Journal of Informetrics 9 (2015) 430–448

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Efficiency and economies of scale and specialization in European universities: A directional distance approach

Cinzia Daraio[a,*,1], Andrea Bonaccorsi[b,2], Léopold Simar[a,c,3]

[a] Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), University of Rome "La Sapienza", Via Ariosto, 25, 00185 Roma, Italy
[b] Department of Energy and Systems Engineering, University of Pisa and Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR), Italy
[c] ISBA, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 12 May 2014
Received in revised form 10 March 2015
Accepted 10 March 2015
Available online 9 April 2015

Keywords:
Efficiency
National academic systems
Disciplinary specialization
Research performance
Teaching and research
Nonparametric and robust frontier estimation

ABSTRACT

In this paper we investigate economies of scale and specialization of European universities. The proposed approach builds on the notion that university production is a multi-input multi-output process different than standard production activity. The analyses are based on an interesting database which integrates the main European universities data on inputs and outputs with bibliometric data on publications, impact and collaborations. We pursue a cross-country perspective; we include subject mix and introduce a robust modeling of production trade-offs. Finally we test the statistical significance of scale and specialization and find that they both have a significant impact on the efficiency of the Humboldt model. Nevertheless, confirming previous findings, specialization has not a significant impact on the efficiency of the research model.

1. Introduction

This paper addresses two contested issues that are at the core of recent debates in higher education and makes the argument that, in order to address them sensibly, there is a need for the integration of existing data and for new elaboration techniques. Thus, although the ultimate issue is a policy one, the approach we suggest makes use of an integrated dataset at European level and applies new techniques.[4] To be more precise: we argue that without an investment into data integration (including data retrieval and data cleaning) and new informetrics, these policy issues cannot be addressed appropriately.

* Corresponding author. Tel.: +39 0677274068.
E-mail addresses: daraio@dis.uniroma1.it (C. Daraio), a.bonaccorsi@gmail.com (A. Bonaccorsi), leopold.simar@uclouvain.be (L. Simar).
[1] Financial support from the "Progetto di Ricerca di Ateneo 2013" of the University of Rome "La Sapienza" is gratefully acknowledged.
[2] The opinions expressed in this paper are personal and do not involve the institution of the second author (ANVUR).
[3] Financial support from the "Inter-university Attraction Pole", Phase VII (No. P7/06) of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.
[4] In Bonaccorsi, Daraio, and Simar (2013) we analyse the impact of scale and specialization on the research efficiency of European universities. In this paper we extend the analysis including additional bibliometric indicators such as Normalized Impact, High Quality Publications, Excellence Rate and International Collaborations. Moreover, we test the impact of scale and specialization by applying state of the art approaches (Daraio & Simar, 2014).

http://dx.doi.org/10.1016/j.joi.2015.03.002
1751-1577/

The two issues under discussion can be formulated as follows:

(a) how does the size of universities influence their efficiency? In other terms, are there economies of scale in higher education?
(b) is there a need to reconsider the main organizational model of universities, which is predominantly based on generalist institutions covering many disciplines? In other terms, are there economies of specialization in higher education?

These two questions come after the higher education system, in advanced countries, has reached the point of massification (i.e. enrolment rates exceeding 50% of the relevant age cohort), while the public budget has not grown correspondingly. Universities are put under pressure to use existing resources, namely staff and funding, in the most efficient way. At the same time there is an increased pressure from the research side: the expectations of society and policy makers on the contribution of research to societal problems have grown significantly, there are new entrants in scientific arena (particularly from Asia) and the competition for funding has increased sharply. This situation creates a classical issue in public policy: we have two valuable goals (serving better mass educational needs and producing good research) between which there is tension. The trade-off between the two goals would require a grounded theory of production, which can be framed in the economic language. If we assume that universities are units of production, then these issues require investigating the existence and importance of economies of scale and specialization. Do we need to increase the size of universities, in order to enhance their efficiency? Do we need to increase the specialization of universities, favoring focused institutions (e.g. technical universities, medical schools, business schools) against the more traditional generalist institutions, covering many unrelated disciplines?

The paper is organized as follows.

In the next section, the relevant literature as well as the main research questions addressed in the paper are outlined. Section 3 describes the main data used in the analysis, providing details on the integration of the different sources. Section 4 provides a simplified graphical illustration of university's activities and their trade-offs. Section 5 provides the methodological background, while Section 6 reports the main results and Section 7 concludes the paper. Appendix A describes the factorial analysis conducted on the data and provides some details on the calculation of gaps.

2. Economies of scale and specialization in higher education

2.1. General introduction

In this section we offer a short and focused survey of the literature.

Economies of scale refer to the reduction of cost per unit of output when the size of operations increases, mainly due to the reduction of unitary fixed costs, but often due also to lower variable costs.

Economies of specialization arise when the cost of producing a specific good by a specialized firm is lower than the cost of the same good made by a firm which produces together two or more goods.

Before entering into the details, let us remind that the issue of economies of scale and specialization can be addressed according to two different approaches.

The first has worked directly with cost functions as the dual of production functions. Here the main difficulty has been the modeling of a production function which is, by definition, not only multi-input (as any production function), but also multi-output. The traditional econometric techniques used to estimate economies of scale in a monoproduct setting were clearly inadequate. After the introduction of a full scale theory of the multi-product firm (Baumol, Panzar, & Willig, 1982), several appropriate econometric techniques have been introduced (see Bonaccorsi & Daraio, 2003, 2004 for an overview).

The second approach is based on the estimation of technical efficiency of the units under analysis, namely the best use of resources (inputs) to realize their outputs. In this line of research, the existence and magnitude of economies of scale and specialization is derived from the difference between the efficiency scores of observed Decision Making Units (DMUs) and the scores that would be obtained if the inputs (and/or outputs) were aggregated. In nonparametric efficiency analysis, traditionally based on a Data Envelopment Analysis (DEA) approach (see e.g. Färe, Grosskopf, & Lovell, 1994), economies of specialization are computed on the base of the comparison of the frontier of specialized firms and the frontier of multiproduct firm constructed from the sum of specialized firms. This approach, however, introduces in the analysis additional assumptions (which rely e.g. on the convexity and additional assumptions on the hypothetical firm, and the sample size bias). Recent works in efficiency analysis (see e.g. Daraio & Simar, 2007) propose the conditional nonparametric analysis to investigate the impact of scale and specialization, which are considered as external–environmental factors that are neither inputs nor outputs under the control of the DMU, but might influence the performance of the units. In this paper we follow the foregoing approach, extending the efficiency methodology to robust and conditional directional distances and implementing a recently introduced test (Daraio & Simar, 2014), based on the bootstrap, to assess the significance of scale and specialization impact.

2.2. Research questions

In the following we proceed with the description of the literature and explicit the main research questions we address in the paper.

2.2.1. Are larger universities more efficient?

It is not surprising that a large literature has addressed the issue of economies of scale in higher education. Brinkman and Leslie (1986) review the first 60 years of empirical studies, most of which from United States. After almost 20 years, Cohn and Cooper (2004) have offered a comprehensive survey of findings from the cost function perspective, while Johnes (2006) has reviewed the technical efficiency literature. In general the literature has addressed the issue of increasing returns to scale in the two core production processes of universities, namely teaching and research.

Teaching is a complex process, whose technology is yet poorly understood. As several authors have noted (e.g. Hanushek, 1986; Johnes, 2006; Worthington, 2001), we really do not have a full scale theory of higher education teaching. Teaching is subject to economies of scale since expanding the size of the class of students expands the output (number of students attending a lecture) while keeping constant the input (the lecturing staff). At the same time teaching also require one-to-one interaction with students, such as examinations and tutoring, for which costs are roughly proportional to the output. The exact combination between these two opposite forces is responsible for the overall effect. As a matter of fact, the existence of economies of scale in undergraduate teaching is largely established in the literature (Cohn, Rhine, & Santos, 1989; Dundar & Lewis, 1995; Glass, Mckillop, & Hyndman, 1995; Hashimoto & Cohn, 1997; Koshal & Koshal, 1999; Laband & Lentz, 2003).

Research is an even less understood production process, for which the arguments for economies of scale are mostly linked to indivisibilities in cognitive capital (minimum scale of research teams) and above all in physical capital (scientific instrumentation). A dedicated literature has examined this issue repeatedly and has been reviewed by SPRU (Science Policy Research Unit, Sassex University) in the early 2000s at a request of the UK government (Von Tunzelmann, Ranga, Martin, & Geuna, 2003). The overall synthesis was that we do not have compelling evidence on the positive impact of the size of research organizations on scientific productivity.

It has also been noted that size may be associated to other factors, such as the pressure for visibility and the quality of the intellectual environment (Bonaccorsi & Daraio, 2005; Qureshi, Hlupic, de Vreede, Briggs, & Nunamaker, 2003; Seglen & Aksnes, 2000). More recently, Carayol and Matt (2006) have stressed that it is not size per se but the adoption of policies for the recruitment of high quality researchers that make a difference. Horta and Lacy (2011) have found that researchers in larger research units have indeed a larger network of scientific contacts and tend to publish more at the international level. Combining the two production processes, a summary of findings from Brinkman and Leslie (1986) is that economies of scale in higher education are pervasive, although they tend to be exhausted at a relatively small scale, in the order of 1000 full time equivalent (FTE) students. Confirming the survey from Brinkman and Leslie (1986) and the results of Cohn et al. (1989), Johnes (2006) find economies of scale at the level of university, but claim that they are exhausted at relatively small size. These results are generally confirmed by stating that the main sources of economies of scale for universities come from undergraduate education, while research contributes little to increasing returns or even is subject to decreasing returns, with postgraduate education somewhat in the middle. Recently Brandt and Schubert (2014), by using data on Germany, show that research is subject to diminishing returns to scale at the level of research team. At the same time, universities offer an umbrella to research teams which is subject to increasing returns to scale, due to shared infrastructures, better efficiency in administrative activities and reputational effects. This might explain the dominant organizational model of universities, based on a number of semi-autonomous research teams, which however accept to operate under the administrative umbrella of universities.

2.2.2. Are specialised universities more efficient?

It is well known that Europe has invented the modern organizational model of universities, called Humboldtian model (Schimank & Winnes, 2000). There are two elements in the model: the coexistence of teaching and research, and the generalist orientation, namely the coexistence of many disciplines within the same institutional umbrella. The coexistence of teaching and research can be explained, in economic terms, with the existence of economies of scope, given that the same staff can produce both outputs and may optimize the use of time budgets by alternating these activities. Most studies on economies of scope between research and teaching confirm this assumption (Johnes, 2004; Longlong, Fengliang, & Weifang, 2009), although it is possible that after a certain level, heavy teaching loads reduce scientific productivity (Izadi, Johnes, Oskrochi, & Crouchley, 2002; Worthington & Higgs, 2011). Much less explored is the issue we analyze here, that is, whether there is an impact on efficiency from the specialization of universities, that is, the orientation to do research and teaching in a few areas, as in specialized universities, as opposed to the traditional model of broad coverage. This is a relatively unexplored issue. Is there an advantage in doing research in a field, provided that there are other fields under the institutional and administrative umbrella of the same university? In Bonaccorsi and Daraio (2007) an investigation on the generalist model of European universities is offered, based on a descriptive analysis. No explanation is given for the prevalence of the generalist model. In this paper we employ directional distance techniques to explore whether efficiency is influenced by various degrees of specialization, using a quantitative variable, which is more informative than previous descriptive analysis. In particular, following Lopez-Illescas, de Moya-Anegon, and Moed (2011), we use the Gini index of institution's disciplinary specialization to characterize generalist versus specialist universities.

2.3. Generalizability and policy relevance of results

While the existing literature, briefly presented in the previous section, deliver a rich array of implications, it mostly comes from country-level studies. Therefore they are subject to serious problems of generalizability, which is a major concern for policy making if decisions must be made based on the evidence of other, poorly comparable, institutional contexts. In addition, existing studies do not offer separate analyses by disciplinary fields. The first wave of studies has been dealing with USA and Anglosaxon countries, partly due to better availability of data, partly as a consequence of major structural reforms of the university system starting in the 1980s in countries such as the United Kingdom, New Zealand, Australia and Canada.

The dominance of Anglosaxon countries in the literature creates an issue of generalizability. The issue at stake is not, as it is often stated, the role of the private sector, which is instead marginal, for example, in UK or Canada. The issue is that, according to OECD, these countries have an institutional framework and labour market conditions that allow a much higher mobility of inputs, such as staff, as well mobility of students. In addition, the autonomy of universities in recruitment decisions is quite high. Placed under conditions of competition for funding, it is likely that universities in these countries enjoy more room for structural adaptation. Not surprisingly, almost all studies on UK and Australia concluded that universities operate at fairly high levels of efficiency. Among multi-country studies the generalizability is still limited, either because of a small set of countries, or because of small numbers of country observations. An example of study with a cross-country perspective is Joumady and Ris (2005), which is however based on a survey of graduates across European countries. Bonaccorsi and Daraio (2007) examined a dozen of countries based on data coming from the Aquameth project, the first research project that collected comparable data on European universities (see Bonaccorsi & Daraio, 2007; Daraio & Bonaccorsi, 2011).

There are also limits in generalizability due to disciplinary differences. Dundar and Lewis (1995) argued that without a careful distinction among disciplines it is impossible to derive meaningful implications. According to them 'the most important problem seems to be that different production technologies among academic disciplines may generate problems in analyzing departmental cost functions. For instance, results can be quite misleading if a single cost function is estimated for both chemistry and English departments because they have quite dissimilar production functions' (Dundar & Lewis, 1995, p. 120). The impact of disciplinary specialization on university performance has been also analysed in Lopez-Illescas et al. (2011) and in Moed, de Moya-Anegon, Lopez-Illescas, and Visser (2011) that rightly emphasized that subject mix should be taken into account in the assessment of university performance.

This paper builds upon the first studies that have explicitly adopted a multi-country perspective (Daraio & Bonaccorsi, 2011), benefiting from the construction of the Eumida dataset (Bonaccorsi, Daraio, & Simar, 2014; Daghbashyan, Deiaco, & McKelvey, 2014). Moving further in the direction of generalizability, this paper also introduces, although only partially, a cross-discipline perspective. This will be done in the modeling part below in which we use the specialization index (SPEC), that is a proxy of the wideness of activities carried out.

3. Data

3.1. Description of datasets

We exploit a large database, recently constructed by the EUMIDA Consortium (European Universities Micro Data, EUMIDA, 2010) under a European Commission tender, supported by DG EAC (Directorate General for Education and Culture), DG RTD (Directorate General for Research and Innovation), and Eurostat.

This database is based on official statistics produced by National Statistical Authorities in all 25 EU countries (with the exception of France and Denmark) plus Norway and Switzerland. The EUMIDA project, relying on the results of the Aquameth project (Bonaccorsi & Daraio, 2007; Daraio et al., 2011) collected two data sets. Data Collection 1 (DC 1) collected a set of uniform variables on all 2457 higher education institutions that are active in graduate and postgraduate education (i.e. universities), but also in vocational training.[5] Accordingly, all institutions delivering ISCED (International Standard Classification of Education) 5a and 6 degrees are included, and the subset of those delivering ISCED 5b degrees that have a stable organization (i.e. mission, budget, staff). Those institutions altogether constitute the perimeter of higher education institutions (HEIs) in Europe.

Data Collection 2 (DC 2) instead included a larger set of variables on the 850 research active institutions that are also doctorate awarding[6]. Interestingly, the number of HEIs research active is 1364, but only 850 of these are also doctorate awarding institutions. This means that a significant portion of research active institutions is found outside the traditional perimeter of universities, that is in the domain of non-university research (particularly in countries with dual higher education systems).

Data refer to 2008, or to 2009 in some cases.

[5] These data are available at: http://datahub.io/it/dataset/eumida (accessed 12 November 2014).

[6] These data are not publicly available. They are available for research purpose only to the Eumida project team.

We integrate the EUMIDA data, in particular the DC 2 dataset, with the Scimago data (Scimago Institutional Rankings, SIR World Report 2011, period analyzed 2005–2009) which include institutions having published at least 100 scientific documents of any type, that is, articles, reviews, short reviews, letters, conference papers, etc., during the period 2005–2009 as collected by Scopus database. From Scimago data we used the following variables:

- number of publications in Scopus (PUB);
- Specialization index (SPEC) of the university that indicates the extent of thematic concentration/dispersion of an institution's scientific output; its values range between 0 and 1, indicating generalist vs. specialized institutions respectively. This indicator is computed according to the Gini Index and in our analysis it is used as a proxy of the specialization of the university. We follow previous bibliometric studies by Lopez-Illescas et al. (2011) and Moed et al. (2011) that showed the usefulness of categorizing universities in generalist versus specialist by means of the Gini index. See also Egghe and Rousseau (1990) for more details on disciplinary specialization indices.
- International Collaboration (IC), a university's output ratio produced in collaboration with foreign institutions.
- High Quality Publications (Q1), a university's ratio of publications published in the first quartile (25%) in their categories, according to the Scimago journal rank indicator.
- Normalized Impact (NI), it shows the relation between an institution's average scientific impact and the world average (that is set to one).
- Excellence Rate (EXC), it is the percentage of publications included in the 10% of the most cited papers in their respective scientific fields.

3.2. Data integration

The integration of the previously described databases has been carried out within the activities of the Smart.CI.EU (Sapienza microdata architecture for education, research and technology studies. A Competence-based data Infrastructure on European Universities). Smart.CI.EU is an experimental data infrastructure created within a research project funded by Sapienza University of Rome and owned at the Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome. Its creation has been made possible by the integration of several data sources coming from different projects.

The matching of the EUMIDA and Scimago databases has been completed in two stages:

- *automatic* matching between the fields "organisation" contained in Scimago world rank 2011 and the fields "institution name" and "English institutions name" from EUMIDA dataset;
- *manual* matching of additional institutions whose denominations were slightly different and not recognised automatically but were clearly recognisable on the base of expert knowledge. Moreover, several cases of changed names (as EUMIDA data refer to year 2008) have been checked case by case from institutions' website.

All the organisations comprised in the higher education sector in Scimago dataset for the countries covered by EUMIDA were matched, with the exception of some institutions for which the lack of some information made it impossible to do the matching, such as one university in Cyprus (because EUMIDA dataset includes only the Greek part of the island); five universities in Spain; four organisations in UK; one college in Ireland; one military university in Poland; five *institutos politecnicos* in Portugal; three universities in Romania.

In a few cases (that are Linnaeus University in Sweden and Aalto University and University of Eastern Finland in Finland) organisations included in the Scimago dataset are the results of a process of merger of two or more HEIs included in EUMIDA. In these cases the total number of publications (output) has been attributed to EUMIDA HEIs proportionally to their respective number of academic staff, while qualitative indicators have been set the same for all.

About university hospitals which are labeled under a separate sector in the Scimago dataset, they were not integrated with the related university organisation in order to avoid distortion due to different organisational setting in different countries. This because it is difficult to disentangle which part of hospital staff and expenditures are included in the related university balance account. Furthermore there is no unambiguous list of university hospitals in Europe and this creates additional problems.

Summing up, as the description reported above clearly shows, the combination of different databases requires several steps and assumptions and should be done carefully. In particular in the Scimago database sometimes a university hospital is separated from its university while in others it is not. Similar problems might arise for those governmental agencies that are strictly linked with universities. It may happen in fact that national authorities that provide data for Eumida dataset will include the personnel counts in the ACSTAF variable while it is not clear which approach is taken by Scimago.[7] Such discrepancies may introduce distortion in the results of the analysis carried out.

[7] We thank an anonymous reviewer for having raised this important point.

Table 1
Definition of inputs, outputs and conditioning factors.

Input Output Conditioning factor	Definition
Input	
NACSTA (x_1)	Number of non academic staff
ACSTAF (x_2)	Number of academic staff
PEREXP (x_3)	Personnel expenditures (PPS)
NOPEXP (x_4)	Non-personnel expenditures (PPS)
FINP	Input factor including: NACSTA, ACSTAF, PEREXP, NOPEXP
Output	
TODEG5 (y_1)	Total Degrees ISCED 5
TODEG6 (y_2)	Total Degrees ISCED 6 (Doctorate)
PUB (y_3)	Number of published papers (Scimago)
IC (y_4)	International Collaboration (Scimago)
NI (y_5)	Normalized Impact (Scimago)
Q1 (y_6)	High Quality Publications (Scimago)
EXC (y_7)	Excellence Rate (Scimago)
FRES	Factor of volume of research including: TODEG6, PUB
FQUAL	Factor of quality of research including: IC, NI, Q1, EXC
Conditioning factors	
SIZE (z_1)	It is the sum of Total Students enrolled ISCED 5 and Total Students enrolled ISCED 6 (used in *log* in the elaborations)
SPEC (z_2)	Proxy of Specialization Gini index of the scientific output (Scimago)

Source: Eumida DC2 and Scimago.

3.3. Variables

In this section we describe the main variables that were retained from the previously integrated databases for the following analysis. It is important to notice that, for the selected variables, not all the European institutions whose data were integrated reported all the information required and this is the reason why the number of observations retained for the elaborations was reduced to 400 observations for which all the information were available.

Table 1 defines and describes the inputs, outputs and conditioning factors that are used in the following analysis.

As commonly used in applied econometrics, the size is used in the analysis as the *log* of the total volume of the activity, that in our case is proxied by the sum of enrolled students at all undergraduate and post-graduate levels (*SIZE*).

Table 2 reports some descriptive statistics (25th percentile, median, average, 75th percentile and standard deviation) on the sample that will be analysed in the paper. Fig. 1 illustrates the nonparametric kernel distributions of the *SIZE* and *SPEC* variables.

From a preliminary data analysis, we found that PUB and TODEG6 were highly correlated; that NACSTA, ACSTAF, PEREXP, NOPEXP were also highly correlated and that IC, NI, Q1, EXC were also highly correlated. We found correlations higher than 85% in all cases, and for this reason, in the analysis we used their aggregating factors, respectively FINP, FRES and FQUAL. Additional details on the factorial data analysis are reported in Appendix A.

Table 2
Descriptive statistics.

Variable	25th	Median	Average	75th Perc	Std
NACSTA	562	1040	1497	1811	1408
ACSTAF	686	1164	1470	1973	1058
PEREXP	54,600,000	103,370,000	142,580,000	187,810,000	121,660,000
NOPEXP	27,250,000	58,100,000	87,111,000	100,320,000	94,925,000
TODEG5	1748	3205	3881	4992	3146
TODEG6	55	126	204	278	215
PUB	1505	3609	5571	7564	5626
IC	33	38	39	44	9
NI	1.10	1.30	1.30	1.50	0.31
Q1	44	53	51	60	13
EXC	11	15	15	20	6
SIZE	10,038	16,718	20,259	24,559	17,485
SPEC	0.60	0.70	0.69	0.80	0.13

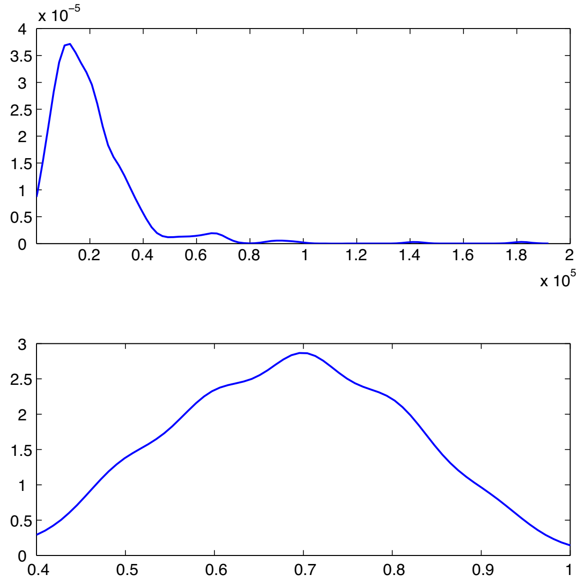

Fig. 1. Nonparametric kernel distribution of *SIZE* (top panel) and *SPEC* (bottom panel).

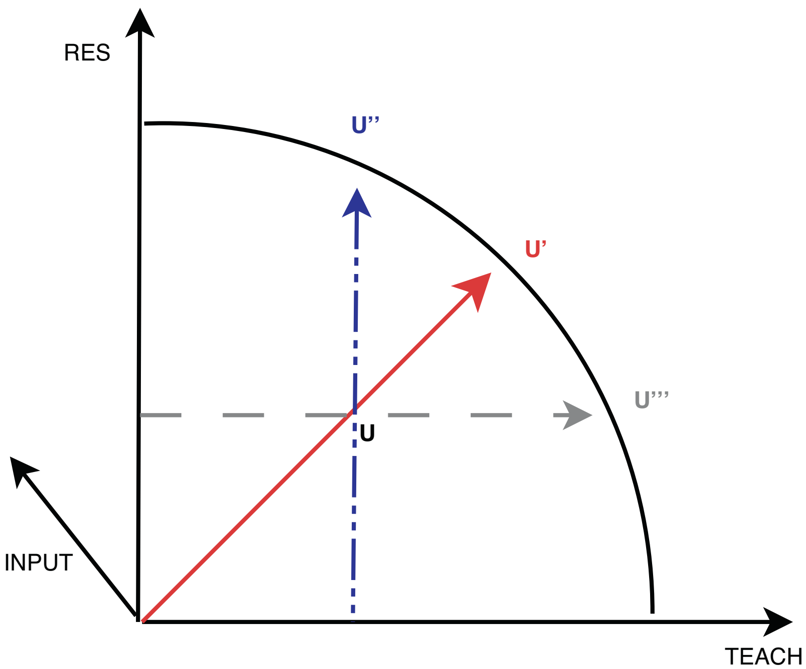

Fig. 2. An illustration of tradeoffs in the academic production.

4. Production models of European universities

In this section we present the modeling strategy of our approach. While this section introduces the main ideas of directional distances through a simple illustration, Section 5 details the methodology of directional distances and their estimation.

Fig. 2 illustrates the flexibility of directional distance functions to model internal trade-offs between dimensions of the academic production. For each unit in the sample, we can assess its performance (or technical efficiency) considering also its input structure, along the research dimension (RES), considering given the teaching that it is carrying out. This corresponds for unit u to move towards u″. Alternatively, we could investigate the performance of u along the teaching dimension (TEACH) keeping constant (or considering given) its research activity (this corresponds to assess the performance of u in reaching the efficient frontier from u to u‴ in Fig. 2). Finally, unit u could be assessed on how it is performing in doing both teaching and research, that corresponds in Fig. 2 to move towards the efficient frontier from u to u′.

This is the basic illustration of the activity. Obviously, the efficiency processes described in Fig. 2 may be affected by some external factors that are, at least in the short run, not under the control of the units. This leads to the inclusion in the analysis of these factors whose potential impact on the performance we are interested in estimating.

In this paper we are going to evaluate the impact of SIZE and SPEC. These factors indeed might influence the probability of each unit (university) of being dominated (that is of lagging far away from the efficient boundary of the production frontier). We apply a directional output distance function, in which the direction to approach the efficient frontier is the same for each university in our sample (‘egalitarian approach’) and it is set to the European median. We think that this choice reflects the important European Research Area pillar of “cooperation and competition” because the comparison in terms of target is with respect to a median value calculated over a highly skewed distribution.

We would like to estimate also the efficiency of the research activity itself, but this was not possible because the available inputs data refer to all the activities of universities including also teaching. We would like also to include information on the third mission activity (i.e. knowledge transfer, collaborations with industry, patents and so on), but data were not available for all the universities in our sample.

We analyse the impact of scale (as proxied by the SIZE variable) and specialization (as proxied by the SPEC variable) on two models of university production in Europe, namely[8]:

Humb model Full model of academic production, in which the targets to reach the frontier are set in terms of teaching, research and quality. The following variables are used: *Input*: FINP, *Outputs*: TODEG5, FRES, FQUAL; *external factors*: SIZE, SPEC.

RES model Research model, in which teaching is considered given, and targets to reach the frontier are set in terms of research and quality. The following variables are used: *Input*: FINP, *Outputs*: TODEG5 is kept constant, FRES, FQUAL; *external factors*: SIZE, SPEC.

In this paper, following also existing literature (e.g. Daghbashyan et al., 2014; Johnes, 2006) we approximate the output of the teaching activities by the total number of degrees produced. Of course, more detailed information about employment rate of graduated students or wages for the first job would provide additional information on the teaching quality and its alignment with the needs of labor market. Unfortunately, comparable data at European level on placement of students are not available.

5. Method: a flexible approach based on directional distances

We apply an activity analysis framework within the theory of production (see Shephard, 1970), in which producing units (hereafter “unit”), realize a set of outputs $Y \in \mathbb{R}^q$ by combining a set of inputs $X \in \mathbb{R}^p$. The technology is characterized by the attainable set T, the set of combination of (x, y) that are technically achievable

$$T = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q | x \,\text{can produce}\, y\}. \tag{1}$$

We know that under the *free disposability* assumption for the inputs and the outputs, the set can be described as[9]:

$$T = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q | H_{XY}(x, y) > 0\}, \tag{2}$$

where $H_{XY}(x, y)$ is the probability of observing a unit (X, Y) dominating the production plan (x, y), i.e. $H_{XY}(x, y) = \text{Prob}(X \le x, Y \ge y)$.

The free disposability we used in this paper is the assumption that if $(x, y) \in T$ then $(\tilde{x}, \tilde{y}) \in T$ for all $\tilde{x} \ge x$ and all $\tilde{y} \le y$. It is a minimal assumption generally made on production processes.

The efficient boundary of T is of interest and several ways have been proposed in the literature to measure the distance of the unit (x, y) to the efficient frontier. One of the most flexible approach is the directional distance introduced by Chambers, Chung, and Färe (1996) (see also Färe, Grosskopf, & Margaritis, 2008). Given a directional vector for the inputs $d_x \in \mathbb{R}^p_+$ and a direction for the outputs $d_y \in \mathbb{R}^q_+$, the directional distance is defined as

$$\beta(x, y; d_x, d_y) = \sup\{\beta > 0 | (x - \beta d_x, y + \beta d_y) \in T\}, \tag{3}$$

or equivalently, under the free disposability assumption (see Simar & Vanhems, 2012)

$$\beta(x, y; d_x, d_y) = \sup\{\beta > 0 | H_{XY}(x - \beta d_x, y + \beta d_y) > 0\}. \tag{4}$$

8 Bonaccorsi et al. (2013) instead analyse the efficiency of a teaching model.

9 See Daraio and Simar (2007) for further details and illustrations.

Hence, we measure the distance of unit (x, y) to the efficient frontier in an additive way and along the path defined by $(-d_x, d_y)$.

This way of measuring the distance is very flexible and generalizes the "oriented" radial measures initiated by Farrell (1957). Indeed by choosing $d_x = 0$ and $d_y = y$ (or $d_x = x$ and $d_y = 0$), we recover the traditional output (reps. input) radial distances. The flexibility is that we might have some elements of the vector d_x and/or of the vector d_y be set to zero, for focusing on the distances to the frontier along certain particular paths (for instance if some inputs or outputs are non-discretionary, not under the control of the manager, etc.).

Consistent nonparametric estimators of Eq. (4) have been proposed in Simar and Vanhems (2012); Daraio and Simar (2014) analyse in details the case when some directions are set to zero, as well as statistical issues in this context.

For a discussion about the choice of a direction, see Färe et al. (2008). The direction can be different for each unit (like in the radial cases) or it can be the same for all the units. Färe et al. (2008) argue that a common direction would be a kind of egalitarian evaluation reflecting some social welfare function. Researchers often select in the latter case $d_x = \mathbb{E}(X)$ and $d_y = \mathbb{E}(Y)$, where in practice empirical averages are chosen.

In this paper we select the same direction for all the units, setting a reference with respect to the European standard. The reference is made with respect to the median value of each output calculated at European level over the analysed sample.

Quantile frontiers for evaluating the performance of firms by using oriented radial measures (input or output) have been extended to directional distance in Simar and Vanhems (2012) and this extension is quite natural after the representation given in (4). In place of looking to the support of the distribution H_{XY} we benchmark the unit against a point which leaves on average $\alpha \times 100\,\%$ of points above the frontier. This benchmark is the α-quantile frontier. Formally the α-order directional distance is defined as

$$\beta_\alpha(x, y; d_x, d_y) = \sup\{\beta > 0 | H_{XY}(x - \beta d_x, y + \beta d_y) > 1 - \alpha\}. \tag{5}$$

Here a value $\beta_\alpha(x, y; d_x, d_y) = 0$ indicates a point (x, y) on the α-quantile frontier, a positive value is a point below the quantile frontier and a negative value is a point above the quantile frontier. We see clearly that when $\alpha \to 1$ we recover the full frontier definition.

The projection of any $(x, y) \in T$ on the estimated α-quantile frontier is given by the points $(\hat{x}_\alpha^\partial, \hat{y}_\alpha^\partial)$ defined as

$$\hat{x}_\alpha^\partial = x - \widehat{\beta}_\alpha(x, y; d_x, d_y)d_x, \quad \text{and} \quad \hat{y}_\alpha^\partial = y + \widehat{\beta}_\alpha(x, y; d_x, d_y)d_y. \tag{6}$$

Since the resulting estimator will not envelop all the data points, the resulting frontier is more robust to outliers and extreme data points than its full version above.

This is the approach we implemented in our empirical analysis.

5.1. Second stage regression: impact of scale and specialization on efficiency

Badin, Daraio, and Simar (2012) propose a general nonparametric methodology to investigate the impact of external–environmental factors on the efficiency scores of units. This framework is completely different with respect to the traditional regression-based framework. The objective is not to estimate the impact of the external factors on the outputs of the analysed units. On the contrary, we estimate nonparametrically a conditional probability law were the external factors can and in general should be related in a linear or even in a nonlinear way with both the inputs and the outputs. We estimate the impact of the external factors on the performance of the units, that is their ability to operate close to their efficient frontier (estimated by using their input–output relationship). This approach allows us to capture also the presence of nonlinear impact of the considered external factors.

Daraio and Simar (2014) extended this methodology to conditional directional distances to investigate the effect of z on the mean of the conditional directional distances.

This method contributes to the literature on the so called two-stage approach, where estimated unconditional efficiency scores (input or output oriented) are regressed in a second stage against the Z variables. However we know from the literature (see Badin, Daraio, & Simar, 2014 for a detailed explanation and more references) that this is valid only under a 'separability' assumption where it is assumed that the frontier of the attainable set is not changing with the values of z.

As indicated in Badin et al. (2012), the use of the estimated conditional efficiency scores for this second stage regression, does not require this restrictive assumption.

Hence, following Daraio and Simar (2014), the flexible second stage regression can be written as the following location-scale nonparametric regression model:

$$\beta_\alpha(X, Y; d_x, d_y | Z = z) = \mu(z) + \sigma(z)\varepsilon, \tag{7}$$

where ε and Z are independent with $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{V}(\varepsilon) = 1$. This model permits to detect the location $\mu(z) = \mathbb{E}(\beta_\alpha(X, Y; d_x, d_y | Z = z))$ and the scale effect $\sigma^2(z) = \mathbb{V}(\beta_\alpha(X, Y; d_x, d_y | Z = z))$.

These two functions can be estimated non parametrically from a sample of observations $\{Z_i, \widehat{\beta}_\alpha(X_i, Y_i; d_x, d_y | Z_i)\}$, $i = 1, \ldots, n$ by using local constant smoothing techniques to guarantee positive estimates of both functions, as suggested by Daraio and Simar (2014). The analysis of the estimated $\widehat{\mu}(z)$ as a function of z will enlighten the potential effect of Z on the average efficiency, with the help of $\widehat{\sigma}(z)$ which may indicate the presence of heteroskedasticity.

5.2. Testing the significance of scale and specialization

Here we apply the approach of Daraio and Simar (2014) which adapted to the efficiency setup the test proposed by Racine (1997). The test statistics is based on the partial derivatives of the mean function $\mu(z) = \mathbb{E}(\beta_\alpha(X, Y; d_x, d_y|Z = z))$ that are:

$$\eta_j(z) = \partial\mu(z)/\partial z_j, \quad \text{for } j = 1, \ldots, r. \tag{8}$$

Without loss of generality, the null hypothesis (H_0) to test is that the first r_1 components of Z do not affect $\mu(z)$ against the alternative hypothesis that some of these components of Z affect $\mu(z)$.

The constructed test statistics is given by:

$$\tau = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r_1}[\eta_j(Z_i)]^2, \quad r_1 \le r. \tag{9}$$

We will reject the null in favor of the alternative when τ is too big. Both the p-value of H_0 and critical values of any size are determined by the bootstrap. See Daraio and Simar (2014) for further details on how the bootstrap is implemented.

We point out that for statistical significant impact we mean that there exists a statistical significant "influence" or "association" of the investigated variables on the performance (input–output relationship) of the analysed units, as we do not consider any causal relation here.

5.3. Analyzing the gaps

It may be useful for policy makers to measure, in original units of the inputs and of the outputs, the estimated distance of an observation to the frontier. This allows us to appreciate the efforts to be achieved in increasing the outputs and reducing the inputs to reach the efficient frontier. For the full frontier these measures are given by what we call the "gaps" to efficiency. They are directly given by:

$$G_x = \widehat{\beta}(x, y; d_x, d_y)d_x, \quad \text{and} \quad G_y = \widehat{\beta}(x, y; d_x, d_y)d_y. \tag{10}$$

For the partial frontiers, the gaps appear as being the difference between (x, y) and the projections on the α-quantile frontier given in (6). They are particularly useful to detect outliers in the direction given by (d_x, d_y). This will be the case in the input direction if $G_{\alpha,x} = \widehat{\beta}_\alpha(x, y; d_x, d_y)d_x$ has some elements with large negative value: the point (x, y) is well below the estimated α-frontier in the input direction, and/or a very large negative value in some elements of the vector $G_{\alpha,y} = \widehat{\beta}_\alpha(x, y; d_x, d_y)d_y$ warns a point being well above the quantile frontier in the output direction.

As explained in Section 4, in the empirical analysis that follows in the next sections we pursue an output orientation approach aiming at estimating the efforts needed to increase the outputs of the units, given the level of their inputs used, and hence we estimate the robust gaps $G_{\alpha,y}$ in terms of percentage values of the analysed outputs.

6. Results

In this section we summarize the main results of the analysis carried out.

6.1. Impact of scale and specialization on efficiency

In this subsection we report the results of the impact of scale and specialization analysis obtained for the Humb model.

Fig. 3 illustrates the results of the nonparametric regression of the estimated $\mu(z)$ in function of SIZE and SPEC.

To provide a graphical illustration in two dimensions of the effect of one variable on the efficiency, one can use partial regression lines of efficiency over one of the variables for a fixed level of the second one. In addition, pointwise bootstrap error bounds can be displayed to appreciate visually the variability of the estimates. This is common practice in nonparametric statistics, as suggested e.g. in Racine (2008).

In Fig. 4 the nonparametric regression of conditional efficiency vs SIZE is reported, so the partial impact of SIZE is represented by fixing the SPEC value at its median level. We observe an inverse U-shaped impact of SIZE (given that SPEC is fixed at its median) already visible in Fig. 3. To read the plot we have to remind that the smaller the level of β_α the greater the efficiency of the unit is. It appears that there is a lot of uncertainty when Z_1 is smaller than 8 (corresponding to exp(8) = 2981 total enrolled students), as pointed out by the large bootstrap error bounds, because there are few and heterogeneous small universities in our sample.

Fig. 4 shows that the partial impact of scale on the efficiency of the European universities analysed is not linear: it appears that size has a negative impact up to a log of SIZE of around 10 (corresponding to a total number of enrolled students of 22,026) and after that it has a positive impact. As we shall see below in this section, the nonlinear impact of size is statistically

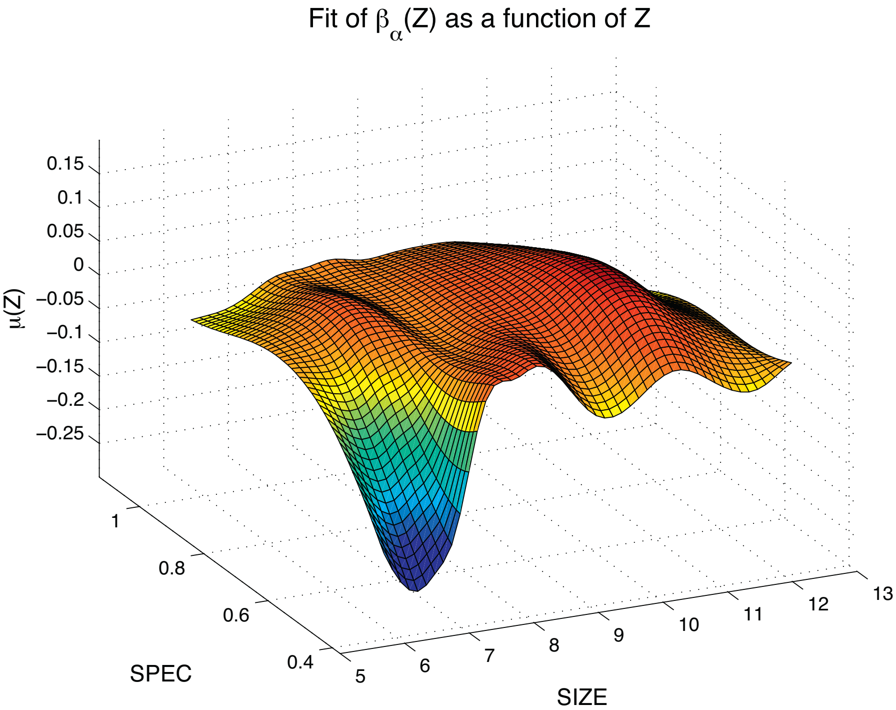

Fig. 3. Humb model. Nonparametric regression of the estimated $\mu(z)$ versus Z_1 = *SIZE* and Z_2 = *SPEC*. Note that *SIZE* is expressed in *log*.

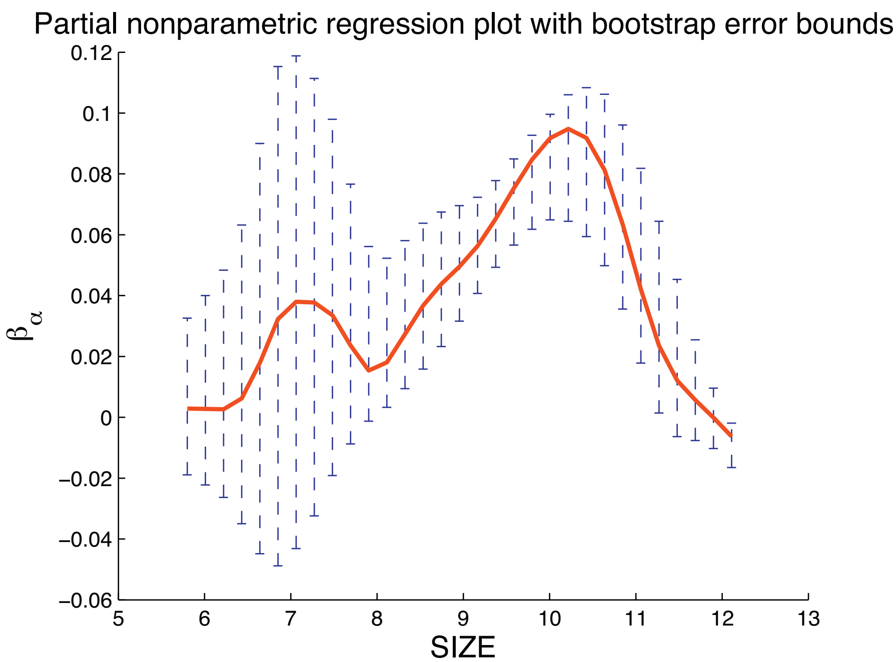

Fig. 4. Humb model. Partial nonparametric regression of conditional efficiency as a function of SIZE, helding constant to its median value SPEC. Bootstrap error bounds are reported to illustrate the variability of the estimates. Note that *SIZE* is expressed in *log*.

significant, considered in isolation, in both HUMB and RES models. Moreover, it is significant also jointly considered with SPEC in the Humb model. We observe that, thanks to the flexibility of our approach, we are able to shed light on the behavior of SIZE over its range of variation, which is not constant but varying, and are able to appreciate the variability of the estimates of its impact.

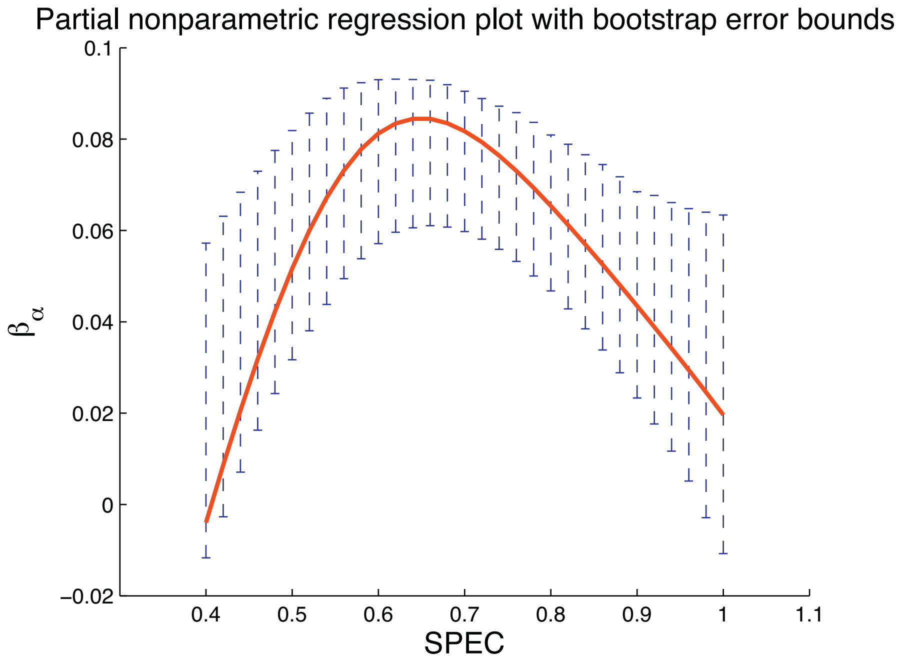

Fig. 5. Humb model. Partial nonparametric regression of conditional efficiency as a function of disciplinary specialization (SPEC), helding constant at its median value SIZE. Bootstrap error bounds are reported to illustrate the variability of the estimates.

In Fig. 5 the nonparametric regression of conditional efficiency vs SPEC is displayed. The partial impact of SPEC is represented by fixing the SIZE value at its median level. Bootstrap error bounds are also reported again to illustrate the variability of the estimates. We observe that SPEC has an even clearer inverse U-shaped impact with respect to SIZE; again to read the plot we have to remind that the smaller the level of β_α the greater the efficiency of the unit is. By inspecting Fig. 5 it appears that also the partial impact of specialization on the efficiency of the European universities analysed is not linear. As observed for size, also specialization has a negative impact up to around 0.65 and after that it has a positive impact. As we shall see below in this section, the nonlinear impact of specialization is statistically significant, considered both in isolation and jointly considered with SIZE in the Humb model, however it is not significant for the RES model. See below for more details.

We run the analysis also for the RES model and obtain very similar results. See Fig. 6.

Here below we report the results of the testing of scale and specialization carried out for the full HUMB MODEL, with $B = 1000$ bootstrap loops. The test has been implemented by following the approach of Daraio and Simar (2014) described above. We investigate the impact of both external factors Z together and also each factor separately.

The obtained results confirm that SIZE and SPEC have a statistically significant impact both together and in isolation. Indeed we have:

size and spec $Z = (Z_1, Z_2)$, the p-value of H_0 at 5% is 0.036, observed value of the test statistics $\widehat{\tau} = 0.0245$, the Bootstrap-based critical value at 5% is 0.0228; given that the p-value of H_0 is less than 0.05 we reject H_0.

size $Z = Z_1$, p-value (at 5%) of H_0 is 0.000001, the observed value of the test statistics $\widehat{\tau} = 0.001619$, the Bootstrap-based 5% critical value is: 0.00079; given that the p-value of H_0 is much less than 0.05 we reject H_0.[10]

spec $Z = Z_2$, p-value (at 5%) of H_0 is 0.031, the observed value of the test statistics $\widehat{\tau} = 0.02291$, the Bootstrap-based 5% critical value is 0.02031; given that the p-value of H_0 is less than 0.05 we reject H_0.

From this analysis, we can conclude that for the Humb model, the nonlinear impact of scale and specialization, illustrated in Fig. 3, is statistically significant both considering size and specialization alone as well as jointly.

[10] Note that at first sight this result may seem in contrast with the large error bounds in Fig. 4. It is not. Remember that Fig. 4 illustrates the error bound of the partial regression of the efficiency on SIZE at a fixed value (median) of SPEC. When integrating the analysis over all values of SPEC we find a significant effect.

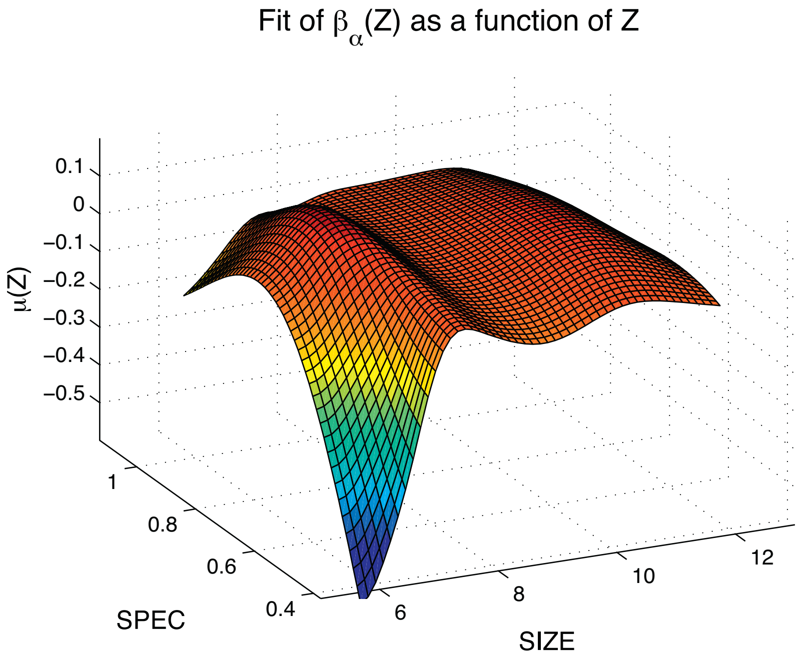

Fig. 6. RES MOD. Nonparametric regression of the estimated $\mu(z)$ versus Z_1 = *SIZE* and Z_2 = *SPEC*. Note that *SIZE* is expressed in *log*.

In the following we summarize the results for the RES model:

size and spec $Z=(Z_1, Z_2)$, the p-value of H_0 (at 5%) is 0.223, observed value of the test statistics $\hat{\tau} = 0.02996$, the Bootstrap-based critical value at 5% is 0.04769; given that the p-value of H_0 is higher than 0.05 we cannot reject H_0.

size $Z=Z_1$, p-value (at 5%) of H_0 is 0.007, the observed value of the test statistics $\hat{\tau} = 0.00202$, the Bootstrap-based 5% critical value is: 0.001245; given that the p-value of H_0 is lower than 0.05 we reject H_0.

spec $Z=Z_2$, p-value (at 5%) of H_0 is 0.186, the observed value of the test statistics $\hat{\tau} = 0.02794$, the Bootstrap-based 5% critical value is 0.0422. Given that the p-value of H_0 is higher than 0.05 we cannot reject H_0.

The obtained results show that the specialization of the universities does not play a significant role on the research performance model.

Interestingly, our results for the RES model seem to support previous results (Moed et al., 2011) which found that the concentration of research among institutions is not associated with better overall performance. Our analysis indeed may reflect previous findings that "it is multidisciplinary research that is the most promising and visible at the international research front, and that this type of research tends to develop better in universities specializing in a particular domain and expanding their capabilities in that domain towards other fields. If specialization is too strong, an institution may not be able to pick up the developments in emerging topics that require a structural contribution from fields in which it hardly shows activity and does not have expertise" (Moed et al. pag. 657). Further research is needed to confirm this hypothesis. Nevertheless, by applying a different approach and integrating bibliometric data with input data at institutional level we were able to find some support to this hypothesis. This result is encouraging and shows the usefulness and the importance of integrating data from different sources to analyze complex input–output relationships at the institutional level.

6.2. *Efficiency results and analysis of gaps*

The analyses have been carried out on the entire European sample. In this section we summarize the obtained results grouping them by country, and report the European average computed over the analysed sample to facilitate the interpretation. We remind again that to read the results, the smaller the level of the efficiency the greater the efficiency of the unit (or group of units) is.

Table 3 reports in the columns: Country, Country code (C. code), number of observations (# obs), number of dominating units (# dom) which is the average number of units which dominates the universities of a given country, empirical estimates of the probability of the universities of a country of being dominated ($\hat{H}_{XY}$, that is defined in Section 5), robust directional

Table 3
Efficiency results for Humb model: averages by country.

Country	C. code	#obs	#dom	$\widehat{H}_{XY}$	$\beta_{\hat{\alpha},XY\|Z}$	Std of $\beta_{\hat{\alpha},XY\|Z}$
Austria	AT	14	4.21	0.0105	0.040465	0.071259
Belgium	BE	4	2.75	0.0069	0.061991	0.087600
Switzerland	CH	11	1.18	0.0029	0.008743	0.028996
Czech Republic	CZ	14	3.00	0.0075	0.042476	0.061663
Germany	DE	71	10.55	0.0263	0.153871	0.158692
Spain	ES	47	6.15	0.0153	0.097338	0.106114
Finland	FI	12	1.75	0.0044	0.012852	0.022466
Hungary	HU	6	27.50	0.0686	0.209560	0.196387
Ireland	IE	10	2.40	0.0060	0.033448	0.045473
Italy	IT	60	4.23	0.0106	0.064099	0.090596
the Netherlands	NL	13	3.46	0.0086	0.048254	0.105020
Norway	NO	8	6.25	0.0156	0.115628	0.112975
Romania	RO	14	1.86	0.0046	0.024394	0.052332
Sweden	SE	17	1.71	0.0043	0.014079	0.037360
Slovakia	SK	4	2.25	0.0056	0.013651	0.027301
United Kingdom	UK	89	1.80	0.0045	0.027551	0.057940
All sample	EU	400	4.96	0.0124	0.068804	

Note: only countries with at least 4 observations are reported in the table. The last line reports the average over the whole analyzed sample.

Table 4
Gaps in percentages for Humb model: averages by country.

Country	#obs	#DEG5	#DEG6	#PUB	IC	Q1	NI	EXC
AT	14	0.13	0.11	0.03	0.03	0.04	0.04	0.04
BE	4	0.04	0.03	0.02	0.05	0.06	0.06	0.05
CH	11	0.02	0.00	0.00	0.01	0.01	0.01	0.01
CZ	14	0.06	0.06	0.11	0.06	0.08	0.07	0.13
DE	71	0.27	0.08	0.11	0.16	0.15	0.15	0.15
ES	47	0.12	0.17	0.14	0.11	0.10	0.12	0.13
FI	12	0.03	0.02	0.03	0.01	0.02	0.01	0.02
HU	6	0.25	0.30	0.26	0.21	0.23	0.32	0.26
IE	10	0.04	0.31	0.22	0.03	0.05	0.04	0.05
IT	60	0.14	0.23	0.08	0.08	0.06	0.07	0.07
NL	13	0.05	0.03	0.02	0.04	0.05	0.04	0.06
NO	8	0.20	1.98	0.24	0.11	0.12	0.11	0.15
RO	14	0.02	0.25	0.19	0.03	0.09	0.05	0.13
SE	17	0.03	0.01	0.01	0.01	0.01	0.01	0.01
SK	4	0.03	0.03	0.04	0.01	0.02	0.02	0.02
UK	89	0.02	0.06	0.06	0.03	0.03	0.03	0.04
EU	400	0.11	0.15	0.09	0.07	0.08	0.08	0.09

Note: only countries with at least 4 observations are reported in the table. The last line reports the average over the whole analyzed sample.

measure of efficiency conditioned to SIZE and SPEC, our Z variables ($\hat{\beta}_{\alpha,XY|Z}$) and the standard deviation of the conditional efficiency scores (Std of $\hat{\beta}_{\alpha,XY|Z}$) in the last column.

The last line of the Table shows the average at European level. An outline of the efficiency analysis results on the Humboldtian model could be obtained by comparing the average performance at national level with the European average.

By inspecting Table 3 it appears that countries that are performing much better than the European average are Switzerland, UK, Sweden, Slovakia, Belgium, Austria, Ireland, the Netherlands, Czech Republic, Romania and Finland. The others follow, as it appears by analysing the average over the country of conditional efficiency scores ($\hat{\beta}_{\alpha,XY|Z}$). Only four countries score above the European average (having hence worst performance), including Germany, which shows a strong influence on the European average. Obviously, these comparisons of country's averages with respect to the European average should be taken with care. It is in fact for this reason that we will report in the following the comparison on a country base, showing the internal variability of the institutions within a country. This finding deserves to be further investigated. Here we observe that there is a great heterogeneity among German universities as well as the other three countries which perform worst than the European average, namely Hungary, Norway and to a certain extent Spain, whose standard deviation of their respective conditional efficiency scores is 0.159 for Germany, 0.196 for Hungary, 0.11 for Norway and 0.106 for Spain. More generally, a certain degree of heterogeneity of the efficiency within countries is observed also for all the European countries, including well performing ones, such as the Netherlands which has a standard deviation of its efficiency scores of 0.105. See also below where we provide some discussion on the results obtained in terms of gaps which are reported in Table 4.

Table 4 reports the estimated gaps in percentage of the outputs produced by the units to reach the robustly estimated efficient frontier.

Table 5
Efficiency results for RES model: averages by country.

Country code	#obs	#dom	$\widehat{H}_{XY}$	$\hat{\beta}_{\alpha,XY\|Z}$	Std of $\hat{\beta}_{\alpha,XY\|Z}$
AT	14	4.21	0.0105	0.051877	0.082760
BE	4	2.75	0.0069	0.061991	0.087600
CH	11	1.18	0.0029	0.008743	0.028996
CZ	13	3.15	0.0079	0.094881	0.129447
DE	71	10.55	0.0263	0.163797	0.165207
ES	47	6.15	0.0153	0.123091	0.138990
FI	11	1.82	0.0045	0.027432	0.060782
HU	6	27.50	0.0686	0.371831	0.249816
IE	10	2.40	0.0060	0.048087	0.081392
IT	60	4.23	0.0106	0.089005	0.148184
NL	13	3.46	0.0086	0.059462	0.113649
NO	8	6.25	0.0156	0.143286	0.130379
RO	9	2.33	0.0058	0.037946	0.062224
SE	16	1.75	0.0044	0.021866	0.047697
SK	3	2.67	0.0067	0.043278	0.074959
UK	86	1.83	0.0046	0.039575	0.074117
EU	387	5.10	0.0127	0.090002	

Note: only countries with at least 4 observations are reported in the table. The last line reports the average over the whole analyzed sample.

Table 6
Gaps in percentages for RES model: averages by country.

Country	#obs	#DEG5	#DEG6	#PUB	IC	Q1	NI	EXC
AT	14	0.00	0.17	0.04	0.04	0.05	0.05	0.06
BE	4	0.00	0.03	0.02	0.05	0.06	0.06	0.05
CH	11	0.00	0.00	0.00	0.01	0.01	0.01	0.01
CZ	13	0.00	0.10	0.21	0.12	0.18	0.15	0.27
DE	71	0.00	0.09	0.13	0.17	0.16	0.16	0.16
ES	47	0.00	0.22	0.18	0.14	0.13	0.15	0.18
FI	11	0.00	0.04	0.04	0.03	0.03	0.03	0.04
HU	6	0.00	0.58	0.62	0.38	0.41	0.57	0.49
IE	10	0.00	0.49	0.36	0.04	0.08	0.06	0.09
IT	60	0.00	0.42	0.14	0.12	0.08	0.10	0.10
NL	13	0.00	0.04	0.02	0.05	0.06	0.05	0.07
NO	8	0.00	4.01	0.40	0.14	0.15	0.14	0.20
RO	9	0.00	0.36	0.29	0.05	0.14	0.08	0.21
SE	16	0.00	0.04	0.04	0.02	0.02	0.02	0.02
SK	3	0.00	0.08	0.14	0.03	0.07	0.07	0.08
UK	86	0.00	0.08	0.08	0.05	0.04	0.04	0.05
EU	387	0.00	0.26	0.14	0.10	0.10	0.10	0.13

Note: only countries with at least 4 observations are reported in the table. The last line reports the average over the whole analyzed sample.

Looking at Table 4 by country, some findings are striking. In the Humb model the targets are expressed in terms of education, research volume and research quality. Switzerland (CH) is the single most efficient country, with negligible gaps in either education and research. Finland, Sweden, the Netherlands, Slovakia and United Kingdom are also countries in which the magnitude of gaps is very small. Among the least efficient countries Hungary and Norway stand up. Other countries exhibit highly differentiated patterns of gap by type of output. For example, Germany looks less efficient in undergraduate education (with a large gap at 0.27), while in postgraduate education and publications it shows higher levels of efficiency. Germany might also improve in the upper tail of scientific production (gap in the EXC indicator= 0.15). Italy, in turn, has large gaps in both undergraduate and postgraduate education, while indicators of research efficiency are much better. Overall, European universities could produce more educational output, both undergraduate and postgraduate, and also get some improvement in research volume and quality.

Nevertheless, we point out that some comparability problems at the country level may still exist, and may result in an extremely high amount of gaps, as is the case for Norway.

Table 5 reports the results of the RES model. In the columns there are: Country code, number of observations (# obs), number of dominating units (# dom), empirical estimates of the probability of being dominated ($\widehat{H}_{XY}$), robust directional measure of efficiency conditioned to SIZE and SPEC, our Z variables ($\hat{\beta}_{\alpha,XY|Z}$) and the standard deviation of the conditional efficiency scores (Std of $\hat{\beta}_{\alpha,XY|Z}$) in the last column.

In the RES model the teaching output at undergraduate level is considered fixed. This model addresses the question whether some improvements in research can be obtained without compromising the educational mission. The last line of Table 5 shows the average efficiency of the RES model at European level. Table 6 shows that it might be possible to increase greatly the doctoral output without compromising the undergraduate education. This is a striking result for Europe. In large countries such as Italy and Spain the gain might be significant. Also, on average 10% improvement in both research volume and quality is attainable without losses in educational output. By comparing the average performance at national level with the European average, it seems that results for the Research Model are similar to the ones obtained in the Humboldtian Model. We can observe an high heterogeneity of university performance within countries, as showed by the high standard deviation of the conditional efficiency scores reported in the last column of Table 5.

Table 6 reports the estimated gaps in percentage of the outputs produced by the units to reach the robustly estimated efficient frontier.

Summing up, the inspection of average efficiency values per country shows large differences due to the national context. Moreover, within each country there is an high degree of heterogeneity in the performance as the high standard deviations (reported in the last column of Table 5) show. The interpretation of these differences will require a dedicated research effort.

A preliminary conjecture could be as follows. In order to make the best use of their inputs, universities should be put in the position to move in their multidimensional strategic space. This space includes inputs and outputs. Efficient universities are those that adjust their mix of inputs in order to achieve the best possible mix of outputs. It is clear that universities do not have full discretionary power over inputs and outputs, as our analysis has clearly recognised. However, national contexts may provide more or less strategic autonomy, that is, may support universities in their strategic positioning or may, on the contrary, create legal and administrative constraints. Supporting the autonomy of universities in strategic positioning is generally associated to two conditions. As for education, it requires that universities are in the position to match appropriately the profile of students to the teaching offering. While this may have different implications in different fields, there is a well known general problem that cuts across fields of education and countries, namely the role of professional education, also called vocational training. According to CEDEFOP (2014), *vocational* education and training "aims to equip people with knowledge, know-how, skills and/or competences required in particular occupations or more broadly on the labour market".

Some countries allocate vocational training to separate institutions, while others add it to the general mission of universities. In the latter case universities have, in general, larger student loads and lower teaching efficiency, given the mismatch between the educational needs of students and the rigidity of the university offering. As for research, efficiency requires that public research funding is allocated according to criteria that give a premium to research quality. This follows the adoption of evaluation exercises, or formula-based funding criteria based on research quality. Universities that are placed in an institutional context based on research quality funding develop over time strategies to improve their positioning. This adjustment may require years, if not decades, to take place. This conjecture might help to explain the findings. Interestingly, the countries that perform well in both models (Humboldtian model and Research model) share, by and large, two institutional features.

On the one hand, they have since many years dual or binary higher education systems, in which vocational training is allocated to non-university institutions (or is delegated to the private sector as in UK). Dual systems are in a better position to adjust their inputs and outputs of education, because university students self-select themselves against a well articulated and prestigious non-university higher education system. On the other hand, many of these countries have implemented since long time university funding systems in which there is a significant performance-based component, largely dependent on research, or in which formalized research assessment exercises have been carried out. Strikingly, no large continental European country shows up among the best performers. Germany has indeed a vibrant vocational training system, but its university sector is somewhat less competitive. Italy and Spain do not have a dual higher education system, so that the efficiency of undergraduate education is reduced by a large number of dropouts. In addition, they have started to implement competitive funding of universities only recently. Thus not only the best performers but also the countries with significant gaps seem to confirm the conjecture. Nevertheless, further research is clearly needed to confirm this conjecture.

7. Conclusions

In this paper we analysed the issue of scale and specialization in European universities by applying state of the art directional distances techniques on an original database built by integrating input/output university data with bibliometric indicators. Moreover, we improved over previous studies adopting a cross-country perspective, applying robust nonparametric estimators and testing for the significance of scale and specialization effects by using the bootstrap.

We find that size and specialization have a significant impact on the efficiency of the Humboldt model, whilst specialization has not a significant impact on the efficiency of the research model. By applying a different approach and integrating bibliometric data with input data at institutional level we were able to find some support to previous research on the importance of *multidisciplinary research that is the most promising and visible at the international research front* (Moed et al., 2011). This evidence is encouraging and shows the usefulness and the importance of integrating data from different sources to analyze complex input–output relationships at the institutional level. Nevertheless, further research is needed to confirm the preliminary findings of this paper.

Although the data we have used come from a feasibility study, that is, even if the data have been extensively examined by experts within the Eumida project, they have not been subject to data quality analysis and systematic checks by the National Statistical Authorities, which however provided them.

Further developments, in progress, are directed to develop a robust methodology for the data quality analysis specifically tailored for input/output data coming from different heterogeneous sources.

Acknowledgments

We thank two anonymous reviewers for their helpful comments and suggestions which allowed us to greatly improve the content of our paper. The usual disclaimers apply.

Appendix A. Technical details on factorial analysis and gaps calculation

It is well known that nonparametric efficiency analysis gains in precision when working in space with lower dimensions (this is the usual "curse of dimensionality" of nonparametric techniques, see e.g. Daraio & Simar, 2007, for a discussion). In the application reported in this paper, the original data are transformed before entering into the analysis, to reduce the dimension of the problem (by using input and/or output factors as defined in Daraio & Simar, 2007). In this case of course, once the gaps have been computed for the variables used in the analysis, the researcher is willing to evaluate the corresponding gaps in the original inputs and outputs. This can be done by transforming back the gaps in the factors into the original units. We briefly explain how to achieve this.

Suppose we are able to reduce the dimension among a selection of inputs, because they are highly correlated. Denote the corresponding matrix of selected inputs by $\widetilde{X}$ that has n rows (the observations) and $\tilde{p}$ columns (we could follow exactly the same procedure for a subset of highly correlated outputs $\widetilde{Y}$). In this method (see e.g. Daraio & Simar, 2007; Härdle & Simar, 2012), the highly correlated $\tilde{p}$ columns can be replaced, without much loss of information by a single new variable through a linear combination. The best linear combination is given by the eigenvector of the matrix $\widetilde{X}'\widetilde{X}$ corresponding to its highest eigenvalue. We call this unique linear combination the "input-factor" $F_{\widetilde{X}}$. The ratios of the largest eigenvalue over the sum of all the $\tilde{p}$ eigenvalues allows us to appreciate the loss of information due to the reduction of dimension. In practice, this ratio should be large, say above 0.85, meaning that more than 85% of the total information shared by the $\tilde{p}$ original inputs is retained in this unique input-factor $F_{\widetilde{X}}$. Note also that if the columns of $\widetilde{X}$ are in different units, we scale them by their standard deviations to obtain unit free variables more adapted to linear combinations. The formal steps of this dimension-reduction are as follows:

[1]If needed, scale the columns of $\widetilde{X}$: $\widetilde{X}_s = \widetilde{X}\mathrm{diag}(1./s_{\tilde{x}})$, where diag(.) is a diagonal matrix, ./ is the Hadamard element-wise division between the vector of ones and the vector $s_{\tilde{x}}$ which is the vector of the empirical standard deviations of the $\tilde{p}$ columns of $\widetilde{X}$.
[2]The input factor is given by $F_{\widetilde{X}} = \widetilde{X}_s a_1$ where $a_1 \in \mathbb{R}^{\tilde{p}}$ is the eigenvector of $\widetilde{X}_s'\widetilde{X}_s$ corresponding to its largest eigenvalue λ_1.
[3]The percentage of inertia of this factor (percentage of information contained in the factor) is given by $\lambda_1/(\lambda_1 + \cdots + \lambda_{\tilde{p}})$. This percentage should be high enough to validate the procedure (say, above 80–85%).

In particular, for the inputs, we replace the 4 scaled inputs by their best (non-centered) linear combination, defined as *FINP*, as described in Table 1. In doing this analysis, we control that the information we loose in aggregating the variables is not too high. We also control the correlation of the resulting univariate input factor with the 4 original inputs, that should be high.

The obtained results are the following: $FINP = 0.48x_1 + 0.56x_2 + 0.52x_3 + 0.44x_4$, where we see that the factor is a weighted average of the 4 inputs. FINP explains 94% of total inertia of original data (correlations of the FINP with the original inputs are 0.93, 0.91, 0.98, 0.92). We follow the same procedure with the outputs. The results for the two factors are: $FRES = 0.70y_2 + 0.71y_3$, $FQUAL = 0.56y_4 + 0.51y_5 + 0.56y_6 + 0.33y_7$, where FRES and FQUAL are defined in Table 1. FRES explains 96% of total inertia of original data (correlations of the FRES with the original data are 0.96 and 0.96), while FQUAL explains 98% of total inertia of original data (correlations of FQUAL with original values are 0.7, 0.9, 0.9, 0.9).

So, in the analysis, the factor $F_{\widetilde{X}}$ will act as a single observed input and will be combined with other inputs (or other input factors) and outputs (or other output factors) along the lines of the techniques developed above. The gaps obtained at the end are thus in the units of the factors $F_{\widetilde{X}}$ used and not in the units of the original variable $\widetilde{X}$. We know that the value of the input factor variable on the efficient frontier is $\widehat{F}^{\partial}_{\widetilde{X}} = F_{\widetilde{X}} + G_F$. It is easy to check that the coordinates of $F_{\widetilde{X}}$ in the original units of $\widetilde{X}_s$ are given by $F_{\widetilde{X}}a_1'$. For the same reason, the coordinates of the frontier points are $\widehat{F}^{\partial}_{\widetilde{X}}a_1'$, so the measure of the gaps in the units of $\widetilde{X}_s$ are given by $G_{\widetilde{X}_s} = G_F a_1'$. Of course we have also to rescale back this solution, if step [1] above has been used. Finally, an estimate of the gaps in the units of the original $\tilde{p}$ input variables, for the n observations is given by:

$$G_{\widetilde{X}} = G_{\widetilde{X}_s}\mathrm{diag}(s_{\tilde{x}}) = G_F a_1' \mathrm{diag}(s_{\tilde{x}}). \tag{11}$$

References

Badin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, *223*, 818–833.

Badin, L., Daraio, C., & Simar, L. (2014). Explaining inefficiency in nonparametric production models: The state of the art. *Annals of Operations Research*, *214*, 5–30.

Baumol, W., Panzar, J., & Willig, R. (1982). *Contestable markets and the theory of industry structure*. New York: Hartcourt Brace Jovanovich.

Bonaccorsi, A., & Daraio, C. (2003). A robust nonparametric approach to the analysis of scientific productivity. *Research Evaluation*, *12*(1), 47–69.

Bonaccorsi, A., & Daraio, C. (2004). Econometric approaches to the analysis of productivity of R&D systems. Production functions and production frontiers. In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research*. Kluwer Academic Publishers.

Bonaccorsi, A., & Daraio, C. (2005). Exploring size and agglomeration effects on public research productivity. *Scientometrics*, *63*(1), 87–120.

Bonaccorsi, A., & Daraio, C. (2007). *Universities and strategic knowledge creation. Specialization and performance in Europe*. Cheltenham, UK: Edward Elgar Publisher.

Bonaccorsi, A., Daraio, C., & Simar, L. (2013). What is the impact of scale and specialization on the research efficiency of European universities? In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Horlesberger, & H. Moed (Eds.), (Vol. 2) *Proceedings of ISSI 2013. Vienna* (pp. 1817–1829). ISBN: 978-3-200-03135-7

Bonaccorsi, A., Daraio, C., & Simar, L. (2014). Scale and research specialization in European universities: A directional distance approach to teaching efficiency. In A. Bonaccorsi (Ed.), *Knowledge, diversity and performance in European higher education* (pp. 292–312). Cheltenham: Edward Elgar.

Brandt, T., & Schubert, T. (2014). Is the university model an organizational necessity? Scale and agglomeration effects in science. In A. Bonaccorsi (Ed.), *Knowledge, diversity and performance in European higher education* (pp. 233–266). Cheltenham: Edward Elgar.

Brinkman, P. T., & Leslie, L. L. (1986). Economies of scale in higher education: Sixty years of research. *Review of Higher Education*, *10*(1), 1–28.

Carayol, N., & Matt, M. (2006). Individual and collective determinants of academic scientists' productivity. *Information Economics and Policy*, *18*(1), 55–72.

CEDEFOP. (2014). *Terminology of European education and training policy – A selection of 100 key terms*. CEDEFOP, Luxembourg: Office for Official Publications of the European Communities. http://www.cedefop.europa.eu/EN/publications/13125.aspx,. Accessed 08.03.15

Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory*, *70*, 407–419.

Cohn, E., & Cooper, S. T. (2004). Multi-product cost functions for universities: Economies of scale and scope. In G. Johnes, & J. Johnes (Eds.), *International handbook on the economics of education* (pp. 579–612). Edward Elgar Publishing.

Cohn, E., Rhine, S. L. W., & Santos, M. C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. *Review of Economics and Statistics*, *71*(2), 284–290.

Daghbashyan, Z., Deiaco, E., & McKelvey, M. (2014). How and why does cost efficiency of universities differ across European countries? An explorative attempt using new microdata. In A. Bonaccorsi (Ed.), *Knowledge, diversity and performance in European higher education*. Cheltenham: Edward Elgar.

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. *Research Policy*, *40*, 148–164.

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis. Methodology and applications*. New York: Springer.

Daraio, C., & Simar, L. (2014). Directional distances and their robust versions: Computational and testing issues. *European Journal of Operational Research*, *237*, 358–369.

Dundar, H., & Lewis, D. R. (1995). Departmental productivity in American Universities: Economies of scale and scope. *Economics of Education Review*, *14*(2), 119–144.

Egghe, L., & Rousseau, R. (1990). Introduction to informetrics: Quantitative methods in library. In *Documentation and information science*. Amsterdam: Elsevier Science Publishers.

EUMIDA (2010). Feasibility study for creating a european university data collection, contract no. RTD/C/C4/2009/0233402, European Commission.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, (A(120)), 253–281.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production frontiers*. Cambridge University Press.

Färe, R., Grosskopf, S., & Margaritis, D. (2008). Efficiency and productivity: Malmquist and more. In H. Fried, C. A. Knox Lovell, & S. Schmidt (Eds.), *The measurement of productive efficiency*. Oxford University Press.

Glass, J. C., Mckillop, D. G., & Hyndman, N. S. (1995). The achievement of scale efficiency in UK universities: A multiple-input multiple-output analysis. *Education Economics*, *3*, 249–263.

Hanushek, E. A. (1986). The economics of schooling: Production and E. *Journal of Economic Literature*, *24*(3), 1141–1177.

Härdle, W., & Simar, L. (2012). *Applied multivariate statistical analysis* (3rd ed., pp. 458). Berlin: Springer-Verlag.

Hashimoto, K., & Cohn, E. (1997). Economies of scale and scope in Japanese private universities. *Education Economics*, *5*, 107–115.

Horta, H., & Lacy, T. A. (2011). How does size matter for science? Exploring the effects of research unit size on academics' scientific productivity and information exchange behaviors. *Science and Public Policy*, *38*(6), 449–460.

Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. *Economics of Education Review*, *21*, 63–71.

Johnes, J. (2004). Efficiency measurement. In G. Johnes, & J. Johnes (Eds.), *International handbook on the economics of education* (2nd ed., Vol. 2, pp. 613–742). Edward Elgar Publishing.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273–288.

Joumady, O., & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, *13*(2), 189–205.

Koshal, R. K., & Koshal, M. (1999). Economies of scale and scope in higher education: A case of comprehensive universities. *Economics of Education Review*, *18*, 269–277.

Laband, D. N., & Lentz, B. F. (2003). New estimates of economies of scale and scope in higher education. *Southern Economic Journal*, *70*, 172–183.

Longlong, H., Fengliang, L., & Weifang, M. (2009). Multi-product total cost functions for higher education: The case of Chinese research universities. *Economics of Education Review*, *28*, 505–511.

Lopez-Illescas, C., de Moya-Anegon, F., & Moed, H. F. (2011). A ranking of universities should account for differences in their disciplinary specialization. *Scientometrics*, *88*, 563–574.

Moed, H. F., de Moya-Anegon, F., Lopez-Illescas, C., & Visser, M. (2011). Is concentration of university research associated with better research performance? *Journal of Informetrics*, *5*(4), 649–658.

Qureshi, S., Hlupic, V., de Vreede, G. J., Briggs, R. O., & Nunamaker, J. (2003). Managing knowledge in a collaborative context: How may intellectual resources be harnessed towards joint effect? *Knowledge and Business Process Management*, 174–198.

Racine, J. (1997). Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics*, *15*(3), 369–378.

Racine, J. S. (2008). *Nonparametric econometrics: A primer*. Now Publishers Inc.

Shephard, R. W. (1970). *Theory of cost and production function*. Princeton, NJ: Princeton University Press.

Schimank, U., & Winnes, M. (2000). Beyond Humboldt? The relationship between teaching and research in European university systems. *Science and Public Policy*, *27*, 398–408.

Seglen, P. O., & Aksnes, D. W. (2000). Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research. *Scientometrics*, *49*(1), 125–143.

Simar, L., & Vanhems, A. (2012). Probabilistic characterization of directional distances and their robust versions. *Journal of Econometrics*, *166*, 342–354.
SIR (Scimago Institutional Ranking) World Report (2011).
Von Tunzelmann, N., Ranga, M., Martin, B., & Geuna, A. (2003). *The effects of size on research performance: A SPRU review, report prepared for the office of science and technology*. Department of Trade and Industry.
Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, *9*(3), 245–268.
Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education*, *61*(4), 387–414.