Statistics and Its Interface Volume 9 (2016) 303–313

Assessing scientific research activity evaluation models using multivariate analysis

Rosaria Romano and Cristina Davino*

The authors of this paper propose a method, based both on confirmatory and exploratory data analysis, aiming to assess the variability arising from the Composite Indicators (CIs) construction process. This research refers to an evaluation exercise very important for universities: the assessment of scientific research. The aim of every evaluation system is to synthesize all the information collected at universities into a unique CI, which will allow comparison of performances or ranks of the objects under evaluation. Since the methodology adopted to construct the CI is just one possible solution among several acceptable alternatives, it is reasonable to wonder about the results from the other options. The proposed approach investigates the impact of the different sources of variability occurring in CIs construction, also taking into account the external information available for each statistical unit. The term CI variability is used in the meaning of CI stability and it refers to differences emerging among CIs obtained using different subjective choices to construct the CI. Furthermore, the stability of the results is assessed through a combination of graphical tools and resampling methods. An empirical analysis is provided to discuss the methodological proposal. The research refers to the 'University Planning and Evaluation 2007–2009' system, implemented by the Italian government to finance public universities.

Keywords and phrases: Composite indicators, Stability, Analysis of variance, Principal component analysis, Scientific research activity.

1. INTRODUCTION

The aim of the paper is to introduce an innovative approach to assess research activity evaluation models, based both on confirmatory and exploratory statistical methods. The empirical reference framework is a system implemented by the Italian government for financing public universities[1], which is based on a set of predefined indicators properly transformed and aggregated into a unique CI, by which universities are ranked and appropriately financed. It is a matter of fact that the obtained CI depends on several subjective choices. These choices are well known in literature as *uncertainty factors* [16] and involve all the steps followed in the CI definition process: definition of the phenomenon to be measured (selection of factors, indicators and statistical units), pre-processing of the original indicators (missing data imputation, indicator transformations) and construction of the CI (identification of the system of weights, selection of the aggregation method). Thus, a study on the assessment of the impact of the different sources of variability occurring in the CI construction is advisable. It is a matter of fact that to each CI corresponds a given ranking of the units and, consequently, choosing among alternative CIs means to advantage or disadvantage some units instead of others. The proposed approach intends to provide decision makers valuable information on the consequences deriving from alternative CIs. The final decision, namely which CI to adopt, is up to the decision maker.

This paper centers on one single component of the Italian funding model, the *scientific research activity evaluation*, for two main reasons: it represents the primary component on which Italian universities are called to invest in the future and the proposed approach has revealed for this component the highest instability to the governmental model as compared to the others.

The contribution of this proposal is the presentation of a statistical approach to assess CIs based on a mixture of multivariate exploratory and confirmatory analyses [11]. The main objective is to investigate the impact of the different subjective choices on the CI variability as well as the related individual differences among the statistical units. More specifically, according to the procedure used to obtain the CI, each unit is assigned a different CI value (and thus a different rank). The aim of the study is to assess which are the reasons of different CI values assigned to the units, thus providing decision makers with information on the impact derived from alternative CI construction strategies. This strategy has already been proposed by [13] in the context of consumers' preferences, and it has been adapted by [4, 6] for the CI sensitivity analysis framework [16] and for the subjective measurement framework [5].

The aim of the present work is to upgrade such an approach to the assessment of research activity evaluation models. Moreover, the proposal is strengthened by the introduction of additional tools to facilitate the discrimination among alternative CIs. Finally, a study of the stability of the multivariate results [8] according to the role played by each

*Corresponding author.

[1] University Planning and Evaluation 2007–2009' system (http://www.istruzione.it/web/universita/programmazione-2007-2009).

Table 1. Statistics of the indicators

I	Min	Max	Q1	Median	Q3	Mean	Variation coefficient	Fisher asymmetry index
$b2$	1.75	13.39	2.97	3.61	4.37	3.98	0.49	2.63
$b3$	0.00	0.44	0.07	0.13	0.20	0.14	0.66	0.84
$b4$	3.21	37.01	13.01	16.18	21.16	17.20	0.42	0.63
$b5$	0.10	0.87	0.60	0.68	0.76	0.66	0.23	-1.14

university in the CI composition is provided. The interpretation of the results is facilitated by the graphical power of the exploratory methods and by the use of resampling methods and confidence ellipses [17] to explore the stability of each observed unit.

This paper is organised as follows. Section 2 introduces the reference framework describing the data used by the Italian government to fund universities. Section 3 is dedicated to the proposed approach: it contains the methodological issues and the main results. Finally, some concluding remarks and further developments are provided.

2. SCIENTIFIC RESEARCH ACTIVITY EVALUATION

University evaluation has become an unavoidable requirement so that various evaluation exercises have been proposed worldwide [15, 3]. In the European countries, the Research Assessment Exercise [14] is one of the most consolidated and formalised assessment processes for the evaluation of the quality of research products. In Italy, even though several evaluation processes have been recommended, a unique and formalized approach has been recently introduced by the National Agency for the Evaluation of Universities and Research Institutes (ANVUR). In addition to ANVUR, several national authorities have taken care about the Italian scientific research evaluation: the Italian Department of Education (MIUR), the National Committee for University System Evaluation (CNVSU) and the Italian Committee for Research Evaluation (CIVR). The proposals differ in several features: the proposer agency, the aim, the subject of the evaluation (e.g. decision support, funding allocation, quality certification), the time of the evaluation (ex-ante, ex-post), the reference model (e.g. number of components), the proposed methodology (e.g. subjective and/or objective indicators, qualitative and/or quantitative indicators, the aggregation method) and the subjects of the evaluation (e.g. universities, departments, teachers). The final result of whatever exercise is to synthesize all the collected information into a unique CI, which will allow the comparison of performances or ranks of the subjects under evaluation.

The present paper does not aim to discuss the main differences among the proposals, but to highlight consequences deriving from alternative choices adopted in the evaluation exercise. In particular, this paper refers to the 'University Planning and Evaluation 2007–2009' system[2], implemented by the Italian government to finance public universities. Such a system guarantees that a part of public funds is assigned to a university according to the evaluation of the improvements made during a given period of time and combines all five components (teaching, research, internationalisation, services, staff) considered necessary to define a university ranking.

This research centers on one single component of the funding model, the scientific research activity. According to the system implemented by the Italian government to finance public universities, the following five indicators can be considered to measure research component in each university:

- $b1$: the ratio between active researchers and the number of researchers;
- $b2$: the average number of Phd fellowships for each Phd course;
- $b3$: the proportion of Phd fellowships with external funding;
- $b4$: the average economic sources for each researcher (millions of Euros);
- $b5$: the proportion of income from external sponsors.

Indicator $b1$ is not considered because it is not available for all the studied universities.

Data refers to the whole set of public universities (59 units).

The main descriptive statistics (Table 1) reveal strong differences among the indicators in central tendency, in variability and in asymmetry. Boxplots of each indicator with respect to the university *dimension* are shown in the Appendix. University *dimension* is measured in terms of the number of students. Each university is classified according to the following groups: *mega* (>40,000 students), *big* (20,000−40,000 students), *medium* (10,000−20,000 students), *small* (<10,000 students) and *polytechnics*. Boxplots in the Appendix reveal, for each indicator, different distributions according to the different university dimension. Moreover, a comparative analysis of the dimension distributions across the indicators shows a remarkable heterogeneity. For instance, polytechnics are positioned on the top values of indicators b4 and b5 highlighting a well known capability of

[2]http://www.istruzione.it/web/universita/programmazione-2007-2009.

these universities to find external funding for financing their researches.

Actually, the approach proposed by the Italian ministry removes from each indicator measured in the current year (2007) the average over the previous three-year period. Let b be one of the four indicators, $\bar{b}$ the average over the previous three-year period, $d_b = b - \bar{b}$ is the difference between them. Such a difference is normalised according to the following transformations:

$$pos_b = d_b - min(d_b) + 1 \quad (1)$$
$$norm_b = \frac{pos_b}{\sum pos_b}$$

It results that each transformed indicator ranges from 0 to 1. Finally, the CI is obtained by simply averaging the four $norm_b$ indicators.

3. THE PROPOSED APPROACH: METHODOLOGY AND RESULTS

The Italian government funding model is based on subjective choices which introduce variability into the CI values [2]. It is thus important to assess how much each methodological choice affects the CI variability and to evaluate the differences and relationships among such alternative CIs.

To this end, the scientific literature offers many proposals: some are based on comparison among the unit rankings derived from alternative strategies of analysis [9]; other authors propose the use of item analysis to compare CIs to each single indicator or to an external assessor not included in the analysis [1]. Saltelli et al. [16] proposed methods able to measure the uncertainty associated with the CI in terms of its variability and the CI sensitivity in terms of the contribution of each factor involved in the construction of the CI on its variability. A technical drawback of some of these methods is their computational cost, as they require many CI simulations. Moreover, these methods provide information on the various uncertainty factors without highlighting the role of the corresponding alternatives and they are based on univariate analysis of each determinant of the CI construction without considering the interactions among them.

The main focus of this paper is to propose an alternative method for assessing CIs, which investigates the impact of the different sources of variability occurring in the construction of the CI, also taking into account the external information available for each statistical unit. In this case study the external information is represented by the university *dimension*. The use of external information is crucial in this type of analysis because it provides additional useful information for effective interpretation of the final results.

The work presented in this paper is embedded in the multivariate framework, combining explicative (Analysis of Variance-ANOVA) and exploratory (Principal Component Analysis-PCA) methods. Following the typical terminology of ANOVA, each issue to be defined to construct a CI is named factor (i.e. dimension selection, variable selection, data transformation, unit selection, weighting method, aggregation method) and its possible alternatives are called levels (i.e. linear and geometric for the aggregation method).

Table 2. Uncertainty factors

Factor	Definition	Levels
t	Transformation	*ministry* (Minis); minmax2 (MinMax2)
a	Aggregation	*linear* (Lin); geometric (Geom)
s	Exclusion	*none*; b2; b3; b4; b5

Let X $(N \times P)$ be the data matrix of the $P = 4$ indicators observed on the $N = 59$ observations (universities) and let us consider the three uncertainty factors: *transformation* (t), *aggregation* (a) and *exclusion* (s), respectively with $I = 2$, $J = 2$ and $K = 5$ levels. A description of the different uncertainty factors and related levels used in this study is provided in Table 2[3]. Besides the Italian ministry method discussed above, the other level of the *transformation* factor is not applied to the simple difference between an indicator and the average over the previous three-year period but to the variation with respect to the average $(d_b/\bar{b})$. It provides a CI in the range $[0, 1]$:

$$MinMax2 : \frac{\frac{d_b}{b} - min(\frac{d_b}{b})}{max(\frac{d_b}{b}) - min(\frac{d_b}{b})} \quad (2)$$

The *aggregation* factor considers the arithmetic mean (linear) and the geometric mean (geometric). Finally, the *exclusion* consists of eliminating one indicator at a time or any one of them. These factors and levels give rise to 20 possible scenarios, namely, 20 possible CIs. In Table 3, the design matrix shows the composition of the CIs derived from the uncertainty factors in Table 2. For example, the CI labeled *MinisLinNone* represents the CI proposed by the Italian Ministry to evaluate scientific research activity, as described in Formula (1). Hereinafter it will be considered a sort of benchmark. An additional factor is represented by the *units* (u) consisting of as many levels as the number of universities and the *external information* is introduced in the model considering the additional factor *dimension* (d) with $Z = 5$ levels.

The CI value assigned to each university changes according to the preferred combination of factors and levels: as a consequence, the position of each university in the general ranking can be different. In Figure 1, for each mega university, the distribution of its positions in the 20 possible rankings is shown. Boxplots highlight relevant variations in the university rankings, but they only provide information on the variability without highlighting the role of the different factors, the respective levels and their interactions. Moreover, such exploration is univariate in the sense that the variability of the CIs is explored for each university individually.

The proposed strategy consists of three main steps:

[3]Levels used by the government model are in italics. Text in brackets corresponds to the labels in the plots.

Table 3. Design matrix

CI	Transformation	Aggregation	Exclusion
MinisLinNone	ministry	linear	none
MinisLin-b2	ministry	linear	b2
MinisLin-b3	ministry	linear	b3
MinisLin-b4	ministry	linear	b4
MinisLin-b5	ministry	linear	b5
MinisGeomNone	ministry	geometric	none
MinisGeom-b2	ministry	geometric	b2
MinisGeom-b3	ministry	geometric	b3
MinisGeom-b4	ministry	geometric	b4
MinisGeom-b5	ministry	geometric	b5
Minmax2LinNone	minmax2	linear	none
Minmax2Lin-b2	minmax2	linear	b2
Minmax2Lin-b3	minmax2	linear	b3
Minmax2Lin-b4	minmax2	linear	b4
Minmax2Lin-b5	minmax2	linear	b5
Minmax2GeomNone	minmax2	geometric	none
Minmax2Geom-b2	minmax2	geometric	b2
Minmax2Geom-b3	minmax2	geometric	b3
Minmax2Geom-b4	minmax2	geometric	b4
Minmax2Geom-b5	minmax2	geometric	b5

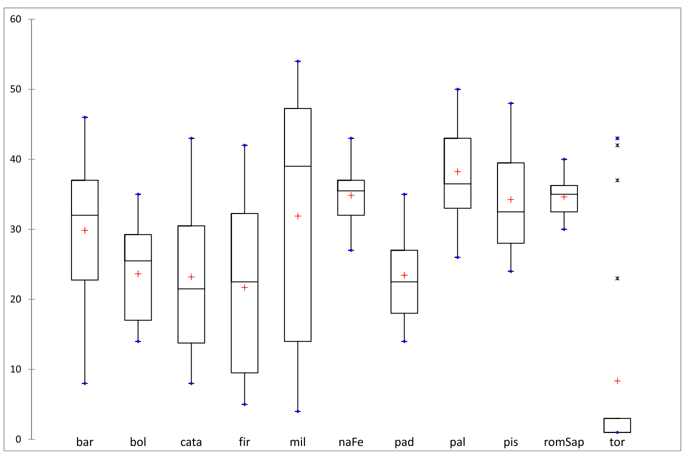

Figure 1. Distributions of the CI rankings for each mega university.

- evaluation of the impact of factors and additional information using ANOVA;
- exploration of interactions among factors and units using PCA;
- assessment of the stability of the multivariate results.

Each step will be described in detail in the following subsections together with the respective main results.

3.1 Evaluation of the impact of factors and additional information using ANOVA

The first stage of the multi-step analysis allows us to investigate the general tendencies of the different factors on the overall set of units. Let $Y(N \times I \times J \times K, 1)$ be the vector

Table 4. ANOVA results

Source	Sum Squares	d.f.	F
Transformation	8.498	1	2114.27
Aggregation	1.820	1	452.95
Exclusion	3.445	4	214.26
University (Dimension)	5.319	54	24.51
Dimension	0.464	4	28.84
Transformation*Aggregation	0.077	1	19.24
Aggregation*Exclusion	0.089	4	5.52
Transformation*Exclusion	0.902	4	57.24
Transformation*Aggregation*Exclusion	0.047	4	2.92
Error	4.429	1102	
Total	25.108	1179	

of CIs obtained by stacking the $I \times J \times K$ CIs observed on the N units. The effect of each uncertainty factor on the CI variability is estimated by means of the following model:

$$y_{ijkzn} = \mu + t_i + a_j + s_k + d_z + u(d)_n + \\ + ta_{ij} + ts_{ik} + as_{jk} + tas_{ijk} + e_{ijkzn} \quad (3)$$

where y_{ijkzn} is the n^{th} observation belonging to the z^{th} group and obtained using the i^{th} $(i = 1,...I)$ level of the t factor, the j^{th} $(j = 1,...J)$ *aggregation* method and the k^{th} $(k = 1,...K)$ level of the s factor. The general mean is represented by μ, while t_i, a_j, s_k are the main effects of the three uncertainty factors and ta_{ij}, ts_{ik}, as_{jk}, tas_{ijk} are their second and third order interaction effects. The main effect of the extra factor represented by the universities is u_n, which is nested in the dimension factor d_z[4], while interactions between universities and uncertainty factors are included in the error term e_{ijkzn}. The university factor is assumed fixed and not random because the analysed universities correspond to the whole population of universities in Italy.

The model in (3) consists in a simultaneous model for all units (i.e. all universities in this study). Results from this ANOVA model show which uncertainty factors strongly affect or do not affect the stability of the CI, whilst the impact of these effects on each single university is encompassed in the residuals. Since the set of units corresponds to the whole population of universities, it makes no sense to evaluate the significance of the ANOVA results. However, the size of the different F ratios provides information on the effect of each uncertainty factor on the variability of the CIs. Model results in Table 4 show that the *transformation* factor plays a prominent role in the variability of the CIs, followed by the other two uncertainty factors. Among the interactions, the most relevant effect is shown by the *transformation*exclusion* combination.

A further investigation of the effects of each factor on the assessment of CIs can be obtained by the analysis of

[4] A factor is nested when subgroups of units match only one of the levels of the nesting factor and not each one of them, as usually happens in a crossed design.

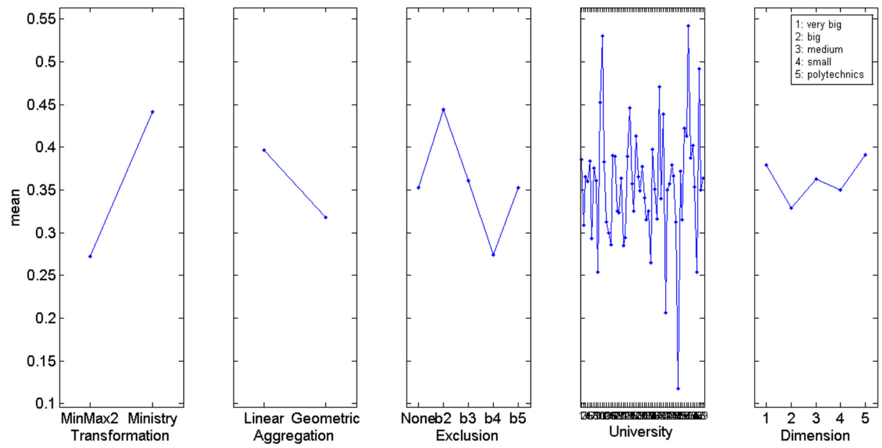

Figure 2. ANOVA means plot.

Figure 2, showing the average values of the CIs according to the different factor levels. The means plot provides the first signs of consequences deriving from alternative choices in the CI construction. For example, the effect played by the *transformation* factor is very different among its levels. Moreover, the *unit* factor shows a great variability among universities, but a deepened exploration of this factor is impossible from the mean plot because of the huge number of units.

3.2 Exploration of interactions among factors and units using PCA

After analysing the effect of the uncertainty factors at the population level (i.e. over all universities) by means of the simultaneous ANOVA in (3), the second step of the analysis is to investigate the effect of the same factors with respect to each single university or groups of universities. This can be achieved by exploiting the potentialities of PCA to explore the effect of each factor on each single statistical unit after removing the general tendency. This analysis is graphically oriented and very flexible with respect to the number of units in the data set.

Specifically, in order to explore individual differences among universities, a PCA is performed on the residuals from the ANOVA model used above:

$$
\begin{aligned}
\hat{e} &= y_{ijkzn} - [\hat{y}_{ijkzn} \\
&= y_{ijkzn} - \mu + t_i + a_j + s_k + \\
&\quad d_z + u(d)_n + \\
&\quad + ta_{ij} + ts_{ik} + as_{jk} + tsa_{ijk}]
\end{aligned}
\tag{4}
$$

Residuals from this model contain information on individual differences among units with respect to the factors plus the random error. In order to run a PCA on the residuals, they have to be rearranged in a data matrix (N, $(I \times J \times K)$) with the units as rows and the CIs as columns. This matrix has a special structure due to the fact that the residuals come from a saturated model, i.e. a model with all uncertainty factors and their interactions. Specifically, the matrix is double-centered, that is, both its rows and columns sum up to 0. The effect of mean centering for each row means that additive differences between universities have been eliminated. On the other hand, the effect of centering the residual data for each column is that for each combination of levels and factors (CI), the values represent the universities' distance to the *average university* for that CI. Those universities having a positive residual value for a CI, have a score on that CI higher than that of the *average university* and vice versa. This means that results from this PCA highlight units with CI values, due to a specific combination of factors, either higher or lower than the average unit. These units will be identified as those which are more sensitive to a specific factor level combination. The impact of the external information is investigated by including it in the PCA as a supplementary variable and projecting it onto the factorial plane.

Results from the PCA show that a solution considering the first two principal components explains 57% of the total variability. An additional percentage of variation equal to 16% is also explained by the third component, however this additional information does not provide a particular advantage in the interpretation of the results but it is included for a comparative study.

Loadings in Figure 3 span the factorial space in all directions, indicating a strong heterogeneity in the universities, according to the different selected combinations of levels and factors[5]. The first principal component, which generally explains the main information, shows how the CIs differ mainly with respect to the levels of the *transformation* factor. This result is coherent with the ANOVA results shown in Figure 2. The second principal component highlights CIs split

[5]The loading belonging to the combination *MinisLinNone* included in a rectangle refers to the CI adopted by the Italian Ministry of Education.

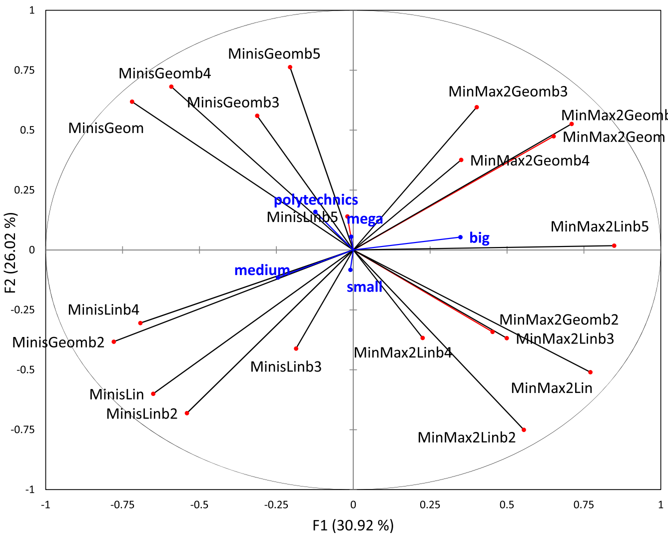

Figure 3. PCA loading plot using the ANOVA residuals (first and second principal component).

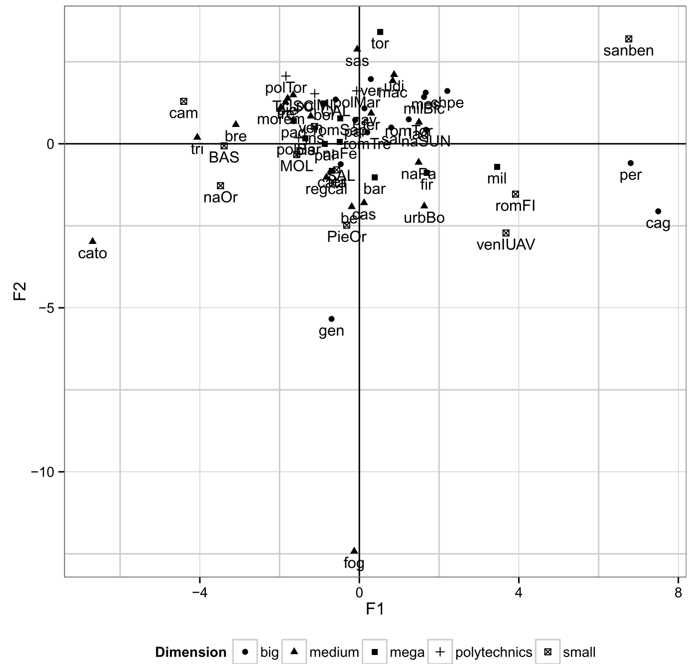

Figure 4. PCA scores plot using the ANOVA residuals.

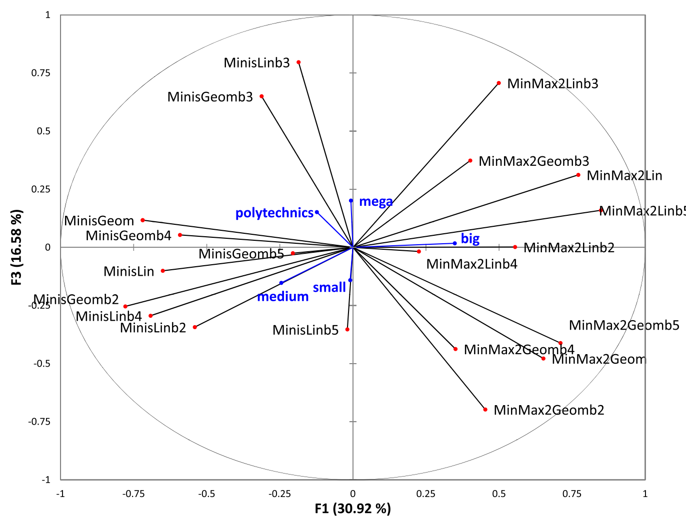

Figure 5. PCA loading plot using the ANOVA residuals (first and third principal component).

according to the *aggregation* factor. It is worth noting that, contrary to the traditional interpretation of PCA results where the basic aim is generally to identify latent dimensions, here, the main objective is to highlight the relations among the alternative CIs in connection with the statistical units.

The projection of the modalities of the external variable (*dimension*) onto the plan spanned by the ANOVA residuals shows that different factor level combinations characterize the five groups of universities. This is underlined by the position of the different modalities which are spanned on the plane (in bold in Figure 3). Specifically, the factorial plan shows which are the CIs characterising each group of universities.

The score plot in Figure 4 highlights differences among universities[6] showing which ones are sensitive to the different factor combinations represented in the related loading plot.

Loading plot crossing the first and the third principal component (Figure 5) shows an additional effect of the *exclusion* factor discriminating between CIs excluding the *b3* indicator on the positive verse of the third component and the *b2* on the opposite side. However, it should be taken into account in this interpretation that this component only explains 16% of the total variability.

Due to the explanatory capability and readability of the PCA plots, it is possible to identify which consequences are related to the choice of a given CI. The proposed approach does not aim to choose the best CI because such a choice is related to the meaning assigned to the best CI concept. Anyway, additional tools are suggested to support the decision.

[6] A table with the names and acronyms of all universities is included in Appendix 1.

Confining our attention to the first factor, the analysis of its correlations with the CIs can be useful if the aim is to select the CI to which the units are more or less sensitive. From Figure 3, it is evident that *MinisLinb5* plays the lowest effect on the units, while *MinisGeomb2* the highest.

Exploiting the added value of the second component on the explained variability, units can be ranked according to their distances from the origin on the first two factors. The aim of Figure 6 is to provide a measure to synthesize the differences and similarities in university sensitivity to the

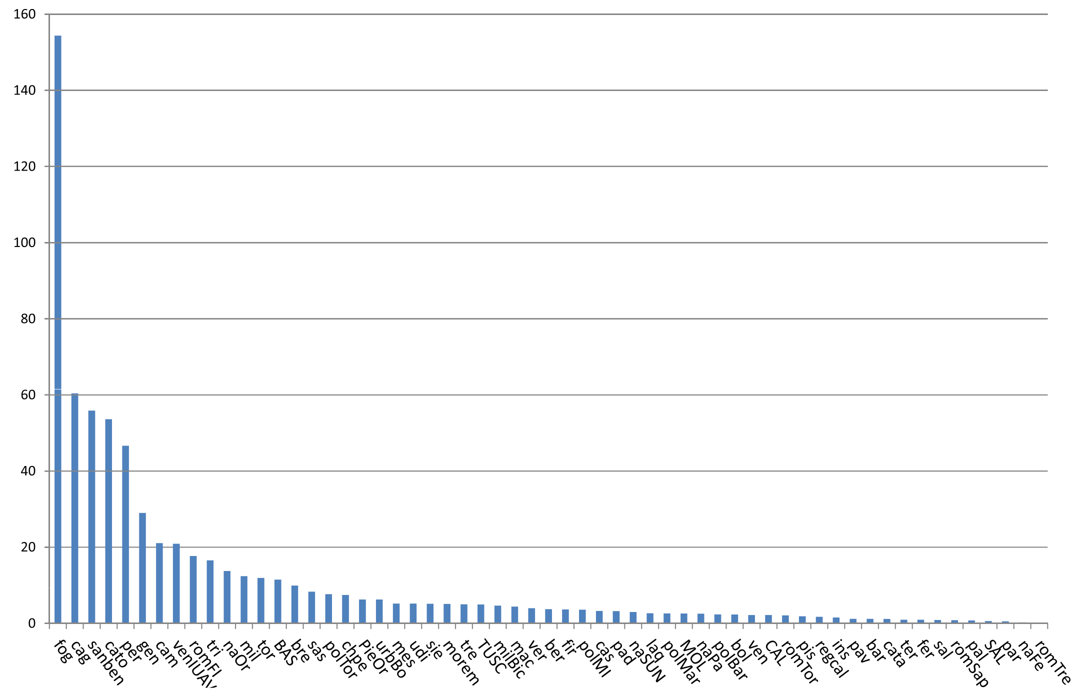

Figure 6. Distance (y-axis) of each university (x-axis) computed on the first 2 factors.

choice of a CI among many. Behind confirming the role of the most sensitive units with the highest coordinates on the score plot, this Figure allows to differentiate the behavior of also those units situated on the origin of the axes which is difficult to read from the score plot.

Finally, a cluster analysis [10, 12] on the first two components of the PCA permits to characterise groups of units sensitive to the same combinations of factors and levels. Figure 7 shows the best obtained partition in five groups on the factorial plane.

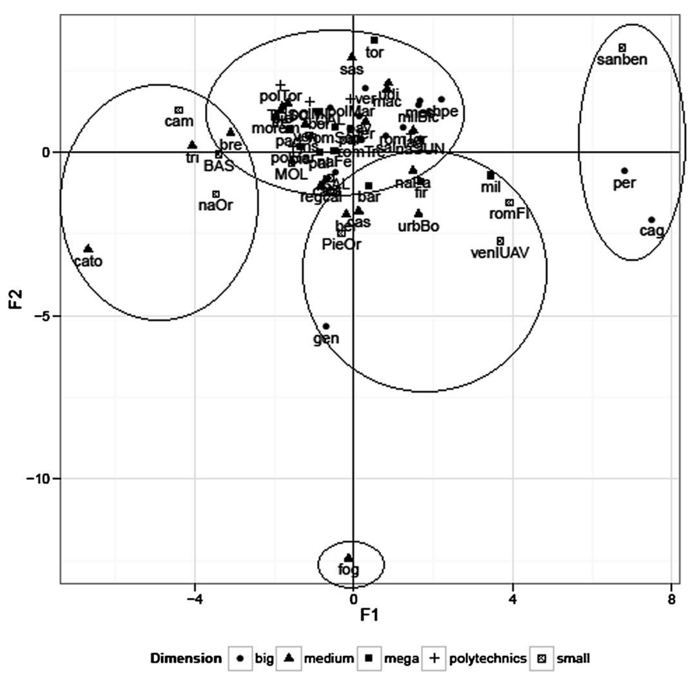

Figure 7. Units represented on the score plot are grouped according to the cluster analysis results.

3.3 Assessment of the stability of the multivariate results

Once the impact of the uncertainty factors on each university has been evaluated, it is reasonable to wonder about the stability of these results. Stability may be investigated both with respect to the role played by a single university and by a group of universities on the obtained results.

To assess if and how much the obtained results depend on each observed unit, a leave-one-out approach can be followed. At this aim, the first (evaluation of the impact factors and additional information) and the second (exploration of interactions among the factors and units) step of the proposed approach are carried out, excluding one university at a time. Figure 8 shows the percentage of variability on the first factorial plane derived from the PCA on the residuals of model (3) obtained by excluding one university at a time (universities are on the horizontal axis and the percentages on the vertical axis. The horizontal line represents the percentage of explained variability obtained on the whole set of units. Universities that most influence the results are included in a rectangle). In addition to the information provided by the score plot, Figure 8 allows us to easily identify

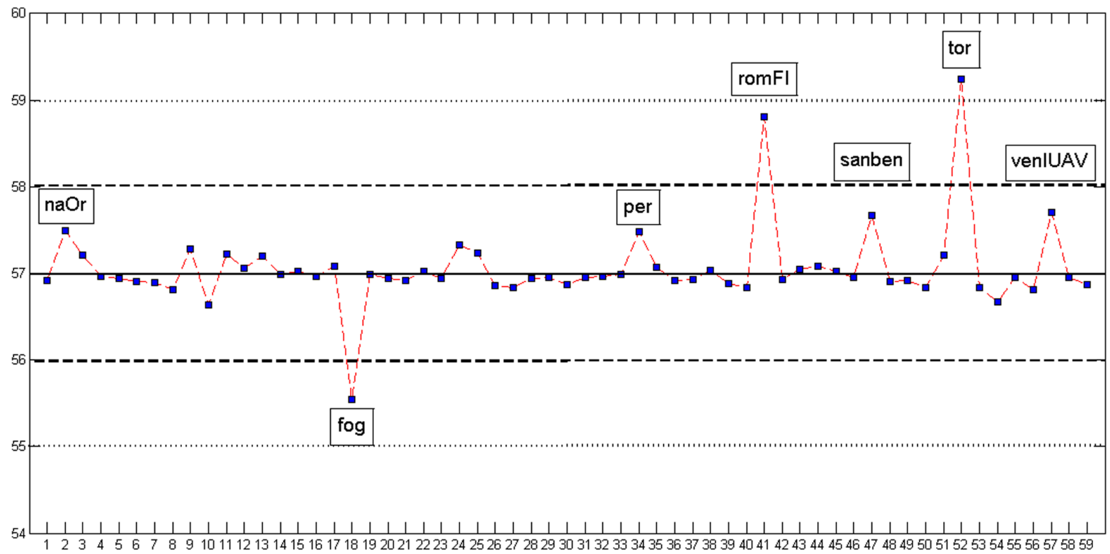

Figure 8. Percentage of variability on the first plane excluding one university at a time.

Table 5. Correlations among the CIs and the second factor from a PCA on the whole sample (All) or excluding one group of universities according to the dimension

CI	All	mega	big	medium	small	polytechnics
MinisGeomb5	**0.763**	**-0.762**	*0.197*	*0.235*	*0.474*	**-0.771**
MinisGeomb4	**0.681**	**-0.713**	*0.366*	**0.737**	**0.965**	**-0.707**
MinisGeom	**0.619**	**-0.657**	*0.588*	*0.374*	**0.937**	**-0.647**
MinMax2Geomb3	0.596	-0.555	-0.295	0.353	-0.157	-0.562
MinisGeomb3	0.560	-0.530	0.398	0.311	0.582	-0.549
MinMax2Geomb5	0.526	-0.507	*-0.740*	0.299	-0.314	-0.509
MinMax2Geom	0.474	-0.501	*-0.661*	0.413	-0.321	-0.462
MinMax2Geomb4	0.376	-0.418	-0.356	0.556	0.111	-0.392
MinisLinb5	0.140	-0.123	0.076	-0.384	-0.131	-0.162
MinMax2Linb5	0.018	0.062	*-0.799*	-0.328	*-0.724*	0.022
MinisLinb4	-0.305	0.306	*0.650*	0.254	0.414	0.255
MinMax2Geomb2	-0.342	0.282	-0.497	-0.071	-0.619	0.339
MinMax2Linb4	-0.367	0.454	-0.190	0.435	-0.205	0.365
MinMax2Linb3	-0.368	0.495	-0.289	-0.295	*-0.712*	0.411
MinisGeomb2	-0.383	0.294	*0.681*	-0.154	0.447	0.337
MinisLinb3	-0.411	0.510	0.462	-0.344	-0.101	0.446
MinMax2Lin	-0.510	0.599	*-0.634*	-0.327	*-0.921*	0.551
MinisLin	**-0.600**	**0.582**	*0.822*	*-0.532*	*0.043*	**0.561**
MinisLinb2	**-0.681**	**0.636**	**0.612**	*-0.573*	*-0.059*	**0.638**
MinMax2Linb2	**-0.750**	**0.790**	*-0.434*	**-0.577**	*-0.832*	**0.777**

if the impact played by each university on the PCA results is positive (excluding the unit, the percentage decreases), null (excluding the unit, the percentage does not vary) or negative (excluding the unit, the percentage increases). The obtained results can be considered almost stable with respect to the units because in the 91.5% of the cases, excluding one unit reduces or increases the percentage of variability by just one percentage point and in 95% of the cases, by two percentage points.

The stability is confirmed if the unit residuals obtained by the leave-one-out procedure are projected onto the score plot as additional points called *supplementary*. If the position of a given university on the plane is very sensitive to the removal of a given unit from the analysis, its position on the plane should change considerably if the given university is removed. It is obvious that units with high coordinates on one or both axes are those expected to be the most influencing. For the sake of brevity, the results for just three units (*sanben*, *cato*, *mac*) are provided. They can be considered sample units because *sanben* and *cato* are very influencing units (high coordinates on both axes) while *mac* is quite close to the origin. To improve the readability of

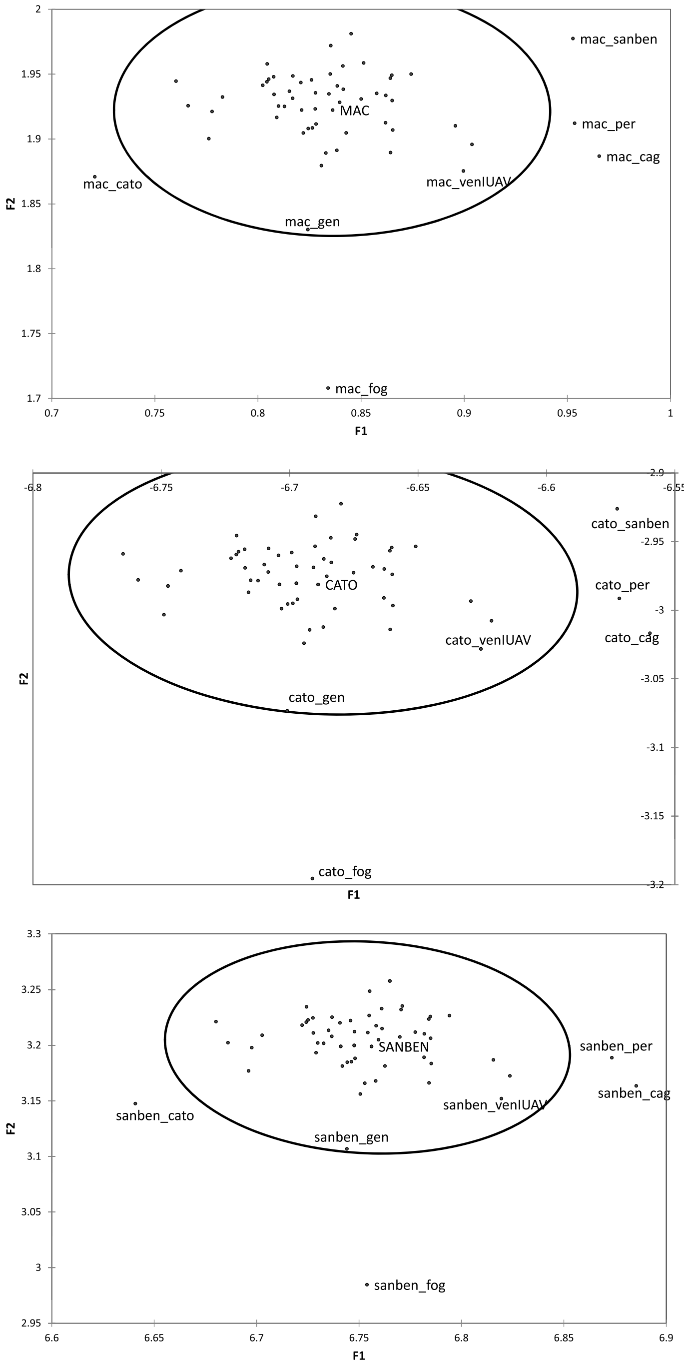

Figure 9. Score plot for cag, cato, mac with their corresponding supplementary points.

the results and to visualize the variability of the results derived from the leave-one-out procedure, confidence ellipses are built containing 95% of the sampled data [17]. The results in Figure 9 show that the position of each of the three units on the original score plot (labeled with capital letters in the figure) does not vary considerably, even if one unit at a time is excluded.

It is well debated that a simultaneous analysis of all universities can be misleading because of the role played by the big universities. The first and second steps of the proposed approach can be carried out by excluding a group of universities at a time according to the *dimension*. The procedure is repeated five times, every time excluding one group of universities (mega, big, medium, small, polytechnics). As the final results provided by the five PCAs do not differ much for the first factor, Table 5 only shows the correlations among the CIs and the second factor (high correlations coherent with or contradictory to the results on the whole sample are highlighted, respectively, in bold or italics). It is worth noting that the greatest differences in results on the whole sample emerge when the group of big or medium universities is excluded from the analysis. This can be considered an interesting result because, once the general central tendency is removed through the use of the residuals, the prevailing typical role of very big universities defaults.

4. CONCLUDING REMARKS AND FURTHER DEVELOPMENTS

The proposed approach aimed to investigate the impact of the different sources of uncertainty in CI construction, both in terms of factors and levels. Moreover, external information was also taken into account and the effect of units (e.g. universities) and their interactions with the factors were evaluated. All such factors were simultaneously analysed through a multidimensional approach combining confirmatory and exploratory methods. Graphical potentialities of multivariate methods and the proposed additional tools for the interpretation can be considered a support for analysts and politicians as they can easily verify the effects of a given policy adopted to construct a CI. Also, the computational capability of the proposed approach guarantees its use in the case of many observations, where classical approaches require an individual inspection of the factors and units. Finally, the stability of the results according to the role played by each unit and dimension was also investigated through resampling methods.

Alternative approaches could also be performed such as the use of linear model with interactions between the university factor and the uncertainty factors but they would provide results very difficult to interpret (the interaction plot in presence of a large number of units becomes unreadable) while results from PCA are graphically oriented and easy to interpret. Moreover, the use of a multidimensional method like PCA allows to visualize individual differences for different university groups in their response to the uncertainty factors, allowing a more in-depth exploration of the interactions.

The methodology proposed by the Italian Ministry of Education to construct CIs is one possible solution among several reasonable alternatives. Notwithstanding their potential, CIs remain a subject of controversy because they lack a standard construction methodology [7].

Further developments of the proposed approach could include the recourse to more complex designs or to complex CIs where the original indicators are grouped in subdimensions.

APPENDIX

Table 6 shows the universities labels used in the plots. Figures 10–13 show the boxplots of each indicator with respect to the university dimension.

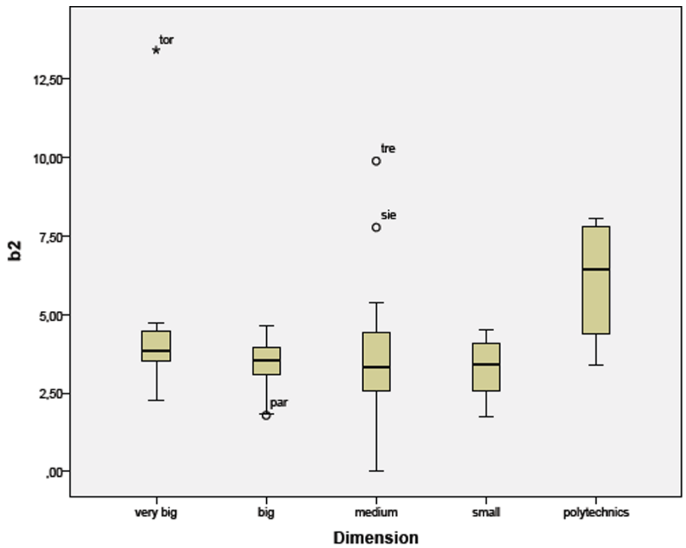

Figure 10. Boxplot of the $b2$ indicator.

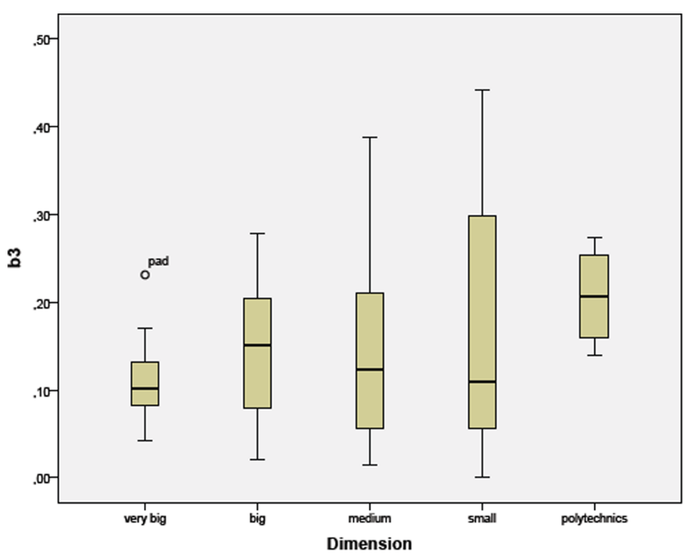

Figure 11. Boxplot of the $b3$ indicator.

Table 6. University Labels

University	Label	Dimension
Ca' Foscari of VENEZIA	ven	3
L'Orientale of NAPOLI	naOr	4
Parthenope of NAPOLI	naPa	3
BARI	bar	1
BASILICATA	BAS	4
BERGAMO	ber	3
BOLOGNA	bol	1
BRESCIA	bre	3
CAGLIARI	cag	2
CAMERINO	cam	4
CASSINO	cas	3
CATANIA	cata	1
CATANZARO	cato	3
CHIETI-PESCARA	chpe	2
CALABRIA	CAL	2
FERRARA	fer	3
FIRENZE	fir	1
FOGGIA	fog	3
GENOVA	gen	2
INSUBRIA	ins	4
L'AQUILA	laq	3
MACERATA	mac	3
Mediterranea of REGGIO CALABRIA	regcal	3
MESSINA	mes	2
MILANO	mil	1
MILANO-BICOCCA	milBic	2
MODENA and REGGIO EMILIA	morem	3
MOLISE	MOL	4
Federico II of NAPOLI	naFe	1
PADOVA	pad	1
PALERMO	pal	1
PARMA	par	2
PAVIA	pav	2
PERUGIA	per	2
PIEMONTE ORIENTALE	PieOr	4
PISA	pis	1
Polytechnic of MARCHE	polMar	5
Polytechnic of BARI	polBar	5
Polytechnic of MILANO	polMI	5
Polytechnic of TORINO	polTor	5
ROMA *Foro Italico*	romFI	4
ROMA *La Sapienza*	romSap	1
ROMA *Tor Vergata*	romTor	2
ROMA TRE	romTre	2
SALENTO	SAL	2
SALERNO	sal	2
SANNIO of BENEVENTO	sanben	4
SASSARI	sas	3
Second University of NAPOLI	naSUN	2
SIENA	sie	3
TERAMO	ter	4
TORINO	tor	1
TRENTO	tre	3
TRIESTE	tri	3
TUSCIA	TUSC	3
UDINE	udi	3
IUAV of VENEZIA	venIUAV	4
URBINO *Carlo BO*	urbBo	3
VERONA	ver	2

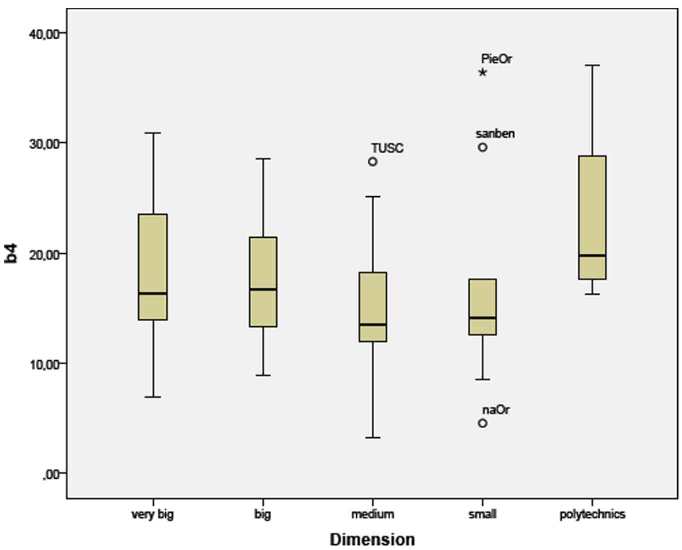

Figure 12. Boxplot of the $b4$ indicator.

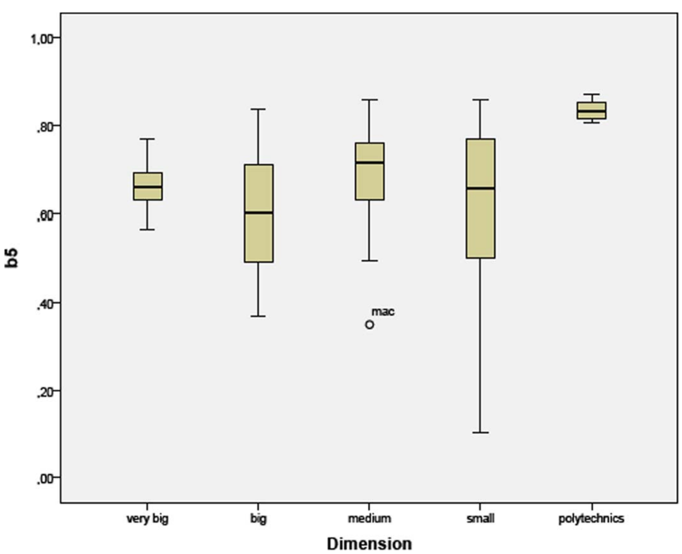

Figure 13. Boxplot of the $b5$ indicator.

Received 19 September 2014

REFERENCES

[1] Adelman I., Morris C. T. (1972) The measurement of institutional characteristics of nations: Methodological considerations. J Dev Stud 8: 111–135.

[2] Booysen F. (2002) An overview and evaluation of composite indicators of development. Soc Indic Res 59 (2): 115–151.

[3] Davino C., Palumbo F., Vistocco D. (2011) Analyzing scientific production through redundancy analysis: The case of the Italian university system. In: Ingrassia, S., Rocci, R., Vichi, M. (eds.) New Perspectives in Statistical Modeling and Data Analysis. Springer, Heidelberg, pp 29–38. MR3075358

[4] Davino C., Romano R. (2011) Sensitivity analysis of composite indicators through Mixed Model Anova. Working Paper n.32. University of Macerata. ISSN 1971–890X.

[5] Davino C., Romano R. (2013) Assessing different scales in subjective measurements. In: Davino, C., Fabbris, F. (eds.) Survey Data Collection and Integration. Springer, Heidelberg, pp 45–59. MR3053015

[6] Davino C., Romano R. (2014) Assessment of composite indicators using the ANOVA model combined with multivariate methods. Soc Indic Res 119 (2): 627–646.

[7] Drewnowski J. (1972) Social indicators and welfare measurement: Remarks on methodology. J. Dev. Stud. 8: 77–90.

[8] Hastie T., Tibshirani R., Friedman F. (2009) The Elements of Statistical Learning. Springer. MR2722294

[9] Jacobs R., Goddard M., Smith P. C. (2005) How robust are hospital ranks based on composite performance measures? Med Care 43: 1177–1184.

[10] Lebart L., Morineau A., Piron, M. (1997) Statistique Exploratoire Multidimensionnelle. Dunod, Paris.

[11] Mardia K. V., Kent J. T., Bibby J. M. (1978) Multivariate Analysis. Academic Press, London. MR0560319

[12] Morineau A. (1984) Note sur la caractrisation statistique d'une classe et les valeurs-test. In: Bulletin Technique du Centre de Statistique et Informatique appliques 2, pp 1–2.

[13] Naes T., Lengard V., Bolling Johansen S., Hersleth M. (2010) Alternative methods for combining design variables and consumer acceptance with information about attitudes and demographics in conjoint analysis. In: Food Qual Prefer. 21: 368–378.

[14] RAE (2008) Research assessment exercise: The outcome. Retrieved December. Ref RAE 01/2008. http://www.rae.ac.uk/results/outstore/RAEOutcomeFull.pdf.

[15] Roberts G. (2003) Review of research assessment: Report by Sir Gareth Roberts to the UK funding bodies issued for consultation May 2003. Ref 2003/22, London Higher Education Funding Councils.

[16] Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Saisana M., Tarantola S. (2008) Global Sensitivity Analysis. The Primer. John Wiley & Sons. MR2382923

[17] Saporta G. (1990) Probabilités Analyse des Données et Statistique. Technip, Paris.

Rosaria Romano
University of Calabria
Department of Economics, Statistics and Finance
Italy
E-mail address: rosaria.romano@unical.it

Cristina Davino
University of Macerata
Department of Political Sciences, Communications and International Relations
Italy
E-mail address: cristina.davino@unimc.it