cognome nome
password

Società Italiana di Statistica (SIS) Pubblicazioni Eventi e Iniziative
Premi e Concorsi SIS Formazione Archivi Link

Atti della XLV Riunione Scientifica (2010)

<<Home>> <<Atti Riunioni Scientifiche>>

XLV Riunione Scientifica
(Padova, 16-18 giugno 2010)

ISBN 978-88-6129-566-7
Editore: CLEUP – Padova

Plenary Session

James J. Heckman, *University of Chicago, USA*
Structural vis-à-vis reduced form estimation of program effects

George Casella, *University of Florida, USA*
Estimation in Dirichlet process random effect models

Paul Boyle, *University of St Andrews, UK*
Space, migration and health: learning from routine data

Enrico Giovannini, *ISTAT, Rome*
Statistics, knowledge and policy: towards a new paradigm

I	Enrico Rettore,	James J. Heckman,	Structural vis-à-vis reduced form estimation of program

Fuzzy Composite Indicators: An Application for Measuring Customer Satisfaction

Sergio Zani, Maria Adele Milioli, Isabella Morlini

Abstract This paper deals with the construction of a fuzzy composite indicator of a latent phenomenon, using a set of manifest variables measured on different scales (quantitative, ordinal and binary). A few criteria for assigning values to the membership function are discussed, as well as for defining the weights of the variables. For ordinal variables we propose a fuzzy quantification method based on the sampling cumulative function and a weight system taking account of the relative frequency of each category. An application to obtain a synthetic measure of customer satisfaction from the results of a survey is presented.

Keywords: imprecise data and fuzzy methods, membership function, quantification of ordinal variables, weighting criteria, overall satisfaction.

1 Introduction

Composite indicators should ideally measure multidimensional concepts which cannot be captured by a single variable. Usually they are formed on the basis of a set of quantitative variables (OCDE, 2008; JRC, 2009). For constructing a composite indicator, in this paper we suggest a method based on fuzzy set theory (Zadeh, 1965; Zimmermann, 1993; Coppi *et al.*, 2006). As mentioned in the literature, fuzzy sets enable to represent vague concepts, e.g. poverty (Cerioli, Zani, 1990; Lemmi, Betti, 2006), well-being (Chiappero Martinetti, 2000; Baliamoune-Lutz, 2004), quality of life (Lazim, Osman, 2009), etc. First we deal with the general problem of obtaining a synthetic fuzzy measure of a latent phenomenon from a set of manifest vari-

Sergio Zani
Dipartimento di Economia, Università di Parma, e-mail: sergio.zani@unipr.it

Maria Adele Milioli
Dipartimento di Economia, Università di Parma, e-mail: milioli@unipr.it

Isabella Morlini
Dipartimento di Economia, Università di Modena, e-mail: isabella.morlini@unimore.it

ables measured on different scales (quantitative, ordinal and binary). In section 2 we present a few criteria to transform the values of a quantitative variable into fuzzy numbers. For ordinal variables we propose a fuzzy quantification method based on the cumulative function and for a set of binary variables we consider the relative frequency of the *symptoms* of the underlying concept. In section 3 we discuss the problem of weighting the individual variables and their aggregation in a composite indicator. Ideally, weights should reflect the contribution of each variable to the latent phenomenon. For ordinal variables we suggest a new criterion taking account of the relative frequency of each category. In section 4 we focus on the specific application of measuring customer satisfaction using ordinal (and binary) variables. The gradual transition from very dissatisfied to really satisfied customers can be captured by a fuzzy index. We apply the suggested methods to a sample of 704 respondents of a survey on the users of a contact center, in order to evaluate their satisfaction for the service. The fuzzy indicator of customer satisfaction allows to obtain a ranking of respondents that can be compared with the traditional ones.

2 Fuzzy transformations of the variables

Let X be a set of elements $x \in X$; a fuzzy subset A of X is a set of ordered pairs:

$$[x, \mu_A(x)] \qquad \forall x \in X \tag{1}$$

where $\mu_A(x)$ is the membership function (m.f.) of x to A in the closed interval [0, 1]. If $\mu_A(x) = 0$ then x does not belong to A, while if $\mu_A(x) = 1$ then x completely belongs to A. If $0 < \mu_A(x) < 1$ then x partially belongs to A and its membership to A increases according to the values of $\mu_A(x)$. Let us assume that the subset A defines the position of each element with reference to achievement of the latent concept e.g. the well-being of a set of countries or the satisfaction of a sample of customers. In this case $\mu_A(x) = 1$ identifies a situation of fully achievement of the target (a country enjoying global well-being or a completely satisfied customer), whereas $\mu_A(x) = 0$ denotes a total failure (a country with no well-being or a very dissatisfied customer).

Consider a set of n units or elements e_i $(i = 1, 2, ..., n)$ and p manifest variables X_s $(s = 1, 2, ..., p)$ reflecting the latent phenomenon. Without loss of generality, let us assume that each variable is positively related with that phenomenon, i.e. it satisfies the property "the larger the better". If a quantitative variable X_s shows negative correlation, we substitute it with a simple decreasing function transformation, e.g. $f(x_{si}) = max(x_{si}) - x_{si}$; in case of an ordinal variable we consider it in reverse order. In order to define the m.f. for each variable it is necessary:

i) to identify the extreme situation such that $\mu_A(x) = 0$ (non membership) and $\mu_A(x) = 1$ (full membership);
ii) to define a criterion for assigning m.f. values to the intermediate modalities of the variable.

Let us assume that X_s is a quantitative variable; for simplicity of notation we omit index s. For that variable X we choose an inferior (lower) threshold l and a superior (upper) threshold u and we define the m.f. as follows:

$$\begin{cases} \mu_A(x_i) = 0, & x_i \leq l \\ \mu_A(x_i) = \frac{x_i - l}{u - l}, & l < x_i < u \\ \mu_A(x_i) = 1, & x_i \geq u \end{cases} \tag{2}$$

In (2) m.f. is a linear function between the values of the two thresholds. Alternatively, we can arrange the values x_i in non-decreasing order according to i and define the following m.f.:

$$\begin{cases} \mu_A(x_i) = 0, & x_i \leq l \\ \mu_A(x_i) = \mu_A(x_{i-1}) + \frac{F(x_i) - F(x_{i-1})}{1 - F[x_{i(l)}]}, & l < x_i < u \\ \mu_A(x_i) = 1, & x_i \geq u \end{cases} \tag{3}$$

where $F(x_{si})$ is the sampling cumulative function of the variable X and $x_{i(l)}$ is the highest value $x_i \leq l$. If $l = x_1 = min(x_i)$ and $u = x_n = max(x_i)$ formula (3) corresponds to the "totally fuzzy and relative approach" suggested by Cheli, Lemmi (1995). In the literature other specifications have been considered. In this paper we are trying to measure the degree of achievement of a certain target. Therefore, the distance $d(x)$ between the value x and the goal is an indicator of the success in achieving the target. If $d(x) = 0$ there is full membership to A, then $\mu_A(x) = 1$. If $d(x) > 0$ then $\mu_A(x) < 1$. So we can write:

$$\mu_A(x) = \frac{1}{1 + d(x)} \tag{4}$$

In general, as highlighted by Zimmerman (1993), the relationship between physical measures and perception takes an exponential form. The distance $d(x)$ can be expressed as:

$$d(x) = e^{-a(x-b)} \tag{5}$$

so that m.f. is defined as follows:

$$\mu_A(x) = \frac{1}{1 + e^{-a(x-b)}} \tag{6}$$

It is worth noting that the parameter a represents the extent of vagueness and the parameter b may be viewed as the point in which the tendency of the subject's attitude changes from rather positive into rather negative. Baliamoune-Lutz (2004) uses m.f. (6) in the problem of measuring the human well-being with a fuzzy approach. If variable X_s is ordinal with k categories, a suitable choice is the following: the m.f. values of the modalities up to the threshold l are equal to 0 (absence of the phenomenon) and those of the modalities $\geq u$ are equal to 1 (complete presence); the intermediate values of m.f. are defined according to the formula (3). For example, consider the scores (on a scale 1-10) of the variable "Overall satisfaction", described in section 4. Choosing l = 5 and u = 9, we obtain the m.f. values indicated in the last column of Table 1. A customer with score equal to 5 does not belong to the set

Table 1 Membership function to the subset of satisfied customers with reference to variable X_8

scores (x_i)	n_i	$F(x_i)$	m.f.(x_i)
≤ 3	7	0.010	0
4	5	0.017	0
5	12	0.034	0
6	40	0.091	0.059
7	161	0.320	0.296
8	240	0.732	0.723
9	93	0.864	1
10	96	1	1
	704		

of satisfied respondents, while a customer with score equal to 8 has a value 0.723 of the m.f. to the set of satisfied users. If variable X_s is binary, one of the modalities can be considered as a symptom of the latent concept; therefore the m.f. is a crisp function assuming only values equal to 0 (absence) and 1 (presence). However, in general we consider a set of q ($q \leq p$) binary variables reflecting several aspects of the phenomenon; in this situation the m.f. can be defined as follows:

$$\mu_A(x_i) = \frac{1}{q} \sum_{s=1}^{q} z_{si} \tag{7}$$

where z_{si}=1 if the corresponding x_{si} denotes presence of the symptom and z_{si}=0 otherwise. Definition (7) is consistent with interpreting membership values as the proportion of "subjects" who rate the i-th unit as an actual member of the fuzzy set A (Cerioli, Zani, 1990).

3 Weighting and aggregation of the variables

Among the steps of the construction of a crisp composite indicator, weighting and aggregation is a major one which directly affects the quality and reliability of the results (OCDE, 2008).

Let us consider the criteria for aggregating the p fuzzy variables, described in Section 2, into a fuzzy composite indicator. The simpler operations for the $i-th$ unit are:

$$intersection: \quad {}_I\mu_A(i) = min[\mu_A(x_{1i}), \mu_A(x_{2i}), ..., \mu_A(x_{pi})] \tag{8}$$

$$union: \quad {}_U\mu_A(i) = max[\mu_A(x_{1i}), \mu_A(x_{2i}), ..., \mu_A(x_{pi})] \tag{9}$$

A general aggregation function is the weighted generalized means (Klir, Folger, 1988, p. 61):

$$\mu_A(i) = \sum_{s=1}^{p} [\mu_A(x_{si})]^{\alpha} \cdot {w_s}^{1/\alpha} \tag{10}$$

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, function (10) is monotonic increasing with α; if $\alpha \to -\infty$, then formula (10) becomes the intersection defined in (8); if $\alpha \to +\infty$, then (10) is equal to the union (9). For $\alpha \to 0$ formula (10) becomes the weighted geometric mean.

The weighting criteria in (10) may be:

- equal weights, that 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 principal components analysis (PCA) for quantitative variables or by nonlinear PCA for ordinal variables; this method of weighting is valid if the first component accounts for a high percentage of the total variance and the weights (loadings) of the variables are proportional to their correlation with the first component (factor) reflecting the underlying concept;
- obtained from expert judgements, e.g. using focus groups;
- determined by an Analytic Hierarchy Process (Kwong, Bay, 2002).

We suggest a criterion for the determination of the weights considering for each variable X_s the fuzzy proportion $g(X_s)$ of the achievement of the target:

$$g(X_s) = \frac{1}{n} \sum_{i=1}^{n} \mu(x_{si}) \tag{11}$$

If X_s is binary, formula (11) coincides with the crisp proportion and in general it may be seen as an index of the proportion of the units having (totally or partially) the latent phenomenon (Cheli, Lemmi, 1995). The normalized weights may be determined as an inverse function of $g(X_s)$, in order to give higher importance to the rare features in the n units . To avoid excessive weights to the variables with low value of $g(X_s)$ we choose (Cerioli, Zani, 1990):

$$w_s = ln \left(\frac{1}{g(X_s)} \right) / \sum_{s=1}^{p} ln \left(\frac{1}{g(X_s)} \right) \tag{12}$$

Through (12) it is possible to attach to each variable a weight sensitive to the fuzzy membership of the units to A according to such variable.

4 A fuzzy indicator of customer satisfaction

Customer satisfaction may be defined as the degree of happiness that a customer experiences with a product or a service and is a function of the gap between expected and perceived quality. But customer satisfaction is a vague concept and consequently fuzzy methods can be used to measure this latent phenomenon (Chien, Tsai, 2000; Darestani and Jahromi, 2009). We apply the methods of the previous sections to the results of a survey on customer satisfaction of the users of the Contact Center of the Municipality of Parma, considering a sample of 704 respondents calling for information. See Zani, Berzieri (2009) for a complete description. The questions on the degree of satisfaction of the users are:

X_1 = Contact at the first call (no=0; yes=1) (CONTACT)
X_2 = Waiting time (too long=1; normal=2; fairly short=3) (WAITING)
X_3 = Courtesy of the operator (COURTESY)
X_4 = Skill of the operator (SKILL)
X_5 = Quality of the provided information (QUALITY)
X_6 = Speed of the information (SPEED)
X_7 = Complete answer (no=1; partially=2; yes=3) (COMPLETE)
X_8 = Overall satisfaction for the service, with scores from 1 to 10 (OVERALL).

All the variables whose modalities are not specified are measured on Likert scale (very dissatisfied=1; dissatisfied=2; neither satisfied nor dissatisfied=3; satisfied=4; very satisfied=5). The cumulative function of the variables X_1 - X_7 and the corresponding m.f. according to formula (3) are presented in Table 2 and for variable X_8 in the previous Table 1. For the variables on Likert scale the inferior threshold is "dissatisfied" and the superior is "very satisfied".

Table 2 Cumulative function F of the modalities of each item in the sample and corresponding membership function to the subset A of satisfied customers.

	X_1			X_2			X_3			X_4			V_5			X_6			X_7	
	F	$m.f.$		F	$m.f.$		F	$m.f.$		F	$m.f$		F	$m.f.$		F	$m.f.$		F	$m.f.$
0	0.08	0	**1**	0.03	0	**1**	0	0	**1**	0	0	**1**	0.01	0	**1**	0.01	0	**1**	0.05	0
1	1	1	**2**	0.29	0.27	**2**	0.01	0	**2**	0.03	0	**2**	0.04	0	**2**	0.04	0	**2**	0.13	0.08
			3	1	1	**3**	0.05	0.05	**3**	0.14	0.12	**3**	0.13	0.09	**3**	0.12	0.08	**3**	1	1
						4	0.46	0.45	**4**	0.60	0.59	**4**	0.57	0.55	**4**	0.58	0.56			
						5	1	1	**5**	1	1	**5**	1	1	**5**	1	1			

We have considered the following weighting systems, with and without the variable X_8, OVERALL (see Table 3):

1. equal weights of the variables (W_1);
2. normalized factor loadings obtained by standard (linear) PCA. (The first PC accounts for 49.7% of the total variance with X_8 and for 50.4% without it) (W_2);
3. normalized factor loadings applying PCA on τ rank correlation matrix. (The first PC explains 46.67% of the total variance with X_8 and 46.44% without it) (W_3);
4. normalized weights as inverse function of the fuzzy proportion of each variable, according to formula (12) (W_4).

The least important variable is always X_1 and could be deleted. The correlation coefficient between W_2 and W_3 is 0.964 considering X_8 and 0.970 without it, but the correlation coefficients of W_4 with the previous weights are in the interval [0.791, 0.909]. Therefore, the last weighting criterion is slightly different from the others. Table 4 shows the frequency distribution of the values of the fuzzy composite indica-

Table 3 Values of W_2, W_3 and W_4.

	With X_8			Without X_8		
	W_2	W_3	W_4	W_2	W_3	W_4
X_1	6.97	6.93	3.59	7.99	7.83	4.38
X_2	9.21	8.84	10.57	10.62	10.05	12.88
X_3	12.29	13.67	13.65	14.17	15.64	16.63
X_4	14.76	15.50	16.18	16.95	17.71	19.71
X_5	15.61	16.08	16.43	17.97	18.25	20.02
X_6	17.73	16.02	16.22	18.13	18.32	19.76
X_7	12.27	10.67	5.43	14.17	12.20	6.62
X_8	13.16	12.29	17.93	-	-	-
Tot	100	100	100	100	100	100

tors with the mentioned weighting criteria. None of the respondents can be regarded as completely unsatisfied, since the values of the composite indicators are at least 0.02: even clients experiencing dissatisfaction for most indicators are found to be not completely dissatisfied for others. On the other hand, 86 respondents (with X_8) and 170 (without X_8) belong to the subset of completely satisfied customers.

Table 4 Frequency distribution of the fuzzy composite indicators with different weights.

	with X_8				without X_8			
classes	W_1	W_2	W_3	W_4	W_1	W_2	W_3	W_4
0.0 ⊢ 0.1	4	7	7	8	5	7	7	7
0.1 ⊢ 0.2	8	12	12	12	7	12	12	14
0.2 ⊢ 0.3	14	13	16	20	15	15	15	18
0.3 ⊢ 0.4	36	36	37	39	30	34	33	35
0.4 ⊢ 0.5	30	31	33	26	35	28	29	30
0.5 ⊢ 0.6	49	49	46	120	26	27	37	81
0.6 ⊢ 0.7	109	158	151	103	91	205	195	142
0.7 ⊢ 0.8	134	96	106	82	161	68	68	69
0.8 ⊢ 0.9	105	77	66	99	122	69	69	66
0.9 ⊢ 1.0	129	136	144	109	42	69	69	72
1.0	86	86	86	86	170	170	170	170
Tot	704	704	704	704	704	704	704	704

Finally, we have compared the suggested fuzzy approach with the traditional (crisp) indicators of customer satisfaction. For lack of space we present only Figure 1. To each score of OVERALL corresponds a distribution of the values of m.f. to the subset of satisfied customer with median increasing function of the scores and a few outliers.

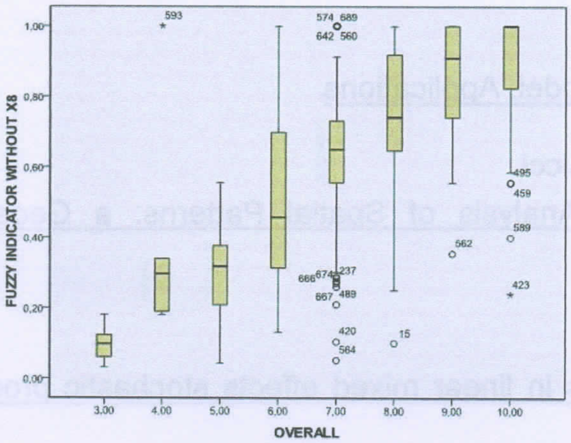

Fig. 1 Boxplots of the fuzzy indicators with weights W4 for each value of variable OVERALL.

References

1. Baliamoune-Lutz M. (2004), On the Measurement of Human Well-Being: Fuzzy Set Theory and Sen's Capability Approach, UNU-WIDER, Helsinki.
2. Cerioli A., Zani S. (1990), A fuzzy approach to the measurement of poverty, in: C. Dagum, M. Zenga (eds.), Income and Wealth Distribution, Inequality and Poverty, Springer, Berlin, 272-284.
3. Cheli B., Lemmi A. (1995), A totally fuzzy and relative approach to the multidimensional analysis of poverty, Economic Notes, 24, n. 1, 115-134.
4. Chiappero Martinetti E. (2000), A multidimensional assessment of well-being based on Sen's functioning approach, Rivista Internazionale di Scienze Sociali, 108, n.2, 207-239.
5. Chien C-J., Tsai, H-H. (2000), Using fuzzy numbers to evaluate perceived service quality, Fuzzy Sets and Systems, vol. 116, 289-300.
6. Coppi R., Gil M.A., Kiers H.A.L. (2006), A fuzzy approach to statistical analysis, Computational Statistics and Data Analysis, 51, 1-14.
7. Darestani A.Y., Jahromi A.E. (2009), Measuring customer satisfaction using a fuzzy inference system, Journal of Applied Sciences, 9, n. 3, 469-478.
8. JRC (2009), An information server on composite indicators and ranking systems, http//:composite-indicators.jrc.ec.europa.eu.
9. Klir G.J., Folger T.A. (1988), Fuzzy Sets, Uncertainity and Information, Prentice-Hall Int., London.
10. Kwong C.K., Bai H. (2002), A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment, J.Intelligent Manufacturing, 13, 367-377.
11. Lazim M.A., Osman M.T.A. (2009), A new Malaysian Quality of Life Index Based on Fuzzy Sets and Hierarchical Needs, Social Indicators Research, 94, n. 3, 499-508.
12. Lemmi A., Betti G. (Eds.) (2006), Fuzzy Set Approach to Multidimensional Poverty Measurement, Springer, New York.
13. OCDE (2008), Handbook on Constructing Composite Indicators, OCDE Publications, Paris.
14. Zadeh L.A. (1965) Fuzzy sets, Information and Control, 8, 338-353.
15. Zani S., Berzieri L. (2009), Measuring customer satisfaction using ordinal variables: an application in a survey on a contact center, Statistica applicata (in press).
16. Zimmermann H.J. (1993), Fuzzy Sets Theory and its Applications, 2nd ed., Kluwer, Boston.