A Framework for Modelling Ordinal Data in Rating Surveys

Maria Iannario *   Domenico Piccolo†

Abstract
In this paper, we generalize a mixture model proposed for ordinal data by specifying subjects' covariates for each component of the mixture. Then, the EM procedure for obtaining the maximum likelihood estimates by means of an effective iterative algorithm has been derived and the asymptotic variance-covariances matrix of parameter estimators has been carried out: these results allow an asymptotically efficient inference. In this respect, we present a step-by-step algorithm in order to perform an effective programming for the implementation of a related software. In addition, we checked the usefulness of the approach by means of a real case study. Some final considerations for future developments conclude the paper.

Key Words: G*e*CUB models, ordinal data modelling, mixture models, asymptotic ML inference

1. Introduction

In statistical surveys, people are often asked to express judgements or evaluations on some topic of interest and, generally, some information about respondents are also collected. The ordered evaluation may concern several aspects of their personal opinions but for simplifying the discussion we assume that answers are in one-to-one correspondence with integers $\{1, 2, \ldots, m\}$ where m is given and known. Then, hereafter, we will speak of ratings and subjects' covariates to refer to ordinal responses and information collected on the respondents, respectively.

Our objective is to explain, fit, and forecast the probability $Pr\,(R = r)$ that a discrete random variable R assumes values $r = 1, 2, \ldots, m$, and we will use subjects' covariates (when available and significant) in order to improve the performance and the interpretation of the estimated probability distribution. For this purpose, we introduce a model in which we assume that the final outcome of the evaluation process is a discrete observation generated by an investigated trait which is intrinsically continuous.

In experimental surveys, the sample data consist of a collection of ordered scores, generally chosen on a Likert scale anchored to the integers $\{1, 2, \ldots, m\}$ for some known m. Thus, respondents choose a qualitative assessment on a graduated sequence of verbal definitions ("extremely satisfied", "very satisfied", ..., "extremely unsatisfied", for instance) which are coded as numbers just for convenience. This circumstance generates ordinal data which require the introduction of specific statistical methods (see Agresti, 2010); most of them rely on the General Linear Models (GLM) framework, introduced by Nelder and Wedderburn (1972), McCullagh and Nelder (1989) and specifically discussed by McCullagh (1980) for ordinal data.

An alternative approach, mainly motivated by the investigation of the psychology of the respondents, have been introduced by Piccolo (2003) and denoted as CUB models (since they are a convex *C*ombination of discrete *U*niform and shifted *B*inomial random variable): such structures have been generalized in several directions (Iannario and Piccolo 2012).

*Department TEOMESUS, Statistical Sciences Unit, University of Naples Federico II, Via Leopoldo Rodinò 22, Naples, I-80138, Italy

†Department TEOMESUS, Statistical Sciences Unit, University of Naples Federico II, Via Leopoldo Rodinò 22, Naples, I-80138, Italy

A peculiar aspect of this class of models is the possibility to explicitly interpret the parameters in relation to the *feeling* of the respondent, the *uncertainty* of the responses and a possible *shelter effect*, also with reference to subjects' and objects' covariates. Such generalized models have been firstly introduced by Corduas *et al.* (2009) and Iannario (2012). Given significant characteristics of the subjects, the approach allows for designing profiles and specifying clusters of respondents (Corduas 2008; 2011).

In this paper, we take into account the inclusion of subjects' covariates on all the components of the extended mixture: this generalization will be denoted as G*e*CUB models. More specifically, we will specify G*e*CUB models, define the corresponding likelihood function when a random sample has been observed, deduce the EM algorithm for the parameter estimation and compute variance-covariances matrix of parameter estimators in order to perform asymptotically efficient statistical inference.

The work is organized as follows: in the next section, we establish notations and specify the G*e*CUB models. Then, in section 3 we derive the maximum likelihood (ML) inference and detail the main steps required for the EM algorithm. The asymptotic variance-covariances matrix of ML estimators is obtained in section 4 whereas in section 5 we consider some simplified G*e*CUB models that are common in real applications. A real case study is discussed and some final remarks conclude the paper. Computational details are deferred to Appendix.

2. Specification of GeCUB models

Suppose that people are requested to rate their opinion/evaluation about an item on a Likert scale which is in a one-to-one correspondence with the support $I(m) = \{1, 2, \dots, m\}$, where m is a prefixed integer. For a given $c \in I(m)$, we will define shifted Binomial, discrete Uniform and degenerate (at $R = c$) random variables, respectively, as:

$$b_r(\xi) = \binom{m-1}{r-1}\xi^{m-r}(1-\xi)^{r-1}; \qquad U_r = \frac{1}{m}; \qquad D_r^{(c)} = \begin{cases} 1, & \text{if } r = c; \\ 0, & \text{otherwise}; \end{cases}$$

for $r = 1, 2, \dots, m$.

These distributions should be considered as the building blocks of the data generating process by which a respondent selects an ordinal modality belonging to $I(m)$. In fact, as depicted in Figure 1, we assume that, when faced to a given item, each respondent adopts a two step strategy:

- first of all, he/she chooses between a simplistic option (*a shelter choice*) consisting in the selection between a modality which he/she considers very attractive (by the nature of verbal wording and/or the numbering of the scale) and a meditated response which requires some thinking about. We assume that this choice happens with probabilities δ and $1 - \delta$, respectively;

- in the second option, he/she selects a modality in the support $\{1, 2, \dots, m\}$ which is the final choice of a balanced decision related to his/her feeling perception or to a totally random choice, with a propensity π and $1 - \pi$, respectively.

This model considers the final choice as a decision between a *shelter* option and a standard CUB model and thus the observed response r is the realization of a random variable R whose probability distribution is defined by:

$$Pr\,(R = r) = \delta\left[D_r^{(c)}\right] + (1-\delta)\left[\pi\, b_r(\xi) + (1-\pi)U_r\right], \quad r = 1, 2, , \dots, m.$$

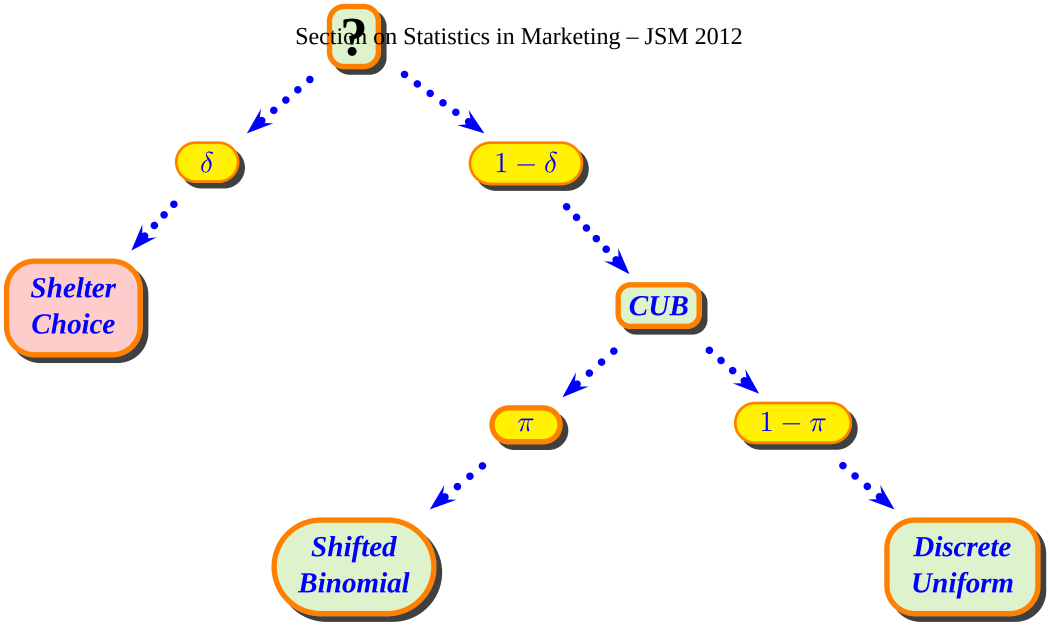

Figure 1: Logical steps in the data generating process of a GeCUB model.

This probability distribution has been denoted CUB model with a *shelter effect* by Iannario (2012a), who discusses its properties, estimation issues and related topics. If we let $\boldsymbol{\theta} = (\pi, \xi, \delta)'$ the parameter vector, the parameter space $\Omega^*(\boldsymbol{\theta})$ is the (left open) unit cube:

$$\Omega^*(\boldsymbol{\theta}) = \{(\pi, \xi, \delta) : \; 0 < \pi \le 1, \;\; 0 \le \xi \le 1, \;\; 0 \le \delta \le 1\}.$$

In this paper, we conjecture that each of the three components of the mixture may be related to subjects' covariates. Suppose that we have information on the n subjects summarized by a set of v variables in the matrix $\boldsymbol{T} = ||t_{ij}, \; i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, v||$. We consider the matrices $\boldsymbol{Y}$, $\boldsymbol{W}$, $\boldsymbol{X}$ obtained by $\boldsymbol{T}$ by selecting convenient columns and introduce the variables Y_0, W_0, X_0 that assume the constant value 1 for all the sample units. Then, we denote by $\boldsymbol{y}_i$, $\boldsymbol{w}_i$, and $\boldsymbol{x}_i$, for $i = 1, 2, \ldots, n$, the i-th row of the $\boldsymbol{Y}$, $\boldsymbol{W}$ and $\boldsymbol{X}$ matrices, respectively, that is:

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

These rows contain all available sample information on the i-th subject and are necessary and sufficient for the model specification. We let $\mathcal{C}_i = (\boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{x}_i)$, $i = 1, 2, \ldots, n$, for convenience.

According to a general paradigm already exploited for standard CUB models, we introduce a logistic link among the parameters and covariates:

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

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)'$ and $\boldsymbol{\omega} = (\omega_0, \omega_1, \ldots, \omega_p)'$, respectively. Alternative links are admissible.

Notice that, given our parameterization, the matrices $\boldsymbol{Y}$, $\boldsymbol{W}$, $\boldsymbol{X}$ may or may not possess an arbitrary number of common columns. In addition, given the finiteness of the covariates, the previous deterministic relationships generate an open unit cube as admissible parametric space $\Omega(\boldsymbol{\theta})$ of a GeCUB model:

$$\Omega(\boldsymbol{\theta}) = \{(\pi, \xi, \delta) : \; 0 < \pi < 1, \;\; 0 < \xi < 1, \;\; 0 < \delta < 1\}.$$

Ratings $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ are realizations of the random sample $(R_1, R_2, \ldots, R_n)'$, where each R_i is identically and independently distributed, for a given $m > 4$ and known c, as a discrete random variable R over the support $I(m)$. For a given i-th subject, for $i = 1, 2, \ldots, n$, the three distributions of the mixture are given by:

$$b_{r_i}(\boldsymbol{\gamma}) = \binom{m-1}{r_i-1} \frac{e^{-\boldsymbol{w}_i\,\boldsymbol{\gamma}(r_i-1)}}{(1+e^{-\boldsymbol{w}_i\,\boldsymbol{\gamma}})^{m-1}}; \qquad U_{r_i} = \frac{1}{m}; \qquad D^{(c)}_{r_i} = \begin{cases} 1, & \text{if } r_i = c\,; \\ 0, & \text{otherwise}\,. \end{cases}$$

Then, a GeCUB model is fully specified by:

$$\begin{cases} Pr\,(R = r_i \mid \mathcal{C}_i,\, \boldsymbol{\theta}) = \delta_i \left[D^{(c)}_{r_i} \right] + (1-\delta_i) \left[\pi_i\, b_{r_i}(\gamma) + (1-\pi_i) U_{r_i} \right]; \\ \pi_i = \dfrac{1}{1+e^{-\boldsymbol{y}_i \boldsymbol{\beta}}}\,; \qquad \xi_i = \dfrac{1}{1+e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}}\,; \qquad \delta_i = \dfrac{1}{1+e^{-\boldsymbol{x}_i \boldsymbol{\omega}}}\,. \end{cases}$$

We denote this structure as a GeCUB (p, q, s) model to indicate the number of corresponding covariates. Figure 2 sketches the relationships among the information present in a GeCUB model.

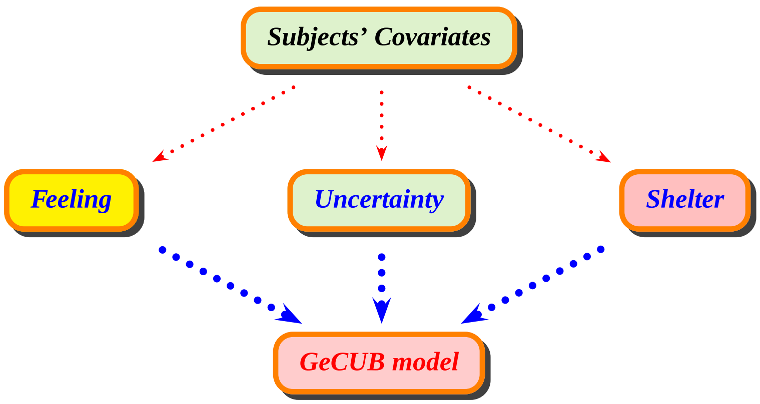

Figure 2: The role of subjects' covariates in a GeCUB model.

We observe that, for a fully specified GeCUB model, the length of the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\omega}')'$ is $(p + q + s + 3)$ and hereafter we will sketch the inferential procedures for a fully specified GeCUB model. In case some or all subjects' covariates are absent, the steps are greatly simplified and, in these circumstances, it is more convenient to refer to specific algorithms, as derived by Piccolo (2006) and Iannario (2012a) for CUB models without and with *shelter effect*, respectively.

3. Maximum likelihood inference

In a mixture distribution, it is useful to formally characterize the notation of the parameters according to their roles, as in MacLachlan and Peel (2000), for instance. Thus, we denote by $\boldsymbol{\theta} = (\boldsymbol{\psi}',\, \boldsymbol{\eta}')'$ the full parameter vector of a GeCUB model and by $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ the parameter vectors of weights (α_g) and of probability distributions $(\mathcal{P}_g)$, respectively, for $g = 1, 2, 3$ (see Table 1).

Given the sample $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ and the information set of covariates $\mathcal{C}_i$, for

Table 1: Notation for the components of the mixture in a GeCUB model.

g	$\alpha_{gi} = \alpha_{gi}(\boldsymbol{\psi}_g)$	$p_{gi} = p_g(r_i;\, \boldsymbol{\eta}_g)$	$\boldsymbol{\psi}_g$	$\boldsymbol{\eta}_g$	Z_{gi}
1	δ_i	$D^{(c)}_{r_i}$	$\boldsymbol{\psi}_1 = \boldsymbol{\omega}$		$\begin{cases} 1, & \text{if } i \in \mathcal{P}_1\,; \\ 0, & \text{otherwise}\,. \end{cases}$
2	$\pi_i(1-\delta_i)$	$b_{r_i}(\boldsymbol{\gamma})$	$\boldsymbol{\psi}_2 = (\boldsymbol{\beta}',\, \boldsymbol{\omega}')'$	$\boldsymbol{\eta}_2 = \boldsymbol{\gamma}$	$\begin{cases} 1, & \text{if } i \in \mathcal{P}_2\,; \\ 0, & \text{otherwise}\,. \end{cases}$
3	$(1-\pi_i)(1-\delta_i)$	U_{r_i}	$\boldsymbol{\psi}_3 = (\boldsymbol{\beta}',\, \boldsymbol{\omega}')'$		$\begin{cases} 1, & \text{if } i \in \mathcal{P}_3\,; \\ 0, & \text{otherwise}\,. \end{cases}$

$i = 1, 2, \dots, n$, the log-likelihood function may be written as:

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

In Appendix A, we will detail the EM procedure for obtaining the ML estimators from $\ell(\boldsymbol{\theta})$. Then, we compute information matrix of GeCUB models in order to perform asymptotic statistical inference. Although both expected and observed information matrix share the same asymptotic properties, we will compute the observed information (that is the negative of the Hessian matrix) since it is preferred for both computational and statistical properties as discussed by Efron and Hinkley (1978), Lloyd (1999, 30-31) and Pawitan (2001, 244-247), among others.

For well behaved log-likelihood functions, such a matrix is generally obtained by numerical computations. However, it is more accurate to derive analytically the second order derivatives as detailed in Appendix B. The negative inverse of the information matrix is the asymptotic variance-covariance matrix of ML estimators.

Most ratings survey are generally based on moderate or large sample size, and thus standard tests of fitting as X^2 are not adequate since they invariably tend to reject the model even in case of good overlapping among observed and expected frequencies. Moreover, in presence of covariates, it is not immediate to define an omnibus test for the data that is able to take their relevance into account (as it happens for dissimilarity indexes, for instance). Thus, we prefer to rely the validation of the model on both the parameters significance and likelihood-based measures as $BIC = -2\,\ell(\hat{\boldsymbol{\theta}}) + (p+q+s+3)\,\log(n)$, for instance.

Simulation experiments (here not reported for brevity) confirmed the adequacy of the ML estimation method for different number of categories (m) and samples of moderate/large size.

4. A real case study

We refer to PLUS, a cross-sectional survey carried out during 2006 by ISFOL (Institute for training of workers, Ministry of Labour and Welfare, Italy) by means of a well structured questionnaire (CATI method) and modelled by CUB models in Iannario and Piccolo (2010).

We will consider the quantitative expression of subjective survival probabilities which has been collected in the following way:

For scientific purposes only, we would like to ask you: "In your opinion, what is the probability that you will reach age 75?" Please provide a value between 100 (certain event) and 0 (impossible event).

All subsequent analyses are based on validated and consistent responses given by a sample of $n = 20,184$ people of age 15-64 years and such information are collected with many subjects' covariates.

Several circumstances act on the expressed subjective survival probabilities: rounding, selective and miscellanea effects and these arguments raised by Iannario and Piccolo (2010) lead to consider these subjective probabilities as a qualitative ordinal judgment about the occurrence of an event.

Then, we will study the expressed evaluation of subjective survival probabilities by means of a 7-point Likert scale as described in Table 2.

Table 2: Ordinal correspondence for the expressed subjective probability.

R	Subjective survival probability	Ordinal interpretation
1	$0.00 \leq Pr\,(S) \leq 0.05$	`IMPOSSIBLE/Almost IMPOSSIBLE`
2	$0.05 < Pr\,(S) \leq 0.25$	`LOW`
3	$0.25 < Pr\,(S) \leq 0.45$	`Moderately LOW`
4	$0.45 < Pr\,(S) \leq 0.55$	`About FIFTY/FIFTY`
5	$0.55 < Pr\,(S) \leq 0.75$	`Moderately HIGH`
6	$0.75 < Pr\,(S) \leq 0.95$	`HIGH`
7	$0.95 < Pr\,(S) \leq 1.00$	`SURE/Almost SURE`

According to the previous wording, we should interpret *uncertainty* as the indecision in formulating the response to the previous question whereas *feeling* is the confidence that the respondent relies on the probability to survive at 75 years.

Previous experience showed that `Age` is a relevant covariate for explaining both uncertainty and confidence to survive at 75 years; then, we will check if this covariate may be usefully exploited for explaining (and quantifying) also the *shelter effect* very pronounced at $R = 7$ (people give an excess of probability to this event and modify the confidence towards this modality during their life).

Since `Age` is a continuous covariate, it is convenient (for computational and statistical purposes) to transform it by considering the deviation of the average after logging, that is:

$$\widetilde{Age_i} = \log(Age_i) - \overline{\log(Age_i)}, \quad i = 1, 2, \ldots, n\,.$$

In fact, this transformation improves convergence and drastically reduces correlations among estimators.

All computations have been implemented by a programm in the GAUSS language by using ML methods and exploiting the EM procedure for convergence (as in Appendix A). Standard errors have been computed by analytical derivation of the observed information matrix with ML estimates plugged in (as in Appendix B).

In Table 3, we list the main inferential results when we fit a sequence of nested models to our data: CUB models without and with covariates, then with a *shelter* effect, and finally by a G*e*CUB model. At last, `Age` turned out to be a relevant covariate for explaining both uncertainty and confidence to survive at 75 years; in addition, this covariate is quite significant for the *shelter effect* and its inclusion really improves the model.

Table 3 confirms the usefulness to fit a G*e*CUB model which consistently preserves the sign and the value of the uncertainty and confidence parameters; in addition, this model improves the fitting as confirmed by log-likelihoods and the values of BIC which regularly decreases from $60,786$ for the CUB model down to $59,797$ for the final G*e*CUB model, despite the increasing number of estimated parameters.

Table 3: Estimation of CUB and GeCUB models for the expressed subjective probabilities.

Models	Covariates	Uncertainty parameters	Confidence parameters	*Shelter* parameters	$\ell(\hat{\boldsymbol{\theta}})$
CUB		$\hat{\pi} =$ 0.867 *(0.005)*	$\hat{\xi} =$ 0.163 *(0.001)*		$-30,383$
CUB	*Constant*	$\hat{\beta}_0 =$ 1.507 *(0.059)*	$\hat{\gamma}_0 = -1.551$ *(0.016)*		
+	$\widetilde{Age}$		$\hat{\gamma}_1 = -0.112$ *(0.023)*		
covariates	$\widetilde{Age^2}$	$\hat{\beta}_2 =$ 1.701 *(0.288)*	$\hat{\gamma}_2 = -0.504$ *(0.070)*		$-30,291$
CUB + *shelter*		$\hat{\pi} =$ 0.886 *(0.005)*	$\hat{\xi} =$ 0.219 *(0.002)*	$\hat{\delta} =$ 0.191 *(0.006)*	$-30,004$
GeCUB	*Constant*	$\hat{\beta}_0 =$ 1.777 *(0.069)*	$\hat{\gamma}_0 = -1.158$ *(0.019)*	$\hat{\omega}_0 = -1.489$ *(0.040)*	
	$\widetilde{Age}$		$\hat{\gamma}_1 =$ 0.223 *(0.030)*	$\hat{\omega}_1 =$ 0.975 *(0.092)*	
	$\widetilde{Age^2}$	$\hat{\beta}_2 =$ 1.936 *(0.345)*	$\hat{\gamma}_2 = -0.572$ *(0.071)*		$-29,864$

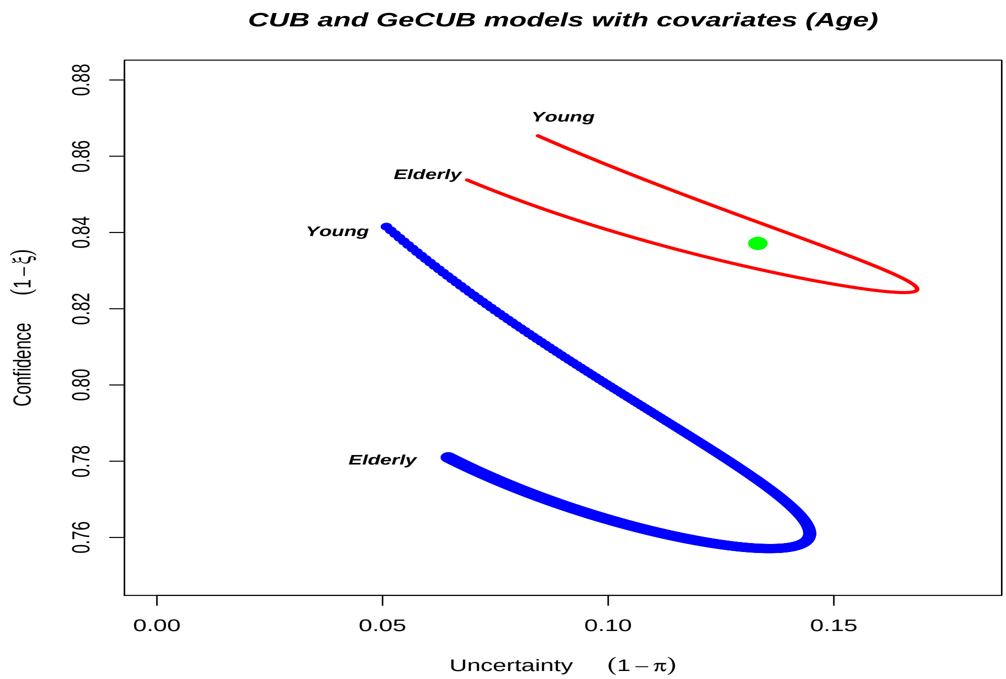

Figure 3: Dynamic visualization of CUB and GeCUB models with respect to Age.

Since a CUB model may be visualized in the parametric space as single point, we plot a sequence of CUB models for selected (π, ξ) to see how the covariate Age modifies both the confidence and uncertainty of respondents with respect to the question of survival at 75 years (Figure 3). More specifically, the single point expresses the location of the estimated CUB model without covariates and confirms the low uncertainty of respondents and the substantial confidence in assessing the probability to survive at age 75. These components varies with the respondents' Age and thus we plot the parametric curve:

$$\pi = \pi(Age); \qquad \xi = \xi(Age); \qquad \text{for Age} \in (15, 64).$$

In this way, we realize that the confidence lowers with increasing Age up to a minimum (estimated at Age=29 years) when it raises again to reach previous values for elderly.

Finally, if we consider that a *shelter effect* is significant and is related to Age (see last model in Table 3) the final GeCUB model may be depicted by the parametric curves:

$$\hat{\pi} = \pi(Age); \quad \hat{\xi} = \xi(Age); \quad \hat{\delta} = \delta(Age); \qquad \text{for Age} \in (15, 64).$$

It seems evident that the inclusion of a peculiar effect for the response at $R = 7$, lowers the *basic* confidence to a considerable degree with respect to a CUB model without considering a *shelter effect*. Moreover, it decreases with Age and after a turning point (now, more realistically located at Age$=40$ years) does not raise so much for elderly. In the plot, the thickness of the second curve is proportional to the increasing value of δ. If one excludes the varying *shelter effect* of Age on the responses the interpretation and the prediction may be biased.

Thus, the inclusion of a covariate in a CUB model allows for a clearer picture since we can see how confidence and indecision jointly change with the respondents' Age. In addition, also the *shelter effect* (well pronounced at $R = 7$) changes the pattern with Age.

The usefulness of considering G*e*CUB models as an improvement of CUB models with covariates becomes even more evident if we consider the profiles of the estimated probabilities for given values of the covariates: this kind of experiment is useful for prediction purpose.

In Figure 4, we present the corresponding estimated profiles of a CUB model distribution without covariates, with covariates, and according to a G*e*CUB model. For effective comparisons, the profiles have been obtained for respondents of 30 and 60 years, respectively.

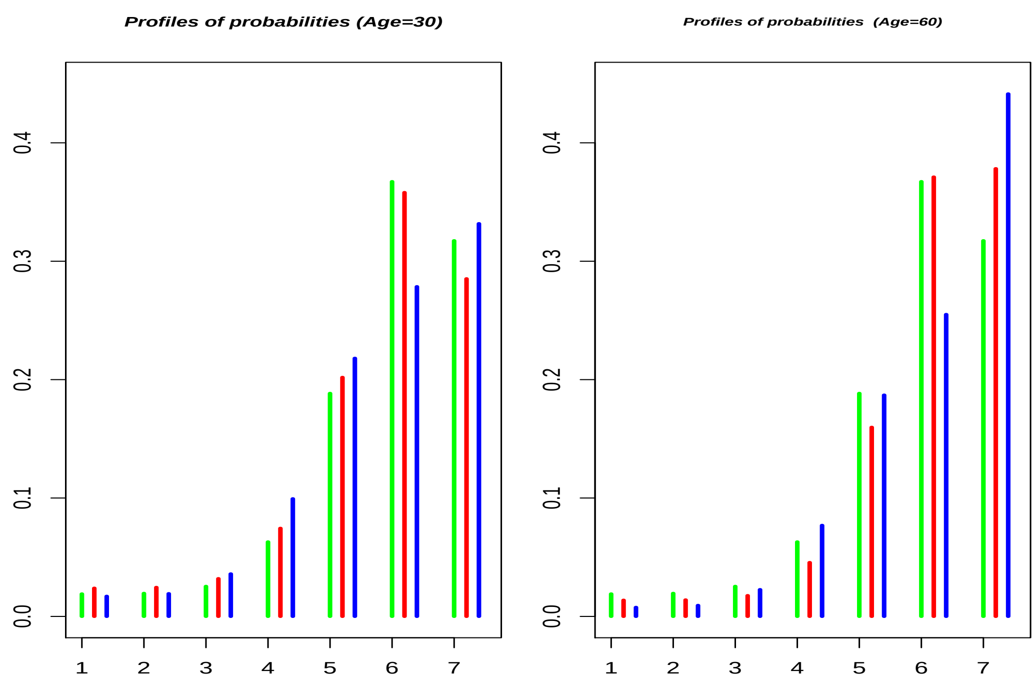

Figure 4: Profile distributions of CUB and G*e*CUB models, given the Age of respondent (left bars: CUB without covariates; intermediate: CUB with covariates; right: G*e*CUB).

The effect of considering a G*e*CUB model is not so high for young but it becomes relevant for elderly since the models without *shelter effect* considerably bias the estimated probability at $R = 6$ and $R = 7$.

5. Further extensions of GeCUB models

It is possible to extend the specification of G*e*CUB models in several directions and we limit ourselves to list a few of them according to the current research which exploits this framework for modelling ordinal data.

- The assumed probability distribution for the feeling has been recently generalized by Iannario (2012c) who introduced a Beta-Binomial distribution to take into account the presence of a possible overdispersion in ordinal data. A similar approach may be also implemented for G*e*CUB models.

- In some circumstances, as in sensometric studies, for instance, it is convenient to introduce some objects' covariates in the link of the parameters since consumers' preferences are undoubtedly conditioned by the sensory characteristics of the item under scrutiny (food or drink). This proposal has been introduced by Piccolo and D'Elia (2008) and may be usefully applied to G*e*CUB models.

- The standard structure of CUB and G*e*CUB models assumes a constant uncertainty whereas some interesting improvements have been recently obtained by Gottard *et al.* (2012) 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 G*e*CUB structure. In this new specification the probability distribution of the uncertainty is supposed to be known on *a priori* basis; thus, this extension does not require further parameters to be estimated.

- When data are organized according to a hierarchical structure, it may be effective to consider multilevel models: according to this line of reasoning, hierarchical CUB models have been introduced by Iannario (2012d). This random effect parameters might be introduced in the G*e*CUB models in order to capture hierarchical structures and the clusters variability.

6. Concluding remarks

In this paper, we have presented the main statistical issues of G*e*CUB models for studying ordinal data. More experience is necessary in order to improve some numerical aspects and we quote, first of all, the opportunity to derive convenient starting values for the EM procedure (as already obtained for CUB model: Iannario 2012b) and to implement more general software to cope with this class of models.

Results obtained by empirical analysis and simulation experiments suggest that a combination of multiple perspectives give higher coverage of the real data compared with the standard model (without extension or with covariates). Further investigations are needed to well understand how the differences between the two perspectives can be statistically detected.

Acknowledgement: The research has been partly supported by PRIN2008-MIUR project: "Modelling latent variables for ordinal data: statistical methods and empirical evidence" (PUC number E61J10000020001) and FARO2011 project, both at University of Naples Federico II. The first Author benefits from a grant received by Fulbright Scholar program while visiting Department of Statistics and Actuarial Science, University of Iowa, USA.

References

Agresti A. (2010) *Analysis of Ordinal Categorical Data*, 2nd edition, J.Wiley & Sons, Hoboken.

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

Corduas M. (2011) A study on University students' opinions about teaching quality: a model base approach to clustering ordinal data, in: Attanasio M. and Capursi V. (eds.), *Statistical Methods for the Evaluation of University Systems*, Physica-Verlag, Springer, Berlin Heidelberg.

Corduas M., Iannario M. and Piccolo, D. (2009) A class of statistical models for evaluating services and performances, in *Statistical methods for the evaluation of educational services and quality of products*, eds. Bini *et al.*, Springer-Verlag, Berlin Heidelberg, 99–117.

Efron B. and Hinkley D. V. (1978) Assessing the accuracy of maximum likelihood estimators: observed versus expected Fisher information, *Biometrika*, 65, 457–482.

Gottard A., Iannario M. and Piccolo D. (2012) *Varying uncertainty in* CUB *models*, submitted.

Iannario M. (2009) Fitting measures for ordinal data models, *Quaderni di Statistica*, 11, 39–72.

Iannario M. (2012a) Modelling *shelter* choices in a class of mixture models for ordinal responses, *Statistical Methods and Applications*, 21, 1–22.

Iannario M. (2012b) Preliminary estimators for a mixture model of ordinal data, *Advances in Data Analysis and Applications*, 6, forthcoming, DOI 10.1007/s11634-012-0111-5.

Iannario M. (2012c) CUBE models for interpreting ordered categorical data with overdispersion, *Quaderni di Statistica*, 14, 137–140.

Iannario M. (2012d) Hierarchical CUB models for ordinal variables, *Communications in Statistics. Theory and Methods*, 41, 3110–3125.

Iannario M. and Piccolo D. (2009) A package in R for CUB models inference, Version 2.0, available at www.dipstat.unina.it

Iannario M. and Piccolo D. (2010) Statistical modelling of subjective survival probabilities, *GENUS*, LXVI, 17–42.

Iannario M. and Piccolo D. (2012) CUB models: Statistical methods and empirical evidence, in *Modern Analysis of Customer Surveys: with applications using R*, Kenett R. S. and Salini S. (eds.), J. Wiley & Sons, Chichester, 231–258.

Karlis D. and Xekalaki E. (2003) Choosing initial values for the EM algorithm for finite mixtures, *Computational Statistics & Data Analysis*, 41, 577–590.

Lloyd C. J. (1999) *Statistical Analysis of Categorical Data*, J.Wiley & Sons, New York.

McLachlan G. and Peel G. J. (2000) *Finite Mixture Models*, J. Wiley & Sons, New York.

McCullagh P. (1980) Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society, Series B*, 42, 109–142.

McCullagh P. and Nelder J. A. (1989) *Generalized linear models*, 2^{nd} edition, Chapman & Hall, London.

Nelder J. A. and Wedderburn R. W. M. (1972) Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370–384.

Pawitan Y. (2001) *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Clarendon Press, Oxford.

Piccolo D. (2003) On the moments of a mixture of uniform and shifted binomial random variables, *Quaderni di Statistica*, 5, 85–104.

Piccolo D. (2006) Observed information matrix for MUB models, *Quaderni di Statistica*, 8, 33–78.

Piccolo D. and D'Elia A. (2008) A new approach for modelling consumers' preferences, *Food Quality and Preference*, 19, 247–259.

Appendix A: 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 $g = 1, 2, 3$:

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

Then, 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 $\mathcal{P}_g$ 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:

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

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:

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

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:

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

These expressions admit close solutions only for simplified G*e*CUB models since the maximization may be greatly simplified when some parameters are absent. Thus, explicit solutions exist for both parameters (π,ξ) in CUB models without covariates (even with a *shelter* effect), or for the parameter π (or ξ, respectively) in a CUB model with covariates only for the feeling (the uncertainty parameter, respectively). Generally, numerical methods are required for a solution.

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

$$\begin{aligned}
\big(\boldsymbol{\beta}'^{(k+1)},\;\boldsymbol{\omega}'^{(k+1)}\big)' &= \underset{\boldsymbol{\beta},\,\boldsymbol{\omega}}{argmax}\;Q_1(\boldsymbol{\beta}^{(k)},\boldsymbol{\omega}^{(k)}); \\
\boldsymbol{\gamma}^{(k+1)} &= \underset{\boldsymbol{\gamma}}{argmax}\;Q_2(\boldsymbol{\gamma}^{(k)}).
\end{aligned}$$

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

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

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

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

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

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

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

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, 47-49) and Karlis and Xekalaki (2003) in general contexts, and confirmed by Iannario (2012b) for CUB models. This issue deserves more studies and extensive experiments; however, in case of large sample size we suggest to start with initial values derived by simplified CUB models (without and with covariates and/or without covariates in the *shelter effect*). In absence of any information we might use $\boldsymbol{\theta}^{(0)} = (0.1, 0.1, \dots, 0.1)'$. However, it is better to start with a random sampling of a subset of the full data set (of $n \leq 200$ subjects, say) and to plug the obtained parameter estimates in the EM procedure as the preliminary ones.

Appendix B: Derivation of the information matrix

The log-likelihood function is expressed by:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\Big[Pr\left(R = r_i \mid \mathcal{C}_i,\, \boldsymbol{\theta}\right)\Big].$$

Thus, all derivatives will be deduced for the likelihood element $\ell_i(\boldsymbol{\theta})$, $i = 1, 2, \ldots, n$.

The computational burden will be reduced in a sensible manner if we adopt a series of sequential settings to allow for a convenient format of the awkward expressions of the formulae.

First, we put:

$$p_i = Pr\left(R = r_i | \mathcal{C}_i;\, \boldsymbol{\theta}\right); \quad \nu_i = 1/p_i; \quad M_i = m - r_i - (m-1)\xi_i(\boldsymbol{\gamma})\,, \qquad i = 1, 2, \ldots, n\,.$$

Then, for any $i = 1, 2, \ldots, n$, we let:

$$\begin{aligned}
A_i(\beta_k) &= \frac{\partial\, p_i}{\partial\, \beta_k}\,\frac{1}{p_i} = y_{ik}\, \pi_i\, (1-\pi_i)\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i\,, \qquad k = 0, 1 \ldots, p;\\
B_i(\gamma_k) &= \frac{\partial\, p_i}{\partial\, \gamma_k}\,\frac{1}{p_i} = w_{ik}\, \pi_i\,(1-\delta_i)\, M_i\, b_{r_i}(\xi_i)\,\nu_i\,, \qquad k = 0, 1 \ldots, q;\\
C_i(\omega_k) &= \frac{\partial\, p_i}{\partial\, \omega_k}\,\frac{1}{p_i} = x_{ik}\,\delta_i\,\big(v_i\, D^{(c)}_{r_i} - 1\big)\,, \qquad k = 0, 1 \ldots, s\,.
\end{aligned}$$

Similarly, for the second derivatives of probabilities, for any $i = 1, 2, \ldots, n$, we let (with indices h, k varying as admissible):

$$\begin{aligned}
D_i(\beta_h, \beta_k) &= \frac{\partial^2\, p_i}{\partial\beta_h\,\partial\beta_k}\,\frac{1}{p_i} = y_{hi}\, y_{ki}\, \pi_i\,(1-\pi_i)\,(1-2\pi_i)\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i\,;\\
G_i(\gamma_h, \gamma_k) &= \frac{\partial^2\, p_i}{\partial\gamma_h\,\partial\gamma_k}\,\frac{1}{p_i} = w_{hi}\, w_{ki}\, \pi_i\,(1-\delta_i)\, b_{r_i}(\xi_i)\,\big[M_i^2 - (m-1)\,\xi_i\,(1-\xi_i)\big]\,\nu_i\,;\\
L_i(\omega_h, \omega_k) &= \frac{\partial^2\, p_i}{\partial\omega_h\,\partial\omega_k}\,\frac{1}{p_i} = x_{hi}\, x_{ki}\,\delta_i\,(1-2\delta_i)\,\big(\nu_i\, D^{(c)}_{r_i} - 1\big)\,;\\
E_i(\beta_h, \gamma_k) &= \frac{\partial^2\, p_i}{\partial\beta_h\,\partial\gamma_k}\,\frac{1}{p_i} = y_{hi}\, w_{ki}\, \pi_i\,(1-\pi_i)\,(1-\delta_i)\, b_{r_i}(\xi_i)\, M_i\,\nu_i\,;\\
F_i(\beta_h, \omega_k) &= \frac{\partial^2\, p_i}{\partial\beta_h\,\partial\omega_k}\,\frac{1}{p_i} = -\,y_{hi}\, x_{ki}\, \pi_i\,(1-\pi_i)\,\delta_i\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i\,;\\
H_i(\gamma_h, \omega_k) &= \frac{\partial^2\, p_i}{\partial\gamma_h\,\partial\omega_k}\,\frac{1}{p_i} = -\,x_{hi}\, w_{ki}\, \pi_i\,\delta_i\,(1-\delta_i)\, b_{r_i}(\xi_i)\, M_i\,\nu_i\,;
\end{aligned}$$

The previous quantities are necessary for computing the negative of the second derivatives of the log-likelihood function. In fact, for any pair of parameters θ_h and θ_k, we get:

$$-\frac{\partial^2\, \log\, p_i}{\partial\theta_h\,\partial\theta_k} = \left(\frac{\partial\, p_i}{\partial\theta_h}\,\frac{1}{p_i}\right)\left(\frac{\partial\, p_i}{\partial\theta_k}\,\frac{1}{p_i}\right) - \frac{\partial^2\, p_i}{\partial\theta_h\,\partial\theta_k}\,\frac{1}{p_i}\,.$$

Taking account of the previous notation and of the symmetry of the derivatives, the observed information matrix may be obtained as:

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix}
\underset{p+1,p+1}{\mathcal{I}_{11}} & \underset{p+1,q+1}{\mathcal{I}'_{21}} & \underset{p+1,s+1}{\mathcal{I}'_{31}}\\
\underset{q+1,p+1}{\mathcal{I}_{21}} & \underset{q+1,p+1}{\mathcal{I}_{22}} & \underset{q+1,s+1}{\mathcal{I}'_{32}}\\
\underset{s+1,p+1}{\mathcal{I}_{31}} & \underset{s+1,q+1}{\mathcal{I}_{32}} & \underset{s+1,s+1}{\mathcal{I}_{33}}
\end{pmatrix}$$

where

$$\mathcal{I}_{11} = \sum_{i=1}^{n}\Big[A_i(\beta_h)A_i(\beta_k) - D_i(\beta_h,\beta_k)\Big]; \qquad \mathcal{I}_{22} = \sum_{i=1}^{n}\Big[B_i(\gamma_h)B_i(\gamma_k) - G_i(\gamma_h,\gamma_k)\Big];$$
$$\mathcal{I}_{33} = \sum_{i=1}^{n}\Big[C_i(\omega_h)C_i(\omega_k) - L_i(\omega_h,\omega_k)\Big]; \qquad \mathcal{I}_{21} = \sum_{i=1}^{n}\Big[A_i(\beta_h)B_i(\gamma_k) - E_i(\beta_h,\omega_k)\Big];$$
$$\mathcal{I}_{31} = \sum_{i=1}^{n}\Big[A_i(\beta_h)C_i(\omega_k) - F_i(\beta_h,\omega_k)\Big]; \qquad \mathcal{I}_{32} = \sum_{i=1}^{n}\Big[B_i(\gamma_h)C_i(\omega_k) - H_i(\omega_h,\omega_k)\Big].$$

Finally, the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of the ML estimators of $\boldsymbol{\theta}$, computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} = (\hat{\pi},\, \hat{\xi})'$, is derived as:

$$\boldsymbol{V}(\boldsymbol{\theta}) = \Big[\mathcal{I}(\hat{\boldsymbol{\theta}})\Big]^{-1}.$$

The asymptotic standard errors of the parameters are computed by taking the square root of each diagonal elements of the diagonal submatrices of $\boldsymbol{V}(\boldsymbol{\theta})$. In addition, nested tests may be obtained by using likelihood ratio critical regions.

If we need to effectively implement the computation of the information matrix $\mathcal{I}(\hat{\boldsymbol{\theta}})$ in a matrix-oriented language (as **R**, *Gauss*©, *Matlab*©, etc.), we suggest to compute the following vectors and matrices.

Let $\boldsymbol{r}$ the vector of observed ordinal ratings, $\nu_i = 1/p_i$ for $i = 1, 2, \dots, n$ and $\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{X}$ the matrices already specified in section 2. Then, define the following vectors:

$$\begin{aligned}
\boldsymbol{a}^* &= ||\pi_i\,(1-\pi_i)\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{b}^* &= ||\pi_i\,(1-\delta_i)\,M_i\,b_{r_i}(\xi_i)\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{c}^* &= ||\delta_i\,\big(D^{(c)}_{r_i} - p_i\big)\,\nu_i||_{i=1,\dots,n}\,.
\end{aligned}$$

The quantities $A_i(\beta_k),\ B_i(\gamma_k)\,C_i(\omega_k)$ may be computed as:

$$\underset{n,p+1}{\boldsymbol{A}} = \boldsymbol{Y} \odot \boldsymbol{a}^*; \qquad \underset{n,q+1}{\boldsymbol{B}} = \boldsymbol{W} \odot \boldsymbol{b}^*; \qquad \underset{n,s+1}{\boldsymbol{C}} = \boldsymbol{X} \odot \boldsymbol{c}^*;$$

where $\boldsymbol{M} \odot \boldsymbol{v}$ denotes the elementwise product between the columns of the matrix $\boldsymbol{M}$ and the vector $\boldsymbol{v}$.

Similarly, if we define the vectors:

$$\begin{aligned}
\boldsymbol{d}^* &= ||\pi_i\,(1-\pi_i)\,(1-2\pi_i)\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{g}^* &= ||\pi_i\,(1-\delta_i)\,b_{r_i}(\xi_i)\,\big[M_i^2 - (m-1)\,\xi_i\,(1-\xi_i)\big]\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{l}^* &= ||\delta_i\,(1-2\delta_i)\,\big(D^{(c)}_{r_i} - p_i\big)\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{e}^* &= ||\pi_i\,(1-\pi_i)\,(1-\delta_i)\,b_{r_i}(\xi_i)\,M_i\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{f}^* &= -\,||\pi_i\,(1-\pi_i)\,\delta_i\,(1-\delta_i)\,\big(b_{r_i}(\xi_i) - U_{r_i}\big)\,\nu_i||_{i=1,\dots,n}\,;\\
\boldsymbol{h}^* &= -\,||\pi_i\,\delta_i\,(1-\delta_i)\,b_{r_i}(\xi_i)\,M_i\,\nu_i||_{i=1,\dots,n}\,;
\end{aligned}$$

the quantities in the observed information matrix $\mathcal{I}(\hat{\boldsymbol{\theta}})$ may be computed as follows:

$$\begin{aligned}
\mathcal{I}_{11} &= \boldsymbol{A}'\,\boldsymbol{A} - \boldsymbol{Y}'\,(\boldsymbol{Y} \odot \boldsymbol{d}^*); &\qquad \mathcal{I}_{22} &= \boldsymbol{B}'\,\boldsymbol{B} - \boldsymbol{W}'\,(\boldsymbol{W} \odot \boldsymbol{g}^*)\\
\mathcal{I}_{33} &= \boldsymbol{C}'\,\boldsymbol{C} - \boldsymbol{X}'\,(\boldsymbol{X} \odot \boldsymbol{l}^*); &\qquad \mathcal{I}_{21} &= \mathcal{I}_{12}{}' = \boldsymbol{B}'\,\boldsymbol{A} - \boldsymbol{W}'\,(\boldsymbol{Y} \odot \boldsymbol{e}^*);\\
\mathcal{I}_{31} = \mathcal{I}_{13}{}' &= \boldsymbol{C}'\,\boldsymbol{A} - \boldsymbol{X}'\,(\boldsymbol{Y} \odot \boldsymbol{f}^*); &\qquad \mathcal{I}_{32} &= \mathcal{I}_{23}{}' = \boldsymbol{C}'\,\boldsymbol{B} - \boldsymbol{X}'\,(\boldsymbol{W} \odot \boldsymbol{f}^*)\,.
\end{aligned}$$