Stat Methods Appl (2016) 25:163–189
DOI 10.1007/s10260-015-0316-9

A generalized framework for modelling ordinal data

Maria Iannario[1] · **Domenico Piccolo**[1]

Accepted: 23 April 2015 / Published online: 12 May 2015

Abstract In several applied disciplines, as Economics, Marketing, Business, Sociology, Psychology, Political science, Environmental research and Medicine, it is common to collect data in the form of ordered categorical observations. In this paper, we introduce a class of models based on mixtures of discrete random variables in order to specify a general framework for the statistical analysis of this kind of data. The structure of these models allows the interpretation of the final response as related to feeling, uncertainty and a possible *shelter* option and the expression of the relationship among these components and subjects' covariates. Such a model may be effectively estimated by maximum likelihood methods leading to asymptotically efficient inference. We present a simulation experiment and discuss a real case study to check the consistency and the usefulness of the approach. Some final considerations conclude the paper.

Keywords Ordinal data · Rating survey · CUB models · *Shelter* choices · G*e*CUB models

1 Introduction

In several applied researches data are collected as categorical ordinal observations. Sometimes they are genuine ordered assessments (judgements, preferences, degree of

✉ Domenico Piccolo
domenico.piccolo@unina.it

Maria Iannario
maria.iannario@unina.it

[1] Department of Political Sciences, University of Naples Federico II, Via Leopoldo Rodinò 22, 80138 Naples, Italy

adhesion to a sentence, etc.) whereas in other circumstances they are categorized for convenience (age of people in classes, measures of objects in block of constant size, education achievement, levels of blood pressure for classifying heart health status, etc.). In both cases, an effective statistical analysis should take the ordinal nature of the responses into account, as discussed by Agresti (2010), Powers and Xie (2000), Tutz (2012), among others. Although the results of this paper may be loosely applied in any context where ordinal data and subjects' covariates are involved, it is more immediate to focus the subsequent discussion in case of rating surveys.

Different lines of attack to the problem have been raised in the literature and some of them stem from the well known historical debate between Pearson and Yule: the main distinctions lie in considering ordinal data as generated by a latent continuous variable or as an intrinsically discrete phenomenon. In fact, the boundary line between these two approaches is not so sharp as it is evident when we face with logistic regression which can be safely introduced within the logic of both paradigms.

In the last decades, several contributions have been proposed and the leading trend is to convey the statistical analysis of ordinal data to the Generalized Linear Models (GLM) framework as proposed by McCullagh (1980) and deepened by Nelder and Wedderburn (1972), McCullagh and Nelder (1989) and extended with several variants by Peterson and Harrell (1990) and Cox (1995), among others. According to this line of reasoning, we model the probability of a response not superior to a given category as a function of selected covariates; in fact, the distribution function induces an ordered constraint among the categories.

An alternative approach, mainly motivated by the investigation of respondents' psychology, has been introduced by Piccolo (2003) and D'Elia and Piccolo (2005) and consists in the so-called CUB models. They have been successfully applied in several fields since they allow for easy interpretation and visualization of the estimation results, and also for designing profiles and specifying clusters of respondents (Corduas 2008a, b, 2011). Then, these models have been extended in several directions and form the basis of the generalization we will pursue in this paper according to the suggestions of Corduas et al. (2009) and the analysis of Iannario (2012a) who support the introduction of a further component, denoted as *shelter effect*. The novelty of the class of models, discussed hereafter and denoted as GeCUB, is the ability to estimate the effect of subjects' covariates for all the components of the extended mixture.

The paper is organized as follows: in the next section, we set notations and motivations for the model whereas in Sect. 3 GeCUB models are specified and their usage is emphasized. Then, the main derivation of the maximum likelihood (ML) estimators is outlined in Sect. 4. A limited simulation experiment is performed in Sect. 5 to confirm the main properties of the ML procedures for finite sample sizes. Section 6 investigates the usefulness and the interpretation of these models in a real case study. Some final remarks conclude the paper.

2 Motivations and notation for the proposed mixture

Sample data consist of a collection of ordered scores $(r_1, r_2, \ldots, r_n)$ anchored to the integers of the support $I_m = \{1, 2, \ldots, m\}$, for some known m. The ordered

evaluation may concern opinions, judgments, degrees of liking/preference, and even a qualitative mapping of some continuous variable, but for simplifying the discussion we assume that responses are some sort of ratings in one-to-one correspondence with integers belonging to I_m. Thus, respondents choose a qualitative assessment on a graduated sequence of verbal definitions (for instance, "extremely dissatisfied", "very dissatisfied", …, "very satisfied", "extremely satisfied") which are coded as numbers just for convenience.

In statistical surveys further information are also collected, and we will speak of ratings and subjects' covariates to refer to ordinal responses and information regarding the respondents, respectively. Our objective is to explain, fit, and predict the probability $Pr\,(R = r)$ that a discrete random variable R assumes values $r = 1, 2, \ldots, m$. When significant, subjects' covariates should improve the performance and the interpretation of such a model. For this purpose, we introduce a probability structure where the final outcome of the evaluation process is a discrete observation generated by an investigated trait which is intrinsically continuous.

In this regard, two possible interpretations are admissible for explaining the mental process by which respondents rate their opinion/evaluation about an item by means of a finite and graduated scale.

According to the *first interpretation*, it may be conjectured that the i-th respondent adopts the following two-step strategy:

– First of all, he/she chooses between a simplistic option (consisting in the selection of a modality which he/she considers very attractive by the nature of verbal wording and/or the numbering of the scale, for example) and a meditated response (which requires some thinking about). We assume that the selection between these two main alternatives happens with probabilities δ_i and $1 - \delta_i$, respectively, for $i = 1, 2, \ldots, n$. This choice may be motivated by a lazy behaviour of the respondent who takes refuge in a category which is judged as convenient, safe, attractive, politically-correct, etc. Such an option has been called *shelter choice* and it may be represented by a degenerate random variable located at $R = c$, where $c \in I_m$ is a known category depending on the specific question at hand:

$$D_r^{(c)} = \begin{cases} 1, & \text{if } \ r = c; \\ 0, & \text{otherwise;} \end{cases} \qquad r = 1, 2, \ldots, m. \tag{1}$$

– If he/she selects the second option, the final selected category is a *balanced decision* between his/her feeling towards the item and a totally random choice, with propensities π_i and $1 - \pi_i$, respectively. This choice assumes a more involved respondent, thus the final decision is a weighted determination between a positive/negative sensation related to the item and a light/heavy indecision/fuzziness. In fact, the selection of an ordered modality among several ones is a very complex mental process since it involves several factors influencing the final choice (Tourangeau et al. 2000). Thus, a simplified version of such a psychological process should limit the analysis only to relevant components. As fully discussed with reference to CUB models (Piccolo 2003; Iannario and Piccolo 2012a), the attractiveness (or

the repulsion) towards the item (= *feeling*) and the indecision (fuzziness) in the response (= *uncertainty*) have been considered as the relevant ones.

We consider *feeling* as an internal/personal attitude concerning the opinion of the subject towards the object and, depending on the circumstances, it may be named as degree of perception, measure of closeness, level of satisfaction/preference, assessment of proficiency, rating of concern, index of selectiveness, pain threshold, risk awareness, subjective probability, degree of confidence, etc.

On the other side, *uncertainty* pertains to the operational modes of the final choice and to the external facts affecting and surrounding the final decision. Thus, uncertainty is not the "randomness" related to the sampling experiment, but it depends on convergent and related factors as: limited set of information about the topic, personal interest/engagement in activities related to the problem, amount of time devoted to the response, nature of the scale in terms of range and wording, tiredness or fatigue for a correct comprehension of the question, willingness to joke and fake, lack of self-confidence, laziness/apathy/boredom of the respondent. Also, the "response style" may be interpreted as a component of uncertainty in the response (see Gottard et al. 2015 for a discussion of these and related topics).

In addition, uncertainty is also related to the "satisficing" behaviour (Simon 1957), which is generated by respondents who choose an adequate answer that may not be the optimal one, in the attempt to minimize the burden of the question (Krosnick 1991). This attitude generates a varying degree of indecision to answer a specific item and it ranges from a complete lack of satisficing (= completely accurate response) to strong satisficing behaviour (= completely random response). Then, we are assuming that uncertainty affects any individual choice and it can be, at worst, constituted by a purely random choice among categories. In intermediate cases, each respondent acts with a *propensity* to adhere to a thoughtful and to a completely random choice, and we will weigh such a propensity with quantities (π_i) and $(1 - \pi_i)$, respectively.

A *second interpretation* may be proposed and again it assumes a two-step strategy for the i-th respondent:

- First of all, he/she decides to activate his/her personal feeling towards the item with a meditated choice (as previously detailed) or to adopt a lazy behavior derived by a global indecision mood with probabilities λ_i and $1 - \lambda_i$, respectively, for $i = 1, 2, \ldots, n$.
- If he/she selects the second option, then he/she may activate a random selection over the support I_m or refuge in a *shelter* category, and this happens with propensities η_i and $1 - \eta_i$, respectively.

The second interpretation is conceptually simpler and it is consistent with the "satisficing" behaviour. In the next section, the equivalence between the two conceptual models will be formally proved.

Turning these interpretations into a statistical framework, several distributions may adequately fit the implied components. The family of CUB models (Piccolo 2003) is characterized by the shifted Binomial and the discrete Uniform random variable for modelling feeling and uncertainty, respectively, as defined by:

$$b_r(\xi_i) = \binom{m-1}{r-1} \xi_i^{m-r} (1-\xi_i)^{r-1}; \qquad p_r^U = \frac{1}{m}; \qquad r = 1, 2, \ldots, m.$$

To support these choices pragmatic and statistical points of view may be advanced. The shifted Binomial distribution involves a single parameter (ξ_i) and presents a modal value located everywhere over the support $\{1, 2, \ldots, m\}$. It allows a parsimonious parameterization when we have to fit observed distributions with different shapes in terms of skewness and flatness. Then, the Binomial distribution (and the shifted one) may be generated by a continuous unimodal distribution by selecting appropriate ordered cutpoints. Thus, this choice is consistent with the common hypothesis that a continuous latent variable moves the final selection of a discrete modality. Finally, the choice of the Binomial random variable may be also justified on the basis of statistical motivations, as detailed in "Appendix 1".

As far as the discrete Uniform distribution is concerned, we adopt this random variable just as the extreme building block for the respondent choice since it is maximally uninformative and maximizes entropy over the class of discrete distribution with a given finite support. In addition, no parameter is added to the model (m is known) and we may judge the resoluteness of the respondent in respect to this extreme choice since it represents the maximum heterogeneity among the responses.

These arguments are behind the introduction of a discrete mixture (Piccolo 2003) defined by

$$Pr\,(R = r) = \pi_i\, b_r(\xi_i) + (1 - \pi_i)\, p_r^U\,, \qquad i = 1, 2, \ldots, n, \tag{2}$$

and called CUB model since it is a convex Combination of discrete Uniform and shifted Binomial random variable.

3 Specification of a GeCUB model

For a given $c \in I_m$ and known m, we will consider the observed response r as the realization of a random variable R whose probability distribution for any i-th subject -according to the *first interpretation*- is defined by:

$$Pr\,(R = r) = \delta_i \left[D_r^{(c)} \right] + (1-\delta_i) \left[\pi_i\, b_r(\xi_i) + (1-\pi_i)\, p_r^U \right], \quad r = 1, 2, , \ldots, m. \tag{3}$$

In absence of covariates, this model has been denoted as CUB model with a *shelter effect* by Iannario (2012a), who discusses properties, estimation issues and related topics.

If we adhere to the *second interpretation*, an alternative specification may be obtained:

$$Pr\,(R = r) = \lambda_i\, b_r(\xi_i) \;+\; (1 - \lambda_i) \left[\eta_i\, p_r^U \;+\; (1 - \eta_i)\, D_r^{(c)} \right], \quad r = 1, 2, , \ldots, m. \tag{4}$$

Given the one-to-one mapping

$$\begin{cases} \lambda_i = \pi_i(1-\delta_i); \\ \eta_i = \dfrac{(1-\pi_i)(1-\delta_i)}{1-\pi_i(1-\delta_i)}; \end{cases} \iff \begin{cases} \pi_i = \dfrac{\lambda_i}{\lambda_i + \eta_i(1-\lambda_i)}; \\ \delta_i = (1-\lambda_i)(1-\eta_i); \end{cases} \qquad i = 1, 2, \ldots, n;$$

it is indifferent to discuss either of GeCUB specifications (3) and (4). Hereafter, we focus on the model (3) since it gives an immediate weight (δ_i) to quantify the *shelter effect*.

In standard CUB random variables, thanks to the one-to-one correspondence between the parameters (π, ξ) and the probability distribution, we plot the estimated models as points in the unit square in order to interpret the behaviour of respondents when faced to different items, for varying circumstances of space, time and contexts. If we wish to add the additional parameter (δ) to this representation we may increase the size of the point (π, ξ) or add an horizontal line starting at (π, ξ) and proportional to δ, for instance.

In presence of covariates, model (3) allows to interpret the parameters in relation to the *feeling* ($1 - \xi_i$) of the respondent, the *uncertainty* ($1 - \pi_i$) of the responses and a possible *shelter effect* δ_i. Briefly, this effect is the weight of the *shelter choice* and it quantifies the increase of probability of the category ($R = c$) with respect to a CUB model (where $\delta_i = 0$). Thus, CUB models are nested into CUB models with a *shelter effect*.

Suppose that information on the n subjects are summarized by a set of v variables and collected in the matrix

$$\boldsymbol{T} = ||t_{ij},\ i = 1, 2, \ldots, n;\quad j = 1, 2, \ldots, v||\,.$$

which summarizes the available subjects' covariates (the so-called concomitant variables). We consider sub-matrices $\boldsymbol{Y}$, $\boldsymbol{W}$, $\boldsymbol{X}$ obtained from $\boldsymbol{T}$ by selecting convenient columns. Then, we denote by $\boldsymbol{y}_i$, $\boldsymbol{w}_i$, and $\boldsymbol{x}_i$, for $i = 1, 2, \ldots, n$, the i-th rows of the $\boldsymbol{Y}$, $\boldsymbol{W}$ and $\boldsymbol{X}$ matrices, respectively, that is:

$$\begin{aligned} &\boldsymbol{y}_i = (y_{i0}, y_{i1}, y_{i2}, \ldots, y_{ip}); \quad \boldsymbol{w}_i = (w_{i0}, w_{i1}, w_{i2}, \ldots, w_{iq}); \\ &\boldsymbol{x}_i = (x_{i0}, x_{i1}, x_{i2}, \ldots, x_{is})\,. \end{aligned}$$

We let: $y_{i0} = w_{i0} = x_{i0} = 1$, for $i = 1, 2, \ldots, n$. These rows contain all available sample information on the i-th subject related to the model components and they are necessary and sufficient for the model specification.

Then, for $i = 1, 2, \ldots, n$, we introduce a direct logistic link among parameters and covariates:

$$\pi_i = \pi_i(\boldsymbol{\beta}) = \frac{1}{1 + e^{-\boldsymbol{y}_i\boldsymbol{\beta}}}\,; \quad \xi_i = \xi_i(\boldsymbol{\gamma}) = \frac{1}{1 + e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}}\,; \quad \delta_i = \delta_i(\boldsymbol{\omega}) = \frac{1}{1 + e^{-\boldsymbol{x}_i\boldsymbol{\omega}}}\,;$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_q)'$ and $\boldsymbol{\omega} = (\omega_0, \omega_1, \ldots, \omega_s)'$, respectively. According to the logistic function: $logit(p) = \log(p/(1-p))$, previous relationships are equivalent to:

$$logit(\pi_i) = \boldsymbol{y}_i\boldsymbol{\beta}\,; \quad logit(\xi_i) = \boldsymbol{w}_i\boldsymbol{\gamma}\,; \quad logit(\delta_i) = \boldsymbol{x}_i\boldsymbol{\omega}\,; \quad i = 1, 2, \ldots, n. \tag{5}$$

Alternative links are admissible but we found that the logistic function is a convenient mapping in most real circumstances. Notice that, given the previous parameterization, the matrices $\boldsymbol{Y}$, $\boldsymbol{W}$, $\boldsymbol{X}$ may or may not possess an arbitrary number of common columns.

To see how a single covariate affects the probability of the response, we may plot this probability mass function for any prefixed value of the discrete covariates or for some specific values of the continuous variables. Alternatively, we may consider the modification of the points in the parametric space for varying values of covariates or to study the behaviour of $(1-\pi_i)$, $(1-\xi_i)$ and δ_i as functions of selected covariates, as it will be pursued in the real case study, for instance.

Finally, a Generalized CUB (=G*e*CUB) model is fully specified by (3) and (5). For this model the length of the vector $(\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\omega}')'$ is $(p+q+s+3)$. If some or all subjects' covariates (or components) are absent, analysis is greatly simplified. In these circumstances, it is more convenient to refer to CUB models without and with *shelter effect*, respectively, as derived by Piccolo (2006) and Iannario (2012a).

A critical point is that the model assumes c as a known constant. In principle, one might test a CUB model with a possible *shelter effect* for any admissible $c = 1, 2, \ldots, m$ and then accept the model with the best fitting and significant parameters. Indeed, in real case studies concerning a specific scientific field, researchers have nearly always accumulated evidence about a category c where people tend to give a response more often than that predicted by the standard model. This happens for psychological motivations, biased or sensible questions, mass media pressure, difficulty of comprehension of the item, desire of privacy, specific wording, and so on. Thus, the knowledge of c is not a severe constraint in most of the current surveys.

4 Statistical inference for the GeCUB model

The sample ratings $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ are considered as realizations of the random sample $(R_1, R_2, \ldots, R_n)'$ where each R_i is independently distributed as a discrete random variable over the support I_m.

In a mixture distribution, it is useful to characterize the notation of the parameters according to their roles. Thus, we denote by $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\eta}')'$ the full parameter vector of a G*e*CUB model where $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ are the parameter vectors of weights (α_g) of the probability distributions $\left(\mathscr{P}_g\right)$, respectively, for the $g = 1, 2, 3$ components (as summarized in Table 1).

Given the sample $\boldsymbol{r}$ and the information set of covariates $\mathscr{C}_i = (\boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{x}_i)$, for $i = 1, 2, \ldots, n$, the log-likelihood function may be written as:

Table 1 Notation for the components of the mixture in the GeCUB model

g	$\alpha_{gi} = \alpha_{gi}(\boldsymbol{\psi}_g)$	$p_{gi} = p_g(r_i;\, \boldsymbol{\eta}_g)$	$\boldsymbol{\psi}_g$	$\boldsymbol{\eta}_g$
1	δ_i	$D_{r_i}^{(c)}$	$\boldsymbol{\psi}_1 = \boldsymbol{\omega}$	
2	$\pi_i(1-\delta_i)$	$b_{r_i}(\boldsymbol{\gamma})$	$\boldsymbol{\psi}_2 = (\boldsymbol{\beta}',\, \boldsymbol{\omega}')'$	$\eta_2 = \boldsymbol{\gamma}$
3	$(1-\pi_i)(1-\delta_i)$	$p_{r_i}^U$	$\boldsymbol{\psi}_3 = (\boldsymbol{\beta}',\, \boldsymbol{\omega}')'$	

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log\left(Pr\left(R = r_i|\mathscr{C}_i,\, \boldsymbol{\theta}\right)\right) = \sum_{i=1}^{n} \log\left(\sum_{g=1}^{3} \alpha_{gi}\, p_g(r_i;\boldsymbol{\eta}_g)\right)\\ &= \sum_{i=1}^{n} \log\Big[\alpha_{1i}\, p_1(r_i;\boldsymbol{\eta}_1) + \alpha_{2i}\, p_2(r_i;\boldsymbol{\eta}_2) + \alpha_{3i}\, p_3(r_i;\boldsymbol{\eta}_3)\Big]\\ &= \sum_{i=1}^{n} \log\Big[\delta_i\, D_{r_i}^{(c)} + \pi_i(1-\delta_i)\, b_{r_i}(\boldsymbol{\gamma}) + (1-\pi_i)(1-\delta_i)\, p_{r_i}^U\Big].\end{aligned}$$

As for all mixture distributions, ML estimators are effectively obtained from $\ell(\boldsymbol{\theta})$ by exploiting the EM procedure proposed by Dempster et al. (1977) and specifically oriented to finite mixtures (McLachlan and Krishnan 2008; McLachlan and Peel 2000). Such a procedure is detailed for GeCUB models in "Appendix 2". Asymptotic inference requires the knowledge of the information matrix for GeCUB models and this step is generally achieved by numerical computations or by simulation devices (bootstrap, for instance). However, it is more accurate to compute the second order derivatives of $\ell(\boldsymbol{\theta})$ by analytic methods.

All estimation procedures have been derived for the parameters of GeCUB models specified by (3). Given the invariance properties of ML estimators (Serfling 1980, p.43), it is immediate to get any estimation result of this model in terms of the alternative specification (4).

Then, the validation of the estimated model relies on several points:

- *parameters significance*: this is achieved by comparing estimates to their standard errors (Wald test). Some caution should be considered when we test on the border of the parametric space since significance must be modified: a detailed account and related references for testing $H_0 : \delta = 0$ for the *shelter effect* are discussed by Iannario (2012a).
- *log-likelihood comparisons*: in presence of nested models, we test the increase in log-likelihoods with respect to the standard χ^2 percentiles to see if the most complex model is a valuable choice. Likelihood ratio tests, deviance and related statistics may be defined as in the current literature (Agresti 2010, pp. 67–75)
- *global indices*: we consider measures as $BIC = -2\,\ell(\hat{\boldsymbol{\theta}}) + (p+q+s+3)\,\log(n)$, for instance, to take into account both the improvement in the log-likelihood and the penalty given by an increase of parameters of the model.
- *residuals diagnostic*: Pearson and relative residuals may be defined and conveniently checked. In addition, further analyses based on the the definition of generalized residuals (Di Iorio and Iannario 2012) may be pursued as well.

A fitting index which compares observed relative frequencies f_r and expected $\hat{p}_r = p_r(\hat{\boldsymbol{\theta}})$ is based on normalized dissimilarity measures as

$$\mathscr{F}^2 = 1 - \frac{1}{2}\sum_{r=1}^{m} |f_r - \hat{p}_r|.$$

It may be interpreted as the proportion of correct predicted responses (Iannario 2009). In presence of discrete covariates with k categories, with obvious notation, this quantity may be generalized by means of

$$\mathscr{F}^2 = 1 - \frac{1}{2}\sum_{j=1}^{k} \frac{n_j}{n} \sum_{r=1}^{m} |f_{rj} - \hat{p}_{rj}|.$$

From a predictive point of view, several problems have to be faced when ordinal data are involved. As a matter of fact, all modelling approaches are able to predict a whole probability distribution given the subjects' covariates; indeed, most of the methods (as involved by log-likelihood computations and analysis of deviance) concern the comparison of predicted and observed proportions of the ordinal categories.

On the other side, the main purpose of the researcher is to predict the rating of a respondent, given his/her characteristics. Thus, we have to synthesize $Pr\left(R = r\,|\,\hat{\boldsymbol{\theta}}, \mathscr{C}_i\right)$ by a predictor $\hat{r}_i$ of r_i, for $i = 1, 2, \ldots, n$. Expectation, modal value (mode) and median of the estimated probability distribution, conditional to selected covariates $\mathscr{C}_i$, are candidates for $\hat{r}_i$. For any selection of a predictor, a Root Mean Square Error (RMSE) is defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(r_i - \hat{r}_i\right)^2}. \tag{6}$$

This measure should be critically considered since it is based on a point estimate; although it is useful to compare different predictors derived by different models, it should not be used to discriminate models belonging to different classes.

5 A simulation experiment

To check the ability of the proposed modelling approach to detect the presence of a possible *shelter* the case of a finite sample size, a limited experimental design has been planned with a subjects' covariate which, for simplicity, we suppose dichotomous. More precisely, in a rating survey with $m = 7$ categories, we assume the existence of two subgroups $\mathscr{G}_0$ and $\mathscr{G}_1$ characterized by the parameters: $\boldsymbol{\theta}_0 = (\pi_0, \xi_0, \delta_0)'$ and $\boldsymbol{\theta}_1 = (\pi_1, \xi_1, \delta_1)'$, respectively, with a *shelter effect* at the fifth category, so that $c = 5$.

Table 2 lists $\boldsymbol{\theta}_i = (\pi_i, \xi_i, \delta_i)'$, $i = 0, 1$ according to CUB and *shelter* parameterization and to the implied G*e*CUB parameters (when the dichotomous covariate is

Table 2 Experimental design for the simulation of GeCUB models

Mod.	Param.	Uncertainty parameters		Feeling parameters		*Shelter* parameters	
1	CUB & sh.	$\pi_0 = 0.550$	$\pi_1 = 0.850$	$\xi_0 = 0.800$	$\xi_1 = 0.400$	$\delta_0 = 0.150$	$\delta_1 = 0.200$
	GeCUB	$\beta_0 = 0.201$	$\beta_1 = 1.534$	$\gamma_0 = 1.386$	$\gamma_1 = -1.792$	$\omega_0 = -1.735$	$\omega_1 = 0.348$
2	CUB & sh.	$\pi_0 = 0.350$	$\pi_1 = 0.750$	$\xi_0 = 0.300$	$\xi_1 = 0.700$	$\delta_0 = 0.100$	$\delta_1 = 0.150$
	GeCUB	$\beta_0 = -0.619$	$\beta_1 = 1.718$	$\gamma_0 = -0.847$	$\gamma_1 = 1.695$	$\omega_0 = -2.197$	$\omega_1 = 0.463$
3	CUB & sh.	$\pi_0 = 0.550$	$\pi_1 = 0.650$	$\xi_0 = 0.200$	$\xi_1 = 0.800$	$\delta_0 = 0.150$	$\delta_1 = 0.050$
	GeCUB	$\beta_0 = 0.201$	$\beta_1 = 0.418$	$\gamma_0 = -1.386$	$\gamma_1 = 2.773$	$\omega_0 = -1.735$	$\omega_1 = -1.210$

explicitly inserted). To express this basic structure in terms of GeCUB models, we exploit the relationships among π_0, π_1 and $\boldsymbol{\beta} = (\beta_0,\ \beta_1)'$ parameters when a dichotomous covariate $D_i,\ i = 1, 2, \ldots, n$ is present:

$$\pi_i = \left[1 + \exp(-\beta_0 - \beta_1\, D_i)\right]^{-1};$$
$$D_i = 0, 1 \implies \beta_0 = \log \frac{\pi_0}{1 - \pi_0};$$
$$\beta_1 = \log \frac{\pi_1}{1 - \pi_1} - \beta_0;$$

similar relationships hold between $\xi_i, \delta_i, i = 0, 1$ and $\boldsymbol{\gamma} = (\gamma_0,\ \gamma_1)', \boldsymbol{\omega} = (\omega_0,\ \omega_1)'$, respectively.

The experiment is characterized by different configurations of location and shape of the probability distributions as shown in Fig. 1, where the contribution of the *shelter effect* at $R = 5$ has been emphasized.

For each simulation run, sample data consist of two samples of $n_0 = n_1 = 500$ observations $(r_i, d_i), i = 1, 2, \ldots, n$ where r_i is generated by the groups $\mathscr{G}_0$ and $\mathscr{G}_1$, and d_i assumes values 0 and 1 for the first and the second subgroups, respectively. Then, the following steps have been performed:

1. Generate n_0 and n_1 observations from the "true" model.
2. From the global sample $(r_1, r_2, \ldots, r_n)$ of $n = n_0 + n_1$ observations, estimate a GeCUB model by the ML method on the basis of the information (r_i, d_i), for $i = 1, 2, \ldots, n$ and collect estimates in the vector

$$\boldsymbol{\theta}^{[j]} = \left(\beta_0^{[j]},\ \beta_1^{[j]},\ \gamma_0^{[j]},\ \gamma_1^{[j]},\ \omega_0^{[j]},\ \omega_1^{[j]}\right)'.$$

3. Repeat 1-2 for $j = 1, 2, \ldots, nsimul = 1000$ times.

We report the estimates of bias and mean square error (MSE) of the parameters in Table 3 and we briefly comment on the main results of this experiment.

- The bias of the estimates is always very limited; thus, MSE is mainly due to the variability of the estimates. In fact, MSE is generally small but for some parameters it deserves some consideration.
- MSE of estimators $\hat{\beta}_1$ seems more extreme in the experiments 1 and 2; the ratios of the parameters to the square root of MSE are 3.182 and 4.516, respectively. These

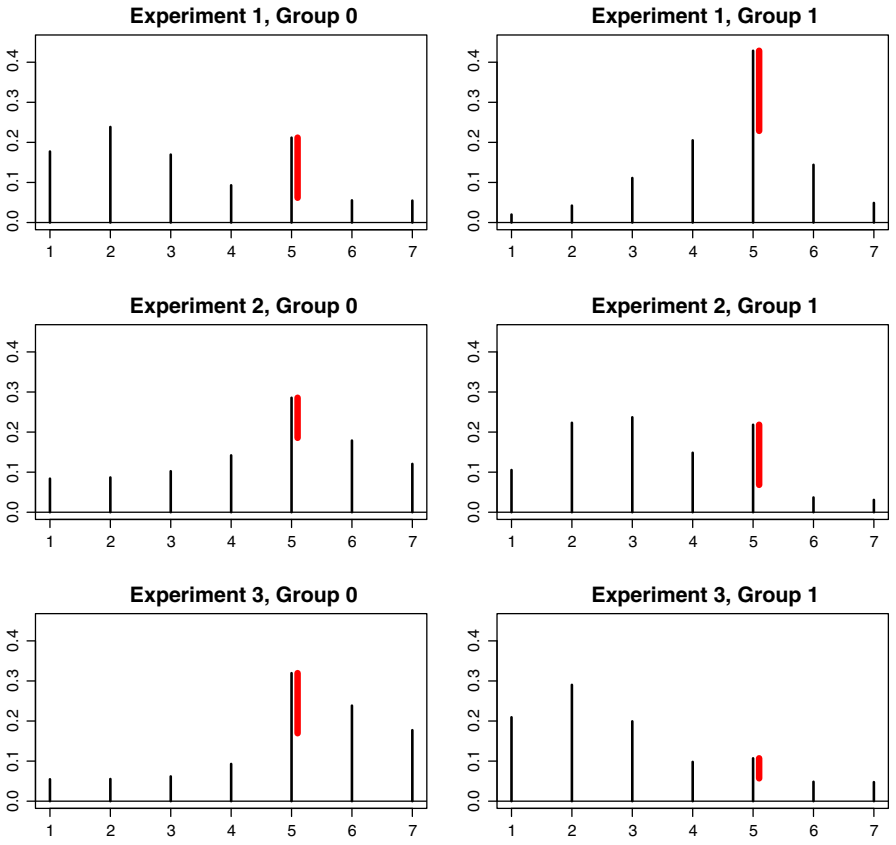

Fig. 1 Population distributions (*shelter* at $R = 5$) of two subgroups: $\mathscr{G}_0$ (*left*) and $\mathscr{G}_1$ (*right*)

Table 3 Bias and mean square error for the simulation experiments

Estimates	Experiment 1		Experiment 2		Experiment 3	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\beta}_0$	−0.00560	0.04414	−0.02063	0.08606	0.01074	0.04384
$\hat{\beta}_1$	0.08720	0.23241	0.03254	0.14475	−0.00709	0.07998
$\hat{\gamma}_0$	0.00313	0.01041	−0.01567	0.02271	−0.00445	0.01741
$\hat{\gamma}_1$	−0.00414	0.01360	0.01828	0.02862	0.00097	0.02508
$\hat{\omega}_0$	−0.01061	0.02894	−0.03626	0.13154	−0.01712	0.06823
$\hat{\omega}_1$	−0.00887	0.07739	0.02362	0.15684	−0.06551	0.32499

MSEs are larger than expected as a consequence of few atypical values observed in these experiments. A similar consideration applies for the MSE of the estimator $\hat{\omega_0}$ in the experiment 2 for which the ratio of the parameters to the square root of MSE is −6.059.

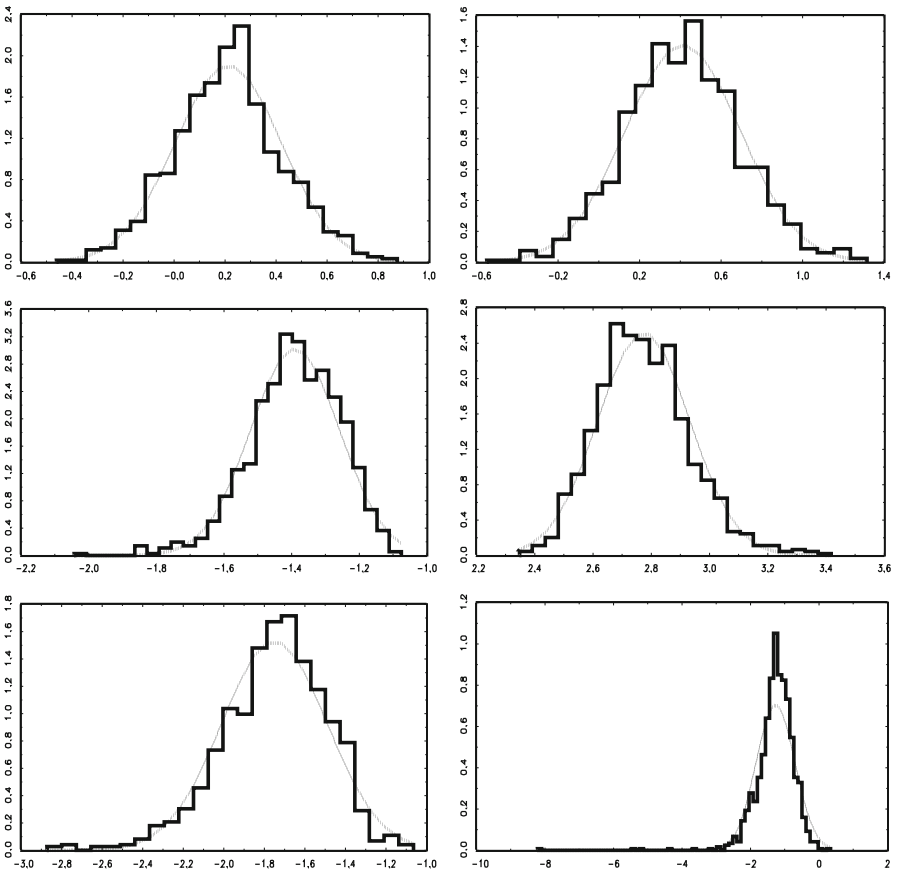

Fig. 2 Histograms and approximating normal distributions for simulated estimates of Experiment 3. *Each panel row* represents the distributions of $\hat{\beta}_i$, $\hat{\gamma}_i$, $\hat{\omega}_i$, for $i = 0$ (*left*) and $i = 1$ (*right*), respectively

– A different problem arises for the estimation of ω_1 in the experiments 2 and 3 for which the mentioned ratios are 1.169 and -2.123, respectively. In these situations, the proportions of the *shelter effect* (estimated by ω_1) at $R = 5$ and expressed by $\delta = 0.15$ and $\delta = 0.05$, respectively, are important with respect to the basic probabilities. Here, the values of $Pr\,(R = 5)$ are 0.218 and 0.107, and the *shelter effect* represents 69 % and 47 % of the probability of the category $R = 5$, respectively. In these cases, a large number of observations is required for a more accurate estimation.
– The asymptotic Normality of all estimators is sufficiently accurate as shown by the histograms reported in Fig. 2 for the more extreme case (Experiment 3). Only the distribution of $\hat{\omega}_1$, for the aforementioned motivations, presents a left tails which is too long for a Gaussian distribution as a consequence of a limited set of atypical values. If we omit them, the resulting distribution is almost perfectly Normal.

Although the parameters of the distributions have been selected in order to scatter different shapes, more extensive simulations are required. We report that

further experiments (here not discussed for brevity) have been performed for different number of categories, different proportions of the groups and varying sample size. They confirmed the adequacy of the ML estimation method and support the ability of the approach to detect different groups for samples of moderate/large size.

6 A real case study

We check the approach so far discussed with a real case study related to the political orientation in a survey planned with the students of University of Naples Federico II and their families and friends, during 2010. Since the research is based on an observational sample the study cannot be considered as representative of the Italian population but it is an instance of the capability of the G*e*CUB model approach in terms of fitting and interpretation of similar results.

It is generally difficult to collect reliable answers about political orientation, and this is particularly true if the research considers a finer disaggregation than a coarse definition of "Conservative", "Moderate", and "Liberal". This happens, for instance, in Italy where the galaxy of political parties is extremely varied; thus, it is important to predict the political orientation of a person by means of related questions and/or different covariates which are strongly related to such an orientation.

The sample data consist of $n = 707$ questionnaires where respondents carefully expressed their (self-assessed) *Political orientation* as an ordinal variable R with $m = 9$ categories, where $1, 5, 9$ stand for "Extremely to Left", "Center" and "Extremely to Right", respectively. In addition, several concomitant variables related to personal socio-demographic and economic situation, opinions, ranking of nationwide newspapers, etc. have been collected. Thus, 48 % of respondents are women and their average age is 38 (derived from a larger group of young university students and a smaller one of their relatives). Then, education is higher than the average of the population since about one half of interviewees has got a (secondary) diploma degree and about 40 % has a university degree.

After a preliminary analysis based on stepwise regression approaches, we found the following covariates as relevant to explain Political orientation: `Age` (=the respondents' age transformed as deviations from the average of logged years), `Rank` (=the ranking assigned to a historic Italian newspaper, "L'Unità", well known for Left positions; here, `Rank` $= 1$ means it is the most preferred, `Rank` $= 7$ means that it is considered the worst), and `Demo` (=a dummy variable which denotes if the respondent has participated to public demonstrations in the last year).

All computations have been implemented by a programm written in the GAUSS language by using ML methods and exploiting the EM procedure for convergence. Standard errors have been computed by analytical derivation of the observed information matrix with ML estimates plugged into: details of these formal developments are reported in Iannario and Piccolo (2012b).

A G*e*CUB model may be considered as the final step of several statistical analyses, including exploratory and correlation methods, CUB models fitting with and without covariates, and CUB model with a dummy to check for a possible *shelter effect* in

Table 4 Estimation of CUB and GeCUB models for the Political orientation

Models	Covariates	*Uncertainty* parameters	*Feeling* parameters	*Shelter* parameters
CUB		$\hat{\pi} = 0.428$ *(0.041)*	$\hat{\xi} = 0.648$ *(0.017)*	
CUB + *shelter*		$\hat{\pi} = 0.374$ *(0.041)*	$\hat{\xi} = 0.717$ *(0.021)*	$\hat{\delta} = 0.089$ *(0.019)*
CUB + covariates	*Constant*	$\hat{\beta}_0 = 0.676$ *(0.159)*	$\hat{\gamma}_0 = 2.018$ *(0.145)*	
	$\widetilde{Age}$	$\hat{\beta}_1 = 1.363$ *(0.463)*		
	Rank		$\hat{\gamma}_2 = -0.362$ *(0.029)*	
	Demo		$\hat{\gamma}_3 = 0.652$ *(0.126)*	
	Gender × *Rank*		$\hat{\gamma}_4 = 0.037$ *(0.019)*	
GeCUB	*Constant*	$\hat{\beta}_0 = 2.133$ *(0.461)*	$\hat{\gamma}_0 = 2.127$ *(0.158)*	$\hat{\omega}_0 = -2.843$ *(0.320)*
	$\widetilde{Age}$			$\hat{\omega}_1 = -4.603$ *(1.629)*
	Rank	$\hat{\beta}_2 = -0.355$ *(0.096)*	$\hat{\gamma}_2 = -0.360$ *(0.033)*	
	Demo		$\hat{\gamma}_3 = 0.641$ *(0.136)*	
	$\widetilde{Age} \times Rank$			$\hat{\omega}_4 = 1.357$ *(0.320)*

a definite category (see Table 4 for the estimates of these different models. Standard errors are in parentheses). Figure 3 summarizes different aspects of these investigations which we briefly comment.

Data set are characterized by a serious uncertainty in the responses since we get $(1-\hat{\pi}) = 0.57$ after fitting a CUB model. The observed distribution shows a prominent *shelter effect* at $R = 5$ (see Fig. 3, top-left panel): an appreciable proportion of people, estimated by $\hat{\delta} = 0.089$, choices an intermediate position which corresponds also to a non-selective choice. This option represents a genuine *shelter option*. By missing this point, one might deduce that the Political orientation of the interviewees is strongly anchored to a "Centre" position (about 1/5 of the respondents), a statement not confirmed by electoral results and other empirical analyses.

Moreover, a significant relationship has been found between the expressed Political orientation and the covariate `Demo`, as confirmed by the estimated CUB model which includes `Demo` as a covariate for feeling: Fig. 3 (bottom-left panel). Similarly, there is a sharp evidence of a connection between Political orientation and the covariate `Rank` as supported by the conditional box-plots of Fig. 3 (top-right panel): here the location of the distribution of the responses changes with `Rank` in a non-linear fashion.

In the framework of CUB models, $(1-\pi)$ can be considered as a direct measure of indecision and π is strongly related to the heterogeneity of the distribution (Iannario 2012c, pp.169–171). Thus, to check if the heterogeneity is related to some subjects' covariate we compute this measure on the ordinal data split by a categorical variable, as `Rank` for instance. We choose the normalized Laakso and Taagepera (1989) index which, for a discrete probability mass function $\{p_1, p_2, \ldots, p_m\}$, is defined by

$$\mathscr{H} = \frac{1}{m-1}\left[\left(\sum_{i=1}^{m} p_i^2\right)^{-1} - 1\right].$$

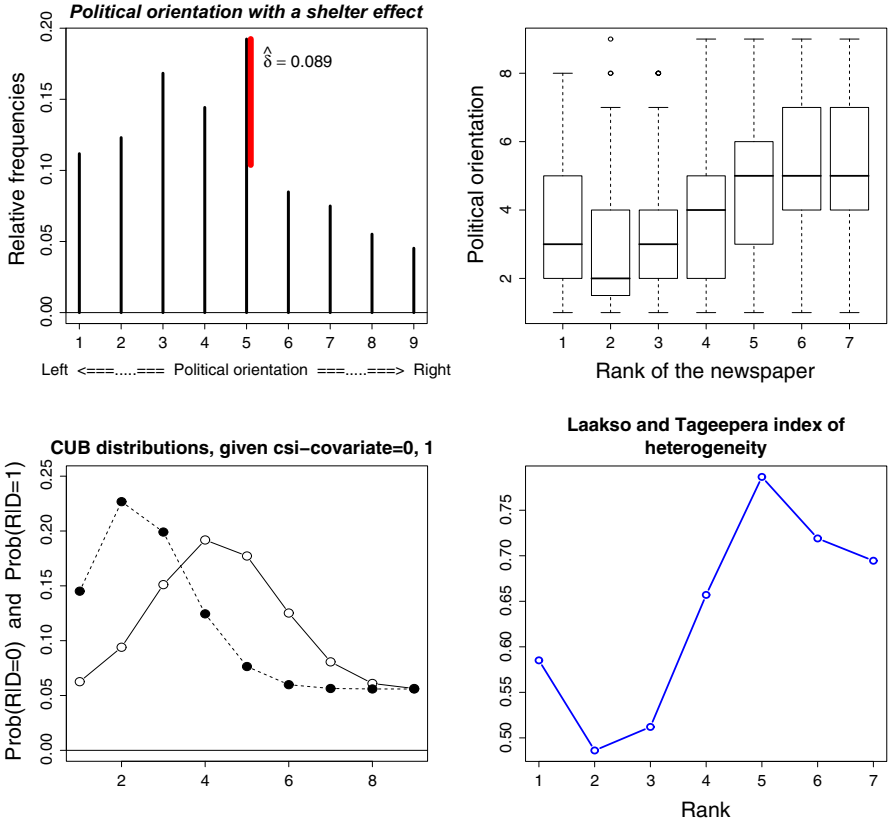

Fig. 3 Observed distribution of Political orientation, with estimated *shelter effect*, (*top-left*). Box-plots of Political orientation with respect to `Rank` (*top-right*). CUB models distributions conditional to `Demo`=0,1 (*bottom-left*). Heterogeneity index of Political orientation with respect to `Rank` (*bottom-right*)

The estimate of this index for the Political orientation conditional to `Rank` is shown in Fig. 3 (bottom-right). Although non-linear, we see a definite increase of $\mathscr{H}$ with the covariate `Rank` so we may expect a relationship of the uncertainty with this variable.

Finally, observe that when $\xi \to 1$ ($\xi \to 0$) people express preference for an extreme left (right) position; thus, the parameter ξ is a direct measure of "Left" orientation.

A CUB model with covariates presents a good fit but it does not take the *shelter effect* into account. In fact, as the log-likelihood and BIC measures confirm (see Table 5), it is relevant to introduce a *shelter effect* with significant covariates by means of a G*e*CUB model.

The final G*e*CUB model interprets such a relationship between the dependent variable and `Age`, `Demo` and `Rank`, with a significant contribution of the interaction between `Age` and `Rank` due to the different personal history of the respondents (which regarded such a newspaper with extreme positive/negative feeling according to their Political orientation). This circumstance is confirmed by the fact that a CUB model with covariates (and no *shelter* component) assumes `Age` as a significant covariate

Table 5 Fitting measures of estimated models for the Political orientation

Models	Number of parameters	$\ell(\hat{\boldsymbol{\theta}})$	BIC	Root mean square error	
				Mode	Mean
CUB	2	−1503.0	3019.1	2.211	2.202
CUB + *shelter*	3	−1492.6	3004.9	2.306	2.196
CUB + covariates	6	−1403.6	2846.6	2.076	2.057
G*e*CUB	8	***−1383.6***	***2819.6***	***2.041***	1.942
POM	12	−1390.0	2858.7	2.045	***1.936***

for explaining uncertainty whereas the introduction of a *shelter* component considers `Age` as an useful covariate to explain this effect. We observe that the introduction of a *shelter effect* removes a significant role of `Gender` and emphasizes the importance of `Rank` for assessing the distribution of the responses.

An immediate visualization consists in plotting the estimated probability distributions of the G*e*CUB models conditional to the significant subjects' covariates, that is `Demo`, `Rank` and `Age`, as in Fig. 4. It is evident that the participation to demonstrations (`Demo`=1) increases the probability of being Left oriented (=low categories of the support). `Rank` is strongly related to Political orientation since a low rank for the selected newspaper is strictly related to Left orientation; noticeably, young respondents who give high consideration to this newspaper also have a considerable *shelter effect*. `Age` has a double effect: with increasing age people moves towards Right and the *shelter effect* becomes so prominent that it accounts for about 50 % of the probability of response.

A more stringent evidence of the *shelter effect* turns out by considering the role of the parameters δ_i in the estimated G*e*CUB model:

$$\hat{\delta}_i = \frac{1}{1 + e^{2.842 + \widetilde{Age_i}\,(4.603 - 1.357\, Rank_i)}}, \qquad i = 1, 2, \ldots, n.$$

The *shelter effect* δ_i adds a positive probability at modality $R = 5$ (the so-called "Centre" choice) and this effect changes with `Age` and with the interaction between `Age` and `Rank`: the interaction acts positively with `Age` if `Rank` $= 1, 2$ (people who substantially agree with the positions of the selected newspaper) and negatively elsewhere. The behaviour of δ_i explains this composite effect for varying `Age` and `Rank`.

Figure 5 suggests that for people aged less than 34 years the *shelter effect* is quite moderate when they have low reputation towards the newspaper (=high rank). A more important contribution is registered for people aged more than 34 years (especially when respondents are elderly): the *shelter effect* systematically increases if they rank the newspaper in a high position (=bad consideration). Thus, the refuge position attracts with more decision people who negatively consider the selected newspaper and are not so young (again this is related to the heated political debate in Italy during the last 50 years).

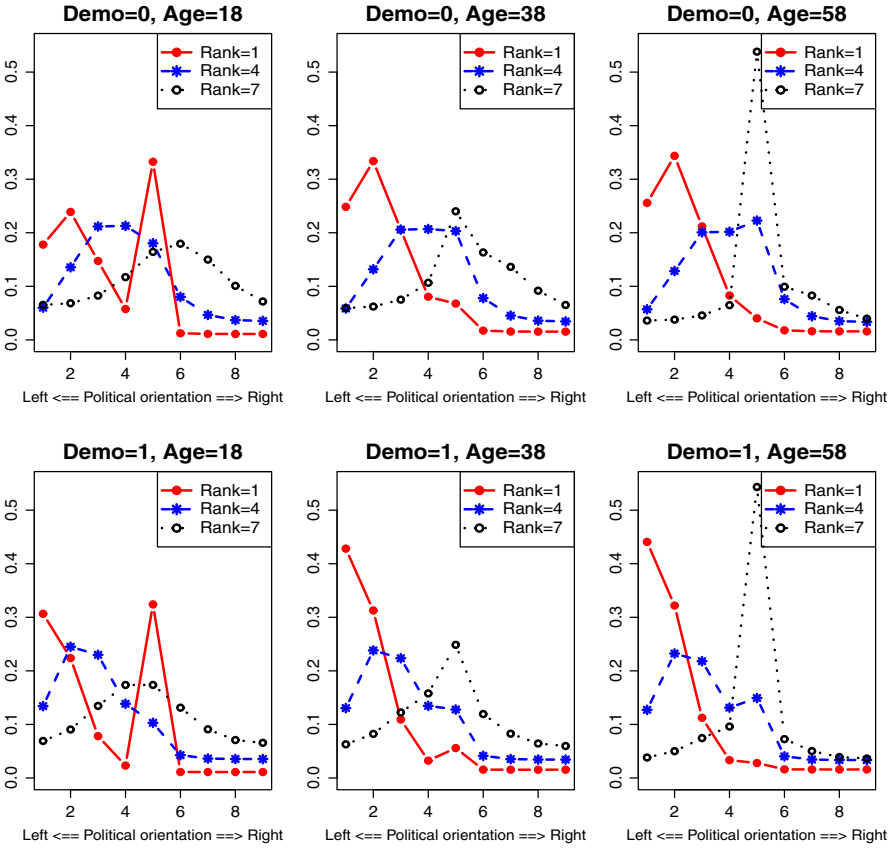

Fig. 4 Estimated GeCUB probability distributions for given subjects' covariates

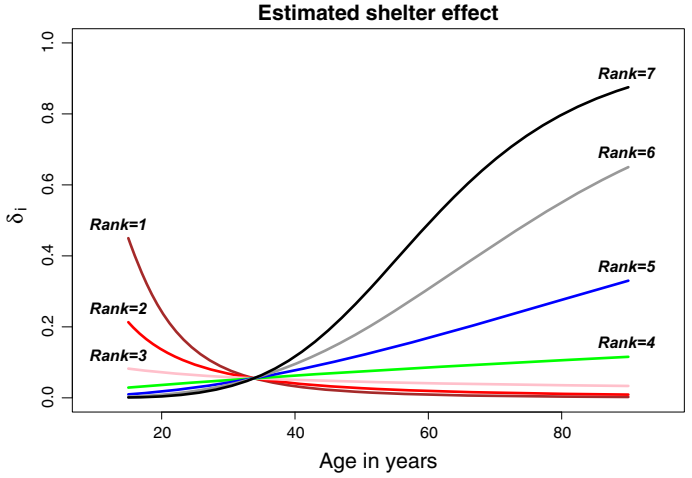

Fig. 5 *Shelter effect* of the estimated GeCUB model for varying `Age` and `Rank`

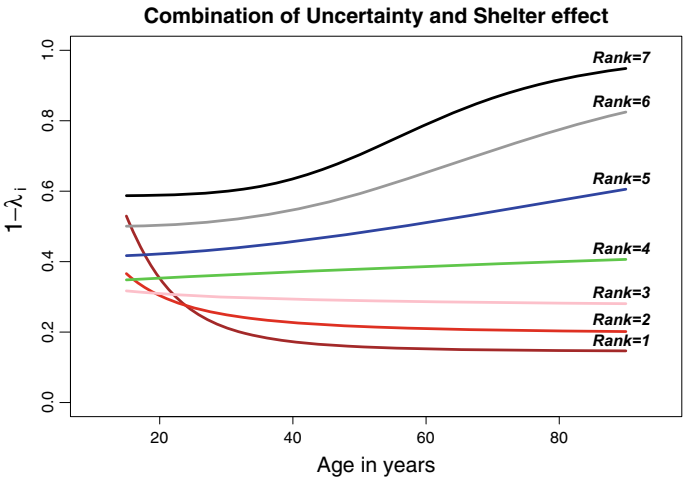

Fig. 6 Global indecision of the estimated G*e*CUB model for varying `Age` and `Rank`

A further interpretation may be derived if we exploit the second specification (4) of a G*e*CUB model where the weight of the combination of uncertainty and *shelter effect* (a sort of global indecision) is measured by $1-\lambda_i$ and estimated as:

$$1-\hat{\lambda}_i \;=\; 1-\left(1-\hat{\delta}_i\right)\,\frac{1}{1+e^{-2.133+0.355\,Rank_i}}\,, \qquad i=1,2,\ldots,n.$$

This quantity is shown in Fig. 6 and confirms that the global indecision suddenly decreases for young people up to 30 years, then it remains substantially constant. On the contrary, this global effect regularly increases with years as far as respondents are right-oriented.

For a comparative analysis, some results obtained by using a more consolidated approach, as the proportional odds models (POM) (Agresti 2010, p.53), are discussed and the main fitting measures are reported in the last line of Table 5. The procedure for selecting significant explanatory variables leads to the set of covariates {`Age`, `Rank`, `Demo`, `Age` × `Rank`} as for G*e*CUB models. The inclusion of `Gender` does not raise significantly the likelihood.

Finally, two expected profiles according to POM and G*e*CUB models (fitted with the same explanatory covariates) are compared to see the effect of the different structures on the probabilities of responses. These results are summarized in Fig. 7 for a young left-oriented respondent who declares to participate to demonstrations (profile A: left panel) and for an elderly right-oriented respondent who does not participate to demonstrations (profile B: right panel). It seems evident that both models capture a similar pattern in the probabilities of responses; however, the G*e*CUB model includes more uncertainty in the expected probabilities and gives special importance to the *shelter* option, which is only partially captured by a POM. In addition, G*e*CUB model relates this effect to significant covariates as already shown in Fig. 5.

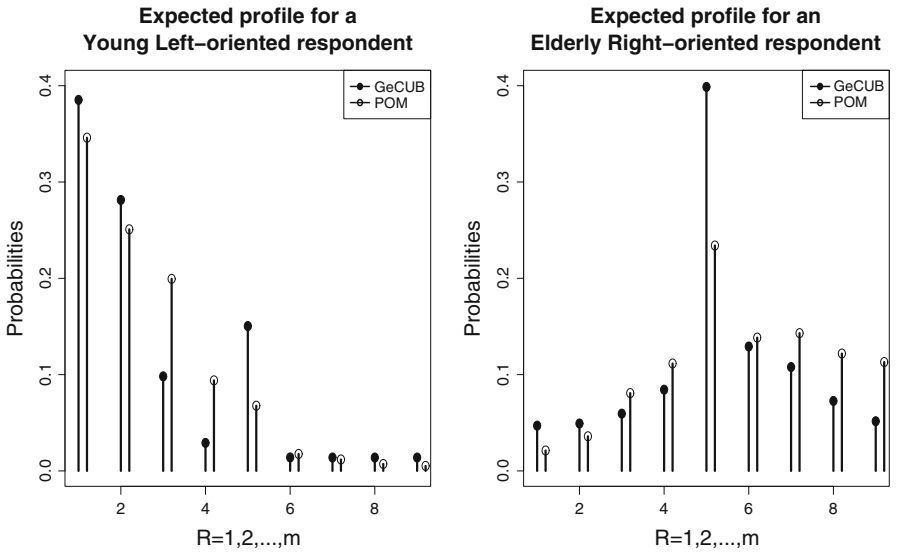

Fig. 7 Expected profiles of responses according to POM and G*e*CUB models. Profile A concerns a young left-oriented respondent who participates to demonstrations (*left panel*). Profile B concerns an elderly right-oriented respondent who does not participate to demonstrations (*right panel*)

Few considerations may be added: first, the estimation of 8 additional cutpoints in POM induces a serious loss in parsimony and, secondly, these models transfer the effect of the respondents' indecision in the global effect of covariates on the responses. This may bias the interpretation of the results and the corresponding prediction of the respondents' behaviour. Thus, although maximum log-likelihood values are definitely comparable, the G*e*CUB model seems more convenient in terms of fitting *and* parsimony (as measured by *BIC*).

The prediction of ordinal data is generally assessed by comparing the observed *proportions* of categorical responses with the fitted ones: this is performed by computing deviance measures which are related to the previous log-likelihood considerations. In fact, these models may be used also for predicting the responses of a *single* respondent given the covariates by means of the estimated significant relationships. More specifically, we use the estimated $\hat{r}_i$ computed from the G*e*CUB model to predict the Political orientation of any respondent, given the knowledge of `Age`, `Rank` and `Demo`. In Table 5, we list the *RMSE* obtained by different estimated models.

In our case study, the modal value as a predictor outperforms the expected value (in fact, modal value is also easier to interpret since it is a proper category). The improvement of a G*e*CUB with respect to a CUB model with covariates is quite moderate; however, from a predictive point of view, G*e*CUB should be preferred if we use the modal value as point predictor for the response and it is comparable to POM if we use the mean.

A common problem arises in the setting of point prediction: although the estimated probability distribution fits the observed ones in a satisfactory measure, it is difficult to predict the single response with high confidence given the appreciable level of uncer-

tainty which surrounds these options. The estimated GeCUB model is able to predict the response exactly or around ± 1 category in more than 60 % of cases. However, the substitution of the whole estimated probability distribution with a single value (mode, mean, or median) misses at least one aspect of the features of the responses: in our data set the observed data are positively skewed and it is almost impossible to predict responses as $R = 7, 8, 9$ (which have been observed in 124 cases, a relative frequency of 0.175). In fact, the prediction of these high categories requires the knowledge of further covariates which are explicitly related to these extreme values.

7 Concluding remarks

GeCUB models share with GLM framework the presence of stochastic and systematic components (given by (3) and (5), respectively) but differentiate in several aspects and a comparative discussion has been pursued by Iannario and Piccolo (2015). The main points are the following:

- Data generating process for GeCUB models concerns a probability mass function, whereas a distribution function is involved in the derivation of cumulative models.
- GeCUB and CUB models do not belong to the exponential family.
- In GeCUB models the link among parameters and covariates is a direct one, without involving moments.
- Parsimony of GeCUB models is an added value since no cutpoints are required in the estimation procedures; thus, GeCUB models should be preferred for larger m.

The specification of GeCUB models may be extended in several directions and we list few of them according to the current research.

- Some interesting relationships between the *shelter choice* and the presence of "don't know" responses in rating surveys have been recently emphasized by Manisera and Zuccolotto (2014a) with special reference to CUB models.
- The Binomial distribution for the feeling has been generalized with the introduction of a Beta-Binomial random variable to take into account a possible overdispersion in the ordinal data (Iannario 2012b, 2014), also with subjects' covariates (Piccolo 2015). A similar approach would lead to GeCUBE models.
- The standard structure of CUB and GeCUB models assumes a constant uncertainty for all categories. Some interesting improvements have been recently obtained by Gottard et al. (2015) who considered a varying uncertainty in the model by specifying an *a priori* distribution for the subjects' indecision. Similar considerations may be pursued by inserting a varying uncertainty in the GeCUB structure and this extension does not require further parameters to be estimated.
- In some circumstances, as in sensometric studies, it is convenient to introduce objects' covariates (Piccolo and D'Elia 2008) in the link of the parameters since consumers' preferences are undoubtedly conditioned by the sensory characteristics of the items under scrutiny (food or beverage, for instance). This proposal may be usefully applied to GeCUB models.
- When data are organized according to a hierarchical structure, it may be effective to consider multilevel models: thus, hierarchical CUB models have been introduced

(Iannario 2012d). This random effect might be extended to the GeCUB models in order to capture hierarchical structures and clusters variability.

To summarize, in this paper the main statistical issues of GeCUB models for studying ordinal data have been presented and this has been pursued according to a general framework derived by an interpretation of the data generating process for this kind of observations. More experience is necessary to implement some design able to detect and test the components of these mixtures. In addition, convenient and effective starting values for the EM procedure are required, as already obtained for CUB models (Iannario 2012c).

Both reported simulations and empirical analysis, although limited, suggest that this framework can accomplish the standard goals of the analysis of ordinal data with an added value in terms of interpretation of the components and their effects, parameters parsimony and immediate graphical facilities. As in any scientific investigation, a multiple perspective for modelling real phenomena should be considered a positive improvement of knowledge.

Acknowledgments Authors thank Associate Editor and referees for their constructive comments. This work has been realized with the support of FIRB2012 project at University of Perugia (code RBFR12SHVV) and the SHAPE project within the frame of Programme STAR (CUP E68C13000020003) at University of Naples Federico II, financially supported by UniNA and Compagnia di San Paolo.

Appendix 1: Justification for the shifted Binomial distribution

For a given m, respondents are requested to rate an item $\mathscr{I}$ over an ordinal scale consisting of the categories $\{A_1, A_2, \ldots, A_m\}$. We introduce the mapping $A_j \leftrightarrow j$, for $j = 1, 2, \ldots, m$, as a convenient simplification.

Let us assume that $\xi \in [0, 1]$ is a real number such that $(1 - \xi)$ is a (normalized) synthesis of the strength of attraction or the common consensus that the respondent feels towards the item $\mathscr{I}$. Depending on the specific context, this may be considered as a measure of agreeableness, preference, liking, etc. For easiness we denote it as *feeling*. Then, $\xi \to 0$ ($\xi \to 1$) means that respondents have a very much positive (negative) opinion towards the item.

Let X be the random variable generated by the selection of an ordinal category $x \in \{1, 2, \ldots, m\}$ such that x increases with the feeling towards the item. Let us define E_j the event "A_x is preferred to A_j" for any $j \neq x$. Then, a respondent selects the category A_x if "A_j are considered too weak with respect to A_x to express his/her preference" for $j = 1, 2, \ldots, x - 1$ and "A_j are considered too strong with respect to A_x to express his/her preference" for $j = x + 1, x + 2, \ldots, m$.

Any typical sequence:

$$E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_{x-1}} \cap \overline{E}_{i_{x+1}} \cap \overline{E}_{i_{x+2}} \cap \cdots \cap \overline{E}_{i_m}$$

generates the selection of the category A_x and it consists of $(x - 1)$ "successes" (when A_x is considered more adequate than A_j to express preference) and of $(m - x)$ "failures" (when A_x is considered less adequate than A_j to express preference).

The realization $X = x$ corresponds to one of those sequences whose number are the permutations with replacement of the $(m-1)$ paired comparisons A_j versus A_x, $\forall j \neq x$ with $(x-1)$ "successes" and $(m-x)$ "failures", that is: $\frac{(m-1)!}{(x-1)!\,(m-x)!} = \binom{m-1}{x-1} = \binom{m-1}{m-x}$, for $x = 1, 2, \ldots, m$.

If we define, for any $x = 1, 2, \ldots, m$,

$$1 - \xi_j \;=\; Pr\left(A_x \text{is preferred to} A_j\right) \;=\; Pr\left(E_j\right), \qquad j \neq x\,,$$

then

$$Pr\,(X = x) = \binom{m-1}{x-1} (1-\xi_1)\,(1-\xi_2)\,\ldots\,(1-\xi_{x-1})\,\xi_{x+1}\,\xi_{x+2}\,\ldots\,\xi_m\,,$$
$$x = 1, 2, \ldots, m\,.$$

The previous argument supports the very nature of a (shifted) Binomial random variable X since it "counts" the number of successes of a sequence of paired comparison experiments, which we assume to be independent to simplify the derivation. In a sense, the proposed (shifted) Binomial random variable counts the "successes" in repeated trials that is, in our interpretation, it *counts* the number of times a category "wins" against the lower ones.

We observe that the Binomial component induces an *ordinal restraint* in the mixture model since each value $X = x$ may be interpreted as the result of cumulated choices against different alternatives. As an instance, for $m = 9$, the event $(X = 7)$ is related to the event $(X = 6)$ since in the former choice we reject more low values than in the latter one.

Independence is not a severe constraint since the sum of non-independent experiments may be approximated by an equivalent sum of independent experiments.

The whole family of CUB models may be derived upon the specification of the sequence ξ_k, $k = 1, 2, \ldots, m$.

1. CUB models (Piccolo 2003) implies: $\xi_1 = \xi_2 = \cdots = \xi_m = \xi$. This is a strong assumption but it may be maintained if one considers the constant ξ as an average of different ξ_j. In fact, all these quantities are high (low) when *feeling* is small (large).
2. CUB models with a *shelter* effect (Iannario 2012a) assume that a category $c \in \{1, 2, \ldots, m\}$ has an additional probability to be chosen since it is considered as a "safer" or "more convincing" choice.
3. CUBE models (Iannario 2014) assume that ξ is a Beta random variable so that X is the predictive distribution (=Beta-Binomial).
4. Non-linear CUB models (Manisera and Zuccolotto 2014b) assume non-constant transition probabilities between a category and the adjacent one.

Notice that $Pr\left(E_j\right)$ is defined in terms of $1 - \xi_j$ (and not of ξ_j) since the first interpretation of this approach has been derived in a ranking context; then, the preference for the item is higher when it is located in the first positions of the scale. As a consequence, the ratings interpretation of ξ_j is reversed.

The shifted Binomial distribution for fitting ordinal data has been first elaborated in a ranking framework as a tool for modelling choices generated by a paired comparison mechanism: D'Elia (2000a, pp. 179–181, 2000b). Then, Piccolo (2003) and D'Elia and Piccolo (2005) extended the approach to preference analysis.

More recently, it has been used by Zhou and Lange (2009) who interpret multiple ratings expressed by panellists on movies as a mixture of a "common consensus" (anchored to respondents by means of an individual shifted Binomial distribution) and a "quirkiness behaviour" which corresponds to *feeling* and *uncertainty*, respectively. Differently from CUB models with subjects' and objects' covariates (Iannario and Piccolo 2012a; Piccolo and D'Elia 2008), their approach requires a huge number of parameters (in fact, the paper is mainly motivated by a computational burden) since no relationships about subjects' characteristics have been introduced. Finally, a discussion and interpretation of the (shifted) Binomial distribution has been advanced by Allik (2014) to support the idea of a genuinely discrete generating process of the responses to a given item in the context of Likert-type personality measures.

Appendix 2: EM algorithm for a GeCUB model

Given the sample data $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$, we introduce the unobservable vector $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n)'$ where $\boldsymbol{z}_i = (z_{1i}, z_{2i}, z_{3i})'$ is a three-dimensional vector such that, for $g = 1, 2, 3$:

$$z_{gi} = \begin{cases} 1, & \text{if the } i\text{-th subject belongs to the } g \text{ component } \mathscr{P}_g; \\ 0, & \text{otherwise.} \end{cases}$$

Then, according to the specification of Sect. 4, the likelihood function of the complete-data vector $(\boldsymbol{r}', \boldsymbol{z}')'$ is given by:

$$L_c(\boldsymbol{\theta}) = \prod_{g=1}^{3} \prod_{i=1}^{n} \left[\alpha_{gi}(\boldsymbol{\psi}_g)\, p_g(r_i;\, \boldsymbol{\eta}_g)\right]^{z_{gi}} ,$$

and the complete-data log-likelihood function is:

$$\ell_c(\boldsymbol{\theta}) = \sum_{g=1}^{3} \sum_{i=1}^{n} \left[z_{gi}\, \log(\alpha_{gi}(\boldsymbol{\psi}_g)) + z_{gi}\, \log(p_g(r_i;\, \boldsymbol{\eta}_g))\right].$$

If we specify starting values $\boldsymbol{\theta}^{(0)}$, the EM algorithm at the $(k+1)$-th iteration is made up of the following steps:

- *E-step*:

The conditional expectation of the indicator random variable Z_{gi}, defined in Table 3, given the observed sample $\boldsymbol{r}$, is:

$$\mathbb{E}\left(Z_{gi} \mid \boldsymbol{r}, \boldsymbol{\theta}^{(k)}\right) = Pr\left(Z_{gi} = 1 \mid \boldsymbol{r}, \boldsymbol{\theta}^{(k)}\right) = \frac{\alpha_{gi}(\boldsymbol{\psi}_g^{(k)})\, p_g(\boldsymbol{r}; \boldsymbol{\eta}_g^{(k)})}{\sum_{j=1}^{3} \alpha_{ji}(\boldsymbol{\psi}_j^{(k)})\, p_j(\boldsymbol{r}; \boldsymbol{\eta}_j^{(k)})} = \tau_{gi}^{(k)} = \tau_{gi}\,,$$

for $g = 1, 2, 3$ and $i = 1, 2, \ldots, n$. Hereafter, when this causes no confusion, we will omit the reference to the iteration number (k) in τ_{gi}. Observe that, for any g, the quantity τ_{gi} is the posterior probability that the i-th subject of the sample with the observed r_i belongs to the g-th component $\mathscr{P}_g$ of the mixture.

Given observed sample $\boldsymbol{r}$ and parameters $\boldsymbol{\theta}$, these probabilities may be assembled in a $3 \times n$ matrix $\boldsymbol{\Pi}$ defined by:

$$\Pi = \begin{pmatrix} \tau_{11} & \tau_{12} & \ldots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \ldots & \tau_{2n} \\ \tau_{31} & \tau_{32} & \ldots & \tau_{3n} \end{pmatrix}.$$

Since the columns of $\boldsymbol{\Pi}$ sum to 1, $\tau_{3i} = 1 - \tau_{1i} - \tau_{2i}, i = 1, 2, \ldots, n$.

The expected log-likelihood of the complete-data vector is obtained as:

$$\begin{aligned}
\mathbb{E}\left(\ell_c(\boldsymbol{\theta}^{(k)})\right) &= \sum_{g=1}^{3}\sum_{i=1}^{n} \tau_{gi}\, \left[\log(\alpha_{gi}(\boldsymbol{\psi}_g^{(k)})) + \log(p_g(r_i;\, \boldsymbol{\eta}_g^{(k)}))\right] \\
&= \sum_{i=1}^{n} \left[\tau_{1i}\, \log(\alpha_{1i}(\boldsymbol{\psi}_1^{(k)})) + \tau_{2i}\, \log(\alpha_{2i}(\boldsymbol{\psi}_2^{(k)})) + \tau_{3i}\, \log(\alpha_{3i}(\boldsymbol{\psi}_3^{(k)}))\right] \\
&+ \sum_{i=1}^{n} \left[\tau_{1i}\, \log(p_1(r_i;\, \boldsymbol{\eta}_1^{(k)})) + \tau_{2i}\, \log(p_2(r_i;\, \boldsymbol{\eta}_2^{(k)})) + \tau_{3i}\, \log(p_3(r_i;\, \boldsymbol{\eta}_3^{(k)}))\right] \\
&= \sum_{i=1}^{n} \tau_{1i}\, \log\left(\delta_i(\boldsymbol{\omega}^{(k)})\right) + \sum_{i=1}^{n} \tau_{2i}\, \log\left[\pi(\boldsymbol{\beta}^{(k)})(1 - \delta_i(\boldsymbol{\omega}^{(k)}))\right] \\
&+ \sum_{i=1}^{n} (1 - \tau_{1i} - \tau_{2i})\, \log\left[(1 - \pi(\boldsymbol{\beta}^{(k)}))(1 - \delta_i(\boldsymbol{\omega}^{(k)}))\right] + Q^*
\end{aligned}$$

where Q^* is independent from $\alpha_{gi}^{(k)}$ parameters. Then, we let:

$$\mathbb{E}\left(\ell_c(\boldsymbol{\theta}^{(k)})\right) = Q_1(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) + Q^*\,.$$

• *M-step*:

At the $(k+1)$-th iteration, we have to maximize the function:

$$\begin{aligned}
Q_1(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) &= \sum_{i=1}^{n} \tau_{1i}\, \log\left(\delta_i(\boldsymbol{\omega}^{(k)})\right) + \sum_{i=1}^{n} \tau_{2i}\, \log\left[\pi(\boldsymbol{\beta}^{(k)})(1 - \delta_i(\boldsymbol{\omega}^{(k)}))\right] \\
&+ \sum_{i=1}^{n} (1 - \tau_{1i} - \tau_{2i})\, \log\left[(1 - \pi(\boldsymbol{\beta}^{(k)}))(1 - \delta_i(\boldsymbol{\omega}^{(k)}))\right]
\end{aligned}$$

with respect to the parameter vector $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\beta}'^{(k)}, \boldsymbol{\omega}'^{(k)})'$.

Similarly, to find the parameter vector $\boldsymbol{\gamma}^{(k)}$, we need to maximize the function:

$$Q_2(\boldsymbol{\gamma}^{(k)}) = \sum_{i=1}^{n} \tau_{2i}\ \log(p_2(r_i;\ \boldsymbol{\eta}_2^{(k)})) = \sum_{i=1}^{n} \tau_{2i}\ \log(b_{r_i}(\boldsymbol{\gamma}^{(k)}))$$
$$\propto -\sum_{i=1}^{n} \tau_{2i}(r_i - 1)(\boldsymbol{w}_i\, \boldsymbol{\gamma}^{(k)}) - (m-1) \sum_{i=1}^{n} \tau_{2i} \log(1 + e^{-\boldsymbol{w}_i\, \boldsymbol{\gamma}^{(k)}})$$

Thus, the maximization step solves in finding parameter vectors such that:

$$\left(\boldsymbol{\beta}'^{(k+1)},\ \ \boldsymbol{\omega}'^{(k+1)}\right)' = \underset{\boldsymbol{\beta},\ \boldsymbol{\omega}}{argmax}\ \ Q_1(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)});$$
$$\boldsymbol{\gamma}^{(k+1)} = \underset{\boldsymbol{\gamma}}{argmax}\ \ Q_2(\boldsymbol{\gamma}^{(k)}).$$

These optimizations, generally, requires numerical methods.

Then, the updated parameter vector $\boldsymbol{\theta}^{(k+1)} = \left(\boldsymbol{\beta}'^{(k+1)}, \boldsymbol{\gamma}'^{(k+1)}, \boldsymbol{\omega}'^{(k+1)}\right)$ will be used and the E- and M-step are repeated until a convergence criterion is satisfied.

The previous derivation may be conveniently expressed by means of a step-by-step implementation (in any formal computer language) as follows. Here, we have to set a fixed tolerance ϵ ($= 10^{-6}$, for instance) and assume that integers m and c are given.

0. $\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\beta}'^{(0)}, \boldsymbol{\gamma}'^{(0)}, \boldsymbol{\omega}'^{(0)}\right)'$; $\quad l^{(0)} = \ell\left(\boldsymbol{\theta}^{(0)}\right)$; $\quad \epsilon = 10^{-6}$.

1. $\alpha_{1i}^{(k)} = \dfrac{1}{1 + e^{-\boldsymbol{x}_i \boldsymbol{\omega}^{(k)}}}$; $\quad \alpha_{2i}^{(k)} = (1 - \alpha_{1i}^{(k)})\ \dfrac{1}{1 + e^{-\boldsymbol{y}_i \boldsymbol{\beta}^{(k)}}}$; $\quad \alpha_{3i}^{(k)} = 1 - \alpha_{1i}^{(k)} - \alpha_{2i}^{(k)}$; $i = 1, 2, \ldots, n$.

2. $p_{1i}^{(k)} = D_{r_i}^{(c)}$; $\quad p_{2i}^{(k)} = p_{2i}(\boldsymbol{\gamma}^{(k)}) = \dbinom{m-1}{r_i - 1} \dfrac{e^{-(r_i - 1)\, \boldsymbol{w}_i \boldsymbol{\gamma}^{(k)}}}{\left(1 + e^{-\boldsymbol{w}_i \boldsymbol{\gamma}^{(k)}}\right)^{m-1}}$; $\quad p_{3i}^{(k)} = \dfrac{1}{m}$; $i = 1, 2, \ldots, n$.

3. $\nu_{gi}^{(k)} = \alpha_{gi}^{(k)}\ p_{gi}^{(k)}$, $\quad g = 1, 2, 3$; $\qquad den_i^{(k)} = \nu_{i1}^{(k)} + \nu_{i2}^{(k)} + \nu_{i3}^{(k)}$; $\quad i = 1, 2, \ldots, n$.

4. $\tau_{gi}^{(k)} = \tau_g(r_i; \boldsymbol{\theta}^{(k)}) = \dfrac{\nu_{gi}^{(k)}}{den_i^{(k)}}$, $\quad g = 1, 2$; $\qquad \tau_{3i}^{(k)} = 1 - \tau_{1i}^{(k)} - \tau_{2i}^{(k)}$; $\quad i = 1, 2, \ldots, n$.

5. $\begin{cases} S_1(\boldsymbol{\omega}^{(k)}) = \sum_{i=1}^{n} \tau_{1i}^{(k)}\ \log\left(\alpha_{1i}^{(k)}\right); \\ S_2(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) = \sum_{i=1}^{n} \tau_{2i}\ \log\left(\alpha_{2i}^{(k)}\right); \\ S_3(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) = \sum_{i=1}^{n} \left(1 - \tau_{1i}^{(k)} - \tau_{2i}^{(k)}\right) \log\left(1 - \alpha_{1i}^{(k)} - \alpha_{2i}^{(k)}\right). \end{cases}$

6. $Q_1(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) = S_1(\boldsymbol{\omega}^{(k)}) + S_2(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}) + S_3(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)})$.

7. $Q_2(\boldsymbol{\gamma}^{(k)}) = -\sum_{i=1}^{n} \tau_{2i}^{(k)}(r_i - 1)(\boldsymbol{w}_i\, \boldsymbol{\gamma}^{(k)}) - (m-1)\ \sum_{i=1}^{n} \tau_{2i}^{(k)} \log(1 + e^{-\boldsymbol{w}_i\, \boldsymbol{\gamma}^{(k)}})$.

8. $\left(\boldsymbol{\beta}'^{(k+1)},\ \boldsymbol{\omega}'^{(k+1)}\right)' = \underset{\boldsymbol{\beta},\,\boldsymbol{\omega}}{argmax}\ Q_1(\boldsymbol{\beta}^{(k)}, \boldsymbol{\omega}^{(k)}); \quad \boldsymbol{\gamma}^{(k+1)} = \underset{\boldsymbol{\gamma}}{argmax}\ Q_2(\boldsymbol{\gamma}^{(k)}).$

9. $\boldsymbol{\theta}^{(k+1)} = \left(\boldsymbol{\beta}'^{(k+1)},\ \boldsymbol{\gamma}'^{(k+1)},\ \boldsymbol{\omega}'^{(k+1)}\right)'; \qquad l^{(k+1)} = \ell\left(\boldsymbol{\theta}^{(k+1)}\right).$

10. $\begin{cases} \text{if } l^{(k+1)} - l^{(k)} \geq \epsilon, & k \to k+1; \quad \text{go to 1;} \\ \text{if } l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k+1)}; \quad \text{stop.} \end{cases}$

Accurate initial values $\boldsymbol{\theta}^{(0)}$ for an effective starting of the EM algorithm would accelerate the convergence of the EM algorithm towards the ML estimates, as emphasized by McLachlan and Peel (2000), pp. 47–49 and Karlis and Xekalaki (2003) in general contexts, and confirmed by Iannario (2012c) for CUB models.

This issue deserves more studies and extensive experiments; in case of large sample size, initial values derived by simplified CUB models (without and with covariates and/or without covariates in the *shelter effect*) may be suggested. In absence of any information we might use $\boldsymbol{\theta}^{(0)} = (0.1, 0.1, \ldots, 0.1)'$. However, it is more effective to start with a random sampling of a subset of the full data set and to plug the obtained parameter estimates into the EM procedure as the preliminary ones.

References

Agresti A (2010) Analysis of ordinal categorical data, 2nd edn. Wiley, Hoboken

Allik J (2014) A mixed-binomial model for Likert-type personality measure. Front Psychol 5:1–13

Corduas M (2008a) Clustering CUB models by Kullback-Liebler divergence. Proceedings of SCF-CLAFAG Meeting, ESI, Napoli, pp 245–248

Corduas M (2008b) A statistical procedure for clustering ordinal data. Quad Stat 10:177–189

Corduas M (2011) A study on University students' opinions about teaching quality: a model based approach to clustering ordinal data. In: Attanasio M, Capursi V (eds) Statistical methods for the evaluation of university systems. Physica-Verlag, Springer, Berlin, pp 67–78

Corduas M, Iannario M, Piccolo D (2009) A class of statistical models for evaluating services and performances. In: Bini M et al (eds) Statistical methods for the evaluation of educational services and quality of products, contribution to statistics. Physica-Verlag, Springer, Berlin, pp 99–117

Cox C (1995) Location-scale cumulative odds models for ordinal data: a generalized non-linear model approach. Stat Medic 14:1191–1203

D'Elia A (2000a) The mechanism of paired comparisons in rank modelling: statistical issues and critical considerations (in Italian). Quad Stat 2:173–203

D'Elia A (2000b) A shifted Binomial model for rankings. In: Nunez-Anton V, Ferreira E (eds) "Statistical Modelling", XV international workshop on statistical modelling, Servicio Editorial de la Universidad del Pais Vasco, pp 412–416

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

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). J R Stat Soc Ser B 39:1–38

Di Iorio F, Iannario M (2012) Residual diagnostics for assessing the fit of CUB models. STATISTICA LXXII, pp 163–172

Gottard A, Iannario M, Piccolo D (2015) Varying uncertainty in CUB models. Submitted

Iannario M (2009) Fitting measures for ordinal data models. Quad Stat 11:39–72

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

Iannario M (2012b) CUBE models for interpreting ordered categorical data with overdispersion. Quad Stat 14:137–140

Iannario M (2012c) Preliminary estimators for a mixture model of ordinal data. Adv Data Anal Classif 6:163–184
Iannario M (2012d) Hierarchical CUB models for ordinal variables. Comm Stat Theory Meth 41:3110–3125
Iannario M (2014) Modelling uncertainty and overdispersion in ordinal data. Comm Stat Theory Meth 43:771–786
Iannario M, Piccolo D (2012a) 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 (2012b) A framework for modelling ordinal data in rating surveys. In: Proceedings of joint statistical meetings, section on statistics in marketing, San Diego, California, pp 3308–3322
Iannario M, Piccolo D (2015) Cumulative and CUB models for ordinal data: a comparative analysis. Submitted
Karlis D, Xekalaki E (2003) Choosing initial values for the EM algorithm for finite mixtures. Comput Stat Data Anal 41:577–590
Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cogn Psychol 5:213–236
Laakso M, Taagepera R (1989) Effective number of parties: a measure with application to West Europe. Compar Polit Stud 12:3–27
McCullagh P (1980) Regression models for ordinal data (with discussion). J R Stat Soc Ser B 42:109–142
McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London
McLachlan G, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, New York
McLachlan G, Peel GJ (2000) Finite mixture models. Wiley, New York
Manisera M, Zuccolotto P (2014a) Modeling "don't know" responses in rating scales. Patt Recogn Lett 45:226–234
Manisera M, Zuccolotto P (2014b) Modeling rating data with nonlinear CUB models. Comput Stat Data Anal 78:100–118
Peterson B, Harrell FE (1990) Partial proportional odds models for ordinal responses variables. Appl Stat 39:205–217
Nelder JA, Wedderburn RWM (1972) Generalized linear models. J R Stat Soc Ser A 135:370–384
Piccolo D (2003) On the moments of a mixture of uniform and shifted binomial random variables. Quad Stat 5:85–104
Piccolo D (2006) Observed information matrix for MUB models. Quad Stat 8:33–78
Piccolo D (2015) Statistical issues for CUBE models with covariates. Comm Stat Theory Meth 44. doi:10.1080/03610926.2013.821487
Piccolo D, D'Elia A (2008) A new approach for modelling consumers' preferences. Food Qual Pref 19:247–259
Powers DA, Xie Y (2000) Statistical methods for categorical data analysis. Academic Press, San Diego, CA
Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
Simon HA (1957) Models of man. Wiley, New York
Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press, Cambridge
Tutz G (2012) Regression for categorical data. Cambridge University Press, Cambridge
Zhou H, Lange K (2009) Rating movies and rating the raters who rate them. Am Stat 63:297–307