Comparing Fuzzy and Multidimensional Methods to Evaluate Well-Being in European Regions

Maria Adele Milioli, Lara Berzieri, and Sergio Zani

Abstract We suggest a new criterion based on fuzzy sets theory in order to evaluate well-being in European regions at NUTS 2 level. With reference to the various domains of this vague and multidimensional concept, a subset of 16 variables available in Eurostat database is selected. After a fuzzy transformation, the variables are aggregated into a fuzzy synthetic indicator, considering different weighting criteria. For each region the fuzzy indicator value, in the range [0, 1], may be interpreted as a membership degree to the subset of the areas with the highest well-being. The results are compared with the ones obtained by principal component analysis (PCA) and k-means cluster analysis applied to the same dataset. Furthermore, the relationships of the fuzzy indicator with GDP per capita and with human development index (HDI) are highlighted. The advantages and the drawbacks of the suggested approach are discussed.

Keywords Cluster analysis • Composite indicators • Fuzzy sets • Membership function • Principal components

1 Introduction

Is increasing GDP per capita a symptom of better life conditions? "Yet Gross Domestic Product measures everything, in short, except that which makes life worthwhile". (Speech excerpt by Robert F. Kennedy, 1968.) The growing interest in the "beyond GDP" ideas has resulted in the construction of several alternative measures of economic development and social progress (e.g. [5]). Well-being and quality of life are the most recurrent terms used to describe these concepts, but in the literature non equivalent definitions and specifications are considered.

M.A. Milioli (✉) • S. Zani
Department of Economics, University of Parma, 43125 Parma PR, Italy
e-mail: mariaadele.milioli@unipr.it; sergio.zani@unipr.it

L. Berzieri
Comune di Parma, 43121 Parma PR, Italy
e-mail: l.berzieri@comune.parma.it

I. Morlini et al. (eds.), *Advances in Statistical Models for Data Analysis*,
Studies in Classification, Data Analysis, and Knowledge Organization,
DOI 10.1007/978-3-319-17377-1_18

"Human development, as an approach, is concerned with what I take to be the basic development idea: namely, advancing the richness of human life, rather than the richness of the economy in which human beings live, which is only a part of it" (Amartya Sen).

Well-being may concern either a single person's life situation (subjective well-being, see, e.g. [9]) or the living conditions of people in a certain area. The two main features of well-being are multidimensionality and vagueness: this latent concept cannot be directly measured, but it can be captured by means of a set of observable variables encompassing different domains. Composite indicators should ideally measure multidimensional concepts which cannot be captured by a simple variable [13]. Furthermore, it is possible to point out the gradual transition from poor to rich living conditions, considering increasing levels of well-being. The measures of well-being should be obtained using multidimensional analysis and fuzzy sets approach, providing a mathematical framework in which this vague concept can be studied.

Most of the researches on well-being are carried out at country level. The recent "Better Life Index" allows to compare well-being across countries, based on 11 topics identifying the areas of material living conditions and quality of life [15]. By narrowing down the analysis at sub-national level, a wide variety of situations emerge across and within the countries.

In this paper we propose the construction of fuzzy composite indicators in order to evaluate well-being in the European regions of the 27 member States, as defined in NUTS 2 (Nomenclature of Territorial Units for Statistics of second level).

Related recent studies on the measurement of the living conditions across European regions are: [2, 4, 16, 19].

The theoretical socio-economic framework that we consider is described in: [1, 8, 14, 20]. Well-being at territorial level may be determined by two main domains: material living conditions (or "economic welfare") that include income and wealth, consumption, jobs and earnings, housing; quality of life, defined as the set of non-monetary attributes of individuals and their opportunities and life chances (health status, education and skills, environmental quality, personal security, etc.). The framework also considers the sustainability over time of the socio-economic conditions and of the natural systems.

Well-being composite indicators are highly sensitive to the variables that are selected, to the methods and weights used in the aggregation: different choices may entail quite different results [18].

Starting from the previous conceptual models and the above mentioned considerations, in Sect. 2 we select a subset of variables available in Eurostat database at NUTS 2 level.

In Sect. 3 we describe the steps for the construction of a fuzzy composite indicator, assumed as a synthetic measure of well-being level in the regions.

In Sect. 4 we present the values of this indicator in the map of European regions and sketch the best and the worst areas. In Sect. 5 the fuzzy sets approach is compared with GDP per capita values and with the results of classical multidimensional methods for dimension reduction and classification of the units: principal

Table 1 Subset of well-being indicators used in the analysis

Well-being indicators	Relationship with well-being
Health and road accidents	
Life expectancy at birth	+
Victims in road accidents (on 100,000 residents)	−
Wealth and free time	
GDP at current market prices (100 = mean value)	+
Family disposable income (100 = mean value)	+
% free time weekly hours	+
Labour market	
Employment rate	+
Unemployment rate	−
Long-term unemployment rate	−
Differences between young and adult unemployment rate	−
Education	
% persons with tertiary education	+
Life-long learning	+
Demography	
Elderly rate	−
% under 10 years old	+
Fertility rate	+
Natural Change rate (mean 2006–2010)	+
Environment	
% land use for residential, commercial and industrial purpose	−

3 The Suggested Fuzzy Indicator

Fuzzy sets theory (e.g. [24]) provides an approach to deal with vague concepts as well-being or quality of life [3, 11]; poverty [7, 12], customer satisfaction [22, 23]. Using the fuzzy approach, the well-being of an area may be interpreted as a question of degree, showing the gradual transition from poor to rich regions: the measure of well-being can be expressed as membership degree to the subset A of the best areas.

Consider a set of n regions r_i $(i = 1, 2, \ldots, n)$ and p manifest variables X_s $(s = 1, 2, \ldots, p)$ reflecting the different aspects of well-being. Without loss of generality, let us assume that each variable is positively related with well-being. If a quantitative variable X_s shows negative correlation, we substitute it with a simple decreasing transformation, e.g. $f(x_{si}) = \max(x_{si}) - x_{si}$.

In order to define the membership function for each variable X_s it is necessary:

1. To identify the extreme situation such that $\mu_A(x) = 0$ (non-membership) and $\mu_A(x) = 1$ (full membership)
2. To define a criterion for assigning membership function values to the intermediate categories of the variable

4 Fuzzy Well-Being Levels in European Regions

We have calculated the values of a first fuzzy composite indicator with equal weights for the 16 variables and of a second fuzzy indicator with weights proportional to the factor loadings of the first principal component (see successive Sect. 5). The correlation between unweighted and weighted indicators is very high ($r = 0.992$) and therefore the classification of the regions obtained by the two criteria is very similar; so we will describe only the unweighted indicator. For this index we have considered also different upper thresholds: 80th and 79th percentile. The correlation with the indicator using 90th percentile is very high: 0.990 and 0.975, respectively. Therefore we present only the results with reference to the 90th percentile upper threshold.

The values of this fuzzy indicator have an interesting interpretation: a value equal to 0 corresponds to a region under the median for all the variables, a value equal to 1 identifies a region over the 90-th percentile for all the variables and a value in the open range (0, 1) may be assumed as membership degree of the region to the subset A of the best areas, i.e. as a fuzzy measure of well-being.

The top ten regions for well-being level are (fuzzy indicator value in brackets):

- Berkshire, Buckinghamshire and Oxfordshire, UK (0.86)
- Stockholm, SE (0.81)
- Noord-Holland, NL (0.80)
- Bedfordshire and Hertfordshire, UK (0.78)
- Zuid-Holland, NL (0.74)
- Flevoland, NL (0.72)
- Gloucestershire, Wiltshire and Bristol/Bath area, UK (0.70)
- Overijssel, NL (0.69)
- Hovedstaden, DK (0.69)

We highlight that no region presents a value equal to one of the fuzzy indicator, i.e. no region shows values greater than the 90-th percentile for all the 16 variables.

The regions with the worst conditions, all with zero values of the fuzzy well-being indicator, are:

- Yuzhen Tsentralen, BG
- Dytiki Makedonia, EL
- Nyugat-Dunántúl, HU
- Dél-Dunántúl, HU
- Dél-Alföld, HU
- Sud-Muntenia, RO

The complete list of the values of the fuzzy indicator may be requested to the first author.

In Fig. 1 we present the map of the values of the fuzzy composite indicator in European regions, according to a partition with five equal classes, based on percentiles. The map is obtained using the program GvSig (http://www.gvsig.org), component analysis (PCA) and k-means cluster analysis, applied to the same dataset.

In Sect. 6 we compare the values of the fuzzy indicator with the ones of human development index (HDI) at NUTS 2 level and we show their non-linear relation.

Concluding remarks in Sect. 7 highlight the additional information of the suggested approach with respect to the traditional ones.

2 The Selection of the Variables

The NUTS 2 classification subdivides the 27 European States into 271 regions. Source of the data is Eurostat's database. This classification corresponds in Italy to the administrative regions, with the exception of Trentino-Alto Adige, divided into the provinces of Trento and Bolzano.

First of all we have erased from the data set 5 units not belonging to European Union: HR1, HR2, HR3—candidate regions in Croatia; IS 00 Iceland (Efta Country); FI 1B Helsinki (new region). We have also deleted the following six regions in other continents: Guyane, Réunion, Martinique, Guadalupe (FR); Melilla, Ceuta (ES).

The selection of the variables has been done starting from the list of all available indicators at NUTS 2 level for European regions (reference year 2010), which is a strong limitation in the definition of the complex concept of well-being. Above all, there is a lack of suitable variables for describing at regional level the aspects of the sustainability, social connection, personal security and subjective well-being.

In order to avoid redundancy for the available domains, a variable selection procedure has been carried out. In most cases, the inclusion of all the variables in a statistical analysis is, at best, unnecessary and, at worst, a serious impediment to the correct interpretation of the data. If two variables are highly correlated, then one of them can often be deleted without the final result being greatly influenced. One way of achieving a simple interpretation is to reduce the number of variables, i.e. to select a subset of the variables to preserve as far as possible the original information. On this topic see, e.g. [17].

Using the criteria of the correlation matrix and PCA, a subset of 16 standardized variables has been selected, with respect to six domains: health and road accidents, wealth and free time, labor market, education, demography, environment. In Table 1 the list of the variables and their relationship, positive or negative, with the global well-being is presented. We point out that the set of the selected variables includes the three aspects considered in the HDI: life expectancy, education and GDP per capita (analysed in Sect. 6).

For each standardized variable X (for simplicity of notation we omit index s), we choose an inferior (lower) threshold l and a superior (upper) threshold u, with l and u finite, and we define the m.f. $\mu_A(x_i)$ as follows:

$$\mu_A(x_i) = \begin{cases} 0 & x_i \leq l \\ \frac{x_i - l}{u - l} & l < x_i < u \\ 1 & x_i \geq u \end{cases}$$

We have chosen: *lower* threshold l = median of the variable; *upper* threshold u = 90th percentile. With this choice the regions with a value of the variable under the median do not belong to the subset A of the best regions, with reference to the considered aspect, and the regions with the 10 % highest values totally belong to the subset of the areas with the highest quality of life.

Among the steps of the construction of a composite index, weighting and aggregation criteria are the most difficult ones as they directly affect the quality and reliability of the results (e.g. [10, 18]). Let us consider the criteria for aggregating p fuzzy variables into a fuzzy composite indicator. A general aggregation function is the weighted generalized mean:

$$\mu_A(i) = \{\sum_{s=1}^{p} [\mu_A(x_{si})]^{\alpha} w_s\}^{1/\alpha}$$

where $w_s > 0$ is the normalized weight that expresses the relative importance of the variables X_s; $(\sum_{s=1}^{p} w_s = 1)$. For fixed arguments and weights, the function is monotonic increasing with α; if $\alpha \to -\infty$, then it becomes the intersection; if $\alpha \to +\infty$, then it is equal to the union. For $\alpha \to 0$ it becomes the weighted geometric mean.

The weighting criteria may be:

- Equal weights, which imply a careful selection of the variables in order to assure a balance of the different aspects of the latent phenomenon
- Factor loadings, obtained by PCA
- Subjective weights obtained by expert judgments, with reference to the importance of the different aspects

Obviously other thresholds, other functions (as the exponential or the cubic ones) and other weights may be considered and in the next section we test the sensitivity of the results obtained using different selection criteria.

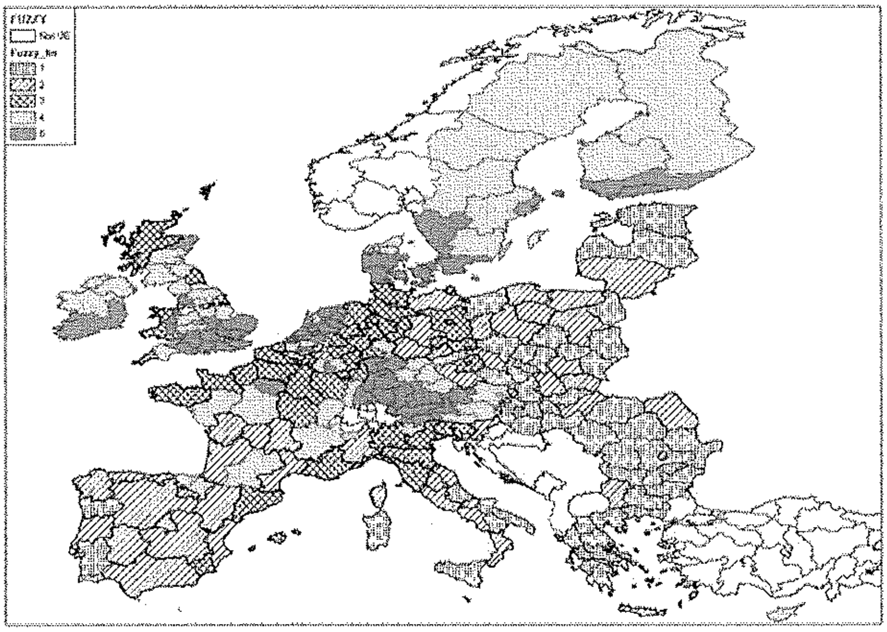

Fig. 1 Map of the European regions according to the values of the fuzzy composite indicator of well-being

a cartographic information system for visualizing the results. The map of the European regions shows the five classes of percentiles by different types of grids: the darkest areas correspond to the best regions.

The lowest levels of well-being are located in the East European countries and in a few regions in Portugal, Spain and southern Italy. The best areas are scattered in different countries of central and northern Europe.

5 Comparison with GDP Per Capita and With Other Multidimensional Indicators

It is interesting to compare the results of the fuzzy multidimensional approach with the traditional indicator of development, i.e. GDP per capita and with the results of other multidimensional methods, applied to the same set of 16 variables. The metropolitan areas of Brussels, Inner and Outer London show too high values for a few variables and may be considered as multidimensional outliers. They have been omitted in the following comparisons and therefore only 257 units are considered.

The correlation between the fuzzy indicator and GDP per capita is moderate ($r = 0.712$) and this restates that GDP is a poor and insufficient criterion for a global evaluation of living conditions of territorial units.

Another comparison method is the classification of the values of the fuzzy indicator values and GDP per capita into a contingency table, considering for each indicator the partition corresponding to five classes of percentiles (Table 2).

The regions in the same percentile class with the two criteria (main diagonal of the matrix) are 42.8 %, and Kendall's tau is 0.620. We highlight that the extreme regions (the worst and the best) are rated in a similar manner on the basis of the two criteria, whereas the regions in the middle of the range present more different classifications.

We have applied PCA to the same set of 16 variables of well-being. The first PC accounts for 37.5 % of the total variance and the second PC for 20.2 %. The percentage explained by the two PC is equal to 57.7 % and is superior to the threshold $0.95^{16} = 0.44$ (Cronbach's alpha = 0.844). The first PC is highly related to the variables measuring income and wealth, education, labour market and life expectancy; the second PC describes demographic domain. The linear correlation between the previous fuzzy indicator and the first PC is sufficiently high ($r = 0.932$) and also the rankings of the regions obtained by the two criteria are similar, but not equal (Spearman's rho = 0.950). The contingency table with reference to the fuzzy indicator values and the scores of the first PC, considering for each indicator the partition corresponding to five classes of percentiles (Table 3), shows that most of the regions (67.7 %) are in the same percentile class with the two criteria, i.e. the two indicators show similar but not equal results (Kendall's tau = 0.847).

Table 2 Contingency table of the values of the fuzzy indicator and GDP per capita

		Percentile classes of GDP per capita					
		1	2	3	4	5	Total
Percentile Classes of the Fuzzy indicator	1	35	14	1	1	0	51
	2	15	17	11	7	2	52
	3	1	13	14	15	8	51
	4	0	7	16	16	13	52
	5	0	1	9	13	28	51
Total		51	52	51	52	51	257

Table 3 Contingency table of the values of the fuzzy and the first PC indicators

		Percentile classes of the first PC					
		1	2	3	4	5	Total
Percentile classes of the fuzzy indicator	1	39	12	0	0	0	51
	2	11	32	9	0	0	52
	3	1	8	30	12	0	51
	4	0	0	12	31	9	52
	5	0	0	0	9	42	51
Total		51	52	51	52	51	257

Table 4 Five clusters of regions by k-means method

Cluster index	Number of regions	Fuzzy values average	First PC scores average	Second PC scores average
3	61	0.089	−1.240	0.178
1	29	0.143	−0.906	0.258
2	37	0.309	0.254	0.974
5	65	0.331	0.323	−1.286
4	65	0.491	1.100	0.437

Finally, we have applied k-means cluster analysis to the 16 standardized variables selecting five groups (for comparison reasons with the previous partitions), ranked according to the average of the values of the fuzzy indicator of the regions in each cluster. The average of the scores of the first and second PC is also presented (Table 4). The 65 regions in cluster n. 4 are the ones with the highest well-being measured by fuzzy and PC indicators.

6 Comparison with HDI

The comparison of the suggested well-being indicators with the results of other researches on this topic at sub-national level is not an easy task, as a consequence of the differences in the choice of variables, methods and territorial units.

We compare our results with the values of HDI computed by Bubbico and Dijkstra [6] for European regions at NUTS 2 level, with reference to 27 EU countries for the year 2007.

HDI is the average of three normalized indices, one in each dimension of human development:

- Life expectancy at birth
- Education
- GDP per capita (PPP US dollars)

The index presents values in the range [0, 100], where 0 is equal to the lowest level of human development and 100 to the highest. The HDI is usually calculated in order to compare the development of the nations all over the world (see, e.g. [21]).

Figure 2 shows the scatterplot with respect to HDI and the fuzzy composite indicator for the same 257 European regions examined in Sect. 5. The relationship between the two indices is moderate and non-linear: $r^2 = 0.597$ for the linear function and $r^2 = 0.720$ for the quadratic function (the cubic function shows a non-significant increase $r^2 = 0.722$). There are also a few bivariate outliers, very far from the curve. In the left side of the figure we can see: 201 = Acores (PT); 202 = Madeira (PT); 205 = Nord-Est of Romania; 208 = Bucuresti (RO) and 224 = Východné Slovensko (SK). Under the curve there is 15 = Brabant Wallon (BE)

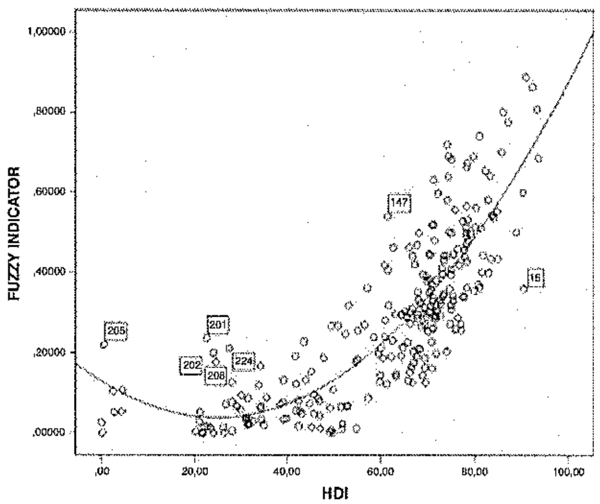

Fig. 2 Scatterplot of the 257 European region with reference to HDI and fuzzy indicator with superimposed quadratic function. The numbers correspond to regions that are bivariate outliers

and over the function 147 = Bolzano (IT). For the seven above mentioned regions the two criteria of well-being evaluation entail quite different results. Deleting these 7 outliers, we obtain a slight improvement in the goodness-of-fit for the quadratic function: $r^2 = 0.750$. The differences between the two indicators may be explained by the sets of variables (3 against 16), the transformation and aggregation criteria, the reference year (2007 and 2010).

7 Concluding Remarks

In this paper we have suggested a criterion based on fuzzy sets theory for the construction of well-being indices at sub-national level. Our fuzzy composite indicator is based on a set of variables describing the various domains of well-being and it presents values in the closed range [0, 1]. The great advantage of this index is its simple and interesting interpretation: a value equal to 0 corresponds to a region under the median for all the variables, a value equal to 1 identifies a region over the 90-th percentile for all the variables and a value in the open range (0, 1) may be assumed as membership degree of the region to the subset of the areas with highest well-being.

The application of the fuzzy indicator to the European regions at NUTS 2 level, considering a set of 16 variables, has pointed out new aspects and better explanations of well-being. The comparison with the results of PCA, applied to the same set of variables, has highlighted that the linear and rank correlation between the previous fuzzy indicator and the first PC are sufficiently high ($r = 0.932$; Spearman rho = 0.950), i.e. the rankings of the regions obtained by the two criteria are similar, but not equal. The correlation of these two composite indicators with GDP per capita is moderate ($r = 0.712$ and $r = 0.759$, respectively) and this confirms the inadequacy of such single variable for a complete description of well-being concept. The relation of the fuzzy indicator with HDI is non-linear ($r^2 = 0.720$ for the quadratic function) and there are a few regions that may be considered as bivariate outliers.

The shortcomings of the suggested approach are related to the following subjective choices in the various steps of the construction of a fuzzy composite indicator:

- Set (or subset) of variables
- Form of the membership function and lower and upper thresholds
- Weights of the variables
- Aggregation criterion.

The sensitivity and robustness of the results with respect to a few different choices in the previous steps have been examined in Sect. 4.

References

1. Allin, P., Hand D.J.: The Wellbeing of Nations: Meaning, Motive and Measurement, Environmetrics. Wiley, New York (2014)
2. Annoni, P., Weziak-Bialowolska, D.: Quality of Life at the Sub-national Level: An Operational Example for the EU. Publications Office of the European Union, Luxembourg (2012)
3. Baliamoune-Lutz, M.: On the measurement of human well-being: fuzzy set theory and Sen's capability approach. In: McGillivray, M., Clarke, M. (eds.) Understanding Human Well-being. United Nation University Press, New York (2006)
4. Berzieri, L., Milioli M.A., Zani S.: A Fuzzy Approach to Measure Well-Being in European Regions. CD, SIS Conference, Brescia (2013)
5. Bleys, B.: Beyond GDP: Classifying alternative measures for progress. Soc. Indic. Res. **109**, 355–376 (2012)
6. Bubbico, R.L., Dijkstra, L.: The European regional Human Development and Human Poverty Indices, Regional Focus, no. 2, European Union, Regional Policy (2011)
7. Cerioli, A., Zani, S.: A fuzzy approach to the measurement of poverty. In: Dagum, C., Zenga, M. (eds.) Income and Wealth Distribution, Inequality and Poverty, pp. 272–284. Springer, Berlin (1990)
8. CNEL-ISTAT, Rapporto BES 2013: il benessere equo e sostenibile in Italia, Roma (2013)
9. Diener, E.: Subjective well-being: the science of happiness and a proposal for a national index. Am. Psychol. **55**, 34–44 (2000)
10. Hsieh, C.-M.: Importance is not important: the role of importance weightings in QOL measures. Soc. Indic. Res. **109**, 267–278 (2012)
11. Lazim, M.A., Abu Osman, M.T.: A new Malaysian Quality of Life index based on fuzzy sets and hierarchical needs. Soc. Indic. Res. **94**, 499–50 (2009)
12. Lemmi, A., Betti. G. (eds): Fuzzy Set Approach to Multidimensional Poverty Measurement. Springer, New York (2006)
13. OECD: Handbook on Constructing Composite Indicators. OECD Publishing, Paris (2008)
14. OECD: Compendium of OECD Well-Being Indicators. OECD Publishing, Paris (2011)

15. OECD: Better Life Index. OECD Publishing, Paris (2013)
16. Okulicz-Kozaryn, A.: Income and well-being across European Provinces. Soc. Indic. Res. **106**, 371–392 (2012)
17. Pacheco J., Casado S., Porras S.: Exact methods for variable selection in principal component analysis: guide functions and pre-selection. Comput. Stat. Data Anal. **57**, 95–111 (2013)
18. Paruolo P., Saisana M. Saltelli A.: Ratings and rankings: voodoo or science? J. R. Stat. Soc. A **176**, **3**, 609–634 (2013)
19. Pittau, M.G., Zelli, R., Gelman, A.: Economic Disparities and Life Satisfaction in European Regions. Soc. Indic. Res. **96**, 339–361 (2010)
20. Stiglitz, J., Sen, A., Fitoussi, J.-P.: Report by the Commission on the Measurement of Economic Performance and Social Progress, Paris (2009)
21. UNDP: Human Development Report 2011. Palgrave Macmillan, New York (2011)
22. Zani, S., Milioli, M.A. Morlini, I.: Fuzzy Methods and Satisfaction Indices. In: Kennett, R.S., Salini, S. (eds.) Modern Analysis of Customer Surveys, pp. 439–456. Wiley, New York (2012)
23. Zani, S., Milioli, M.A. Morlini, I.: Fuzzy Composite Indicators: An Application for Measuring Customer Satisfaction. In: Torelli, N., Pesarin, F., Bar-Hen, A. (eds.) Advances in Theoretical and Applied Statistics, pp. 243–253. Springer, New York (2013)
24. Zimmermann, H.J.: Fuzzy Sets Theory and its Applications, 4th edn. Kluwer, Boston (2001)