

The Evaluation of University Departments. A Case Study for Firenze

Lucia Buzzigoli · Antonio Giusti ·
Alessandro Viviani

Published online: 27 October 2009
© International Atlantic Economic Society 2009

Abstract Over the last two decades, the interest to assess the quality of university teaching and research has considerably grown. This paper presents a study concerning the evaluation of the departments of the University of Firenze using Data Envelopment Analysis. It shows several applications with different variables choices to assess the performance both in teaching and in research activities. The reliability of the preferred specification was verified with a heuristic experiment, using different variables and a different number of variables. Particular attention is given to the problem of data availability and quality (e.g. for research output assessment).

Keywords University evaluation · Efficiency analysis · Data envelopment analysis

JEL C10 · I00 · O50

Introduction

Recently, it has been considered more and more important to assess the quality of university research and teaching activities. The attention to quality enforcement is due to a different attitude towards the academic world, and this attitude has produced a new concern “about the increasing cost of funding university-based research [...] and the need to obtain ‘value for money’ for public expenditure on higher education” (OECD 1987). According to the Lisbon Agenda, Europe hopes to become “the most

L. Buzzigoli · A. Giusti · A. Viviani (✉)
Department of Statistics, University of Firenze, Viale Morgagni, 59, Firenze 50134, Italy
e-mail: viviani@ds.unifi.it

competitive and dynamic knowledge-based economy in the world”, yet, except for some excellent situations, European universities suffer from serious problems—the first being the lack of adequate funding (Lambert and Butler 2006).

In Italy, the evaluation process of academic activity, enforced in 1999, is currently relevant because universities are subjected to budget cuts by the government. In this context, a rising portion of the allocated financial resources is meant to be calculated on the base of effectiveness and efficiency indicators.

The evaluation among and inside universities (among departments and/or faculties) is a difficult task, because it is hard to identify and measure inputs and outputs of the underneath production process, especially when the research areas in which the different units are involved are dissimilar. An additional difficulty is given by the necessity to simultaneously consider the different inputs and outputs: in this step the evaluation process should aim to minimize the subjective component that is normally involved in multifactor approaches.

Many papers present a large number of performance evaluations applied to universities, faculties or departments. All of that research suffers from the complexity of university activity and organization and tries to find a solution with very different approaches. For example, as far as international comparisons are concerned, the Academic Ranking of World Universities compiled by the Shanghai Jiao Tong University (Shanghai Jiao Tong University 2007) is based on research output and citations plus the affiliations of Nobel Prize winners, while the Times Higher Education Supplement annually publishes a ranking of the world’s top 200 universities; this ranking is based on both quantitative data and on a ‘peer review’ system (Ioannidis et al. 2007). For Italy we can mention Agasisti and Dal Bianco (2006), who made a study on the efficiency of the 58 Italian public universities.

Substantially, the different approaches to evaluate efficiency in higher education are: the one based on indicators and their combination, the one based on models (for instance, by means of stochastic frontier approach), and the one based on linear programming methods.

In this paper, we made an application on the departments of the University of Firenze, and the work is focused on the efficiency in resource utilization. When comparing departments of different kinds, which can have very different production processes, the evaluation of academic performance can be rather difficult. For this reason, we decided to measure the efficiency of departments relative to a frontier constructed over the data by means of Data Envelopment Analysis (DEA), a method that doesn’t make explicit a specific production function and finds the frontier empirically.

The structure of this paper is the following. In the second section we briefly present the university system in Italy; in the third one we make a sketch of DEA. In the fourth section we describe the characteristics of the University of Firenze. In the fifth section we indicate the DEA specification, and in the sixth one we present the results of our exercise. Some concluding remarks complete the paper.

The University System in Italy

The Italian university system has been radically reformed since 1999 according to the objectives of the so-called Bologna Process, which aims at building a more

competitive European Higher Education Area by 2010. The reform process involves three main dimensions: the organization, the evaluation procedure and the funding scheme.

The university organization is based on faculties and departments. Faculties coordinate teaching activity, distributing responsibilities and workloads among teachers and researchers. Departments promote and manage research inside sectors that are homogenous by objectives or methods and gather teachers of the pertinent courses; furthermore, they organize doctoral programs, support teaching and carry out consultancy work. This kind of organization has important consequences on data capture activity for performance evaluation.

The teaching and research activities are monitored by a number of special bodies. Two of them are nominated by the Ministry of Education (MIUR): the National Evaluation Committee for the University System (CNVSU) and the National Committee for Evaluation of Research (CIVR). In each university there is an Internal Evaluation Committee (NUV) that has the duty to autonomously assess internal performance and to supply data and analysis to the CNVSU and to the CIVR. In 2004, the CIVR produced an evaluation of research products for each Italian university and for the different scientific areas (CIVR 2007).

The national funding system regards both ordinary endowment funds and research grants. In the first area the new legislative framework has introduced (in a significant component) some performance indicators in teaching and research as determinants of allocation of funds among universities. Research grants are partly originated by the MIUR (PRIN grants) and partly attracted (by Universities, Departments and single researchers) from private and public institutions (UE, banks, local authorities, etc.).

The Method

The DEA method,¹ introduced by Dantzig (1951) and Farrell (1957) and improved by Charnes et al. (1978), is a technique used to measure the performance of n production units or, more generally, of DMUs (Decision Making Units); given n production units transforming a number of inputs in a number of outputs, it identifies a non parametric piece-wise linear frontier—for each unit separately—which represents the best practice in input/output transformation.

DEA is a nonparametric method because no model specification is made for the production process. It considers simultaneously several inputs and several outputs and it does not require any pre-specified set of weights for them; the efficiency of each DMU is measured in relative terms and is determined by empirically comparing the input/output structure of the unit with all the others and finding the set of weights that are able to maximize the efficiency of the given DMU with respect to the others. Briefly, the procedure checks whether a theoretical best producer can be found which is able to produce the same quantity of outputs of the DMU with a smaller quantity of inputs (input oriented DEA) or that can produce a bigger quantity of outputs with the same quantity of inputs (output oriented DEA). In this sense, DEA

¹ The literature on DEA is really vast; for the model description we basically refer to Coelli et al. (2005).

is an extreme point approach, while other methods adopt a central tendency approach (comparing DMUs with an average DMU).

The DEA problem is solved separately for each of the DMUs by a linear programming algorithm that builds a non-parametric piece-wise surface (or frontier) over the data.² Assuming we have I DMUs producing M outputs with N inputs, the input oriented problem can be written in analytical terms as the following problem of minimum:

$$\begin{aligned} \min_{\theta, \lambda} \quad & \theta \\ \text{s.t.} \quad & Q\lambda \geq q_i \\ & \theta x_i \geq X\lambda \\ & \lambda \geq 0 \end{aligned}$$

Where θ is a scalar representing the efficiency of the i -th DMU; Q is the $M \times I$ output matrix; X is the $N \times I$ input matrix; λ is a $I \times 1$ vector of constants; the column vectors x_i and q_i respectively contain the inputs and the outputs for the i -th DMU.

The solution of the problem, θ , is 1 when the DMU under study is input efficient. When the resulting θ is less than 1 the difference $1-\theta$ represents the reduction in inputs (radial movement) which is necessary to project the inefficient DMU on the frontier. The vector λ contains the weights to apply to the DMUs which are efficient (and therefore represent the units of reference for the considered DMU) to build the optimal theoretical producer which is the benchmark for the DMU. Therefore, the first inequality states that the theoretic DMU should produce at least as many outputs as the considered DMU. The purpose of the second inequality is to find how much the input of the considered DMU could be decreased, according to the value of efficiency θ (because the model is input oriented). The point $(X\lambda, Q\lambda)$ represents the projection of the DMU on the frontier.

When the piece-wise frontier has sections which are parallel to the axes, the projection of inefficient points on the frontier can give solutions which in some cases are not efficient, because it is possible to further reduce one of the inputs (or increase one of the outputs) and produce the same outputs (or use the same inputs); this proportional reduction (augmentation) is called input (output) slack and can be evaluated with various methods.

This kind of model is associated with a technology with constant returns to scale (CRS, like in Charnes et al. 1978). DEA can also be applied with variable returns of scale (VRS; Banker et al. 1984), adding a convexity constraint to the above-mentioned system: $\mathbb{1}'\lambda = 1$ ($\mathbb{1}$ is a $I \times 1$ vector of ones). The hypothesis of VRS is often more realistic, because the dimension of the DMU can affect its productivity and, therefore, also its efficiency. When the size of the DMUs is heterogeneous each DMU should be compared only with the DMUs with a similar size. Note that in the VRS case the estimated frontier is closer to the original points; therefore VRS efficiency scores are higher than the corresponding CRS ones, which can suffer from scale inefficiency. Scale efficiency can be calculated as the ratio between total technical efficiency under CRS and pure technical efficiency under VRS.

² We used DEAP, a free software developed by T. Coelli that can be downloaded at the web site www.uq.edu.au/economics/cepa/deap.htm.

After the introduction of DEA, many different versions appeared in the literature (Adler et al. 2002), but for our application, we refer to the classical ones: the CCR (assuming constant returns of scale) and BCC (assuming variable returns on scale). The choice of the more adequate specification for our application (input/output oriented, CRS or VRS) will be made in accordance to the specific purposes of the analysis and to the characteristics of the data we are going to elaborate.

In the literature on academic performance, DMUs are often universities, while at local level the comparison is often at faculty levels (Flagg et al. 2004; McMillan and Chan 2005; etc.); the analysis of academic efficiency based on departments is less common, but it is of great interest (Moreno and Tadepali 2002; Kao and Hung 2008). On the one hand, at the department level we can observe all the dimensions of academic activity: teaching, research and administrative services that support them. When the DMUs are the faculties, however, only the teaching activity can be evaluated. Therefore, the DMUs considered in the analysis are the 70 departments of the University of Firenze.

The University of Firenze: The Available Data and The Choice of Variables

The university administrative files contain the core data on teaching, research and administrative activities: the number of teachers and researchers, number of students, administrative staff, expendable funds, PhD students, research grants, etc.. Some of these information is currently produced at the department level (like the number of teachers and the amount of grants), while others are available only at faculty level (like number of students). Some other information, which is published only at the faculty level, can be obtained from microdata: this is the case, for instance, of teaching workload.

In the academic year 2006/2007, the University of Firenze was the sixth university in Italy in terms of number of students (59,729) and of number of teachers and researchers³ (2,271); it included 12 faculties with 105 first level degree courses, 110 second level degree courses, and 95 PhD courses, and it was structured in 70 departments. For organizational purposes, the departments are classified in five scientific areas: bio-medical, scientific, technological, socio-economic, and liberal arts.

The amount of teaching activity is related to the number of teachers (Table 1). The average number of teachers of the departments is about 33; the minimum is 8, and the maximum is 104. 50% of departments has 30 people or less.

The teaching burden is usually measured in credits: since the amount of credits represents the workload for students and not for teachers, and because each degree course may have a different amount of hours per credit (from 5 to 15), in our application credits have been converted into lecture hours. Figure 1 presents the number of teaching hours per teacher per department.

Another way to quantify the teaching burden should be the number of students, but this kind of information is available only for faculties or for degree courses.

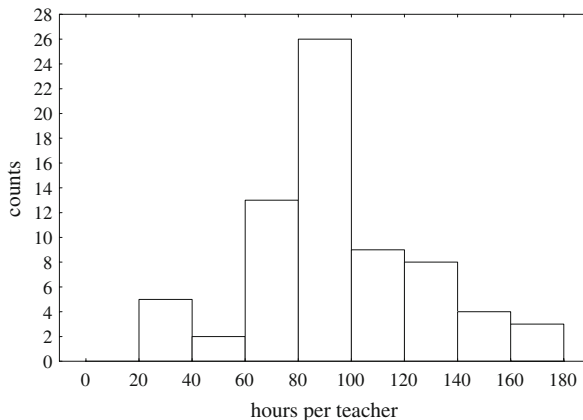
³ Source: MIUR database: <http://statistica.miur.it>.

Table 1 Departments per area and per number of teachers

No. of teachers	8–20	21–30	31–40	41–50	51–60	61–70	71–104	Total	Avg. no. of teachers
Bio-medical	6	5	3	2	0	1	1	18	30.9
Scientific	2	3	1	3	1	0	2	12	38.7
Technological	11	0	1	5	1	1	0	19	29.6
Socio-economic	1	5	3	1	1	0	0	11	30.7
Liberal arts	2	1	4	2	1	0	0	10	36.0
Total	22	14	12	13	4	2	3	70	32.6

The academic staff is essential also in the evaluation of research activity, together with the amount of external grants and research achievement. As far as research grants are concerned, departments benefit from a number of external sources. Here, we consider PRIN government grants, because their amount is an important indicator of university research capability: these funds promote the reputation of the departments attracting external research staff and more grants. Since these funds have a two-year time horizon, and members of a financed project cannot participate to the next year's call, we have considered the average of the funds for two academic years (2005/2006 and 2006/2007). The frequency distribution of PRIN funds is represented in Fig. 2: nearly one fourth of the departments receive less than € 100,000 and more than 62% receive less than € 300,000.

Research achievement is usually estimated by means of research products evaluation, based both on quantity and quality issues. Since publications have different quality and impact, and the different research areas do not have a common basis for comparison, in this paper we refer to a two-step productivity/quality index, which derives from CIVR evaluation and has been estimated by the NUV of the University of Firenze. The first step consists in calculating a "CIVR index" at department level, adding up the CIVR evaluations of the products of the researches

**Fig. 1** Teaching hours per teacher per department

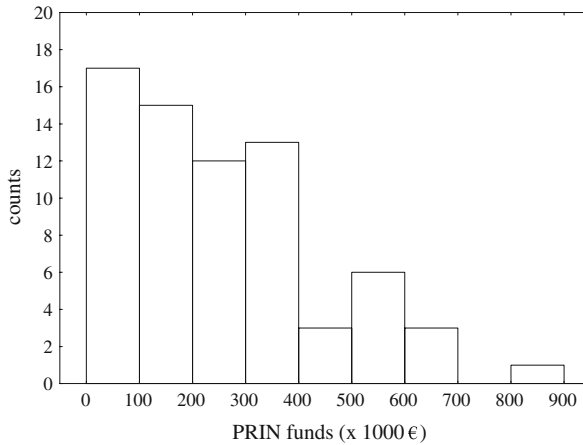


Fig. 2 Distribution of PRIN funds

by the departments (every product gets an evaluation among 0.4—limited— and 1.0—excellent). This absolute index does not consider the different number of researchers in the departments and that the total production of each department should be compared and ranked in national panels of corresponding research areas. In the second step the index is properly weighted to consider both the dimension of departments and their position in the national area rankings. We call this measure Research Productivity Index (RPI) (NUV 2007). The index doesn't have a fixed range. Figure 3 shows the distribution of RPI per department: more than 50% of them obtained two or less and about 95% got less than six.

The ordinary endowment funds have been considered as a measure of administrative activity which supports teaching and research: the average value is €35,000, but the values are rather spread among the departments (from €9,508 to €112,186—Fig. 4).

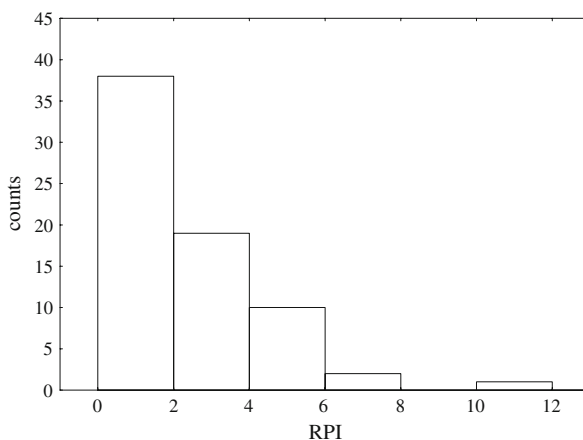


Fig. 3 Frequency distribution of RPI per department

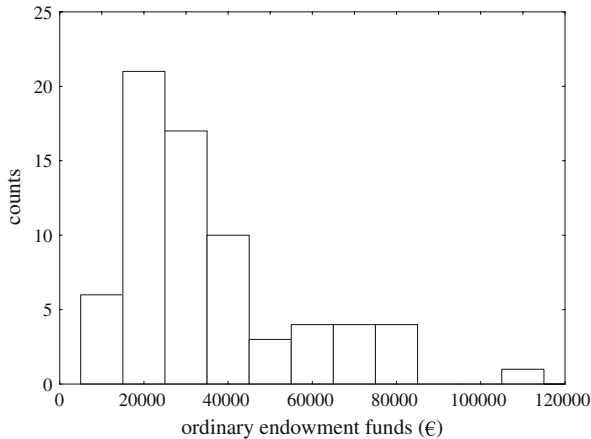


Fig. 4 Ordinary endowment funds per department

Other information can be drawn from administrative datasets at department level—for example, the number and the professional level of administrative staff, the number of Ph.D. students and research work contracts, European Union research grants, number of theses, etc. However, as we will see later, they will not be so useful in our analysis.

The Model Specification

The results of DEA depend heavily on the set of input and output variables chosen in the application. Teaching and research are essentially connected to both input and output, while administrative activity is strictly considered as an input factor.

Various methods have been proposed to facilitate the choice of variables, and they can be classified in a variety of ways: judgemental (using the opinion of experts) versus quantitative (such as regression or correlation techniques); ex-ante (when the variables are chosen prior the DEA application) versus ex-post (when the selection is made after considering the results stemming from different models); statistical versus non-statistical, etc. For a recent, concise review of methods see Wagner and Shimshak (2007). A general consensus on specific methods is still missing because the final choice is often determined by the actual availability of variables at the fixed DMU level.

Moreover, two other problems must be faced. The first one relates to the circular meaning of some core variables. This is the case, for instance, of research grants, which can be considered as a measure of research output (because they are assigned to virtuous researchers) and also as representative of an input because they support research projects. In our application, research grants represent an output variable, because we want to stress their importance as indicators of departments' research capability. In particular, we refer to PRIN grants, which are assigned periodically to all Italian universities with the same procedure. The second issue refers to the measure of intangible aspects, which are particularly relevant in output specification.

Let's consider, for instance, the problems in evaluating the quality of teaching and of research, or of the academic reputation. In this work, RPI can be considered the best measure available of research quantity and quality, while there is no information about teaching quality.

According to the previous considerations, the variables used in the DEA application are the following: MIUR grants, RPI, and teaching hours as output variables, as well as number of teaching personnel and the amount of operating expenses as input variables. In particular, the teaching personnel represents an essential input both for teaching and research, while operating expenses is considered as an input supporting all the different activities of the departments.

The average of each variable for the departments of each area are reported in Table 2.

A cost saving approach appears to be a more rationale managerial objective, because in Italy the university sector suffers from a diffused financial crisis which is going to produce a reduction in the available resources: in this framework an input oriented DEA is preferable.

The Results

A preliminary note is necessary before the presentation of the results for the proposed DEA specification (in the following denoted as model D1): since DEA results depend heavily on variables choice, we have to evaluate D1 appropriateness by means of a reliability analysis.

On the one hand, we modified the variables maintaining their total number. In particular, in the alternative DEA specification labelled D1a, instead of using the number of teachers, we used the 'weighted' teaching personnel where full professors are weighted 1, associate professors 0.7 and researchers 0.5, according to their relative cost. On the other hand, in the DEA specification labelled D1b, we modified the number of variables of model D1, adding the output variable number of examinations, as a proxy for the number of students, which are not available at department level.

Since the interest was mainly focussed on the stability of rankings and not on the stability of efficiency values, we computed Spearman ranking coefficient of D1 against D1a and D1b. Figure 5 contains the scatterplots for efficiency measures of

Table 2 Variables used in the DEA application (avg. per department per area)

AREA	Grants avg	Teach.Hours avg	Teachers avg	Ord.Funds avg
Bio-medical	230.9	2,486	30.9	35,211.6
Scientific	323.8	3,650	38.7	47,758.6
Technological	260.7	3,451	29.5	37,268.5
Socio-economic	197.1	3,007	30.7	25,250.8
Liberal arts	269.6	3,061	36.0	26,472.4
Total	255.1	3,111	32.6	35,107.1

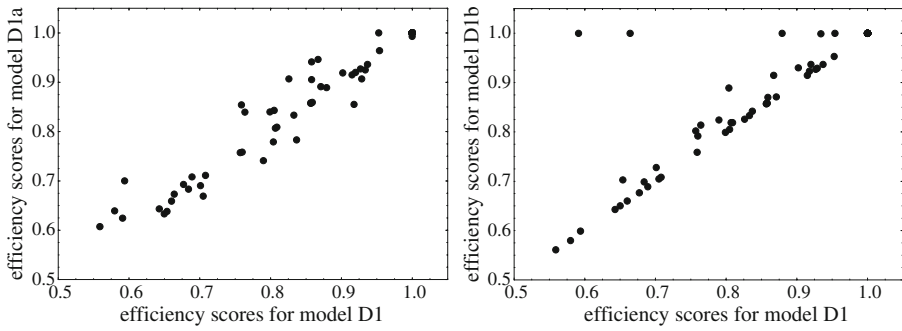


Fig. 5 Efficiency scores: D1 vs D1a model and D1 vs D1b model

D1 vs D1a and D1 vs D1b. The corresponding Spearman coefficients are $\rho_S(D1, D1a) = 0.98$ and $\rho_S(D1, D1b) = 0.88$.

In the D1b case the DMUs no.41 and 69 have a low efficiency score in D1, but are efficient in D1b; after erasing these two outlier observations, the Spearman coefficient becomes 0.98.

Other model specifications gave similar results; therefore, the D1 specification seemed sufficiently stable. We thus briefly consider the main results for the 70 departments.

Table 3 presents a summary of efficiency scores resulting from D1: the efficient DMUs are 24 (34% of DMUs) and the average efficiency for the 70 DMUs is 0.861, while the average for the 46 non-efficient DMUs is 0.789. We can also see that only 15 of the 70 departments have an efficiency score less than 0.75 (21.4%).

To evaluate the differences among the five research areas and to appreciate the dissimilarity amid all the DMUs and the not efficient ones, Table 4 presents the efficiency scores per area. The technological area has the maximum number of peers (53% of technological departments) and the highest efficiency average. It has the best results even when considering only the non-efficient DMUs. We can also see that in the Technological, Socio-economic, and Liberal Arts areas, the differentiation

Table 3 Model D1: departments per VRS efficiency scores

Efficiency scores	Count	Cum.	%	% Cum.
.55≤x<.60	4	4	5.7	5.7
.60≤x<.65	1	5	1.4	7.1
.65≤x<.70	7	12	10.0	17.1
.70≤x<.75	3	15	4.3	21.4
.75≤x<.80	6	21	8.6	30.0
.80≤x<.85	7	28	10.0	40.0
.85≤x<.90	8	36	11.4	51.4
.90≤x<.95	8	44	11.4	62.9
.95≤x<1.0	2	46	2.9	65.73
1.0	24	70	34.3	100.0

Table 4 Model D1: efficiency scores statistics per area for all the 70 DMUs and for the not efficient DMUs

AREA	N. of peers	% (on DMUs per area)	All the DMUs			Not efficient DMUs		
			avg	N	Std dev	avg	N	Std dev
Bio-medical	4	22	0.83	18	0.14	0.79	14	0.12
Scientific	2	17	0.81	12	0.14	0.77	10	0.12
Technological	10	53	0.93	19	0.11	0.85	9	0.11
Socio-economic	4	36	0.85	11	0.14	0.76	7	0.10
Liberal arts	4	40	0.86	10	0.15	0.77	6	0.12
Total	24	34	0.86	70	0.14	0.79	46	0.11

among all the DMUs and the non-efficient ones are quite larger than in the Bio-medical and in the Scientific areas.

A more general observation concerns the possibility to identify an optimal dimension of DMUs in terms of resources and/or activities. The results of our application do not give information about this aspect, according to one or more variable. In particular, Table 5 shows the number of teaching personnel of the peers in the different areas and points out that peers have very different size (from 8 to 77).

Table 6 shows some statistics for the actual DMUs and the virtual DMUs obtained with input adjustment according to D1 model: they highlight the amount of the possible reduction in the number of teachers and in the total of ordinary funds.

In order to verify whether teaching and research have a different impact on DMUs efficiency, we developed two other DEA versions with the same input variables as D1, but different outputs: model D2 is teaching oriented and its output variables are number of exams and teaching hours; model D3 is research oriented and its output variables are grants and RPI.

D2 identified 11 peers, while D3 17. The Spearman correlation coefficient between the efficiency rankings of both models with D1 ranking are: $\rho_S(D1,D2) = 0.42$, $\rho_S(D1,D3) = 0.71$; while $\rho_S(D2,D3) = -0.0066$. The corresponding scatter is in Fig. 6 and shows a particular arrow shape, because many DMUs that are fully efficient in one model are not efficient in the other. Moreover, some of D2 peers also

Table 5 Model D1: number of peers and teaching personnel in peers per area

AREA	N	% of peers	Number of teaching personnel
Bio-medical	4	22	8; 12; 33; 44
Scientific	2	17	12; 77
Technological	10	53	12; 14; 15; 17; 18; 20; 42; 43; 64
Socio-economic	4	36	15; 25; 33
Liberal arts	4	40	17; 38; 39; 44
Total	24	34	

Table 6 Model D1 input variables: statistics for actual and virtual DMUs

		Avg	Median	Sum	Min	Max	Std
Teaching	Actual	32.6	27.5	2,279.0	8	104	18.1
	Virtual	27.1	24.6	1,895.8	8	77	14.5
Ord.Funds	Actual	35,107.1	30,335.0	2,457,495.0	9,508	112,166	21,155.7
	Virtual	28,977.1	24,303.0	2,028,394.0	9,508	78,270	17,665.1

have a very low efficiency score in D3 and vice versa. Therefore, in the departments of Firenze University there is a trade-off between teaching and research efficiency.

This problem also affects the dimension of the virtual DMUs. Table 7 compares some statistics computed on teaching personnel and ordinary funds in the actual DMUs and in the virtual DMUs obtained with the three models. The approach, which implies the highest reduction in inputs, is D2, the teaching oriented one.

As far as the dimension of peers is concerned, considering only teaching or only research, the number of teaching personnel turns out to be rather diversified (Table 8). Three DMUs are peers in all the models used in the analysis (D1, D2 and D3); all the other peers of D3 are also peers in D1, while only five of the peers in D2 are also peers in D1.

Concluding Remarks

In this paper we used DEA to evaluate the efficiency of the departments of the University of Firenze. We implemented several applications with different variable choices to assess DMUs’ performance, both in teaching and in research activities. We made the evaluation through ranks and verified the reliability of the preferred

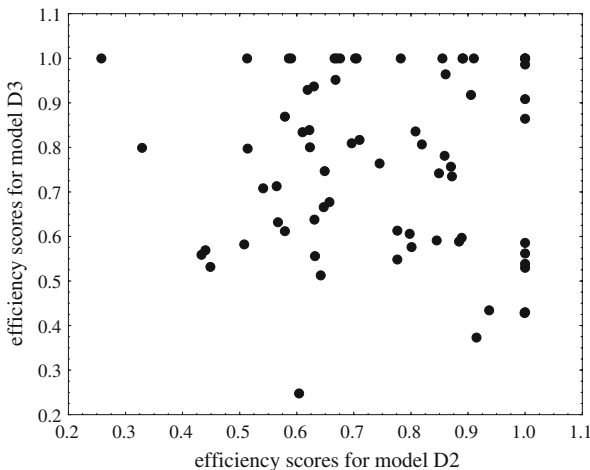


Fig. 6 Efficiency scores: D3 model vs D2 model

Table 7 Teaching personnel and on ordinary funds: statistics for actual and virtual DMUs

	Teaching personnel						Ordinary funds					
	Avg	Median	Sum	Min	Max	Std	Avg	Median	Sum	Min	Max	Std
Actual DMUs	32.6	27.5	2,279	8	104	18.1	35,107.1	30,335	2,457,495	9,508	112,166	21,155.7
Virtual DMUs D1	27.1	24.6	1,896	8	77	14.5	28,977.1	24,303	2,028,394	9,508	78,270	17,665.1
Virtual DMUs D2	22.9	17.8	1,601	8	64	12.9	22,925.0	18,795	1,604,748	9,508	72,691	14,309.9
Virtual DMUs D3	23.0	20.0	1,611	8	77	12.4	25,285.8	20,237	1,770,009	9,508	78,270	16,164.4

Table 8 Models D2 and D3: number of peers and teaching personnel of peers per area

AREA	D2			D3		
	N	% of peers	Number of teaching personnel	N	% of peers	Number of teaching personnel
Bio-medical	3	16.7	8 ; 33; 48	3	16.7	8 ; 12; 44
Scientific	1	8.3	12	2	16.7	12 ; 77
Technological	4	21.1	<i>17</i> ; 18; 46; <i>64</i>	6	31.6	<i>12</i> ; <i>14</i> ; <i>15</i> ; <i>15</i> ; <i>20</i> ; <i>42</i>
Socio-economic	1	9.1	22	3	27.3	<i>25</i> ; <i>25</i> ; <i>33</i>
Liberal arts	2	20.0	38; 39	3	30.0	<i>17</i> ; 39 ; <i>44</i>
Total	11	15.7		17	24.3	

bold = peer in D1, D2, D3; *italics* = peer in D1, and in D2 or in D3

specification with a heuristic experiment, using different variables and a different number of variables. The resulting scores were dissimilar, but there were no valuable variations in the ranks across the different specifications.

The DEA method doesn't refer to a specific theoretical model that can help in interpreting results, but some patterns were empirically identified: peers do not have a typical dimension, and peers of the model related only to teaching activity differ from peers of the model related only to research. Moreover, all the last ones are also peers in the general model including both teaching and research.

Inefficiency can be related, in general, to the fact that teaching personnel have different impacts on the results. The number of academic staff is usually linked to teaching activity and has a weaker link to research activity and, in particular, to related outputs (at least as measured in our application). This means that at present, in the University of Firenze, the allocation of academic staff is more related to teaching activity rather than to research results.

The measure of research output is undoubtedly one of the most critical aspects of this kind of analysis, because there is no general agreement on it and different measures can produce very different results. In our case the problem is even more difficult because the departments belong to different research areas: this implies that the limited number of input and output variables actually refers to a heterogeneous mix of disciplines and activities and this may affect the results. Different areas need different kinds of research facilities (laboratories, libraries, equipments, etc.) and require a different amount of grants; RPI is not influenced by the research characteristics but only by research results. In this sense PRIN grants and RPI can give a satisfactory synthesis of research outcome.

Furthermore, besides the problems related to measurement issues and consequently to variable specification, the analysis encourages some possible research developments from both an empirical and a methodological point of view. For example, both a temporal and a cross-universities approach would be necessary. In the first case the analysis could point out efficiency improvement or decrease in the different DMUs while in the second case the bias caused by the discipline mix could be partly mitigated comparing departments of different universities with similar research/teaching areas.

Finally, further performance specifications could be developed, expanding efficiency evaluations to effectiveness evaluations (e.g. Golany and Tamir 1995), according with the recent Italian laws on budget government allocation to universities.

References

- Adler, N., Friedman, L., & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, 140, 249–265.
- Agasisti, T., & Dal Bianco, A. (2006). Data envelopment analysis to the Italian university system: theoretical issues and policy implications. *International Journal of Business Performance Management*, 8(4), 344–367.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Models for estimation of technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078–1092.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.
- CIVR (2007) *Relazione finale. VTR 2001-2003*, Roma.
- Coelli, T. J., Prasada Rao, D. S., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). New York: Springer.
- Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. In T. Koopman (Ed.), *Activity analysis of production and allocation*. New York: John Wiley and Sons, Inc.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A, CXX, Part, 3*, 253–290.
- Flagg, A. T., Allen, D. O., Field, K., & Thurlow, T. W. (2004). Measuring the efficiency of British universities: a multi-period data envelopment analysis. *Education Economics*, 12(3), 231–249.
- Golany, B., & Tamir, E. (1995). Evaluating efficiency-effectiveness-equality trade-offs: a data envelopment analysis approach. *Management Science*, 41(7), 1172–1184.
- Ioannidis, J., Patsopoulos, N. A., Kavvoura, F. K., Tatsioni, A., Evangelou, E., Kouri, I., et al. (2007). International ranking systems for universities and institutions: a critical appraisal. *BMC Medicine*, 5, 30.
- Kao, C., & Hung, H. T. (2008). Efficiency analysis of university departments: an empirical study. *Omega*, 36, 653–664.
- Lambert, R., & Butler, N. (2006). *The future of European universities. Renaissance or decay?* London: Centre for European Reform.
- McMillan, M. L., & Chan, W. H. (2005). University efficiency: A comparison and consolidation results from stochastic and non-stochastic methods. Working Paper Series, 2005-04, University of Alberta, Department of Economics.
- Moreno, A. A., & Tadepali, R. (2002). Assessing academic department efficiency at a public university. *Managerial and Decision Economics*, 23, 385–397.
- NUV-Nucleo di Valutazione dell'Università di Firenze (2007). *Rapporto 2005/2006*, Università di Firenze.
- OECD. (1987). *Universities under scrutiny*. Paris: OECD.
- Shanghai Jiao Tong University (2007). *Academic ranking of world universities*, Institute of Higher Education, Shanghai Jiao Tong University.
- Wagner, J. M., & Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives. *European Journal of Operational Research*, 180, 57–67.