University departments evaluation: a multivariate approach

Monacciani, Fabiana

Department of Social Sciences, University "L'Orientale", Naples, Italy

15 April 2010

University Departments evaluation: a multivariate approach[1]

Abstract

Aim of the paper is to present a new model, based on multivariate statistic analyses, allowing to express a synthetic judgement on Departments activities by taking into consideration the whole set of indicators describing them both as aggregations of researchers and as University autonomous organs.
The model, based on Principal Component Analysis and Cluster Analysis, allows both to explain the determinants of Departments performances, and to classify them into homogeneous groups.
The paper shows the results obtained by testing the proposed model on University of Naples "L'Orientale" Departments, using data extracted by the 2007 assessment report to the Ministry of University and Research.

Keywords: Evaluation, Departments, Multivariate statistics.

The evaluation of scientific research at different levels: state of the art and advances

In the last years, the issue about the evaluation of scientific research has become quite debated both in economic literature (Hirsch, 2005; Marcuzzo and Zacchia, 2007) and in scientific international forums (EUA, 2009; OECD, 1997; CSS, 2006).
Research evaluation procedures may refer to different situations and could be classified according to various parameters:
a) aim of evaluation (financing allocation, career progression, retributions);
b) subject evaluated (individuals, Departments, Universities, Research Country Systems);
c) evaluative methodology (peer review, bibliometric criteria, mixed methods);
d) objects to evaluate (articles, monographs, patents, other research products);
e) disciplines and scientific areas (technological, scientific, socio- humanistic).
Each of these evaluation fields bring crucial distinctions and require appropriate analyses and specific proposals; by the way, this paper mainly focuses on points b) and c).
Depending on the evaluation level - single researchers, Departments, Universities, Countries -, assessment procedures and used evaluation tools are very different from each other.
Country-level evaluation, for example, is usually performed through benchmarking procedures, mainly based on descriptive statistics (OECD, 2009), taking into consideration input indicators such as GERD (Gross Domestic Expenditure on R&D) and human resources devoted to research, and output data like patents, publications, laureates (Rizzuto, 2003) and publications citations (King, 2004); in other cases, though, Country rankings are made according to aggregated indicators, such as the EIS (European Innovation Scoreboard) (European Commission, 2009) and the GCI (Growth Competitiveness Index) (World Economic Forum, 2009; Dubini, 2005) indexes, synthesizing several aspects of national research performance on innovation and competitiveness.
University-based evaluations, instead, are normally implemented in order to rank such institutions with reference to a weighted sum based synthetic score, taking into consideration their performance in research and teaching activities. The ranking of Higher Education Institutions (HEIs) has earned a lot of attention in recent times. Prospective students, especially post graduates, use rankings to get an idea of a University 's relative performance; this, in turn, encourages public bodies to pay attention to rankings when allocating funds to HEIs.
At present, two of the most important international University rankings are, for example, the ones produced annually by the Shanghai Jiao Tong University, the ARWU (Academic Ranking of World Universities), and

by the Times Higher Education, the THES (Times Higher Education Supplement) World University Ranking. Despite their popularity, such rankings have come under some criticism both regarding their methodology and choice of variables (Liu and Cheng, 2005;Van Raan, 2005), both concerning the fact that the use of an average score to measure performance has a determining influence on the ranking (Vinke, 2009).

Some of the most important Italian experiences in evaluating Universities performances, instead, are those performed since 2009 by MIUR (Italian Ministry for University and Research) for the 7% FFO (Ordinary Financing Fund) attribution (MIUR, 2009), based on the evaluation of Universities research performance and teaching efficacy, and the one made by CIVR (Italian Committee for Research Evaluation) during the 2000-2003 Triennial Research Evaluation (VTR) *(CIVR,2003)*, based only on research performance measured through publications peer review quality assessment.

As for international rankings methodologies, such procedures, too, have been hardly criticized (Macerata University, 2009; Lippi and Peracchi, 2007).

At present, the main debates about the evaluation of scientific research are focused on performance assessment of single researchers, in order to judge about their quality and their career progressions. Referring to this issue, the most widespread opinion is that the evaluation of research activity should be done quantitatively (Hicks, 2007), by using numerical indexes, mainly based on bibliometric data, whose advantage is that of being simply constructible and easily comparable (Narin and Hamilton, 1996; Van Leeuwen, 2006); according to others (Figà Talamanca, 2009), though, qualitative panel evaluation of self-selected publications or quali-quantitative evaluation (informed peer review)[2], by using a mixture of the two former approaches, could help overcoming the ever cited limits of bibliometric evaluation[3] (Adler *et al*,2008; Frey and Rost, 2008; Moed, 2008; Seglen, 1997)[4].

While for Country, University and researcher-based assessment, vast literature and numerous applicative examples could be found, the evaluation of Departments performance, instead, is not yet a wide explored field.

Referring to such evaluation level, almost all independent studies consider only bibliometric, technometric or productive indices (Narin and Hamilton, 1996; Cugini and Michelon, 2009), and describe Departments performance by synthetic scores, calculated by simply summing up the individual score gained by each researcher belonging to them.

According to Italian University Council (CUN) (MIUR, 2010), too, the comparison of Departments operating in the same scientific field could be plausibly made mainly basing on bibliometric indexes[5].

With reference to administrative praxis, the most well known Department evaluation procedure, the UK Research Assessment Exercise (RAE, 2007), is exclusively based on the mere evaluation of researchers publications, and funding allocation among Departments is defined according to the relative score gained by each.

According to us, such procedures, have the problem of giving only partial indications of Department productivity and efficiency (Guan and Wang, 2004) as they take into consideration only one aspect of Departments activities, namely the publication performance of single researchers, while they completely omit the evaluation of others.

Differently from the UK, in Italy Departments performance evaluation is remitted to University internal evaluation panels, the NVAs[6] (Nuclei di Valutazione di Ateneo), generally aiming at assessing two main aspects: 1) Departments capacity in attracting external financing; 2) Department performance in research activities. The two aspects are normally measured by respectively quantifying the amount of external financing gained by all the belonging researchers, and the number of publications produced by them (sometimes weighted according to the typology of publication, the number of citations etc.); only few cases of self selected publications qualitative evaluation could be found in practice (Polytechnic of Turin; University of Trento, University of Pavia, University of Siena). In some cases, though, other indicators describing Department's performances, as the number of research projects presented, the resources spent for publications etc., are taken into consideration.

Starting from indicators referred to each researcher, Department's indicators are normally calculated either as the mean value of researchers belonging to it, or as the sum of the latters (MIUR, 1998); Departments global performance, then, is described through the use of statistic descriptive tools.

Even though such procedure is more exhaustive than that based only on the evaluation of publications quality, we think that two main limits could still be identified:

1) Departments are only treated as a mere sum of researchers whose scientific performance should be evaluated; no consideration is ever given, instead, to the fact that Departments are Universities' autonomous bodies, whose tasks and performances must be assessed in a whole;
2) the use of descriptive statistic methods to evaluate Departments is completely unsuitable as it is incapable both of synthesizing the whole available information and to give an overall judgement on Departments performance.
Referring to such limits, we think that advances in Department evaluation could be done by taking into consideration, in addition to bibliometric and researcher-related indexes, other indicators able to describe some important Department features such as: coordination and research promotion capacity, research policy setting competence, diffusion of research products stimulation capacity, young researchers training efficiency, ability in promoting collaborations and integration among its researchers and with other research bodies. For such reason, some *ad hoc* indicators able to describe Departments performance as unitary and independent subjects (distinct by its belonging researchers), endowed by autonomous and distinct personality, should be defined.
In addition, instead of descriptive statistics, multivariate analyses could be more profitably used to express a synthetic judgement on Departments performance by taking into consideration a wide set of indicators describing the various aspects in which Departments activities concretizes[7].
In this framework, the aim of this paper is to present a new model, based on multivariate statistics, allowing to synthesize Departments performance by taking into consideration a set of diversified indicators, and to classify them into homogeneous clusters, in order to point out the main differences occurring between the most and the less "virtuous" ones.
The idea of a multidimensional model for evaluating Departments performance results from the awareness that the already used methods are still far from grasping the real polyhedral essence of Departments performance, relying mainly on publication data, and it is an attempt to overcome the limits of the most widespread evaluation procedures.

Methodology

The model is based on Principal Component Analysis (PCA) and Cluster Analysis (CA) (Gallo, 2007; Gherghi and Lauro, 2004).
The first aims to describe variability among many observed variables in terms of fewer unobserved variables called "factors". Such factors are not directly observable and are modeled as linear combinations of the observed variables, plus "error" terms.
The information gained about the interdependencies can be used later to reduce the set of original variables. PCA is very useful when relations among statistic unities cannot be efficiently interpreted because of the high number of variables considered. If original space dimension is reduced, it is possible to interpret a reduced number of components instead of a high number of variables.
In other words, PCA helps to synthesize and interpret a complex phenomenon, in our case the "performance" of Departments.
Cluster analysis methodology, then, allows to identify Departments with homogeneous characteristics, by considering the multiplicity of possible determinants of Departments performance as variables. Such a technique allows to identify the groups of Departments that are characterized by the same distinctive elements - and, therefore, have interior homogeneous characteristics - and that present the highest heterogeneity with other groups. As a result, a homogeneous class will be described as a combination of variables, allowing both to identify the strengths and the weaknesses of each group and to understand the main differences occurring among the most and the less "virtuous" Departments.
Both PCA and CA can be performed by statistical package SPAD.

Case study

The proposed model has been tested on data extracted by the University of Naples "L'Orientale" NVA's 2007 assessment report to the MIUR; in this article, though, in order to avoid reputational effects of

“L’Orientale”’s Departments linked to the output of the model, input data and results are not directly referable to the real situation.
Analysis is conducted on 30 indicators, identified as able to describe some key factors of Departments performance: the first 18 indicators describe Departments as aggregations of researchers[8], while the last 12 indicators describe Departments as University bodies[9] with their own personality and tasks.
Such indicators have been classified into 5 categories, according to the different aspects they are related to (tab. 1):

- the category "research performance" groups the indicators more directly related to Departments research activities. In particular, indicators from 1 to 9 refer to Department performance in PRIN[10] projects; indicator 10 is referred to researchers' productivity, measured in terms of weighted publications; indicators from 11 to 14 are related to financial aspects linked with research ability of Department members (external resources attracted, financial resources granted);
- the category "dimension" contains three indicators describing Departments size: number of researchers belonging to them and number of scientific sectors covered; indicator 17 considers as a size proxy the percentage of athenaeum financing addressed to each Department;
- the category "teaching" contains one indicator describing the working burden deriving from teaching activities (number of CFU taught by Department researchers);
- the category "body tasks" contains eight indicators describing Department ability in training young researchers (indicators from 19 to 21 take into consideration department effort in young researchers training, through Phd and research scholarships), in setting research policy and in coordinating research activity (indicators from 22 to 26);
- the category "organs functioning" groups a set of 4 indicators describing the efficiency in Departments organs functioning.

In order to obtain a synthetic judgment on Departments performance and an indication of the multiple relations existing among the variables, so as to better understand the analyzed phenomenon, Principal Component Analysis has been performed on dataset in table 1.
Analysis made on active variables[11], brought to the construction of six principal components, with a total eigenvalue higher than one.
To show the results on a two dimension plan, consideration has been given only to the first two axes, thus maximizing the phenomenon explained variance - as we can see in table 2, in fact, the first two factorial axes have a very high explanatory relevance, (about 58% of phenomenon variability) as they express more than the half of information given by the 30 original variables -.

Tab 1: Dataset

Variables					Variable labels	Departments								
						A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Departments as aggregations of researchers	Research		1	N° researcher participating to presented PRIN/tot Department's researchers	RES_PRES_PRIN/DEP_RES	106	115	84	97	**132**	102	74	101	66
			2	N° researchers participating to approved PRIN/tot Department's researchers	RES_APP_PRIN/DEP_RES	121	132	0	**134**	130	92	108	73	52
			3	N°researchers participating to positive not financed PRIN/tot Department's researchers	RES_POS_NOT_FIN_PRIN/DEP_RES	97	103	139	91	**132**	123	36	94	92
			4	N° researchers participating to rejected PRIN/ tot Department's researchers	RES_REJ_PRIN/DEP_RES	97	107	118	28	138	70	103	**185**	24
			5	N° researchers participating to rejected PRIN/ N° researchers participating to presented PRIN	RES_REJ_PRIN/RES_PRES_PRIN	92	93	**141**	29	105	68	139	183	36
			6	N° PRIN presented/tot Department's researchers	PRIN_PRES/DEP_RES	133	138	38	90	**191**	127	60	49	81
			7	N° PRIN approved/tot Department's researchers	PRIN_APPR/DEP_RES	**145**	125	0	130	177	112	105	47	31
			8	N° weighed PRIN presented/tot Department's researchers	W_PRIN_PRES/DEP_RES	123	148	30	81	**201**	125	55	56	81
			9	N° weighed PRIN approved/tot Department's researchers	W_PRIN_APPR/DEP_RES	113	142	0	129	**177**	139	82	46	24
			10	Total weighed publications score/ researcher	W_PUB/RES	105	63	115	123	115	**174**	25	79	133
			11	Research external financing/tot Department's researchers	EXT_FIN/DEP_RES	21	151	80	**231**	55	149	75	4	28
			12	Research external financing/ tot.University financing	EXT_FIN/TOT_UNI_FIN	19	142	83	**231**	49	146	81	4	36
			13	External research activity incomes/tot. incomes	EXT_R_ACT_INC/TOT_INC	0	**142**	0	0	61	0	0	16	24
			14	Research expenses/tot Department's researchers	RES_EXP/DEP_RES	67	88	75	118	**254**	192	58	49	32
	Dimension		15	N° scientific fields	SCI_FI	130	**177**	55	143	75	68	48	116	89
			16	Tot. Department's researchers/ Tot. University researchers	DEP_RES/TOT_UNI_RES	87	**158**	71	152	71	81	68	152	59
			17	University financing to Department/ Tot. University financing	UNI_FIN_DEP/TOT_UNI_FIN	97	**168**	69	152	81	83	64	141	45
	Teaching		18	teaching burden (CFU)/tot Department's researchers	TEACH/DEP_RES	77	103	118	93	77	**130**	88	103	117
Departments as University bodies	Dept. tasks	young researchers training trough phd and research fellowships	19	N° Phd students/ tot Department's researchers	PHD/DEP_RES	150	133	**144**	83	131	121	41	58	40
			20	N° post-doc and research fellows/ tot Department's researchers	POSTDOC/DEP_RES	118	49	144	68	**180**	159	113	68	131
			21	N° research fellows/ N° Phd students	POSTDOC/PHD	79	37	100	82	138	131	**276**	116	331
		research policy sertting, coordination, promotion	22	N° departmental Commissions/ tot Department's researchers	COMM/DEP_RES	140	90	104	**123**	86	92	112	80	107
			23	N° meetings tot.departmental Commissions/ tot. departmental Commissions	MEET_COMM/COMM	**134**	88	109	**134**	77	79	123	81	102
			24	N° researchers belonging to departmental Commissions/ tot. Department's researchers	RES_COMM/DEP_RES	**144**	76	132	109	81	98	134	88	92
			25	N° scientific manifestations or conferences (days)/ tot. Department's researchers	CONF/DEP_RES	109	**160**	80	120	117	111	102	89	98
			26	Scientific manifestations and conferences expenses/ tot. Department's researchers	CONF_EXP/DEP_RES	129	**170**	77	115	134	108	106	78	120
	organs functioning		27	Presence rate to collegial organs meetings	PRES_RATE_MEET	**178**	89	102	120	78	87	125	76	109
			28	Procedures observance (qualitative judgement 1-10)	PROC	**165**	87	107	128	80	90	132	81	106
			29	N° collegial organs meetings occurred/ n° collegial organs meetings convened	MEET_OCC/MEET_CONV	**157**	92	103	110	84	89	128	82	108
			30	N° collegial organs meetings minuted/ n° collegial organs meetings occurred	MIN/MEET_CONV	**160**	90	108	111	84	91	140	79	113

* Indicators referring to PRIN are calculated as the mean of years 2004-2007.

**The highest values are in green, while the lowest in grey.

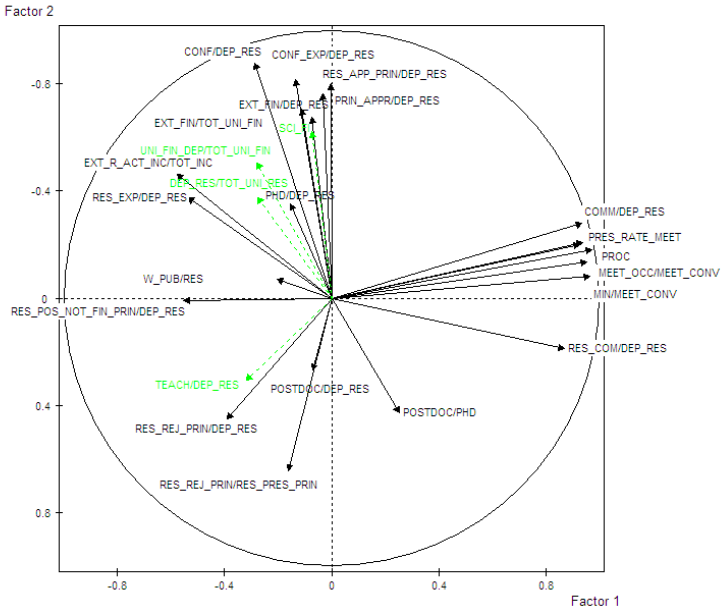
*** Considering the size difference among Departments, in order to make a comparison among them, all indicators have been transformed in index numbers, by setting University mean value equal to 100. In this way all the indicators have been transformed in undimensional data and could be read as deviation from Athenaeum mean value.

Tab. 2: Eigenvalues associated to factorial axes

Number	Eigenvalue	Percentage	Cumulated percentage
1	7,48	34,04	34,04
2	5,21	23,69	57,73
3	3,12	14,18	71,91
4	2,52	11,49	83,41
5	1,77	8,08	91,50
6	1,21	5,52	97,03
7	0,52	2,37	99,41
8	0,12	0,58	100,00

First two factorial axes have been used to build the factorial plan (fig. 1), showing original variables projection on new latent variables.

Fig 1: Factorial plan*



*Green variables are descriptive and, thus, not influencing factorial axes

Following PCA factorial plan interpretation rules[12], we can see that first factorial axe is characterized by a strong correlation with variables displayed in the right side of the plan, expressing good functioning of Department organs: for such reason, this axe could be interpreted as "Department organizational performance"; second factorial axe, instead, is characterized by a strong positive correlation with indicators expressing both Department capacity in financing its research activity through external channels and a high success in PRIN projects approval; on the contrary, the down side of the axe is characterized by a low success in PRIN. For these reasons, second axe could be interpreted as "Department research activity performance"

Table 3 shows characterizing variables and characterization degree for each axe[13].

Tab 3 - Variables characterizing factorial axes

Factor 1

Coordinates	Variable label	Mean	St. Dev.
-0,58	EXT_R_ACT_INC/TOT_INC	27,11	44,95
-0,55	RES_POS_NOT_FIN_PRIN/DEP_RES	100,71	28,39
-0,54	RES_EXP/DEP_RES	103,69	69,33
-0,39	RES_REJ_PRIN/DEP_RES	96,56	48,16
-0,29	CONF/DEP_RES	109,56	21,6
0,93	COMM/DEP_RES	103,78	18,16
0,94	PRES_RATE_MEET	107,11	29,98
0,95	MEET_OCC/MEET_CONV	105,89	22,75
0,96	MIN/MEET_CONV	108,44	25,36
0,97	PROC	108,44	26,89

Factor 2

Coordinates	Variable label	Mean	St. Dev.
-0,88	CONF/DEP_RES	109.56	21.60
-0,81	CONF_EXP/DEP_RES	115.22	26.98
-0,8	RES_APP_PRIN/DEP_RES	93.51	42.50
-0,76	PRIN_APPR/DEP_RES	96.85	54.87
-0,7	EXT_FIN/DEP_RES	88.32	70.51
0,19	RES_COMM/DEP_RES	106.00	23.62
0,27	POSTDOC/DEP_RES	114.42	42.42
0,43	POSTDOC/PHD	143.26	91.16
0,45	RES_REJ_PRIN/DEP_RES	96.56	48.16
0,64	RES_REJ_PRIN/RES_PRES_PRIN	98.41	47.63

Concerning correlation among variables, we can see that factorial plan shows a positive correlation among all variables related to Departments organs functioning; on the contrary, no correlation exists among the latters and research quality (measured both in terms of PRIN approved and in terms of external financing per researcher): this means that research performance is not influenced by the way in which Departments organs work.

A positive correlation exists, instead, between conferences (number and organizational expenses) and external financing attraction: this fact could be interpreted in a sense that conferences are the occasion for publicizing Departments activities and skills and for catching the interests of external actors.

A moderate positive correlation exists between the number of Phd students per researcher and the number of participants in PRIN projects presented and approved. This could mean that Departments with more Phd students per researcher are more propositional than those with a lower value of such ratio: this may happen because young researchers feel a higher discomfort than the permanent ones - due to a scarce availability of financial resources to carry on research activities - and this fact spurs them to strive to obtain PRIN funds.

If widening the analysis to descriptive variables, other interesting relations could be found.
First of all, Department size and researchers productivity are strictly correlated: this could be interpreted in a sense that good quality research is often the result of confrontation of different opinions and points of view, and such confrontation is much easier if it takes place within the same Department, where the occasions for discussing are more frequent than those occurring outside it: in this sense, bigger Departments seem to be advantaged than smaller ones.
Almost no correlation exists, instead, between teaching activity and research quality (measured in terms of external financing per researcher); this fact totally disproof the widespread opinion that a trade-off between teaching and research activity exists.

Tab 4 – Correlation matrix

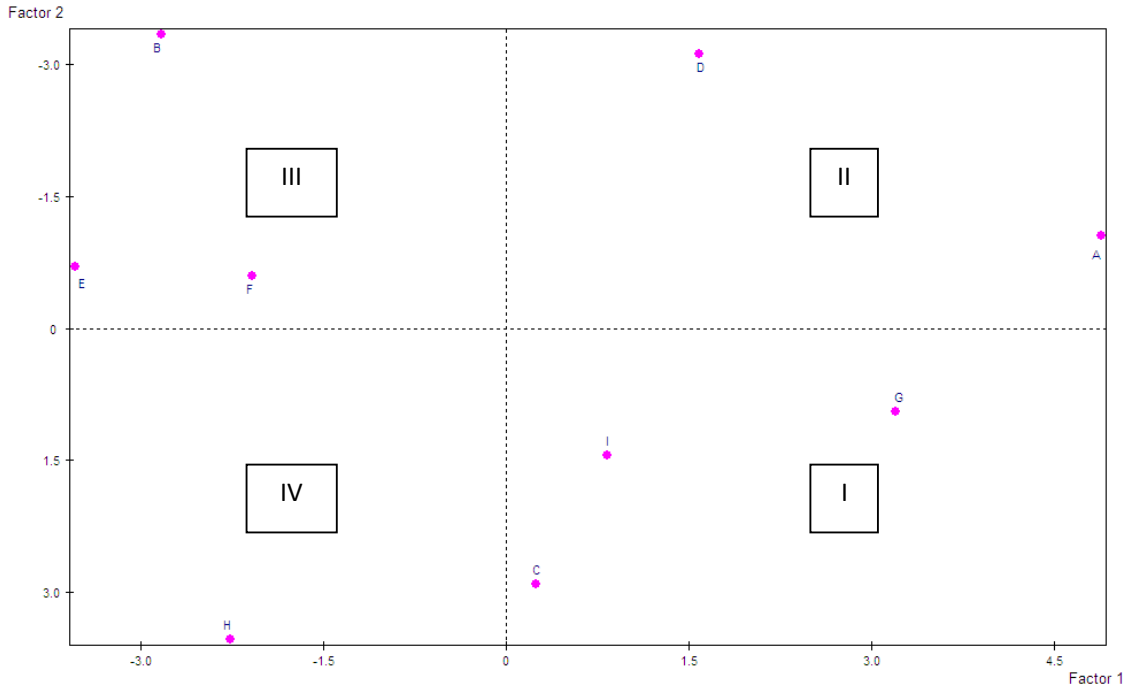
	RES_APP_PRIN/DEP_RES	RES_POS_NOT_FIN_PRIN/DEP_RES	RES_REJ_PRIN/DEP_RES	RES_REJ_PRIN/RES_PRES_PRIN	PRIN_APPR/DEP_RES	W_PUB/RES	EXT_FIN/DEP_RES	EXT_FIN/TOT_UNI_FIN	EXT_R_ACT_INC/TOT_INC	RES_EXP/DEP_RES	PHD/DEP_RES	POSTDOC/DEP_RES	POSTDOC/PHD	COMM/DEP_RES	MEET_COMM/COMM	RES_COMM/DEP_RES	CONF/DEP_RES	CONF_EXP/DEP_RES	PRES_RATE_MEET	PROC	MEET_OCC/MEET_CONV	MIN/MEET_CONV
RES_APP_PRIN/DEP_RES	1																					
RES_POS_NOT_FIN_PRIN/DEP_RES	-0.30	1																				
RES_REJ_PRIN/DEP_RES	-0.08	0.14	1																			
RES_REJ_PRIN/RES_PRES_PRIN	-0.30	-0.06	0.92	1																		
PRIN_APPR/DEP_RES	0.93	-0.07	-0.02	-0.28	1																	
W_PUB/RES	-0.22	0.67	-0.43	-0.56	-0.05	1																
EXT_FIN/DEP_RES	0.36	0.05	-0.50	-0.53	0.29	0.20	1															
EXT_FIN/TOT_UNI_FIN	0.32	0.01	-0.54	-0.54	0.25	0.20	0.99	1														
EXT_R_ACT_INC/TOT_INC	0.37	0.16	0.18	-0.02	0.29	-0.28	0.14	0.09	1													
RES_EXP/DEP_RES	0.38	0.53	0.06	-0.19	0.62	0.44	0.31	0.27	0.15	1												
PHD/DEP_RES	0.85886	0.71	0.19	-0.002	0.32	0.29	0.15	0.10	0.25	0.45	1											
POSTDOC/DEP_RES	-0.28	0.43	-0.03	-0.06	0.04	0.49	-0.29	-0.28	-0.32	0.55	0.25	1										
POSTDOC/PHD	-0.27	-0.49	-0.32	-0.12	-0.32	-0.07	-0.38	-0.33	-0.30	-0.27	-0.74	0.33	1									
COMM/DEP_RES	0.15	-0.33	-0.52	-0.40	0.15	-0.003	0.08	0.11	-0.45	-0.33	0.08	-0.05	0.03	1								
MEET_COMM/COMM	0.10	-0.48	-0.45	-0.25	0.03	-0.21	0.17	0.20	-0.46	-0.45	-0.08	-0.23	0.05	0.93	1							
RES_COMM/DEP_RES	-0.21	-0.29	-0.11	0.13	-0.12	-0.16	-0.12	-0.09	-0.67	-0.36	0.11	0.14	0.06	0.77	0.79	1						
CONF/DEP_RES	0.74	-0.02	-0.17	-0.38	0.65	-0.15	0.53	0.49	0.80	0.27	0.29	-0.40	-0.41	-0.07	-0.12	-0.46	1					
CONF_EXP/DEP_RES	0.69	-0.02	-0.25	-0.47	0.65	-0.11	0.27	0.23	0.77	0.23	0.29	-0.18	-0.19	0.06	-0.05	-0.38	0.91	1				
PRES_RATE_MEET	0.17	-0.38	-0.35	-0.23	0.18	-0.12	-0.10	-0.09	-0.39	-0.39	0.11	-0.05	0.03	0.96	0.86	0.79	-0.07	0.10	1			
PROC	0.15	-0.45	-0.37	-0.21	0.15	-0.15	-0.02	0.0003	-0.48	-0.39	0.03	-0.07	0.05	0.97	0.93	0.85	-0.13	0.007	0.98	1		
MEET_OCC/MEET_CONV	0.14	-0.46	-0.30	-0.15	0.15	-0.22	-0.16	-0.14	-0.39	-0.43	0.05	-0.02	0.12	0.93	0.85	0.82	-0.11	0.08	0.98	0.97	1	
MIN/MEET_CONV	0.08	-0.51	-0.32	-0.14	0.10	-0.24	-0.16	-0.14	-0.43	-0.43	-0.01	0.03	0.22	0.91	0.85	0.85	-0.16	0.03	0.96	0.96	0.99	1

In order to have a first, rough idea of Departments performance with reference to the new synthetic variables, a conjoint reading of both statistic unities and variables has been made by projecting the formers on factorial plan[14].
Figure 2 shows that the first two factorial axes divide the plan in four sectors with the following characteristics:

I) good organizational performance / low research performance
II) good organizational performance / good research performance
III) weak organizational performance / good research performance
IV) weak organizational performance / low research performance

According to statistic unities positioning on factorial plan, a first synthetic evaluation of Departments global performance could be easily made.

Fig 2 – Statistic unities projection on factorial plan



As last step, in order to understand the specific features of Departments having similar positioning on factorial plan, and to point out the main differences occurring among those located in different quadrants, Cluster Analysis has been performed.
Such analysis showed that our Departments could be grouped into three homogenous clusters, whose main strengths and weaknesses have been identified.
Department membership to each class is verified through the V-Test, a test allowing to determine, for each variable, the difference occurring between the cluster mean of the observed variable and the general mean for all statistic unities.
In detail, first class groups good performing Departments and is characterized by a good functioning of Departments organs and by a low participation of researchers in PRIN projects.

Tab 5 - Classes description

Tab 5a: CLASS 1 / 3

Characteristic variables	Cluster mean	Overall mean	Test-value
MEET_COMM/COMM	123,25	103,00	2,33
COMM/DEP_RES	120,50	103,77	2,32
PROC	132,75	108,44	2,28
MIN/MEET_CONV	131,00	108,44	2,241
MEET_OCC/MEET_CONV	125,75	105,88	2,20
PRES_RATE_MEET	133,00	107,11	2,18
...	...	...	...
RES_POS_NOT_FIN_PRIN/DEP_RES	79,06	100,71	-1,92

Second class groups worst Departments and is characterized by bed performances in PRIN projects and by low capacities both in organizing conventions and in attracting external financing.

Tab 5b:CLASS 2 / 3

Characteristic variables	Cluster mean	Overall mean	Test-value
RES_REJ_PRIN/RES_PRES_PRIN	162,09	98,40	2,02
RES_REJ_PRIN/DEP_RES	151,57	96,56	1,72
...	...	...	... ...
CONF/DEP_RES	84,5000	109,556	-1,75
PRIN_APPR/DEP_RES	23,67	96,85	-2,01
RES_APP_PRIN/DEP_RES	36,48	93,50	-2,02
CONF_EXP/DEP_RES	77,50	115,22	-2,11

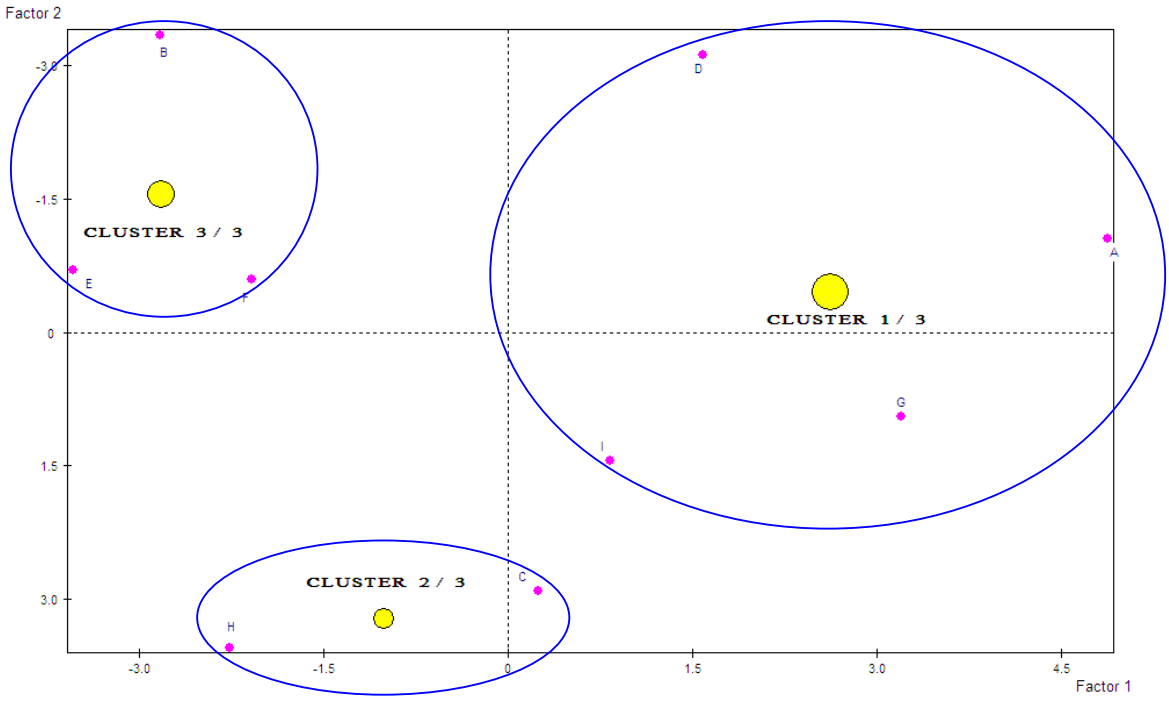
Third class groups intermediate performing Departments and is characterized by pro-capite research expenses higher than the University average value and by organs functioning worse than other clusters.

Tab 5c: CLASS 3 / 3

Characteristic variables	Cluster mean	Overall mean	Test-value
RES_EXP/DEP_RES	177,93	103,69	2,14
CONF/DEP_RES	129,33	109,55	1,83
EXT_R_ACT_INC/TOT_INC	67,82	27,10	1,81
...	...	...	...
MEET_COMM/COMM	81,33	103,00	-1,97

Fig 3 shows factorial plan displaying both statistic unities and clusters barycentre; this figure can help better understanding each group positioning with reference to factorial axes.

Fig 3: Cluster representation on factorial axes



Conclusions

Apart from specific results related to the case study, the main advantage of the proposed methodology is that it allows a global and synthetic evaluation of Departments without wasting any available information, and taking into consideration all the aspects considered as relevant.

The flexibility of the model allows to tailor the evaluation in relation to Athenaeum necessities and to the width of the analyzed context; the output of the model, then, could be used for different purposes, according to the reasons underlying Departments assessment.

In an Athenaeum based evaluation, for example, the choice of variables to analyze can be led by the necessity of identifying the "hidden" determinants of departmental research perfomance and the output of the model can be used to help University governmental organs devising the best policies to improve athenaeum research quality.

The model, though, could be easily applied to wider contexts than the University one, for example to make comparisons among all the University Departments of a Region, or to perform a nation-wide analysis of Departments with a specific scientific skilfulness; in this case, the choice of variables to analyze could be done basing on available data at any level of aggregation (e.g. MIUR database could be used for nation based analyses) and the output of the model could help defining national, regional or sectoral resource allocation among research units in relation to their positioning in the identified clusters.

In addition, if indicators time series are available, the model allows, at any level, to perform a "dynamic" cluster, in order to visualize research performance of Departments over time; such results could be used, then, to set up a financial incentive system, linked to the shift from a “worse” cluster to a “better” one.

Notes

1. The paper has been presented to the 15th Spring Meeting of Young Economists (SMYE 2010), 15-17 April 2010, Luxembourg City, Luxembourg.
2. The use of quantitative parameters by peer evaluators is important in order to avoid that absolutely subjective elements may cause both fully arbitrary evaluations and huge discrepancies among judgments expressed by different evaluators.
3. According to the Italian University Council (CUN), the adoption of bibliometric parameters to evaluate individuals is a "*conceptual statistic mistake*". For a demonstration, see MIUR (2010).
4. Probably, in a medium-long term period (5-10 years), the evolution of communications technologies, even in the scientific field, related to internet and to electronic publishing, and the development of social network, will bring the diffusion of new evaluative systems, such as the one launched by Nature, called "open control", according to which manuscripts are "exposed" on internet for a short time period and each reader can make comments before editorial board decides to publish it, on the base of readers opinions.
5. CUN demonstrates that when the number of evaluated subjects, considered as an aggregate, grows, statistic laws ensures that the reliability of quantitative analyses becomes higher (MIUR, 2010) than that performed on single researchers.
6. Italian University evaluation system has been set out by Decree 204 of 5 June 1998, that recognizes academic financial, managerial, didactic and scientific autonomy by the State Administration. In a context where Universities are autonomous bodies but mainly financed by public funds, Italian law has provided the creation of internal evaluation panels, namely the "Nuclei di Valutazione di Ateneo (NVA)", whose aim is evaluating University administrative, teaching and research activities, and reporting such information to CNVSU (National Committee for the Evaluation of the University System), a specialized body of Italian Ministry for Education, University and Research, acting as advisor to the Ministry itself.
7. The attempt to make a multidimensional analysis of research performance has already been made by a research group for CNVSU (Bini et al., 2008), but it is focused only at a University and Faculty level; no deeper disaggregation at Department level has been made.
8. Data extracted by NVA 2007 relation.
9. Data obtained by interviews to Departments Administrative Directors.
10. PRIN are research projects annually co-financed by the Italian Ministry of Research.
11. All variables in table 1 have been considered as active, except those related to the category "Dimension" and "Teaching" (indicators 15 - 18) .
12. PCA factorial plan interpretation rules are the following:
    - Correlation between variables is shown by the angle between the vectors representing them: the narrower it is, the highest is the correlation between variables;
    - Principal components are latent variables; they can be compared with original variables in terms of angle. Such comparison allows to understand the meaning of latent variables;
    - Comparison between variables is correct only if they are well represented on factorial plan: the further is the vector from the axes origin (the highest is the square of the coordinate of the point on the axe), the best represented is the variable (relative contribution);
    - The higher is the variable coordinate on the axe, the highest is the influence of the former on the latter (absolute contribution);
    - Supplementary variables do not influence axes determination (their projection on factorial plan do not alter its interpretation except for absolute contribution).
13. Variable describing publication score is not well represented -as vector representing such variable is very small-, and, with its slope of 45°both with the first and the second axe, it also plays a small role in determining them. On the contrary, such variable is highly positively characterizing the third axe, whose role, considering its low explanatory significance (only 14% of phenomenon variability), is not considered in this analysis. Other important variables, too, such as those describing Department ability in young researchers training (through Phd and research scholarships) have a very low influence in determining the first two factorial axes; on the contrary , they highly characterize the fourth axe, whose explanatory capacity is very low (only 11% of phenomenon variability) and, thus, not considered in this analysis.
14. Principal Component Analysis allows to project statistic unities on factorial plan: the latter, even if not identical to the one showing the variables, is perfectly stackable to it, so as to allow a conjoint reading of both statistic unities and variables.

References

Adler R., Ewing J., Taylor P. 2008. *Citation Statistics*, Report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), 6/11/2008.

CSS - Consiglio Italiano per le Scienze Sociali 2006. *La valutazione della ricerca*. *Libro bianco*. Venezia, Italy: Marsilio,.

Cugini A., Michelon G. 2009. La misurazione della performance nei Dipartimenti universitari. *Economia & Management*, n° 4 (July/August).

Dehon C, McCathie A, Verardi V. 2009. Uncovering Excellence in Academic Rankings. A case study on the Shanghai Ranking. *Working Paper Mimeo*.

Dubini P. 2005. Misurare la competitività dei sistemi Paese: il modello del World Economic Forum. *Economia e Management*, n°6.

EUA - European University Association 2009. *Institutional Evaluation Programme: Guidelines for participating institutions*. Belgium, Brussels: EUA Editions.

European Commission, Enterprise Directorate-General 2009, European Innovation Scoreboard (EIS). *PRO INNO Europe Paper*, n° 15, Bruxelles, Belgium.

Figà Talamanca A. 2009, *E' lecito usare l'Impact Factor per valutare la qualità della ricerca?,* relation presented to the conference *The problem of evaluation in Social Sciences*, Bergamo University, 25 November 2009, Bergamo, Italy.

Frey B.S., Rost K. 2008. Do Rankings Reflect Research Quality?, *WP* n. 390, Institute for Empirical Research in Economics, University of Zurich.

Gallo M. 2007. The scaling problems in service quality evaluation. *Metodološki Zvezki*, vol. 2, 165-176.

Gherghi M., Lauro C. 2004. *Appunti di Analisi dei Dati Multidimensionali. Metodologia ed esempi*. Naples, Italy: RCE Edizioni,.

Guan, J.; Wang, J. 2004. Evaluation and interpretation of knowledge production efficiency. *Scientometric*, n° 59, 131-155.

Hicks D. 2007. *The role of quantitative research evaluation in developing international research excellence*, relation presented to the conference organized by the Italian Society of Economists (SIE) on *Research Evaluation: experiences and methods*, Ancona University, 2-3 April 2007, Ancona, Italy.

Hirsch E.J. 2005, An index to quantify an individual's scientific research output, *PNAS*, vol. 102, n° 46, 16569-16572.

King D.A. 2004, The scientific impact of nations, *Nature*, 430, 15 July 2004.

Lippi M., Peracchi F. 2007. *Il primo esercizio italiano di valutazione della ricerca: una prima valutazione*, relation presented to the conference organized by the Italian Society of Economists (SIE) on *Research Evaluation: experiences and methods*, Ancona University, 2-3 April 2007, Ancona, Italy.

Liu N.C., Cheng Y. 2005. Academic Ranking of World Universities - Methodologies and Problems, *Higher Education in Europe*, 30, n°2.

Macerata University 2009. *Se questa vi sembra una valutazione*, Macerata, Italy: EUM - Edizioni Università di Macerata.

Marcuzzo M.C., Zacchia G. 2007, L'econlit e gli strumenti per la valutazione della ricerca economica in Italia. *Rivista italiana per gli economisti*, n°. 2., 279-310.

MIUR 1998. *Indicazioni per la preparazione delle relazioni dei Nuclei di valutazione interna e insieme minimo di indicatori*, doc 11/98.

MIUR 2009. D.M. 23 settembre 2009 prot. 45/2009, *Decreto criteri di ripartizione del Fondo di Finanziamento Ordinario (FFO) delle Università per l'anno 2009*.

MIUR, CUN 2010. *Valutazione della ricerca*, Prot. 612, 30/3/2010.

Moed, H.F. 2008. UK Research Assessment Exercises: Informed judgments on research quality or quantity?. *Scientometrics*, n° 74, 141-149.

Narin, F.; Hamilton, K.S. 1996. Bibliometric performance measures. *Scientometrics*, n° 36, 293-310.

Neary J.P., Mirrlees J.A., Tirole J 2003. Evaluating Economics Research in Europe: An Introduction. *Journal of the European Economic Association*, Vol 1 , n° 6, p. 1247.

OECD 1997. *The evaluation of scientific research: selected experiences*, Paris.

OECD 2009. *Science Technology and Industry Scoreboard 2009*, Paris

RAE - Research Assessment Exercise 2006. *Assessment panel's criteria and working methods*, doc. RAE 01/06, in http://www.rae.ac.uk.

Remler D. K., Pema E. 2009. Why do Institutions of Higher Education Reward Research While Selling Education?. *NBER Working Paper* No. 14974.

Rizzuto, C. 2003. Finanziamenti, progetti, centri di eccellenza, FAR, FISR, distretti tecnologici, in *La ricerca universitaria: esperienze, modelli, proposte*, eds. Casciotti C. and Mosconi G., p. 96-115, Italy,Rome: CRUI editions.

Seglen P. O. 1997. Why the impact factor of journals should not be used for evaluating research, *BMJ*, 314-497.

Van Leeuwen T. 2006. Descriptive Versus Evaluative Bibliometrics. Monitoring and Assessing of National R&D Systems, in *Handbook of Quantitative Science and Technology Research*, part 3, 373-388, Netherlands: Springer.

Van Raan A. 2005. *Challenges in Ranking of Universities*, Invited paper for the *First International Conference on World Class Universities*, Shanghai Jaio Tong University, Shanghai, June 16-18, 2005

Vinke P. 2009. University Rankings, in *Ranking Universities,* eds. Dehon C., Jacobs D., Vermandele C., pp.11-25, Belgium, Bruxelles: Editions de l'Université de Bruxelles.

World Economic Forum 2009, *The Global Competitiveness Report 2009-2010*.