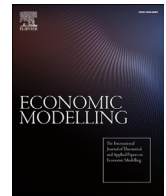


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Quality as a latent heterogeneity factor in the efficiency of universities[☆]

Cinzia Daraio^{a,*}, Léopold Simar^{a,b}, Paul W. Wilson^{c,d}^a Department of Computer, Control and Management Engineering, A. Ruberti (DIAG) Sapienza University of Rome, via Ariosto, 25, 00185, Rome, Italy^b Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), Université Catholique de Louvain, Voie du Roman Pays 20, B 1348, Louvain-la-Neuve, Belgium^c Department of Economics, Clemson University, Clemson, SC, 29634, USA^d School of Computing, Division of Computer Science, Clemson University, Clemson, SC, 29634, USA

ARTICLE INFO

JEL classification:

C1
C13
C14
D24
I23

Keywords:

Nonparametric efficiency
Performance assessment
Benchmarking
Directional distances
Conditional efficiency
Observed and unobserved heterogeneity
Separability condition
European universities
Quality

ABSTRACT

In this paper we show the usefulness of recent advanced nonparametric efficiency techniques to model the performance of universities in the presence of observed and unobserved heterogeneity. Using directional distances for benchmarking purposes, we identify a latent heterogeneity factor related to the human capital of the universities and their management, that is independent from their size, and interpret the identified latent factor as a “quality” factor of the universities. After testing the significance of this latent factor, we investigate its impact on the boundary of the production set (efficient frontier) and on the distances of the units from the efficient frontier. The frontier and the efficiency distribution of our European Universities sample appear influenced by our estimated latent quality factor. We investigate these impacts from various points of view, including the trade-off between our quality factor and efficiency as well as the roles of size and specialization.

1. Introduction

Efficiency, performance evaluation and benchmarking exercises abound in the empirical literature. There are many performance evaluation systems both (i) for business and (ii) for the public sphere, that with the advent of the so called *New Public Management* have affected particularly the educational sector.¹ A common approach in both practices is to define one or more Key Performance Indicators (KPIs) and compare them across different units. While this approach is useful in very simple cases, it has some drawbacks: it presumes constant returns to scale, it does not facilitate a comprehensive view of the unit under

analysis that accounts for all inputs and outputs, and different KPIs may point to different ideal units. It is difficult to evaluate an organization's performance when there are multiple performance metrics related to a system or operation. The difficulties are further enhanced when the relationships among the performance metrics are complex and involve unknown trade-offs.

In all cases, and in particular for the analysis of the performance of services, it is important to describe a general model of production on the base of which to run the empirical analyses. Performance is a broad concept which includes both productivity and efficiency. The productivity of a unit is defined as the ratio of its outputs to its inputs. Efficiency

[☆] The financial support of the Italian Ministry of Education and Research (through the PRIN Project N. 2015RJARX7), of Sapienza University of Rome (through the Sapienza Awards no. PH11715C8239C105 and no. RM11916B8853C925), and of the EU Horizon 2020 RISIS2 Project (grant agreement N 824091) is gratefully acknowledged.

* Corresponding author.

E-mail addresses: daraio@diag.uniroma1.it (C. Daraio), leopold.simar@uclouvain.be (L. Simar), pww@clemson.edu (P.W. Wilson).

¹ On business performance evaluation systems see Neely et al. (1995); Neely et al. (2000); Neely (2002). On the New Public Management, see Lane (2000) and Van Dooren, Bouckaert and Halligan (2015). On the impact of New Public Management on the educational sector in period of crisis, see Aparicio et al. (2018).

<https://doi.org/10.1016/j.econmod.2021.03.004>

Received 13 October 2020; Received in revised form 12 February 2021; Accepted 5 March 2021

Available online 20 March 2021

0264-9993/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

is instead the distance between the outputs/inputs ratio of a unit and the outputs/inputs ratio of the best possible or *efficient frontier* for the unit. As discussed in Daraio and Simar (2007, p. 14), productivity and efficiency are two *cooperating* concepts for analysing the performance of producing units.

Frontier efficiency analysis, introduced and developed in the economics of production, operational research and management science, and based on nonparametric quantitative methods (e.g. see Bogetoft, 2012; Zhu, 2014), are widely used in the context of performance evaluation and benchmarking for many reasons. First, it offers a rigorous analytical framework for representing a general model of production. Second, because of their empirical orientation and nonparametric nature, typical nonparametric efficiency estimators such as Data Envelopment Analysis (DEA, Farrell, 1957; Charnes et al., 1978) and Free Disposal Hull (FDH, Deprins et al., 1984), do not require a priori assumptions about the functional relationships between inputs and outputs. Third, frontier efficiency analysis allows identification of best practices as a means to improve performance and increase productivity. Finally, frontier efficiency analysis is particularly valuable for service operations, where identifying benchmarks or standards is more difficult than in a manufacturing context.

Nonparametric efficiency analysis is more and more used in studies involving best-practice identification in the nonprofit sector including education, higher education, the healthcare sector, in the regulated sector and in the private sector. Robust nonparametric techniques, based on the so-called partial frontiers, have also been introduced (see e.g. Daraio and Simar, 2007 for an introduction) to overcome some of the limits of the traditional nonparametric approach, namely the influence of extreme values and outliers. When directional distance functions introduced by Chambers et al. (1996) are used, the target is then defined as the virtual unit obtained by the projection of the evaluated unit to the efficient frontier along the chosen direction. The directional distance function approach provides a general and flexible way to use a benchmarking model as a *learning lab* (Bogetoft, 2012). By changing the direction of improvement the user can learn about the possibilities available and choose a production target based on this interaction. Recent surveys (e.g., Emrouznejad and Yang, 2017) show an increasing trend in applications of nonparametric efficiency analysis in all kind of services.

A major challenge in benchmarking and performance assessment of services is accounting for heterogeneity. One of the main criticisms of benchmarking analyses of all kinds is that they are not able to adequately take into account the peculiarities of the assessed units. The quantitative evaluations and comparison should take into account the main features of the assessed units, or in other words, should account for their *heterogeneity* and the efficiency analysis should include possibly *quality* dimensions. “Quality” is a difficult concept to define precisely. However, in this paper we propose to consider it as an *unobserved* factor of *heterogeneity* connected to the intellectual capital of the units assessed that is linked to the quality indicators commonly used in the literature. We will come back to this later in the application on European universities.

Quantitative studies of efficiency in the education sector have grown over the past two decades. While the earliest analyses of efficiency in the service sector (e.g. Ruggiero, 1996) have been mostly concerned with comparing input to output quantities, subsequent studies have tried to integrate output heterogeneity using various methods (e.g. Färe et al., 2006). The survey carried out by Worthington (2001) highlights the fact that despite the importance of the issue of efficiency, quantitative applications with frontier methods were not so numerous at the time. In a recent survey of 2017, De Witte and Lopez-Torres (2017, p. 356) after making a rich analysis of the inputs, outputs and contextual factors used in the numerous existing studies in the field of education efficiency, highlight in their conclusions that “researchers have to work with rather poor proxies” and that it is necessary to invest in “better and more detailed data on human resources”.

In this paper, we show the usefulness of the latest available nonparametric efficiency analysis techniques to model the heterogeneity of service and knowledge organizations as universities are, taking into account both observed heterogeneity factors and “Quality”. We consider the latter as an unobserved heterogeneity factor within the performance assessment (efficiency) of universities. More specifically, in the paper we (i) illustrate the usefulness of directional distances for benchmarking universities; (ii) identify a latent heterogeneity factor related to the human capital of the universities and their management, that is independent from their size; (iii) interpret the identified latent factor and find out that it may be considered as a “Quality” factor or score of the universities. Then, we test if this latent factor has an impact on the production set and analyze this impact on the boundary of the production set (efficient frontier) and on the distances of the units from the efficient frontier.

The paper is organized as follows. Section 2 summarizes existing related literature and details the contribution of the paper. Section 3 introduces the main concepts and notations. Section 4 introduces latent heterogeneity in the efficiency estimation. Section 5 illustrates the proposed approach on European Universities data, while the last section summarizes and concludes the paper. The two Appendices report the description of (i) the statistical test of the impact of latent heterogeneity on the efficiency (see Appendix A) and (ii) the procedure to recover gaps in original units (see Appendix B).

2. Existing literature and contribution

The previous section illustrates the main advantages of nonparametric methods for benchmarking purposes of service in general, as compared to traditional KPIs. Here we summarize the existing literature in terms of (i) how the production process is modelled, (ii) methodological approaches used, specifically in efficiency studies on universities; and (iii) critical issues related to the modeling of “quality” in knowledge organizations. We also describe the main contribution of this paper.

2.1. Modelling issues

The topic of efficiency modeling taking into account the heterogeneity of factors and the production process is of considerable importance. In addition, the inclusion of quality is crucial for all production sectors, in particular for knowledge organizations, such as universities, for which the human factor is crucial for the achievement of outcomes and performance. Existing methods and applied works on this topic are varied and increasingly numerous.

We can observe a change in the focus of current studies on the evaluation of the efficiency in the service sector. While the earliest studies have been mostly concerned with comparing input to output quantities, subsequent studies have tried to integrate output heterogeneity using various methods. More recent studies attempt to include in the empirical analyses also external or environmental variables that are neither inputs nor outputs but might affect the productive performance. Ruggiero (1996) is one of the first to point out that existing measures of technical inefficiency obtained through linear programming models in the public sector do not properly control for environmental variables that affect production and to show that the consequence of not controlling for these fixed factors are biased estimates of technical efficiency. Färe et al. (2006) model and compute productivity, including a measure of quality, of public education analysing Sweden public schools. They include a proxy for quality of the inputs, namely experience of teachers and found that quality matters for productivity growth changes. Lee and Kim (2014) propose a DEA-based approach to aggregate and benchmark different measures of service quality.

There are indeed different ways to include heterogeneity in efficiency analysis. The most-used within the nonparametric efficiency literature are (i) *one-stage approaches*, in which the contextual-environmental variables are included in the efficiency estimation as

inputs or outputs depending on the role they have in the production process; (ii) *two-stage approaches*, in which the (*unconditional*) efficiency scores are estimated including only the inputs and outputs and afterwards are regressed, in a second stage, against heterogeneity (contextual and/or environmental) variables; and (iii) *conditional approaches* that include heterogeneity variables by conditioning to their values the production process.

By considering an heterogeneity factor as an output, according to the one-stage approach, we are not able to empirically investigate whether the factor has a positive or negative, or a mixed impact on the production process. It is well known that the two-stage approach suffers from different limitations (see Simar and Wilson, 2007 and 2011) and is based on the so called *separability assumption* which, as will be seen below, assumes that the heterogeneity factor does not affect the efficient frontier of the best practice, but may affect only the distribution of the distances of the units from this efficient frontier. The conditional approach of Daraio and Simar (2007) extended in Badin et al. (2012, 2014), developed and applied in e.g. Halkos and Tzeremes (2011, 2013), Verschelde and Rogge (2012), Matousek and Tzeremes (2016), Minviel and De Witte (2017), and Mastromarco and Simar (2018), may be helpful in disentangling the impact of heterogeneity factors on efficiency without relying on the restrictive assumptions of the one-stage and two-stage approaches, as we will see below in Section 5 in the application to European universities.

2.2. State of the art of university efficiency

As noted in Section 1, Worthington (2001) in surveying existing efficiency studies in education pointed out that quantitative applications with frontier methods were not so numerous at the time. This may be due to the problems mentioned by Johnes (2006), Jones et al. (2009), and Lu (2012) that usually arise in studies of universities' efficiency. Universities carry out a complex production process. They realize different activities, such as teaching, research and knowledge transfer (the so called third mission), by combining different resources: human capital, financial stocks and infrastructures. Their activities, realized within an heterogeneous environment, produce heterogeneous outputs, such as undergraduate degrees, PhD degrees, scientific publications, citations, service contracts, patents, spin off and so on. In this process, size and subject mix also play an important role (e.g. Daraio et al., 2015 a,b and the references cited therein). Johnes (2006) investigates the measurement of efficiency in the higher education sector, highlighting the advantages and drawbacks of different methods focusing on Data Envelopment Analysis (DEA) that, for its flexibility and easiness in handling multidimensional production processes, is an attractive technique for measuring the efficiency of higher education institutions. Jones et al. (2009) highlight that the most valuable resources in any university are the expertise of its faculty and staff (intellectual capital) and extend existing methods to measure the intellectual capital in a university. These methods rely on comprehensive databases created by ad-hoc surveys. Lu (2012) applies a two-stage approach, based on a truncated-regression, to analyze whether intellectual capital influences the efficiency of universities and finds that intellectual capital influences both teaching and research efficiency.

Existing studies have been surveyed and analysed in many recent works, including Grosskopf et al. (2014), Johnson and Ruggiero (2014), Johnes (2015), Nigsch and Schenker-Wicki (2015), Klumpp (2015), Rhaïem (2017) and De Witte and Lopez-Torres (2017). Grosskopf et al. (2014) stress that education is a key policy sector due to its links to human capital, growth and innovation. They point out that the measurement of efficiency in services in general and education in particular is challenging, showing some advantages of frontier models in addressing some of the existing challenges. Johnson and Ruggiero (2014) investigate the educational efficiency of 604 school districts in Ohio, including external factors within a one-stage approach, and find that non-discretionary inputs are critical in explaining relative efficiency in pub-

lic sector applications.

Johnes (2015) observes that DEA and related non-parametric methods continue to be used in all sectors of education, including universities as units of analysis but also individual academic departments or programmes within an institution, or central administration or services across universities. Nigsch and Schenker-Wicki (2015) present a survey of the main applications of DEA to higher education, highlighting obtained results, methodological contributions, and main limitations. Klumpp (2015) discusses university management and policy and the relevant questions of which inputs and which outputs to use, and how to include quality in this framework. Rhaïem (2017) surveys 102 empirical studies published between 1990 and 2012 focusing on technical efficiency in academic research production. Around 11% of the surveyed papers analysed the efficiency in a multi-country framework, 18% analysed UK, the most studied country in the period, 12% China, around 10% Italy, 7% USA, around 5% Australia and Germany, followed by lower percentages in other country studies. European multi-country studies surveyed in Rhaïem (2017) include Wolszczak-Derlacz and Parteka (2011) and Wolszczak-Derlacz (2017) preceded by the pioneering works carried out within the European Project Aquameth, reported by Bonaccorsi and Daraio (2007), Daraio et al. (2011) and Bonaccorsi et al. (2014). According to Rhaïem (2017), over the 102 articles surveyed, only three consider some indicators of quality in the modelling of the efficiency of universities. These three are Warning (2004), Kempkes and Pohl (2008) and Sav (2012), which include as indicator of quality the percentage of student faculty ratio (as a measure of the potential teaching quality) and Sav (2012) which includes also the percentage of faculty employed that have received tenure to produce scholarly research activity. De Witte and Lopez-Torres (2017) survey 223 studies and offer many detailed tables with inputs, outputs and exogenous variables reporting the related papers in which the variables have been used. We refer the reader to the existing surveys cited above to have additional detailed information on the variables, data and specifications empirically applied by scholars to assess the efficiency of education and higher education institutions in particular.

Daraio (2018) summarizes the main methodological approaches that are usually applied to assess those institutions, their main limitations, and outlines proposals of improvement. Daraio (2019) presents econometric approaches to research productivity and efficiency, and investigate the potential of econometric approaches for research assessment. Each approach has advantages and disadvantages. The main advantage of non-parametric approaches is that they can handle multiple inputs and outputs in a simple way, while most stochastic approaches are based on univariate outputs. In addition, non-parametric approaches do not require any assumptions about the functional form or specification of the error term, from this their non-parametric nature come from, while stochastic methods need assumptions on these features. On the other hand, non-parametric methods also have shortcomings which include (i) deterministic nature, as they assume that all deviations from the frontier are due to inefficiency; (ii) the standard tools for statistical inference do not work in this nonparametric set up; (iii) sensitivity to extremes or outliers; and (iv) curse of dimensionality (need for thousands of observations to provide estimates of the inefficiency levels with an acceptable precision). In this paper we use state of the art techniques in nonparametric and robust efficiency which overcome the main shortcomings cited above and which allow inclusion of observed and unobserved heterogeneity factors in the analysis. To our knowledge, this is the first such application in education.

2.3. Quality of knowledge organizations

When knowledge organizations are investigated, the attempt to include "quality" generates measurement and conceptual difficulties. The concept of quality of higher education institutions is difficult and problematic. Its modeling in quantitative analysis is compelling and challenging. Regarding measurement issues, Seth et al. (2005), in their

Table 1
Selected references on “quality” in higher education.

Description	References
Conceptualization of “quality”	Harvey and Green (1993); Sarrico et al. (2010) Williams and de Rassenfosse (2018)
Total Quality in HE	Lewis and Smith (1994)
Quality Assurance and regulation in HE	Westerheijden et al. (2007)
Total Quality Management in Education	Sallis (2002)
Quality Management in HE	Manatos et al. (2016); Sarrico (2018a, 2018b)
Econometric modelling of Quality	Daraio (2017, 2019))
Human capital management and efficiency in HE	Kucharčíková et al. (2015)
Academic Quality (reputation and excellence)	Paradeise and Thoenig (2015)

review of various quality models, note that “often the outcomes may be guided by the way quality of service is being measured (Seth et al., 2005, p. 945)”. They observe that the quality outcome and measurement is dependent on the type of service setting, situation, time and need factors. One of their concluding remarks is that the answer to the question “How to quantify and measure quality of service?” is one of the research issues to further address in future research.

Regarding conceptual difficulties, the task of defining quality in higher education is rather tricky, due to the complexity of the matter (Sarrico, 2018a,b; Sarrico et al., 2010): “A consensus seems to have emerged in recent years that attempts to define quality can be regarded as an *unrewarding venture*, since quality does not appear to exist as something unique and absolute in higher education” (Sarrico et al., 2010, p. 40). “Quality” can have several different meanings, including quality as academic excellence or quality as value for money. Quality seems to be not only an elusive concept, but also a complex one that can be perceived in very different ways (Westerheijden et al., 2007). According to this perspective, quality is seen as a multidimensional concept that should take into account all these different perspectives about higher education and its quality, going from quality to *qualities* of higher education (Blackmur, 2007). Daraio (2017) proposes an overarching concept of quality to develop models for the quantitative assessment of research and higher education, based on a conceptual framework made by three dimensions: theory, methodology and data. From this framework it clearly appears the challenging role of the econometric modeling of quality from a methodological perspective.

Human capital is relevant to increase productivity and output of organizations as it includes natural ability, innate skills, knowledge, experience, talent and inventiveness. In the context of university education, it has been observed by Kucharčíková et al. (2015, p. 52) observe that there are several approaches for how to measure the value of human capital, but a single methodology has not yet been adopted. This is because on the one hand there is a problem of quantification of knowledge, ability, skills, motivation and talent. On the other hand, the main models proposed in the literature, based on accounting, “have not achieved wider application in practice, due to largely subjectivism, uncertainty and lack of replicability” (Kucharčíková et al., 2015, p. 52).

Paradeise and Thoenig (2015, pp. 1–2) state that “Academic quality still remains a black box not only with regard to assessing the outputs, but also in terms of the formal and informal social, cultural and organizational processes adopted by specific university governance regimes”. Paradeise and Thoenig (2015) identify two components of quality: reputation (internal component, the *elitist oligarchy*) and excellence (external component, rankings and *Top of the Pile* model). Quality is linked to the academic staff, it is a combination of the “iron law of talent”, and of a “post-excellence” quality which rests in administrators and faculty. Table 1 provides a summary of the literature on quality in higher education without any claim of completeness.

2.4. Our contribution

After the helicopter view of existing studies of efficiency in higher education given above, we can now summarize our contribution. We contribute to the existing literature: by (i) showing in the illustration on a sample of European universities (see Section 5) that the methodology proposed in this paper, based on a nonparametric and robust benchmarking framework including observed and unobserved heterogeneity for knowledge organizations (i.e. universities) (described in Sections 3 and 4) represents a significant advance in the existing literature on the efficiency of universities summarized above; by (ii) extending the separability test proposed by Daraio et al. (2018) to a directional distance framework; by (iii) proposing an empirical definition of “quality” as an unobserved factor related to the human capital of knowledge organizations; and by (iv) providing empirical evidence on the impact of the estimated latent “quality” on the efficient frontier and on the distances of the units from the efficient frontier. We investigate these impacts from various points of view, including the trade-off between our quality factor and efficiency as well as the roles of size and specialization.

The contribution of our paper can be also considered as an attempt to relax the rigidity of current evaluation models which impose precise definitions and standardization of the dimensions in which the activities are organized. This is very difficult for activities related to human capital such as services and in particular of knowledge organizations, as universities are. Vidaillet (2013, p. 120) observes that “working implies cultivating some secrets.” Therefore, in evaluating performance, factors and characteristics not directly observed, related to the human capital involved, must also be considered. *Intangibles and intellectual capital* have always been considered as relevant factors to the productivity and competitiveness of the private sector as well as of the public sector (Guthrie and Dumay, 2015; Dumay et al., 2015; Secundo et al., 2018). Nevertheless, the measurement of intellectual capital (Bryl, 2018) is an emerging research area in *knowledge management* (Tiwana, 2000; Alavi and Leidner, 2001; Liebowitz, 2012). Being at its infant stage, it still lacks a rigorous methodology for being assessed and remains difficult to be directly measured and included in a more general performance measurement system. Therefore, our attempt is to show that existing up-to-date nonparametric and robust directional distance measures, including observed and unobserved heterogeneity factors, can be a first step towards a more comprehensive model of performance which includes also “quality”. Although conceptually difficult, we show that “quality” can be operationalized in the efficiency calculation as a latent heterogeneity factor related to the human capital.

3. Frontiers and conditional frontier models

This section introduces and summarizes the basic setup and notation for frontier, conditional frontier models and their robust version. Here we present a comprehensive summary of concepts developed in Cazals et al. (2002), Daraio and Simar (2005), Simar and Wilson (2007, 2011), Bădin et al. (2012, 2014), Daraio et al. (2018, 2020) and Simar and Vanhems (2012). The reader who is familiar with these approaches can

easily skip this section. Below, Section 4 introduces the methodology to include latent heterogeneity in this setup.

3.1. Introducing heterogeneity in frontier models

Production may be characterized by a process generating a vector of inputs and outputs defined over an appropriate probability space. Let $X \in \mathbb{R}^p$ denote inputs and $Y \in \mathbb{R}^q$ the outputs. Define the attainable set

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} \mid x \text{ can produce } y\} \quad (3.1)$$

as the set of values (x, y) which are technically possible.

The attainable set Ψ is the support of the joint distribution of (X, Y) which can be described, e.g. by the joint probability $H_{XY}(x, y) = \text{Prob}(X \leq x, Y \geq y)$, which is the probability of finding a unit (X, Y) dominating the point (x, y) . As shown in Cazals et al. (2002),

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} \mid H_{XY}(x, y) > 0\} \quad (3.2)$$

under the assumption of free disposability.²

In the presence of external or environmental factors $Z \in \mathcal{Z} \subset \mathbb{R}^r$ that may introduce heterogeneity by influencing the production process, the probability space to consider has to be augmented. Thus, we consider the random variables X, Y, Z and we denote by \mathcal{P} the support of the joint distribution of (X, Y, Z) . Let Ψ^z denote the support of (X, Y) given that $Z = z$. Thus the attainable set for units facing external conditions $Z = z$ is

$$\begin{aligned} \Psi^z &= \{(x, y) \in \mathbb{R}^{p+q} \mid x \text{ can produce } y \text{ if } Z = z\}, \\ &= \{(x, y) \in \mathbb{R}^{p+q} \mid H_{XY|Z}(x, y \mid z) > 0\} \end{aligned} \quad (3.3)$$

where $H_{XY|Z}(x, y \mid z) = \text{Prob}(X \leq x, Y \geq y \mid Z = z)$. The variables in Z can affect the production process either (i) only through Ψ^z the support of (X, Y) , or (ii) only through the conditional distribution (X, Y) given Z , affecting e.g. only the probability of a unit to reach its optimal boundary, or (iii) through both. It is easy to see that $\Psi = \cup_{z \in \mathcal{Z}} \Psi^z$, so that $\Psi^z \subseteq \Psi$, for all $z \in \mathcal{Z}$. In the very particular case where the joint support of (X, Y, Z) can be written as a cartesian product $\mathcal{P} = \Psi \times \mathcal{Z}$, then Z will have no impact on the boundaries of Ψ and $\Psi^z = \Psi$ for all $z \in \mathcal{Z}$ (this is called the “separability condition” in this literature; see for example, Simar and Wilson, 2007, 2011). In the latter case, Z may eventually influence the production process only through the probability of reaching its optimal boundary.

The performance of a unit operating at level (x, y) can be measured by its distance to its optimal boundary defining a measure of efficiency. Several measures have been proposed in the literature (see e.g. Fried et al., 2008). We will focus our presentation to flexible directional distances (see e.g. Chambers et al., 1998; Färe et al., 2008). The choice of the directions $d_x \in \mathbb{R}_+^p$ and $d_y \in \mathbb{R}_+^q$ for measuring the distance from the efficiency boundary of unit operating at level (x, y) allows us to analyze different strategies of the units to reach the efficient frontier. The directional distance is defined by

$$\begin{aligned} \beta(x, y; d_x, d_y) &= \sup\{\beta > 0 \mid (x - \beta d_x, y + \beta d_y) \in \Psi\}, \\ &= \sup\{\beta > 0 \mid H_{XY}(x - \beta d_x, y + \beta d_y) > 0\} \end{aligned} \quad (3.4)$$

where the second equality assumes free disposability of inputs and outputs (see Simar and Vanhems, 2012). Note that $\beta(x, y; d_x, d_y) \geq 0$ for $(x, y) \in \Psi$ and that a value of zero indicates a unit (x, y) on the efficient boundary. It measures the distance of the unit (x, y) toward the boundary of Ψ along the path determined by (d_x, d_y) . Similarly, for conditional measures we add the conditioning on $Z = z$ to obtain

$$\beta(x, y; d_x, d_y \mid z) = \sup\{\beta > 0 \mid H_{XY|Z}(x - \beta d_x, y + \beta d_y \mid z) > 0\}. \quad (3.5)$$

² The free disposability of inputs and outputs assumes that if $(x, y) \in \Psi$, then $(\tilde{x}, \tilde{y}) \in \Psi$ for all (\tilde{x}, \tilde{y}) such that $\tilde{x} \geq x$ and $\tilde{y} \leq y$. In a sense, it assumes the possibility of wasting resources.

It is well known that the particular case $d_x = 0$ and $d_y = y$ allows us to recover the popular output-oriented radial measures of Farrell-Debreu and of Shephard (the input-oriented case is given by $d_x = x$ and $d_y = 0$). Note that the additive nature of directional distances permits negative input and output quantities, which is not the case for radial distances.

Nonparametric estimators are obtained by substituting the nonparametric estimators \hat{H}_{XY} and $\hat{H}_{XY|Z}$ in the expressions above (we give more details in Section 3.3). As shown in Cazals et al. (2002), Daraio and Simar (2005) and Simar and Vanhems (2012) this allows us to recover the Free Disposal Hull (FDH, Deprins et al., 1984) estimators of the efficiency measures and even the Data Envelopment Analysis (DEA, Farrell, 1957, Charnes et al., 1978) estimators if we convexify the FDH estimator of the attainable set (see Simar et al. (2012) for their statistical properties). All of these nonparametric estimators have well-known asymptotic properties: to summarize, they suffer from the curse of dimensionality, and practical inference for individual efficiencies requires bootstrap techniques (see Simar and Wilson, 2015, and the references therein for a recent survey).³

The analysis of the effect of Z on efficiency is based on the investigation of the ratios of the conditional on the unconditional efficiency scores (Daraio and Simar, 2005, 2007). Bădin et al. (2012, 2014) show that in the output orientation an increasing shape of the ratios (unconditional divided by conditional efficiency scores) as a function of Z corresponds to an unfavorable (negative) effect of Z , while the opposite is true for a decreasing trend (positive effect of Z). Daraio and Simar (2014) extend this approach to directional distances, considering the differences between unconditional and conditional efficiency scores, and show that an increasing trend of these differences implies a negative impact of Z on the frontier, while a decreasing trend of these differences points to a positive impact of Z .

3.2. Robust approach: partial frontiers

The nonparametric estimators (FDH or DEA type) are envelopment estimators in the sense that the corresponding estimate of Ψ (or of Ψ^z) envelops the cloud of observed data points and so they are quite sensitive to extreme values and outliers. This is the major interest of the robust version of these estimators developed for radial measures (for an overview see Daraio and Simar, 2007). Simar and Vanhems (2012) extend these concepts to directional distances. The idea is to define a less extreme boundary as benchmark, here we define a partial-frontier by contrast to the full-frontier used above. It allows us to measure the distance of a unit to a partial-frontier allowing, by construction, some data points to be outside this partial-frontier. Two ways have been suggested in the literature: the order- α quantile frontier and the order- m partial frontier. An introduction and an overview on these methods may be found in Daraio and Simar (2007). In this summary we give only some intuitive definitions for the case of one output, one input and with the output orientation (e.g. $d_x = 0$ and $d_y = 1$) and for the unconditional to Z case. In the next section we describe the most general cases.

For any $\alpha \in (0, 1]$ the order- α measure of efficiency is given by

$$\beta_\alpha(x, y; 0, 1) = \sup\{\beta \mid S_{Y|X}(y + \beta \mid x) > 1 - \alpha\}, \quad (3.6)$$

where $S_{Y|X}(y \mid x) = \text{Prob}(Y \geq y \mid X \leq x) = H_{XY}(x, y)/F_X(x)$ is the conditional survival function of Y given $X \leq x$. We remark that if $\alpha \rightarrow 1$, we are back to the usual full frontier measure (for $d = (0, 1)$). So for $\alpha < 1$, the benchmark frontier for the unit (x, y) (i.e. where

³ For instance, for the FDH case we will follow below, the rate of convergence of the efficiency estimates is of the order $n^{1/(p+q)}$ which becomes much less than the usual parametric rate of convergence ($n^{1/2}$) when the dimension of the problem is $p + q > 2$.

$\beta_\alpha(x, y; 0, 1) = 0$) corresponds to the α -quantile of the conditional distribution of the output among the population of units using less inputs than x . So $\beta_\alpha(x, y; 0, 1)$ can take negative values if y is large and the unit lies above this conditional quantile.

The order- m frontier in output orientation is defined, for any integer m , as

$$\phi_m(x) = \mathbb{E} [\max(Y_1, \dots, Y_m) | X \leq x], \quad (3.7)$$

where the Y_j are independent and identically distributed (iid) realizations of the output Y , conditionally on $X \leq x$. Hence $\beta_m(x, y; 0, 1) = \phi_m(x) - y$ which can take negative values for large values of y . Here, as $m \rightarrow \infty$, we are back to the usual full-frontier measure. So the benchmark frontier is the expected value of the maximum output among m peers drawn from the population of units using less inputs than x . It can be shown that when Y takes only positive values

$$\phi_m(x, y) = \int_0^\infty [1 - (1 - S_{Y|X}(y | x))^m] dy. \quad (3.8)$$

Nonparametric estimators are obtained by *plugging-in* the empirical version of the conditional survival function ($\hat{S}_{Y|X}(y | x)$) in the previous equation. They share interesting properties, in particular they achieve the parametric \sqrt{n} rate of convergence independently of the dimension of the problem. Their robustness properties rely on the fact that for large α or m we estimate a partial frontier not far from the full one, but for $\alpha < 1$ and finite m , the estimators will not envelop all the data points and so are robust to extreme data points and outliers. Comparisons of the two concepts from a robustness point of view can be found in [Daouia and Ruiz-Gazen \(2006\)](#) and [Daouia and Gijbels \(2011\)](#). In particular, it is known that once the order- α quantile frontiers break down for large chosen tail probability levels, they become less resistant to outliers than the order- m frontiers. Also, the asymptotic theory when conditioning on latent heterogeneity has been established in [Simar et al. \(2016\)](#) only for the order- m case. Consequently, we use the order- m approach for our analysis below.

Nonparametric frontier estimation, conditional and unconditional, and their robust versions, are widely applied. Examples include [Verschelde and Rogge \(2012\)](#), [Varabyova and Schreyögg \(2017\)](#), [Matousek and Tzeremes \(2016\)](#), [Minviel and De Witte \(2017\)](#).

3.3. Nonparametric estimators of DDF: a summary

In this section we summarize the main computational aspects linked to the nonparametric estimation of the directional distance functions introduced in the previous section. Nonparametric estimators are obtained by plugging nonparametric estimates of H_{XY} or of $H_{XY|Z}$ into these expressions. For a sample of observations $S_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$, they are given by

$$\hat{H}_{XY}(x, y) = n^{-1} \sum_{i=1}^n 1I(X_i \leq x, Y_i \geq y), \quad (3.9)$$

$$\hat{H}_{XY|Z}(x, y | z) = \frac{\sum_{i=1}^n 1I(X_i \leq x, Y_i \geq y)K_h(Z_i - z)}{\sum_{i=1}^n K_h(Z_i - z)}, \quad (3.10)$$

where $1I(\cdot)$ is the indicator function and where $K_h(Z_i - z)$ is a product kernel with bandwidths h determined by standard least squares cross validation techniques (see e.g. [Li et al., 2013](#)).

Under the free disposal assumption only, the FDH estimators are given by

$$\hat{\beta}_{FDH} = \max_{\beta} \{ \beta | x - \beta d_x \geq X_i, y + \beta d_y \leq Y_i, i = 1, \dots, n \}, \quad (3.11)$$

$$\hat{\beta}_{FDH}^z = \max_{\beta} \{ \beta | x - \beta d_x \geq X_i, y + \beta d_y \leq Y_i, i \in I(z, h) \}, \quad (3.12)$$

where $I(z, h) = \{i | |Z_i - z| \leq h\}$ and the inequality $|Z_i - z| \leq h$ has to be understood component by component. In other words, the conditional FDH is a localized version (only based on observations i such that

Z_i is around the value z) of the unconditional FDH; the localization is governed by the bandwidths vector h (see [Bădinetal., 2019](#)).

If in addition, we want to add the assumption of convexity of the attainable sets, we have the DEA estimators

$$\hat{\beta}_{DEA} = \max_{\beta} \left\{ \beta | x - \beta d_x \geq \sum_{i=1}^n \gamma_i X_i, y + \beta d_y \leq \sum_{i=1}^n \gamma_i Y_i, \gamma_i \geq 0, \sum_{i=1}^n \gamma_i = 1 \right\}, \quad (3.13)$$

$$\hat{\beta}_{DEA}^z = \max_{\beta} \left\{ \beta | x - \beta d_x \geq \sum_{i \in I(z, h)} \gamma_i X_i, y + \beta d_y \leq \sum_{i \in I(z, h)} \gamma_i Y_i, \gamma_i \geq 0, \sum_{i \in I(z, h)} \gamma_i = 1 \right\}. \quad (3.14)$$

Additional, practical details are given by [Simar and Vanhems \(2012\)](#) and [Simar et al. \(2012\)](#). See also [Simar and Wilson \(2015\)](#) for a recent survey. For the variants, conditional and unconditional, of robust measures see [Daraio et al. \(2020\)](#) where Matlab codes are provided. All of these estimators have been implemented in the current version of the FEAR package for R introduced by [Wilson \(2008\)](#).

4. Inclusion of latent heterogeneity

As observed in [Simar and Wilson \(2007, 2011\)](#) and in [Daraio et al. \(2018\)](#), neglecting heterogeneity factors Z that are not separable may introduce problems. This happens when the boundary of the attainable set varies with Z (i.e., if $\Psi^z \neq \Psi$ for some $z \in \mathcal{Z}$). In fact, the problem is that the boundary of Ψ considered by ignoring these factors may be not achievable for units facing particular external conditions described by Z and hence, benchmarking units against such boundary has little economic meaning. We have to consider the boundary of Ψ^z for units facing condition $Z = z$.

The problem is the same if we suspect that some unobserved (latent) factor of heterogeneity may affect the boundary of the attainable set. In our illustration below we will use an output orientation and we propose to use the approach suggested by [Simar et al. \(2016\)](#), which allows identification of a latent factor linked to some input (the converse would follow similar developments).

Suppose without loss of generality that this latent heterogeneity factor, V , is linked to the input X^1 and that we can write the link through the nonparametric model

$$X^1 = \varphi(W, V), \quad (4.1)$$

where W is an *auxiliary* variable, correlated to X^1 but independent of V . This model is nonseparable in V and has been studied in the econometrics literature (see e.g. [Matzkin, 2003](#)). The classical assumptions of the model are as follows: monotonicity (increasing) of φ with respect to V and without loss of generality V is uniformly distributed on $[0, 1]$ (it is just a matter of scaling V such that it can be interpreted as a quantile). It is known that under these assumptions V is identified by the conditional distribution of X^1 given W , i.e.,

$$V = F_{X^1|W}. \quad (4.2)$$

Thus, we can see the latent heterogeneity variable V as the part of X^1 which is independent of W . The choice of the input X^1 and of the auxiliary variable W are crucial to identify the latent heterogeneity variable we are interested in. We may identify latent factors using a different auxiliary variable for each input ([Simar et al., 2016](#)) or we could even use the same auxiliary variable for identifying latent heterogeneity factors linked to different inputs. As pointed in [Simar et al. \(2016\)](#), it has to be noticed that the function φ is unknown and in nonseparable models like (4.1) V plays the role of residual. Under the monotonicity assumption, V is identified by (4.2) and since V is uniform on $[0, 1]$, φ can

Table 2
Variables about European Universities used in the illustration.

Role	Acron.	Description	Source
Inputs	ACAD	Total number of academic staff	ETER
	NONAC	Total number of non-academic staff	ETER
	TEXP	Total expenditures in Euro PPP ^a	ETER
Outputs	TDEG	Total number of degrees ISCED5-7 ^b	ETER
	PUB	Total number of publications	SCIMAGO
	PHD	Total number of PhD degrees	ETER
“unobserved” Het. factor	V	estimated by $\hat{V}_i \in [0, 1]$ (see below)	our elab.
Observed Het. factor	$Z = SPEC$	Degree of specialization $\in [0, 1]$	SCIMAGO
Auxiliary variable	SIZE	Total number of enrollments	ETER
Observed “quality”	%REVTHIRD	Share of third party funds	ETER
	F. Year	Foundation year	ETER
partial indic.	%IC	International Collaboration rate	SCIMAGO
	NI	Normalized Citation Impact	SCIMAGO
	%Q1	High “quality” Publication ratio	SCIMAGO
	%Exc.	Excellence ratio	SCIMAGO
	%EwL.	Excellence with Leadership ratio	SCIMAGO
	WR	Scimago World Ranking	SCIMAGO
	RR	Scimago European Ranking	SCIMAGO

^a PPP stands for Purchasing Power Parity.

^b ISCED is the International Standard Classification of Education maintained by the UNESCO. ISCED 5 is short cycle tertiary education, ISCED 6 corresponds to bachelor’s level and ISCED 7 to Master’s level.

be interpreted as a quantile function. This is a nice duality property of these nonseparable models. The choice of the uniform distribution for V is not a limitation since it is just a matter of rescaling V , but if we rescale it in another way, then we lose the natural interpretation in terms of quantile function and cdf (cumulative distribution function).⁴ We will see below how to estimate these unknown quantities.

The approach above may work *mutatis mutandis* in many setups. In the application to the activity of European Universities, we will choose X as the total number of academic staff and W as total enrolled students that represents a proxy for the size of the university. The identified V is what remains from the academic personnel X once we have accounted for its volume component W , and in that sense, we can interpret V as a factor related to the quality of the human capital of the universities and their management, once we have removed the impact of the size. In practice, we can check that our identified latent factor behaves as expected by model (4.1), i.e. if our estimated V is independent of W . We should also check empirically if the identified latent factor V may be related to some known partial indicators of quality to give empirical support to our guess above (see Section 5). This approach to estimate latent heterogeneity factors identifying what remains from the volume of the human capital once we have accounted for its size component could be extended and tested also in other contexts and different services. This is left to further research.

5. Application on European universities

In this section we illustrate the proposed methodology by analysing the efficiency of European universities. We first introduce the data. After that, we estimate an unobserved heterogeneity factor identifying what remains from the volume of the human capital once we have

⁴ Since the nonparametric function φ in (4.1) is monotone increasing with respect to V , any monotonic, increasing transformation of V could be included in φ , but the interpretation will depend on the specific transformation that is used. For example, we might model the latent factor \tilde{V} as being $N(0, 1)$ instead of uniformly distributed by defining $\tilde{V} = \Phi^{-1}(V)$ where Φ^{-1} is the quantile function of the $N(0, 1)$ distribution. But then we lose equation (4.2), and this would modify the function φ and its natural interpretation.

accounted for its size component. After that, we check that our identified latent factor behaves as expected by model (4.1). We also check empirically if the identified (unobserved factor) V may be related to some known partial indicators of quality. Finally we estimate the efficiency and complete the benchmarking analysis.

5.1. Data and variables specification

Our data have been collected within the European Project ETER (European Tertiary Education Register) and have been validated by national statistical authorities.⁵ The ETER data considered refer to year 2011 (academic year 2011/2012). They include as inputs total number of academic staff (ACAD) and total number of non-academic staff (NONAC), total expenditures (TEXP) that is the sum of all expenditures (includes expenditure for personnel, non-personnel, capital and unclassified expenditure); as outputs total number of degrees (TDEG) in all the educational levels without the PhDs which are considered as an additional output (PHD). As additional variables, that are neither inputs nor outputs, but which may affect the production process, we consider the share of Third party funding (in PPP) over Total revenues (in PPP, indicated as %REVTHIRD), the foundation year (F. Year) i.e. the year when the institution was established. The total enrolled students ISCED 5–7 plus PhD students is used as a proxy for SIZE and will serve as an auxiliary variable in the following analysis. The list of all the variables used in this paper and their sources are summarized in Table 2.

These data were integrated with other data on the scientific activity of universities collected from the Scopus bibliometric database in the Scimago Global 2013 Rank (SCIMAGO in Table 2), whose data refer to outputs realized in the years 2007–2011. These scientific publications data include total number of publications (PUB) considered as an output which includes the total number of documents published in scholarly journals indexed in Scopus, the specialization index (SPEC) that

⁵ For additional information and to download the data, see the project website: <http://eter.joanneum.at/imsdas-eter/> where one can find also additional information on the variables and the Data Quality Report.

Table 3
Inputs, Outputs and Environmental variables used in the analysis: averages by country.

Country (Code)	#obs	ACAD	NONACAD	TEXP	TDEG	PUB	PHD	SPEC	SIZE
Belgium (BE)	5	297065158.52	2851.11	1425.54	5002.40	330.80	12685.60	0.50	9.40
Switzerland (CH)	11	334058152.35	2251.07	1115.27	2417.00	315.09	8884.00	0.65	9.23
Cyprus (CY)	1	107912583.34	389.00	566.00	1525.00	43.00	2862.00	0.71	8.81
Germany (DE)	73	431951227.71	2009.73	2458.53	2801.77	351.82	7352.86	0.65	9.63
Denmark (DK)	8	310038255.68	2106.25	1694.25	4023.50	193.12	8438.25	0.66	9.47
Hungary (HU)	7	300646738.90	1310.14	3225.29	4325.43	134.71	3927.14	0.70	10.05
Ireland (IE)	10	163345454.22	929.21	814.84	3826.90	136.80	4019.30	0.62	9.42
Italy (IT)	60	261088844.47	1448.43	959.42	4890.20	181.73	5989.38	0.67	9.87
Lithuania (LT)	4	354822075.78	955.25	927.75	3416.50	47.00	2292.00	0.80	9.41
Luxembourg (LU)	1	112655414.25	741.98	306.53	803.00	57.00	1643.00	0.74	8.46
Malta (MT)	1	104048589.04	791.00	759.00	3345.00	19.00	897.00	0.67	9.22
Netherlands (NL)	13	412966543.19	1868.71	1381.97	5836.38	295.46	14897.46	0.62	9.78
Norway (NO)	10	193484267.88	1298.14	938.31	2417.30	127.60	4955.40	0.70	8.94
Portugal (PT)	17	88363541.43	860.31	599.77	2757.24	115.06	2968.94	0.69	9.24
Sweden (SE)	20	224089812.76	1215.27	920.08	2478.10	166.30	6234.75	0.67	9.62
United Kingdom (UK)	96	279935092.91	1339.79	1517.76	6464.06	261.30	6801.55	0.64	9.73
Europe (EU)	337	298163446.67	1538.62	1506.55	4411.41	241.41	6760.74	0.66	9.64

indicates the extent of thematic concentration/dispersion of an institution's scientific output (with values between 0 and 1, indicating generalist vs. specialized institutions respectively), that will be considered as a Z variable, and other variables considered as observed partial quality indicators, that are International Collaboration Institution's output ratio (%IC), Normalized Impact of citations (NI), High quality Publications Ratio (publications in the first 25% of the distribution % Q1), Excellence Rate (percentage of publications among the most 10% of highly cited publications, %Exc.), Excellence with Leadership (%EwL) that indicates the amount of documents in the Excellence rate in which the institution is the main contributor, the placement of the institution in the Scimago ranking at world level (WR), the placement of the institution in the Scimago ranking at regional level (where region = Europe, RR). From these sources we have the data available for $n = 337$ European universities. See Table 2 for the list of variables we use in our application and their sources.

Due to the limited size of the available sample, and due to the high correlation between the three inputs and between the two research outputs (PUB and PHD), we use the dimension reduction based on factor analysis, suggested in Daraio and Simar (2007) and analysed by Monte-Carlo analysis in Wilson (2018). See Appendix B for more technical details. The resulting input factor, denoted FX , is determined by the first eigenvector of the second moment matrix of the three inputs $u_x = (0.57230.62180.5346)'$, which can roughly be interpreted as an average of the scaled inputs; it explains 96% of the total inertia and so little information is lost by using this single input factor. Its correlations with the three original inputs are 0.9777, 0.9474 and 0.9325 respectively. For the two research outputs we have similar results with $u_y = (0.69860.7155)'$ which explains 97% of the total inertia. The resulting output research factor, denoted FY , has correlations 0.9676 and 0.9691 with PUB and PHD, respectively. So we end up with 337 observations with one input $X = FX$ and two outputs $Y = (TDEG, FY)$ the first one being the teaching activity and the second summarizing the research activity.

Table 3 shows some descriptive analysis based on the average values at the country level of the inputs, outputs and external variables used in the empirical illustration. The first column of Table 3 reports the country and its acronym, while the second column reports the number of observations (universities) considered in each country and the last column reports the average of the SIZE variable in log units. To give an overview on the analysed sample of European universities and its variety we report Tables 4 and 5. Table 4 reports some descriptive statistics (average) on Quality Indicators and \hat{V} , while Table 5 illustrates some descriptive statistics (average) on the variables used in the efficiency analysis, namely FX , Y_1 , FY , \hat{V} and $SPEC$.

The directional distance function approach provides a general and flexible way to use a benchmarking model as a *learning lab* (see Bogetoft, 2012), as noted in Section 1. By changing the direction of improvement, the user can learn in an *interactive manner* about the possibilities available and choose a production target or budget based on this interaction. Addressing strategic issues through directional distances for outputs (because $d_x = 0$), we compare an egalitarian centralized path (median direction: $d_y = \text{med}(Y)$), as often used in analysis with directional distances, with the results obtained by using autonomous paths (individual directions). This will allow us to analyze the difference of *centralized* directions towards a given output mix (egalitarian direction) versus *autonomous* directions of improvement selected by the units (individual directions) of the European Humboldtian university model of education production of teaching and research (Schimank and Winnes, 2000).

For identifying a latent heterogeneity factor V , we decide to select the input factor and try to identify the part of FX which is independent of the SIZE of the university, which acts as an auxiliary variable according to the model described in Section 4. Due to the asymmetric nature of the size of universities, that is distributed as a lognormal, we work rather with $W = \log(\text{SIZE})$, which formally does not change anything, but simplifies the nonparametric estimation of $F_{X|W}$, avoiding huge universities isolated with large values of W_i .

5.2. Estimation and results

5.2.1. Unobserved heterogeneity factor

We start our analysis by the estimation of a latent heterogeneity factor V_i . First, once the values of \hat{V}_i are obtained, we check whether the assumption of independence between V and the instrument W is reasonable. As observed by Simar et al. (2016), the theory for a test of independence has still to be provided, but we can at least compute the various correlations between \hat{V}_i and W_i and inspect the p -values for the hypothesis that these correlations could be zero (as they would in case of independence). The results are shown in Table 6 and clearly indicate that the assumption of independence seems to be reasonable.⁶

Then we check if the identified latent factor can be interpreted as related to some observed partial quality factors. This is done by looking to the correlations (Pearson) between \hat{V}_i and some proxies suggested in

⁶ As requested by an anonymous referee, we also computed the Kolmogorov-Smirnov distance between the joint distribution of $F_{\hat{V}W}$ and the product of the marginal distributions $F_{\hat{V}}F_W$ and obtained the value 0.0193. The bootstrap algorithm described in Li et al. (2009) provided a p -value = 0.912, confirming, as a robustness check, that the assumption of independence is reasonable.

Table 4
Quality Indicators and \hat{V} : averages by country.

Country	#obs	%REVTHIRD	%IC	NI	%Q1	%Exc	%EwL	WR	RR	F.Year	\hat{V}
BE	5	0.53	52.56	1.56	54.24	19.05	8.62	517.00	180.80	1837.60	0.74
CH	11	22.58	58.34	1.74	55.35	21.43	9.93	808.91	292.55	1851.64	0.74
CY	1	5.26	65.93	1.51	45.91	18.08	6.49	1125.00	385.00	1989.00	0.61
DE	73	19.99	40.47	1.38	46.28	15.42	7.08	843.07	306.68	1810.52	0.65
DK	8	25.87	50.68	1.69	49.03	20.11	9.85	1029.75	391.75	1893.50	0.58
HU	7	40.46	39.65	0.93	36.23	9.22	3.34	986.86	49.57	1946.14	0.50
IE	10	31.27	46.48	1.29	42.87	15.81	8.46	1271.60	485.00	1892.30	0.45
IT	60	12.39	36.32	1.38	48.69	15.32	6.06	961.77	347.68	1773.30	0.32
LT	4	43.69	26.98	0.92	19.66	8.74	5.03	1420.75	91.25	1861.50	0.67
LU	1	14.32	68.65	1.48	30.68	17.46	8.13	1641.00	591.00	2003.00	0.75
MT	1	6.90	43.70	0.99	27.20	10.26	5.35	2213.00	861.00	1769.00	0.52
NL	13	26.84	46.24	1.77	60.41	21.47	10.28	274.92	89.92	1827.15	0.69
NO	10	19.16	46.67	1.38	51.27	16.19	7.21	1402.30	541.10	1962.80	0.59
PT	17	8.92	38.86	1.10	36.41	12.18	5.11	1637.76	636.12	1931.53	0.31
SE	20	26.47	44.42	1.38	48.60	15.73	7.12	1185.55	448.55	1922.55	0.34
UK	96	13.81	39.89	1.44	48.59	16.57	8.11	1062.10	388.74	1864.60	0.48
EU	337	17.82	41.30	1.40	47.45	16.01	7.34	1010.00	360.80	1847.75	0.50

Table 5
Variables used in efficiency analysis: averages by country.

Country	#obs	FX	Y_1	FY	\hat{V}	SPEC
BE	5	2.6456	5002.4000	2.2341	0.7353	0.4980
CH	11	2.2969	2417.0000	1.7929	0.7367	0.6500
CY	1	0.6388	1525.0000	0.4175	0.6142	0.7100
DE	73	2.8646	2801.7671	1.7329	0.6518	0.6545
DK	8	2.3792	4023.5000	1.4125	0.5842	0.6638
HU	7	2.4952	4325.4286	0.7800	0.5002	0.6986
IE	10	1.1306	3826.9000	0.7954	0.4493	0.6180
IT	60	1.6636	4890.2000	1.1247	0.3248	0.6687
LT	4	1.5926	3416.5000	0.3687	0.6736	0.8025
LU	1	0.7404	803.0000	0.3281	0.7507	0.7400
MT	1	0.9116	3345.0000	0.1460	0.5163	0.6700
NL	13	2.3605	5836.3846	2.3692	0.6939	0.6208
NO	10	1.4332	2417.3000	0.8683	0.5914	0.7030
PT	17	0.8570	2757.2353	0.6259	0.3147	0.6876
SE	20	1.4482	2478.1000	1.1082	0.3427	0.6745
UK	96	1.8487	6464.0625	1.4275	0.4791	0.6442
EU	337	1.9879	4411.4095	1.3688	0.5022	0.6569

Table 6
Correlations of \hat{V}_i with W_i and p -values.

	Pearson	Spearman	Kendall
Correlations	-0.0187	0.0311	0.0236
p -values	0.7329	0.5695	0.5186

the literature to indicate some partial quality indicators of the university output production (see Moed, 2017 and the references in Table 1). The results are shown in Table 7, where we also give the correlations with the two outputs (Y_1 is teaching (TDEG) and Y_2 is our research factor (FY)). We can see that all the correlations have the expected sign and are when needed clearly different from zero. The negative correlations in Table 7 are as expected: decreasing ranks indicate increasing quality (i.e., higher values of \hat{V}). It appears we have the same effect for the age of the university: quality increases with the age, and hence decreases with the year of foundation.

We see that the estimated latent factor V can be interpreted as *the hidden component of the resources*, after the elimination of the size component, that contributes to the quality of the university. Interestingly, the same results have been obtained if we estimate the latent factor not of the aggregated input factor (FY) but only of the total number of Academic Staff (ACAD). This could confirm that the estimated latent factor is mainly related to the unobserved or difficult to measure quality of the human capital and in particular of the academic staff of the

Table 7
Correlations of \hat{V}_i with outputs and some observed partial indicators of "quality". Output Y_1 is the number of degrees ISCED5-7 and output Y_2 is the research factor.

Variable	Correlation
Y_1	0.0609
Y_2	0.5817
%REVTHIRD	0.4305
%IC	0.4405
NI	0.4508
%Q1	0.4785
%EXC	0.4751
%EWL	0.3549
WR	-0.6139
RR	-0.6050
F. Year	-0.2626

universities. The quality of the academic staff is made by internal quality (*elitism and reputation*) and external quality (*excellence and rankings*) according to Paradeise and Thoening (2015).

Summing up, the estimated *unobserved* quality factor of universities is linked to their resources (input), in particular to the academic staff. We will investigate in Section 5.2.2 if it plays a role on the efficiency of the higher education systems (and which kind of role, i.e. if it is *comple-*

mentary to or a substitute for efficiency), and afterwards we will assess its impact on the benchmarking frontier, including also an observed factor of heterogeneity (Z) that is subject mix or specialization of the higher education institutions.

Now the role of our identified latent quality factor on the production process is still an open question. Does it act as a hidden input, or as a latent output? Does it influence the shape of the production possibilities (attainable set) of universities and/or the distribution of their efficiency scores? These questions are addressed in the next section.

5.2.2. Frontier estimation and benchmarking

Before starting our analysis, we performed a test of convexity due to Kneip et al. (2016) and the convexity assumption was highly rejected (with a p -value = 0.0000166). In all the frontier analysis then we use FDH-based estimators. These do not rely on the convexity assumption of the attainable set Ψ .

We test the separability condition where (Z, V) have no influence on the boundary of the attainable set. We perform the test of the separability, first for V and Z themselves and then jointly for (Z, V) . We use the multiple splitting procedure suggested by Simar and Wilson (2020). For each sample split, we obtain a test statistic T_j and a corresponding p -value p_j . We perform 1000 splits, and consider the sample average \bar{T} of the T_j , and the Kolmogorov-Smirnov (KS) statistic for the distribution of the p_j which is uniform under the null hypothesis. Neither the T_j nor the p_j are independent, but the bootstrap described by Simar and Wilson (2020) provides valid p -values for both \bar{T} and the KS statistic. We obtain $\bar{T} = 2.0252$ and $KS = 0.5294$ with corresponding p -values less than 10^{-9} in both cases. Therefore we reject the separability condition without hesitation.⁷ The test provides clear evidence that the variables $SPEC = Z$ and V the identified unobserved quality factor, hereafter labeled as $UQUAL = V$, modify (have an impact on) the shape of the efficient benchmarking boundary.

We investigate the effects of our variables (Z, V) on potential shifts of the frontier by analyzing the nonparametric regression surface of estimates of $\beta(x, y; 0, d) - \beta(x, y; 0, d | z, v)$ on (z, v) as explained in Bădin et al. (2012) and Daraio and Simar (2014). Fig. 1 displays the results and illustrates the way in which the two variables affect the shift of the efficient frontier by looking to the local linear regression of the differences $\hat{\beta}(x, y; 0, d) - \hat{\beta}(x, y; 0, d | z, v)$ on (z, v) (see e.g. Bădin et al., 2012, and Daraio and Simar (2014)).

Of course the efficiency measures depend also on the input level x , so we should analyze these differences as a function of (z, v) for fixed levels of x . We follow the strategy of Florens et al. (2014) and fix three levels of the input factor at its 3 quartiles (Q_1, Q_2, Q_3); we then take all the available measures for the observations (X_i, Y_i, Z_i, V_i) such that $|X_i - Q_k| \leq h_x, k = 1, \dots, 3$, where h_x is the normal reference rule bandwidth for X . This yields three subsamples with 66, 85 and 48 observations respectively. From these we build the 3 local linear estimates of the regression of $\hat{\beta}(x, y; 0, d) - \hat{\beta}(x, y; 0, d | z, v)$ on (z, v) . The results are displayed in Fig. 1.

Fig. 1 shows that the effect on the efficient benchmarking frontier (shift) is present for all the values of X , but is much more important for the large units (with high level of staff). We see also that the latent quality factor \hat{V} has a bigger effect than the specialization ($SPEC$). This effect (the shift) is more important for universities with high quality factor indicating a trade-off between quality and the efficiency of production.

A careful analysis of Fig. 1 shows the existence of some interesting interactions between $SPEC$, latent quality factor and university size. In particular, for the median case (middle panel of Fig. 1) the impact on the efficient frontier appears to be monotonically decreasing in $SPEC$ for low level of the latent quality factor \hat{V} . Specialization has a positive impact on the efficient frontier for low level of quality. On the contrary,

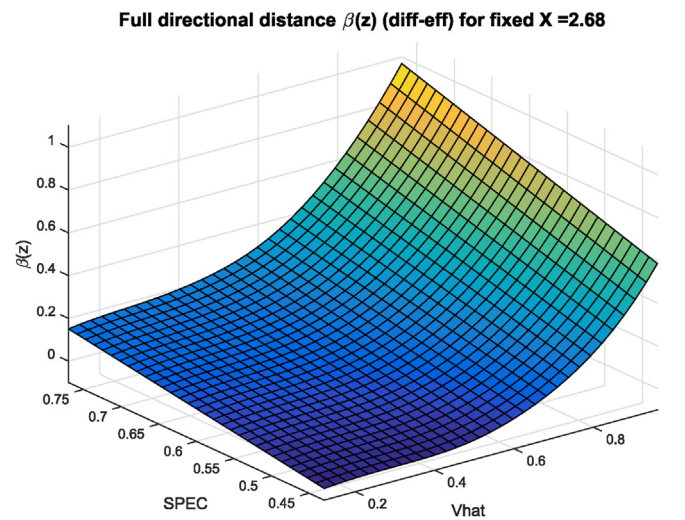
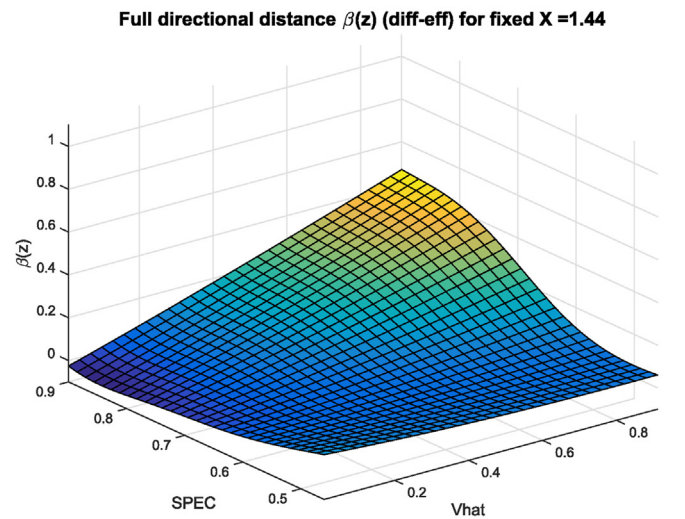
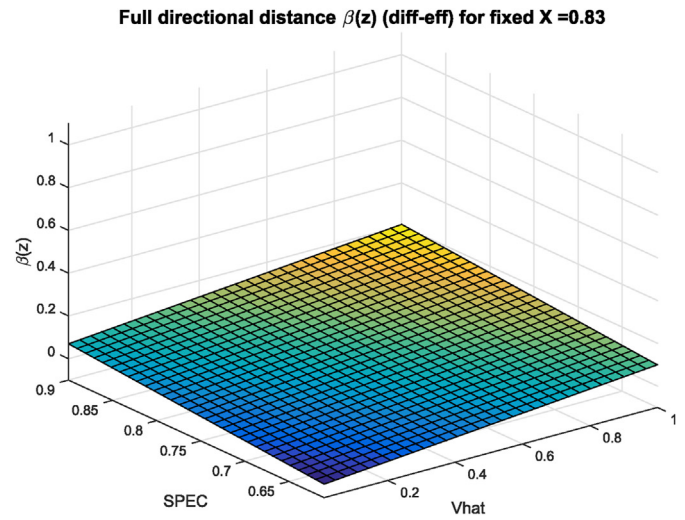


Fig. 1. Impact of $\hat{V} = UQUAL$ and $Z = SPEC$ on the shift of the full frontier $\beta(x, y; 0, d) - \beta(x, y; 0, d | z, v)$, where $d = \text{med}(y)$ for fixed values of the Input Factor (FX) at the 3 quartiles: from top to bottom, small, median and large level of labor (ACAD).

⁷ Using only 100 splits gave quite similar results.

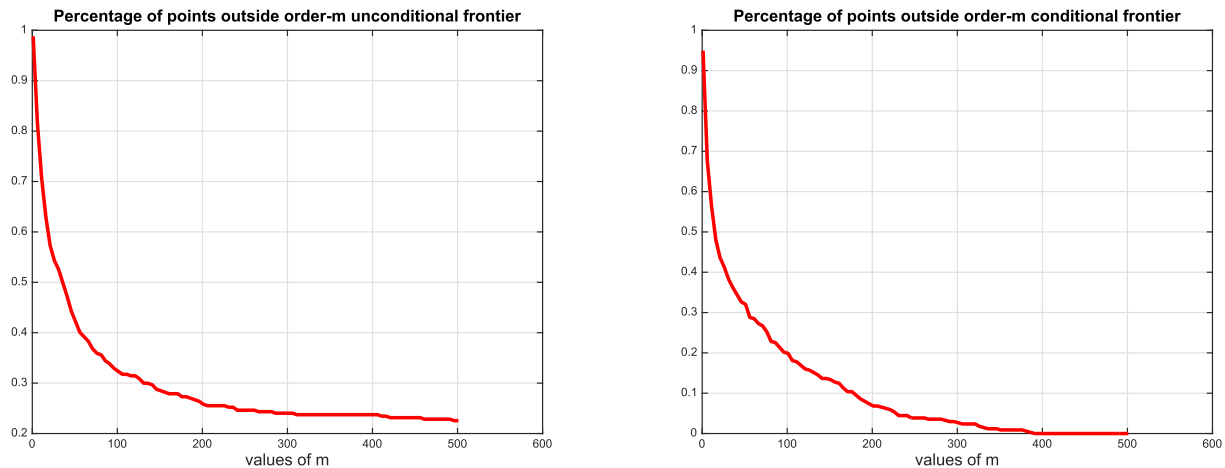


Fig. 2. Percentage of points outside the order- m frontier. From the left panel (unconditional efficiencies), we select $m = 310$, around 24% points still outside the frontier. On the right panel, conditional on (V, Z) , with $m = 310$, only around 2% points outside the conditional frontier.

Table 8
Estimates of Efficiency, direction is egalitarian: averages by country and standard deviations of the conditional measures $\hat{\beta}(x, y|z)$ and $\hat{\beta}_m(x, y|z)$.

Country	#obs	$\hat{\beta}(x, y)$	$\hat{\beta}(x, y z)$	std	$\hat{\beta}_m(x, y)$	$\hat{\beta}_m(x, y z)$	std
BE	5	0.1687	0.1152	0.1293	0.1196	0.1152	0.1293
CH	11	0.5883	0.2051	0.2207	0.5129	0.2051	0.2207
DE	73	0.9908	0.6996	0.6140	0.8801	0.6887	0.6066
DK	8	0.7121	0.4228	0.3848	0.6213	0.4179	0.3797
HU	7	1.0406	0.5463	0.4533	0.9870	0.4237	0.2954
IE	10	0.1293	0.0637	0.0990	0.1159	0.0637	0.0990
IT	60	0.1976	0.1137	0.1788	0.1504	0.1060	0.1681
LT	4	0.7334	0.2923	0.2242	0.7021	0.2923	0.2242
NL	13	0.3250	0.0579	0.0959	0.2190	0.0576	0.0954
NO	10	0.5508	0.4408	0.5428	0.5045	0.4373	0.5412
PT	17	0.1219	0.0723	0.1059	0.1075	0.0721	0.1059
SE	20	0.3445	0.2262	0.2866	0.3191	0.2260	0.2863
UK	96	0.0972	0.0621	0.1305	0.0184	0.0522	0.1156
EU	337	0.4072	0.2582		0.3374	0.2488	

Table 9
Estimates of Efficiency, direction is autonomous: averages by country and standard deviations of the conditional measures $\hat{\beta}(x, y|z)$ and $\hat{\beta}_m(x, y|z)$.

Country	#obs	$\hat{\beta}(x, y)$	$\hat{\beta}(x, y z)$	std	$\hat{\beta}_m(x, y)$	$\hat{\beta}_m(x, y z)$	std
BE	5	0.1609	0.0648	0.0857	0.1443	0.0648	0.0857
CH	11	0.3912	0.1699	0.2238	0.3411	0.1699	0.2238
DE	73	0.6984	0.4914	0.4472	0.6416	0.4880	0.4480
DK	8	0.5091	0.2981	0.3417	0.4633	0.2944	0.3408
HU	7	1.1710	0.4907	0.3665	1.1074	0.3996	0.2903
IE	10	0.1721	0.1264	0.2065	0.1618	0.1233	0.1982
IT	60	0.2779	0.1638	0.2909	0.2574	0.1579	0.2879
LT	4	1.6082	0.4993	0.4257	1.5668	0.4719	0.3993
NL	13	0.2042	0.0375	0.0615	0.1562	0.0372	0.0608
NO	10	0.7585	0.3205	0.3045	0.7342	0.3115	0.3059
PT	17	0.3204	0.2158	0.3035	0.3122	0.2147	0.3038
SE	20	0.4443	0.2684	0.2906	0.4223	0.2665	0.2882
UK	96	0.0896	0.0496	0.0998	0.0668	0.0457	0.0976
EU	337	0.3804	0.2245		0.3486	0.2184	

the impact (shift) is monotonically increasing in SPEC for high level of \hat{V} . This means that increasing specialization has a negative impact on the efficient frontier of median universities with high level of quality. Specialization then, can be used to increase technical efficiency only for low quality levels; for high quality levels, an increase in specialization has a negative impact on the frontier of production possibilities. This is true for mid-sized universities. By analyzing the large universities (bottom panel of Fig. 1), we observe that the differences are increas-

ing monotonically with SPEC for both low and high qualitative values. This means that for large universities, specialization always has a negative impact on the frontier of production possibilities, regardless of the quality level.

To analyze the impact of (Z, V) on the distribution of the efficiency scores, we use robust estimators of the frontier to avoid that extreme data points or outliers hide some effects (see Daraio and Simar, 2007, for simple examples of these situations). We choose to perform the

robust analysis by using the order- m partial frontiers. One might perform a similar analysis using the order- α quantile frontier. Comparisons of the two concepts from a robustness point of view can be found in Daouia and Ruiz-Gazen (2006) and Daouia and Gijbels (2011). We prefer to focus the presentation with the order- m case for two reasons. First for robustness properties: once the quantile based frontiers break down they become definitively less resistant to outliers than the order- m frontiers. Second, the asymptotic theory linked with the identification of latent factors and its use in frontier estimation has been done in Simar et al. (2016) for order- m only. We conjecture that the same theory is valid for order- α , but it is only a conjecture, so we prefer to do the analysis with the order- m robust frontiers.

We select a value for the order m using the standard methods suggested in the literature (see e.g. Daraio and Simar, 2007; Daouia and Gijbels, 2011), i.e. by looking to the percentages of points lying above the estimated order- m frontier, as a function of m . Of course this curve will converge to zero as $m \rightarrow \infty$. This is shown in the left panel of Fig. 2, where the curve indicates a shoulder effect (i.e., becomes more “flat”), indicating a far larger value of m is needed for the points outside or above the order - m frontier to fall under the (larger m) frontier. This suggests that the points lying above our order - m frontier are extreme data points and potential outliers. Here we select $m = 310$, letting around 24% of the data points outside the frontier.⁸

Interestingly, when drawing the analogous picture for the conditional to (Z, V) order- m frontier in the right panel of Fig. 2, we see that with $m = 310$ almost all the points are under the frontier except eight of them (around 2%). This indicates that most of the heterogeneity which was present in the input \times outputs space has mostly disappeared when conditioning on (Z, V) . In the latter cases, the order- m estimates will be very similar to the full conditional frontier (for $m \rightarrow \infty$, i.e. the conditional FDH frontier). This will be confirmed in the tables of results shown below.

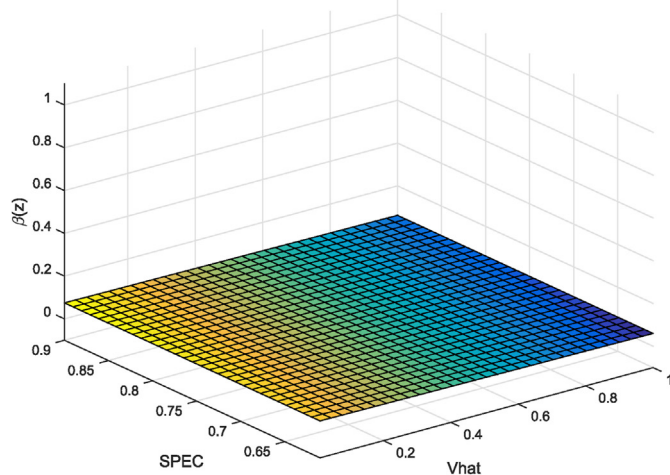
We focus on the comparison of the averages of the efficiency scores by country, provided in Tables 8 and 9. Each table shows by column the country, the number of observations ($\#obs$), averages of the full unconditional ($\hat{\beta}(x, y)$) and conditional ($\hat{\beta}(x, y | z)$) efficiency scores, their corresponding robust versions ($\hat{\beta}_m(x, y)$ and $\hat{\beta}_m(x, y | z)$) and their standard deviation (std).⁹ The difference between the two tables lies in the direction chosen for reaching the efficient frontier. In Table 8 the directional vector is the same for all the universities (egalitarian centralized path) and is fixed at the European median level ($med(Y)$). While in Table 9 the direction is different for each university (with individual directions given by the values of Y) showing *autonomous* paths.

Considering the values of robust conditional efficiency ($\hat{\beta}_m(x, y | z)$) and remembering that closer to zero is the value of $\hat{\beta}_m(x, y | z)$ the higher is the level of efficiency, we can compare the average values reported in Tables 8 and 9. We note that in some countries (Belgium, Switzerland, Germany, Denmark, Netherlands, Norway and United Kingdom) passing from the egalitarian direction (Table 8) to the autonomous one (Table 9) we observe an increase in efficiency (reduction of the $\hat{\beta}_m(x, y | z)$ value), while for the other countries (Ireland, Italy, Lithuania, Portugal and Sweden) we have a reduction in efficiency (increase in the value of $\hat{\beta}_m(x, y | z)$) associated with the transition from the same direction for all (centralized path) to autonomous direction. Hungary remains almost unchanged. This is a striking result that may point to existing differences in the governance systems of the higher education national systems: more *differentiated higher education systems*

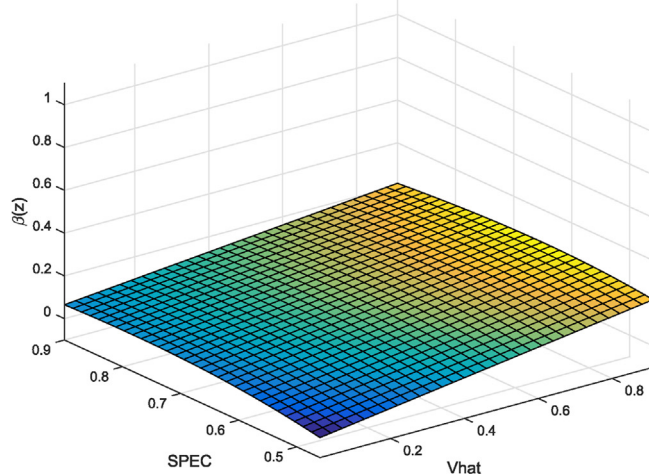
⁸ In this analysis we focus on the analysis of the impact of the heterogeneity and quality factors on the efficient frontier. It could be interesting investigate what happens if we consider the impact of those factors on the average production function that can be obtained by setting the level of $m = 1$. This further analysis, based on a different paradigm of production, namely average production function, is left for future work.

⁹ Countries with only one university such as Cyprus, Luxemburg and Malta are not displayed.

Order- m ($m=310$) directional distance $\beta(z)$ (cond-eff) for fixed $X=0.83$



Order- m ($m=310$) directional distance $\beta(z)$ (cond-eff) for fixed $X=1.44$



Order- m ($m=310$) directional distance $\beta(z)$ (cond-eff) for fixed $X=2.68$

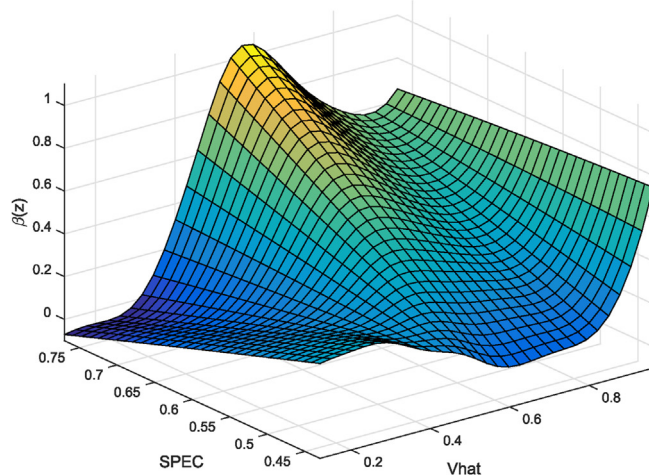


Fig. 3. Impact of $\hat{V} = UQUAL$ and $Z = SPEC$ on conditional order- m efficiency measures $\beta_m(x, y; 0, d|z, v)$, where $d = med(y)$ for fixed values of the Input Factor at the 3 quartiles: from top to bottom, small, median and large levels of labor.

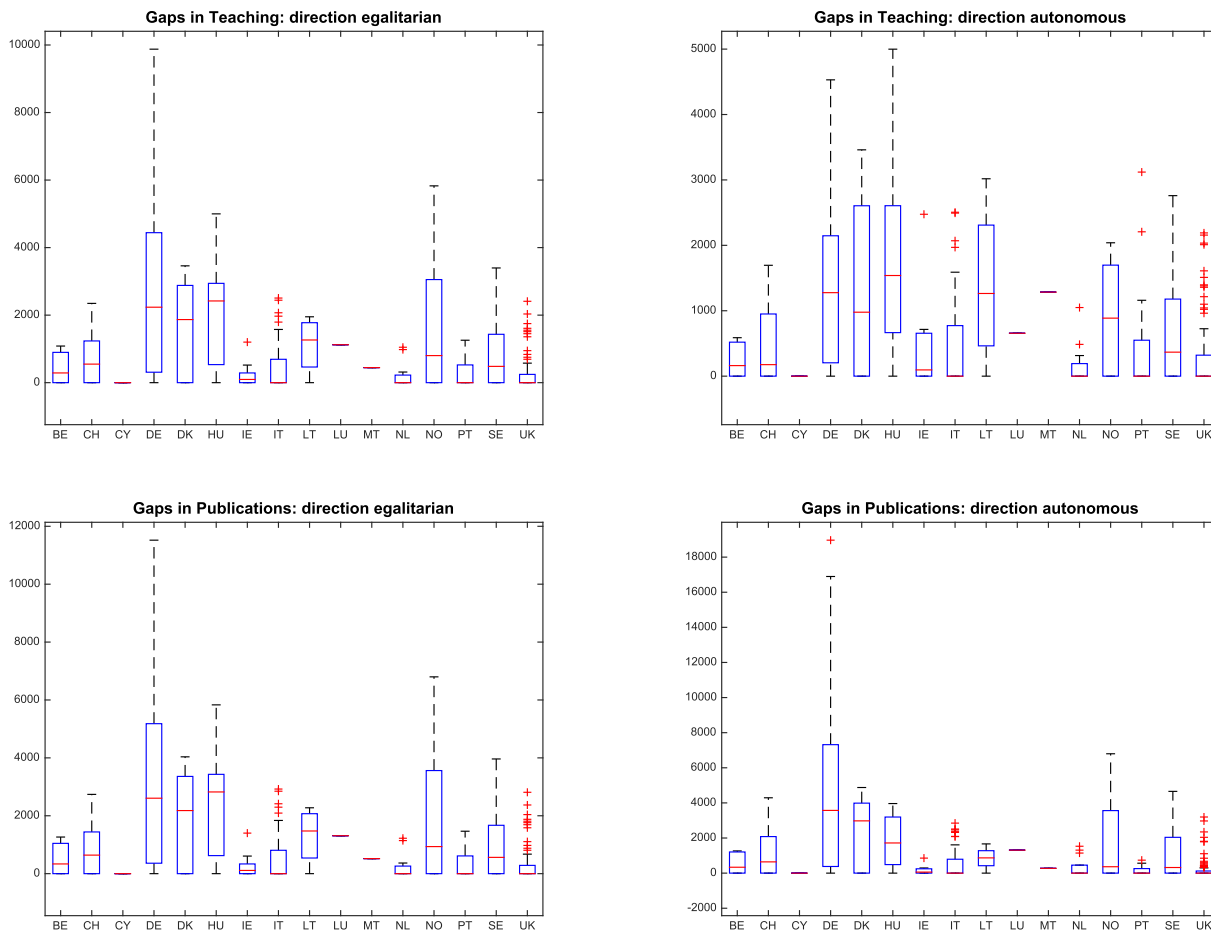


Fig. 4. Estimated gaps in the outputs. Left panels report the boxplots of the European countries considered following an egalitarian centralized path (median direction). Right panels show the boxplots obtained by selecting autonomous path (individual directions).

including Switzerland, Netherlands and United Kingdom benefit from the autonomy in the choice of the path to follow in order to reach the best practice frontier, while *undifferentiated higher education systems* such as Italy and Portugal are not able to fully exploit their autonomy because of governance constraints. Of course this is just a conjecture that should be empirically validated with additional research and is outside the scope of the present paper. The inclusion in the analysis of variables on the governance of higher education systems may represent an interesting line for further research. In aggregate, Europe improves its level of efficiency by moving from the same direction for all to the autonomous one (see the last row of Tables 8 and 9 corresponding to EU).

In the next step, we analyze the impact of (Z, V) on the efficiency measures $\beta_m(x, y | z, v)$ (see Bädin et al., 2012; Daraio and Simar, 2014). As above for Fig. 1 the efficiency measures depends on the input level x , so we analyze $\hat{\beta}_m(x, y | z, v)$ as a function of (z, v) for fixed levels of x at its three quartiles (Q_1, Q_2, Q_3). From the three subsamples, as above we build the three local linear estimates of the regression of $\hat{\beta}_m(x, y | z, v)$ on (z, v) . The results are displayed in Fig. 3.

Globally, efficiency decreases ($\beta_m(x, y; 0, d | z, v)$ increases) when X increases. We see an almost flat impact for $X = Q_1$ (first quartile of small universities with low academic staff). We observe a slight negative effect of the quality factor on efficiency (as V increases, $\beta_m(x, y; 0, d | z, v)$ increases) for $X = Q_2$ median-sized universities. There is also a modest effect of the specialization (SPEC). It seems that there is a *trade-off* between quality factor and efficiency: when the quality factor (V) increases universities may decrease their efficiency levels (the value of $\beta_m(x, y; 0, d | z, v)$ increases), they may produce less

of their output mix. In addition, for big universities (large staff number corresponding to the third quartile of the distribution, $X = Q_3$), there is an interaction between degree of specialization (SPEC) and quality factor: we observe a different effect for specialized university than for generalist ones, pointing globally to a *trade-off of quality vs efficiency* except for generalist (unspecialized) universities (with lower values of SPEC) which seem to combine efficiency and quality well.

Finally, Fig. 4 gives, for each country, the boxplots of the gaps for each university to reach the frontier according to the egalitarian and autonomous directions. They are given in the original units of the outputs, even for the research outputs that were transformed in the output factor (FY) in the analysis. See Appendix B for a detailed explanation of how to construct the gaps in the original units. The boxplots confirm the results reported in Table 8 and in Table 9, but in addition give an idea of the efforts required to eliminate the gaps and to reach the efficient frontier in terms of the original units of the outputs.

6. Conclusions

In this paper we describe the importance of overcoming the limitations of a performance evaluation based on Key Performance Indicators in the field of services in general, as well as for the evaluation of universities. We show the importance of an assessment based on robust and nonparametric efficiency analysis techniques capable of including both observed and unobserved (or latent) heterogeneity factors. We propose a nonparametric procedure to estimate *unobserved* quality features, test their impact on the performance and analyze it, in a state-of-the-art

nonparametric performance evaluation model based on up-to date conditional and robust frontier estimation techniques.

Analysing European Universities we identify a latent heterogeneity factor and interpret this as a kind of “quality” factor of the universities. After testing the significance of the latent “quality” factor, we investigate its impact on the boundary of the production set or efficient frontier and on the distances of individual units from the efficient frontier. We find that the estimated latent quality factor has an impact on the efficient frontier. It seems that there is a *trade-off* between quality and efficiency: when quality increases, universities may decrease their efficiency levels and may produce less of their output mix. In addition, for big universities, there is an interaction between degree of specialization and quality: we observe a different effect for specialized universities than for generalist ones, pointing globally to a *trade-off of quality vs efficiency* except for generalist (unspecialized) universities which seem to combine efficiency and quality well. We calculate and compare the efficiency measures related to different paths towards the efficient frontier, selecting different directions towards the benchmarking frontier. We compare an egalitarian centralized path (median direction), with the results obtained by using autonomous paths (individual directions). It seems that more *differentiated higher education systems* includ-

ing Switzerland, Netherlands and UK benefit from the autonomy in the choice of the path to follow in order to reach the best practice frontier, while *undifferentiated higher education systems* such as Italy and Portugal are not able to fully exploit their autonomy because of governance constraints. Although these results are interesting, to consolidate them additional research and the extension of the investigations to updated data and a broader sample, including e.g. US universities, are required and are left for further research. In our examination of the activity of European universities, we identify a latent quality variable related to the human capital of the universities and their management, that is independent from their size. We believe that this approach for estimating latent quality factors, and our specific choice of identifying it as what remains from the volume of the human capital or labour once we have eliminated its size component, could be particularly interesting in the area of quantitative assessment of intangibles, intellectual capital and knowledge management. It would be interesting to extend and test the proposed approach also in other contexts and different services¹⁰. This, however, is left to further research.

Declaration of competing interest

None.

A. Statistical issues and separability test

Nonparametric estimators of the unknown functions in (4.1) and (4.2) are obtained from a sample of observations (X_i^1, W_i) by the following estimator

$$\begin{aligned} \hat{V}_i &= \hat{F}_{X^1|W}(X_i^1 | W_i) \\ &= \frac{\sum_{k=1}^n \mathbb{1}(X_k^1 \leq X_i^1) K_{h_w}(W_i - W_k)}{\sum_{k=1}^n K_{h_w}(W_i - W_k)}, \end{aligned} \quad (\text{A.1})$$

of V_i , where $\mathbb{1}(\cdot)$ is the indicator function, $K_{h_w}(W_i - W_k) = (1/h_w)K((W_i - W_k)/h_w)$ and $K(\cdot)$ is an usual kernel function (we use an Epanechnikov kernel).¹¹ Statistical properties of such estimators are derived in Li et al. (2013), in particular it is shown that the bandwidth determined by leave-one-out least-squares cross-validation has the optimal order $n^{-1/5}$. Note that an estimate of the function φ defined in (4.1) is obtained by the corresponding quantiles of the cdf $\hat{F}_{X^1|W}$.

Theorem 2.1 in Li et al. (2013) indicates that the error of estimation $(\hat{V}_i - V_i)$ has an asymptotic normal distribution, with a bias term and a variance that have rather complicated expressions, but we could use the bootstrap to evaluate for each $i = 1, \dots, n$ a probability interval of level γ (e.g. $\gamma = 0.95$) for V_i . We should use here the bias-corrected percentile method (see Efron and Tibshirani, 1993) to account for the bias term and to achieve intervals included between the natural bounds $[0, 1]$.

Once the latent heterogeneity factor has been estimated, we can use the values \hat{V}_i as an additional variable (like the observed external factor Z_i), and as shown in Simar et al. (2016), the fact that we use \hat{V}_i in place of V_i does not affect the asymptotic statistical properties of the nonparametric frontier estimators, nor of the resulting estimators of the conditional efficiency measures such as $\hat{\beta}(x, y; 0, d_x, d_y | z, v)$, computed from the sample $\{(X_i, Y_i, Z_i, \hat{V}_i)\}_{i=1}^n$, where $d_x = 0$ since we have chosen the output orientation.

The effect of (Z_i, V_i) on the efficiency measures is an empirical question. First we can test the separability assumption for (Z_i, V_i) (does the boundary of the attainable set depends on (z, v)) and in a second stage we can analyze the links between the conditional efficiency scores with (Z_i, V_i) , by using appropriate nonparametric regressions (see e.g. Daraio and Simar, 2014).

In general setups, for testing separability by using directional distances we suggest taking a fixed direction d (that may contain some zeros for inactive variables). This allows to give an interesting interpretation of the test statistics derived in Daraio et al. (2018). By doing so, the directional distances may be interpreted at a constant (the inverse of the norm of the direction vector, $\|d\|$ which does not depend on (x, y)) as the Euclidean distance between the point under evaluation and its projection in the direction d on the efficient frontier. We have $\beta(x, y; d_x, d_y) = \|d\|^{-1} \|\Psi^d(x, y) - (x, y)\|$ and similarly $\beta(x, y; d_x, d_y | z, v) = \|d\|^{-1} \|\Psi^{d, z, v}(x, y) - (x, y)\|$. So the test statistics we use for the test (see Daraio et al., 2018) is an estimator of $\mathbb{E}_{XYZV}(\beta(X, Y; d_x, d_y)) - \mathbb{E}_{XYZV}(\beta(X, Y; d_x, d_y | Z, V))$ (where for the first term, the expectation in Z, V is just an abuse of notation since $\beta(X, Y; d_x, d_y)$ does not depend on Z, V). This quantity can be interpreted as a constant multiplied by the expected value of the Euclidean distances between the projections of random (X, Y, Z, V) on the unconditional and on the conditional frontiers. We reject the null hypothesis (separability: (Z, V) has no influence on the frontier) if an estimator of this expected distance is too large.

For practical application, first split the sample $S_n = \{(X_i, Y_i, Z_i, \hat{V}_i)\}_{i=1}^n$ randomly into two independent sub-samples, S_{1, n_1}, S_{2, n_2} such that

¹⁰ The Matlab code for the implementation and extension of our approach in a broader set of application contexts is available upon requests to the authors. All the computations can also be performed using the current version of the FEAR package for R described by Wilson (2008).

¹¹ As noted by an anonymous referee, other kernels with compact support could be used like the biweight or the triweight kernels, as described in Silverman (1986) and Fan and Gijbels (1996), since they put less mass on the boundary of the windows. However, the Epanechnikov shares some optimality properties for density estimation and regression problems.

$n_1 = \lfloor n/2 \rfloor$, $n_2 = n - n_1$, $S_{1,n_1} \cup S_{2,n_2} = S_n$, and $S_{1,n_1} \cap S_{2,n_2} = \emptyset$. The n_1 observations in S_{1,n_1} are used for the unconditional estimates, while the n_2 observations in S_{2,n_2} are used for the conditional estimates.

After splitting the sample, compute for the chosen direction $d = (d_x, d_y)$, the estimators

$$\hat{\mu}_{n_1} = n_1^{-1} \sum_{(X_i, Y_i) \in S_{1,n_1}} \hat{\beta}(X_i, Y_i; d | S_{1,n_1}) \tag{A.2}$$

and

$$\hat{\mu}_{c,n_2,h} = n_2^{-1} \sum_{(X_i, Y_i, Z_i, \hat{V}_i) \in S_{2,n_2,h}^*} \hat{\beta}(X_i, Y_i; d | Z_i, \hat{V}_i, S_{2,n_2}), \tag{A.3}$$

where $S_{2,n_2,h}^*$ in (A.3), is a random subsample from S_{2,n_2} of size $n_{2,h} = \min(n_2, n_2 h^{r+1})$. Here to simplify the notation, h^{r+1} denotes the product of the bandwidths for the $r + 1$ conditioning variables (Z_i, \hat{V}_i) obtained by least squares cross validation when computing the estimator of $H_{X,Y|Z,V}$. Consistent estimators of the variances in the two independent samples are given by

$$\hat{\sigma}_{n_1}^2 = n_1^{-1} \sum_{(X_i, Y_i) \in S_{1,n_1}} (\hat{\beta}(X_i, Y_i; d | S_{1,n_1}) - \hat{\mu}_{n_1})^2 \tag{A.4}$$

and

$$\hat{\sigma}_{c,n_2,h}^2 = n_2^{-1} \sum_{(X_i, Y_i, Z_i, \hat{V}_i) \in S_{2,n_2,h}^*} (\hat{\beta}(X_i, Y_i; d | Z_i, \hat{V}_i, S_{2,n_2}) - \hat{\mu}_{c,n_2,h})^2 \tag{A.5}$$

(respectively), where the full (sub)samples are used to estimate the variances.

Now the final form of test statistics depends on the value of $p + q$. As explained below, in our application we will use the FDH estimators so the rate of convergence is n^κ , where $\kappa = 1/(p + q)$.¹² Then, if $\kappa \geq 1/3$,

$$T_{1,n} = \frac{(\hat{\mu}_{n_1} - \hat{\mu}_{c,n_2,h}) - (\hat{B}_{\kappa,n_1} - \hat{B}_{\kappa,n_2,h}^c)}{\sqrt{\frac{\hat{\sigma}_{n_1}^2}{n_1} + \frac{\hat{\sigma}_{c,n_2,h}^2}{n_{2,h}}}} \xrightarrow{\mathcal{L}} N(0, 1) \tag{A.6}$$

under the null. Alternatively, for larger values of $p + q$, when $\kappa < 1/2$,

$$T_{2,n} = \frac{(\hat{\mu}_{n_{1,\kappa}} - \hat{\mu}_{c,n_{2,h,\kappa}}) - (\hat{B}_{\kappa,n_1} - \hat{B}_{\kappa,n_2,h}^c)}{\sqrt{\frac{\hat{\sigma}_{n_{1,\kappa}}^2}{n_{1,\kappa}} + \frac{\hat{\sigma}_{c,n_{2,h,\kappa}}^2}{n_{2,h,\kappa}}}} \xrightarrow{\mathcal{L}} N(0, 1) \tag{A.7}$$

under the null, where $n_{1,\kappa} = \lfloor n_1^{2\kappa} \rfloor$ with $\hat{\mu}_{n_{1,\kappa}} = n_{1,\kappa}^{-1} \sum_{(X_i, Y_i) \in S_{n_{1,\kappa}}^*} \hat{\beta}(X_i, Y_i; d | S_{n_1})$, and $S_{n_{1,\kappa}}^*$ is a random subsample of size $n_{1,\kappa}$ taken from S_{n_1} . For the conditional part, we have similarly and as described in the preceding section, $n_{2,h,\kappa} = \lfloor n_{2,h}^{2\kappa} \rfloor$, with $\hat{\mu}_{c,n_{2,h,\kappa}} = n_{2,h,\kappa}^{-1} \sum_{(X_i, Y_i, Z_i, \hat{V}_i) \in S_{n_{2,h,\kappa}}^*} \hat{\beta}(X_i, Y_i; d | Z_i, \hat{V}_i, S_{n_2})$ where $S_{n_{2,h,\kappa}}^*$ is a random subsample of size $n_{2,h,\kappa}$ from S_{n_2} . Here the terms \hat{B}_{κ,n_1} and $\hat{B}_{\kappa,n_2,h}^c$ are estimators of the corresponding bias correction. They are obtained by a generalized jackknife method described in Daraio et al. (2018); without these bias corrections, the above results do not hold (the limiting normal distributions will have an unknown mean different from zero).

Given a random sample S_n , one can compute values $\hat{T}_{1,n}$ or $\hat{T}_{2,n}$ depending on the value of $(p + q)$.¹³ The null should be rejected whenever $1 - \Phi(\hat{T}_{1,n})$ or $1 - \Phi(\hat{T}_{2,n})$ is less than the desired test size, e.g., 0.1, 0.05, or 0.01, where $\Phi(\cdot)$ denotes the standard normal distribution function.¹⁴

B. Recovering gaps in original units

Due to the small number of observations and the high correlations between the inputs or between the outputs, Daraio and Simar (2007) have suggested a way to reduce the dimension of the input/output space by using tools of factor analysis. Wilson (2018) has shown in an extensive Monte-Carlo analysis the advantages of these methods from an econometric point of view (reduction of bias, MSE, etc.). We used in our application this tool for reducing the 3 inputs in one input factor FX and to reduce two of the 3 outputs in one output factor. So at the end we work to estimate efficient frontier and efficiency measures one input $X = FX$ and 2 outputs $Y = (TDEG, FY)$. Since we are in an output oriented framework $d_x = 0$ and $d_y = med(Y)$ the resulting estimated gaps (distance from the units to their projection on the efficient frontier in the given direction) are thus in these transformed units. In our case, we have only gaps in the output space. We summarize here how to rebuild the gaps in the original output units from the gaps on the factor.¹⁵

To be more general consider the set $Y_i^{(A)} \in \mathbb{R}^k$ of the k outputs to be aggregated that define the $(n \times k)$ data matrix $Y^{(A)} = (Y_1^{(A)}, Y_2^{(A)}, \dots, Y_n^{(A)})'$. Let $a \in \mathbb{R}^k$, with $a' a = 1$ be the eigenvector of the 2nd moments matrix of the data $Y^{(A)} Y^{(A)'} corresponding to its largest eigenvalue. The value of the output factor for the i th observation is given by$

$$FY_i = a' Y_i^{(A)}, \text{ for } i = 1, \dots, n. \tag{B.1}$$

¹² For computing the directional distance estimators we used the fast and exact algorithms described in Daraio et al. (2020).

¹³ Note that when $p + q = 3$ we can use both statistics, but it is better to use the test statistics $T_{2,n}$ involving errors of approximation in the underlying Central Limit Theorem of smaller order (see Remark 4.1 in Daraio et al., 2018).

¹⁴ Note that the splitting can be repeated large number of times as proposed in Simar and Wilson (2020).

¹⁵ In case where d_x would be different from zero for an input factor, the procedure would be quite similar.

Geometrically, FY_i is the orthogonal projection of the data point $Y_i^{(A)}$ on the unit vector a .

Now the projection of FY_i on the efficient frontier in the direction d_y is given by

$$FY_i^\delta = FY_i + \beta d_{FY}, \quad (\text{B.2})$$

where $\beta \geq 0$ is the estimation of the used directional distance and d_{FY} is the element of d_y corresponding to the output FY . So the gap for this output for the unit i is given by

$$G_i = FY_i^\delta - FY_i = \beta d_{FY}. \quad (\text{B.3})$$

The question is how to recover a reasonable value of $Y_i^{(A),\delta}$ such that $a'Y_i^{(A),\delta} = FY_i^\delta$. There is no unique solution to this question but we identify a solution by choosing a direction vector in the original units $Y^{(A)}$ to project the point on the efficient frontier. Suppose we select a direction vector $d_{Y^{(A)}} \geq 0$, then the gaps in the original units for the outputs $Y_i^{(A)}$ will be defined as $\delta d_{Y^{(A)}}$ where δ is the unique solution of the equation

$$a' \left[Y_i^{(A)} + \delta d_{Y^{(A)}} \right] = FY_i^\delta. \quad (\text{B.4})$$

The solution is obviously given by

$$\delta = \frac{FY_i^\delta - FY_i}{a'd_{Y^{(A)}}} = \frac{G_i}{a'd_{Y^{(A)}}}. \quad (\text{B.5})$$

Finally, the gaps in the original units in the direction $d_{Y^{(A)}}$ are given by

$$G_{Y_i^{(A)}} = \frac{G_i}{a'd_{Y^{(A)}}} d_{Y^{(A)}}. \quad (\text{B.6})$$

A natural choice for $d_{Y^{(A)}}$ could be $Y_i^{(A)}$ to look to the gaps in a radial way. In this case, one would have

$$G_{Y_i^{(A)}} = \frac{FY_i^\delta - FY_i}{FY_i} Y_i^{(A)}. \quad (\text{B.7})$$

Another natural choice, as in Daraio et al. (2015a, 2015b), is to keep the idea that these outputs are considered along the direction given by FY , i.e. $d_{Y^{(A)}} = a$. In this case, that we follow in our application, since $a'a = 1$, we have

$$G_{Y_i^{(A)}} = G_i a. \quad (\text{B.8})$$

Of course, if before the factor analysis the outputs $Y^{(A)}$ have been scaled by their standard deviations (to be unit free), at the end we have to rescale the gaps $G_{Y_i^{(A)}}$ accordingly.

References

- Alavi, M., Leidner, D.E., 2001. Knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Q.* 107–136.
- Aparicio, J., Lopez-Torres, L., Santin, D., 2018. Economic crisis and public education. A productivity analysis using a Hicks-Moorsteen index. *Econ. Modell.* 71, 34–44.
- Blackmur, D., 2007. The public regulation of higher education qualities: rationale, processes, and outcomes. In: *Quality Assurance in Higher Education*. Springer, Dordrecht, pp. 15–45.
- Bogetoft, P., 2012. *Performance Benchmarking: Measuring and Managing Performance*. Springer Science & Business Media.
- Bonaccorsi, A., Daraio, C. (Eds.), 2007. *Universities and Strategic Knowledge Creation: Specialization and Performance in Europe*. Edward Elgar Publishing.
- Bonaccorsi, A., Daraio, C., Simar, L., 2014. Scale and research specialization in European universities: a directional distance approach to teaching efficiency. In: *Knowledge, Diversity and Performance in European Higher Education*. Edward Elgar Publishing.
- Bryl, L., 2018. Intellectual capital measurement. In: *Encyclopedia of Information Science and Technology*, fourth ed. IGI Global, pp. 5056–5066.
- Bădin, L., Daraio, C., Simar, L., 2012. How to measure the impact of environmental factors in a nonparametric production model. *Eur. J. Oper. Res.* 223 (3), 818–833.
- Bădin, L., Daraio, C., Simar, L., 2014. Explaining inefficiency in nonparametric production models: the state of the art. *Ann. Oper. Res.* 214 (1), 5–30.
- Bădin, L., Daraio, C., Simar, L., 2019. A bootstrap approach for bandwidth selection in estimating conditional efficiency measures. *Eur. J. Oper. Res.* (in press).
- Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *J. Econom.* 106, 1–25.
- Chambers, R.G., Chung, Y.H., Färe, R., 1996. Benefit and distance functions. *J. Econ. Theor.* 70, 407–419.
- Chambers, R.G., Chung, Y.H., Färe, R., 1998. Profit, directional distance functions and nerlovian efficiency. *J. Optim. Theor. Appl.* 98, 351–364.
- Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* 2, 429–444.
- Daouia, A., Gijbels, I., 2011. Robustness and inference in nonparametric partial frontier modeling. *J. Econom.* 161 (2), 147–165.
- Daouia, A., Ruiz-Gazen, A., 2006. Robust nonparametric frontier estimators: qualitative robustness and influence function. *Stat. Sin.* 16 (4), 1233.
- Daraio, C., 2017. A framework for the assessment of Research and its Impacts. *J. Data Inf. Sci.* 2 (4), 7–42 2017.
- Daraio, C., 2018. Nonparametric methods and higher education. In: Teixeira, P., Shin, J. (Eds.), *Encyclopedia of International Higher Education Systems and Institutions*. Springer, Dordrecht.
- Daraio, C., 2019. Econometric approaches to the measurement of research productivity. In: Glänzel, W., Moed, H.F. (Eds.), *Springer Handbook of Science and Technology Indicators*. Schmoeh H. and Thelwall M., pp. 633–666.
- 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., 2007. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer, New-York.
- Daraio, C., Simar, L., 2014. Directional distances and their robust versions: computational and testing issues. *Eur. J. Oper. Res.* 237, 358–369.
- Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B., Bach, L., Bogetoft, P., et al., 2011. The European university landscape: a micro characterization based on evidence from the Aquameth project. *Res. Pol.* 40 (1), 148–164.
- Daraio, C., Bonaccorsi, A., Simar, L., 2015a. Rankings and university performance: a conditional multidimensional approach. *Eur. J. Oper. Res.* 244, 918–930.
- Daraio, C., Bonaccorsi, A., Simar, L., 2015b. Efficiency and economies of scale and specialization in European universities. A directional distance approach. *J. Inf.* 9, 430–448.
- Daraio, C., Simar, L., Wilson, P.W., 2018. Central limit theorems for conditional efficiency measures and tests of the “separability” condition in nonparametric, two-stage models of production. *Econom.* J. 21, 170–191.
- Daraio, C., Simar, L., Wilson, P.W., 2020. Fast and efficient computation of directional distance estimators. *Ann. Oper. Res.* 288, 805–835.
- De Witte, K., Lopez-Torres, L., 2017. Efficiency in education: a review of literature and a way forward. *J. Oper. Res. Soc.* 68 (4), 339–363.
- Deprins, D., Simar, L., Tulkens, H., 1984. Measuring labor inefficiency in post offices. In: Pestieau, M.M.P., Tulkens, H. (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*. North-Holland, Amsterdam, pp. 243–267.
- Dumay, J., Guthrie, J., Puntillo, P., 2015. IC and public sector: a structured literature review. *J. Intellect. Cap.* 16 (2), 267–284.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, London.
- Emrouznejad, A., Yang, G.L., 2017. A Survey and Analysis of the First 40 Years of Scholarly Literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

- Färe, R., Grosskopf, S., Forsund, F.R., Hayes, K., Heshmati, A., 2006. Measurement of productivity and quality in non-marketable services: with application to schools. *Qual. Assur. Educ.* 14 (1), 21–36.
- Färe, R., Grosskopf, S., Margaritis, D., 2008. Efficiency and productivity: malmquist and more. In: Fried, Harold, Knox Lovell, C.A., Schmidt, Shelton (Eds.), *The Measurement of Productive Efficiency*, second ed. Oxford University Press.
- Farrell, M.J., 1957. The measurement of the productive efficiency. *J. R. Stat. Soc. Ser. A, CXX*, Part 3, 253–290.
- Florens, J.P., Simar, L., Van Keilegom, I., 2014. Frontier estimation in nonparametric location-scale models. *J. Econom.* 178, 456–470.
- Fried, H., Lovell, C.A.K., Schmidt, S., 2008. *The Measurement of Productive Efficiency*, second ed. Oxford University Press.
- Grosskopf, S., Hayes, K.J., Taylor, L.L., 2014. Applied efficiency analysis in education. *Econ. Bus. Lett.* 3 (1), 19–26.
- Guthrie, J., Dumay, J., 2015. New frontiers in the use of intellectual capital in the public sector. *J. Intellect. Cap.* 16 (2), 258–266.
- Halkos, G.E., Tzeremes, N.G., 2011. Modelling the effect of national culture on multinational banks' performance: a conditional robust nonparametric frontier analysis. *Econ. Modell.* 28 (1–2), 515–525.
- Halkos, G.E., Tzeremes, N.G., 2013. A conditional directional distance function approach for measuring regional environmental efficiency: evidence from UK regions. *Eur. J. Oper. Res.* 227 (1), 182–189.
- Harvey, L., Green, D., 1993. Defining quality. *Assess Eval. High Educ.* 18 (1), 9–34.
- 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., 2015. Operational research in education. *Eur. J. Oper. Res.* 243 (3), 683–696.
- Johnson, A.L., Ruggiero, J., 2014. Nonparametric measurement of productivity and efficiency in education. *Ann. Oper. Res.* 221 (1), 197–210.
- Jones, N., Meadow, C., Sicilia, M.A., 2009. Measuring intellectual capital in higher education. *J. Inf. Knowl. Manag.* 8 (2), 113–136.
- Kempkes, G., Pohl, C., 2010. The efficiency of German universities—Some evidence from nonparametric and parametric methods. *Appl. Econ.* 42 (16), 2063–2079.
- Klump, M., 2015. Efficiency in Higher Education: Requirements, Theory, Methods and Decision Areas. *Theory and Method in Higher Education Research*, pp. 93–118.
- Kneip, A., Simar, L., Wilson, P.W., 2016. Testing hypotheses in nonparametric models of production. *J. Bus. Econ. Stat.* 34, 435–456.
- Kucharčíková, A., Tokarčíková, E., Blašková, M., 2015. Human capital management aspect of the human capital efficiency in university education. *Procedia-Soc. Behav. Sci.* 177, 48–60.
- Lane, J.E., 2000. *New Public Management*. Taylor & Francis US.
- Lee, H., Kim, C., 2014. Benchmarking of service quality with data envelopment analysis. *Expert Syst. Appl.* 41 (8), 3761–3768.
- Lewis, R.G., Smith, D.H., 1994. *Total Quality in Higher Education*. Total Quality Series. St. Lucie Press, 100 E. Linton Blvd. Suite 403B, Delray Beach, FL 33483.
- Li, Q., Maasoumi, E., Racine, J.S., 2009. A nonparametric test for equality of distributions with mixed categorical and continuous data. *J. Econom.* 148, 186–200.
- Li, Q., Lin, J., Racine, J.S., 2013. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *J. Bus. Econ. Stat.* 31 (1), 57–65.
- Liebowitz, J. (Ed.), 2012. *Knowledge Management Handbook: Collaboration and Social Networking*. CRC Press.
- Lu, W.M., 2012. Intellectual capital and university performance in Taiwan. *Econ. Modell.* 29 (4), 1081–1089.
- Manatos, M.J., Sarrico, C.S., Rosa, M.J., 2016. The integration of quality management in higher education institutions: a systematic literature review. *Total Qual. Manag. Bus. Excel.* 28 (1), 159–175.
- Mastromarco, C., Simar, L., 2018. Globalization and productivity: a robust nonparametric world frontier analysis. *Econ. Modell.* 69, 134–149.
- Matousek, R., Tzeremes, N.G., 2016. CEO compensation and bank efficiency: an application of conditional nonparametric frontiers. *Eur. J. Oper. Res.* 251 (1), 264–273.
- Matzkin, R.L., 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71 (5), 1339–1375.
- Minviel, J.J., De Witte, K., 2017. The influence of public subsidies on farm technical efficiency: a robust conditional nonparametric approach. *Eur. J. Oper. Res.* 259 (3), 1112–1120.
- Moed, H.F., 2017. *Applied Evaluative Informetrics*. Springer International Publishing.
- Neely, A. (Ed.), 2002. *Business Performance Measurement: Theory and Practice*. Cambridge University Press.
- Neely, A., Gregory, M., Platts, K., 1995. Performance measurement system design: a literature review and research agenda. *Int. J. Oper. Prod. Manag.* 15 (4), 80–116.
- Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M., Kennerley, M., 2000. Performance measurement system design: developing and testing a process-based approach. *Int. J. Oper. Prod. Manag.* 20 (10), 1119–1145.
- Nigsch, S., Schenker-Wicki, A., 2015. Frontier efficiency analysis in higher education. In: *Incentives and Performance*. Springer, Cham, pp. 155–170.
- Paradeise and Thoenig, 2015. In: *Search of Academic Quality*. Springer.
- Rhaim, M., 2017. Measurement and determinants of academic research efficiency: a systematic review of the evidence. *Scientometrics* 110 (2), 581–615.
- Ruggiero, J., 1996. On the measurement of technical efficiency in the public sector. *Eur. J. Oper. Res.* 90, 553–565.
- Sallis, E., 2002. *Total Quality Management in Education*. Kogan Page Ltd London.
- Sarrico, C., 2018a. Performance indicators in higher education. In: Teixeira, P., Shin, J. (Eds.), *Encyclopedia of International Higher Education Systems and Institutions*. Springer, Dordrecht.
- Sarrico, C., 2018b. Performance and quality management in higher education. In: Teixeira, P., Shin, J. (Eds.), *Encyclopedia of International Higher Education Systems and Institutions*. Springer, Dordrecht.
- Sarrico, C.S., Rosa, M.J., Teixeira, P.N., Cardoso, M.F., 2010. Assessing quality and evaluating performance in higher education: worlds apart or complementary views? *Minerva* 48 (1), 35–54.
- Sav, T.G., 2012. Stochastic cost inefficiency estimates and rankings of public and private research and doctoral granting universities. *J. Knowl. Manag. Econ. Inf. Technol.* 4 (3), 11–29.
- Schimank, U., Winnes, M., 2000. Beyond Humboldt? The relationship between teaching and research in European university systems. *Sci. Publ. Pol.* 27 (6), 397–408.
- Secundo, G., Lombardi, R., Dumay, J., 2018. Intellectual capital in education. *J. Intellect. Cap.* 19 (1), 2–9.
- Seth, N., Deshmukh, S.G., Vrat, P., 2005. Service quality models: a review. *Int. J. Qual. Reliab. Manag.* 22 (9), 913–949.
- Silverman, B.W., 1986. *Density Estimation and Data Analysis*. Chapman and Hall, New York.
- Simar, L., Vanhems, A., 2012. Probabilistic characterization of directional distances and their robust versions. *J. Econom.* 166, 342–354.
- Simar, L., Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *J. Econom.* 136 (1), 31–64.
- Simar, L., Wilson, P.W., 2011. Two-stage DEA: *caveat emptor*. *J. Prod. Anal.* 36 (2), 205.
- Simar, L., Wilson, P.W., 2015. Statistical approaches for non-parametric frontier models: a guided tour. *Int. Stat. Rev.* 83, 77–110.
- Simar, L., Wilson, P.W., 2020. Hypothesis testing in nonparametric models of production using multiple sample splits. *J. Prod. Anal.* 53 (3), 287–303.
- Simar, L., Vanhems, A., Wilson, P.W., 2012. Statistical inference with DEA estimators of directional distances. *Eur. J. Oper. Res.* 220, 853–864.
- Simar, L., Vanhems, A., Van Keilegom, I., 2016. Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. *J. Econom.* 190, 360–373.
- Tiwana, A., 2000. *The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System*. Prentice Hall PTR.
- Van Dooren, W., Bouckaert, G., Halligan, J., 2015. *Performance Management in the Public Sector*. Routledge, New York.
- Varabyova, Y., Schreyögg, J., 2017. Integrating quality into the nonparametric analysis of efficiency: a simulation comparison of popular methods. *Ann. Oper. Res.* 1–28.
- Verschele, M., Rogge, N., 2012. An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: evidence from a conditional Data Envelopment Analysis approach. *Eur. J. Oper. Res.* 223 (1), 214–225.
- Vidaillet, B., 2013. *Évaluez-moi! Évaluation au travail: les ressorts d'une fascination*. Le Seuil.
- Warning, S., 2004. Performance differences in German higher education: empirical analysis of strategic groups. *Rev. Ind. Organ.* 24 (4), 393–408.
- Westerheijden, D.F., Stensaker, B., Rosa, M.J. (Eds.), 2007. *Quality Assurance in Higher Education: Trends in Regulation, Translation and Transformation*, vol. 20. Springer Science & Business Media.
- Williams, R., de Rassenfosse, G., 2018. Quality of higher education systems. In: Teixeira, P., Shin, J. (Eds.), *Encyclopedia of International Higher Education Systems and Institutions*. Springer, Dordrecht.
- Wilson, P.W., 2008. FEAR: a software package for frontier efficiency analysis with R. *Soc. Econ. Plann. Sci.* 42 (4), 247–254.
- Wilson, P.W., 2018. Dimension reduction in nonparametric models of production. *Eur. J. Oper. Res.* 267, 349–367.
- Wolszczak-Derlacz, J., 2017. An evaluation and explanation of (in) efficiency in higher education institutions in Europe and the US with the application of two-stage semi-parametric DEA. *Res. Pol.* 46 (9), 1595–1605.
- 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.C., 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*, vol. 213. Springer.