Evaluation of Academic Activity through Synthetic Indicators

Roberto Gismondi, Massimo Alfonso Russo

Abstract The new Italian regulation for the distribution of the Ordinary Financing "public" Fund (*FFO*) is included in the framework of the reform of the University System financing politics. A part of the *FFO* is bound by a specific merit rating. The performance evaluation consists in a periodical monitoring of results achieved by each University, compared with the 3 years operating programme, which contains the main goals, in terms of improvement of the public service quality. This task is based on a *Quality aggregate index* (*Qai*), which is a weighted synthesis of twenty-one parameters, measured on a scale of ratios and classified into five homogeneous macro-ares. However, the actual proposal for the *Qai* quantification may generate misleading results with respect to the original purpose. In this paper, both a preliminary transformation of parameters – aimed at reducing asymmetry and variability of their variation ranges – and the use of ratios between performances referred to two following periods instead of differences are proposed. Moreover, simple statistical techniques for weighting groups of variables are introduced. A final empirical attempt based on 2008 data compares the actual and the alternative methodologies.

Key words: Dimensionality reduction, Index numbers, Multivariate analysis, Ranking

1 Basic rules for the evaluation of academic activity in Italy

The recent Italian regulation (L. 43/2005) for the distribution of the Ordinary Financing "public" Fund (*FFO*) is included in the framework of the reform of the University System financing politics and pursues the public service improvement goal. This

---

Roberto Gismondi,
ISTAT, Via Tuscolana 1788, 00173, Roma; email: gismondi@istat.it
Massimo Alfonso Russo
Università di Foggia, Via Papa Giov. Paolo II 1, 71100, Foggia; email: m.russo@unifg.it

approach is strengthened by the recent law L. 1/2009, which aims at bridging the present internal system of resource distribution for higher education and scientific research to a "qualitative" logic of productivity, already adopted in other industrialised countries. The need to identify a comparable evaluation system among the European countries is well known. However, a common and agreed evaluation procedure is not available yet. On this field, the Italian attempt is one of the first proposal characterised by both a methodological background and a relative simplicity. At present, only the assignment of one third of *FFO* is related to a specific evaluation (*ex-post*) of each single Italian University. Such an evaluation is inspired by the couple of issue goals-results. Each University determines (*ex-ante*) its goals, as regards the 3 years operating programme, while the results are quantified (*ex-post*) by a specific Quality aggregate index (*Qai*). The Italian Government (Ministry of University) has fixed both parameters and criteria in order to quantify the *Qai*. The index is based on 21 parameters, measured as scales of ratios and classified in 5 homogeneous macro-areas[2] (groups).

Table 1: List of the computable parameters by group (Gr.)

Gr.	**Parameter**	**Gr.**	**Parameter**
A3	Average full tenured teachers by degree course	D1	% of students who took part to international mobility programs
B2	Average grants by doctorate courses	D2	% of foreign students enrolled to teacher-training courses
B3	% of doctorate grants not financed by University	D3	% of foreign students enrolled to doctorate courses
B4	Average financial funds by full tenured professor or researcher	D4	Relative weight of incomes derived from contracts with external bodies
B5	% of research funds not financed by University	E1	% incidence of personnel costs on the whole balance incomes
C1	Students enrolled in the second year divided by students enrolled in the first year (the year before)	E2	Relative incidence of vacancies satisfied with teachers previously not belonging to the Athenaeum
C2	% of enrolled students who attended stages external to their University	E3	% weight of "organic points" used for courses with a ratio students/teachers larger than the national median
		E4	% of "organic points" used for enrolling new researchers

For a more detailed description of indicators, see MIUR (2008).

The recourse to *performance* analyses founded on rankings which synthesise the content of a multivariate database has becoming a quite common standard for many social and economic fields of applied research, especially as regards efficiency evaluations. The case of academic activity can be linked to not new problem concerning

[2] The 5 macro-areas are: A) Degree Curricula (3 parameters); B) Scientific Research (5); C) Services for students (5); D) Internationalization (4); E) Academic and Administrative Staff (4). The following table 1 reports the only parameters which can be effectively measured at the actual stage, broken down according to the following 5 groups. The description reported in the table 1 is quite synthetic, more details are available in MIUR (2008).

the choice of ranking criteria for multivariate data (Barnett, 1976; Leibowitz and Hyman, 1999; Giudici and Avrini, 2002; D'Esposito and Ragozini, 2004).

In this context, in section 2 we resume the basic methodology and propose an alternative procedure based on transformations of the original parameters; in particular, a class of not linear transformations is introduced. In section 3, weights for the 5 groups of variables are assigned through 2 alternative weighting systems, both based on the logic to underweight groups characterised by a large internal correlation level. Section 4 resumes the main results of the empirical application to real data concerning the performance evaluation referred to 2008, with some perspective conclusions.

2 Performance evaluation

The basic methodology used for the evaluation of academic activity in Italy is resumed hereafter. Some simplifications and adaptations have been introduced, without a significant distortion of the underlying logic which the complete algorithm is based on (MIUR, 2008). We suppose to measure on n units k variables (parameters), all positively correlated with a positive performance. Variables are classified into G logical groups, where the *g-th* group includes k_g variables. If $x_{g_h i(t)}$ is the value of the *h-th* variable belonging to the *g-th* group on the *i-th* unit at a time t, and w_h (h=1,2,…, k) represents a series of given group weights summing up to one, the scores assigned to the *i-th* unit as regards the group g and as a whole are given by, respectively[3]:

$$S_{gi}^0 = \sum_{h=1}^{k_g} [x_{g_h i(t)} - x_{g_h i(t-1)}] / k_g \qquad\qquad S_i^0 = \sum_{g=1}^{G} S_{gi}^0 w_g \;. \tag{1}$$

If $\bar{x}_{g_h(t)} = \sum_{i=1}^{n} x_{g_h i(t)} / n$, the correspondent indicators for the whole system are given by:

$$S_g^0 = \sum_{h=1}^{k_g} [\bar{x}_{g_h(t)} - \bar{x}_{g_h(t-1)}] / k_g \qquad\qquad S^0 = \sum_{g=1}^{G} S_g^0 w_g \;. \tag{2}$$

Finally, the index which expresses the level of performance with respect to the whole system (*Qai*: *Quality aggregate index*) is:

$$Qai_i^0 = (S_i^0 / S^0) \cdot 100 \tag{3}$$

where in the case under study i=1,2,…,56, G=5, k=15, g_1=1, g_2=4, g_3=2, g_4=4, g_5=4 and w_h=0.200 for each h[4]. According to (3), a unit has a performance better than the whole system average if *Qai*>100.

A general alternative procedure can be based on: i) the use of ratios between variables z which are functions of the original correspondent variables x; ii) the recourse to a geometric mean in order to obtain each group score. Given the previous

3 As regards the following formula, in the real context t=2008 and (t-1) is the average 2004-2006.
4 More precisely, each university can apply weights such that $0.10 \le w_h \le 0.30$, under the condition that the sum of weights is equal to one.

definitions, the scores S assigned to the *i-th* unit as regards the group g and as a whole are given by:

$$S_{gi} = \sqrt[k_g]{\prod_{h=1}^{k_g} [z_{g_h i(t)} / z_{g_h i(t-1)}]^{1/k_g}} \qquad S_i = \sum_{g=1}^{G} S_{gi} w_g \qquad \text{for each } i \in g \text{ , } g=1,\ldots,G, \quad (4)$$

while the analogous indicators for the whole system are given by:

$$S_g = \sqrt[k_g]{\prod_{h=1}^{k_g} [z_{g_h(t)} / z_{g_h(t-1)}]^{1/k_g}} \qquad S = \sum_{g=1}^{G} S_g w_g \text{ .} \quad (5)$$

As regards the final *Qai*, that remains formally similar to (3).

The geometric mean is more robust than the arithmetic mean in presence of potentially outlier observations, even though it can not be used in presence of null or negative values. As regards the use of ratios between observations referred to time t and (t-1) instead of differences, we note how indexes let the possibility to deal with pure numbers not dependent from the original measurement units.

Two simple options for the new transformed variable z are given by the two following positions (groups labels are omitted):

$$z_i = x_i \quad (6)$$

$$z_i = (x_i - m_x)(M_x - m_x)^{-1} \text{ .} \quad (7)$$

Of course, the use of (6) is equivalent to calculate indexes between times t and (t-1) using the original variables. The main risk concerned with ratios between original variables is that they can lead to serious under-evaluations of variables characterised by small variation ranges, mainly when m_x (the lowest x level) and M_x (the largest x level) are quite different among variables.

The *normalisation* (7) should fit better when variables are characterised by small variation ranges and/or when variation ranges are quite different among variables (Delvecchio, 1995, pp.133-134; Spada and Russo, 2006). As each linear transformation, that allows to keep the same ratio between observations with a different origin (*proportionality*). In addition, transformed variables maintain the same asymmetry level and the same linear correlation, which characterises the original x-variables. However, also normalisation results could be affected by potential outlier values for m_x and/or M_x (Aiello and Attanasio, 2004). Moreover, effects of normalisation is less immediately clear when original variables are characterised by different mean, minimum and maximum (Gismondi and Russo, 2007). As a consequence, a further proposal is based on the following not linear transformation[5]:

$$\frac{z_i - m_z}{M_z - z_i} = \frac{p_x (x_i - m_x)}{(1 - p_x)(M_x - x_i)} \quad (8a)$$

where p_x is a *smoothing* coefficient for the difference between x and its minimum, ranging from 0 to 1. The idea underlying (8a) is that the highest is p_x, the lowest will be the relative weight assigned to (M_x-x). When p_x=0,5, after the transformation in the new

[5] Not linear transformations do not necessarily satisfy the *proportionality* property.

z-scale the ratio between the distance between the transformed z and its minimum and the distance between its maximum and the transformed z must be equal to the same ratio measured on the original variable x. Moreover, when M_z=1 and m_z=0, we get:

$$z_i = \frac{p_x(x_i - m_x)}{p_x(x_i - m_x) + (1 - p_x)(M_x - x_i)} . \qquad (8b)$$

If the new variable z must range within [0,1] (p_x=0.5), z is given by the usual rule (7).

Given x, M_x and m_x, z will be as much higher as p_x will be high, meaning a larger weight assigned to the difference between x and its minimum rather than the difference between the maximum and x. That can be useful when the minimum is more representative of the whole x-distribution rather than the maximum, as it should happen when x-distribution is affected by a strong positive *asymmetry*, as in many empirical contexts. The evaluation of distance of minimum and maximum from the whole distribution is a particular case of the wider problem concerned with distance of one point from a cluster of points (Leti, 1979, p.68).

Broadly speaking, the choice of p_x depends on how much it is more important to remark distance from the lowest value instead of the gap respect to the highest one. In case of perfect symmetry, one should put p_x=0.5. Various criteria can be applied (Gismondi and Russo, 2007); herein we consider a single formula on the basis of which p_x will vary in the interval [0,1]. If $q_{(0.50)}$ is the median, we have:

$$p_{x(1)} = 0.5\left(1 + \frac{\mu_x - q_{x(0.50)}}{\sigma_x}\right) \qquad (9)$$

Formula (9) is strictly connected with the *Pearson's second asymmetry coefficient*, given by $3(\mu_x - q_{x(0,50)})/\sigma_x$. If a distribution is symmetric (for instance, in presence of uniform or normal distributions), the previous formula turns out to be equal to 0.5. The coefficient p_x will be higher than 0.5 when the mean is higher than the median, as it happens in the case of positive asymmetry[6]. However, a well known limit of these asymmetry indicators is that they can be zero even when asymmetry is not null.

3 Weighting groups of indicators

The weights w_g assigned to each group of variables may be determined on the basis of a more objective statistical criterion. In particular, the basic idea is to apply larger weights to groups characterised by a lower average level of linear correlation between couples of variables belonging to that group. This strategy reduces the weight of groups affected by a too large internal level of linear correlation between couples of variables.

A first method is based on the calculation of the complement to one of the average squared correlation (or absolute correlation) between each of the couples of different variables inside each group, normalised through the division by the sum of the previous complements from 1 to G. In symbols, we have:

[6] The link between asymmetry and the distribution form is not new, having been formalised by Pearson according to the "criterion K" (Delvecchio, 1995, pp.256-257).

$$w_{g(*)} = \left[1 - \sum_{h=1}^{k_g} \sum_{\substack{l=1 \\ l \neq h}}^{k_g} r_{hl}^2 \Bigg/ \binom{k_g}{2} \right] \Bigg/ \left[\sum_{g=1}^{G} \left(1 - \sum_{h=1}^{k_g} \sum_{\substack{l=1 \\ l \neq h}}^{k_g} r_{hl}^2 \Bigg/ \binom{k_g}{2} \right) \right] \quad (10)$$

so that groups including some variables very correlated each other will be penalised.

Another method is based on the principal component analysis (*PCA*) and is founded on eigenvalues (Jolliffe, 2002). On the basis of *PCA*, inside each group *g* one identifies the k_g eigenvalues λ_{g_h}, for $h=1,2,\ldots,k_g$ and $g=1,2,\ldots,G$. A synthetic index expressing the average level of internal correlation among variables included in this sub-group is:

$$\gamma_g = \sum_{h=1}^{k_g} \left| \lambda_{g_h} - 1 \right| \Big/ k_g \,. \quad (11)$$

This index will reach its minimum *min* – equal to 0 – when all variables in group *g* are perfectly uncorrelated (that is, when $\lambda_{g_h} = 1$ for each *h*), while it will reach its maximum *max* – equal to $2(k_g–1)/k_g$ – when one eigenvalue is equal to k_g and all the remaining $(k_g–1)$ are equal to 0. As a consequence, the draft index $(1-\gamma_g)$, expressing the degree of linear independence between the couples of variables belonging to the *g-th* group, will vary between $–(k_g–2)/k_g$ and 1. A final *relative index of linear independence* for the *g-th* group, varying in the range [0,1], can be obtained through normalisation:

$$\gamma_g^{'} = \frac{(1-\gamma_g) - min}{max - min} = \frac{2(k_g - 1) - k_g \gamma_g}{2(k_g - 1)} \,. \quad (12)$$

A final weight – varying in the range [0,1] and such that the sum of weights over *g* will be equal to one – for the *g-th* group will then given by:

$$w_{g(**)} = \gamma_g^{'} \Big/ \sum_{g=1}^{G} \gamma_g^{'} \,. \quad (13)$$

If a group includes only one variable (as for g_1), its weight can be put equal to the original one (0.200) and the other new weights $w_{g(1)}$ can be rescaled so that the final sum of all weights is one.

The application to the context under study showed that linear correlations inside groups are quite low: levels larger than 0.30 characterise only b1 and b3 (0.378), c3 and c4 (0,549), d1 and d3 (0.459), d1 and d4 (0.411). As a consequence, the new weights derived from formulas (10) and (13) are quite similar to the original ones (0,200): for the 5 groups they are, respectively, 0.200, 0.202, 0.197, 0.202, 0.199 according to (10) and 0.200, 0.197, 0.209, 0.199, 0.195 according to (13). These results confirm that the basic, subjective rule which assigns the same weight to each group is fully justified from an objective statistical point of view.

4 Main empirical results and conclusions

An empirical attempt has been carried out on the basis of data referred to 2008 (*t*), compared with the base period 2004-2006 (*t*-1). The methods compared are: *Qai*0: that is the basic method based on formulas (1) and (2); *Qai*1: the method based on ratios between the original variables (formulas (4), (5) and (6)); *Qai*2: the method based on ratios between normalised variables (formulas (4), (5) and (7)); *Qai*3: the method based on a not linear normalisation (formulas (4), (5), (8b) and (9)). Table 2 reports rankings obtained with each method. The ranking produced according to *Qai*0 has two main peculiarities: 1) it presents several cases for which different universities have the same ranking position; 2) it is quite different with respect to the other methods' rankings.

Table 2: Rankings derived from 4 methodologies (each group weight is 0.200)

University	**Ranking**				***University***	**Ranking**			
	***Qai*0**	***Qai*1**	***Qai*2**	***Qai*3**		***Qai*0**	***Qai*1**	***Qai*2**	***Qai*3**
1	1	9	1	1	**31**	17	54	58	58
2	2	40	9	10	**32**	17	57	42	44
3	3	6	17	16	**33**	17	58	44	45
4	3	7	27	27	**34**	34	5	4	2
5	3	13	3	3	**35**	34	8	15	14
6	3	21	24	20	**36**	34	14	21	15
7	3	22	22	24	**37**	34	15	5	5
8	3	27	31	28	**38**	34	26	35	36
9	3	32	2	4	**39**	34	29	54	50
10	10	3	33	32	**40**	34	31	11	13
11	10	4	20	21	**41**	34	35	38	42
12	10	10	8	7	**42**	34	36	45	41
13	10	17	14	17	**43**	34	45	50	53
14	10	20	10	9	**44**	34	49	56	56
15	10	25	13	11	**45**	34	50	29	35
16	10	34	36	34	**46**	34	52	59	59
17	17	2	28	25	**47**	34	56	53	54
18	17	11	39	38	**48**	34	59	43	52
19	17	18	46	46	**49**	49	1	18	19
20	17	19	26	26	**50**	49	12	37	37
21	17	23	6	6	**51**	49	24	23	23
22	17	28	40	39	**52**	49	38	41	47
23	17	30	7	8	**53**	49	41	19	22
24	17	33	16	18	**54**	49	48	47	49
25	17	37	32	31	**55**	55	16	52	40
26	17	42	51	51	**56**	55	39	48	43
27	17	43	34	33	**57**	55	51	25	29
28	17	44	30	30	**58**	55	53	57	57
29	17	46	12	12	**59**	59	55	55	55
30	17	47	49	48					

*Qai*0: basic method based on (1), (2); *Qai*1: method based on (4), (5), (6); *Qai*2: method based on (4), (5), (7); *Qai*3: method based on (4), (5), (8b), (9).

The basic indicator *Qai*0 is quite steady and it does not discriminate units under study properly: that is also confirmed by its very small range of variation (from 0.15 to 0.22) and a coefficient of variation lower than *Qai*1. On the other hand, the ranking derived by *Qai*1 is quite different from *Qai*0, as well as rankings got with *Qai*2 and *Qai*3 are similar each other, but different from *Qai*0 and *Qai*1.

The previous evidences are confirmed by the linear co-graduation coefficients reported in table 3 (figure in bold are coefficients larger than 0.5). Rankings referred to *Qai*2 and *Qai*3 have the largest co-graduation (0.984); they are more similar to *Qai*1 (both linear co-graduations are larger than 0.5) rather than to *Qai*0, that is itself more correlated with *Qai*2 and *Qai*3 than with *Qai*1. As a matter of fact, the use of the alternative groups' weighting systems (10) and (13) does not lead to significant changes in the correspondent rankings concerning *Qai*1, *Qai*2 and *Qai*3 (since new weights are still near 0.200, as remarked at the end of section 3).

As a conclusion, results are more sensitive with respect to the parameters transformation than to the group weight selection. On the whole, the index *Qai*2 based on normalisation may be preferred, because of its flexibility and statistical coherence.

Table 3: Co-graduation coefficients between couples of rankings compared

	Qai0	**Qai1**	**Qai2**	**Qai3**	**Qai1***	**Qai2***	**Qai3***	**Qai1****	**Qai2****	**Qai3****
Qai0	1.000	0.300	0.428	0.434	0.298	0.428	0.436	0.300	0.428	0.434
Qai1		1.000	**0.555**	**0.621**	1.000	**0.554**	**0.622**	1.000	**0.555**	**0.621**
Qai2			1.000	**0.984**	**0.557**	1.000	**0.985**	**0.555**	1.000	**0.984**
Qai3				1.000	**0.622**	**0.984**	1.000	**0.621**	**0.984**	1.000
Qai1*					1.000	**0.556**	**0.623**	1.000	**0.557**	**0.622**
Qai2*						1.000	**0.984**	**0.554**	1.000	**0.984**
Qai3*							1.000	**0.622**	**0.985**	1.000
Qai1**								1.000	**0.555**	**0.621**
Qai2**									1.000	**0.984**
Qai3**										1.000

Legenda: *: group weights (10); **: group weights (13). Bold: co-graduations > 0.5.

References

1. Aiello, P., Attanasio, M.: How to Transform a Batch of Single Indicators to Make Up a Unique One? Atti della XLII riunione scientifica della Società Italiana di Statistica (Sessioni plenarie e specializzate), pp.327-338 (2004).
2. Barnett, V.: Ordering of Multivariate Data, Journal of the Royal Statistical Society, Ser.A, 139, pp.318-354 (1976).
3. Delvecchio, F.: Scale di misura e indicatori sociali, Cacucci Editore (1995).
4. Gismondi, R., Russo, M.A.: Methodological Proposals for a Qualitative Evaluation of Italian Durum Wheat Varieties, Università degli studi di Foggia, DSEMS, quaderno 23 (2007).
5. Giudici, P., Avrini, P.: Modelli statistici per la costruzione di indicatori della qualità della vita: aspetti metodologici, Rivista di statistica ufficiale, 1, pp.61-80 (2002).
6. Grilli, L., Russo, M.A.: Decision Making in Financial Markets through Multivariate Ordering Procedure, in Mathematical and Statistical Methods in Insurance and Finance, (C. Perna e M.Sibillo Editors), Springer, pp.139-147 (2007).
7. Jolliffe, I.T.: Principal Components Analysis - Second Edition, Springer (2002).
8. Leibowitz, S.G., Hyman, J.B.: Use of Scale Invariance in Evaluating Judgement Indicators, Environmental Monitoring and Assessment, 58, pp.283-303 (1999).
9. Leti, G.: Distanze ed indici statistici, La Goliardica (1979).
10. MIUR: Programmazione e valutazione delle Università per il triennio 2007-2009. Primo quadro informativo degli indicatori, http://www.miur.it/UserFiles/2482.pdf (2008).