Journal of Informetrics 2 (2008) 183–194

Contents lists available at ScienceDirect

Journal of Informetrics

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

Measuring scientific performance of public research units for strategic change

Mario Coccia*

Italian National Research Council and Max Planck Institute of Economics, Germany

ARTICLE INFO

Article history:
Received 15 October 2007
Received in revised form 2 April 2008
Accepted 3 April 2008

Keywords:
Scientometrics
Public research institutes
Scientific performance

ABSTRACT

This research presents a new metrics to measure and assess the scientific performance of public research institutes, which improves models based on standard multivariate techniques. These models called Research Lab Evaluation (RELEV) adjusted are successfully applied to Italian public research institutes, operating in five scientific fields. In addition, the paper presents a comparison between this method and the Data Envelopment Analysis to show some analogies and differences in the results.

1. Introduction

[... in the calculation of research units performance not only the output is important, the input matters too ...]

Recent changes in the international economic environment, and the growing scientific basis for contemporary technologies due to the information and communication revolution, will make public research institutions even more important for the growth of modern economies. Scientific institutions are complex and dynamic organizations, whose main purpose is producing and disseminating scientific research within national economic systems to generate inventions and innovations, which are more and more important to the competitiveness of countries. The measurement and evaluation of scientific performances of public research institutes are critical activities to improve their organizational behaviour and strategic change.[1] In fact, in Italy (Coccia, 2001, 2004, 2005; Silvani, Sirilli, & Tuzi, 2005), Spain (Cruz-Castro & Sanz-Menéndez, 2006), Germany (Krull, 1995), France (Larédo, 2002; Vilkas, 2004), the United Kingdom (Harris & Kaine, 1994), Finland (Luwel, Noyons, & Moed, 1999), the United States of America (Crow & Bozeman, 1998) and so on, several researches focus on suitable metrics of research units that support the funding and other decisions of policy makers concerning these institutions to improve their organizational efficiency. Universities and similar institutions are evaluated either by peer review, which has some drawbacks due to its subjectivity and high cost, or by bibliometrics, which is cheaper and more objective than peer review, although biased (Geisler, 2000). Some studies have measured performance at the organizational level in universities and public research institutions (e.g. Coccia, 2001, 2004, 2005). However, these approaches have some drawbacks because of unit, technique and context of measurement; some improvements are, therefore, needed. In particular, governments need

* Correspondence address: CERIS-CNR, Institute for Economic Research on Firm and Growth, Collegio Carlo Alberto - via Real Collegio, n. 30, 10024 Moncalieri (Torino), Italy. Tel.: +39 011 68 24 925; fax: +39 011 68 24 966.
E-mail address: m.coccia@ceris.cnr.it.

[1] In general terms, change involves an attempt to alter the current way of thinking and acting by the organization's membership. More specifically, *strategic change* involves an attempt to change current modes of cognition and action to enable the organization to take advantage of important opportunities or to cope with consequential environmental threats (Gioia & Chittipeddi, 1991).

1751-1577/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2008.04.001

suitable metrics to identify which research institutions are high performers in homogeneous research fields and which are not, in order to efficiently allocate public funds and improve their scientific production. The problem I propose to tackle is: how is it possible to separate high performing from low performing research units within each research field? To answer this important managerial and economic question, an alternative model is presented to measure and evaluate the scientific research performance of Italian public research institutes, in order to support their strategic change.

2. Theoretical framework

Measuring scientific performance of public research units is a research field of scientometric literature (Geisler, 2000; Ingwersen & Larsen, 2005; Moed, Glanzel, & Schmoch, 2004; Torres-Salinas & Moed, 2007; Van Raan, 1988). Two main paths are:

(a) measuring performance of academic workers using indices such as Hirsch's *h*-index (Egghe & Rousseau, 2006).
(b) measuring performance of scientific institutions: teaching institutions (universities and colleges) and research institutions (such as public research units). The assessment of research institutions is based on indices indicating production, productivity, and impact of research groups (Braun & Glänzel, 2000; Luwel et al., 1999; Neri & Rodgers, 2006; Rodgers & Valadkhani, 2006). These techniques may not be sufficient to measure all the main aspects of public research bodies, since they do not consider the financial resources within public research institutions that are the sources of scientific production. As a consequence, bibliometrics (Moed, 2005) and standard techniques give only partial results in relation to the scientific performance of research units. The new literature puts forward indicators that measure several elements of research activities within public research bodies and are accurately synthesized through forms of clustering (Coccia, 2001, 2004, 2005; Geisler, 2000). Geisler (2000) suggests that the large numbers of indicators and measures assigned to various outputs of scientific organizations would be excellent candidates for some form of aggregation. The idea would be to create macro-indicators, based on the computation of normalized weights of indicators for all the outputs.

Coccia's approach (2001) argues that, in the calculation of scientific institutes' performance, the output is not the only important thing, input matters too. Therefore, he measures and evaluates the scientific performance of Italian research institutes using a series of indices based on financial, human, and scientific aspects, which consider all the organizational dimensions of research bodies.

In fact, Coccia (2001) uses *k* variables based on the following indexes:

o A = index of public funding to research bodies = $(\alpha_1, \alpha_2, \ldots, \alpha_n)$;
o B = index of self-financing = $(\beta_1, \beta_2, \ldots, \beta_n)$;
o X = index of personnel in training = $(\chi_1, \chi_2, \ldots, \chi_n)$;
o Δ = index of teaching activities by researchers = $(\delta_1, \delta_2, \ldots, \delta_n)$;
o E = index of publications with domestic diffusion = $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$;
o Φ = index of ISI[2] publications = $(\phi_1, \phi_2, \ldots, \phi_n)$;
o Γ = technometric index based on patents = $(\gamma_1, \gamma_2, \ldots, \gamma_n)$.

Let i be a research institute so that $i \in \{1, 2, \ldots, n\}$, let $(\alpha_i, \beta_i, \chi_i, \delta_i, \varepsilon_i, \phi_i, \gamma_i)$ be evaluation indices of the ith institute, then the Research Laboratory Evaluation (RELEV) function of the ith research unit is the following combination:

$$\Omega_{\text{relev}}(i) = 3 - \left(\frac{1}{\max A}\right) \times \alpha_i + \left(\frac{1}{\max B}\right) \times \beta_i + \left(\frac{1}{\max X}\right) \times \chi_i + \left(\frac{1}{\max \Delta}\right) \times \delta_i + \left(\frac{1}{\max E}\right) \times \varepsilon_i + 2\left(\frac{1}{\max \Phi}\right) \times \phi_i + (1 \text{ se } \gamma_i > 0;\ 0 \text{ se } \gamma_i = 0)$$

A model (0–1) is applied to vector Γ, this is 1 if the number of patents is at least 1, 0 if there are not patents. The reason for this is to avoid penalizing research bodies operating in social or mathematical sciences which do not produce patents, as instead occurs in other sciences (physics, chemistry, ...). This RELEV model I (Coccia, 2001) measures R&D performance on various dimensions and gives a single output, the R&D performance score = ω_i, which represents the total research performances of each research unit. The ω_i score synthesizes the financial, technological, and scientific aspects of the ith research institute. The weakness of this model is that almost all the operators have the same weight in the function, except index of ISI publications. Therefore, a second RELEV model is elaborated by Coccia (2004) using discriminant analysis with a direct method, i.e. all the output variables are entered into the model. In this case, the weights of the variables are

[2] Since ISI publications generate greater diffusion of knowledge, they are given in the RELEV function a double weight in comparison to those published in a domestic context.

produced by the technique itself. The model is based on the following discriminant function:

$$M = -5.178 + 1.389X_1 + 1.347X_2 + 1.007X_3 + 0.483X_4 - 0.00871X_5 \quad \text{(direct method)}$$

where: X_1: self-financing (€); X_2: training (no. of trainees); X_3: teaching (no. of courses); X_4: ISI publications (number); X_5: domestic publications (number).

The application of this second model on 200 Italian public research institutes, using data from year 2001, shows that 22.5% of the institutes are high performers.

Coccia (2005) presents the RELEV model III adjusted, which uses discriminant analysis with the Wilks Method and considers all research units within the Italian National Research Council (CNR). According to this model, 33.33% of Italian public research units are high performers.

The weakness of this approach is that there is only one research scientometric function for institutes operating in several scientific fields. This cannot be a suitable technique when it is necessary to evaluate the performance of research units, which operate in several research fields having different behaviour and structure in terms of scientific production, process and goals. This new method is needed since the reorganization of the Italian science sector, as well as in other countries, has been changing the organization and governance of research units.

3. New metrics to measure scientific performance of research units

The new adjusted Research Laboratory Evaluation (RELEV) model version IV, described here, improves models by Coccia (2001, 2004, 2005). It is based on a new sample and on the application of discriminant analysis with the Wilks Method (Everitt & Dunn, 1991), separating research units operating in different scientific fields. This model reduces the correlation among explanatory variables and represents a more reliable scientometric tool to assess the total research performance of institutes operating in different scientific fields. To introduce this new metrics, the following definitions are stated:

A research unit is a system of interacting and interrelated elements (e.g. researchers, technicians, public and private funds, information, equipment, etc.) to achieve a common purpose represented by the production of scientific research and knowledge (Coccia, 2001). The terms *research unit or institute* are used as synonyms in this study.

Performance of research institute: fixed specific objectives, it is the scientific production of a research unit over time, in comparison to other research units of the same scientific field within the same national system of innovation.

The *organizational efficiency of a research institute* is the best use of internal resources (such as economic and human resources) to produce scientific research output, e.g. publications, proceedings, patents, and so on.

Scientometrics of research units measure performance of research institutions, by using statistics and mathematical approaches, with the purpose of improving their organizational efficiency.

Remark: scientific performance as well as organizational efficiency of research institutes is measured by performance indicators such as publications per researcher. This paper considers *High performance institutes* (HPIs) as a proxy of the most efficient research units and *Low performance institutes* (LPIs) as a proxy of inefficient ones.

The new adjusted RELEV model version IV creates five mathematical functions focusing on Basic, Life, Earth, Social–Human Sciences and Technological, Engineering and Information Sciences. This distinction is due to differences between these research fields in terms of organization, scientific production process, outputs, scientific equipment, and behaviour towards the market. The following models are applied to a case study on Italy but they can be easily generalised, since scientific organizations tend to have similar behaviours (Laudel, 2006).

3.1. RELEV model adjusted version IV

The first methodological step of the adjusted RELEV model IV is to identify, within the research units operating in a scientific field, two groups of institutes represented by:

Group A: High scientific performance research institutes "HPIs"

Institutes in group A are organizations that combine scientific excellence and high domestic and international visibility. These research units have high productivity in relation to publications, proceedings, etc. These institutes are leaders in strategic scientific programs for Italian and European research plans and they have intense relationships with universities and other institutions.

Group B: Low scientific performance research institutes "LPIs"

Group B includes research institutes that have low scientific productivity and mainly administrative and bureaucratic culture (Crow & Bozeman, 1998), despite a world scenario that emphasises strategic change.

Table 1
Outputs and inputs of public research institutes

Output (t)	Input ($t-1$)
X_1 = ISI publications[a]	Researchers
X_2 = Non-ISI publications	Research fellows
X_3 = Proceedings	

[a] Number of publications in journals classified by the Institute of Scientific Information (ISI) in the Science Citation Index. These publications are an objective indicator of papers published in world's leading research journals in the sciences, technology, and Social–Human Sciences.

These sets are determined in an exogenous way using studies (Coccia, 2001, 2004, 2005) and information provided by the Italian National Research Council's (CNR's) Board of Directors (CNR Report, 2004). After having defined these two groups of institutions, the research investigates whether it is possible to predict the positioning of an institute, taken from a given population, in either of the above groups A or B, on the basis of key variables.

The approach considers the main outputs and inputs of public research units (Table 1).

In the calculation of performance of research units, the output is not the only important element, input matters too. Therefore, the following step is to calculate the productivity per ith institute, dividing the output t (described in Table 1) by the inputs t-1 (i.e. researchers + research fellows), since economic and human resources invested at time t-1, they generate scientific production at time t. In fact, organizational efficiency is measured by a performance indicator represented by the average productivity of the ith institute:

$$X_j = \text{Average productivity of } i\text{th institute} = \frac{\text{Output}_t}{\text{Inputs}_{t-1}}, \quad j = 1, 2, 3$$

These three indicators are considered independent variables of the adjusted RELEV model version IV. The discriminant function has the following form:

$$\omega(i) = \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3$$

where: $\omega(i)$ = research performance of ith Institute; δ_j = scalar values; X_j = indicators (j = 1, 2, 3). Another variable used in the model is a dummy variable, which can have only two values, that is $M_k = 1$ for HPIs = High scientific Performance Institutes, grouped in the A set, and $M_k = 2$ for LPIs = Low scientific Performance Institute (group B). For each case, this variable expresses its inclusion in either group A or group B.

Discriminant analysis is useful for situations where one wants to build a predictive model of group membership based on observed characteristics for each case. The procedure generates a discriminant function based on linear combinations of the predictor variables that provide the best discrimination between the groups. The functions are generated from a sample of cases for which group membership is known; the functions can then be applied to new cases with measurements for the predictor variables but unknown group membership. The discriminant analysis method uses the stepwise variable selection and adopts, as selection rule, the minimization of the value of Wilks' lambda statistic that chooses variables for the equation on the basis of how much they lower Wilks' lambda. At each step, the variable that minimizes the overall Wilks' lambda is entered. Wilks' lambda is a test statistic used in multivariate analysis of variance (MANOVA) to test whether there are differences between the means of identified groups of subjects on a combination of independent variables; in addition we use the F value.[3] A variable is entered into the model if its F value is greater than the entry value and is removed if the F value is less than the removal value. Entry must be greater than removal, and both values must be positive. In our case, the maximum number of steps is 6, the minimum level of tolerance is 0.001, the minimum partial F to be entered is 3, the maximum partial F to be removed is 2. The analysis is carried out by using the statistics software SPSS® (Statistical Package for Social Sciences), which provides all results described and analyzed in the following section.

4. Empirical analysis

The adjusted RELEV models version IV use data from the Italian National Research Council (CNR) in the 2004, 2005, 2006 period, published in 2007. CNR is the body that promotes, coordinates, and disciplines scientific research in Italy, in order to increase the scientific and technological progress of the Country.

The organizational structure of the Italian CNR is based on about 100 research units which have roughly 190 decentralised units. They belong to five scientific fields: Basic Sciences; Life Sciences; Earth Sciences; Social and Human Sciences; Technological, Engineering, and Information Sciences (Table 2).

[3] The F value is calculated dividing the between-group variance by the within-group variance. F values are used to compare distances between groups in the multivariate space.

Table 2
Distribution of research units among research fields

Area	Number of research units
Basic Sciences	28
Life Sciences	33
Earth Sciences	10
Social and Human Sciences	19
Technological, Engineering, and Information Sciences	18
Total	108

4.1. Results for each scientific field

The first methodological step of this new metrics based on discriminant analysis is to identify, within the population of each area, the groups of high and low performing research units by means of an exogenous selection according to information described in the methodology. Group A of high performers and group B of low performers have homogeneous internal characteristics, so that they can be used to classify the remaining institutes in HPIs and LPIs. The selection rule of the adjusted RELEV model IV, i.e. minimizing Wilks' lambda, chooses some of the following explicative variables that can change according to each scientific field:

(1) ISI publications (productivity).
(2) Non-ISI publications (productivity).
(3) Proceedings (productivity).

4.1.1. Basic Sciences

The canonical correlation of the discriminant function (Wilks' method) is high and equal to $R_{c1}^{1} = 0.981$ (Table 3).

The unstardardized canonical discriminant function is instead given by:

$$F_{\text{Basic}} = -3.341 + 7.34(\text{Proceedings} - \text{Productivity}) - 21.01(\text{Publications Non-ISI} - \text{Productivity})$$

The cut-off-point $\bar{Y}$ based on centroids (Table 4) is given by:

$$\bar{Y} = \frac{\bar{Y}_A + \bar{Y}_B}{2} = \frac{4.08 - 4.08}{2} = 0$$

As the cut-off-point is 0 (zero), it can be stated that the institutes which have a positive score from the unstardardized canonical discriminant function F_{Basic} are HPIs, whereas those with a negative score are LPIs. The Percentage of "grouped" cases correctly classified by this function is 100.00%.

4.1.2. Life Sciences

The canonical correlation of the discriminant function (Wilks' method) is high and equal to $R_{c1}^{1} = 0.998$ (Table 5).

The unstardardized canonical discriminant function is instead given by:

$$F_{\text{Life}} = -14.21 + 13.42(\text{Proceedings} - \text{Productivity}) + 6.237(\text{Publications ISI} - \text{Productivity})$$

Table 3
Canonical discriminant functions for Basic Sciences

Function	Eigenvalue	λ_1/λ_j	Canonical correlation	Wilks' lambda	Chi-square	Degree of freedom (d.f.)	Significance
1. Wilks' stepwise method	24.93	100	0.981	0.04	9.77	2	0.008

Table 4
Unstardardized canonical discriminant function evaluated at group means (group centroids[a]) for Basic Sciences

Group	Function 1 for Basic Sciences
A: $\bar{Y}_1$ = centroid group A	−4.08
B: $\bar{Y}_2$ = centroid group B	+4.08

[a] For each group in our sample, we can determine the location of the points representing the means for all variables in the multivariate space defined by the variables in the model. These points are called group *centroids*.

Table 5
Canonical discriminant functions for Life Sciences

Function	Eigenvalue	λ_1/λ_j	Canonical correlation	Wilks' lambda	Chi-square	Degree of freedom (d.f.)	Significance
1. Wilks' stepwise method	268.94	100	0.998	0.004	16.79	2	0.000

Table 6
Unstardardized canonical discriminant function evaluated at group means (group centroids) for Life Sciences

Group	Function 1 for Life Sciences
A: $\bar{Y}_1$ = centroid group A	−13.39
B: $\bar{Y}_2$ = centroid group B	+13.39

The cut-off-point $\bar{Y}$ based on centroids (Table 6) is given by:

$$\bar{Y} = \frac{\bar{Y}_A + \bar{Y}_B}{2} = \frac{13.39 - 13.39}{2} = 0$$

As the cut-off-point is 0 (zero), it can be stated that the institutes which have a positive score from the unstardardized canonical discriminant function F_{Life} are HPIs, whereas those with a negative score are LPIs. 100.00% of original group cases are correctly classified.

4.1.3. *Earth Sciences*

The canonical correlation of the discriminant function (Wilks' method) is high and equal to $R^1_{c1} = 1.000$ (Table 7).

The unstardardized canonical discriminant function is instead given by:

$$F_{\text{Earth}} = -67.86 + 18.76(\text{Proceedings} - \text{Productivity}) + 110.19(\text{Publications Non-ISI} - \text{Productivity})$$

The cut-off-point $\bar{Y}$ based on centroids (Table 8) is given by:

$$\bar{Y} = \frac{\bar{Y}_A + \bar{Y}_B}{2} = \frac{47.09 - 47.09}{2} = 0$$

As the cut-off-point is 0 (zero), it can be stated that the institutes which have a positive score from the unstardardized canonical discriminant function F_{Earth} are HPIs, whereas those with a negative score are LPIs. 100.00% of original group cases are correctly classified.

4.1.4. *Social and Human Sciences*

The canonical correlation of the discriminant function (Wilks' method) is high and equal to $R^1_{c1} = 1.000$ (Table 9).

The unstardardized canonical discriminant function is instead given by:

$$F_{\text{Social and human}} = -29.191 + 17.137(\text{Proceedings} - \text{Productivity}) + 13.667(\text{Publications Non-ISI} - \text{Productivity})$$

The cut-off-point $\bar{Y}$ based on centroids (Table 10) is given by:

$$\bar{Y} = \frac{\bar{Y}_A + \bar{Y}_B}{2} = \frac{23.32 - 23.32}{2} = 0$$

Table 7
Canonical discriminant functions for Earth Sciences

Function	Eigenvalue	λ_1/λ_j	Canonical correlation	Wilks' lambda	Chi-square	Degree of freedom (d.f.)	Significance
1. Wilks' stepwise method	4436.34	100	1.000	0.000	8.398	2	0.015

Table 8
Unstardardized canonical discriminant function evaluated at group means (group centroids) for Earth Sciences

Group	Function 1 for Earth Sciences
A: $\bar{Y}_1$ = centroid group A	−47.09
B: $\bar{Y}_2$ = centroid group B	+47.09

Table 9
Canonical discriminant functions for Social and Human Sciences

Function	Eigenvalue	λ_1/λ_j	Canonical correlation	Wilks' lambda	Chi-square	Degree of freedom (d.f.)	Significance
1. Wilks' stepwise method	1087.85	100	1.000	0.001	6.99	2	0.030

Table 10
Unstardardized canonical discriminant function evaluated at group means (group centroids) for Social and Human Sciences

Group	Function 1 for Social and Human Sciences
A: $\bar{Y}_1$ = centroid group A	−23.32
B: $\bar{Y}_2$ = centroid group B	+23.32

Table 11
Canonical discriminant functions for Technological, Engineering, and Information Sciences

Function	Eigenvalue	λ_1/λ_j	Canonical correlation	Wilks' lambda	Chi-square	Degree of freedom (d.f.)	Significance
1. Wilks' stepwise method	29.907	100	0.984	0.032	5.146	1	0.023

Table 12
Unstardardized canonical discriminant function evaluated at group means (group centroids) for Technological, Engineering, and Information Sciences

Group	Function 1 for Technological, Engineering, and Information Sciences
A: $\bar{Y}_1$ = centroid group A	−3.867
B: $\bar{Y}_2$ = centroid group B	+3.867

As the cut-off-point is 0 (zero), it can be stated that the institutes which have a positive score from the unstardardized canonical discriminant function $F_{\text{Social and human}}$ are HPIs, whereas those with a negative score are LPIs. 100.00% of original group cases are correctly classified.

4.1.5. Technological, Engineering and Information Sciences

The canonical correlation of the discriminant function (Wilks' method) is high and equal to $R^1_{c1} = 0.984$ (Table 11).

The unstardardized canonical discriminant function is instead given by:

$$F_{\text{Technological, Engineering and Information}} = -7.214 + 4.192(\text{Proceedings} - \text{Productivity})$$

The cut-off-point $\bar{Y}$ based on centroids (Table 12) is given by:

$$\bar{Y} = \frac{\bar{Y}_A + \bar{Y}_B}{2} = \frac{3.867 - 3.867}{2} = 0$$

As the cut-off-point is 0 (zero), it can be stated that the institutes which have a positive score from the unstardardized canonical discriminant function $F_{\text{Technological, Engineering and Information}}$ are HPIs, whereas those with a negative score are LPIs. 100.00% of original group cases are correctly classified.

4.2. DEA and comparison of results with the adjusted RELEV model version IV

To verify the analysis, the institutes' performance is also measured by DEA (Data Envelopment Analysis), considering that each institute has a process oriented to research production. DEA is a nonparametric method in operations research and economics for the estimation of production frontiers. It is used to empirically measure the productive efficiency of decision making units. DEA is based on a Linear Programming methodology to measure the efficiency of multiple Decision Making Units (DMUs) when the production process presents a structure of multiple inputs and outputs. DEA efficiency scores are defined as the ratios of the weighted sum of outputs and the weighted sum of inputs, where weights are endogenously attributed by DEA. If we consider a set of n DMUs consuming M (r = 1, . . ., M) inputs and producing S (i = 1, . . ., S) outputs, the efficiency of DMU a, is measured by solving:

$$\max h_a = \left(\frac{\sum_r u_r y_{ra}}{\sum_i v_i x_{ia}} \right)$$

subject to:

$$\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1\,, \quad j = 1, \ldots, n;\ r = 1, \ldots, S;\ i = 1, \ldots, M \qquad u_r, v_i > 0$$

where u_r, v_i are the output and input weights; y_{rj} and x_{ij} are the amounts of the rth output produced and the ith input consumed by unit j. The solution gives, for each DMU, the relative technical efficiency (TE) score [with TE = 100 indicating maximum efficiency and (100 − TE) equal to inefficiency]. According to the standard theoretical background, DEA is performed with the hypothesis of CRS (constant returns of scale).

As known, DEA (Cooper, Seiford, & Tone, 2000; Sexton, 1986) was primarily used to measure the efficiency of similar but autonomous Decision Making Units (DMU), i.e. any multi-inputs/outputs firms in the private or public sector. Checking on each DMU separately, DEA determines (by means of the linear programming technique) whether or not there is some combination of all the other DMUs in the system that can achieve the output of the jth unit with fewer inputs. As DEA-based efficiency estimation does not require any production function specification, it has become extremely popular in the analyses relating to public sectors, such as education, justice, and police forces among others. Some of the benefits of DEA are: no need to explicitly specify a mathematical form for the production function; proven to be useful in uncovering relationships that remain hidden for other methodologies; capable of handling multiple inputs and outputs; capable of being used with any input–output measurement; the sources of inefficiency can be analyzed and quantified for every evaluated unit.

Table 13
Comparison between DEA and discriminant analysis (adj. RELEV model version IV) among scientific sciences

	DEA[a]		F1[b]—discriminant analysis with Wilks' lambda (adj. RELEV model IV)[c]	
	Efficient (%)	Inefficient (%)	High performers institutes—HPIs (%)	Low performers institutes—LPIs (%)
Basic Sciences	21.4	78.6	46.15	53.85
Life Sciences	21.2	78.8	30.30	69.70
Earth Sciences	20.0	80.0	30.00	70.00
Social and Human Sciences	15.8	84.2	38.89	61.11
Technological, Engineering, and Information Sciences	27.8	72.2	29.41	70.59
Arithmetic mean	21.24	78.76	34.95	65.05
	Number	Number	Number	Number
Basic Sciences	6	22	12	14
Life Sciences	7	26	10	23
Earth Sciences	2	8	3	7
Social and Human Sciences	3	16	7	11
Technological, Engineering, and Information Sciences	5	13	5	12
Total	23	85	37	67

[a] Data 2003–2004.
[b] This function uses the *F* value: minimum partial *F* to be entered is 3, maximum partial *F* to be removed is 2.
[c] Data 2004–2005.

For instance, Rizzi (2001) presents an analysis of the efficiency of the academic departments at the University of Venice in Italy, using DEA and Deterministic Frontier analysis; the results are quite different for the two interpretative models, both with respect to the level of efficiency and to the ranking of departments. On the other hand, Agasisti and Dal Bianco (2006) consider the problem of determining technical efficiency for 58 Italian public universities, using DEA. They find a core of universities that perform well for various input and output specifications.

DEA is carried out with the hypothesis of constant returns of scale and is applied to the CNR dataset using the following variables:

- *Inputs*: Researchers, technicians, administrative staff.
- *Outputs*: Number of ISI and non-ISI publications, Proceedings at conferences.

By applying the RELEV adjusted model IV and DEA on the CNR research units belonging to scientific fields, Table 13 shows the percentage of efficient and inefficient research units, HPI and LPI Groups, among scientific fields. Conversely, Table 14 shows the comparison of inputs and outputs (using the arithmetic mean) for HPIs and LPIs among scientific fields.

5. Discussion and lessons learned

This new scientometric model is successfully applied to Italian public research institutes. Coccia (2004) shows that 22.5% of the total institutes in CNR are high performers (Discriminant Analysis with Direct Method), whereas Coccia (2005) finds that about 33.33% of Italian public research units are high performers (Discriminant Analysis with Wilks' Method, using a scientometric function for all the CNR research institutes). This research refines the results through a more reliable scientometric tool based on discriminant analysis with Wilks' lambda (DA) that is compared to DEA. In addition, the results are divided per scientific field, since the institutes operating in several areas have different organizational behaviours, equipment and scientific process/production. Results change according to each scientific field and technique applied: DEA shows that research units with higher efficiency (in percentage) are in the Technological, Engineering, and Information Sciences, and Basic Sciences, whereas lower percentage of efficiency is found in Social and Human Sciences. Conversely, DA shows that high performers are present in Basic and Social–Human Sciences. The RELEV function version IV refines the results of other studies and shows that the average number of HPIs is about 35 per cent of the total, when the arithmetic mean for all scientific fields is considered; this value is higher than Coccia's (2004) and close to Coccia's (2005), although the present paper shows values that range between 29.41 and 46.15%, depending on scientific field (Table 13). On the other hand, DEA shows lower values of HPIs (i.e. 21.24% of the total) in comparison to DA (Table 13), with a range between 15.8 (minimum) in Social and Human Sciences and 27.8 (maximum) in Technological, Engineering, and Information Sciences.

These results are due to the higher productivity of some scientific fields in comparison to others (Table 15): i.e. the highest productivity of ISI publications in the 2004–2005 period is in Basic Sciences, the lowest is in Social and Human Sciences; the maximum productivity of non-ISI Publications is in Social and Human Sciences, the minimum in Life Sciences; Productivity of proceedings is maximum in Technological, Engineering, and Information Sciences, minimum in Life Sciences (Table 15). In addition, a key finding is that HPIs are those with lower number of personnel and public funds (Table 14). This result is consistent with the economic literature (Coccia & Rolfo, 2007), which confirms the high performance of small-sized labs.

Table 14
Comparison of inputs and outputs by arithmetic mean between HPIs and LPIs per scientific fields

Arithmetic mean	Basic Sciences		Life Sciences		Earth Sciences		Social and Human Sciences		Technological, Engineering, and Information Sciences	
	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)
2004 – inputs										
Researchers	30.67	43.71	44.60	40.24	33.75	76.20	13.86	19.36	19.00	33.57
Research fellows	11.67	5.57	5.20	7.48	2.25	25.40	0.86	2.36	0.00	3.21
Technicians	11.00	21.52	19.00	22.28	24.50	41.80	9.43	6.82	7.00	22.57
Administrative staff	3.00	6.14	5.80	6.92	6.75	14.00	3.71	2.27	4.00	6.14
Total personnel	56.33	76.95	74.60	76.92	67.25	157.40	27.86	30.82	30.00	65.50
Public funds	485.67	794.57	608.80	846.16	553.25	1481.20	269.71	190.18	373.00	641.93
External funds	1487.33	1101.76	1333.00	2861.64	3315.75	3639.60	172.29	343.09	384.00	1764.93
External funds previous year	725.67	830.67	864.60	1136.56	1083.75	1971.00	282.00	318.18	344.00	860.00
2005 – inputs										
Researchers	33.08	40.93	42.30	39.00	44.33	57.29	16.57	16.91	20.80	36.50
Research fellows	4.58	4.57	8.20	8.00	0.33	21.57	2.57	3.09	0.00	4.75
Technicians	15.83	19.79	17.50	21.52	34.00	28.14	8.86	6.82	13.20	22.33
Administrative staff	4.67	6.43	5.80	7.35	12.67	10.00	2.29	3.55	4.40	5.92
Total personnel	58.17	71.71	73.80	75.87	91.33	117.00	30.29	30.36	38.40	69.50
Public funds	865.83	1011.93	780.90	1040.13	1142.33	1254.57	284.29	226.09	555.80	715.25
External funds	1556.58	1355.71	1840.80	3641.91	5280.33	3096.14	331.14	153.09	1004.40	2321.67
External funds previous year	856.00	804.07	737.10	912.17	1987.67	1559.57	215.86	147.73	356.40	1498.25
2005 – outputs										
Patents	2.67	1.19	1.60	0.84	0.00	0.60	0.00	0.00	0.00	0.64
Publications ISI	116.67	61.43	66.80	40.64	28.50	78.20	1.86	1.64	22.00	26.64
Publications non-ISI	4.00	5.43	26.60	4.68	23.25	34.60	26.00	7.45	17.00	16.43
Proceedings	78.33	11.00	47.60	1.88	47.50	29.80	17.57	10.18	54.00	30.14
Books	1.33	3.62	24.00	5.44	15.00	19.40	19.14	16.55	1.00	6.43
Reports	1.00	3.33	7.40	2.92	33.00	7.60	3.00	6.73	5.00	18.71
Results of projects	0.67	0.95	1.60	0.24	7.50	0.00	2.86	0.36	0.00	7.86
Applied results	0.00	2.48	8.80	0.32	1.25	3.20	0.00	2.73	0.00	2.79
Abstracts	12.00	10.43	28.20	23.08	48.00	76.80	12.29	9.45	0.00	6.93
Publishers activity	0.33	2.38	3.40	0.48	1.00	1.80	4.71	0.91	0.00	2.00
2006 – outputs										
Patents	2.33	3.00	1.60	1.48	0.00	0.00	0.00	0.00	0.80	0.83
Publications ISI	71.58	84.57	73.80	46.43	33.67	78.57	5.86	0.45	22.60	27.83
Publications non-ISI	5.58	5.93	15.20	10.48	24.00	20.14	15.86	13.00	6.40	20.17
Proceedings	56.17	10.14	38.80	5.22	53.00	48.29	31.14	5.36	43.60	46.33
Books	3.58	8.71	15.30	5.39	26.00	28.00	12.71	27.36	2.80	12.00
Reports	8.33	4.29	7.60	5.39	64.67	27.71	11.86	7.18	34.40	32.00
Results of projects	1.58	3.36	2.80	0.61	6.67	3.43	1.29	0.64	11.60	9.42
Applied results	1.67	1.86	0.70	1.61	1.33	4.43	1.86	0.45	3.60	5.75
Abstracts	34.00	42.50	24.50	32.61	39.00	86.57	21.86	13.00	5.20	16.00
Publishers activity	0.58	5.00	1.70	0.78	5.33	3.43	3.29	3.55	0.40	0.92

Table 15
Comparison of productivity per scientific field, and HPIs vs. LPIs

Productivity	Basic Sciences		Life Sciences		Earth Sciences		Social and Human Sciences		Technological, Engineering, and Information Sciences	
	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)	High performers institutes (HPIs)	Low performers institutes (LPIs)
2004										
Publications ISI-to-researchers + research fellows ratio	**2.76**	1.25	1.34	0.85	0.79	0.77	*0.13*	0.08	1.16	0.72
Publications non-ISI-to-researchers + research fellows ratio	*0.09*	0.11	0.53	0.10	0.65	0.34	**1.77**	0.34	0.89	0.45
Proceedings-to-researchers + research fellows ratio	1.85	0.22	*0.96*	0.04	1.32	0.29	1.19	0.47	**2.84**	0.82
2005										
Publications ISI-to-researchers + research fellows ratio	**1.90**	1.86	1.46	0.99	0.75	1.00	*0.31*	0.02	1.09	0.67
Publications non-ISI-to-researchers + research fellows ratio	*0.15*	0.13	0.30	0.22	0.54	0.26	**0.83**	0.65	0.31	0.49
Proceedings-to-researchers + research fellows ratio	1.49	0.22	*0.77*	0.11	1.19	0.61	1.63	0.27	**2.10**	1.12

Bold value is the maximum; *italics* value is the minimum.

Therefore, the research policy of consolidation based on the larger size of institutes, applied by the Italian Governments since 2001, has been increasing inefficiency. In fact, although there now are around 100 new institutes, these often have several (2–6) decentralized units spread over the territory and far from the institute – headquarters. This situation creates some diseconomies of scale, due to the increased costs of co-ordination of such decentralized units. Although large sizes may be more economical in some circumstances, there certainly are limits above which size becomes a synonym of inefficiency. In fact, the Italian consolidation of research units generates a large amount of administrative activities at the institutes – headquarters and therefore to manage several decentralized units is increasing organization bureaucracy (Crow & Bozeman, 1989; Gornitzka, Kyvik, & Larsen, 1998) that affects negatively scientific production and efficiency.

This approach, confirmed by DEA, shows that the biggest Italian public research body has a low average number of HPIs and the main cause is the lack of both a long-term national research strategy and a research policy shared by Italian governments of different coalitions. In fact, in the last decade governments have been implementing confused and ambiguous reorganizations that reduce scientific performances and increase general uncertainty. This lack of consistent research policies in Italy that do not consider research evaluation results (*ex ante* and *ex post*) to allocate public funds, have been reducing the efficiency of the Italian CNR. In addition, because of high public debt Italian Governments have been shrinking public research unit budgets both the efficient and inefficient scientific bodies, and this has been increasing the negative performance of the whole Italian system of innovation, thus having a negative impact on the competitiveness and economic growth of the overall economic system.

6. Concluding remarks

The measurement and the evaluation of scientific research are important and useful activities but they present two problems. Firstly, research is a public good that is difficult to quantify by prices (Arrow, 1962). Secondly, the goals of scientific institutions are more complex than those of private businesses. Universities or public research units must maximise their prestige, which in turn is a function of other variables that are not easily measured. The justification for the existence of performance indicators is mainly based on their implications in terms of financial resources. In fact, in order to stimulate scientific productivity, scientometric indicators need be positively correlated to financing, which need be higher for efficient and high performing research units. The problem that can arise is whether the performance indicators and the statistical–mathematical analysis alone can be sufficient to evaluate the performance of scientific organizations (Ball & Wilkinson, 1994). Johnes and Taylor (1990), for instance, compare five measures of performance in UK universities during the 1980s and find that no university performed either consistently well or consistently badly across all indicators. Moreover, they examine 11 different types of publications and calculate departmental rankings on the basis of some 120 different weighting schemes. The rank position achieved by departments proves to be very sensitive to the choice of the weights vector. Recent developments in linear programming are a partial solution to this problem. Data Envelopment Analysis (DEA) allows the evaluation of technical efficiency in each department without imposing an arbitrary weighting function (Sexton, 1986). However, Johnes (1992) discusses some problems and limitations concerning DEA. The discriminant analysis with Wilks' method applied in this research chooses the variables and the weights (through a mathematical algorithm) with the aim of reaching a high discrimination between HPIs and LPIs within each scientific field. The discriminant analysis (DA) as well as DEA do not require the specification of the functional form relating to inputs and outputs. Moreover, the weights are generated by the DEA and DA models themselves. The coefficients or weights developed via DEA are unique to the individual organization under evaluation, whereas in the discriminant analysis the weights are unique for all the institutes in each scientific field. DEA and DA also show different results since these techniques operate in different manners, nonetheless confirming the high number of low performing research units within the Italian CNR.

To reduce the risk of wrong research evaluation, the mathematical techniques have to be associated with other qualitative approaches such as *peer review* that, though subjective, offer opportunities for a more complete analysis of institutions with respect to the indicators of performance taken individually. Quantitative and qualitative approaches in scientometrics should be complementary in order to achieve a correct research evaluation. In the absence of performance indicators, using only subjective judgment could lead to distorted and exaggerated results, while trusting only to statistic techniques could be dangerous if there are not indications concerning how they operate and which indicators they use. To sum up, performance indicators are effective management techniques for research units and they should be used to provide valid support for scientific and research policies to allocate public funds to high performing research units. This strategy is crucial to increase the efficiency of the National System of Innovation and the competitiveness of modern economies.

Acknowledgements

I am very grateful to Prof. Secondo Rolfo for useful comments and to Dr. Fabrizio Tuzi for data. Great many thanks go to Prof. Luca Gnan, Andy Neely, Michael, Giorgio Sirilli, Miranda Cuffaro, Fabrizio Erbetta for useful suggestions. I have also benefited from comments by participants at: the 35th Australian Conference of Economics; the XLIII Conference of Società Italiana di Statistica; the 15th IAMOT conference; the scientometrics and infometrics conference (Madrid); the British Academy Management Conference; the 2003 IEEE Conference; the 2003 International Conference on Management of R&D, held at Indian Institute of Technology, as well as I have received useful suggestions by two anonymous referees. I am also thankful to Silvana Zelli and Diego Margon for research assistance. Any errors or omissions in the text are my sole responsibility. The views expressed herein are those of the author and not necessary those of Ceris.

References

Agasisti, T., & Dal Bianco, A. (2006). Data envelopment analysis of the Italian university system: Theoretical issues and policy implications. *International Journal of Business Performance Management*, *8*(4), 344–367.
Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson (Ed.), *The rate and direction of inventive activity: Economic and social factors*. Princeton: Princeton University Press.
Ball, R., & Wilkinson, R. (1994). The use and abuse of performance indicators in UK higher education. *Higher Education*, *27*, 417–427.
Braun, T., & Glänzel, W. (2000). Chemistry research in Eastern Central Europe (1992–1997): Facts and figures on publication output and citation impact for 13 countries. *Scientometrics*, *49*(2), 187–214.
CNR Report (2004). *Research results 2003, Budget 2004–2005*. Roma: D'Anselmi editore, Hoepli.
Coccia, M. (2001). A basic model for evaluating R&D performance: Theory and application in Italy. *R&D Management*, *31*(4), 453–464.
Coccia, M. (2004). Models for measuring the research performance and identifying the productivity of public research institutes. *R&D Management*, *34*(3), 267–280.
Coccia, M. (2005). Scientometric model for the assessment of the scientific research performance within the public institutes. *Scientometrics*, *65*(3), 97–311.
Coccia, M., & Rolfo, S. (2007). How Research policy changes can affect the organization and productivity of public research institutes. *Journal of Comparative Policy Analysis. Research and Practice*, *9*(3), 215–233.
Cooper, W. W., Seiford, L. M., & Tone, K. (2000). *Data Envelopment Analysis*. Boston/Dordrecht/London: Kluwer Academic Publishers.
Crow, M., & Bozeman, B. (1989). Bureaucratization in the Laboratory. *Research Technology Management*, *32*(September–October (5)), 30–31.
Crow, M., & Bozeman, B. (1998). *Limited by design. R&D Research units in the U.S. National Innovation System*. New York: Columbia University Press.
Cruz-Castro, L. & Sanz-Menéndez, L. (2006). *Research evaluation in transition: Individual versus organisational assessment in Spain* (Working paper). Unidad de Politicals Comparadad (CSIC), 06-12.
Egghe, L., & Rousseau, R. (2006). Infometric model for Hirsch index. *Scientometrics*, *69*(1), 121–129.
Everitt, B. S., & Dunn, G. (1991). *Applied multivariate data analysis*. London: Edward Arnold., pp. 219–220.
Geisler, E. (2000). *The metrics of science and technology*. Westport/Connecticut/London: Quorum Books.
Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal*, *12*(6), 433–448.
Gornitzka, Å., Kyvik, S., & Larsen, I. M. (1998). The bureaucratization of universities. *Minerva*, *36*(1), 21–47.
Harris, G., & Kaine, G. (1994). The determinants of research performance: A study of Australian university economists. *Higher Education*, *27*, 191–201.
Ingwersen, P., & Larsen, B. (Eds.). (2005). *Proceedings of the 10th International conference of the international society for scientometrics and infometrics*. Stockholm, Sweden: Karolinska University Press.
Johnes, G. (1992). Performance indicators in higher education: A survey of recent work. *Oxford Review of Economic Policy*, *8*(2), 19–34.
Johnes, G., & Taylor, J. (1990). *Performance indicators in higher education*. Buckingham: Open University Press.
Krull, W. (1995). The Max Planck experience of evaluation. *Scientometrics*, *34*(3), 441–450.
Larédo, P. (2002). L'évaluation de la recherche en France: Éléments de cadrage. *Présentation pour la session du Conseil scientifique du CNRS sur l'évaluation*, 1–2. Juillet, Paris.
Laudel, G. (2006). The art of getting funded: How scientists adapt to their funding conditions. *Science and Public Policy*, *33*(7), 489–504.
Luwel, M., Noyons, C. M., & Moed, F. (1999). Bibliometric assessment of research performance in Flanders: Policy background and implications. *R&D Management*, *29*(2), 133–141.
Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.
Moed, H. F., Glanzel, W., & Schmoch, U. (Eds.). (2004). *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*. London: Kluwer Academic.
Neri, F., & Rodgers, J. R. (2006). Ranking australian economics departments by research productivity. *The Economic Record*, *82*(September (Special Issue)), S74–S84.
Rizzi, D. (2001). L'efficienza dei dipartimenti dell'Università Ca' Foscari di Venezia via DEA e DFA. *Economia Pubblica*, *3*, 103–136.
Rodgers, J. R., & Valadkhani, A. (2006). A multidimensional ranking of Australian Economics Departments. *The Economic Record*, *82*(March (256)), 30–43.
Sexton, T. R. (1986). The methodology of data envelopment analysis. In R. H. Silkman (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. San Francisco: Jossey Bass.
Silvani, A., Sirilli, G., & Tuzi, F. (2005). R&D evaluation in Italy: More needs to be done. *Research Evaluation*, *14*(3), 207–215.
Torres-Salinas, D., & Moed, H. F. (Eds.). (2007). *Proceeding of ISSI 2007*.
Van Raan, A. F. J. (Ed.). (1988). *Handbook on science and technology indicators*. North Holland: Amsterdam.
Vilkas, C. (2004). L'évaluation au CNRS. *La Revue pour l'histoire du CNRS*, *11*(November).