ORIGINAL PAPER

CrossMark

A model-based fuzzy analysis of questionnaires

E. Di Nardo[1] · R. Simone[2]

Abstract
In dealing with veracity of data analytics, fuzzy methods are more and more relying on probabilistic and statistical techniques to underpin their applicability. Conversely, standard statistical models usually disregard to take into account the inherent fuzziness of choices and this issue is particularly worthy of note in customers' satisfaction surveys, since there are different shades of evaluations that classical statistical tools fail to catch. Given these motivations, the paper introduces a model-based fuzzy analysis of questionnaire with sound statistical foundation, driven by the design of a hybrid method that sets in between fuzzy evaluation systems and statistical modelling. The proposal is advanced on the basis of CUB mixture models to account for uncertainty in ordinal data analysis and moves within the general framework of Intuitionistic Fuzzy Set theory to allow membership, non-membership, vagueness and accuracy assessments. Particular emphasis is given to defuzzification procedures that enable uncertainty measures also at an aggregated level. An application to a survey run at the University of Naples Federico II about the evaluation of Orientation Services supports the efficacy of the proposal.

Keywords Uncertainty · CUB models · Intuitionistic fuzzy sets · Fuzzy composite indicators

1 Introduction

Consider an experimental design aimed to investigate an unobservable trait of a population through measurements of opinions, judgements, or preferences. Then, once

✉ R. Simone
rosaria.simone@unina.it

E. Di Nardo
elvira.dinardo@unito.it

1   Department of Mathematics "G. Peano", University of Turin, Via Carlo Alberto, 10, 10123 Turin, Italy

2   Department of Political Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22, 80138 Naples, Italy

a questionnaire is administered and data are collected as ratings on an ordinal scale, policy makers should not disregard to take into account the fuzziness of the outcomes. Indeed, latent constructs like customer satisfaction are inherently vague. For instance, if on a scale ranging from $1 =$ 'completely dissatisfied' to $7 =$ 'completely satisfied', the rater marks $R = 6$ ($R = 2$, resp.), how strong should our confidence be that he/she is actually satisfied (dissatisfied)? How confident can the scholar be about the resulting classification? In the following, we consider the satisfaction of respondents as latent phenomenon, but the proposed investigation can be applied more generally for the analysis of agreement and preferences.

A classical way to assess the imprecision and the uncertainty of evaluations is offered by Fuzzy Sets Theory (Zadeh 1965). When a respondent marks "very satisfied" about a particular service, he/she is producing a judgment on the veracity of the statement "the quality of service is high", whose plausibility can be encoded in a Fuzzy measure. The Intuitionistic Fuzzy Set theory (IFS) (Atanassov 1986) proposes to convey the dual assessment of the membership grade through a non-membership function. Then the residual degree of indecision and its complementary to one are considered as measures of uncertainty and accuracy, respectively, measuring the veracity of the expressed rates. Usually, fuzzy methods for evaluation are preferred over simple descriptive analysis, but they lack of a sound inferential background that could enhance their applicability. Moreover, they are strongly grounded on subjective choices, which makes reproducibility of analysis and conclusions unstable. Thus, the integration with a proper statistical tool could underpin fuzzy methods' reliability.

The present paper aims to pursue this task by fostering the application of a suitable statistical method within the fuzzy framework. Due to its psychological and probabilistic structure, an appealing candidate for our purpose is the class of CUB models (D'Elia and Piccolo 2005; Iannario and Piccolo 2012). This rationale conceives the data generating process driving latent perceptions into discrete evaluations as the combination of two components: the *feeling*, responsible for the level of agreement/pleasantness towards the item under investigation, and the *uncertainty*, accounting for the overall nuisance affecting a fully meditated response (laziness, difficulties in understanding the question, ignorance of the topics, wording and length of the scale etc). CUB models are then defined as a two-component mixture distribution: a shifted Binomial for feeling and a discrete Uniform for uncertainty, which explains the acronym CUB : *C*ombination of a *Uniform* and a shifted *B*inomial.

Then, the proposal is a new Fuzzy evaluation system for ratings in the setting of Intuitionistic theory in order to properly account for uncertainty in the data as meant by the CUB paradigm: dually, the heuristic definition of CUB uncertainty as a measure of the intrinsic fuzziness of the decision process is properly justified. As for all fuzzy schemes (Lalla 2005), the resulting model-based fuzzy system places itself as a support tool to the analysis of ratings.

The advantage of the proposed method is manifold: it is more objective, since fuzzy functions are structured on both data and inferential procedures and simultaneously they are designed to be more sensitive to measurement errors; it is able to discriminate among items of a questionnaire not only at an aggregated level (that is when respondents are grouped according to their choices) and it lends itself to broader extensions to consider further sources of the fuzziness blurring the response distribution. In partic-

ular, the resulting modeling allows us to include the so-called *shelter effect*, occurring when a proportion of respondents identifies a category as a *refuge* option for non-meditated choices (Iannario 2012; Iannario and Piccolo 2016), yielding to inflation of frequency in some categories.

Our proposal considers membership and non-membership functions of spline type (Marasini et al. 2015), grounded on sampled information and limiting the subjectivity of parameters choice. The application of the empirical distribution function within the fuzzy evaluation system is supported by the literature (Cheli and Lemmi 1995; Zani et al. 2013) and, in these terms, it allows us to take into account also the feeling component as meant by CUB models.

In the end, a Fuzzy analysis of questionnaire is completed with a so-called *defuzzification* procedure. This last step consists in computing synthetic measures that encode a simultaneous examination of all items across respondents (Van Leekwijck and Kerre 1999): specifically, fuzzy functions have to be suitably weighted and aggregated to produce fuzzy composite indicators. In Marasini et al. (2015), different criteria and quantification methods are discussed. In particular, the weights associated with items can be either uniform or assigned by experts who are in charge of discriminating the items by assessing their relative importance to the universe of discourse. Here we propose an aggregator belonging to the class of Intuitionistic Weighted Aggregator Means (IWAM) (Beliakov et al. 2011), designed to balance both satisfaction and dissatisfaction for indecision and unpredictability.

With regard to the literature on CUB models, the novelty of the paper is the multi-item perspective offered by the defuzzification procedure, in which dimensions affected by a larger uncertainty are recognized a weaker importance. Although CUB models are designed to run an item-by-item investigation, first multidimensional perspectives are given in (Andreis and Ferrari 2013; Corduas 2015), whereas multi-object analysis are discussed in Iannario and Piccolo (2012), Capecchi et al. (2016) and multi-item aggregation is pursued with a model-based composite indicator in Capecchi and Simone (2018). Very recently, a multivariate extension of CUB models is proposed in Colombi and Giordano (2016).

The work is organized as follows: Sect. 2 provides a short overview of the IFS theory, with a focus on composite indicators given in Sect. 2.2. CUB models are shortly described in Sect. 3 and the proposed CUB -Fuzzy evaluation system is introduced in Sect. 3.1, where a new Fuzzy composite indicator is defined in terms of the CUB uncertainty parameter. Finally, in Sect. 4, the proposal is illustrated on the basis of a survey collected at the University of Naples Federico II about the evaluation of Orientation Services. The discussion is pursued by assuming a comparative perspective with standard methodologies for multi-item analysis. Summarizing remarks and some notes on future developments end the paper. The whole analysis has been run within the R environment: the code is available upon request from Authors.

2 Fuzzy systems: Intuitionistic theory

Let X be the universe of discourse. For instance, assume that we are investigating customers' satisfaction and a questionnaire is designed to that purpose: then X is the

set of all customers, which is observed through the respondents to the survey. A Fuzzy set A consists of a subset of X endowed with a *membership* function μ_A assessing the degree of membership to the set A,

$$\mu_A : X \longrightarrow [0, 1], \qquad x \longmapsto \mu_A(x),$$

in such a way that $\mu_A(x) = 1$ if and only if x is certainly an element of A, while $\mu_A(x) = 0$ if and only if x is certainly not. For the illustrative example above, A is the subset of the satisfied customers. Assume that answers to an item are collected on an Likert-type scale with $m = 10$ categories that can be coded with equispaced integer scores, and that only categories $j \geq 6$ have a positive wording. Evaluations are not *crisp*, and classification of respondents should be elastic, thus it is an over-simplification to consider certainly satisfied those users whose rate expresses satisfaction, regardless of the position along the scale. Certainly, $j = 6$, $j = 8$ and $j = 10$ have to be associated with different degrees of belonging to A. A Fuzzy evaluation system will frame this circumstance by assigning increasing levels of membership to increasing scores. Then, it is widely acknowledged that in studies like those on customer satisfaction, it is of foremost importance to accompany evaluation of satisfaction with measures of unsatisfaction and dissatisfaction. In this vein, the rationale of Intuitionistic Fuzzy Sets (IFS) puts forth a theory to supply the Fuzzy analysis with a *non-membership* function (Atanassov 1986):

$$\nu_A : X \longrightarrow [0, 1], \qquad x \longmapsto \nu_A(x),$$

expressing the dual assessment of the non-membership grade of an element x to A, in such a way that if $\nu_A(x) = 1$, then x is certainly not an element of A.

Membership and non-membership values should be defined in such a way that $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ (Atanassov 2012), thus a measure of the residual indecision about the statement "$x \in A$" is given by the *hesitancy degree* or *Fuzzy uncertainty* function:

$$u_A(x) = 1 - \mu_A(x) - \nu_A(x). \tag{1}$$

Fuzzy uncertainty function (1) parallels the usual confidence band. Indeed the *Interval-Valued Fuzzy Set* (IVFS) for an IF singleton $< \mu_A, \nu_A >$ is the function (Atanassov and Gargov 1989):

$$M_A : X \longrightarrow \mathcal{B}[0, 1], \qquad x \longmapsto [\mu_A(x), 1 - \nu_A(x)],$$

where $\mathcal{B}[0, 1]$ denotes the Borel set of sub-intervals of the unit interval. Then the hesitancy degree is the range of $M_A(x) = [\mu_A(x), \mu_A(x) + u_A(x)]$. For subsequent purposes, let us underline that some IFS evaluation systems first characterize the hesitancy degree and then obtain the non-membership function as:

$$\nu_A(x) = 1 - \mu_A(x) - u_A(x). \tag{2}$$

Within IFS, there are indicators aiming to summarize an item performance when aggregated among subjects, and to derive information about the latent phenomenon

when aggregated among items: in particular we refer to the fuzzy score and the fuzzy accuracy (Xu 2007). The *Fuzzy score* function (3) indicates how strong is the classification of membership with respect to the classification of non-membership by computing how far apart membership and non-membership statements are, that is:

$$s(x) = \mu_A(x) - \nu_A(x) \in [-1, 1]. \tag{3}$$

The *Fuzzy accuracy* function, instead, measures the extent to which the fuzzy classification of membership and non-membership is encompassed:

$$a(x) = \mu_A(x) + \nu_A(x) = 1 - u_A(x) \in [0, 1], \tag{4}$$

in the sense that $a(x) = 1$ denotes that no undefined state is contemplated other than membership and non-membership (conversely, $a(x) = 0$ indicates a fully incomplete fuzzy statement).

2.1 Spline fuzzy systems

When discussing the IFS framework for questionnaire analysis, a benchmark approach is the one delivered in Marasini et al. (2015, 2016, 2017). Let us consider a balanced Likert-type ordinal scale with an odd number m of choices, with an indifference point i_p located at the middle category. This choice is convenient since the indifference point thresholds membership and non-membership grades, although the setting here established could be easily extended to scales of even length. The scale is coded into integer categories, say $1, 2, \ldots, m$, so that a rate $r = 1$ corresponds to the most unsatisfied choice; conversely, $r = m$ corresponds to an extremely satisfied answer.

According to Marasini et al. (2015), the spline membership function is of the type:

$$\mu_A(r) = \begin{cases} 0, & 1 \leq r < a, \\ \dfrac{1}{2} - \dfrac{1}{2}\left(2\dfrac{i_p - r}{b - a}\right)^{\epsilon}, & a \leq r \leq i_p, \\ \dfrac{1}{2} + \dfrac{1}{2}\left(2\dfrac{r - i_p}{b - a}\right)^{\epsilon}, & i_p \leq r \leq b, \\ 1, & b < r \leq m, \end{cases} \tag{5}$$

with $\epsilon > 0$, and $b - a$ denoting the range of non-crisp responses (notice that for the sake of the subsequent discussion, we define membership directly on the ordinal scale, whereas the approach in Marasini et al. (2015) works on the latent continuous measurement scale). Spline parameters ϵ, η, θ are chosen according to the sampling experiment, the strength and vagueness of the wording of the scale and its length. For instance, in Marasini et al. (2015) the authors advocate to adopt a linear spline ($\epsilon = 1$) for items within a certain section of the questionnaire, and a quadratic spline ($\epsilon = 2$) for items within another one, measured on a scale whose wording can be perceived vaguer in the central part of the scale, or in cases in which there is a non-linear step

between subsequent categories. In the latter cases, at least a quadratic spline should be recommended.

The hesitancy degree is defined from (5) with:

$$u_A(r) = \mu_A(r)^\theta (1 - \mu_A(r))^\eta, \quad \theta, \eta \geq 1, \tag{6}$$

and then the non-membership function is derived according to (2). This definition is meant to convey both membership and its residual assessment to uncertainty measurement: for balanced scale, one sets $\theta = \eta$. As a result, without any prior experts' assessments on the values of parameters, the Fuzzy spline functions (5) and (2) are equal for all items evaluated on a common scale.

Although interesting, we will not rely on Definition (6) for the uncertainty function, but implement a fuzzy uncertainty that carries a specific statistical interpretation in the spirit of freeing the stakeholders from preliminary subjective assessments.

2.2 Fuzzy composite indicators

Consider a latent phenomenon to be measured by K observable variables, as the items of a questionnaire. Assume the questionnaire has been filled out by n respondents, who have chosen among m ordered alternatives, available for each item. Let $\mathbf{r}_j = (r_{j,1}, r_{j,2}, \ldots, r_{j,K})$ be the row vector of ratings given by the j-th respondent for $j = 1, 2, \ldots, n$, to the K items, and denote with $\mu_A^{(k)}(\cdot), \nu_A^{(k)}(\cdot)$ the membership and non-membership functions for the k-th item, respectively. If seeking for a composite fuzzy value for each respondent, the IWAM (Intuitionistic Weighted Aggregator Mean) is defined as the pair:

$$< \mu_A(\mathbf{r}_j), \nu_A(\mathbf{r}_j) > = < \sum_{k=1}^{K} w_k \, \mu_A^{(k)}(r_{j,k}), \sum_{k=1}^{K} w_k \, \nu_A^{(k)}(r_{j,k}) >, \tag{7}$$

where $\{w_1, \ldots, w_K\}$ is a given system of weights such that $\sum_{k=1}^{K} w_k = 1$, establishing the relative importance of items. Such values could be used to perform a fuzzy clustering of responses, where belonging of each observation to a cluster is decided on the basis of the fuzzy composite score $\mu_A(\mathbf{r}_j) - \nu_A(\mathbf{r}_j)$, for instance. Each aggregated value is considered an IFS singleton $< j, \mu_A(\mathbf{r}_j), \nu_A(\mathbf{r}_j) >$, thus a final composite score can be obtained by considering uniform weights for subjects:

$$< \bar{\mu}, \bar{\nu} > = < \frac{1}{n} \sum_{j=1}^{n} \mu_A(\mathbf{r}_j), \frac{1}{n} \sum_{j=1}^{n} \nu_A(\mathbf{r}_j) > . \tag{8}$$

Then, according to (1) and (8), the *uncertainty* (or *hesitancy degree*) is computed as the global residual degree of indeterminacy of the fuzzy assessment:

$$\bar{u} = 1 - \bar{\mu} - \bar{\nu}, \tag{9}$$

whereas the overall Fuzzy score and Fuzzy accuracy are given respectively by:

$$\bar{s} = \bar{\mu} - \bar{\nu}, \qquad \bar{a} = \bar{\mu} + \bar{\nu}.$$

Different choices of weights in (7) give different indicators. In the framework of composite indicators, the choice of a weighting system is of primary importance. As a matter of fact, several applications suggest to choose weights depending on the loadings of the first principal component or factor. Nevertheless, such choice is consistent only if that variable explains a large proportion of the variability. For this reason, and aiming to a fuzzy system that is model-based and thus not subjective, we will propose a system of weights that is driven by data through estimation procedure: in this sense, it can be considered a safer option.

3 CUB models

Let R be the rating random variable modelling the response distribution to an item of a questionnaire, measured on a scale with m ordered categories coded as integers from 1 up to m. A CUB distribution CUB (π, ξ) for R consists in the following two-component mixture with parameters $(\pi, \xi) \in (0, 1] \times [0, 1]$:

$$\mathbb{P}\big(R = r \mid \pi, \xi\big) = \pi \, b_r(\xi) + (1 - \pi) \, h_r \,, \quad r = 1, 2, \ldots, m \,,$$

where $b_r(\xi)$, $r = 1, 2, \ldots, m$ for $m > 3$ denotes the shifted Binomial distribution with parameter $1 - \xi$:

$$b_r(\xi) = \binom{m-1}{r-1} \xi^{m-r} (1 - \xi)^{r-1}, \quad r = 1, 2, \ldots, m \,,$$

and $h_r = \dfrac{1}{m}$ is the discrete Uniform distribution over the given support. The parameter ξ is referred to as the *feeling* parameter since $1 - \xi$ measures the preference of a category over the lower ones in a sequence of pairwise comparisons (D'Elia 2000; Iannario and Piccolo 2016). The parameter π, instead, is called the *uncertainty* parameter since $1 - \pi$ charges for the inherent fuzziness arising when perception translates into an evaluation, and thus measures the overall uncertainty of the respondent's assessment. The role of the uncertainty component within the CUB rationale has been usually considered as an expression of the inherent indeterminacy of human decisions, generating fuzziness and thus representing a source of unpredictability of the evaluation process. As a by-product and since the Uniform distribution represents the least informative model, its weight in the mixture aims at catching the level of heterogeneity in the data.

It should be emphasized that the choice of the Uniform distribution for the uncertainty component adheres to the baseline CUB paradigm: under this assumption, π is an inverse indicator of heterogeneity. Departing from the defining specification, other choices can be supported to model uncertainty in the data: response styles and

category-specific measurement errors can be suitably specified by adjusting this component (see Gottard et al. 2016; Simone and Tutz 2018). These extensions do not affect the distinctive trait of the CUB fuzzy evaluation system since this is grounded on the mixing weight π for the deliberate choice. Alternative distributions for the uncertainty component would simply change interpretation of results and penalize data for response styles or more specific forms of uncertainty. In the following, we will focus on some particular circumstances for illustrative purposes.

For instance, in order to further disentangle the fuzziness charged by the uncertainty component, one may contemplate a *shelter effect* concentrated at category $c \in \{1, \ldots, m\}$ in the CUB mixture distribution when inflation in c is observed (Iannario 2012). Let us consider a degenerate random variable $D_r^{(c)}$ such that $\mathbb{P}(D_r^{(c)} = r) = 1$ if $r = c$ and $\mathbb{P}(D_r^{(c)} = r) = 0$ otherwise. Then, the CUB distribution CUB (π, ξ, δ) with *shelter effect* at $r = c$ is:

$$\mathbb{P}(R = r \mid \pi^\star, \xi, \delta) = \delta \, D_r^{(c)} + (1 - \delta) \left[\pi^\star b_r(\xi) + (1 - \pi^\star) \, h_r \right], \quad r = 1, 2, \ldots, m, \tag{10}$$

for $m > 4$. The additional parameter δ quantifies the importance of the *shelter effect*. When testing its significance, it may be useful to deal with (10) according to the following equivalent parameterization:

$$\mathbb{P}(R = r \mid \pi_1, \pi_2, \xi) = \pi_1 \, b_r(\xi) + \pi_2 h_r + (1 - \pi_1 - \pi_2) \, D_r^{(c)}, \quad r = 1, 2, \ldots, m,$$

with $\pi_1 = \pi^\star(1 - \delta) > 0$, $\pi_2 = (1 - \pi^\star)(1 - \delta) \geq 0$. When the inclusion of a *shelter effect* in the model yields to a significant improvement of the fit (to be checked with a Likelihood Ratio Test, for instance), the overall level of inaccuracy has to convey both the heterogeneity accounted by the Uniform distribution and the *shelter effect* as measured by the parameter δ. Since π_1 is the mixture coefficient corresponding to a deliberate choice, the measure of the overall uncertainty in this augmented case corresponds to $1 - \pi_1 = \pi_2 + \delta$ taking into account the *shelter effect*. In order to provide a general framework not limited to cases where the shelter is significant, we shall use the notation π in place of π_1, since in that cases the whole discussion holds for baseline CUB models. Notice that CUB models paradigm assumes a linear step between adjacent categories: for non-linear versions, see Manisera and Zuccolotto (2014).

For our computation we shall rely on the Maximum Likelihood (ML) estimates $\hat{\pi}, \hat{\xi}, \hat{\delta}$ of π, ξ, δ, respectively (equivalently $\hat{\pi}_1, \hat{\pi}_2, \hat{\xi}$) obtained by running the Expectation-Maximization algorithm (D'Elia and Piccolo 2005; Piccolo 2006) as implemented in the R package CUB (Iannario et al. 2018).

3.1 CUB-fuzzy evaluation system

The idea to use CUB model parameters in computing membership functions stems from the preliminary work (Di Nardo and Simone 2016), but the method here presented is more accurately designed.

For a preliminary investigation and comparison with the methods introduced in Sect. 2, we have focussed on balanced Likert-type scales of odd length with indifferent

point at the midpoint. Suppose the scale is oriented in such a way that "the greater the score, the higher the feeling", that is, there is a positive relation between the latent phenomenon and the scale. Here, *negative* and *positive* refer to expression of satisfaction, so that $r < i_p$ ($r > i_p$) corresponds to a *negative* (*positive*) evaluation.

Definition 1 For a given item of the questionnaire, the CUB -Fuzzy membership function is:

$$\mu_A(r) = \begin{cases} 0, & 1 \leq r \leq l_b, \\ \dfrac{\hat{\pi}}{2} - \dfrac{\hat{\pi}}{2} \dfrac{F(i_p) - F(r)}{F(i_p) - F(l_b)}, & l_b + 1 \leq r \leq i_p, \\ \dfrac{\hat{\pi}}{2} + \dfrac{\hat{\pi}}{2} \dfrac{F(r) - F(i_p)}{F(u_b - 1) - F(i_p)}, & i_p \leq r \leq u_b - 1, \\ 1, & u_b \leq r \leq m, \end{cases}$$

where $F(r)$ denotes the empirical distribution function of the given variable, $\hat{\pi}$ is estimated from a CUB model fitted to the data[1] and l_b (u_b, resp.) is a fixed lower (upper, resp.) bound to threshold the categories corresponding to crisp negative (positive, resp.) scores.

In full generality, the setting of l_b and u_b may be affected by the wording of the scale, the problem under investigation and/or a preliminary analysis of the data. For $l_b = 1$ and $u_b = m$, the membership function (12) corresponds to the totally fuzzy and relative approach given in Cheli and Lemmi (1995). This choice allows us to penalize uniformly each category and it is best-suited for our purpose of accounting for heterogeneity, and thus it is the natural choice for the CUB -Fuzzy proposal.

Definition 1 is a linear spline in the distribution function. Specifically and compared with (5), Definition 1 relies on the CUB uncertainty parameter, but also considers a spline transformation of the empirical distribution function for the item rather than of ordinal categories as in (5). Indeed, measuring distances between categories via their differences may be inappropriate since results depend on the chosen scores: this issue is particularly relevant when the same latent trait is assessed in different groups, locations or times for comparison purposes. Most importantly, it is not necessary to specify spline degrees ϵ, since $\hat{\pi}$ will charge for all the unspecified effects and vagueness of the evaluation, as that derived from the nature of the scale (Iannario 2015).

The rationale and probabilistic genesis behind Definition 1 can be summarized as follows:

(i) the updating of the category r is penalized with the mixing weight for the feeling component $\hat{\pi}$ since it establishes the accuracy of the preference part of the model by adjusting its importance for heterogeneity and diverse sources of imprecision in the data;

(ii) for $r > i_p$ ($r < i_p$, resp.) the frequency of the category r is normalized taking into account the set of positive (negative, resp.) non-crisp choices;

---

[1] $\hat{\pi}$ is replaced by $\hat{\pi}_1$ when *shelter effect* is considered.

(iii) the greater is the heterogeneity (that is, as $\hat{\pi} \to 0$), the less meaningful is the contribution of the relative frequencies to the membership degrees.

The choice of normalizing the updating contribution $F(r) - F(i_p)$ with $F(u_b - 1) - F(i_p)$ for the categories $i_p \leq r \leq u_b - 1$ can be explained as follows: the categories $r \geq u_b$ are certainly associated with membership to A (in our case, A is the set of satisfied users) as $\mu_A(r) = 1$. Symmetric arguments apply to the choice $F(i_p) - F(l_b)$ for lower categories $l_b + 1 \leq r \leq i_p$. Hence, the shades of membership across intermediate positive categories should rather be computed starting from the indifference point and excluding the categories being assigned crisp membership degrees. Moreover, the choice of halving $\hat{\pi}$ and distinguishing between left and right non-crisp sides of the scale is due to weight for the dual contribution of each category to the assessment of membership and non-membership.

The CUB -Fuzzy proposal stems from the central idea of giving to $1 - \hat{\pi}$ a proper definition as measure of fuzziness of the decision process. Thus we assume the range of the IVFS constantly equal to $1 - \hat{\pi}$ for each category r, in agreement with the role that uncertainty plays in CUB models.

Definition 2 For a given item of the questionnaire, the CUB -Fuzzy uncertainty function for A is defined as:

$$u_A(r) = \begin{cases} 0, & 1 \leq r \leq l_b \text{ and } u_b \leq r \leq m, \\ 1 - \hat{\pi}, & l_b + 1 \leq r \leq u_b - 1. \end{cases}$$

From (4), the Fuzzy accuracy function results to be:

$$a(r) = \begin{cases} 1, & 1 \leq r \leq l_b \text{ and } u_b \leq r \leq m, \\ \hat{\pi}, & l_b + 1 \leq r \leq u_b - 1, \end{cases}$$

catching the propensity to assume a meditated response mechanism. Indeed, given the mixture definition, π is a direct indicator of reliability of predictions under the feeling component, which could be adjusted to incorporate also overdispersion (Iannario 2014) or a more general specification (Tutz et al. 2017). Thus, the choice for $u(r) = 1 - \pi$ under the CUB -Fuzzy system implies that the assessment of membership and non-membership of score r is penalized by the unpredictability of responses under the feeling model. In addition, π can be interpreted as a measure of propensity between a well-structured response-behaviour and a random choice: the close $\pi \to 1$, the stronger the frequency distribution can be legitimately used for a fuzzy evaluation system.

From (2), the non-membership function $v_A(r)$ is given by:

$$v_A(r) = \begin{cases} 1, & 1 \leq r \leq l_b, \\ \dfrac{\hat{\pi}}{2} + \dfrac{\hat{\pi}}{2} \dfrac{F(i_p) - F(r)}{F(i_p) - F(l_b)}, & l_b + 1 \leq r \leq i_p, \\ \dfrac{\hat{\pi}}{2} - \dfrac{\hat{\pi}}{2} \dfrac{F(r) - F(i_p)}{F(u_b - 1) - F(i_p)}, & i_p < r \leq u_b - 1, \\ 0, & u_b \leq r \leq m. \end{cases} \tag{11}$$

Let us remark that, as the CUB -Fuzzy uncertainty decreases (that is, the more $\hat{\pi}$ approaches 1), the more the non-membership function (11) increases towards 1 by moving from the indifference point to the first category, and similarly decreasing in the opposite direction of the scale. If the scale orientation is opposite, then the definition of membership and non-membership should be simply switched.

The middle point of the scale is then equally mirrored both in the membership and non-membership scores, as

$$\mu_A(i_p) = \nu_A(i_p) = \frac{\hat{\pi}}{2}.$$

In this way, the indifference expressed by the respondent choosing i_p corresponds to equi-preference of categories, since π is an inverse indicator of heterogeneity. Then, for each rating, the degrees of membership and non-membership are equally split around the indifference point i_p, by halving the weight of the uncertainty parameter. Note that for distributions with low heterogeneity and thus with higher concentration, one has that $\pi \to 1$ and $\mu_A(i_p) = \nu_A(i_p) \to 1/2$, as for the spline approach recalled in Sect. 2.1. In view of the defuzzification procedure, the membership and non-membership degrees are defined in such a way that the accuracy is lower for the items affected by higher heterogeneity, regardless of the level of feeling. Indeed, for increasing heterogeneity (that is, as $\hat{\pi} \to 0$), from Definition 1 and (11) we have:

$$\mu_A(r), \nu_A(r) \to 0, \ u_A(r) \to 1, \ a(r) \to 0, \qquad r = l_b + 1, \ldots, u_b - 1,$$

so that the residual fuzziness $u_A(r)$ increases over the accuracy $a(r)$; accordingly, we are let to negligible membership/non-membership values.

The usage of the empirical distribution function for a Fuzzy evaluation system is also the key of the approach pursued in Zani et al. (2013), which accomplishes a questionnaire analysis in a standard Fuzzy Sets (FS) framework, grounded solely on the membership function:

$$\mu_A(r) = \begin{cases} 0, & 1 \le r \le l_b, \\ \mu_A(r-1) + \dfrac{F(r) - F(r-1)}{1 - F(l_b)}, & l_b < r < u_b, \\ 1, & u_b \le r \le m. \end{cases} \qquad (12)$$

In the forthcoming discussion, this classical FS method will be referred to as the *empirical* Fuzzy system.

3.2 Scoring uncertainty

For the CUB -Fuzzy evaluation system, we propose the IWAM (7) as aggregation index, but with weights $\{w_k\}$ depending on the CUB uncertainty parameter. This choice meets the well-acknowledged recommendation to assign weights that are larger for the more explanatory items, as in Marasini et al. (2015). Here, explanatory is meant as related to accuracy in the assessment of the fuzzy trait and it is inversely related to uncertainty.

Thus, the rationale of the CUB -Fuzzy evaluation system is to penalize items with higher estimated heterogeneity, them being less reliable and explanatory for the assessment of membership and non-membership to A. In this regard, we shall consider the Fuzzy proportion of uncertainty function (1):

$$g(X_k) = \frac{1}{n} \sum_{j=1}^{n} u_A^{(k)}(r_{j,k}), \text{ for } k = 1, \ldots, K. \tag{13}$$

and apply an inverse transform to impute low weight to more uncertain items:

$$w_k = \ln\left(\frac{1}{g(X_k)}\right) \bigg/ \sum_{l=1}^{K} \ln\left(\frac{1}{g(X_l)}\right), \text{ for } k = 1, \ldots, K \tag{14}$$

Here the logarithm transform is taken only to prevent excessive values for very low uncertainty. This weighting scheme has been already used in the Fuzzy Set literature (that is, only with reference to membership functions) (Zani et al. 2012, 2013) to assess the capabilities of each category r in expressing satisfaction across items:

$$\tilde{\mu}_A(r) = \sum_{k=1}^{K} w_k \, \mu_A^{(k)}(r), \quad r = 1, 2, \ldots, m. \tag{15}$$

In that case, the weights have been based on the Fuzzy proportion of the achievement of the target (in our case, respondents' satisfaction):

$$g(X_k) = \frac{1}{n} \sum_{j=1}^{n} \mu_A^{(k)}(r_{j,k}), \text{ for } k = 1, \ldots, K, \tag{16}$$

for which the transformation (14) prevents from giving higher importance to the *rare features* among subjects.

More generally, as we are considering that all items are collected on the same ordinal scale, from Definition 2 with r replaced by $r_{j,k}$, the proportion $g(X_k)$ in (13) has the following closed form.

Proposition 1 *If $\hat{\pi}^{(k)}$ is the estimated CUB uncertainty parameter of the k-th item, then*

$$g(X_k) = \left(1 - \hat{\pi}^{(k)}\right) \left(F^{(k)}\left(u_b^{(k)} - 1\right) - F^{(k)}\left(l_b^{(k)}\right)\right), \text{ for } k = 1, \ldots, K, \tag{17}$$

where $F^{(k)}(\cdot)$ is the empirical distribution function of ratings on the k-th item.

Note that $F^{(k)}\left(u_b^{(k)} - 1\right) - F^{(k)}\left(l_b^{(k)}\right)$ is the percentage of respondents for which $l_b^{(k)} < r_{j,k} < u_b^{(k)}$, thus whose fuzzy evaluation on the k-th item is not crisp.

If $l_b^{(k)} = 1$ and $u_b^{(k)} = m$, Eq. (17) simplifies in the CUB -Fuzzy uncertainty function $\bar{u}^{(k)}$:

$$\bar{u}^{(k)} = \frac{1}{n} \sum_{j=1}^{n} u_A^{(k)}(r_{j,k}) = \left(1 - \hat{\pi}^{(k)}\right) \left(F^{(k)}(m-1) - F^{(k)}(1)\right) \qquad (18)$$

when aggregating the k-th item among respondents. Then the Fuzzy uncertainty score $\bar{u}$ in (9) can be written as:

$$\bar{u} = \sum_{k=1}^{K} w_k \bar{u}^{(k)} = \sum_{k=1}^{K} w_k \left(1 - \hat{\pi}^{(k)}\right) \left(F^{(k)}(m-1) - F^{(k)}(1)\right).$$

In this sense, the CUB uncertainty parameters are given a precise fuzzy interpretation also at the aggregated level.

4 A case study

Fuzzy methods for questionnaire analysis are particularly appealing in evaluation studies. Motivated by this feature, we show how the CUB -fuzzy proposal can be applied on the assessment of satisfaction for the Orientation Services at University of Naples Federico II. The survey was administered from 2002 to 2008 across all the 13 Faculties and aimed at measuring the satisfaction towards the service[2] across different dimensions of the trait. On a balanced $m = 7$ point Likert scale: 1 = 'extremely unsatisfied', 2 = 'very unsatisfied', 3 = 'unsatisfied', 4 = 'indifferent', 5 = 'satisfied', 6 = 'very satisfied', 7 = 'extremely satisfied', the following measurements were collected:

– satisfaction on the acquired information (informat);
– evaluation of the willingness of the staff (willingn);
– adequacy of time-table of opening hours (officeho);
– evaluation of the competence of the staff (compete);
– global satisfaction (global).

The present discussion will concern the data collected in 2002, consisting of $n = 2179$ observations. Motivations for our choice include the fact that the evaluation of University courses, offices and institutions is a popular topic for fuzzy analysis; in addition, the first available wave was chosen since these data allows us to discuss and illustrate all the nuances of the proposal.

In the first part of the section, CUB models are fitted to the data: the estimation procedure takes into account the *shelter effect*, if significant. Then, the CUB -Fuzzy system introduced in Sect. 3.1 is compared with the empirical and the spline approaches recalled in Sects. 3.1 and 2 within the classical and IF settings, respectively.

---

[2] Data are available at http://www.labstat.it/home/research/resources/cub-data-sets-2/.

Table 1 Parameter estimates: CUB model no shelter effect (standard errors in parentheses)

	informat	willingn	officeho	compete	global
$\hat{\pi}$	0.794 *(0.016)*	0.857 *(0.012)*	0.680 *(0.019)*	0.802 *(0.015)*	0.868 *(0.013)*
$\hat{\xi}$	0.181 *(0.005)*	0.117 *(0.004)*	0.197 *(0.006)*	0.164 *(0.005)*	0.171 *(0.004)*

4.1 CUB models estimation

The ML estimates of CUB parameters for the chosen data are summarized in Table 1. Overall, there is a moderate level of uncertainty and an extreme positive feeling across the items, the highest satisfaction being expressed for `willingn` $(1 - \hat{\xi} = 0.8833)$, the lowest for `officeho` $(1 - \hat{\xi} = 0.8029)$. However, there are certain items for which uncertainty is not negligible and an evaluation system should properly consider these differences. In particular, `officeho` is the item with the highest estimated uncertainty $(1 - \hat{\pi} = 0.3198)$, followed by `informat` and then by `compete`.

Table 2 shows the estimation results when including shelter effects: globally, the previous comments continue to hold, but uncertainty is disclosed in more details and feeling estimates are corrected. In particular, notice that `informat` has a different shelter category $(c = 5)$ compared with all other items (for which $c = 7$). Despite the shelter category is the same for the last four items, its effect is more prominent for `willingn` $(\hat{\delta} = 0.194)$, associated also with the lowest weight of the Uniform distribution $(\hat{\pi}_2 = 0.123)$. Instead, even if significant, the *shelter effect* within `global` is the weakest; moreover, this item corresponds to the strongest attitude towards a more meditated choice, resulting in the highest level of accuracy $(\hat{\pi}_1 = 0.828)$. For the sake of brevity, statistical results on the model selection (based on the BIC criterion) are skipped and are available on demand.

The rationale of the CUB -Fuzzy evaluation system is to provide statistical models for rating data with some veracity analytics in the spirit of a fuzzy analysis of questionnaire. Then, the proposal is not advanced to be a direct competitor of ordinary techniques; nevertheless, it is worth to notice how its performances match with outcomes of standard procedures.

In this respect, the first two components as obtained from a PCA are sufficient for our purposes since they account for more than 80% of the total variability (Fig. 1). Results confirm that item `officeho` plays a distinctive role in the assessment of the latent satisfaction, thus it should be properly discriminated and weighted.

Notice that Likert-scale categories are ordered measurements that can be thought of as cutpoints of a latent continuum. Thus, it is not always adequate to fit a PCA or Factor analysis directly on a data matrix like that from a rating survey, unless a proper correlation is obtained. In this regard, polychoric correlation is a validated choice especially when the number of categories is moderate.

Table 2 Parameter estimates: CUB model with shelter at category c (standard errors in parentheses)

|	`informat` ($c = 5$)	`willingn` ($c = 7$)	`officeho` ($c = 7$)	`compete` ($c = 7$)	`global` ($c = 7$)
$\hat{\pi}^{\star}$	0.749 (0.017)	0.848 (0.012)	0.705 (0.022)	0.804 (0.016)	0.869 (0.013)
$\hat{\xi}$	0.153 (0.006)	0.157 (0.005)	0.265 (0.009)	0.199 (0.008)	0.183 (0.007)
$\hat{\delta}$	0.085 (0.014)	0.194 (0.011)	0.145 (0.014)	0.117 (0.020)	0.048 (0.022)
$\hat{\pi}_1$ (Accuracy)	0.685 (0.021)	0.683 (0.015)	0.602 (0.019)	0.710 (0.021)	0.828 (0.022)
$\hat{\pi}_2$	0.230 (0.015)	0.123 (0.012)	0.252 (0.019)	0.173 (0.015)	0.125 (0.013)
$1 - \hat{\pi}_1$ (Uncertainty)	0.315	0.317	0.398	0.290	0.172
$1 - \hat{\xi}$ (Feeling)	0.847	0.843	0.735	0.802	0.817
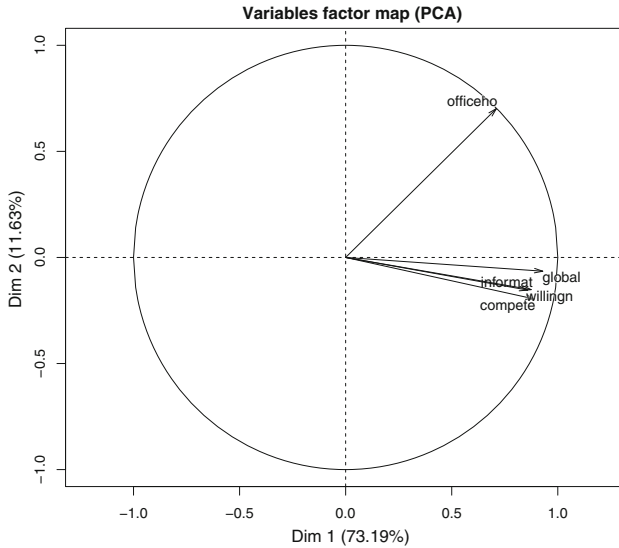
Fig. 1 PCA: Variable factor map

4.2 Classical FS: CUB-fuzzy versus empirical

Figure 2 plots the membership function of the CUB -Fuzzy evaluation system (1) against the empirical one (12) for each item.

For the CUB -Fuzzy approach, the higher the value of $\hat{\pi}_1$ is, the faster the membership degrees increase moving from the indifference point to the maximum of the scale. In addition, notice that the two methods behave quite differently, especially for willingn and global having the lowest estimated heterogeneity ($\hat{\pi}_2 = 0.123$ and $\hat{\pi}_2 = 0.125$, respectively): this indicates a more prominent attitude of the CUB -Fuzzy membership function to correctly discriminate among different levels of heterogeneity, also when moderate.

In order to aggregate membership values, Table 3 shows the two systems of weights $\{w_k\}$ employable in computing the aggregator index (15): we refer to (14) paired with (13) and with (16) for the CUB -Fuzzy model (dotted line) and for the empirical model (solid line), respectively. For comparative purposes, normalized variables loadings derived for the first principal component (PCA1) are also reported.

For the CUB -Fuzzy evaluation system, even if at aggregated level results do not substantially vary at aggregated level for different weights (see Sect. 4.4), it turns out that weights based on (16) do not suitably penalize officeho and informat having a weak importance due to the highest observed uncertainty among the items (for example $1 - \hat{\pi}_1 = 0.398$ for officeho in Table 2). Instead, for the weights based on (13), the lowest value is attained exactly for officeho ($w_3 = 0.154$). Notice that willingn is assigned a higher weight than informat (w.r.t. the CUB -Fuzzy system of weights), though it shows a higher uncertainty and a lower feeling, comparatively. This is explained by willingn having the most prominent shelter effect at $c = 7$, indicating a strong tendency of the distribution to be concentrated at
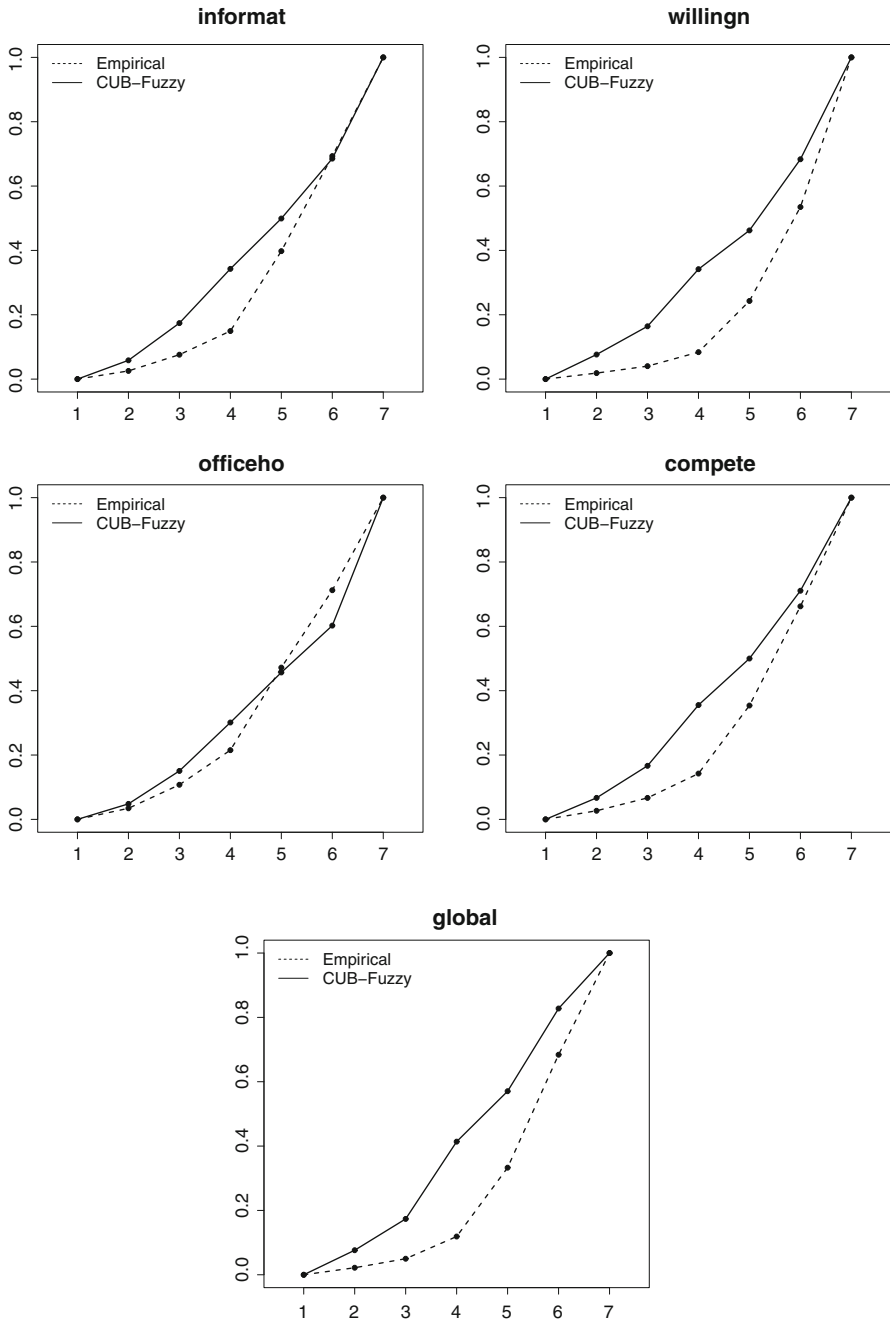
Fig. 2 Comparison of membership functions for the CUB -Fuzzy model (solid line) versus the empirical model (dashed line)

Table 3 Weights systems

	informat	willingn	officeho	compete	global
w_k based on (16)	0.216	0.152	0.287	0.196	0.146
w_k based on (13)	0.183	0.212	0.154	0.198	0.254
PCA1	0.209	0.201	0.138	0.216	0.236

higher categories. Instead, the shelter at $c = 5$ for `informat` acts by deflating the weight of importance (to assess membership, non-membership, etc.) since it is closer to the indifference point and thus tends to penalize a positive evaluation of satisfaction.

Figure 3 shows how the CUB -Fuzzy system scales membership at aggregated level more coherently when compared to the expressed `global` satisfaction. Specifically, we have run both a CUB -Fuzzy evaluation and an empirical system on the first 4 items, omitting `global` satisfaction. Then, we have stratified the aggregated membership values for the two fuzzy systems across increasing level of `global` satisfaction (by choosing weights accordingly: the inverse fuzzy proportion of uncertainty for the CUB -Fuzzy proposal and the inverse fuzzy proportion of membership for the empirical one). Then, read on the y-axis, it appears that the CUB -Fuzzy proposal is more convincing in aggregating the information withdrawn from the first four items if used as a proxy of the global satisfaction. In other words, aggregated values for membership under the CUB -Fuzzy scheme are increasingly more consistent with increasing level of global satisfaction.

4.3 IFS: CUB-fuzzy versus spline

Without any prior assessment of experts on specific values for the parameters, membership (5), fuzzy uncertainty (6) and non-membership function (2) will be equal for all items, as shown in Table 4, where $\epsilon = 1$, $a = 1$, $b = m - 1$ and $\theta = \eta$ have been set to account for the balanced scale as recommended by Marasini et al. (2015). In particular, in the following we have tested the CUB -Fuzzy proposal against the spline uncertainty with $\theta = \eta = 1$ due to its interpretation as a risk measure: indeed, in this case the spline uncertainty fuzzy function corresponds to the variance of a Bernoulli random variable whose success trial (the membership to A) has probability of occurrence set to $\mu_A(r)$. Dually, the fuzzy uncertainty prescribed by the CUB -Fuzzy system accounts for risk in terms of heterogeneity. In Fig. 4, such spline membership and non-membership values are compared with those obtained with the CUB -Fuzzy proposal. Observe that the membership given in Definition 1 is more accurate and naturally shaped to the data. What is constant over categories in the CUB -Fuzzy model is the Fuzzy uncertainty function in Definition 2 (that is $1 - \hat{\pi}_1$ in Table 2). Indeed $1 - \hat{\pi}_1$ (equivalently, $1 - \hat{\pi}$ when *shelter effect* is not significant) measures the overall *fuzziness*, independently from the membership and non-membership degrees; dually, $\hat{\pi}_1$ quantifies the level of accuracy of the Fuzzy evaluation system. This feature is only partially accomplished by the spline method as the Fuzzy uncertainty function has symmetric values around the indifference point, see Table 4. Notice that in general,
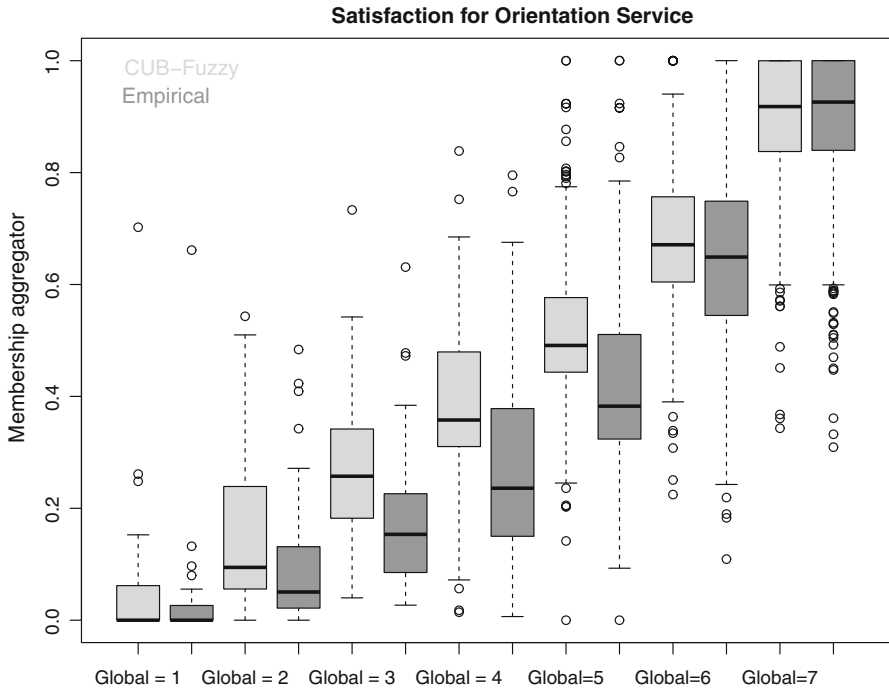
Satisfaction for Orientation Service

Fig. 3 Boxplots of aggregated membership for increasing levels of `global` satisfaction: comparison between the CUB -Fuzzy (light grey) and Empirical system (dark grey)

Table 4 Spline fuzzy functions

	$R = 1$	$R = 2$	$R = 3$	$R = 4$	$R = 5$	$R = 6$	$R = 7$
Membership degree	0	0.10	0.30	0.50	0.70	0.90	1
Non-membership degree	1	0.81	0.49	0.25	0.09	0.01	0
Uncertainty degree	0	0.09	0.21	0.25	0.21	0.09	0

the uncertainty function of an IF evaluation system is set in such a way that it attains a maximum at the indifference category if available. Instead, we consider the uncertainty as uniformly spread along the scale, so that it can be considered as a feature of the item and not of a single category.

Comparisons between the two methods based, for example, on the composite indicator (15) are meaningless since for the spline one $\tilde{\mu}_A(r) = \mu_A(r) = \mu_A^{(k)}(r)$ for all k. For this reason, we propose to aggregate the k-th item uniformly across respondents, achieving a complete IFS evaluation system. More specifically, by keeping the notation introduced in Sect. 3.2, for the membership function we compute $\bar{\mu}^{(k)} = \frac{1}{n} \sum_{j=1}^{n} \mu_A^{(k)}(r_{j,k})$ both for the (linear) spline and the CUB -Fuzzy evaluation systems, see Table 5. The same is done for the non-membership functions as well as the Fuzzy score and accuracy measures. As we see from Table 5, the spline approach
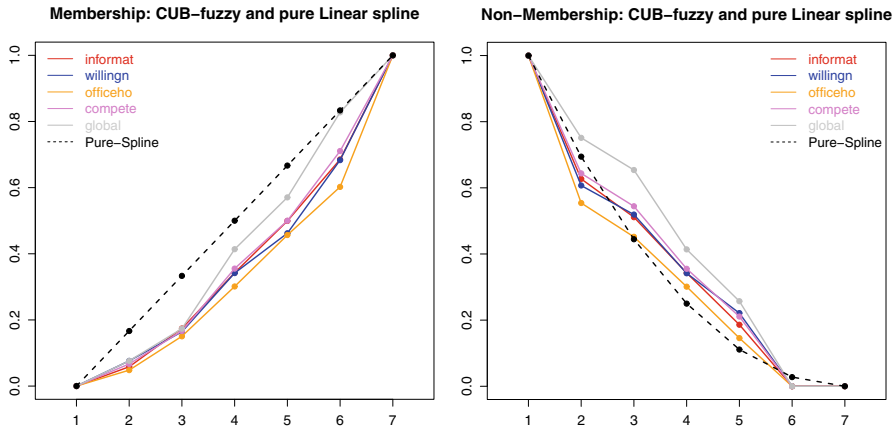
Fig. 4 CUB-Fuzzy and spline membership/non-membership functions

Table 5 Fuzzy functions aggregated per item

Function	Method\Item:	informat	willingn	officeho	compete	global
Membership	CUB-Fuzzy	0.651	0.743	0.571	0.681	0.752
	Spline	0.780	0.853	0.733	0.797	0.814
Non-membership	CUB-Fuzzy	0.136	0.091	0.157	0.131	0.132
	Spline	0.113	0.072	0.151	0.104	0.086
Fuzzy uncertainty	CUB-Fuzzy	0.212	0.166	0.272	0.188	0.116
	Spline	0.107	0.075	0.116	0.099	0.100
Fuzzy score	CUB-Fuzzy	0.515	0.652	0.414	0.550	0.620
	Spline	0.667	0.781	0.581	0.693	0.728
Fuzzy accuracy	CUB-Fuzzy	0.788	0.834	0.728	0.812	0.884
	Spline	0.893	0.925	0.884	0.901	0.900

does not sufficiently discriminate the different levels of uncertainty among the items, yielding to a narrow range for both the fuzzy accuracy and uncertainty. Conversely, the CUB-Fuzzy proposal offers a major flexibility in grading Fuzzy indicators according to the observed uncertainty. In particular, we stress that the accuracy of the CUB-Fuzzy approach is penalized for items with higher global uncertainty in the sense of CUB models, while it increases for items corresponding to a weaker indeterminacy. For instance, with reference to Table 2, the maximum estimated overall indeterminacy corresponds to officeho ($1 - \hat{\pi}_1 = 0.398$), whereas the minimum corresponds to global ($1 - \hat{\pi}_1 = 0.172$). As a result, these items coherently are assigned the minimum and maximum levels of the accuracy, respectively, while the spline model is more restrictive in accounting for this variability.

Figures 5 and 6 show how the CUB-Fuzzy system scales -on the *y* axis- membership at aggregated level in a comparable way as the linear spline conditional on increasing scores for global satisfaction, whereas it is more adequate than the quadratic spline
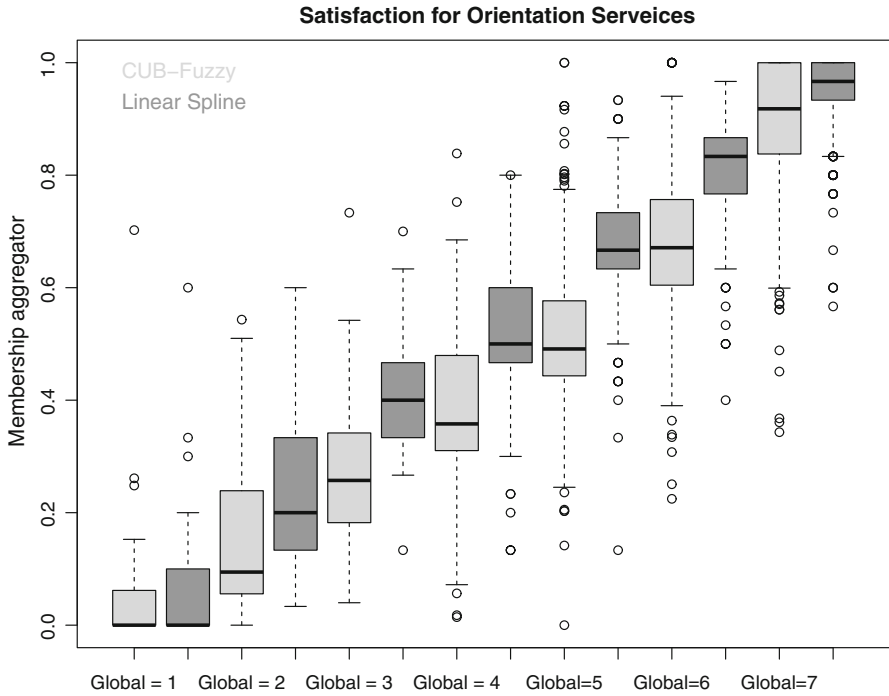
Fig. 5 Boxplots of aggregated membership for increasing levels of `global` satisfaction: comparison between the CUB -Fuzzy (light grey) and linear spline systems (dark grey)

system. Spline memberships have been aggregated with uniform weights across items. Comparable results are obtained if normalized variable loadings derived from factor analysis are considered.

Finally, the ultimate step of the Fuzzy evaluation procedure is to provide a measure of the overall uncertainty by aggregating the functions given in Table 5 among items: as example, for the membership function we compute $\bar{\mu} = \sum_{k=1}^{K} w_k \bar{\mu}^{(k)}$ with $\bar{\mu}^{(k)}$ given in Table 5 (first row); similarly, for all fuzzy functions. For the CUB -Fuzzy method we employ the weights (14) paired with (13). For the spline method, instead, we shall consider uniform weights also across the items (Marasini et al. 2015). The resulting Fuzzy composite aggregators are reported in Table 6.

Thus, we can conclude that the proposed evaluation system based on CUB models is safer in assigning fuzzy values, being designed to account for heterogeneity and stylistic responses in the data. Nevertheless, it does not miss to provide a global positive picture (in terms of membership and accuracy).

In conclusion, in order to disclose the different perspectives offered by the CUB -Fuzzy analysis of questionnaire, we combined a two-component PCA analysis with a k-means clustering on the data-matrix according to a so-called *tandem scheme* (Arabie and Hubert 1994). Specifically, we compared the closeness of the derived classification with that of a k-means algorithm applied to the IWAM aggregators (7): $k = 5$ was set for the k-means algorithm to identify certainly unsatisfied ($R = 1$), fairly unsatisfied
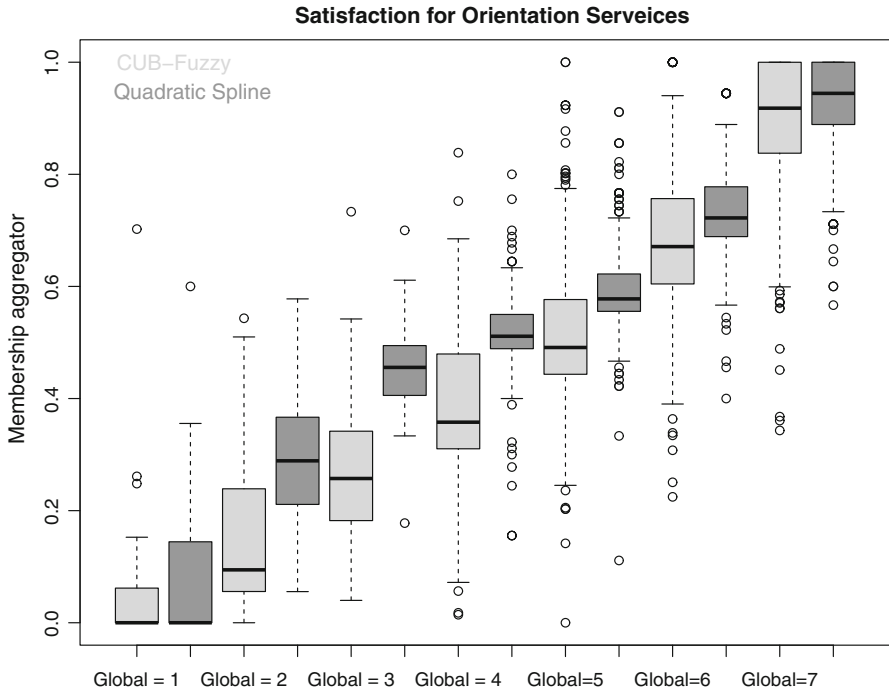
Fig. 6 Boxplots of aggregated membership for increasing levels of `global` satisfaction: comparison between the CUB -Fuzzy (light grey) and quadratic spline systems (dark grey)

Table 6 Fuzzy composite indicators (aggregation of items)

	Weights	Membership	Non-membership	Uncertainty	Score	Accuracy
CUB -Fuzzy	Log inverse of Fuzzy unc.	0.690	0.128	0.182	0.562	0.818
Spline	Uniform weights	0.795	0.105	0.099	0.690	0.900

Table 7 Cohen's κ: agreement between the classification obtained from k-means on the first two PCA components and that obtained from k-means on membership and non-membership aggregators (7) for different fuzzy methods

	Lower	Estimate	Upper
CUB -Fuzzy	0.45	0.49	0.53
Linear Spline	0.33	0.36	0.38
Quadratic Spline	0.18	0.21	0.24

($R = 2, 3$), indifferent ($R = 4$), fairly satisfied ($R = 5, 6$), and certainly satisfied ($R = 7$) respondents. Table 7 reports the Cohen's κ measure to assess agreement between the two corresponding classifications, along with lower and upper confidence bounds:

Results do not substantially vary if considering the polychoric correlation to run PCA. Notice that fuzzy clustering has a precise meaning in the literature (Everitt et al. 2011), which is not involved in the present analysis: this perspective will be the subject of future investigation.

4.4 Sensitivity analysis

A sketch of sensitivity analysis is here accomplished to validate the CUB -Fuzzy proposal against the other fuzzy alternatives considered.

First, read top to bottom, Figs. 7, 8 and 9 display membership and non-membership functions for decreasing levels of heterogeneity and for right-tailed, symmetric and left-tailed rating distributions generated from varying CUB distributions: $\xi = 0.8, \xi = 0.5, \xi = 0.1$, respectively, and for each of them $\pi = 0.2$ (top), $\pi = 0.4, \pi = 0.6, \pi = 0.8$ (bottom). Imagine that the distributions correspond to measurements on a scale 1="*extremely dissatisfied*" up to 7="*completely satisfied*". Then Figs. 7, 8 and 9 correspond to overall dissatisfied, overall neutral and overall satisfied respondents, respectively. It appears evident that the CUB -Fuzzy proposal, being naturally shaped to the data, is safer and more integrated with the observed scores. In addition, for data with low heterogeneity (the bottom panels), it is globally intermediate between the linear and quadratic splines. The price of estimation–which can be promptly run by means of the R package 'CUB' (Iannario et al. 2018)–is compensated with no need of prior choice for spline degrees and parameters, and with a more flexible and versatile tool for uninformative circumstances.

Secondly, we assess distances between aggregated Intuitionistic Fuzzy sets obtained with different weighting and fuzzy systems. Specifically, the (normalized) Hamming distance between IFS sets is computed (Szmidt and Kacprzyk 2000). Briefly, if

$$B = \{< x, \mu_B(x), \nu_B(x) > | x \in X\}, \qquad C = \{< x, \mu_C(x), \nu_C(x) > | x \in X\}$$

are two IF evaluation systems defined for the universe of discourse X, then the (normalized) Hamming distance between B and C is defined as:

$$d_H(B, C) = \frac{1}{2n} \sum_{i=1}^{n} \left(|\mu_B(x_i) - \mu_C(x_i)| + |\nu_B(x_i) - \nu_C(x_i)| + |u_B(x_i) - u_C(x_i)| \right),$$

with n being the number of observations. Back to our case study, in order to support the consistency of the CUB -Fuzzy proposal, we will compute the distance between the aggregated IWAM fuzzy sets (7) corresponding to different choices of the weighting system. For $l = 1, 2$, let $\boldsymbol{w}^{(l)} = \{w_1^{(l)}, \ldots, w_K^{(l)}\}$ be two alternative choices for weights for which:

$$\tilde{\mu}_A^{(l)}(\boldsymbol{r}_i) = \sum_{k=1}^{K} w_k^{(l)} \mu_A^{(k)}(r_{i,k}), \qquad \tilde{\nu}_A^{(l)}(\boldsymbol{r}_i) = \sum_{k=1}^{K} w_k^{(l)} \nu_A^{(k)}(r_{i,k})$$
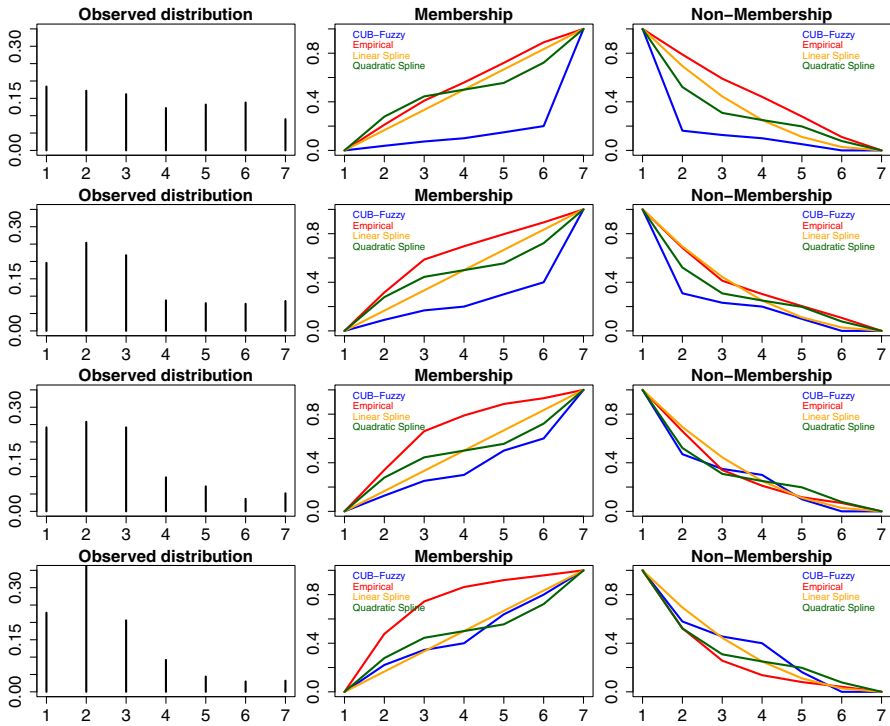
Fig. 7 Comparison of membership and non-membership for right-tailed rating distributions ($\xi = 0.8$) corresponding to decreasing levels of heterogeneity: $\pi = 0.2$ (top) up to $\pi = 0.8$ (bottom)

are the corresponding IWAM for membership and non-membership functions across K items. Then, let:

$$B = \{< \boldsymbol{r}_i, \tilde{\mu}_A^{(1)}(\boldsymbol{r}_i), \tilde{v}_A^{(l)}(\boldsymbol{r}_i) > |i = 1, \ldots, n\}$$
$$C = \{< \boldsymbol{r}_i, \tilde{\mu}_A^{(2)}(\boldsymbol{r}_i), \tilde{v}_A^{(2)}(\boldsymbol{r}_i) > |i = 1, \ldots, n\}.$$

As reported in Table 8, the restrained values of the distance support the relative indifference between the fuzzy proportion of uncertainty and membership values according to (14) for the CUB -Fuzzy Proposal, and with weights derived from PCA-type procedures. Nevertheless, we acknowledge that weights based on the fuzzy proportion of uncertainty are the most natural and always applicable choice for CUB -Fuzzy systems, and that weights based on PCA-type are an acceptable option only if the first component explains an appreciable amount of variability. In addition, some of the CUB -Fuzzy indicators are stable with respect to the reversion of the scale, that is for samples $m - r_{j,k} + 1$ with $j = 1, \ldots, n$ and $k = 1, \ldots, K$. Indeed, the CUB random variable R is reversible: if $R \sim \mathrm{CUB}\,(\pi, \xi)$ over the m categories, then $m - R + 1 \sim$ CUB $(\pi, 1 - \xi)$ (D'Elia and Piccolo 2005). If the scale is balanced and $R \sim \mathrm{CUB}\,(\pi, \xi)$, with shelter at c measured by δ, then $m - R + 1 \sim \mathrm{CUB}\,(\pi, 1 - \xi)$ with shelter at $m - c + 1$ measured by δ as well. Due to this property, for reversed ratings the CUB -Fuzzy accu-
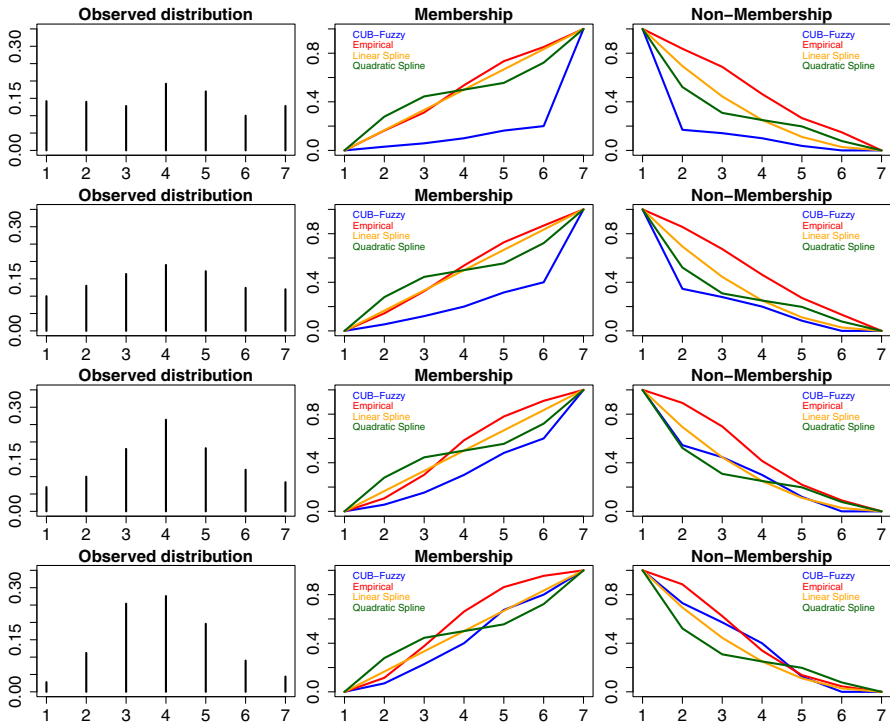
Fig. 8 Comparison of membership and non-membership for symmetric rating distributions ($\xi = 0.5$ for decreasing levels of heterogeneity: $\pi = 0.2$ (top) up to $\pi = 0.8$ (bottom)

racy and uncertainty functions remain unchanged, differently from the membership and non-membership degrees, being dependent on the distribution functions. Instead, the spline Fuzzy functions would change only at an aggregated level. In addition, also weights (14) paired with (13) are invariant, being directly related to CUB uncertainty parameters. Conversely, if paired with the Fuzzy proportion of membership degrees (16), such weights would vary.
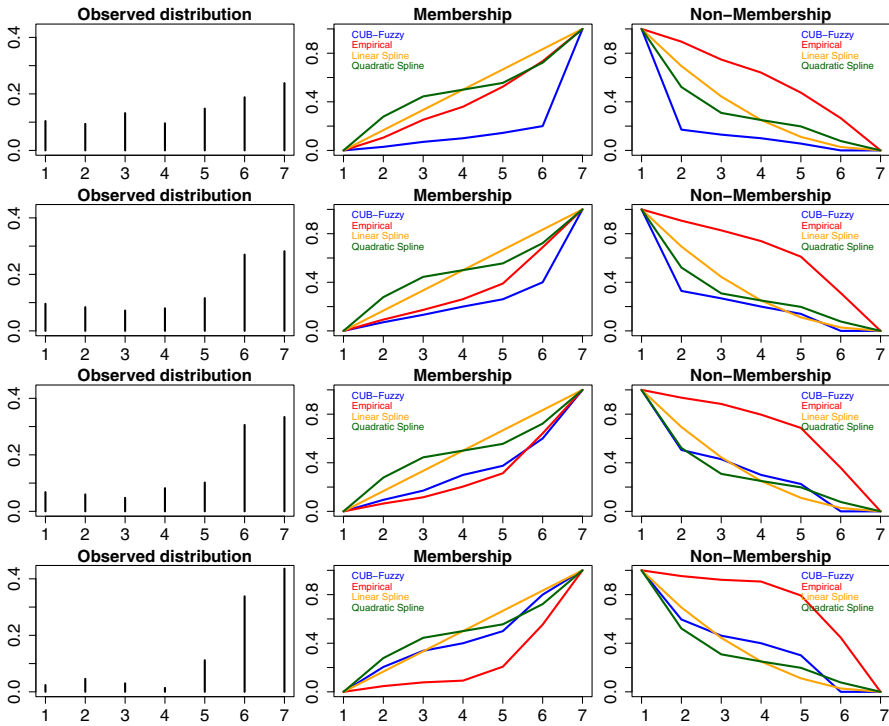
Fig. 9 Comparison of membership and non-membership for left-tailed rating distributions ($\xi = 0.1$) for decreasing levels of heterogeneity: $\pi = 0.2$ (top) up to $\pi = 0.8$ (bottom)

Table 8 Normalized Hamming distance between IWAM aggregators (7) of the CUB-Fuzzy system with different weights	$d_H(B, C)$	$\boldsymbol{w}^{(2)}$		
	$\boldsymbol{w}^{(1)}$	(14) with (16)	PCA loadings	Uniform
	(14) with (13)	0.028	0.006	0.011

5 Comments and conclusions

The present contribution fosters the application of CUB mixture models for ordinal rating data to account for uncertainty of choices within a fuzzy analysis of questionnaires. The resulting CUB -Fuzzy procedure is well-suited for broad applications since it allows to deal with veracity of rating by means of the assessment of membership and non-membership grades.

From the statistical modelling point of view, the proposal sheds new light on the CUB paradigm, conveying its vague definition of uncertainty into a precise frame. From the point of view of fuzzy analysis, it roots membership and non-membership assessments on the basis of sound statistical procedures, thus fuzzy functions gain reliability. Moreover, the proposal is build in such a way that the data structures and relations are preserved, and -when pertinent- results match with traditional methods.

The proposal stems from the spline approach introduced in Marasini et al. (2015), suitably adjusted with the CUB uncertainty parameter to let the spline approach be more insightful and freed of subjectivity of parameters choice. The methodology here introduced meets also a classical proposal based on the empirical distribution function (Cheli and Lemmi 1995; Zani et al. 2013), which is adjusted within a general IFS framework. The procedure provides a refined tool to account for heterogeneity and other forms of nuisances, as meant by the rationale of CUB uncertainty. In particular, the CUB model uncertainty $1 - \pi$ is equal to the Fuzzy hesitancy level for each item, and the accuracy function results to be more sensitive to different sources of indeterminacy as heterogeneity and *shelter effect* (see Table 6). As a result, the CUB uncertainty measure is validated as an effective Fuzzy composite indicator.

Summarizing, the spline methods for fuzzy analysis of questionnaire is a valuable methodology, whose main criticisms are the subjectivity of choices and the lack of a statistical foundation, where diversity in scale-usages needs to be taken into account. The CUB -Fuzzy system overcomes these pitfalls by defining fuzzy functions anchored to the empirical distribution functions and adjusted for uncertainty in the data. Encoding the uncertainty degrees implies that the resulting CUB -Fuzzy system is freed of subjectivity of parameters values for spline functions, without the need of choosing between a linear or a quadratic splines since the vagueness of responses induced by the scale is automatically charged by the uncertainty parameter. In addition, it is grounded on ML estimation and in this sense it is more robust. In the same vein, in order to let the CUB -Fuzzy analysis of questionnaire adhere to respondents' subjectivity and not to that of scholars and judges, the uncertainty parameter can be estimated on subjective basis (π_i) and linked to responses drivers (covariates Y_i) by means of a logistic transform:

$$logit(\pi_i) = \beta_0 + Y_i \, \boldsymbol{\beta}.$$

Future developments in this direction will be adressed to take into account the Hesitant Fuzzy Set framework too, as proposed in Marasini et al. (2015) and Torra (2010). Nevertheless, it is worth to underline that the CUB -fuzzy proposal can be enriched by the specification of covariates to disclose response profiles, but it is valid *per se*, conversely to some other traditional methods.

In conclusion, let us underline that the proposal does not advance a brand new statistical model, rather it is tailored to boost the idea that any statistical model should be prone to offer some veracity analytics of data. Traditional models are able to discriminate sharply between satisfied and dissatisfied respondents, at the price of more involved model-specifications and no measure of uncertainty (and, dually, no measure of accuracy/reliability of ratings). Conversely, the proposal offers a multifaceted interpretation of results, with both *local* modelling (as for cumulative and partial credit models, for instance), and global assessments (feature that is inherited by CUB models). For instance, it allows discrimination of items in terms of their capabilities of identifying satisfied and unsatisfied respondents (and, in full generality, members and not-members of latent classes). This advantage can be appreciated the more data are heterogeneous.

References

Andreis F, Ferrari PA (2013) On a copula model with CUB margins. QdS J Methodol Appl Stat 15:33–51

Arabie P, Hubert L (1994) Cluster analysis in marketing research. In: Bagozzi RP (ed) Advanced methods of marketing research. Blackwell, Oxford, pp 160–189

Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20:87–96

Atanassov KT (2012) On intuitionistic fuzzy sets theory. Springer, New York

Atanassov KT, Gargov G (1989) Interval valued intuitionistic fuzzy sets. Fuzzy Sets Syst 31:343–349

Beliakov G, Bustince H, Goswamic DP, Mukherjeed UK, Pale NR (2011) On averaging operators for Atanassov's intuitionistic fuzzy sets. Inf Sci 181:1116–1124

Capecchi S, Simone R (2018) A proposal for a model-based composite indicators: experience on perceived discrimination in Europe. Soc Indic Res https://doi.org/10.1007/s11205-018-1849-9

Capecchi S, Endrizzi I, Gasperi F, Piccolo D (2016) A multi-product approach for detecting subjects' and objects' covariates in consumer preferences. Br Food J 118:515–526

Cheli B, Lemmi A (1995) A totally fuzzy and relative approach to the multidimensional analysis of poverty. Econ Notes 24:115–134

Colombi R, Giordano S (2016) A class of mixture models for multidimensional ordinal data. Stat Model 16:322–340

Corduas M (2015) Analyzing bivariate ordinal data with CUB margins. Stat Model 15:411–432

D'Elia A (2000) A shifted Binomial model for rankings. In: Proceedings of the 15th international workshop on statistical modelling. New Trends in Statistical Modelling (IWSM), Bilbao, Spain, pp 412–416

D'Elia A, Piccolo D (2005) A mixture model for preference data analysis. Comput Stat Data Anal 49:917–934

Di Nardo E, Simone R (2016) CUB models: a preliminary fuzzy approach to heterogeneity. In: Pratesi M, Perna C (eds) Proceedings of the 48th scientific meeting of the Italian Statistical Society (ISBN: 9788861970618), pp 1–10

Everitt BS, Landau S, Leese M, Stahl D (2011) Cluster analysis, 5th edn. Wiley series in probability & statistics. Wiley, Chichester

Gottard A, Iannario M, Piccolo D (2016) Varying uncertainty in CUB models. Adv Data Anal Classif 10(2):225–244

Iannario M (2012) Modelling shelter choices in a class of mixture models for ordinal responses. Stat Methods Appl 21:1–22

Iannario M (2014) Modelling uncertainty and overdispersion in ordinal data. Commun Stat Theory Methods 43(4):771–786

Iannario M (2015) Modelling scale effects and uncertainty in rating surveys. EJASA 8(3):329–345

Iannario M, Piccolo D (2012) CUB models: statistical methods and empirical evidence. In: Kenett RS, Salini S (eds) Modern analysis of customer surveys: with applications using R. Wiley, Chichester, pp 231–258

Iannario M, Piccolo D (2016) A generalized framework for modelling ordinal data. Stat Methods Appl 25:163–189

Iannario M, Piccolo D, Simone R (2018) CUB: a class of mixture models for ordinal data. (R package version 1.1.2). http://CRAN.R-project.org/package=CUB

Lalla M (2005) Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity. Qual Quant 38(5):577–601

Manisera M, Zuccolotto P (2014) Modeling rating data with non linear CUB models. Comput Stat Data Anal 78:100–118

Marasini D, Quatto P, Ripamonti E (2015) Intuitionistic Fuzzy sets in questionnaire analysis. Qual Quant 50:767–790

Marasini D, Quatto P, Ripamonti E (2016) Fuzzy analysis of students' ratings. Eval Rev 40:122–141

Marasini D, Quatto P, Ripamonti E (2017) Inferential confidence intervals for fuzzy analysis of teaching satisfaction. Qual Quant 51(4):1513–1529

Piccolo D (2006) Observed information matrix for MUB models. Quad Stat 8(1):33–78

Simone R, Tutz G (2018) Modelling uncertainty and response styles in ordinal data. Stat Neerl 72(3):224–245

Szmidt E, Kacprzyk J (2000) Distance between intuitionistic fuzzy sets. Fuzzy Sets Syst 114:505–518

Torra V (2010) Hesitant fuzzy sets. Int J Intell Syst 25:529–539

Tutz G, Schneider M, Iannario M, Piccolo D (2017) Mixture models for ordinal responses to account for uncertainty of choice. Adv Data Anal Classif 11(2):281–305

Van Leekwijck W, Kerre EE (1999) Defuzzification: criteria and classification. Fuzzy Sets Syst 108:159–178

Xu Z (2007) Intuitionistic fuzzy aggregation operators. IEEE Trans Fuzzy Syst 15(6):1179–1187

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Zani S, Milioli MA, Morlini I (2012) Fuzzy methods and satisfaction indices. In: Kenett RS, Salini S (eds) Modern analysis of customer surveys: with applications using R. Wiley, Chichester, pp 439–456

Zani S, Milioli MA, Morlini I (2013) Fuzzy composite indicators: an applications for measuring customer satisfaction. In: Torelli N, Pesarin F, Bar-Hen A (eds) Advances in theoretical and applied statistics. Springer, Heidelberg, pp 243–253