Quaderni di Statistica
Vol.8, 2006

Observed information matrix for MUB models

Domenico Piccolo
Dipartimento di Scienze Statistiche
Università degli Studi di Napoli Federico II
E-mail: domenico.piccolo@unina.it

Summary: In this paper we derive the observed information matrix for MUB models, without and with covariates. After a review of this class of models for ordinal data and of the E-M algorithms, we derive some closed forms for the asymptotic variance-covariance matrix of the maximum likelihood estimators of MUB models. Also, some new results about feeling and uncertainty parameters are presented. The work lingers over the computational aspects of the procedure with explicit reference to a matrix-oriented language. Finally, the finite sample performance of the asymptotic results is investigated by means of a simulation experiment. General considerations aimed at extending the application of MUB models conclude the paper.

Keywords: Ordinal data, MUB models, Observed information matrix.

1. Introduction

"Choosing to do or not to do something is a ubiquitous state of activity in all societies" (Louvier *et al.*, 2000, 1). Indeed, "Almost without exception, every human beings undertake involves a choice (consciously or sub-consciously), including the choice not to choose" (Hensher *et al.*, 2005, xxiii).

From an operational point of view, the finiteness of alternatives limits the analysis to discrete choices (Train, 2003) and the statistical interest in this area is mainly devoted to generate probability structures adequate to interpret, fit and forecast human choices[1].

[1] An extensive literature focuses on several aspects related to economic and market-

Generally, the nature of the choices is qualitative (or categorical), and classical statistical models introduced for continuous phenomena are neither suitable nor effective. Thus, qualitative and ordinal data require specific methods to avoid difficulties in the interpretation and/or loss of efficiency in the analysis of real data.

In this area, we investigated a probability model that seems capable to produce interpretable results and good fitting. The feasibility of efficient maximum likelihood methods and the implementation of accurate numerical algorithms are essential requirements in order to apply these new models.

The paper is organized as follows: in the next section, we establish notations for ordinal data and in sections 3-4 we introduce the logic and the main properties of MUB models (without and with covariates, respectively). Then, section 5 presents the E-M algorithms steps to obtain the maximum likelihood (ML) estimates of the parameters. In sections 6-8 we derive the information matrices while section 9 discusses the related numerical algorithms. A special emphasis has been devoted to the minimization of the computation efforts to implement the formal results. Section 10 investigates the possibility to make direct inference on the MUB parameters when the model requires some covariates for a better fitting and in section 11 we check the finite sample performance of the asymptotic results presented in this work by means of a simulation study. Some concluding remarks end the paper.

2. Ordinal data

Human choices are placed within a list of alternatives, that we may call “items” or “objects”, as: brands, candidates, services, topics, sentences, situations, teams, songs, recreation sites, colors, professions, etc. From a formal point of view, the very nature of the items is not relevant, although we need to specify them to relate human behavior to some se-

---

ing management: Franses and Paap (2001). In this regard, a relevant issue is the problem of discriminating and/or combining *stated* and *revealed preferences*, as discussed by Louvier *et al.* (2000) and Train (2003, 156-160).

lected covariates.

In our context, ordinal data (=*ranks*) are integers that express preferences or evaluations of a group of raters towards a well defined list of items/objects/ services[2]. In this regard, we emphasize situations where ordinal data are used both for preference and evaluation studies; these contexts are substantially different but share so many structural similarities to suggest common analyses.

More specifically, in *preference analysis* the ranks express the location of an "object" in a given list. Instead, in *evaluation analysis* the ranks express the subject's perceived level of a given "stimulus" (sensation, opinion, perception, awareness, appreciation, feeling, taste, etc.).

In essence, the main similarities among preferences and evaluations experiments are the following:

- Both analyses are expressed by ordered integer values.
- The answer is a categorical variable transformed into an integer.
- Psychological, sociological, environmental effects influence (and bias) the answers.
- The extreme answers are more reliable and persisting than the middle ones[3].
- The answer is the result of a paired or sequential selection process.

Instead, the main differences among preference and evaluation experiments are the following:

[2] Ordinal data may be also generated by several different situations. Subjects may be assigned to categories, as it often happens in Medicine, Psychology and Environmental Sciences, for instance. Moreover, a continuous variable may be classified in classes and then a sequential code is given to each class, as it is common in Marketing and Finance researches.

[3] Indeed, especially for those items that do not generate strong liking or disliking feelings, it seems plausible to assume that the elicitation mechanism exhibits a greater uncertainty. This fact can be easily shown if we consider, for instance, a repeated ranking of several items by the same group of raters. While the extreme ranks are expected to remain unchanged, it is very plausible that the intermediate ranks will change somewhere.

- In preference analyses, the output is a set of related ordered responses: thus, we study only the marginal distribution of the preference towards a single object.

- The evaluations of several items are (generally) strongly correlated, since they express a common (positive/negative) judgement towards the problem at hand.

In order to avoid frequent repetitions, in the rest of this article, we will limit the discussion to preference analysis although the approach and the applications are by far more general than the preference context.

The problem is set as follows: we ask a group of n subjects (=raters, judges) to order a set of $m > 1$ "objects": $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_j \dots, \mathcal{O}_m$ giving order (=rank) 1 to the preferred one, order 2 to the second best, ..., order m to the least preferred object. No ties are allowed.

The experiment produces a $n \times m$ matrix of the assigned ranks to the m objects (the columns) by the n raters (the rows):

$$\begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,j} & \dots & r_{1,m} \\ r_{2,1} & r_{2,2} & \dots & r_{2,j} & \dots & r_{2,m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n,1} & r_{n,2} & \dots & r_{n,j} & \dots & r_{n,m} \end{bmatrix}$$

In this paper, we consider only the collection of ranks assigned by the n raters to a prefixed $\mathcal{O}_j = \mathcal{O}$ object. Then, we may denote simply by $\boldsymbol{r} = (r_1, r_2, \dots, r_n)'$ the expressed ranking towards a prefixed object $\mathcal{O}$. Indeed, we study *only* the *marginal distribution* of a multivariate random variable generated by the m-th observations of the previous matrix[4].

We interpret $\boldsymbol{r}$ as an *observed sample* of size n, that is the realization of a *random sample* $(R_1, R_2, \dots, R_n)$. This random sample is the collection of n independent and identically distributed discrete random variables $R_i \equiv R \sim f(r; \boldsymbol{\theta})$, where the probability (mass) function is characterized by a parameter vector $\boldsymbol{\theta}$.

[4] We emphasize this point since it restricts and characterizes our approach with respect to many different models for ordinal data presented in the statistical literature.

3. A probability model for ordinal data

Many probability models and statistical tools have been proposed for describing the ranking process and/or analyzing ranks data[5]. The issue of ordinal data modeling is generally included in the domain of qualitative and multinomial models[6].

In this area, the standard approach is motivated by a generalized linear model (GLM) developed by Nelder and Wedderburn (1972), McCullagh (1980) and McCullagh and Nelder (1989). Generally, the *distribution* or the *survival functions* are introduced as tools for expressing the probability mechanism that relates the ordered responses to some covariates, via a latent variable which transforms continuous variables into discrete responses.

Instead, our approach is motivated by a direct investigation of the psychological process that generates the choice mechanism among m alternatives. This approach has led us to a series of results[7] which have been conveyed into a final proposal defining a class of probability distributions

---

[5] Among the several references, we list: Amemiya (1981); Maddala (1983); Agresti (1984; 1996; 2002); Fligner and Verducci (1993); Marden (1995); Fox J. (1997, 475-478); Lloyd (1999, 338-352; 355-367); Greene (2000, 875-880); Mittelhammer et al. (2000, 584-585); Power and Xie (2000), 201-222; Cramer (2001); Franses and Paap (2001, 112-132); Dobson (2002, 143-148); Woolridge (2002, 505-509); Davison (2003, 507-510). For discrete-choice models, we refer to: Agresti (2002, 298-300); Train (2003, 163-168).

[6] In most cases, the topics is discussed as a specification of qualitative unordered models, also with some extreme position: "The application of ordered multinomial response models seems to be the exception rather than the rule.": Mittelhammer et al. (2000, 585).

[7] The starting point for this kind of models is a paper by D'Elia (1999; 2003a) where an Inverse Hypergeometric (IHG) random variable has been introduced to explain the sequential choice of several objects. Then, the need for a probability distribution with an intermediate mode suggested the study of a shifted Binomial model: D'Elia (2000a; 2000b). Finally, in order to improve both the interpretation and the flexibility, MUB models were proposed in 2003 and first published by D'Elia and Piccolo (2005a). Since all the proposed models included subjects' covariates, then the *logic* of the GLM models was exploited. Of course, most of these results may be generalized without substantial modifications if we include in the models both *choices' covariates* and *chooser's covariates*: Agresti (2002, 292).

called MUB models.

In essence, while most of the proposed models for ordinal data are based on log-transformations of probabilities (distribution functions, adjacent categories and continuation-ratio logits, etc.), we will model explicitly the probability of an ordinal choice, and then we will relate its parameters to the subjects' covariates.

In our opinion, a probability model for the random variable R should be *adequate* for representing the psychological choice mechanism, *consistent* with the ordinal feature of the data, *parsimonious* in that it contains as few parameters as possible, and *realistic* and *flexible* to be fitted to different empirical phenomena.

More specifically, one should consider that the rater choice may be a *thoughtful* or *instinctive* one, and this choice may result from a *paired* or *sequential* comparison of the objects[8].

Anyway, the final choice is the result of *two hierarchical steps*:

– a *global* and immediate evaluation of the feeling (the subject agrees, certainly disagrees or is indifferent);

– a *local* and reflective setting for expressing the final rank, within the previous global assessment.

Several models satisfy these requirements; therefore, the choice mechanism is the result of a *feeling* towards the object and the *uncertainty* in the choice of the rank. Of course, these components interact in the choice mechanism with different weights.

In this respect, we notice that both feeling and uncertainty are *continuous* and *latent* random variables, whereas ordinal data are expressed as discrete responses taking a value in $\{1, 2, \dots, m\}$. Thus, to attain efficient transformations of continuous variables into a discrete set, we need some preliminary investigation about the nature of these components.

• *Feeling* is the result of a continuous random variable that becomes a discrete one, since the subject is compelled to express the preferences into m prefixed bins. Now, the judgement process is intrinsically continuous

[8] It seems evident that the sequential choice is preferred for several items whereas a paired choice is more useful for few items.

and, since this perception depends on several causes, it can be assumed to follow a Gaussian distribution.

We recall that a *latent variable approach* for the analysis of ordinal data[9] assumes that the records are generated by an unobserved continuous random variable (say R^*), generally Normally distributed; then, a correspondence with a discrete ordinal random variable R is given by means of ordered threshold parameters[10]:

$$\begin{array}{llll} -\infty & < R^* \leq & \alpha_1 & \\ \alpha_1 & < R^* \leq & \alpha_2 & \\ \ldots\ldots & & & \Longleftrightarrow \\ \alpha_{m-2} & < R^* \leq & \alpha_{m-1} & \\ \alpha_{m-1} & < R^* < & +\infty & \end{array} \qquad \begin{array}{l} R = 1 \\ R = 2 \\ \ldots\ldots\ldots \\ R = m-1 \\ R = m \end{array}$$

Following this idea, a suitable model for achieving the mapping of the unobserved continuous variable R^* into a discrete random variable defined on the support $r = 1, 2, \ldots, m,$ may be introduced by the ***shifted Binomial*** distribution[11]. Indeed, Figure 1 shows how, by *varying the ordered thresholds*, a standard Normal random variable can be made discrete, obtaining different features (mode, skewness, tails, etc.) that are well fitted by a shifted Binomial random variable.

• *Uncertainty* is a vaguer component that needs some clarification. Indeed, uncertainty *is not* the stochastic component related to the sampling experiment (such that different people generates different rankings). But, instead, uncertainty *is* the result of several factors, intrinsically related to the choice mechanism, such as the *knowledge* or *ignorance* of the problems and/or the characteristic of the objects, the personal *interest* or

[9] General references to latent variable models are Everitt (1984), Bartholomew (1987) and Sammel *et al.* (1997). For ordinal data, the latent variable approach is widely discussed by Moustaki (2000; 2003), Moustaki and Knott (2000), Jöreskog and Moustaki (2001), Moustaki and Papageorgiou (2004), Cagnone *et al.* (2004), Huber *et al.* (2004).

[10] The threshold parameters $(\alpha_1, \alpha_2, \ldots, \alpha_{m-1})$ are also called *cutpoints*, and it is convenient to assume: $\alpha_0 = -\infty$; $\alpha_m = +\infty$. They are unobserved but may be estimated by ML methods.

[11] We prefer to work with the *shifted* random variable since usually the choice set $\{1, 2, \ldots, m\}$ is more common than the Binomial support, that starts with 0.

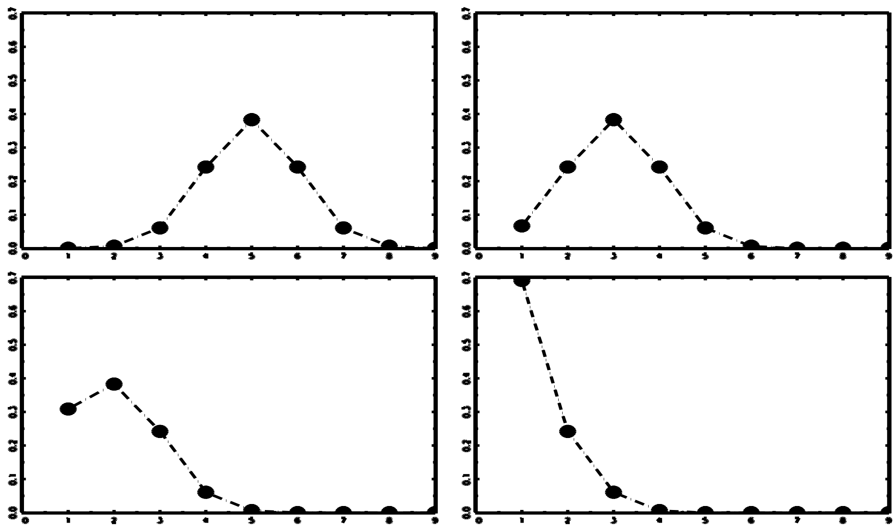

Figure 1: *Shifted Binomial distributions generated by* $Z \sim N(0,1)$.

engagement in similar activities, objects, opinions, etc., the *time spent* for elaborating the decision, the *laziness* or *apathy* of the subject, and so on.

Under the circumstance that the subject shows a complete indifference (=*equipreference*) towards a given item, then it seems appropriate to model ranks by means of a discrete Uniform random variable U with probability mass: function defined by:

$$Pr\,(U=r)=\frac{1}{m},\quad r=1,2,\ldots,m\,,$$

and the choice is the result of a complete randomized mechanism where the item has a constant probability to be given any rank $r \in [1,m]$.

For this reason, we choose the ***discrete Uniform*** distribution as a building block for modeling the uncertainty in the ordinal modeling[12].

Finally, we propose to take into account the *composite nature of the*

[12] We remember that the discrete Uniform random variable maximizes the entropy, among all the discrete distributions with finite support $\{1,2,\ldots,m\}$, for a fixed m.

elicitation process by means of a *mixture model*, where the feeling and uncertainty components are adequately weighted[13].

Thus, we assume that the rank r is the realization of a random variable R that is a ***M***ixture of a ***U***niform and a shifted ***B***inomial random variable.

Formally, for a fixed and known integer $m > 1$, we define the (discrete) MUB random variable R with parameters π and ξ, on the finite support $\{r : r = 1, 2, \ldots, m\}$, and denote it by $R \sim MUB(\pi, \xi)$, if and only if its probability distribution is:

$$P_r\,(R = r) = \pi \underbrace{\left[\binom{m-1}{r-1}(1-\xi)^{r-1}\xi^{m-r}\right]}_{\textit{feeling}} + (1-\pi) \underbrace{\left[\frac{1}{m}\right]}_{\textit{uncertainty}} .$$

Since $\pi \in [0, 1]$ and $\xi \in [0, 1]$, the parametric space of R is the unit square $[0, 1] \times [0, 1]$.

The MUB model turns out to be a flexible stochastic structure since the distribution varies largely as the parameters π and θ vary (D'Elia and Piccolo, 2005a). In this regard, we list some features of the MUB distribution which will be useful for our discussion:

- It may admit a mode at any value of the support $\{1, 2, \ldots, m\}$.
- It is a symmetric random variable if and only if $\xi = \frac{1}{2}$.
- It is a ***reversible*** random variable, since:

$$R \sim MUB(\pi, \xi) \Longrightarrow (m + 1 - R) \sim MUB(\pi, 1 - \xi)\,.$$

- It is consistent with the hypothesis that the population is made by two sub-groups of raters (an informed/reflexive set and a more uninformed/instinctive one) and their relative ratio is $\pi/(1 - \pi)$.
- It emulates many theoretical distributions:
  - A *Uniform* distribution, if $\pi = 0$;

[13] Recent applications of mixture models to ranked data are discussed by Gormley and Murphy (2006) with reference to college applications data; see also Marden (1995).

- A *Shifted Binomial* distribution, if $\pi = 1$;
- An *Inverse HyperGeometric* distribution, if $\pi \to 1$ *and* ξ tends to 0 or 1;
- A *Normal* distribution, if $\xi \to \frac{1}{2}$ *and* $m \to \infty$;

For our purposes, the moments of this distribution are not relevant since the sequence $\{1, 2, \ldots, m\}$ is only a *proxy* for a qualitative ordering, and no metric property should be attached to these integer values. However, in some contexts and for comparison purposes, it may be useful to know that the first two cumulants[14] of R are:

$$\mathbb{E}(R) = \pi\ (m-1)\left(\frac{1}{2} - \xi\right) + \frac{(m+1)}{2}\,;$$

$$Var(R) = (m-1)\left\{\pi\,\xi\,(1-\xi) + (1-\pi)\left[\tfrac{m+1}{12} + \pi\,(m-1)\left(\tfrac{1}{2} - \xi\right)^2\right]\right\}.$$

As far as the interpretation of the parameters of the MUB model is concerned, we remember that the π parameter measures the uncertainty through the quantity: $(1-\pi)/m$, which is the *measure of the uncertainty* distributed over all the support. Also, the ratio $\pi/(1-\pi)$ measures the relative weight of the basic feeling and the uncertainty components, respectively.

On the other hand, both π and ξ are related to the liking towards the object. The exact meaning of ξ changes with the setting of the analysis and, being the MUB model reversible, it mostly depends on how the ratings have been codified (the greater the rating the greater the feeling, or viceversa). According to the context that we defined for the responses, the ξ parameter may be a *degree of perception*, a *measure of closeness*, a *rating of concern*, an *index of selectiveness*, and so on.

---

[14] Even though R has been introduced with reference to (qualitative) ordinal data, a complete study of the first four cumulants of the MUB random variable is pursued by Piccolo (2003b). Notice that some caution is needed when we compare ordinal data by $\mathbb{E}(R)$ since, for a given expectation, many values of the parameters (π, ξ) are admissible. In fact, when both π and ξ increase towards 1, the mean value $\mathbb{E}(R)$ converges to 1, and then the MUB model implies a greater preference for the given object. Thus, it is not strictly correct to relate the expected preference *only* to the parameter ξ. For a graphical evidence, see the next Figure 2.

Indeed, the presence of both parameters in the MUB model produces an extremely flexible tool for fitting real data sets and for explaining different judgement choices.

Of course, the introduction of the *raters' covariates* for relating both the feeling and the uncertainty to the subject's characteristics improves both the results and the interpretation. In this way, the covariates allow to link the main features of the raters to the rank they assigned to a given item; both the feeling and the uncertainty can be explained by means of subjects' specific covariates, yielding a deeper insight in the preference data analysis. This allows a sound interpretation of the liking/disliking behavior to be used for inferential and predictive purposes and/or for characterizing meaningful subsets of the population.

In this regard, the standard GLM approach (where the inverse link function explains the expectation of the relevant random variable via a linear function of selected covariates[15]) cannot be pursued here. In fact, for a given m, several couples of different (π, ξ) generates the same expectation as long as they obey:

$$\pi\left(\frac{1}{2}-\xi\right)=\frac{\mathbb{E}\left(R\right)-\frac{m+1}{2}}{m-1}\,.$$

To be explicit, Figure 2 shows some different MUB models characterized by the same expectation; in these comparisons, we selected $m = 9$ and set $\mathbb{E}\left(R\right) = 6$.

For these models, as detailed in Table 1, the mode (that is the most probable value for the ordinal variable R) and the probability of values around the expectation, that is $Pr\left(5 \leq R \leq 7\right)$, are completely different; thus, one takes different decisions when faced with each of these four models, although they have the same expectation[16]

[15] Nelder and Wedderburn (1972); Agresti (2002, 116-117); Dobson (2002, 43-54). In fact, our approach is more related to the class of general models discussed by King *et al.* (2000, 348).

[16] In this regard, we observe that the mode of the MUB distribution, that is: $Mo = 1 + [m(1-\xi)]$, coincides with the mode of the shifted Binomial component of the mixture. Thus, the mode is more immediately related to the feeling measure.

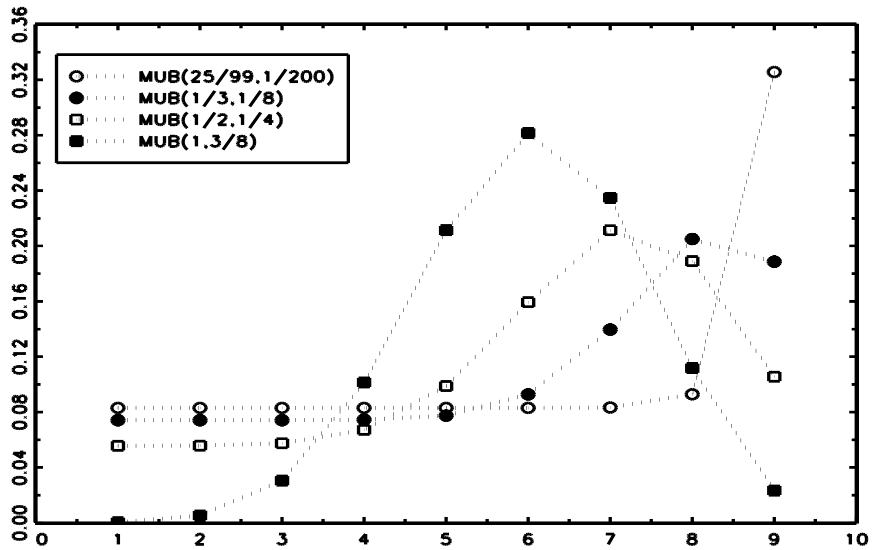

Figure 2: *Different* MUB *models with constant expectation:* $\mathbb{E}(R) = 6$*.*

Table 1. Location measures for different MUB *models*

Models	$\mathbb{E}(R)$	*Mode*	$Pr(5 \leq R \leq 7)$
$MUB(\frac{25}{99}, \frac{1}{200})$	6	9	0.249
$MUB(\frac{1}{3}, \frac{1}{8})$	6	8	0.310
$MUB(\frac{1}{2}, \frac{1}{4})$	6	7	0.469
$MUB(1, \frac{3}{8})$	6	6	0.728

It should be evident that MUB models implied by a given expectation may be substantially different; thus, it seems preferable to look for a link among model parameters and subjects' covariates, without a direct reference to the expectation of this random variable.

4. MUB *models with covariates*

It is reasonable to assume that the main components of the choice mechanism (that is, feeling and uncertainty) vary with the subjects' characteristics as, for instance, gender, age, income, profession, level of education, etc. Thus, it seems worth relating these to subjects' covariates by means of an explicit modelling procedure[17].

From a formal point of view, in MUB models with covariates we will assume that the *uncertainty* parameter π is a function of p subjects' covariates $Y_1, Y_2, \dots, Y_p$ and/or, similarly, the *feeling* parameter ξ is a function of q subjects' covariates $W_1, W_2, \dots, W_q$. The Y's variables (or a subset of them) may also coincide with the W's variables (or a subset of them).

Then, in order to classify MUB models according to the presence/ absence of covariates, we introduce the following terminology:

Models	*Covariates*	*Parameter vectors*	*Parameter space*
• MUB-00	no covariates	$\boldsymbol{\theta} = (\pi, \xi)'$	$[0,1] \times [0,1]$
• MUB-10	covariates for π	$\boldsymbol{\theta} = (\boldsymbol{\beta}', \xi)'$	$\mathbb{R}^{p+1} \times [0,1]$
• MUB-01	covariates for ξ	$\boldsymbol{\theta} = (\pi, \boldsymbol{\gamma}')'$	$[0,1] \times \mathbb{R}^{q+1}$
• MUB-11	covariates for π and ξ	$\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$	$\mathbb{R}^{p+q+2}$

Suppose we have a sample of ordinal data $\boldsymbol{r} = (r_1, r_2, \dots, r_n)'$, and assume that for each of the n units we have several measurements on the

[17] The problem of choosing the significant covariates for a MUB model is still open since traditional correlation methods are not effective. In this regard, a study on association indexes and predictability measures for ordinal data might be fruitfully pursued. Finally, some encouraging results has been obtained by tree-based methods as discussed by Cappelli and D'Elia (2006a; 2006b).

subjects summarized in the following matrices:

$$\boldsymbol{Y} = \begin{pmatrix} 1 & y_{11} & y_{12} & \dots & y_{1p} \\ 1 & y_{21} & y_{22} & \dots & y_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & y_{i1} & y_{i2} & \dots & y_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix}; \ \boldsymbol{W} = \begin{pmatrix} 1 & w_{11} & w_{12} & \dots & w_{1q} \\ 1 & w_{21} & w_{22} & \dots & w_{2q} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w_{i1} & w_{i2} & \dots & w_{iq} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w_{n1} & w_{n2} & \dots & w_{nq} \end{pmatrix}.$$

To make the notation more compact, we introduce the variables Y_0 and W_0 that assume the constant value 1 for all the sample units; they specify the constants (baselines) of the model.

We denote by $\boldsymbol{y}_i$ and $\boldsymbol{w}_i, i = 1, 2, \dots, n$ the i-th row of the $\boldsymbol{Y}$ and $\boldsymbol{W}$ matrices, respectively, that is:

$$\boldsymbol{y}_i = (y_{i0}, y_{i1}, y_{i2}, \dots, y_{ip}); \qquad \boldsymbol{w}_i = (w_{i0}, w_{i1}, w_{i2}, \dots, w_{iq})\,;$$

and let:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'; \qquad \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)'\,.$$

In order to specify a correspondence among the real valued $\boldsymbol{Y}$ and $\boldsymbol{W}$ and the $\pi \in (0, 1)$ and $\xi \in (0, 1)$ parameters, D'Elia (2003a) and Piccolo (2003a) proposed a logistic mapping defined by:

$$(\pi \mid \boldsymbol{y}_i) = \frac{1}{1 + e^{-\boldsymbol{y}_i \boldsymbol{\beta}}} = \left[1 + e^{-\sum_{s=0}^{p} \beta_s\, y_{is}}\right]^{-1};$$
$$(\xi \mid \boldsymbol{w}_i) = \frac{1}{1 + e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}} = \left[1 + e^{-\sum_{t=0}^{q} \gamma_t\, w_{it}}\right]^{-1}.$$

Notice that similar mappings[18] might be investigated as, for instance, the *probit* function: $\Phi(z) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}\,z^2}$ and the *complementary log-log* function: $F(z) = \exp\{-\exp(-z)\}$. However, in the following, we will limit the study to the logistic function.

[18] In this context the common considerations on the transformations in categorical modeling apply; thus, logit and probit appear quite similar while complimentary log-log should be an alternative when the impact of covariates is expected to cause some asymmetric behavior in the response variable.

Finally, the general MUB model with covariates is explicitly defined, for any $i = 1, 2, \ldots, n$, by the probability distribution:

$$P_r\left(R = r_i \mid \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \frac{1}{1 + e^{-\boldsymbol{y}_i \boldsymbol{\beta}}} \left[\binom{m-1}{r_i - 1} \frac{\left(e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}\right)^{r_i - 1}}{\left(1 + e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}\right)^{m-1}} - \frac{1}{m} \right] + \frac{1}{m} \, .$$

We again observe that this approach is logically related to GLM since we relate the parameters of a model to subjects' covariates. However, we are not introducing a *link function* between the expectation and the covariates, and our probability distribution does not belong to the exponential family. Indeed, the more relevant difference is that we are relating explicitly the parameters of the response distribution to the covariates while in the GLM approach the relationship is established between a transformation of the distribution function and the covariates.

5. The maximum likelihood estimation via the E-M algorithm

In this section we discuss the computational steps involved in the E-M algorithm[19] for ML estimation of the parameters in the MUB model[20].

The main advantage of the E-M algorithm for MUB models is that the involved log-likelihood splits into two functions where the observed and the unobserved quantities (that is the probability π that the observation comes from one of the two sub-populations) are well defined and neatly separated. Thus, it is possible to activate two alternating steps of Expectation and Maximization converging to the ML estimate.

We denote by $\boldsymbol{\theta}$ the parameter vector for all MUB models; thus, $\boldsymbol{\theta}$ consists of two elements in the case of a MUB model without covariates

[19] A discussion of the E-M algorithm in a general statistical framework is contained in McLachlan and Krishnan (1997) and Jorgenson (2002) and, for mixture models, in McLachlan and Peel (2000).

[20] The derivation of this algorithm for the MUB-00 model was obtained by D'Elia and Piccolo (2005a), while the E-M algorithms for MUB-10 and MUB-01 models are in D'Elia (2003b). Finally, the E-M algorithm for the general MUB-11 model has been fully developed by Piccolo (2003a). Notice that Tables $2-5$, here reported, correct some misprints contained in Piccolo (2003a).

and of two parameters vectors in the case of MUB models with covariates, as shown in the previous section.

First, we define the log-likelihood function $\ell(\boldsymbol{\theta})$ of the MUB model without covariates, that is:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{r=1}^{m} n_r \log\left\{Pr\left(R = r|\boldsymbol{\theta}\right)\right\} \\ &= \sum_{r=1}^{m} n_r \log\left\{\pi\left[b\left(r;\xi\right) - \frac{1}{m}\right] + \frac{1}{m}\right\}, \end{aligned}$$

where n_r are the observed frequencies of $(R = r)$, $r = 1, 2, \ldots, m$ and we denote by

$$b\left(r;\xi\right) = \binom{m-1}{r-1}(1-\xi)^{r-1}\xi^{m-r}, \quad r = 1, 2, \ldots, m,$$

the shifted Binomial distribution.

The log-likelihood function of the MUB-10 model, where the π parameter is explained by some covariates, that is $\pi = f\left(\boldsymbol{Y}, \boldsymbol{\beta}\right)$, is:

$$\ell(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left(1 + e^{-\boldsymbol{y}_i\boldsymbol{\beta}}\right) + \sum_{i=1}^{n} \log\left(b\left(r_i;\xi\right) + \frac{e^{-\boldsymbol{y}_i\boldsymbol{\beta}}}{m}\right).$$

The log-likelihood function of the MUB-01 model, where the ξ parameter is explained by some covariates, that is $\xi = g\left(\boldsymbol{W}, \boldsymbol{\gamma}\right)$, is:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left\{\pi\left[\binom{m-1}{r_i-1}\frac{e^{-\boldsymbol{w}_i\boldsymbol{\gamma}(r_i-1)}}{\left(1 + e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}\right)^{m-1}} - \frac{1}{m}\right] + \frac{1}{m}\right\}.$$

Finally, if both the π and the ξ parameters are explained by covariates, then $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')'$ and $\pi = f\left(\boldsymbol{Y}, \boldsymbol{\beta}\right)$ and $\xi = g\left(\boldsymbol{W}, \boldsymbol{\gamma}\right)$, and the log-likelihood function of the general MUB-11 model is:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left\{\frac{1}{1 + e^{-\boldsymbol{y}_i\boldsymbol{\beta}}}\left[\binom{m-1}{r_i-1}\frac{\left(e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}\right)^{r_i-1}}{\left(1 + e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}\right)^{m-1}} - \frac{1}{m}\right] + \frac{1}{m}\right\}.$$

All the steps required for the E-M procedures are reported in the following Tables 2–5, where we define the sample mean of the responses by $\overline{R}_n = \sum_{i=1}^{n} r_i/n$. In order to stress the correspondence among the four tables, we have left a blank line where a single step does not apply.

First of all, we present the algorithm in absence of covariates (MUB-00: Table 2); then, the procedure is shown when only one set of covariates is present (MUB-10: Table 3 and MUB-01: Table 4); finally, the general MUB model when both the parameters are functions of two (different or coincident) sets of covariates is presented (MUB-11: Table 5).

Table 2. E-M Algorithm for a MUB model without covariates.

$\boldsymbol{MUB}$**-00** $\boldsymbol{Model\ with\ (\pi, \xi)}$	
$\boldsymbol{\theta} = (\pi, \xi)'$; $\epsilon = 10^{-6}$; $dim(\boldsymbol{\theta}) = 2.$	
$\ell(\boldsymbol{\theta}) = \sum_{r=1}^{m} n_r \log\left\{\pi\left[b(r;\xi) - \frac{1}{m}\right] + \frac{1}{m}\right\}.$	
$Steps$	
0	$\boldsymbol{\theta}^{(0)} = \left(\pi^{(0)}, \xi^{(0)}\right)' = \left(\frac{1}{2}, \frac{m-\overline{R}_n}{m-1}\right)'$; $l^{(0)} = \ell\left(\boldsymbol{\theta}^{(0)}\right).$
1	$b\left(r;\xi^{(k)}\right) = \binom{m-1}{r-1}\left(1-\xi^{(k)}\right)^{r-1}\left(\xi^{(k)}\right)^{m-r}, r = 1, 2, \dots, m.$
2	$\tau\left(r;\boldsymbol{\theta}^{(k)}\right) = \left[1 + \frac{1-\pi^{(k)}}{m\,\pi^{(k)}\,b\left(r;\xi^{(k)}\right)}\right]^{-1}, r = 1, 2, \dots, m.$
3	$\overline{R}_n\left(\boldsymbol{\theta}^{(k)}\right) = \frac{\sum_{r=1}^{m} r\,n_r\,\tau\left(r;\boldsymbol{\theta}^{(k)}\right)}{\sum_{r=1}^{m} n_r\,\tau\left(r;\boldsymbol{\theta}^{(k)}\right)}.$
4	$\pi^{(k+1)} = \frac{1}{n}\sum_{r=1}^{m} n_r\,\tau\left(r;\boldsymbol{\theta}^{(k)}\right).$
5	$\xi^{(k+1)} = \frac{m-\overline{R}_n\left(\boldsymbol{\theta}^{(k)}\right)}{m-1}.$
6	
7	$\boldsymbol{\theta}^{(k+1)} = \left(\pi^{(k+1)}, \xi^{(k+1)}\right)'.$
8	$l^{(k+1)} = \ell\left(\boldsymbol{\theta}^{(k+1)}\right).$
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \rightarrow k+1;\ go\,to\,1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k+1)};\ stop. \end{cases}$

Table 3. E-M Algorithm of a MUB model with covariates for π.

$\boldsymbol{MUB}$-10 $\boldsymbol{Model\ \ with\ \ \pi = f\,(\beta; y)}$	
$\boldsymbol{\theta} = (\boldsymbol{\beta}', \xi)'$; $\epsilon = 10^{-6}$; $dim\,(\boldsymbol{\theta}) = p_\pi + 2$.	
$\ell\,(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left(1 + e^{(-\boldsymbol{y}_i\boldsymbol{\beta})}\right) + \sum_{i=1}^{n} \log\left(b\,(r_i; \xi) + \dfrac{e^{(-\boldsymbol{y}_i\boldsymbol{\beta})}}{m}\right)$.	
Steps	
0	$\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\beta}'^{(0)}, \xi^{(0)}\right)' = \left(0.1, \ldots, 0.1,\ \frac{m-\overline{R}_n}{m-1}\right)'$; $l^{(0)} = \ell\left(\boldsymbol{\theta}^{(0)}\right)$.
1	$b\left(r_i; \xi^{(k)}\right) = \binom{m-1}{r_i-1}\left(1-\xi^{(k)}\right)^{r_i-1}\left(\xi^{(k)}\right)^{m-r_i}$, $i = 1, 2, \ldots, n$.
2	$\pi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}}$; $\tau\left(r_i; \boldsymbol{\theta}^{(k)}\right) = \left[1 + \frac{e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}}{m\,b\left(r_i;\xi^{(k)}\right)}\right]^{-1}$, $i = 1, 2, \ldots, n$.
3	$\overline{R}_n\left(\boldsymbol{\theta}^{(k)}\right) = \frac{\sum_{i=1}^{n} r_i\,\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)}{\sum_{i=1}^{n}\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)}$.
4	$Q_1\left(\boldsymbol{\beta}^{(k)}\right) = -\sum_{i=1}^{n}\left\{\log\left(1 + e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}\right) + \left(1 - \tau\left(r_i; \boldsymbol{\theta}^{(k)}\right)\right) e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}\right\}$.
5	$\boldsymbol{\beta}^{(k+1)} = argmax_{\boldsymbol{\beta}}\, Q_1\left(\boldsymbol{\beta}^{(k)}\right)$.
6	$\xi^{(k+1)} = \frac{m-\overline{R}_n\left(\boldsymbol{\theta}^{(k)}\right)}{m-1}$.
7	$\boldsymbol{\theta}^{(k+1)} = \left(\boldsymbol{\beta}'^{(k+1)},\ \xi^{(k+1)}\right)'$.
8	$l^{(k+1)} = \ell\left(\boldsymbol{\theta}^{(k+1)}\right)$.
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \rightarrow k+1;\ go\ \ to\ \ 1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k+1)};\ stop. \end{cases}$

Table 4. E-M Algorithm of a MUB model with covariates for ξ.

$\boldsymbol{MUB}$-01 $\boldsymbol{Model\ \ with}\ \boldsymbol{\xi}=\boldsymbol{g}\left(\boldsymbol{\gamma};\boldsymbol{w}\right)$	
$\boldsymbol{\theta}=\left(\pi,\boldsymbol{\gamma}'\right)';\ \epsilon=10^{-6};\ dim\left(\boldsymbol{\theta}\right)=p_{\xi}+2.$	
$\ell\left(\boldsymbol{\theta}\right)=\sum_{i=1}^{n}\log\left\{\pi\left[\binom{m-1}{r_i-1}\frac{e^{(-\boldsymbol{w}_i\boldsymbol{\gamma})(r_i-1)}}{\left(1+e^{(-\boldsymbol{w}_i\boldsymbol{\gamma})}\right)^{m-1}}-\frac{1}{m}\right]+\frac{1}{m}\right\}$	
$Steps$	
0	$\boldsymbol{\theta}^{(0)}=\left(\pi^{(0)},\boldsymbol{\gamma}'^{(0)}\right)'=\left(\frac{1}{2},\,0.1,\ldots,0.1\right)';\ l^{(0)}=\ell\left(\boldsymbol{\theta}^{(0)}\right).$
1	$\xi_i^{(k)}=\frac{1}{1+e^{-\boldsymbol{w}_i'\boldsymbol{\gamma}^{(k)}}};\ b\left(r_i;\boldsymbol{\gamma}^{(k)}\right)=\binom{m-1}{r_i-1}\frac{e^{-(r_i-1)\,\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}}{\left(1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}\right)^{m-1}}\ ,\ i=1,2,\ldots,n.$
2	$\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)=\left[1+\frac{1-\pi^{(k)}}{m\,\pi^{(k)}\,b\left(r_i;\boldsymbol{\gamma}^{(k)}\right)}\right]^{-1},\ i=1,2,\ldots,n.$
3	
4	$\pi^{(k+1)}=\frac{1}{n}\sum_{i=1}^{n}\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right).$
5	$Q_2\left(\boldsymbol{\gamma}^{(k)}\right)=-\sum_{i=1}^{n}\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)\left\{(r_i-1)\,\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}+(m-1)\log\left[1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}\right]\right\}.$
6	$\boldsymbol{\gamma}^{(k+1)}=argmax_{\boldsymbol{\gamma}}\,Q_2\left(\boldsymbol{\gamma}^{(k)}\right).$
7	$\boldsymbol{\theta}^{(k+1)}=\left(\pi^{(k+1)},\ \boldsymbol{\gamma}'^{(k+1)}\right)'.$
8	$l^{(k+1)}=\ell\left(\boldsymbol{\theta}^{(k+1)}\right).$
9	$\begin{cases} if\ l^{(k+1)}-l^{(k)}\geq\epsilon, & k\rightarrow k+1;\ go\ \ to\ \ 1;\\ if\ l^{(k+1)}-l^{(k)}<\epsilon, & \hat{\boldsymbol{\theta}}=\boldsymbol{\theta}^{(k+1)};\ stop.\end{cases}$

Table 5. E-M Algorithm for a general MUB model with covariates.

$\boldsymbol{MUB}$-11 $\boldsymbol{Model\ \ with}\ \boldsymbol{\pi} = \boldsymbol{f}(\boldsymbol{\beta};\boldsymbol{y})\,;\ \boldsymbol{\xi} = \boldsymbol{g}(\boldsymbol{\gamma};\boldsymbol{w})$	
$\boldsymbol{\theta} = (\boldsymbol{\beta}';\boldsymbol{\gamma}')'\,;\ \ \epsilon = 10^{-6};\ dim(\boldsymbol{\theta}) = p_\pi + p_\xi + 2.$	
$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left\{\frac{1}{1+e^{(-\boldsymbol{y}_i\boldsymbol{\beta})}}\left[\binom{m-1}{r_i-1}\frac{e^{(-\boldsymbol{w}_i\boldsymbol{\gamma})(r_i-1)}}{(1+e^{(-\boldsymbol{w}_i\boldsymbol{\gamma})})^{m-1}} - \frac{1}{m}\right] + \frac{1}{m}\right\}$	
$Steps$	
0	$\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\beta}'^{(0)};\boldsymbol{\gamma}'^{(0)}\right)' = (\,0.1,\ldots,0.1;\,0.1,\ldots,0.1)'\,;\ l^{(0)} = \ell\left(\boldsymbol{\theta}^{(0)}\right).$
1	$\xi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}};\ b\left(r_i;\boldsymbol{\gamma}^{(k)}\right) = \binom{m-1}{r_i-1}\frac{e^{-(r_i-1)\,\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}}{\left(1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}\right)^{m-1}}\,,\ i = 1,2,\ldots,n.$
2	$\pi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}};\ \tau\left(r_i;\boldsymbol{\theta}^{(k)}\right) = \left[1+\frac{e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}}{m\,b\left(r_i;\xi^{(k)}\right)}\right]^{-1},\ i = 1,2,\ldots,n.$
3	
4	$Q_1\left(\boldsymbol{\beta}^{(k)}\right) = -\sum_{i=1}^{n}\left\{\log\left(1+e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}\right) + \left(1-\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)\right)e^{-\boldsymbol{y}_i\boldsymbol{\beta}^{(k)}}\right\}.$
5	$Q_2\left(\boldsymbol{\gamma}^{(k)}\right) = -\sum_{i=1}^{n}\tau\left(r_i;\boldsymbol{\theta}^{(k)}\right)\left\{(r_i-1)\,\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)} + (m-1)\log\left[1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}^{(k)}}\right]\right\}.$
6	$\boldsymbol{\beta}^{(k+1)}{=}argmax_{\boldsymbol{\beta}}\ Q_1\left(\boldsymbol{\beta}^{(k)}\right);\ \boldsymbol{\gamma}^{(k+1)}{=}\ argmax_{\boldsymbol{\gamma}}\ Q_2\left(\boldsymbol{\gamma}^{(k)}\right).$
7	$\boldsymbol{\theta}^{(k+1)} = \left(\boldsymbol{\beta}'^{(k+1)},\,\boldsymbol{\gamma}'^{(k+1)}\right)'.$
8	$l^{(k+1)} = \ell\left(\boldsymbol{\theta}^{(k+1)}\right).$
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \to k+1;\ go\ \ to\ \ 1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k+1)};\ stop. \end{cases}$

We observe that, in Tables 2-3, step 3, the conditional average rank $\overline{R}_n\left(\boldsymbol{\theta}^{(k)}\right)$ is the average rank weighted by the *a posteriori* probability that each observed rank originates from the first component distribution $b\left(r;\xi\right)$.

Secondly, according to the Bayes' theorem, the quantities:

$$\tau\left(r;\boldsymbol{\theta}^{(k)}\right) = \left[1+\frac{1-\pi^{(k)}}{m\,\pi^{(k)}\,b\left(r;\xi^{(k)}\right)}\right]^{-1} = \frac{\pi^{(k)}\,b\left(r;\,\xi^{(k)}\right)}{Pr\left(R=r\right)},$$

express the probability that the r-*th* unit belongs to the first component (the shifted Binomial population) given that $(R=r)$, $\forall r=1,2,\ldots,m$.

Finally, a main problem of the E-M algorithm is the choice of a convenient set of starting values for the estimates, since this procedure is generally slower than the second order convergence rates of the ML routines.

In the previous algorithm, we set the starting values according to the following criteria:

- for π, we choose the midrange of the parameter space;
- for ξ, we choose the moment estimator, given $\pi=1$, that is $\overline{R}_n$;
- for the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ vectors, we choose arbitrary small values (e.g., 0.1).

Of course, when some *a priori* information are available, it is convenient to choose more appropriate initial values; in fact, Piccolo (2003b) showed that moment estimates of the parameters for the MUB-00 model are suitable starting values in order to accelerate the convergence of the E-M algorithm[21].

The asymptotic variance-covariance of ML estimators has been derived by D'Elia (2003a) for MUB-01 and MUB-10 models, under the statement that the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ were asymptotically uncorrelated.

[21] Indeed, many different proposals have been suggested for this aim (McLachlan and Krishnan, 1997, 70-73); however, in our extensive experience –in estimating models for real data sets and running simulations experiments– we never need modifications of the previous stated rules.

This assumption has been adopted also by Piccolo (2003a) in the general MUB-11 model.

However, parameter estimators of the MUB models are correlated and, in some circumstances, the effect of ignoring the asymptotic correlation might be sensible; thus, in this paper, we remove this simplification and obtain the correct results[22].

6. *The information matrix of the* MUB*-00 model*

It is well known that the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of the ML estimators $\hat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ of a random variable $R \sim f(x; \boldsymbol{\theta})$ is obtained by inverting the negative of the expectation of the second derivatives (the Hessian) of the log-likelihood function $\ell(\boldsymbol{\theta})$.

Then, the *expected information matrix* $\boldsymbol{I}(\boldsymbol{\theta})$ and the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ are related by:

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\partial^2\, \ell(\boldsymbol{\theta})}{\partial\, \boldsymbol{\theta}\; \partial\, \boldsymbol{\theta}'}\right); \quad \boldsymbol{V}(\boldsymbol{\theta}) = \left[\boldsymbol{I}(\boldsymbol{\theta})\right]^{-1} .$$

An alternative method, which shares the same asymptotic properties, is based on the *observed information matrix* $\mathcal{I}(\boldsymbol{\theta})$, that is the Hessian computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Several Authors have supported statistical inference based on the observed information with respect to the expected one; in this regard, the main contribution is Efron and Hinkley (1978), while discussions with empirical evidences are reported by Lloyd (1999, 30-31) and Pawitan (2001, 244-247). Indeed, the observed information should be relevant when inferential statements are related to the sample under consideration, and thus it should deserve more importance for assuming statistical decisions.

[22] Empirical results on several case studies show that the effect of ignoring the correlation among the parameters estimators may be relevant, since it leads to a general understatement of standard errors. We report that this effect is less dramatic for the ξ parameter (and for the parameters of the covariates explaining ξ). However, we remark that any inferential consideration based on likelihood functions and/or deviances is unaffected by the amount of the parameters correlations.

With reference to the MUB model without covariates, D'Elia and Piccolo (2005a) obtained the explicit formulae[23] for the *expected* information matrix $\boldsymbol{I}(\boldsymbol{\theta})$, where the parameters are $\boldsymbol{\theta} = (\pi, \xi)'$. In this work, instead, we deduce the *observed* information matrix $\mathcal{I}(\boldsymbol{\theta})$ when the model is characterized by the parameters $\boldsymbol{\theta}$, and also when the parameters π and ξ are functions of the covariates observed on the sample units.

The matrix $\mathcal{I}(\boldsymbol{\theta})$ is obtained by explicit computation of the second derivatives of the log-likelihood function defined by:

$$\ell(\boldsymbol{\theta}) = \log \prod_{i=1}^{n} Pr\left(R = r_i \mid \boldsymbol{\theta}\right) = \sum_{r=1}^{m} n_r \, \log\left\{p_r(\boldsymbol{\theta})\right\},$$

where n_r are the sampling frequencies of $(R = r)$, $r = 1, 2, \ldots, m$ and the probability distribution $p_r(\boldsymbol{\theta}) = Pr\left(R = r \mid \pi, \, \xi\right)$ has been defined in section 3.

From an inferential point of view, in the MUB model without covariates, the set of information contained in $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ is strictly equivalent to the set $(n_1, n_2, \ldots, n_m)'$. In fact, for a given sample size n, the random sample $(N_1, N_2, \ldots, N_{m-1})'$ is a minimal sufficient statistic for $\boldsymbol{\theta}$.

However, in order to achieve formal results which are more homogenous and comparable with respect to MUB models with covariates, in the following expressions, we show also the formulae obtained by using all the sample data $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$. Thus, we will use the symbol r when it is useful to refer to the value of the observed random variable R or for grouped data (and we need also the corresponding frequencies $n_r, \; r = 1, 2, \ldots, m$), and we will use the symbol $r_i, \; i = 1, 2, \ldots, n$ when the ordinal value is referred to the sample observation.

[23] The *expected* information matrix for the MUB model without covariates has been obtained, by exploiting a well known result by Rao (1973, 367-368), which concerns score and information functions for grouped data.

After some algebra, we found that[24]:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi^2} = -\frac{1}{\pi^2} \sum_{r=1}^{m} n_r \, (1-q_r)^2 = -\frac{1}{\pi^2} \sum_{i=1}^{n} (1-q_i)^2 \, ;$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi \, \partial \xi} = +\frac{1}{\pi} \sum_{r=1}^{m} n_r \, v_r \, q_r \, q_r^* = +\frac{1}{\pi} \sum_{i=1}^{n} v_i \, q_i \, q_i^* \, ;$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \xi^2} = -\sum_{r=1}^{m} n_r q_r^* \left[u_r - (1-q_r^*) v_r^2\right] = -\sum_{i=1}^{n} q_i^* \left[u_i - (1-q_i^*) v_i^2\right] ;$$

where the quantities v_r, u_r and q_r, q_r^*, for $r = 1, 2, \ldots, m$, are defined by:

$$v_r = \frac{m-r}{\xi} - \frac{r-1}{1-\xi} \, ; \qquad u_r = \frac{m-r}{\xi^2} + \frac{r-1}{(1-\xi)^2} \, ;$$

$$q_r = \frac{1}{m \, p_r(\boldsymbol{\theta})} \, ; \qquad q_r^* = \frac{\pi \, b_r(\xi)}{p_r(\boldsymbol{\theta})} = 1 - (1-\pi) \, q_r \, ;$$

and, similarly, the quantities v_i, u_i and q_i, q_i^*, for $i = 1, 2, \ldots, n$, are defined by:

$$v_i = \frac{m-r_i}{\xi} - \frac{r_i-1}{1-\xi} \, ; \qquad u_i = \frac{m-r_i}{\xi^2} + \frac{r_i-1}{(1-\xi)^2} \, ;$$

$$q_i = \frac{1}{m \, p_i(\boldsymbol{\theta})} \, ; \qquad q_i^* = \frac{\pi \, b_i(\xi)}{p_i(\boldsymbol{\theta})} = 1 - (1-\pi) \, q_i \, .$$

Then, 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 obtained as:

[24] The formal expressions presented here are aimed at minimizing the computational effort; for instance, all the derivatives are expressed in terms of the probabilities $p_i(\boldsymbol{\theta})$, and their transformations, and so on.

$$\boldsymbol{V}(\boldsymbol{\theta}) = \left[\mathcal{I}(\hat{\boldsymbol{\theta}})\right]^{-1} = -\begin{pmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi\, \partial \xi} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi\, \partial \xi} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \xi^2} \end{pmatrix}^{-1}_{(\boldsymbol{\theta}=\hat{\boldsymbol{\theta}})} .$$

7. The information matrix of the MUB*-11 model*

In MUB models with covariates the log-likelihood function has to be expressed with reference to the sample units $i = 1, 2, \dots, n$, since each of them conveys different information about the values of the covariates.

In this regard, it is convenient to introduce some simplifying notations[25] to obtain the second derivatives of $\ell(\boldsymbol{\theta})$.

Thus, for any $i = 1, 2, \dots, n$, we let:

$$k_i = \binom{m-1}{r_i - 1}; \qquad b_i = e^{-\boldsymbol{y}_i \boldsymbol{\beta}}; \qquad c_i = e^{-\boldsymbol{w}_i \boldsymbol{\gamma}};$$
$$E_i(\boldsymbol{\beta}) = \frac{1}{1+b_i}; \qquad B_i(\boldsymbol{\gamma}) = k_i\, \frac{(c_i)^{r_i-1}}{(1+c_i)^{m-1}}.$$

In this way, the probability distribution for the general MUB model may be written[26] as:

$$p_i(\boldsymbol{\theta}) = E_i(\boldsymbol{\beta})\, \left\{B_i(\boldsymbol{\gamma}) - \frac{1}{m}\right\} + \frac{1}{m}; \quad i = 1, 2, \dots, n,$$

and the related log-likelihood function is expressed by:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[E_i(\boldsymbol{\beta})\, \left\{B_i(\boldsymbol{\gamma}) - \frac{1}{m}\right\} + \frac{1}{m}\right] .$$

[25] The objective of these settings is to separate in the MUB modeling the role of the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters, respectively. Then, we will omit explicit references to these parameters when no confusion occurs.

[26] Notice that the E_i are functions only of the $\boldsymbol{\beta}$ parameters, and the B_i are functions only of the $\boldsymbol{\gamma}$ parameters.

We also let:

$$F_i(\boldsymbol{\gamma}) = \frac{1}{1+c_i}; \qquad a_i(\boldsymbol{\gamma}) = (r_i - 1) - (m-1)(1-F_i)\,;$$
$$\widetilde{E}_i = E_i(1-E_i) = \frac{b_i}{(1+b_i)^2}; \qquad \widetilde{F}_i = F_i(1-F_i) = \frac{c_i}{(1+c_i)^2}\,;$$

and

$$E_i = E_i(\boldsymbol{\beta}); \qquad B_i = B_i(\boldsymbol{\gamma}); \qquad F_i = F_i(\boldsymbol{\gamma}); \qquad a_i = a_i(\boldsymbol{\gamma});$$

when no confusion arises.

In the sequel, we will need the following derivatives, for $i = 1, 2, \ldots, n$:

$$\begin{aligned}
\frac{\partial\, b_i}{\partial \beta_s} &= -y_{is}\, e^{-\boldsymbol{y}_i \boldsymbol{\beta}} = -y_{is}\, b_i; \quad s = 0, 1, 2, \ldots, p\,;\\
\frac{\partial\, c_i}{\partial \gamma_t} &= -w_{it}\, e^{-\boldsymbol{w}_i \boldsymbol{\gamma}} = -w_{it}\, c_i; \quad t = 0, 1, 2, \ldots, q;\\
\frac{\partial\, E_i}{\partial \beta_s} &= y_{is}\, E_i(1-E_i); \quad s = 0, 1, 2, \ldots, p\,;\\
\frac{\partial\, B_i}{\partial \gamma_t} &= -w_{it}\, a_i\, B_i; \quad t = 0, 1, 2, \ldots, q;\\
\frac{\partial\, a_i}{\partial \gamma_t} &= (m-1)\, w_{it}\, F_i(1-F_i); \quad t = 0, 1, 2, \ldots, q.
\end{aligned}$$

We generalize the symbols of the previous section, by defining:

$$q_i^* = 1 - (1-E_i)q_i; \quad \widetilde{Q}_i = q_i^*(1-q_i^*); \quad i = 1, 2, \ldots, n.$$

where $q_i = 1/\left(m\, p_i(\boldsymbol{\theta})\right);\ i = 1, 2, \ldots, n$, has been already defined with reference to MUB-00 models.

After lengthy algebraic calculations, we can finally obtain the second derivatives of the log-likelihood function. For any $s = 0, 1, 2, \ldots, p$ and

$t = 0, 1, 2, \dots, q$, they are expressed, respectively, by:

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_s \, \partial \beta_t} &= -\sum_{i=1}^{n} y_{is} \, y_{it} \left\{ \widetilde{E}_i - \widetilde{Q}_i \right\} ; \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_s \, \partial \gamma_t} &= -\sum_{i=1}^{n} y_{is} \, w_{it} \, a_i \, \widetilde{Q}_i \, ; \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_s \, \partial \gamma_t} &= -\sum_{i=1}^{n} w_{is} \, w_{it} \, \left\{ (m-1) \, q_i^* \, \widetilde{F}_i - a_i^2 \, \widetilde{Q}_i \right\} \, .
\end{aligned}$$

8. The information matrix of the MUB-01 and MUB-10 models

The general results obtained for the MUB-11 model require some modifications in the cases of MUB-10 and MUB-01 models, where only π or ξ are functions of covariates. Thus, in the following table, we collect the symbols we need for these models:

Symbols	*Model MUB-10*	*Model MUB-01*
$p_i(\boldsymbol{\theta})$	$E_i(\boldsymbol{\beta}) \left\{ B_i(\xi) - \frac{1}{m} \right\} + \frac{1}{m}$	$\pi \left\{ B_i(\boldsymbol{\gamma}) - \frac{1}{m} \right\} + \frac{1}{m}$
$B_i(\boldsymbol{\gamma})$	$B_i(\xi)$	$\binom{m-1}{r_i - 1} \frac{(e^{-\boldsymbol{w}_i \boldsymbol{\gamma}})^{r_i - 1}}{(1 + e^{-\boldsymbol{w}_i \boldsymbol{\gamma}})^{m-1}}$
$E_i(\boldsymbol{\beta})$	$\frac{1}{1 + e^{-\boldsymbol{y}_i \boldsymbol{\beta}}}$	π
$F_i(\boldsymbol{\gamma})$	ξ	$\frac{1}{1 + e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}}$
q_i	$\frac{1}{m \, p_i(\boldsymbol{\beta}, \xi)}$	$\frac{1}{m \, p_i(\pi, \boldsymbol{\gamma})}$
q_i^*	$1 - q_i \, \{1 - E_i(\boldsymbol{\beta})\}$	$1 - q_i \, \{1 - \pi\}$

Finally, the information matrices for the MUB-10 and MUB-01 models are expressed by the following formulae, where the dimensions of each component matrices have been indicated:

- $\boxed{MUB\textbf{-10}}$

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_s\partial\beta_t}\right]_{(p+1,p+1)} & \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_s\,\partial\,\xi}\right]_{(p+1,1)} \\ \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\xi\,\partial\,\beta_t}\right]_{(1,p+1)} & \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\xi^2}\right]_{(1,1)} \end{pmatrix}_{(\boldsymbol{\theta}=\hat{\boldsymbol{\theta}})}$$

where:

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_s\partial\beta_t} &= \sum_{i=1}^{n} y_{is}\, y_{it} \left\{\widetilde{E}_i - \widetilde{Q}_i\right\}\ ; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_s\,\partial\,\xi} &= -\sum_{i=1}^{n} y_{is}\, v_i\, \widetilde{Q}_i\,; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\xi\,\partial\,\beta_t} &= -\sum_{i=1}^{n} y_{it}\, v_i\, \widetilde{Q}_i\,; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\xi^2} &= \sum_{i=1}^{n} \left\{u_i\, q_i^* - v_i^2\, \widetilde{Q}_i\right\}\ . \end{aligned}$$

- $\boxed{MUB\textbf{-01}}$

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\pi^2}\right]_{(1,1)} & \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\pi\,\partial\,\gamma_t}\right]_{(1,q+1)} \\ \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\gamma_s\,\partial\,\pi}\right]_{(q+1,1)} & \left[\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\gamma_s\,\partial\gamma_t}\right]_{(q+1,q+1)} \end{pmatrix}_{(\boldsymbol{\theta}=\hat{\boldsymbol{\theta}})}$$

where:

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\pi^2} &= \frac{1}{\pi^2} \sum_{i=1}^{n} \{1 - q_i\}^2\ ; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\pi\,\partial\,\gamma_t} &= \frac{1}{\pi} \sum_{i=1}^{n} w_{it}\, a_i\, q_i\, q_i^*\,; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\gamma_s\,\partial\,\pi} &= \frac{1}{\pi} \sum_{i=1}^{n} w_{is}\, a_i\, q_i\, q_i^*\,; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\gamma_s\,\partial\gamma_t} &= \sum_{i=1}^{n} w_{is}\, w_{it}\, \left\{(m-1)\, q_i^*\, \widetilde{F}_i - a_i^2\, \widetilde{Q}_i\right\}\ . \end{aligned}$$

9. A unifying scheme for numerical implementation

In the previous sections, the matrix $\boldsymbol{V}(\boldsymbol{\theta}) = \left[\mathcal{I}(\hat{\boldsymbol{\theta}})\right]^{-1}$ has been obtained for each type of MUB model.

Specifically, the matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of a MUB model without covariates has dimensions (2×2) and it is computed by the formulae obtained in section 6. Instead, for MUB models with covariates, the matrix $\boldsymbol{V}(\boldsymbol{\theta})$ is evaluated by means of the formulae presented in sections 7-8.

In this section, we now present a unifying scheme that allows us to simplify the numerical implementation of the variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ in a matrix-oriented language[27].

We consider the following *inputs* and *output*:

INPUT	*OUTPUT*
• the fixed value of m • the sample vector $\boldsymbol{r}$ • the sample matrices $\boldsymbol{Y}, \boldsymbol{W}$ • the ML estimates $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}$	• the variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$

and we suggest to proceed as follows:

1. Compute probabilities $p_i(\boldsymbol{\theta}),\ i = 1, 2, \dots, n$;

2. Compute the following vectors of length n, according to the definitions given in sections 6-8:

Models	*Vector elements*
MUB-00	$q_i,\ q_i^*,\ v_i,\ u_i$
MUB-10	$q_i,\ q_i^*,\ v_i,\ \widetilde{E}_i,\ \widetilde{Q}_i$
MUB-01	$q_i,\ q_i^*,\ a_i,\ \widetilde{E}_i,\ \widetilde{F}_i,\ \widetilde{Q}_i$
MUB-11	$q_i,\ q_i^*,\ a_i,\ \widetilde{F}_i,\ \widetilde{Q}_i$

[27] The implementation we will discuss about is related to the $GAUSS$ language. Minor modifications are necessary for programming in R (and S-plus) or $MATLAB$ languages.

3. Define the vectors $\widetilde{\boldsymbol{f}}$, $\widetilde{\boldsymbol{g}}$, $\widetilde{\boldsymbol{g}}_{10}$, $\widetilde{\boldsymbol{g}}_{01}$, $\widetilde{\boldsymbol{h}}$, whose elements for $i = 1, 2, \dots, n$ are specified in the following table:

Models	*Vectors*	*Elements*
MUB-10 , MUB-11	$\widetilde{\boldsymbol{f}}$	$\widetilde{f}_i = \widetilde{E}_i - \widetilde{Q}_i$
MUB-11	$\widetilde{\boldsymbol{g}}$	$\widetilde{g}_i = a_i\, \widetilde{Q}_i$
MUB-10	$\widetilde{\boldsymbol{g}}_{10}$	$\widetilde{g}_{10,i} = v_i\, \widetilde{Q}_i$
MUB-01	$\widetilde{\boldsymbol{g}}_{01}$	$\widetilde{g}_{01,i} = \dfrac{a_i\, q_i\, q_i^*}{\pi}$
MUB-01 , MUB-11	$\widetilde{\boldsymbol{h}}$	$\widetilde{h}_i = (m-1)\, q_i^*\, \widetilde{F}_i - a_i^2\, \widetilde{Q}_i$

Given the sample data $\{\boldsymbol{r}, \boldsymbol{Y}, \boldsymbol{W}\}$, the *observed information matrix* $\mathcal{I}(\hat{\boldsymbol{\theta}})$, computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, is obtained by:

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \mathcal{I}_{11}(\hat{\boldsymbol{\theta}}) & \mathcal{I}_{12}(\hat{\boldsymbol{\theta}}) \\ \mathcal{I}_{21}(\hat{\boldsymbol{\theta}}) & \mathcal{I}_{22}(\hat{\boldsymbol{\theta}}) \end{pmatrix} = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}'_{12} & \mathcal{I}_{22} \end{pmatrix}.$$

where each sub-matrix is specified as follows[28]:

[28] It is convenient to introduce the *element-wise matrix product* $\boldsymbol{A} \odot \boldsymbol{v}$ between a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{v}$ (with the same number of columns).

Formally, if $\boldsymbol{A}$ is a $(n \times p)$ matrix and $\boldsymbol{v}$ is a $(n \times 1)$ vector, we define the *element-wise matrix product* as the matrix $\boldsymbol{B} = \boldsymbol{A} \odot \boldsymbol{v}$ of dimensions $(n \times p)$ matrix whose elements

Models	$\mathcal{I}_{11}$	$\mathcal{I}_{12} = \mathcal{I}'_{21}$	$\mathcal{I}_{22}$
MUB-00 $(2,2)$	$\frac{1}{\pi^2}\sum_{i=1}^{n}\{1-q_i\}^2$ $(1,1)$	$-\frac{1}{\pi}\sum_{i=1}^{n} v_i q_i q_i^*$ $(1,1)$	$\sum_{i=1}^{n}\left\{u_i\, q_i^* - v_i^2\, \widetilde{Q}_i\right\}$ $(1,1)$
MUB-10 $(p+2,p+2)$	$\boldsymbol{Y}'(\boldsymbol{Y}\odot\widetilde{\boldsymbol{f}})$ $(p+1,p+1)$	$-\boldsymbol{Y}'\widetilde{\boldsymbol{g}}_{10}$ $(p+1,1)$	$\sum_{i=1}^{n}\left\{u_i\, q_i^* - v_i^2\, \widetilde{Q}_i\right\}$ $(1,1)$
MUB-01 $(q+2,q+2)$	$\frac{1}{\pi^2}\sum_{i=1}^{n}\{1-q_i\}^2$ $(1,1)$	$\widetilde{\boldsymbol{g}}'_{01}\boldsymbol{W}$ $(1,q+1)$	$\boldsymbol{W}'(\boldsymbol{W}\odot\boldsymbol{h})$ $(q+1,q+1)$
MUB-11 $(p+q+2,p+q+2)$	$\boldsymbol{Y}'(\boldsymbol{Y}\odot\widetilde{\boldsymbol{f}})$ $(p+1,p+1)$	$\boldsymbol{Y}'(\boldsymbol{W}\odot\widetilde{\boldsymbol{g}})$ $(p+1,q+1)$	$\boldsymbol{W}'(\boldsymbol{W}\odot\widetilde{\boldsymbol{h}})$ $(q+1,q+1)$

Finally, the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of the ML estimators is computed for any MUB model as:

$$\boldsymbol{V}(\boldsymbol{\theta}) \;=\; \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}'_{12} & \mathcal{I}_{22} \end{pmatrix}^{-1} \;=\; \left[\mathcal{I}(\hat{\boldsymbol{\theta}})\right]^{-1}\,.$$

10. Inference on feeling and uncertainty parameters

The formulae presented in the previous sections are suitable for statistical decisions concerning the parameter $\boldsymbol{\theta}$ in the MUB models.

b_{ij} are obtained by the relationships:

$$\{b_{ij}\} = \{a_{ij}\, v_i\}, \qquad i = 1, 2, \ldots, n;\; j = 1, 2, \ldots, p\,.$$

This operation is quite common in matrix-oriented languages.

In some applications, the inferential interest focuses directly on the uncertainty (π) and feeling (ξ) parameters, while the previous asymptotic expressions only refer to $\boldsymbol{\beta}$ and/or $\boldsymbol{\gamma}$, respectively. This problem is relevant when the population is stratified and we use the stratification factor as a covariate for one or both MUB parameters;for instance, usually, stated preferences differ according to gender, occupation, residence, etc. Thus, it will be interesting to account for such variables when making inference on π and ξ conditional to the gender, say[29].

It is convenient to embed the above problem in a general framework, and to derive the asymptotic standard errors for π and ξ, respectively, given a subject's ***profile*** $\boldsymbol{d}_i = (d_{0i}, d_{1i}, \dots, d_{ki})$, for any $i = 1, 2, \dots, n$.

Here, we denote by "profile" the values assumed by the subject's covariates that are present in the estimated MUB model; thus, in a sense, the profile characterizes the subject with relation to the stated choice.

As usual, the first components $d_{0i} = 1$, and we denote by k the number of subject's covariates required for specifying the profile.

Then, we apply the delta method[30] to the estimators defined by:

$$\hat{\pi}_i = \pi(\hat{\boldsymbol{\beta}}) \mid \boldsymbol{d}_i = \frac{1}{1 + e^{-\boldsymbol{d}_i \hat{\boldsymbol{\beta}}}}; \qquad \hat{\xi}_i = \xi(\hat{\boldsymbol{\gamma}}) \mid \boldsymbol{d}_i = \frac{1}{1 + e^{-\boldsymbol{d}_i \hat{\boldsymbol{\gamma}}}}\,.$$

Hereafter, when there is no risk of confusion, we omit the reference

[29] A real situation where this approach has been successfully applied to MUB models is discussed in D'Elia (2007).

[30] Specifically, let $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)'$. For an asymptotically unbiased ML estimator $T_n = g(\boldsymbol{\theta})$, the asymptotic variance of T_n is approximated as:

$$Var(T) \simeq [\boldsymbol{\delta}'\, \boldsymbol{V}(\boldsymbol{\theta})\, \boldsymbol{\delta}\,]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\,,$$

where

$$\boldsymbol{\delta} = \left(\frac{\partial}{\partial \theta_1}\, g(\boldsymbol{\theta}),\; \frac{\partial}{\partial \theta_2}\, g(\boldsymbol{\theta}), \dots, \frac{\partial}{\partial \theta_k}\, g(\boldsymbol{\theta}) \right)'$$

and $\boldsymbol{V}(\boldsymbol{\theta}) = \|Cov(\theta_s,\, \theta_t)\|$ is the variance-covariance matrix of the estimator $\hat{\boldsymbol{\theta}}$.

For the multivariate version of this approach refer to: Rao (1973, 388); Serfling (1980, 122-124); Casella and Berger (2002, 241-245) and, for categorical data, to: Agresti (2002, 73-74).

to the given profile. Thus, we let:

$$
\begin{aligned}
\boldsymbol{\delta}_{\boldsymbol{\beta}} &= \left(\frac{\partial}{\partial \beta_0}\, \pi(\hat{\boldsymbol{\beta}}),\; \frac{\partial}{\partial \beta_1}\, \pi(\hat{\boldsymbol{\beta}}), \dots, \frac{\partial}{\partial \beta_p}\, \pi(\hat{\boldsymbol{\beta}}) \right)' ;\\
\boldsymbol{\delta}_{\boldsymbol{\gamma}} &= \left(\frac{\partial}{\partial \gamma_0}\, \xi(\hat{\boldsymbol{\gamma}}),\; \frac{\partial}{\partial \gamma_1}\, \xi(\hat{\boldsymbol{\gamma}}), \dots, \frac{\partial}{\partial \gamma_q}\, \xi(\hat{\boldsymbol{\gamma}}) \right)' ;
\end{aligned}
$$

where, for any $i = 1, 2, \dots, n$:

$$
\begin{aligned}
\frac{\partial}{\partial \beta_s}\, \pi(\hat{\boldsymbol{\beta}}) &= d_{si}\, \frac{e^{-\boldsymbol{d}_i\, \hat{\boldsymbol{\beta}}}}{\left(1 + e^{-\boldsymbol{d}_i\, \hat{\boldsymbol{\beta}}}\right)^2} = d_{si}\, \hat{\pi}_i\, (1 - \hat{\pi}_i); \quad s = 0, 1, 2, \dots, p;\\
\frac{\partial}{\partial \gamma_t}\, \xi(\hat{\boldsymbol{\gamma}}) &= d_{ti}\, \frac{e^{-\boldsymbol{d}_i\, \hat{\boldsymbol{\gamma}}}}{(1 + e^{-\boldsymbol{d}_i\, \hat{\boldsymbol{\gamma}}})^2} = d_{ti}\, \hat{\xi}_i\, (1 - \hat{\xi}_i); \quad t = 0, 1, 2, \dots, q\,.
\end{aligned}
$$

If $\boldsymbol{V}(\boldsymbol{\beta})$ and $\boldsymbol{V}(\boldsymbol{\gamma})$ are the variance-covariance matrices of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ estimators, respectively, then –for a well defined profile– the asymptotic variances of $\pi(\hat{\boldsymbol{\beta}})$ and $\xi(\hat{\boldsymbol{\gamma}})$ are given by:

$$
\begin{aligned}
Var(\pi(\hat{\boldsymbol{\beta}})) &\simeq \left[\boldsymbol{\delta}'_{\boldsymbol{\beta}}\, \boldsymbol{V}(\boldsymbol{\beta})\, \boldsymbol{\delta}_{\boldsymbol{\beta}}\,\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} ;\\
Var(\xi(\hat{\boldsymbol{\gamma}})) &\simeq \left[\boldsymbol{\delta}'_{\boldsymbol{\gamma}}\, \boldsymbol{V}(\boldsymbol{\gamma})\, \boldsymbol{\delta}_{\boldsymbol{\gamma}}\,\right]_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} .
\end{aligned}
$$

These results are useful for any MUB model with covariates. For the sake of simplicity, we develop in detail the case of a MUB-11 model when a single dichotomous covariate, assuming $0/1$ values[31], is related both to π and ξ. As a consequence, the only admissible subjects' profiles are:

$$
\boldsymbol{d}_0 = (1,\, 0)'; \qquad \boldsymbol{d}_1 = (1,\, 1)'\,.
$$

Then, omitting explicit reference to the subjects, the conditional pa-

[31] This situation is common, for instance, when the covariate *Gender* explains a different behavior of subjects' preference in terms of both feeling and uncertainty.

rameters are defined in the following scheme:

$$\hat{\pi}_0 = \pi(\hat{\boldsymbol{\beta}}) \mid d_0 = \frac{1}{1+e^{-\hat{\beta}_0}} \quad ; \quad \hat{\pi}_1 = \pi(\hat{\boldsymbol{\beta}}) \mid d_1 = \frac{1}{1+e^{-\hat{\beta}_0-\hat{\beta}_1}}$$
$$\hat{\xi}_0 = \xi(\hat{\boldsymbol{\gamma}}) \mid d_0 = \frac{1}{1+e^{-\hat{\gamma}_0}} \quad ; \quad \hat{\xi}_1 = \xi(\hat{\boldsymbol{\gamma}}) \mid d_1 = \frac{1}{1+e^{-\hat{\gamma}_0-\hat{\gamma}_1}}$$

Given the profile $\boldsymbol{d}_0 = (1,\, 0)'$, for any $i = 1, 2, \ldots, n$, the vectors of the derivatives, are:

$$\boldsymbol{\delta}_{\boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial}{\partial\beta_0}\,\hat{\pi}_0 \\ \\ \frac{\partial}{\partial\beta_1}\,\hat{\pi}_1 \end{pmatrix} = \begin{pmatrix} \hat{\pi}_0(1-\hat{\pi}_0) \\ \\ 0 \end{pmatrix} = \hat{\pi}_0(1-\hat{\pi}_0)\,\boldsymbol{d}_0';$$

$$\boldsymbol{\delta}_{\boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial}{\partial\gamma_0}\,\hat{\xi}_0 \\ \\ \frac{\partial}{\partial\gamma_1}\,\hat{\xi}_1 \end{pmatrix} = \begin{pmatrix} \hat{\xi}_0(1-\hat{\xi}_0) \\ \\ 0 \end{pmatrix} = \hat{\xi}_0(1-\hat{\xi}_0)\,\boldsymbol{d}_0'\,.$$

Similarly, given the profile $\boldsymbol{d}_1 = (1,\, 1)'$, for any $i = 1, 2, \ldots, n$, the vectors of the derivatives, are:

$$\boldsymbol{\delta}_{\boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial}{\partial\beta_0}\,\hat{\pi}_1 \\ \\ \frac{\partial}{\partial\beta_1}\,\hat{\pi}_1 \end{pmatrix} = \begin{pmatrix} \hat{\pi}_1(1-\hat{\pi}_1) \\ \\ \hat{\pi}_1(1-\hat{\pi}_1) \end{pmatrix} = \hat{\pi}_1(1-\hat{\pi}_1)\,\boldsymbol{d}_1';$$

$$\boldsymbol{\delta}_{\boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial}{\partial\gamma_0}\,\hat{\xi}_1 \\ \\ \frac{\partial}{\partial\gamma_1}\,\hat{\xi}_1 \end{pmatrix} = \begin{pmatrix} \hat{\xi}_1(1-\hat{\xi}_1) \\ \\ \hat{\xi}_1(1-\hat{\xi}_1) \end{pmatrix} = \hat{\xi}_1(1-\hat{\xi}_1)\,\boldsymbol{d}_1'\,.$$

These expressions may be stated, for any profile $j = 0, 1$, in a compact way as:

$$\boldsymbol{\delta}_{\boldsymbol{\beta}}^{(j)} = \hat{\pi}_j\,(1-\hat{\pi}_j)\,\boldsymbol{d}_j';$$
$$\boldsymbol{\delta}_{\boldsymbol{\gamma}}^{(j)} = \hat{\xi}_j\,(1-\hat{\xi}_j)\,\boldsymbol{d}_j'\,.$$

If we partition the variance-covariance matrices $\boldsymbol{V}(\boldsymbol{\beta})$ and $\boldsymbol{V}(\boldsymbol{\gamma})$ as:

$$\boldsymbol{V}(\boldsymbol{\beta}) = \begin{pmatrix} v_{00}^{(\boldsymbol{\beta})} & v_{01}^{(\boldsymbol{\beta})} \\ \\ v_{01}^{(\boldsymbol{\beta})} & v_{11}^{(\boldsymbol{\beta})} \end{pmatrix}; \qquad \boldsymbol{V}(\boldsymbol{\gamma}) = \begin{pmatrix} v_{00}^{(\boldsymbol{\gamma})} & v_{01}^{(\boldsymbol{\gamma})} \\ \\ v_{01}^{(\boldsymbol{\gamma})} & v_{11}^{(\boldsymbol{\gamma})} \end{pmatrix};$$

the asymptotic variances of the π and ξ estimators, for the given profiles, are:

$$\begin{aligned} Var(\hat{\pi}_j) &\simeq \left[\boldsymbol{\delta}_{\boldsymbol{\beta}}'^{(j)}\, \boldsymbol{V}(\boldsymbol{\beta})\, \boldsymbol{\delta}_{\boldsymbol{\beta}}^{(j)}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\ &= \left[\hat{\pi}_j\,(1-\hat{\pi}_j)\right]^2\, d_j' \begin{pmatrix} v_{00}^{(\boldsymbol{\beta})} & v_{01}^{(\boldsymbol{\beta})} \\ \\ v_{01}^{(\boldsymbol{\beta})} & v_{11}^{(\boldsymbol{\beta})} \end{pmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} d_j; \\ Var(\hat{\xi}_j) &\simeq \left[\boldsymbol{\delta}_{\boldsymbol{\gamma}}'^{(j)}\, \boldsymbol{V}(\boldsymbol{\gamma})\, \boldsymbol{\delta}_{\boldsymbol{\gamma}}^{(j)}\right]_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} \\ &= \left[\hat{\xi}_j\,(1-\hat{\xi}_j)\right]^2\, d_j' \begin{pmatrix} v_{00}^{(\boldsymbol{\gamma})} & v_{01}^{(\boldsymbol{\gamma})} \\ \\ v_{01}^{(\boldsymbol{\gamma})} & v_{11}^{(\boldsymbol{\gamma})} \end{pmatrix}_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} d_j\,. \end{aligned}$$

Given the dichotomous nature of the covariates, the previous expressions may be further simplified. For instance, if $j=0$, the asymptotic variances become:

$$\begin{aligned} Var(\hat{\pi}_0) &\simeq \left[\hat{\pi}_0\,(1-\hat{\pi}_0)\right]^2 \left(v_{00}^{(\boldsymbol{\beta})}\right)_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}; \\ Var(\hat{\xi}_0) &\simeq \left[\hat{\xi}_0\,(1-\hat{\xi}_0)\right]^2 \left(v_{00}^{(\boldsymbol{\gamma})}\right)_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}; \end{aligned}$$

and so on.

We apply the previous formulae to a real data set of $n = 354$ sample units related to the ranking of the concern for the immigration in the city of Naples. D'Elia and Piccolo (2005b) obtained a MUB-11 model whose estimates are reported in the following table.

Parameters estimates	Standard errors
$\hat{\beta}_0 = 1.599$	0.244
$\hat{\beta}_1 = -0.801$	0.354
$\hat{\gamma}_0 = -1.468$	0.409
$\hat{\gamma}_1 = -0.055$	0.016

The covariate for the uncertainty parameter π is the *Gender* (=0 for Males, =1 for Females) while the covariate for the parameter[32] ξ is the *Age* of the subject (expressed in years and in the interval $(18, 55)$ for the observed sample). The corresponding estimated parameters are all high significant.

We compare two profiles by defining an *elderly man* ($Gender = 0$ and $Age = 55$ years) and a *young woman* ($Gender = 1$ and $Age = 19$ years); thus, in this example, we have a different profile for the variables Y (concerning π) and W (concerning ξ), respectively.

$$\begin{array}{lrcl} \text{Elderly Man :} & \boldsymbol{d}_0 & = & (y_{00} = 1,\, y_{01} = 0 \mid w_{00} = 1,\, w_{01} = 55)' \\ \text{Young Woman :} & \boldsymbol{d}_1 & = & (y_{10} = 1,\, y_{11} = 1 \mid w_{10} = 1,\, w_{11} = 19)' \end{array}$$

The results are shown in the following table (we report in parenthesis the asymptotic standard errors of the estimates), where the 95% confidence intervals based on the previous asymptotic approximations are also presented.

Profiles	$\hat{\pi}$	$C.I.(\pi)$	$\hat{\xi}$	$C.I.(\xi)$
Elderly Man $(1,\, 0 \mid 1,\, 55)'$	0.832 (0.034)	$[0.764 - 0.900]$	0.011 (0.005)	$[0.0002 - 0.022]$
Young Woman $(1,\, 1 \mid 1,\, 19)'$	0.690 (0.061)	$[0.568 - 0.811]$	0.075 (0.012)	$[0.052 - 0.098]$

Notice that the two profiles imply different parameters as confirmed by Figure 3, where we plot the MUB probability distributions conditioned by the profiles $\boldsymbol{d}_0$ and $\boldsymbol{d}_1$, respectively.

[32] Notice that, in this study, the "feeling" is indeed a *measure of concern*.

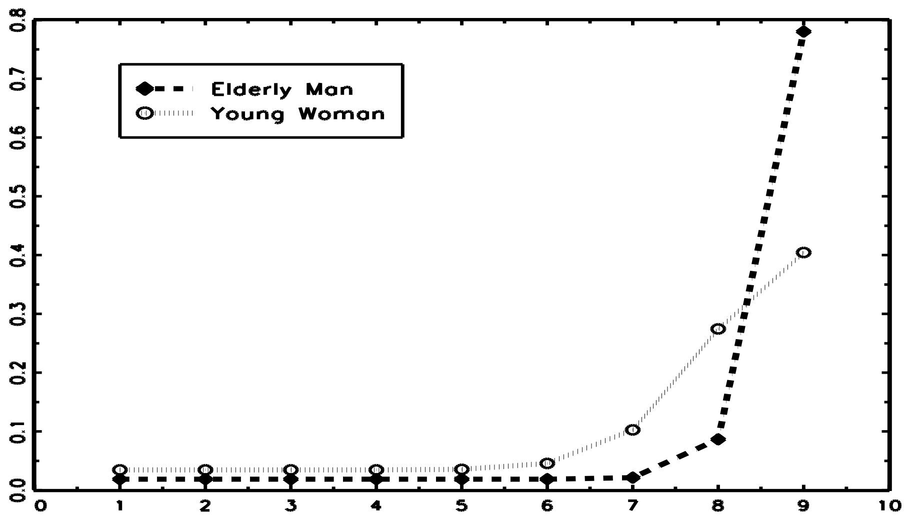

Figure 3: *Profiles of Immigration concern according to MUB models.*

However, the confidence intervals for the uncertainty parameter (π) have an overlapping region while the corresponding intervals for the feeling parameter (ξ) are well disjoint among the two profiles. These considerations support the conclusion that while similar in the uncertainty, the concerns for the immigration problem expressed by young women and elderly men are significantly different.

11. Finite sample performance behavior

In this section we perform a simulation experiment to assess the finite sample size performance of the asymptotic standard errors as computed by expected and observed information matrices for the MUB-00 model[33].

Given $m = 9$, we choose a MUB model with $\pi = 0.3$ and $\xi = 0.8$;

[33] An extensive simulation experiment for the MUB-00 model has been carried out by D'Elia (2004) using the expected information matrix.

such a model is characterized by a positive skewness with a mode at $R = 2$ and a secondary mode at $R = 3$. The mean value of this random variable is $\mathbb{E}\,(R) = 4.280$ and the uncertainty factor is $(1-\pi)/m = 0.078$.

The results that we present are based on 500 simulations; however, 100 simulations were sufficient to produce stable results. With regard to the sample size, the experiment has been repeated for $n = 100, 200, 300, 500, 1000, 5000$.

For each simulation experiment and for both parameters, we show the theoretical standard error computed by assuming that the parameters are known (using the expected information as in D'Elia and Piccolo, 2005a); then, we compute the average and the standard deviation of the ML estimates obtained in each run; finally, we present the average of the standard errors computed by means of the expected and observed information matrices, respectively.

$n \rightarrow$ Statistics	100	200	300	500	1000	5000
AVER($\hat{\pi}$)	0.3078	0.3063	0.3025	0.3021	0.2988	0.3019
AVER($\hat{\xi}$)	0.7992	0.7995	0.8013	0.7993	0.8006	0.7993
STDEV($\hat{\pi}$)	0.0935	0.0673	0.0544	0.0413	0.0306	0.0136
AVSTERREXP($\hat{\pi}$)	0.0943	0.0671	0.0548	0.0426	0.0301	0.0135
AVSTERROBS($\hat{\pi}$)	0.0944	0.0671	0.0548	0.0427	0.0302	0.0135
STDEV($\hat{\xi}$)	0.0511	0.0364	0.0277	0.0221	0.0148	0.0069
AVSTEXP($\hat{\xi}$)	0.0495	0.0340	0.0276	0.0212	0.0150	0.0066
AVSTDOBS($\hat{\xi}$)	0.0496	0.0342	0.0276	0.0213	0.0151	0.0066

For a full understanding of the previous table, we list the abbreviations used for each estimated parameters:

– AVER: the average of the estimated parameters obtained by the simulations;

– STDEV: the standard deviation of the estimated parameters obtained by the simulations;

– AVSTDEXP: the average of the standard deviation based on the expected information matrix of the estimated parameters in the simulations;

– AVSTDOBS: the average of the standard deviation based on the observed information matrix of the estimated parameters in the simulations.

The results of the simulation experiments enhance the following points[34]:

- The main differences among the standard errors are registered only for small sample sizes[35].

- Generally, all the values shown in the previous table are similar, confirming that standard errors based on expected and observed information matrices are very close. Anyway, the standard errors based on observed information matrices are generally larger than those derived by the expected ones, although their difference is quite small.

- The simulated distributions of the ML estimates for π and ξ are adequately approximated by a Normal distribution, also for small sample size; this confirms that the correctness of the classical asymptotic theory still holds in statistical decisions concerning the MUB parameters estimators.

- In all situations, the standard errors of the ξ estimators are lower than the corresponding standard errors of the π estimators and their ratio ranges between 2 and 4.

- The simulated distributions of the standard errors derived by the expected and observed matrices are similar in all cases; however,

[34] Some considerations also derive from box-plots, kernel density estimates and further statistical measures that we omit to show for space constraints.

[35] Strictly speaking, the results for $n = 100$ are not reliable since, for $m = 9$, we need at least 150-200 units to get significant results in MUB modeling. As a matter of fact, when $n \leq 100$, it is highly probable that some ordinal value might not be observed.

those related to ξ denote a positive skewness (caused by some extreme values) while those related to π denote a low kurtosis (because of the high concentration of the results around the mean value).

Thus, we can apply the asymptotic results of previous sections, even for moderate sample size. Moreover, our confidence is always higher for the feeling parameter ξ than for the uncertainty parameter π.

Finally, we suggest the use of the observed information matrix as a starting point for asymptotic statistical decisions on the MUB model parameters since this quantity is related to the observed sample and thus is able to capture the features of the data in a better way.

12. Concluding remarks

In this paper, we obtained the asymptotic variance-covariance of ML estimators for any MUB models, derived from the observed information matrix, and some results about direct inference on the feeling and uncertainty parameters in MUB models with covariates.

The simulation experiment has confirmed that the difference between the standard errors evaluated from the expected and observed information matrices for the MUB model without covariates is not generally significant, even in the case of moderate sample sizes.

Further topics to be explored, for a fully efficient implementation of MUB modeling, include the following issues:

- Simulated comparisons of classical ordinal and MUB models.
- Heteroscedastic behavior of the choice mechanism and inclusion of overdispersion effects.
- Efficient methods for the selection of significant covariates.
- Explicit inclusion of objects' covariates in MUB models.
- Significance levels of fitting measures for MUB models.

All these issues require further investigations about the related statistics distributions, and in several cases non-parametric techniques may be suggested.

Finally, it is worthwhile to mention that, hitherto, MUB models have been successfully applied in several different areas:

- the *preference analysis* towards: colors (expressed by young people, children, air force cadets) in D'Elia *et al.* (2001); the cities where to live in D'Elia and Sitzia (2002); the future professions (to be chosen by Political Sciences students) in D'Elia (2003b);

- the *consumers' preference* of: olive oil brands in Del Giudice and D'Elia (2001), and of salmon in Europe in D'Elia and Piccolo (2007); the *evaluation surveys* about Orientation services and University teaching in D'Elia (2001) and D'Elia and Piccolo (2002; 2006);

- the *levels of perception* of: chronic pain in D'Elia (2007); word synonymy in Cappelli and D'Elia (2004; 2006c); humor in Balirano and Corduas (2006);

- the *grade of concern* for: the main problems in a metropolitan area in D'Elia and Piccolo (2005b).

We remark that the MUB modeling approach is a general framework for the statistical interpretation of ordinal data, and it is not strictly related to preferences and evaluations analyses. For instance, given a random sample of related geographical units (boroughs, municipalities, districts, counties, regions, etc.), we can give an interpretation of the *ranking of the political parties* as a function of socio-economic, cultural and environmental variables, by estimating MUB models with covariates.

Acknowledgements: This work benefitted from the research structures of the Dipartimento di Scienze Statistiche, Università di Napoli Federico II, and CFEPSR, Portici. Critical suggestions and amendments by two referees are gratefully acknowledged; a special thanks is due for their careful checking of the formal developments contained in this paper. Of course, usual disclaimers apply. The main result of the paper was obtained while the Author was involved in the Erskine Fellowship Programme during a three months stage at University of Canterbury, Christchurch, New Zealand.

References

Agresti A. (1984), *Analysis of Ordinal Categorical Data*, J. Wiley & Sons, New York.

Agresti A. (1996), *An Introduction to Categorical Data Analysis*, J. Wiley & Sons, New York.

Agresti A. (2002), *Categorical Data Analysis*, 2^{nd} edition, J. Wiley & Sons, New York.

Amemiya T. (1981), Qualitative response models: a survey, *Journal of Economic Literature*, XIX, 1483-1536.

Balirano G., Corduas M. (2006a), Detecting Semiotically Expressed Humor in Diasporic TV Productions, *HUMOR: International Journal of Humor Research*, forthcoming.

Balirano G., Corduas M. (2006b), Statistical methods for the linguistic analysis of a humorous TV sketch show, *Quaderni di Statistica*, 8, this issue.

Bartholomew D. J. (1987), *Latent Variable models and Factor Analysis*, M. Dekker, New York.

Cagnone S., Gardini A. and Mignani S. (2004), New developments of latent variable models with ordinal data, *Atti della XLII Riunione Scientifica SIS*, Bari, vol.I, 1-12.

Cappelli C., D'Elia A. (2004), La percezione della sinonimia: un'analisi statistica mediante modelli per ranghi, in: *Le poids des mots - Actes de JADT2004* (Prunelle G., Fairon C. and Dister A. Eds.), Presses Universitaires de Louvain, Belgium, 229-240.

Cappelli C., D'Elia A. (2006a), Preliminary identification of relevant covariates in models for ordinal data: a strategy based on tree methods, in: (Croux C., Riani M., Chiandotto B., Bini M. and Bertaccini B., Editors) *ROBCLA-06 Book of Abstracts*, Università di Firenze, 23-24.

Cappelli C., D'Elia A. (2006b), A tree-based method for variable selection in models for ordinal data, *Quaderni di Statistica*, in press.

Cappelli C., D'Elia A. (2006c), Grouping near-synonymy of a dictionary entry: thesauri and perceptions, submitted to *Computational Linguistics*.

Casella S., Berger R. L. (2002), *Statistical Inference*, 2nd edition, Duxbury, Australia.

Cramer J. S. (2001), *An Introduction to the Logit Model for Economists*, 2^{nd} edition, Timberlake Consultants Ltd., London.

Davison A. C. (2003), *Statistical Models*, Cambridge University Press, Cambridge.

Del Giudice T., D'Elia, A. (2001), Valorizzazione dell'olio extravergine di oliva meridionale: una proposta metodologica per l'analisi delle preferenze, *Rivista di Economia Agraria*, LVI, 571-609.

D'Elia A. (1999), A proposal for ranks statistical modelling, in: H. Friedl, A. Berghold, G. Kauermann (eds.), *Statistical Modelling*, Graz-Austria, 468-471.

D'Elia A. (2000a), Il meccanismo dei confronti appaiati nella modellistica per graduatorie: sviluppi statistici ed aspetti critici, *Quaderni di Statistica*, 2, 173-203.

D'Elia A. (2000b): A Shifted Binomial Model for Rankings, in Statistical Modelling - XV International Workshop on Statistical Modelling, (Nunez-Anton, V. and Ferreira, E. eds.), Servicio Editorial de la Universidad del Pais Vasco, 412 - 416.

D'Elia A. (2001), Efficacia didattica e strutture universitarie: la valutazione mediante un approccio modellistico, in: *Processi e metodi statistici di valutazione*, SIS, ECRA, Roma, 21-24.

D'Elia A. (2003a), Modelling ranks using the Inverse Hypergeometric distribution, *Statistical Modelling: an International Journal*, 3, 65-78.

D'Elia A. (2003b) A mixture model with covariates for ranks data: some inferential developments, *Quaderni di Statistica*, 5, 1-25.

D'Elia A. (2004), Finite sample performance of the E-M algorithm for ranks data modelling, *Statistica*, LXIII, 41-51..

D'Elia A. (2007) A statistical modelling approach for the analysis of TMD chronic pain data, *Statistical Methods in Medical Research*, accepted for publication.

D'Elia A., Mauriello E. and Sitzia, N. (2001), Uno studio sulla scelta dei colori: esperienze e metodi statistici, *Atti del Convegno Nazionale su Matematica, Formazione Scientifica e Nuove Tecnologie*, Montevarchi, 85 - 96.

D'Elia A., Piccolo D. (2002), Problemi e metodi statistici nei processi di valutazione della didattica, *Atti della Giornata di Studio su "Valutazione della Didattica e dei Servizi nel Sistema Universitario"*, Università di Salerno, 105-127.

D'Elia A., Piccolo D. (2005a), A mixture model for preference data analysis, *Computational Statistics & Data Analysis*, 49, 917-934.

D'Elia A., Piccolo D. (2005b), Uno studio sulla percezione delle emergenze metropolitane: un approccio modellistico, *Quaderni di Statistica*, 7, 121-161.

D'Elia A., Piccolo D. (2006), Analyzing evaluation data: modelling and testing for homogeneity, in: S.Zani, A.Cerioli, M.Riani and M.Vichi editors: *Data Analysis, Classification and the Forward Search*, Springer, Berlin, 299-307.

D'Elia A., Piccolo D. (2007), A new approach for modelling consumers' preferences: some evidence from the Euro-salmon project, *Food and Quality Preference*, submitted for publication.

D'Elia A., Sitzia N. (2002), Selecting between Non-Nested Models for Preference Data, *Atti della XLI Riunione Scientifica SIS*, 129-132, CLEUP, Padova.

Dobson A. J. (2002), *An Introduction to Generalized Linear Models*, 2^{nd} edition, Chapman & Hall/CRC, Boca Raton.

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

Everitt B. S. (1984), *An Introduction to Latent Variable Models*, Chapman and Hall, New York.

Fligner M. A., Verducci J. S. (1999), *Probability Models and Statistical Analyses of Ranking Data*, Springer-Verlag, New York.

Fox J. (1997), *Applied Regression Analysis, Linear Models, and related Methods*, Sage Publications, Thousand Oaks, California.

Franses P. H., Paap R. (2001), *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

Gormley I.C., Murphy T. B. (2006), Analysis of Irish third-level college applications data, *Journal of the Royal Statistical Society, Series A*, 169, 361-379.

Greene W. H. (2000), *Econometric Analysis*, 4^{th} edition, Prentice Hall International, Inc., New Jersey.

Hensher D.A., Rose J. M. and Greene W. H. (2005), *Applied Choice Analysis. A Primer*, Cambridge University Press, Cambridge.

Huber P., Ronchetti E. and Victoria-Feser M.P. (2004), Estimation of generalized linear latent variable models, *Journal of the Royal Statistical Society, Series B*, 66, 893-908.

Jöreskog K. G., Moustaki I. (2001), Factor analysis of ordinal variables: a comparison of three approaches, *Multivariate Behavioral Research*, 36, 347-387.

Jorgenson M. (2002), E-M algorithm, in: A.H. El-Shaarawi and W.W. Piegorsch (editors), *Encyclopedia of Environmetrics*, vol.2, 637-653.

King G., Tomz M. and Wittenberg J. (2000), Making the most of statistical analyses: improving interpretation and presentation, *American Journal of Political Science*, 44, 341-355.

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

Louviere J. J., Hensher D.A. and Swait J. D. (2000), *Stated Choice Methods. Analysis and Applications*, Cambridge University Press, Cambridge.

Maddala G. S. (1995), *Limited-dependent and Qualitative Variables in Econometrics*,, Cambridge University Press, Cambridge.

Marden J. I. (1995), *Analyzing and Modelling Rank Data*, Chapman & Hall, London.

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

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

McLachlan G., Krishnan G. J. (1997), *The EM Algorithm and Extensions*, J. Wiley & Sons, New York.

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

Mittelhammer R. C., Judge G. G. and Miller D. J. (1995), *Econometric Foundations*, Cambridge University Press, Cambridge.

Moustaki I. (2000), A latent variable model for ordinal data, *Applied Psychological Measurement*, 24, 211-223.

Moustaki I. (2003), A general class of latent variable model for ordinal manifest variables with covariate effects on the manifest and latent variables, *British Journal of Mathematical and Statistical Psychology*, 56, 337-357.

Moustaki I., Knott M. (2000), Generalized latent trait models, *Psychometrika*, 65, 391-411.

Moustaki I., Papageorgiou I. (2004), Latent class models for mixed variables with applications in Archaeometry, *Computational Statistics & Data Analysis*, 48, 1-17.

Nelder J. A., 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. (2003a), Computational issues in the E-M algorithm for ranks model estimation with covariates, *Quaderni di Statistica*, 5, 27-48. Corrections in: *Quaderni di Statistica*, 6, 199.

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

Power D. A., Xie Y. (2000), *Statistical Methods for Categorical Data Analysis*, Academic Press, London.

Rao C. R. (1973), *Linear Statistical Inference and Its Applications*, 2^{nd} edition, J. Wiley & Sons, New York.

Sammel M. D., Ryan L. M. and Legler J. M. (1997), Latent variable models for mixed discrete and continuous outcomes, *Journal of the Royal Statistical Society, Seris B*, 59, 667-678.

Serfling R. S. (1980), *Approximation Theorems of Mathematical Statistics,* J. Wiley & Sons, New York.

Train K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Woolridge J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.