for the heterogeneity of Spanish HEIs (i.e., medium- and large-sized public universities and small-sized public universities) are presented in Table 3 (figures not in italics). Efficiency measurement was performed in the

whole population of public universities (second column) as well as in the sub-populations of universities (third column). The fourth (last) column shows the meta-technology ratio (MTR), which essentially measures how close a group frontier is to the meta-frontier.

The technical efficiency of a group-specific frontier is evaluated using the output-oriented CCR model as the first step.<sup>11</sup> We run an output-oriented CCR model on the pooled sample of medium-large and small institutions as a second step. Importantly, the meta-frontier envelops the two group frontiers.<sup>12</sup> Then, we measured the gap between the group frontier and the meta-frontier by getting the so-called MTRs. For each DMU, we divided its efficiency score measured against the pooled frontier (common meta-frontier) by the efficiency score measured against its own group's frontier.<sup>13</sup>

Table 3 (second column) shows that the average efficiency of the Spanish public system of higher education for undergraduate teaching is about 93%: about 95% in the case of medium- and large-sized institutions and about 92% in the case of small ones. There is a difference of about 3% in overall efficiency, and of this percentage, 2.5% represents an efficiency advantage due to differences in the scale of operation, including environmental factors (e.g., medium- and large-sized universities are strongly research-oriented). The Mann-Whitney test is carried out to verify the validity of our result. The null hypothesis is rejected. There is a statistically significant difference between the meta-technology ratios of the two groups.

## 7.2 A bootstrapped non-parametric meta-frontier approach

Table 3 also presents bootstrapped DEA efficiency scores (*figures in italics*) to obtain better decision information about efficiency. From an applied point of view, bias-corrected DEA scores enable us to refine DEA scores from an “artificial efficiency” just inducted by estimation problems rather than management practices (Curi et al. 2011). Another cause for the potential bias of the DEA efficiency estimators is associated with the non-measurement of random error and thus the improper meaning of overall deviation from the frontier as inefficiency (Keramidou et al. 2011). To calculate the bootstrapped DEA estimates, we used Algorithm #2, proposed by Simar and Wilson (2007), which has been implemented in Stata® 17 software by Badunenko and Tauchmann (2017) with *simarwilson*. The results show that the efficiency scores without correction of bias that are obtained from the LP problem formulated in Eq. (2) are upward-biased, or, if we want, inefficiency

<sup>11</sup> These estimates were already shown in Table 2, although in the meta-frontier approach, the efficiency scores vary between 0 and 1. To do this, we must calculate the inverse of the efficiency scores of each DMU and then average these scores.

<sup>12</sup> Spanish HEIs, as an industry, operate under CRS in each group. We carried out returns-to-scale testing (Badunenko and Mozharovskyi 2016) and we were not able to reject the null hypothesis at the 5% level that the global technology is CRS. Therefore, we used CRS TE scores (output-oriented). We also carried out returns-to-scale testing (Badunenko and Mozharovskyi 2016) on the pooled sample of institutions (46 universities), and we were not able to reject the null hypothesis that the global technology is CRS either. Tests on the type of returns to scale exhibited globally by the technology have been carried out using the *npstrts* command in Stata® 17 software, as proposed by Badunenko and Mozharovskyi (2016).

<sup>13</sup> The traditional meta-frontier approach in O'Donnell et al. (2008) obtains the technology gap ratio by comparing the technical efficiencies under two kinds of frontiers in radial DEA models (i.e., the meta-frontier and the group frontier).

**Table 3** Comparison of teaching mean efficiency between medium-large and small Spanish public universities, 2015/2016

	Pooled universities (common meta-frontier)	Small-, and medium- and large-sized universities separately (group frontier) <sup>a)</sup>	Distance between the group frontier and the meta-frontier
Total DMUs assessed (=46 universities)			
Average efficiency	0.929		
Average efficiency [ <i>bootstrapping</i> ]	<i>0.890</i>		
Small-sized universities (=25 universities)			
Average efficiency	0.915	0.950	0.963
Average efficiency [ <i>bootstrapping</i> ]	<i>0.875</i>	<i>0.916</i>	<i>0.955</i>
Medium- and large-sized universities (=21 universities)			
Average efficiency	0.945	0.956	0.989
Average efficiency [ <i>bootstrapping</i> ]	<i>0.907</i>	<i>0.923</i>	<i>0.983</i>
Difference in efficiency	3.0%	0.6%	2.5% <sup>b)</sup>
Difference in efficiency [ <i>bootstrapping</i> ]	<i>3.2%</i>	<i>0.7%</i>	<i>2.8%</i> <sup>c)</sup>
			The efficiency advantage of medium- and large-sized universities is statistically significant at the 5% significance level

<sup>a)</sup> As a general rule, efficiency levels measured relative to one frontier (e.g., small-sized universities) cannot be compared with efficiency levels measured relative to another frontier (the medium- and large-sized universities frontier) (O'Donnell et al. 2008).

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

H<sub>0</sub>: means do not differ between groups

<sup>b)</sup> z=2.712 p-value<0.01

<sup>c)</sup> z=2.856 p-value<0.01

Figures not in italics: output-oriented technical efficiency measures under constant returns to scale (CRS).

Figures in italics: bias-corrected output-oriented efficiency scores under constant returns to scale (CRS). Simar and Wilson (2007) (Algorithm # 2).

Source: author's calculations

is estimated downwards.<sup>14</sup> The average efficiency is now 89% for undergraduate teaching (second column); the average efficiency is about 91% for medium- and large-sized HEIs and about 88% for small ones. However, the differences in average efficiency between the two groups of universities are similar to those obtained previously, which points to an efficiency measurement over a meta-frontier and group frontiers shifted upward in parallel.<sup>15</sup>

### 7.3 Explaining the distance between the meta-frontier and group-specific frontiers

As we have shown in the last column of Table 3, the average meta-technology ratios (MTRs) for the two groups of universities are statistically different from each other. The highest average MTR is for medium- and large-sized Spanish public universities, which means that these institutions are closer to the meta-frontier (MF) than their small-sized counterparts. Note that a higher (lower) MTR value implies a smaller (larger) technology gap between the individual frontier and the MF, which in the context of undergraduate teaching provision means greater (lesser) technical efficiency once internal factors affecting efficiency within each group have been discounted. What explains the average efficiency advantage of large and medium universities over small ones? In the business context, this may be related to the fact that small firms have less access to technology (e.g., Battese et al. 2004; Moreira and Bravo-Ureta 2010). The meta-frontier analysis has also been used to explain the disparity between efficiency-driven and innovation-driven countries and to examine the technology gaps between the two groups of countries (Mitropoulos and Mitropoulos 2023). However, the HEIs in our sample are all public institutions that use similar technology to convert inputs into teaching outputs (i.e., traditional classroom instruction). But universities may differ in other features that shape the educational production process, which might explain differences in student academic productivity between both groups. Thus, we must look for some specificity that allows us to explain the heterogeneity between the two groups of public HEIs and, ultimately, the differences in the academic productivity of undergraduate students. That variable, for our sample of undergraduate education producers, was the location on the campus of a school of medicine.

In Table 4, individual bootstrapped MTRs (dependent variable) are regressed on a binary explanatory variable that takes the value 1 if the HEI has a school of medicine and the value 0 otherwise. The estimated coefficient is positive and statistically significant, which means that institutions with a school of medicine are closer to the MF. The proportion of large and medium universities in our sample with medical schools is 85.7%, while the proportion of small universities with medical schools is 48%. A two-sample test of proportions confirms that the percentage of the former is significantly higher than the percentage of the latter at the 5% significance level. In fact, medium- and large-sized

<sup>14</sup> According to Johnes (2006), “the production frontier estimated by DEA may not in fact be the true frontier which could be achieved if the sectors were truly efficient; it is merely the observed production frontier for the sector. If this is the case, then overall levels of efficiency are overestimated by DEA” (p. 279). In other words, DEA can tell us how well we are doing compared to our peers but not compared to a “theoretical maximum”.

<sup>15</sup> Scale efficiency was practically equal to 1, calculated as the ratio of the bias-corrected output-oriented CRS efficiencies to the bias-corrected output-oriented VRS efficiencies.

**Table 4** How close to the meta-frontier are Spanish public universities in their production of undergraduate teaching? A bootstrapped Tobit regression model

	Ob- served Coef.	Boot- strap Std. Err.	z	$P >  z $	[95% Conf. Interval]	
The university has a school of medicine (= 1 yes)	0.0344	0.0148	2.32	0.021	0.0053	0.0635
Constant	0.9483	0.0129	73.54	<0.001	0.9230	0.9735

Number of obs. = 46 (right-censored=6)

Replications=2,000

Wald chi2(1)=5.36

Prob. > chi2=0.0206

Source: author's estimates using Stata® 17

Spanish public universities are the ones that have traditionally offered a medical degree. Many small-sized Spanish public universities were born in the 1980 and 1990 s as a result of segregation from the traditional large universities, but they do not offer a degree in Medicine.<sup>16</sup> It is worth noting that Spanish universities that have medical schools attract the best high school graduates to health sciences degrees such as Medicine and Nursing, which translates into greater academic productivity for students (Hernández-Armenteros and Pérez-García 2017).

## 8 Discussion

Policymakers and university managers need to know how well their institutions are functioning. In this regard, the use of performance indicators for evaluating the activities of universities has a long tradition in educational evaluation (e.g., Ball and Halwachi 1987). In practice, however, the DEA has been one of the most widely used methodologies for measuring the efficiency of universities as a performance indicator in higher education. Given the importance of bachelor's degree programs in Spain (called *grados*), this article focused on the efficiency measurement of undergraduate teaching provision, that is, whether Spanish HEIs are maximizing their teaching outputs. The latter are frequently defined in terms of examination performance using, for instance, students' grade point average (GPA) scores (e.g., Jayanthi et al. 2014). Nevertheless, this research focused on assessing whether Spanish public universities are maximizing their teaching outputs, measured by the number of credits passed by their students, from their educational inputs (measured by the number of credits enrolled by their students and faculty).

By using outputs and inputs at the institutional level, we can identify technically efficient institutions that may serve as benchmarks in the higher education sector. However, a major challenge in benchmarking and performance assessment is accounting for DMUs' heterogeneity. Potentially, the non-homogeneous DMU issue could be handled by break-

<sup>16</sup> For an international reader, we must clarify that the degree in Medicine in Spain is part of undergraduate degree programs, although with a duration of six years instead of four years, which is the usual time for a bachelor's degree.

ing the set of DMUs into multiple groups and then doing a separate DEA analysis for each group. In this article, we propose to divide the sample of DMUs (i.e., universities) into groups based on scale homogeneity since Spanish public universities range in size from “too small” to “too large”. The cluster analysis methodology was proposed and applied to this end. In this way, a DMU is evaluated against only “true” peers. Within each group, for each inefficient university, DEA showed its efficient peers of similar scale to emulate, showing reasonable comparison results. Specifically, within the group of medium- and large-sized universities, Universidad Autónoma de Madrid, Universidad Rey Juan Carlos, and Universitat Politècnica de València are the main benchmarks for inefficient universities in this group. Within the small-sized universities, Universidad Carlos III de Madrid, Universitat de Lleida, and Universitat Pompeu Fabra are the main benchmarks for inefficient universities in this group.

What are the managerial implications of these results? For instance, what can inefficient universities learn from efficient ones? Inefficient universities could eliminate certain bachelor’s degrees (or reduce the number of places) for which student performance is very low, taking their peers into account. One limitation of this study, however, is that the information about degrees was very aggregated. We took into account social sciences and humanities degrees, on the one hand, and hard sciences, engineering, and health sciences degrees, on the other. Ideally, with the information available, the current exercise should be carried out by specific degrees (e.g., Economics, Political Science, Nursing, etc.). This article, nonetheless, has important policy implications since policymakers can judge more fairly the overall efficiency of the Spanish higher education system. For example, Universidad Carlos III de Madrid was judged inefficient in terms of teaching, both scale-inefficient and management-inefficient, in the DEA analysis with all universities together (Appendix Table A3). However, this university is considered efficient when it is evaluated with similar-in-size institutions, and it is in fact a benchmark for the inefficient ones in its group (Appendix Table A6 and Table 1).

Assessing the performance of universities on efficiency-based rankings is a tool to enhance the accountability and transparency of public institutions as well. However, pooling institutions of very different sizes can lead to unfair ranks among HEIs. The latter is important since students tend to rate reputation and position in rankings as crucial determinants in their choice of institution (Hazelkorn 2015). Furthermore, the results from the meta-frontier framework carried out in this article are relevant for policymakers to learn about why certain characteristics are likely to exert an impact on the way in which HEIs are able to transform their inputs into outputs.

## 9 Conclusion

Policymakers, as well as the media, devote considerable attention to higher education outcomes. Performance evaluation of traditional university goals of teaching and research has indeed been a frequent topic in applied work over the past decades. Accordingly, this paper aimed to assess the performance of Spanish public universities with up-to-date data after the Bologna reform. Performance evaluation focused on the measurement of the technical efficiency of undergraduate teaching provision. Technical efficiency was measured using the DEA approach. DEA provides an avenue to explore the efficiency of

converting multiple inputs into multiple outputs, and it has been widely used in the field of university performance evaluation.

Nonetheless, this article showed that immediate benchmarks generated by the DEA method when assessing how efficiently educational inputs are being used by traditional Spanish public universities to maximize their teaching outputs may be of little informative value when pooling HEIs of very different sizes. While this issue challenges the homogeneity assumption of DMUs, DEA studies have paid relatively little attention to university size when assessing the performance of HEIs. In order to fill this research gap, this article proposes novel, effective methods for evaluating university performance and identifying practical benchmarks for improving the operations of inefficient performers. Specifically, DEA and cluster analysis are applied for the evaluation of the performance of traditional Spanish public universities.

Based on the methodology proposed in this article, the undergraduate teaching production processes of medium-large- and small-sized universities are evaluated separately. In particular, we computed DMUs' efficiency measured against their own group's frontier. Within each group, for each inefficient university, DEA showed its peers of similar scale to emulate, showing reasonable comparison results. In addition, for correct comparison of inter-group average efficiencies, we used the non-parametric meta-frontier framework, which has been proposed in the literature, to compare the technical efficiencies of entities that may be classified into different groups. When carrying out an inter-group analysis using the meta-frontier framework, our results showed that there is an efficiency advantage for medium- and large-sized universities over small ones in providing undergraduate teaching. A bootstrapped, non-parametric meta-frontier approach also verified this latter result. One of the factors that could explain the differences in relative efficiency between both groups of universities is the location of a school of medicine on the campus. Spanish universities that have medical schools, which are mainly large and strongly research-oriented, attract the best high school graduates to health sciences degrees, which translates into greater undergraduate students' academic performance.

One limitation of this study, however, is that the information about degrees was very aggregated. Future research might focus on measuring the technical efficiency of providing specific degrees and examining the main consequences of inefficiencies. For instance, poor student performance will translate into an "excessive" time (more than four years) to produce four-year bachelor's degrees.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12351-024-00816-x>.

**Funding** Funding for open access charge: Universidad de Granada / CBUA.

## Declarations

**Conflict of interest** The author declares that he has no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line



to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abbott M, Doucouliagos C (2003) The efficiency of Australian universities: a data envelopment analysis. *Econ Educ Rev* 22(1):89–97
- Adler N, Friedman L, Sinuany-Stern Z (2002) Review of ranking methods in the data envelopment analysis context. *Eur J Oper Res* 140(2):249–265
- Agasisti T, Dal Bianco A (2009) Reforming the university sector: effects on teaching efficiency—evidence from Italy. *High Educ* 57(4):477–498
- Agasisti T, Pérez-Esparrells C (2010) Comparing efficiency in a cross-country perspective: the case of Italian and Spanish State universities. *High Educ* 59(1):85–103
- Agasisti T, Egorov A, Serebrennikov P (2023) Universities' efficiency and the socioeconomic characteristics of their environment—evidence from an empirical analysis. *Socio-Economic Plann Sci* 85:101445
- Ahn T, Charnes A, Cooper WW (1988) Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-economic Plann Sci* 22(6):259–269
- Altbach PG (2015) What counts for academic productivity in research universities? *Int High Educ* 79:6–7
- An Q, Meng F, Ang S, Chen X (2018) A new approach for fair efficiency decomposition in two-stage structure system. *Oper Res Int Journal* 18:257–272
- Athanassopoulos AD, Shale E (1997) Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. *Educ Econ* 5(2):117–134
- Avkiran NK (2001) Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-economic Plann Sci* 35(1):57–80
- Badunenko O, Mozharovskyi P (2016) Nonparametric frontier analysis using Stata. *Stata J* 16(3):550–589
- Badunenko O, Tauchmann H (2017) 'SIMARWILSON': Module to perform Simar & Wilson (2007) efficiency analysis. Statistical Software Components. Retrieved from <https://ideas.repec.org/c/boc/bocode/s458156.html>
- Ball R, Halwachi J (1987) Performance indicators in higher education. *High Educ* 16(4):393–405
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manage Sci* 30:1078–1092
- Banker RD, Cooper WW, Seiford LM, Zhu J (2011) Returns to scale in DEA. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.), *Handbook on data envelopment analysis* (pp. 41–70). Springer. [https://doi.org/10.1007/978-1-4419-6151-8\\_2](https://doi.org/10.1007/978-1-4419-6151-8_2)
- Battese GE, Rao DS, O'Donnell CJ (2004) A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *J Prod Anal* 21(1):91–103
- Berbegal-Mirabent J, Lafuente E, Solé F (2013) The pursuit of knowledge transfer activities: an efficiency analysis of Spanish universities. *J Bus Res* 66(10):2051–2059
- Bougnol ML, Dulá JH (2006) Validating DEA as a ranking tool: an application of DEA to assess performance in higher education. *Ann Oper Res* 145(1):339–365
- Breu TM, Raab RL (1994) Efficiency and perceived quality of the nation's top 25 National Universities and National Liberal arts Colleges: an application of data envelopment analysis to higher education. *Socio-Economic Plann Sci* 28:33–45
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat* 3:1–27
- Carrington R, Coelli TIM, Rao DP (2005) The performance of Australian universities: conceptual issues and preliminary results. *Economic Papers: A Journal of Applied Economics and Policy* 24(2):145–163
- Casanova JR, Cervero A, Núñez JC, Almeida LS, Bernardo A (2018) Factors that determine the persistence and dropout of university students. *Psicothema* 30(4):408–414
- Cazals C, Florens JP, Simar L (2002) Nonparametric frontier estimation: a robust approach. *J Econ* 106(1):1–25
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2:429–444

- Charnes A, Cooper WW, Rhodes E (1981) Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Manage Sci* 27(6):668–697
- Clermont M, Schaefer J (2019) Identification of outliers in data envelopment analysis. *Schmalenbach Bus Rev* 71(4):475–496
- Coelli TJ (1996) Assessing the performance of Australian universities using data envelopment analysis. University of New England, Department of Econometrics
- Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer
- Cohn E, Rhine S, Santos M (1989) Institutions of higher education as multi-product firms: economies of scale and scope. *Rev Econ Stat* 71(2):284–290
- Compagnucci L, Spigarelli F (2020) The third mission of the university: a systematic literature review on potentials and constraints. *Technol Forecast Soc Chang* 161:120284
- Cook WD, Harrison J, Imanirad R, Rouse P, Zhu J (2015) Data envelopment analysis with non-homogeneous DMUs. In: Zhu J (ed) *Data envelopment analysis: a handbook of models and methods*. Springer, pp 309–340
- Curi C, Gitto S, Mancuso P (2011) New evidence on the efficiency of Italian airports: a bootstrapped DEA analysis. *Socio-Economic Plann Sci* 45(2):84–93
- Daraio C (2020) Nonparametric methods and higher education. In: Teixeira PN, Shin JC (eds) *The international encyclopedia of higher education systems and institutions*. Springer, pp 2109–2114
- Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *J Prod Anal* 24(1):93–121
- Daraio C, Simar L, Wilson PW (2021) Quality as a latent heterogeneity factor in the efficiency of universities. *Econ Model* 99:105485
- De la Torre EM, Agasisti T, Perez-Esparrells C (2017) The relevance of knowledge transfer for universities' efficiency scores: an empirical approximation on the Spanish public higher education system. *Res Evaluation* 26(3):211–229
- De Witte K, López-Torres L (2017) Efficiency in education: a review of literature and a way forward. *J Oper Res Soc* 68(4):339–363
- De Witte K, Marques RC (2010) Incorporating heterogeneity in non-parametric models: a methodological comparison. *Int J Operational Research* 9(2):188–204
- Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. *Eur J Oper Res* 132(2):245–259
- Emrouznejad A, Yang GL (2018) A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Plann Sci* 61:4–8
- Fandel G (2007) On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds judged using DEA efficiency measures. *Eur J Oper Res* 176:521–533
- Färe R, Grosskopf S, Lovell CAK (1985) *The measurement of efficiency of production*. Springer Science & Business Media
- Farrell MJ (1957) The measurement of productive efficiency. *J Royal Stat Society: Ser (General)* 120(3):253–281
- Flegg AT, Allen DO, Field K, Thurlow TW (2004) Measuring the efficiency of British universities: a multi-period data envelopment analysis. *Educ Econ* 12(3):231–249
- Fried HO, Lovell CAK, Schmidt SS (eds) (2008) *The measurement of productive efficiency and productivity growth*. Oxford University Press
- García-Aracil A (2013) Understanding productivity changes in public universities: evidence from Spain. *Res Evaluation* 22(5):351–368
- Gerami J, Mozaffari MR, Wanke PF, Correa H (2022) A novel slacks-based model for efficiency and super-efficiency in DEA-R. *Oper Res Int Journal* 22:3373–3410
- Golany B, Roll Y (1989) An application procedure for DEA. *Omega* 17(3):237–250
- Gralka S, Wohlrabe K, Borrmann L (2019) How to measure research efficiency in higher education? Research grants vs. publication output. *J High Educ Policy Manag* 41(3):322–341
- Guironnet JP, Peypoch N (2018) The geographical efficiency of education and research: the ranking of U.S. universities. *Socio-Economic Plann Sci* 62:44–55
- Hajiagha SHR, Hashemi SS, Mahdiraji HA (2016) Fuzzy C-means based data envelopment analysis for mitigating the impact of units'. *Heterogeneity Kybernetes* 45(3):536–551
- Hazelkorn E (2015) *Rankings and the reshaping of higher education. The battle for world-class excellence*, 2nd edn. Palgrave Macmillan
- Hernández-Armenteros J, Pérez-García JA (2017) *La Universidad española en cifras 2015/2016*. CRUE

- Huntington-Klein N, Gill A (2021) Semester course load and student performance. *Res High Educat* 62(5):623–650
- Jayanthi SV, Balakrishnan S, Ching ALS, Latiff NAA, Nasirudeen AMA (2014) Factors contributing to academic performance of students in a tertiary institution in Singapore. *Am J Educational Res* 2(9):752–758
- Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. *Econ Educ Rev* 25(3):273–288
- Johnes J (2016) Performance indicators and rankings in higher education. In: Barnett R, Temple P, Scott P (eds) *Valuing higher education: an appreciation of the work of Gareth Williams*. UCL IOE Press. Ch. 4
- Johnes J, Taylor J (1990) *Performance indicators in higher education: UK universities*. Open University Press and the Society for Research into Higher Education
- Johnes G, Tone K (2017) The efficiency of higher education institutions in England revisited: comparing alternative measures. *Tert Educ Manag* 23:191–205
- Johnes J, Yu L (2008) Measuring the research performance of Chinese higher education institutions using data envelopment analysis. *China Econ Rev* 19:679–696
- Kao C, Hung HT (2008) Efficiency analysis of university departments: an empirical study. *Omega* 36(4):653–664
- Kazemi S, Mavi RK, Emrouznejad A, Kiani Mavi N (2021) Fuzzy clustering of homogeneous decision making units with common weights in data envelopment analysis. *J Intell Fuzzy Syst* 40(1):813–832
- Keramidou I, Mimis A, Pappa E (2011) Estimating technical and scale efficiency of meat products industry: the Greek case. *J Appl Sci* 11(6):971–979
- Koopmans TC (1951) Analysis of production as an efficient combination of activities. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation*. Cowles Commission for Research in Economics: Monograph No. 13 (33–97). Wiley
- Kounetas K, Androulakis G, Kaisari M, Manousakis G (2023) Educational reforms and secondary school's efficiency performance in Greece: a bootstrap DEA and multilevel approach. *Oper Res Int Journal* 23(1):9
- Liu JS, Lu LY, Lu WM, Lin BJ (2013) A survey of DEA applications. *Omega* 41(5):893–902
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1., pp. 281–297). University of California Press
- Mastromarco C, Toma P, Daraio C (2022) Teaching efficiency of Italian universities: A conditional frontier analysis. In D. Checchi, T. Jappelli, & A. Uricchio (Eds.), *Teaching, research and academic careers: An analysis of the interrelations and impacts* (pp. 319–358). Springer
- McMillan ML, Datta D (1998) The relative efficiencies of Canadian universities: a DEA perspective. *Can Public Policy/Analyse De Politiques* 24(4):485–511
- Milligan GW, Cooper MC (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50(2):159–179
- Mitropoulos P, Mitropoulos A (2023) Evaluating efficiency and technology gaps of the national systems of entrepreneurship using stochastic DEA and club convergence. *Oper Res Int Journal* 23(1):1
- Moncayo-Martínez LA, Ramírez-Nafarrate A, Hernández-Balderrama MG (2020) Evaluation of public HEI on teaching, research, and knowledge dissemination by data envelopment analysis. *Socio-Economic Plann Sci* 69:100718
- Moreira VH, Bravo-Ureta BE (2010) Technical efficiency and metatechnology ratios for dairy farms in three southern cone countries: a stochastic meta-frontier model. *J Prod Anal* 33(1):33–45
- Moreno-Gómez J, Calleja-Blanco J, Moreno-Gómez G (2020) Measuring the efficiency of the Colombian higher education system: a two-stage approach. *Int J Educational Manage* 34(4):794–804
- Murillo-Zamorano LR (2004) Economic efficiency and frontier techniques. *J Economic Surveys* 18(1):33–77
- Nazarko J, Šaparauskas J (2014) Application of DEA method in efficiency evaluation of public higher education institutions. *Technological and Economic Development of Economy* 20(1):25–44
- Ng YC, Li SK (2000) Measuring the research performance of Chinese higher education institutions: an application of data envelopment analysis. *Educ Econ* 8(2):139–156
- Nigsch S, Schenker-Wicki A (2015) Frontier efficiency analysis in higher education. In: Welpel IM, Woltersheim J, Ringelhan S, Osterloh M (eds) *Incentives and performance: governance of research organizations*. Springer, pp 155–170

- Nunamaker TR (1985) Using data envelopment analysis to measure the efficiency of non-profit organizations: a critical evaluation. *Manag Decis Econ* 6:50–58
- O'Donnell CJ, Rao DP, Battese GE (2008) Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics* 34(2):231–255
- OECD (2017) *Benchmarking higher education system performance: conceptual framework and data*. Enhancing higher education system performance. OECD Publishing
- Olesen OB, Petersen NC, Podinovski VV (2015) Efficiency analysis with ratio measures. *Eur J Oper Res* 245(2):446–462
- Rauhvargers A (2011) *Global university rankings and their impact. EUA Report on Rankings 2011*. European University Association
- Rostamzadeh R, Akbarian O, Banaitis A, Soltani Z (2021) Application of DEA in benchmarking: a systematic literature review from 2003–2020. *Technological and Economic Development of Economy*. 27(1):175–222
- Ruiz JL, Segura JV, Sirvent I (2015) Benchmarking and target setting with expert preferences: an application to the evaluation of educational performance of Spanish universities. *Eur J Oper Res* 242(2):594–605
- Sadri E, Harsef F, Hajiaghaei-Keshteli M, Siyahbalaii J (2022) Evaluation of the components of intelligence and greenness in Iranian ports based on network data envelopment analysis (DEA) approach. *J Modelling Manage* 17(3):1008–1027
- Salas-Velasco M (2020a) The technical efficiency performance of the higher education systems based on data envelopment analysis with an illustration for the Spanish case. *Educ Res Policy Pract* 19(2):159–180
- Salas-Velasco M (2020b) Measuring and explaining the production efficiency of Spanish universities using a non-parametric approach and a bootstrapped-truncated regression. *Scientometrics* 122(2):825–846
- Samoilenko S, Osei-Bryson KM (2008) Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. *Expert Syst Appl* 34(2):1568–1581
- Seiford LM, Zhu J (1999) An investigation of returns to scale in data envelopment analysis. *Omega* 27(1):1–11
- Shephard RW (1971) *Theory of cost and production functions*. Princeton University Press
- Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *J Econ* 136(1):31–64
- Sinuany-Stern Z, Hirsh A (2021) The relative efficiencies of higher education in OECD countries. In: Sinuany-Stern Z (ed) *Handbook of operations research and management science in higher education*. Springer, pp 481–512
- StataCorp (2021) *Stata for Windows* (Version 17) [Computer software]
- Steinley D (2006) K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 59(1):1–34
- Tavares RS, Angulo-Meza L, Sant'Anna P, A (2021) A proposed multistage evaluation approach for higher education institutions based on network data envelopment analysis: a Brazilian experience. *Eval Program Plan* 89:101984
- Thanassoulis E, De Witte K, Johnes J, Johnes G, Karagiannis G, Portela CS (2016) Applications of data envelopment analysis in education. In: Zhu J (ed) *Data envelopment analysis*. Springer, pp 367–438
- Titus MA, Eagan K (2016) Examining production efficiency in higher education: The utility of stochastic frontier analysis. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 31, pp. 441–512). Springer
- Villano RA, Tran CDT (2018) Performance of private higher education institutions in Vietnam: evidence using DEA-based bootstrap directional distance approach with quasi-fixed inputs. *Appl Econ* 50(55):5966–5978
- Visbal-Cadavid D, Martínez-Gómez M, Guijarro F (2017) Assessing the efficiency of public universities through DEA: a case study. *Sustainability* 9(8):1416
- Wolszczak-Derlacz J, Parteka A (2011) Efficiency of European public higher education institutions: a two-stage multicountry approach. *Scientometrics* 89(3):887–917
- Worthington A (2001) An empirical survey of frontier efficiency measurement techniques in education. *Educ Econ* 9(3):245–268
- Zhu J (2014) *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets*, 3rd edn. Springer

---

## Authors and Affiliations

### Manuel Salas-Velasco<sup>1</sup>

---

✉ Manuel Salas-Velasco  
msalas@go.ugr.es

<sup>1</sup> Department of Applied Economics | Campus Cartuja, University of Granada,  
Granada 18071, Spain