

# Assessment of Composite Indicators Using the ANOVA Model Combined with Multivariate Methods

Cristina Davino · Rosaria Romano

Accepted: 2 December 2013 / Published online: 14 December 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** The synthesis of a number of single observed indicators into a unique composite indicator involves various subjective choices related, for instance, to the type of combination (linear, non-linear) and to the aggregation method (simple average, geometric average) used in its construction. Thus, it is clearly important to analyse the variability of a composite indicator according to all possible alternatives before employing it in any decision-making process. Within such a framework, in this paper, we present a new approach based on a combination of explorative and confirmative analyses aiming to investigate the impact of different subjective choices on the variability of composite indicators. This new approach also allows the analysis of the related individual differences among the statistical units and the use of external information on the same units to enhance the interpretation of the final results.

**Keywords** Composite indicators · Stability · Analysis of variance · Principal component analysis

## 1 Introduction

The requirement for the synthesis of univariate indicators by means of composite indicators (CIs) is becoming increasingly urgent in all contexts where the object of analysis cannot be directly observed and measured due to the presence of several different concurrent factors acting as determinants.

---

C. Davino  
University of Macerata, piazza Oberdan 3, 62100 Macerata, Italy  
e-mail: cristina.davino@unimc.it

R. Romano (✉)  
University of Calabria, Arcavacata di Rende, 87036 Cosenza, Italy  
e-mail: rosaria.romano@unical.it

Listed below are several examples of CIs proposed by international organisations in different fields, e.g. social, environmental and economic:

- *Human Development Index*:<sup>1</sup> takes into account three main dimensions of human development (life expectancy, education, income).
- *Economic Competitiveness Index*:<sup>2</sup> measures the ability of a nation to guarantee favorable economic conditions for firm competitiveness.
- *European Satisfaction Index*:<sup>3</sup> considers customer satisfaction based on four factors (image, customer expectation, perceived quality, perceived value) and generates two consequences (loyalty and complaints).
- *Environmental Sustainability Index*:<sup>4</sup> is a composite index tracking socio-economic, environmental and institutional indicators that characterise and influence environmental sustainability at the national scale.

These and other examples of CI. have given rise to increasing interest among politicians, workers from different socio-economic sectors, researchers, news agencies and the general public. It is a matter of fact that CIs are recognised as fundamental tools according to which important political decisions, often aiming to share financial resources, are made. They are also widely used to communicate the relative performance of countries.

Whatever the applicative context, the construction of CIs involves stages where several decisions have to be taken (Cherchye et al. 2007). The first requirement pertains to the characterisation of the dimensions underlying the concept to be measured. Once these have been identified, the quantitative and qualitative variables (observable indicators) able to measure each dimension must be specified. A pre-processing of the univariate indicators is then performed to deal, for example, with missing values, coding and scale transformation of raw values. Finally, several aggregation methods and systems of weights can be adopted.

The scientific literature offers many contributes aiming to deal with each phase of the construction of CIs:

- **Problem definition**  
The researcher decides to refer to a specific meaning of the concept to be analysed thus adopting a personal conceptual definition. That means he or she establishes a personal theoretical framework where the concept to be measured is contextualised. The number and nature of the components (dimensions and sub-dimensions) that will decompose the CI depend on this definition (McGranahan et al. 1972).
- **Data selection**  
The choice of the set of indicators depends on the theory defined in the problem phase but also on considerations provided by experts and politicians.
- **Data pre-processing**  
Data reduction statistical techniques can help to refine a defined set of indicators. Moreover, the complexity of the observed phenomenon leads to the use of heterogeneous indicators (both qualitative and quantitative or expressed in different measurement units) that cannot be directly aggregated into the CI. Several transformations can be adopted (e.g. centering, z-scores, normalisation within the bounds of a

<sup>1</sup> <http://hdr.undp.org/en/statistics/hdi/>.

<sup>2</sup> <http://www.cforic.org/pages/european-competitiveness.php>.

<sup>3</sup> [http://relationmonitor.dk/uk/analysis/about\\_our\\_research/effektprofil\\_analysis/european\\_customer\\_satisfaction\\_index\\_ecsi](http://relationmonitor.dk/uk/analysis/about_our_research/effektprofil_analysis/european_customer_satisfaction_index_ecsi).

<sup>4</sup> <http://envirocenter.yale.edu/?page=environmental-sustainability-index>.

given range, percentages of the total population, ranks, logarithms) according to the variable nature and to the next aggregation method (Drewnowski 1972; McGranahan et al. 1972; Nardo et al. 2005).

- CI construction and analysis

The choice of the method used to aggregate the set of indicators mainly depends on the theory defined in the problem phase and on the relationship hypothesised among the indicators (linear or non-linear). Whatever the selected aggregation method, the analysis phase also involves another crucial decision concerning the system of weights to be adopted. The simplest and most common strategy consists in giving the same role to each indicator in the CI construction, but consultation with experts (Moldan et al. 1997; Saaty 1987) or the use of multivariate techniques (Lebart et al. 1984) can suggest completely different solutions.

Alternative schedules to the previously described steps have also been proposed; for example McGranahan et al. (1972) and Booyesen (2002) consider four steps (selection, scaling, weighting and aggregation), while Nardo et al. (2005) provide a long checklist for building CIs.

As the values of a CI depend on the subjective choices adopted for its construction, it is advisable to evaluate if and how each phase of the CI construction influences the final results. The aim of this paper is to develop a new approach for assessing CIs, thus providing decision makers with information on the impact derived from alternative CI construction strategies. The proposed approach does not aim for an automatic selection of the best CI, since each choice could be the optimal choice for some units and the worst for some others.

Several contributions have been proposed in the literature to assess if each methodological choice affects CI variability. For example, some of them are based on the comparison among the unit rankings derived from alternative strategies of analysis (Jacobs et al. 2005). The rationale behind these methodologies is that one of the main purposes of CIs is to allow the comparison of several units (e.g. countries); however, comparing rankings does not allow us to highlight the sources of potential instability.

Other authors (Adelman and Morris 1972) propose the use of item analysis to compare CIs to each single indicator or to an external evaluator not included in the analysis.

Nardo et al. (2005) introduce methods able to measure the uncertainty associated with the CI in terms of its variability and CI sensitivity in terms of the contribution of each factor involved in the CI construction on its variability. For example, variance-based methods represent a widespread class of methods used to perform sensitivity analyses. They provide, for each uncertainty factor (namely each necessary choice to construct a CI), a measure of the impact of the given factor on the CI. A technical drawback of variance-based methods is their computational cost since they require many CI simulations. Moreover, these methods provide information on the different uncertainty factors without highlighting the role of the corresponding alternatives.

The main limits of the current state of the art in the assessment of CIs are related to the univariate analysis of each determinant of the CI construction without considering the interactions among them. This is also the reason why it is difficult to make a comparison between these methods and the approach proposed in this paper, which instead is embedded in the theoretical framework of multivariate analysis both explorative and confirmative.

The present contribution is based on a combination of analysis of variance (ANOVA) models (McCulloch and Searle 2001) and principal component analysis (PCA) (Mardia

et al. 1979). This strategy was previously proposed by Naes (Endrizzi et al. 2011; Naes et al. 2010) in the context of consumers' preferences and it was adapted by Davino and Romano (2011) for CI analysis and for the subjective measurement framework (Davino and Romano 2013). The aim of the present work is to properly extend such an approach to the CI framework in order to assess if each methodological choice affects CI variability and to evaluate the differences and relationships among such alternative CIs. Moreover, a study of the stability of the multivariate results is also provided through the use of resampling methods (Efron 1982).

The combination of confirmative and explorative methods allows the introduction of an innovative solution for CI analysis based on a multivariate approach where individual differences among the observations are highlighted, interactions among alternative choices in the construction of CIs are taken into account and the interpretation of the results are enhanced by graphical tools.

The proposed approach is described by means of a case study based on the *Quality of Life Index* proposed by the Italian national daily business newspaper *Il Sole 24 Ore*.<sup>5</sup>

## 2 The Proposed Approach: ANOVA-PCA Based Method

The main focus of this paper is to propose an alternative method for assessing CIs. This method investigates the impact of the different sources of variability occurring in the CI construction, also taking into account the external information available for each statistical unit (e.g. continent, dimension, etc.). The use of external information is critically important in this type of analysis since it provides additional useful information for accurate interpretation. This paper is embedded in a multivariate framework, combining confirmative (ANOVA) and explorative (PCA) methods.

Following the typical terminology of ANOVA, each issue to be defined to construct a CI is called a factor (i.e. dimension selection, variable selection, unit selection, data transformation, weighting method, aggregation method) and its possible alternatives are called levels (i.e. linear and geometric for aggregation method, etc.). The overall combination of factors and levels produces the so-called *design matrix*, which shows the number and the type of generated CIs.

The proposed strategy consists of two main steps:

1. evaluation of the impact of factors and additional information by ANOVA;
2. exploration of interactions among factors and units by PCA.

The first step allows the investigation of the general tendencies of the different factors on the overall set of units. The second step exploits the potentialities of the PCA to explore the effect of each factor on each single statistical unit after removing the general tendency. This latter stage is graphically oriented and very flexible with respect to the number of units in the data set. The two steps are sequentially connected because further analysis of the variability of the CIs among the units is advised only if results from the ANOVA in the first step highlight interactions among the units and main factors. The proposed strategy is enhanced with the introduction of additional tools for the interpretation of the results and the analysis of their stability.

<sup>5</sup> <http://www.ilssole24ore.com/english-version/front-page.shtml>.

## 2.1 Evaluation of the Impact of Factors and Additional Information by ANOVA

ANOVA is a useful method when the objective is an assessment of the impact of certain controllable factors (categorical variables) on a specific response (continuous variable). The impact is significant if the variability between the groups defined by the factor levels (categories) is much larger than the variability within the groups. An ANOVA model is equivalent to a linear model where the response variable becomes the dependent variable, and each of the factors is transformed into a dummy variable according to the number of levels.

Given the data matrix  $\mathbf{X}$  ( $N \times P$ ) of  $P$  indicators observed on  $N$  units, let us consider for simplicity only two factors  $Z'$  and  $Z''$ , respectively, with  $I$  and  $J$  levels. The *units* factor  $U$  will consist of as many levels as the number of observed units, and it is nested in the *external information* factor  $\delta$  with  $M$  levels. 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. The *full* ANOVA model can then be written as:

$$y_{ijmn} = \mu + z'_i + z''_j + \delta_m + u_n(\delta) + z'z''_{ij} + z'u_{in} + z''u_{jn} + e_{ijmn}, \quad (1)$$

where  $y_{ijmn}$  is the  $n$ th observation obtained using the  $i$ th ( $i = 1, \dots, I$ ) level of the  $z'$  factor, the  $j$ th ( $j = 1, \dots, J$ ) level of the  $z''$  factor and the  $m$ th category of the external factor. In model (1), the general mean is represented by  $\mu$ , while  $z'_i$  and  $z''_j$  are the main effects of the two factors and  $z'z''_{ij}$  is their interaction effect. The main effect of the factor represented by the units and nested in the external information factor  $\delta$  is  $u_n$ , while  $z'u_{in}$  and  $z''u_{jn}$  are the interactions between units and the two main factors. Finally, the term  $e_{ijmn}$  is the random error. In the rest of the paper the factors  $z'$  and  $z''$  are called *main factors* as opposed to the unit factor.

Model (1) corresponds to a simultaneous ANOVA for all statistical units and it is estimated by stacking in a pile the matrix containing the different combinations of factor levels and the corresponding CIs obtained for each statistical unit. This type of modeling data permits the estimation of all relevant main effects and interactions and then models the individual differences among the units as a sum of an additive unit effect and various interactions between the units and the main factors.

Results from model (1) show which factors strongly affect or do not affect the stability of the CI and also the impact of these effects on each single unit. In order to better explore such differences and similarities among the units, a PCA exploiting all the advantages of the factorial methods is performed on the residuals of an ANOVA model without the interactions between the units and the uncertainty factors, as shown in the next sub-section.

## 2.2 Exploration of Interactions Among Factors and Units by PCA

A study of the differences among the units in their CI values has already been achieved by introducing the *units* effect as an additive factor in the ANOVA (model 1). In fact, the main effects and the interaction effects plots graphically show such differences. However, in the case of many statistical units, these plots are unreadable, and more sophisticated explorative methods are required. PCA is appropriate in this context since it allows the synthesis of multivariate data in a few linear combinations to be plotted by means of factorial planes. Specifically, PCA will be used to explore individual differences of the residuals obtained in a *reduced* ANOVA model with only main factors and the unit effect:

$$y_{ijmn} = \mu + z'_i + z''_j + z'z''_{ij} + \delta_m + u_n(\delta_m) + e_{ijmn}. \quad (2)$$

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 ( $N \times (I \times J)$ ) 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 containing all main factors and their interactions. Specifically, the matrix is double-centred, that is, both its rows and columns sum up to 0. The effect of mean centring for each row means that the additive differences between units have been eliminated. On the other hand, the effect of centring the residual data for each column is that for each combination of levels and factors (CI), the values represent the distances of the units to the *average unit* for that CI. Those units having a positive residual value for a CI have a score on that CI higher than that of the *average unit* 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 planes obtained by the residuals from model (2).

### 3 Case Study

The dataset used to describe the proposed approach is based on a survey proposed every year by the business newspaper *Il Sole 24 Ore* on the *Quality of Life* in the Italian provinces. A total of 36 indicators were observed in 107 provinces in 2010.<sup>6</sup> The indicators aim to describe six main aspects of the complex concept *Quality of Life*, each measured by the same number of indicators. The different aspects with their related indicators are listed below:

- standard of living:  
savings, retirement, inflation, GNP, housing, spending
- business and work:  
bankruptcies, firms, inscriptions-cancellations, protests, youth occupations, female occupations
- environment and services:  
nurseries, infrastructures, public health, climate, ecology, justice
- criminality:  
burglaries, automobile theft, extortions, robberies, fraud, murders
- population:  
density, foreigners, graduates, births, senior citizens, divorces
- spare time:  
books, restaurants, shows, sports, volunteer work, cinema.

Provinces can be classified according to four *Areas*: (1) north-west, (2) north-east, (3) center, (4) south and islands. Information from these areas will be used as *external information* in modeling data.

Results from the PCA on the standardised raw data highlight the similarities and differences among the provinces with respect to the original indicators. Factorial planes in

<sup>6</sup> [http://www.ilssole24ore.com/speciali/qvita\\_2010/home.shtml](http://www.ilssole24ore.com/speciali/qvita_2010/home.shtml).

Figs. 1 and 2<sup>7</sup> explain 46% of the total variation. The first axis is related to both *economic* and *social well-being*. On the positive direction it discriminates between provinces with cultural and sportive activities, high levels of spending and savings, high percentages of young and female employees but also a large number of divorces. On the negative direction, there are provinces behaving in the opposite way and most of them belong to the south area. The second axis is related to the *liveableness*; large provinces (e.g. NA, RM, MI) are all situated on the positive direction. They have in common the typical problems of any big city-high density. However, the large provinces in the south are mostly characterised by high levels of cars stealing, bankrupts and protests, whereas the provinces situated in other areas present high levels of infrastructures, robberies, books consumption and high houses prices.

The final aim of the survey is to measure the *Quality of Life* by synthesising the 36 observed indicators into a unique CI and rank the 107 Italian provinces accordingly. However, as discussed in Sect. 1, there are many subjective choices in the construction of a CI, thus requiring an appropriate analysis of its variability.

### 3.1 Assessing the *Quality of Life* Index

The Italian newspaper *Il Sole 24 Ore* proposes to construct a CI for measuring the *Quality of Life* by a simple linear aggregation of the original 36 indicators transformed into ordinal variables. In particular, for each indicator, a score equal to 1000 is assigned to the unit with the best value; the remaining units are rescaled proportionally.

It is a matter of fact that several alternatives to the one proposed by the Italian newspaper are possible. The first step for assessing the *Quality of Life* index is the selection of the subjective choices to take into account in the study. Three main factors are considered in the present analysis.

The first factor is related to the normalisation method, and three different alternatives are considered:

- index numbers: each value is divided by a reference value, e.g. the average overall units;
- standardised values: each value is divided by the standard deviation after subtraction of the average;
- scores: proposed by *Il Sole 24 Ore*.

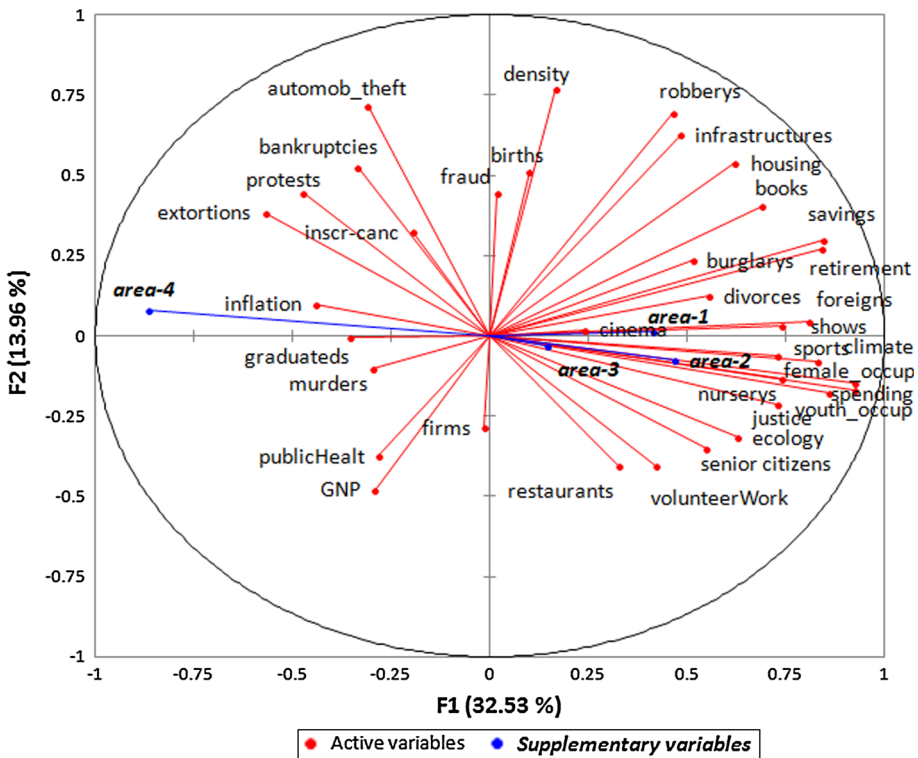
The second factor corresponds to the set of weights to be used in the linear combination of the 36 indicators to calculate the CIs. The two considered levels are:

- equal weights: the coefficients of the linear combinations are all equal to 1;
- PCA weights: the coefficients of the linear combinations are the components of the first eigenvector from the PCA on the standardised indicators.

The third factor regards the possibility of including all 36 indicators in the CI construction or excluding each of the six dimensions, one at time. Such a factor aims to evaluate the impact of the variable and dimension selection on the CI.

The combination among the levels of the three factors produces a design matrix with 42 different CIs. The factors with their related levels are listed in Table 1; labels used in the following result figures and tables are in brackets.

<sup>7</sup> Labels used for provinces are described in the “Appendix”.



**Fig. 1** First factorial plane of variables from the PCA on the *Il Sole 24 Ore* indicators. (Color figure online)

Before implementing the ANOVA-PCA approach introduced in Sect. 2, all 42 CIs are rescaled to the range  $[0, 1]$  in order to make reasonable comparisons among them.

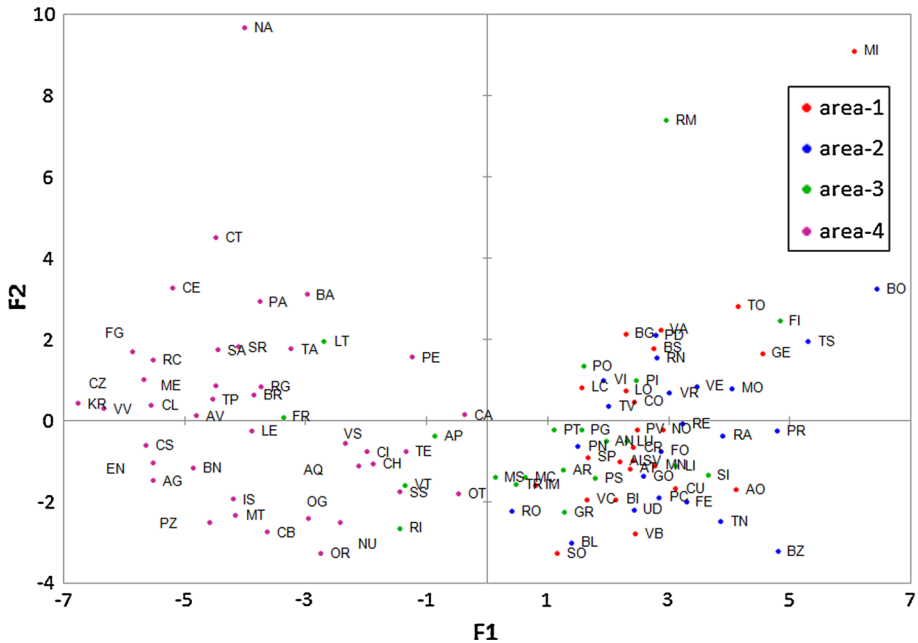
The assessment of the CIs obtained using the different level combinations is univariately explored by means of a boxplot for each province, together with information on the minimum and maximum ranks. For brevity, results for only *Area 1* are reported in Fig. 3, which shows large variations for both the CIs values and the ranks for every province. A simple univariate analysis, such as that shown in Fig. 3, reveals that for each unit, alternative CIs can provide quite different results both in the CI value and in the overall ranking position.

On the other hand, the univariate analysis provides information only on the variability without highlighting the role of the different factors and the respective levels.

In order to evaluate the effect of the factors on the variability of the CIs, a *full* ANOVA model must be estimated in step 1 (Eq. 1). Results for the *full* ANOVA model with all main factors, external information, unit factor and interactions are presented in Table 2. Some terms are marked by a symbol (^) because they do not have full rank, display zero degrees of freedom and missing  $p$  values. This can happen when there are missing factor combinations, like in nested models, and the model has higher-order terms. However, information on the effect of these factors can be recovered by estimating a simpler model like the *reduced* model introduced in step 2 (Eq. 2), where all terms are estimated (Table 3).

Since the set of units corresponds to the entire population of all Italian provinces, it makes no sense to evaluate the significance of the ANOVA results. The relevant result





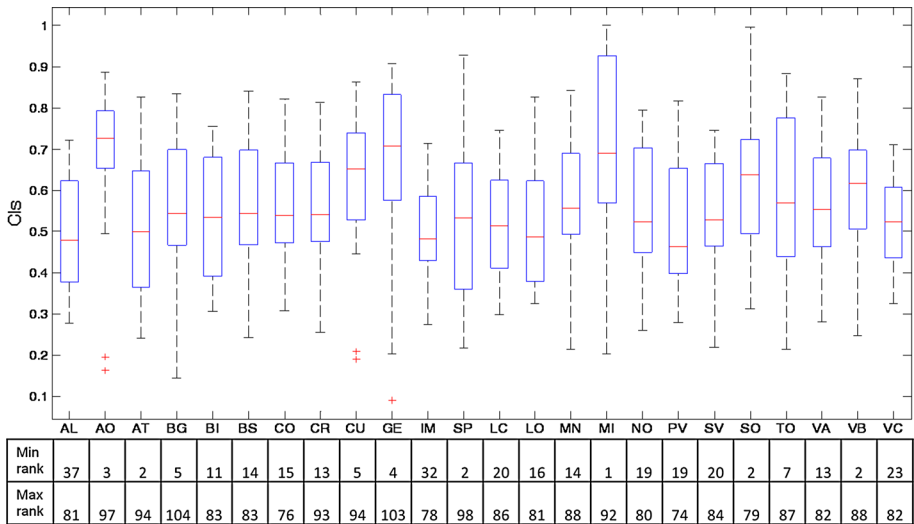
**Fig. 2** First factorial plane of units from the PCA on the *Il Sole 24 Ore* indicators. (Color figure online)

**Table 1** Factors and levels

Factors	Levels
Normalisation (Norm)	index numbers (Ni), standardisation (S), scores by <i>Il Sole 24 Ore</i> (Pun)
Weighting (Weight)	equal (Ew), principal component analysis (ACP)
Inclusion (Inclusion)	all (all), all-d1 (D1), all-d2 (D2), all-d3 (D3), all-d4 (D4), all-d5 (D5), all-d6 (D6)

from the model is the size of the F- test, which provides information on the contribution of the main effects and interactions in explaining the variation of the CIs. As can be seen, all of them provide an interesting contribution but the main factors most responsible for higher variation in the CIs are the *Weight* and the *Area*. The interaction between *Norm* and *Weight* is also important, while the unit factor (*Province*) interacts mostly with the *Weight* factor. This information can be further investigated by Fig. 4, showing the average values of the CIs according to the different levels of the various factors included in the model and Fig. 5 representing the interactions. As mentioned in Sect. 2.2, the graph of the interactions becomes complicated to read when the levels of one of the factors are numerous, as in the case of the factor *Province* which has 107 levels.

The *unit* factor points out individual differences in CI values between provinces and in their behavior with respect to the different choices occurring in the CI construction. This result suggests that the analysis continue by implementing the second step of the proposed strategy, which consists of exploring residuals from the *reduced* model using only the main factors, their interactions and the unit factor.



**Fig. 3** Boxplot of provinces in *Area 1* and minimum and maximum positions in the overall ranking

Residuals from the *reduced* model are computed and arranged into a matrix with the provinces as rows and the different combinations of the three factors as the columns. PCA is then run on these data, including *Area* as a supplementary categorical variable. The related score and loading plots are shown in Figs. 6 and 7. Loadings span the factorial space in all directions, indicating a strong heterogeneity in the Italian provinces according to the different selected combinations of factor levels. Most of the variation (55%), which is explained by the first factor, is related to the choice of the weights used in the aggregation phase. In fact, we see that all CIs with the *PCA weight* method lie on the positive direction of the factor, while those related to the *equal weight* method are positioned on the opposite direction. These results correspond to the results from the previous step, where the *Weight* factor showed a very high F- test. Now the role of its levels are also explored.

Different from the traditional interpretation of PCA results, where the basic aim is generally to identify latent dimensions, in this study, the main objective is to highlight a structure of relations among the CIs and their relations with the statistical units. Specifically, the factorial planes show groups of CIs acting in the same way on different groups of provinces.

The score plot shows which countries are sensitive to the different factor combinations represented in the related loading plot. For instance, the CI used by *Il Sole 24 Ore* (PunEwAll) enhances the position in the ranking of Crotona (KR) and Oristano (OR), which lie in the same zone of the factorial plane. The projection of the modalities of the external variable (*geographic area*) onto the plan spanned by the ANOVA residuals shows that different factor level combinations characterises the four areas.

Due to the explanatory capability and readability of the PCA plots, it is possible to identify the consequences related to the choice of a given CI. For example, a choice to adopt as a system of weights that of *equal weights* (Ew), would benefit the southern provinces since these would assume higher values of the CI.

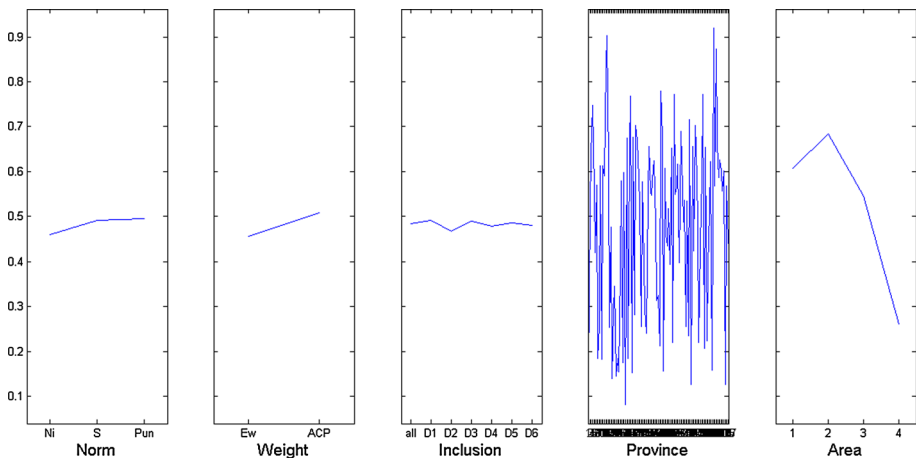
It is possible to extract additional information from factorial planes through the use of some graphical tools that enrich the interpretation of the results. For instance, exploiting

**Table 2** Results from the full ANOVA (Terms marked with ^ are not full rank)

Factors	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
Norm^	0	0	0	0	NaN
Weight^	0	0	0	0	NaN
Inclusion^	0	0	0	0	NaN
Province(Area)	51.199	103	0.4971	120.84	0
Area	139.53	3	46.5099	11306.78	0
Norm*Weight	5.072	2	2.536	616.52	0
Weight*Inclusion	0.681	6	0.1136	27.61	0
Norm*Inclusion	1.063	12	0.0886	21.54	0
Norm*Province	4.411	212	0.0208	5.06	0
Weight*Province	18.059	106	0.1704	41.42	0
Inclusion*Province	3.744	636	0.0059	1.43	0
Error	14.002	3404	0.0041		
Total	242.306	4493			

**Table 3** Results from the reduced ANOVA

Factors	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
Norm	1.107	2	0.5535	59.98	0
Weight	3.19	1	3.1896	345.64	0
Inclusion	0.249	6	0.0415	4.5	0.0002
Province(Area)	51.199	103	0.4971	53.87	0
Area	139.53	3	46.5099	5040.05	0
Norm*Weight	5.072	2	2.536	274.82	0
Weight*Inclusion	0.681	6	0.1136	12.31	0
Norm*Inclusion	1.063	12	0.0886	9.6	0
Error	40.216	4358	0.0092		
Total	242.306	4493			

**Fig. 4** Full ANOVA means plot

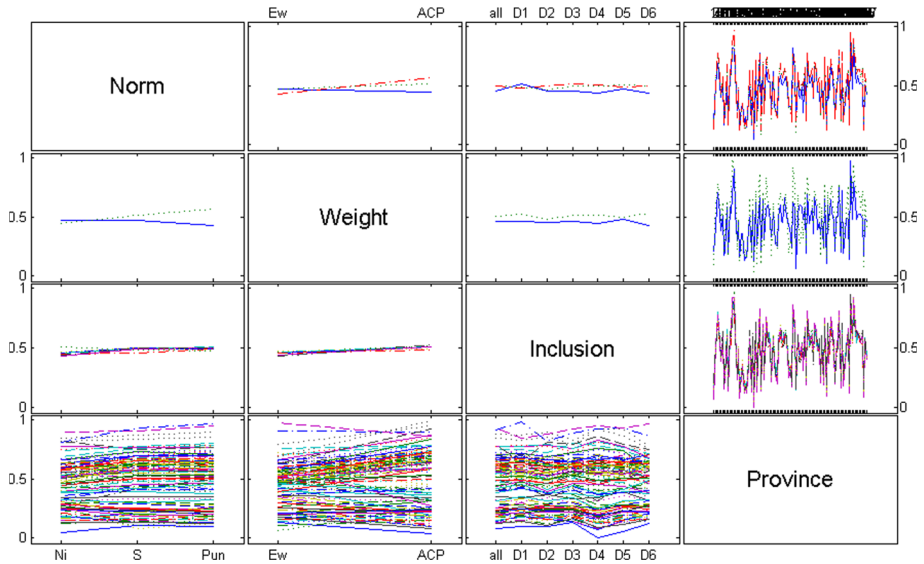


Fig. 5 Full ANOVA interactions plot. (Color figure online)

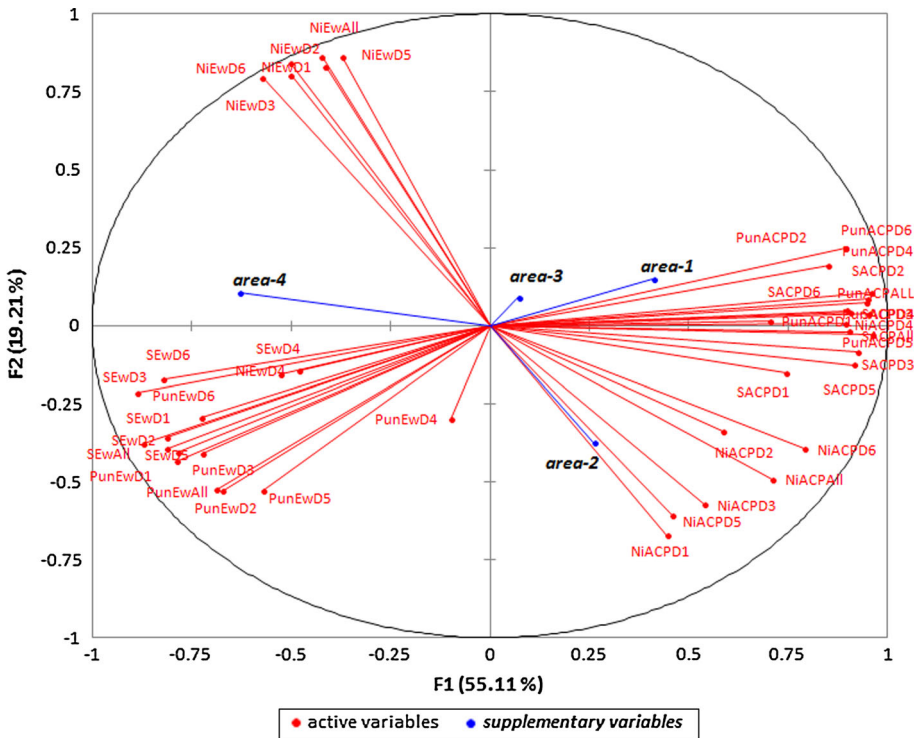
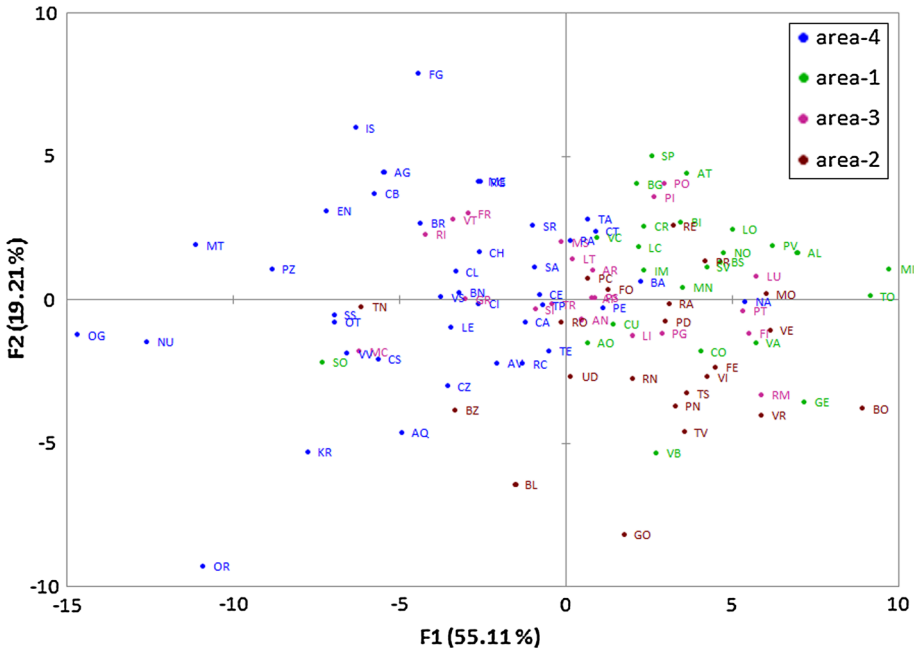


Fig. 6 First factorial loading plot from the PCA on the ANOVA residuals. (Color figure online)



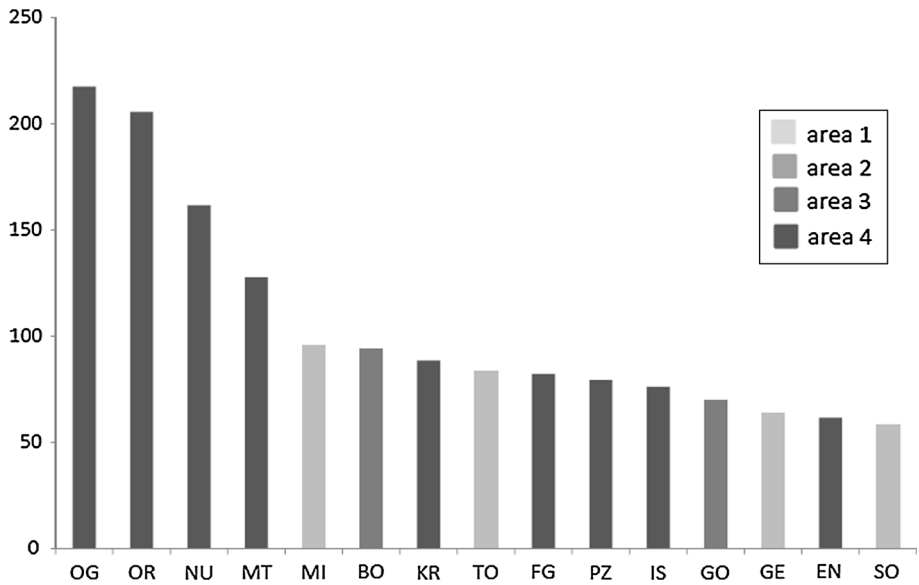
**Fig. 7** Firts factorial score plot from the PCA on the ANOVA residuals. (Color figure online)

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. Figures 8 and 9 respectively show the 15 provinces most and least sensitive to the choice of different CIs. Provinces are colour coded according to the geographical area to which they belong. Besides confirming the role of the most sensitive units with the highest coordinates on the score plot, these figures allow the differentiation of the behaviour of those units also situated on the origin of the axis, which is difficult to read from the score plot.

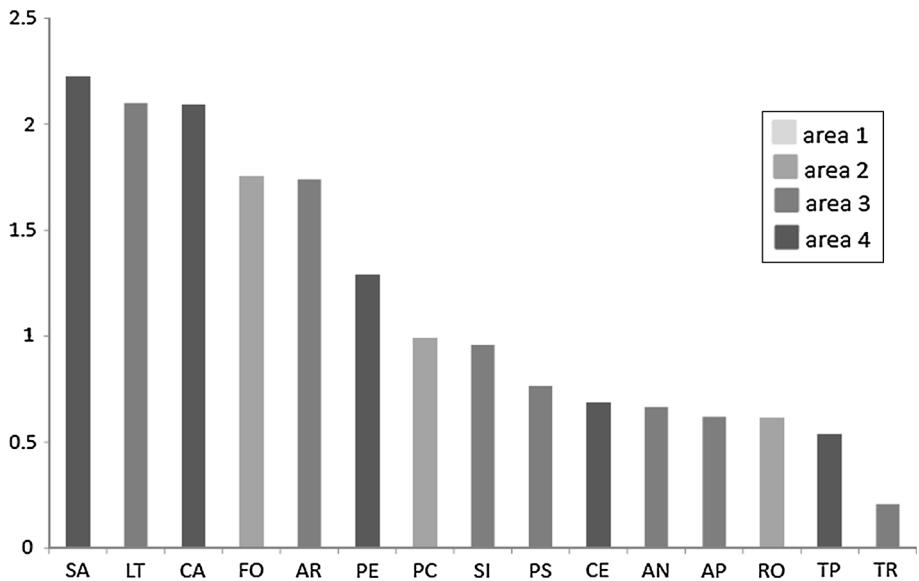
### 3.2 Validation and Stability of the Multivariate Results

A validation of the proposed approach with other methods available in the literature is difficult to perform. The main reason is that these methods are based on univariate statistical analysis, while the innovative approach of the proposal is the multivariate analysis between the possible CIs obtained from the design matrix.

One way to validate the method may be to compare the obtained results with those derived from the original CIs. For example, let us consider the most extreme units on the first factor of the PCA performed on the *reduced* ANOVA residuals: Ogliastra (OG) on the negative direction and Milano (MI) on the positive direction. Loadings in Fig. 6 shows that OG is characterised by high values of all the CIs obtained using a system of equal weights (Ew), while MI is favoured if a weight system based on a PCA of the original indicators is adopted. These results also reveal in Figs. 10 and 11, where for the two considered provinces, the values of the indicators are shown (CIs obtained using a system of equal weights are in blue, while CIs obtained using a system of PCA weights are in red).

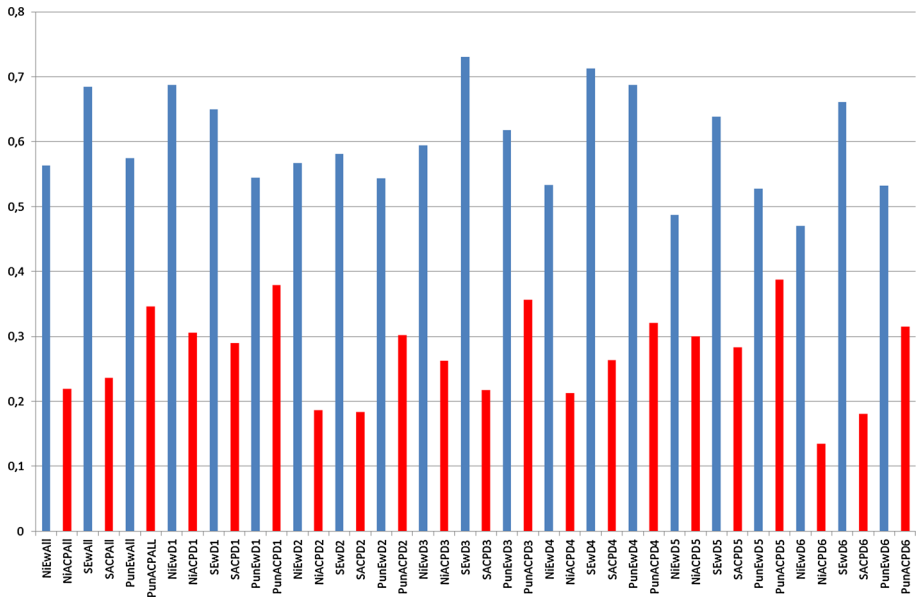


**Fig. 8** Provinces with the highest distance from the origin computed on the first two factors

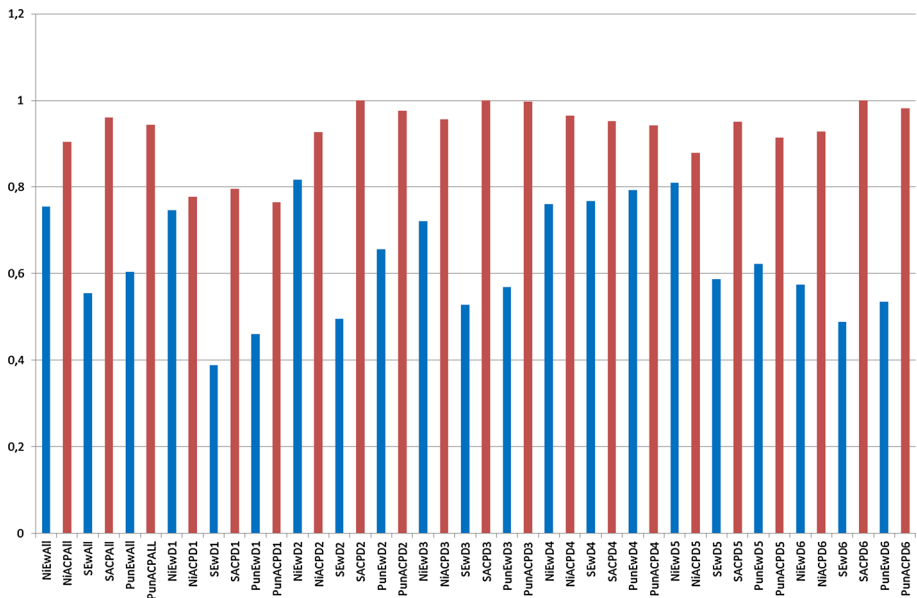


**Fig. 9** Provinces with the lowest distance from the origin computed on the first two factors

The innovative aspect of the method lies in its ability to summarize the information on each individual statistical units, highlighting groups of units with similar behavior with respect to the set of generated CIs. This is possible by considering the structure of multivariate relations among the different CIs.

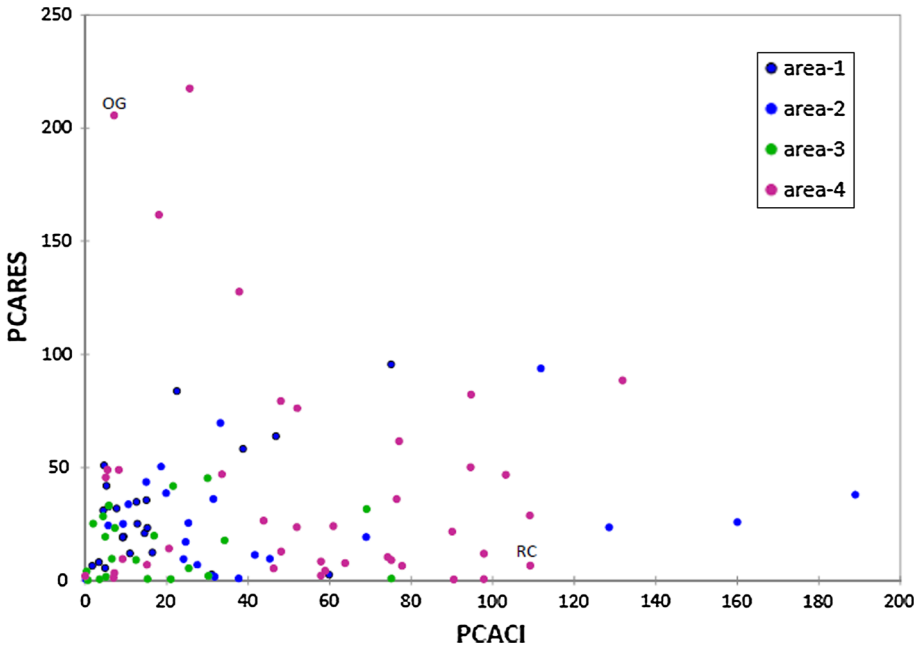


**Fig. 10** CIs of the Ogliastra (OG) province



**Fig. 11** CIs of the Milano (MI) province

A comparison between results from the PCA on the reduced ANOVA residuals and on the raw CIs highlight the innovative aspect of the proposed approach. Specifically, a scatter plot of the sum of squared coordinates on the first factorial plane for each province from the two PCAs is shown in is shown in Fig. 12. Considering two provinces labeled on the



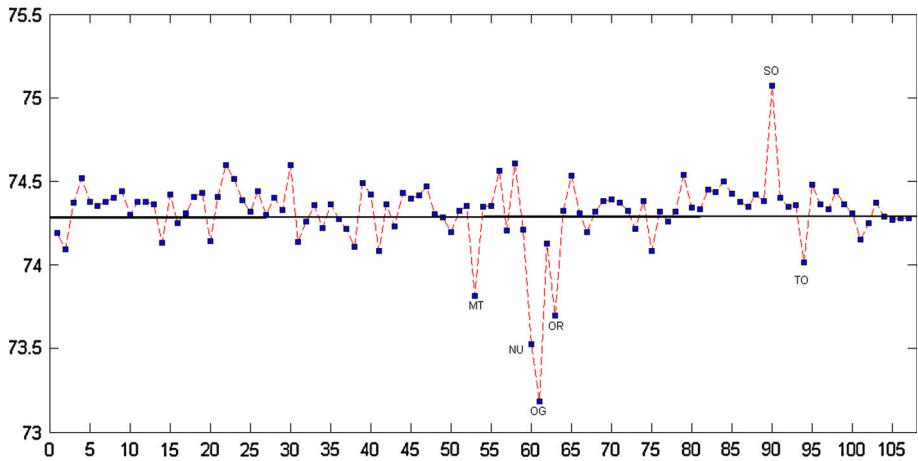
**Fig. 12** Scatter plot of provinces comparing PCA results on reduced ANOVA residuals (PCARES) and raw CIs (PCACI). (Color figure online)

plot, it is possible to show two different situations. The OG province presents a higher value on the PCA from the reduced ANOVA residuals, which means that it is particularly influenced by a set of factor level combinations. Such strong interaction between OG and the factors combination is hidden in the PCA on the raw CIs, where only information on the general tendency is predominant. The RC province, on the other hand, plays an important role in the PCA on the raw CIs, but does not present high variability with respect to the different factor combinations.

Once the impact of the alternative CIs on each province has been evaluated, it is reasonable to wonder about the stability of these results. In the present paper the stability is investigated with respect to the role played by a single province, but it could also consider the role played by a group of provinces.

To assess if and how much the obtained results depend on each observed unit, a leave-one-out approach can be followed. The first (evaluation of the impact factors and additional information) and the second (exploration of interactions among the factors and units) steps of the proposed approach are carried out, excluding one province at a time. Figure 13 shows the percentage of variability on the first factorial plane derived from the PCA on the ANOVA residuals obtained by excluding one province at a time (provinces are on the horizontal axis and the percentages are on the vertical axis). The horizontal line represents the percentage of explained variability obtained on the whole set of units. Labels of the provinces that most influence the results are highlighted). In addition to the information provided by the score plot, Fig. 13 allows us to easily identify if the impact by each province 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





**Fig. 13** Percentage of variability on the first plane excluding one province at a time

percentage increases). The obtained results can be considered almost stable with respect to the units because in most of the cases, excluding one unit reduces or increases the percentage of variability by just one percentage point, and only in a few cases by two percentage points.

## 4 Conclusion

The approach introduced in this paper aims to propose an assessment exercise that should provide decision makers valuable information about the consequences derived from the different subjective choices in the construction of CIs. The final decision, namely which CI to adopt, is up to the decision maker.

The methodological strengths and the advantages in the practical utility of the method can be summarised by the following points:

- External information (the four areas in the real data application): the inclusion of variables not directly involved in the construction of the CIs but related to the phenomenon under investigation clarifies the interpretation of the results.
- Unit factor (provinces in the real data application): including an additional factor into the ANOVA model with as many levels as the number of units allows the evaluation of the impact of the differences among the units on the CI variability.
- Multivariate analysis: a simultaneous analysis of all the CIs derived from the combination of factors and levels can reveal effects hidden by the usual univariate evaluation. Moreover, the proposed approach combines inferential and exploratory methods.
- Interactions: the analysis on the residuals of the ANOVA permits highlighting of the interactions among CIs, factors and levels after general tendencies have been removed.
- Computational issues and graphical potentiality: the method is not affected by the number of units, where classical approaches require an individual inspection of the factors and units. The explicative capability of graphs derived from PCA is also exploited.

- Validation and stability of the results: main results of the proposed method can be validated comparing them with the univariate results obtained on the raw CIs.

**Table 4** Labels of provinces used in plots

Provinces	Labels	Provinces	Labels
Agrigento	AG	Alessandria	AL
Ancona	AN	Aosta	AO
Arezzo	AR	Ascoli Piceno	AP
Asti	AT	Avelino	AV
Bari	BA	Belluno	BL
Benevento	BN	Bergamo	BG
Biella	BI	Bologna	BO
Bolzano	BZ	Brescia	BS
Brindisi	BR	Cagliari	CA
Caltanissetta	CL	Campobasso	CB
Carbonia-Iglesias	CI	Caserta	CE
Catania	CT	Catanzaro	CZ
Chieti	CH	Como	CO
Cosenza	CS	Cremona	CR
Crotone	KR	Cuneo	CU
Enna	EN	Ferrara	FE
Firenze	FI	Foggia	FG
Forlì	FO	Frosinone	FR
Genova	GE	Gorizia	GO
Grosseto	GR	Imperia	IM
Isernia	IS	La Spezia	SP
L'Aquila	AQ	Latina	LT
Lecce	LE	Lecco	LC
Livorno	LI	Lodi	LO
Lucca	LU	Macerata	MC
Mantova	MN	Massa Carrara	MS
Matera	MT	Medio Campidano	VS
Messina	ME	Milano	MI
Modena	MO	Napoli	NA
Novara	NO	Nuoro	NU
Ogliastra	OG	Olbia-Tempio	OT
Oristano	OR	Padova	PD
Palermo	PA	Parma	PR
Pavia	PV	Perugia	PG
Pesaro	PS	Pescara	PE
Piacenza	PC	Pisa	PI
Pistoia	PT	Pordenone	PN
Potenza	PZ	Prato	PO
Ragusa	RG	Ravenna	RA

**Table 4** continued

Provinces	Labels	Provinces	Labels
Reggio Calabria	RC	Reggio Emilia	RE
Rieti	RI	Rimini	RN
Roma	RM	Rovigo	RO
Salerno	SA	Sassari	SS
Savona	SV	Siena	SI
Siracusa	SR	Sondrio	SO
Taranto	TA	Teramo	TE
Terni	TR	Torino	TO
Trapani	TP	Trento	TN
Treviso	TV	Trieste	TS
Udine	UD	Varese	VA
Venezia	VE	Verbania	VB
Vercelli	VC	Verona	VR
Vibo Valentia	VV	Vicenza	VI
Viterbo	VT		

Moreover, a leave-one-out procedure permits to study if and how much the obtained results depend on each unit.

## Appendix

See Table 4.

## References

- Adelman, I., & Morris, C. T. (1972). The measurement of institutional characteristics of nations: Methodological considerations. *Journal of Development Studies*, 8, 111–135.
- Booyens, F. (2002). An overview and evaluation of composite indicators of development. *Social Indicators Research*, 59(2), 115–151.
- Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to benefit of the doubt. *Composite Indicators. Social Indicators Research*, 82, 111–145.
- Davino, C., Romano, R. (2011). Sensitivity analysis of composite indicators through mixed Model Anova. Working Paper no.32. Univeristy of Macerata. ISSN 1971-890X.
- Davino, C., & Romano, R. (2013). Assessing different scales in subjective measurements. In C. Davino & F. Fabbris (Eds.), *Survey data collection and integration* (pp. 45–59). Heidelberg: Springer.
- Drewnowski, J. (1972). Social indicators and welfare measurement: Remarks on methodology. *Journal of Development Studies*, 8, 77–90.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.
- Endrizzi, I., Menichelli, E., Johansen, S.M.B., Olsen, N.V., & Naes, T. (2011). Handling of individual differences in conjoint analysis. *Food Quality and Preference*, 22(3), 241–254.
- Jacobs, R., Goddard, M., & Smith, P.C. (2005). How robust are hospital ranks based on composite performance measures?. *Medical Care*, 43(12), 1177–1184.
- Lebart, L., Morineau, A., & Warwick, K. (1984). *Multivariate descriptive statistical analysis*. New York: Wiley.
- McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. United States of America: Wiley.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

- McGranahan, D.V., Richard-Proust, C., Sovani, N.V., & Subramanian, M. (1972). Contents and measurement of socioeconomic development. A Staff Study of the United Nations Research Institute for Social Development (UNRISD), 3–136, New York: Praeger.
- Moldan, B., Billharz, S., & Matravers, R. (1997). *Sustainability indicators: Report of the project on indicators of sustainable development*. SCOPE 58, Chichester and New York: Wiley.
- Naes, T., Lengard, V., Johansen, S.M.B., & Hersleth, M. (2010). Alternative methods for combining design variables and consumer acceptance with information about attitudes and demographics in conjoint analysis. *Food Quality and Preference*, 21, 368–378.
- Nardo, M., Saisana, M., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. EC Joint Research Centre and OECD Statistics Directorate and the Directorate for Science, Technology and Industry.
- Saaty, R.W. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, 9, 161–176.