Journal of Classification
https://doi.org/10.1007/s00357-025-09529-y

ORIGINAL RESEARCH

Multilevel Latent Class with CUB Models

Nicola Piras[1] · Silvia Columbu[1] · Julien Jacques[2]

Received: 23 October 2024 / Accepted: 16 December 2025

Abstract
We propose a latent class model for ordinal data with CUB (combination of discrete uniform and shifted binomial) distributions in the case of multilevel structures of the data. The CUB model is a powerful approach to the analysis of ordinal data, where the elicitation process is thought to be governed by a feeling parameter and an uncertainty parameter. Ordinal data are common across different research fields and may present a multilevel structure with units nested within groups. The model we present extends the framework of multivariate CUB models for model-based clustering to multilevel data, either hierarchical or cross-classified. Numerical experiments on simulated data highlight the added value of assuming a CUB model to account for ordinal information; the procedure's interest is also shown through a real data application.

Keywords Multilevel latent class · Ordinal data · CUB

1 Introduction

Latent class (LC) analysis (McLachlan & Peel, 2000) is a useful model-based approach for data clustering of units on the basis of observations arising, usually, from a set of categorical indicators.

When the data have a multilevel structure with units nested within higher-level observations, the model must be adapted to account for this structure. Latent class models can be extended by defining multilevel finite mixture models that consider two levels of mixtures with separate discrete latent variables for lower-level and higher-level units. Lower-level units (individuals) belong to LCs at the lower level, and higher-level units (groups) belong to LCs at the higher level.

Examples of multilevel data can be found in different research fields. In social science, we can think of employees belonging to the same organizations, individuals living in the

✉ Nicola Piras
nicola.piras97@unica.it

Silvia Columbu
silvia.columbu@unica.it

Julien Jacques
julien.jacques@univ-lyon2.fr

1 Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy

2 Laboratoire ERIC, Université Lumière Lyon 2, Lyon, France

Published online: 06 January 2026

same regions, and customers of the same stores. In medical research, examples include data on patients nested within centers, respondents nested within regions, children nested within families, and repeated measures nested within subjects. Such structures are known as multilevel hierarchical. Sometimes, the units can be simultaneously nested within two or more higher levels — for example, patients nested within centers and medical departments, or customers nested within chain stores and regions — and the groupings are instead cross-classified between them. We refer to this structure as multilevel cross-classified (CC).

For hierarchical data, a multilevel version of the latent class (MLC) model was proposed in Vermunt (2003) and Vermunt (2008). The extension to the cross-classified version (MLC-CLC) has recently been proposed in Columbu et al. (2025).

The MLC and MLCCLC model formulations allow for not only clustering of lower-level units but also of higher-level units. In the case of employees nested into organizations, the model provides a simultaneous clustering for the organizations (higher-level units) and a clustering for the employees (lower-level units) within the different classes of organizations.

The parameters of the latent class model can be estimated by maximum likelihood using the expectation maximization (EM) algorithm (e.g., Dempster et al., 1977). The structure of the MLC model requires modifying the E step of the EM algorithm to account for membership variables at the various levels (see Vermunt, 2003). The introduction of this modification does not affect the usual convergence properties of EM. When the structure is cross-classified, the log-likelihood becomes untractable, and the application of usual EM algorithms is unfeasible; therefore, in Columbu et al. (2025) it has been proposed to apply instead a stochastic variation of the EM.

Typically, latent class analysis is performed when the observations are categorical. One particular type of categorical data is ordinal data (Agresti, 1990), which occur when the categories are ordered. Ordinal data take values from a set of ordered categories, coded as integer numbers, that express evaluations of a group of raters towards a well-defined list of items. They are widely used to express the subject's perceived level of perception, appreciation, or feeling. Such data are very frequent in different fields: in medicine and psychology with questionnaires to evaluate services or opinions on an ordinal scale (Selosse et al., 2019); in social science when assessing the level of educational attainment (i.e., high school graduation, bachelor's degree, master's degree, and PhD). In all these situations, there is a natural order in the information. Sometimes ordered data arises from the categorization of continuous variables into classes with ordered labels; for instance, this happens in marketing and finance (Amato et al., 2024) when evaluating products or different advertising. Usual approaches for analyzing this kind of data are based on generalized linear models (GLMs) and their extensions (e.g., McCullagh, 1980; McCullagh & Nelder, 1989; Ananth & Kleinbaum, 1997). A possible alternative is that of the CUB (Combination of Uniform and Binomial) model, introduced in D'Elia and Piccolo (2005) and Piccolo (2006), defined as the convex combination of a uniform distribution, to take into account the uncertainty in the choice of the rating, and a shifted binomial one to model the propensity towards the ordered categories. This paradigm is an alternative to that of cumulative models (Piccolo et al., 2019).

In the model-based clustering framework, less attention has been devoted to the case of ordinal data compared to that of nominal ones. Often, the ordered nature is ignored, and data are transformed into quantitative or nominal. However, both choices, i.e., introducing an artificial and subjective distance between categories or omitting the order between categories, introduce bias and should be preferably avoided. The first contributions in which the ordinal information was integrated in the latent class frame considered log-linear formulations with the introduction of covariates, where ordinal regression models are used to link the classes to the responses (Heinen, 1996; Formann, 1992; Oberski & Vermunt, 2015). Other contributions have proposed the introduction of specific models; for instance, in Biernacki and Jacques (2016), the authors proposed relying on a distribution tailored to the ordered nature of the data, defining the binary ordinal search (BOS) model. In this model, the data-generating process is assumed to be a stochastic binary search algorithm in a sorted table, so the resulting distribution is governed by two meaningful parameters (position and precision). Another alternative approach is that of McParland and Gormley (2016), which assumes that an ordinal variable is the discretization of a latent continuous variable. Finally, Grilli et al. (2014) proposed the use of a CUB model to handle multimodality with ordinal data. This has also recently been extended (see Ventura et al., 2024) to multivariate data in the context of model-based clustering.

In this paper, we define a multilevel version of the latent class CUB model for model-based clustering of multivariate data. The hierarchical structure for CUB models has been previously investigated in Iannario (2012a) within the framework of multilevel regression, introducing continuous random effects for the groupings. This is the first time the CUB distribution has been proposed for multilevel latent class analysis in hierarchical and cross-classified scenarios.

The paper is structured as follows. Section 2 presents the multilevel latent class CUB model (MLC-CUB) distinguishing the case of hierarchical and cross-classified data; Section 3 presents the inferential procedures adopted; in Section 4 some simulation results are presented to asses the efficiency of parameter estimation and the classification accuracy; Section 5 considers an application to a dataset on evaluation in Italian university, where the MLC-CUB model is applied in the hierarchical frame to simultaneously cluster students within faculties and faculties; Section 6 concludes the paper with final considerations and observations.

2 Multilevel Mixture Models with CUB Distribution

Hierarchical Version The MLC model (see Vermunt, 2003 and Vermunt, 2008) is defined by extending the standard LC to account for the additional nesting level. In particular, the nesting of units implies that each subject j belongs to one of the K groups, indexed as k with $1 \leq k \leq K$. The number of lower-level units within the higher-level unit k is denoted by n_k. The response of the lower-level unit j within the higher-level unit k in the indicator i is denoted with y_{ijk}, $\mathbf{y}_{jk}$ will be used to refer to the full vector of responses of case j in group k, while $\mathbf{y}_k$ refers to the full vector of responses for group k.

The membership variable at the lower level is denoted by X_{jk}, and a particular class by ℓ ($\ell = 1, \ldots, L$). The latent component at the higher level is denoted by W_k, and a particular component by h ($h = 1, \ldots, H$). In what follows, we will denote $\pi_h = P(W_k = h)$ and $\pi_{\ell|h} = P(X_{jk} = \ell | W_k = h)$.

The hierarchical mixture model consists of two parts:

1) The first part connects the observations belonging to the same group, and in the complete data form can be expressed as:

$$f(\mathbf{y}_k, W_k = h) = \pi_h f(\mathbf{y}_k | W_k = h) = \pi_h \prod_{j=1}^{n_k} f(\mathbf{y}_{jk} | W_k = h),$$

where groups are assumed to belong to one of H latent classes with prior probabilities equal to π_h and observations within a group are assumed to be mutually independent;

2) The second part is similar to the structure of a standard finite mixture model conditionally to the group classes:

$$f(\mathbf{y}_{jk}|W_k = h) = \sum_{\ell=1}^{L} \pi_{\ell|h} f(\mathbf{y}_{jk}|X_{jk} = \ell) = \sum_{\ell=1}^{L} \pi_{\ell|h} \prod_{i=1}^{I} f(y_{ijk}|X_{jk} = \ell)$$
$$= \sum_{\ell=1}^{L} \pi_{\ell|h} \prod_{i=1}^{I} f(y_{ijk}; \boldsymbol{\theta}_{i\ell}). \quad (1)$$

We have imposed the constraint for which parameters defining the conditional distributions for the response variables ($\boldsymbol{\theta}_{i\ell}$) are independent from the higher-level latent classes (W_k), i.e., $\boldsymbol{\theta}_{i\ell h} = \boldsymbol{\theta}_{i\ell}$. This constraint could eventually be relaxed. Model conditions are summarized in the graphical model in Fig. 1, where it is clear that there are no direct arrows from W_k to the observed variables and only lower-level latent variables X_{jk} depend directly on W_k. This structure, used in almost all applications aiming to simultaneously cluster higher and lower-level units, implies that the clustering of higher-level units is obtained from the information contained in multiple lower-level responses via their class memberships.

As written in the introduction, ordinal data are categorical data in which the categories have an order. They take values into a set of ordered categories that may express, for instance, the evaluation or feeling that subjects have about an item in a survey. In the case of i items, each subject expresses their preferences in m_i ordered categories, coded as $\{1, 2, \ldots, m_i\}$. The combination of uniform and shifted binomial variables (CUB) (see D'Elia & Piccolo, 2005) is a suitable model for these data; they assume that the elicitation process is the mixture of two distinct components: a feeling on the item and the uncertainty in the choice of the rating. Thinking of a preferences' order as the result of a paired comparison criterion (D'Elia, 2000), the feeling can be modeled assuming a shifted binomial (SB) distribution with support $\{1, 2, \ldots m_i\}$ and parameter $\xi_{i|\ell}$ for each item i:

$$P_{SB}(y_{ijk}|X_{jk} = \ell) = \binom{m_i - 1}{y_{ijk} - 1} (1 - \xi_{i|\ell})^{y_{ijk}-1} \xi_{i|\ell}^{m_i - y_{ijk}},$$

where $y_{ijk} = \{1, 2, \ldots, m_i\}$ are the observations for individual j and variable i.

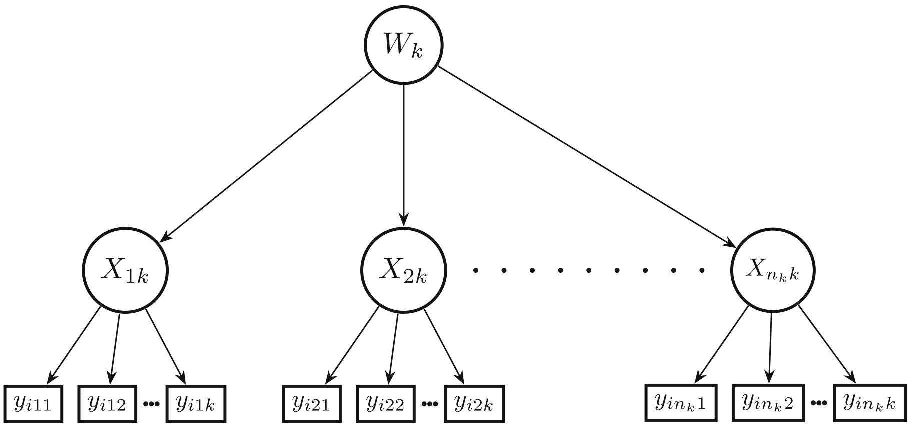

Fig. 1 Graphical model for MLC

The corresponding mean and variance parameters assume the expressions

$$E(SB|X_{jk}=\ell)=\xi_{i|\ell}+m_i(1-\xi_{i|\ell}), \qquad \mathrm{Var}(SB|X_{jk}=\ell)=(m_i-1)\xi_{i|\ell}(1-\xi_{i|\ell}). \quad (2)$$

We note that Var(SB) is maximized when $\xi_{i|\ell}=1/2$, i.e., there is the greatest uncertainty in assigning a rating to the item.

When the subject shows a complete indifference (or equi-preference) towards a given item, it seems appropriate to model ratings by means of a discrete uniform random variable (U):

$$P_U(y_{ijk})=\frac{1}{m_i}.$$

To model the elicitation process for ordinal data as the result of two components –feeling and uncertainty –the CUB model is defined as a mixture of a shifted binomial and a discrete uniform random variable. The two components of the CUB distribution have weights that depend on $\pi_{i|\ell}$ and $(1-\pi_{i|\ell})$, respectively. The mixing $(1-\pi_{i|\ell})$ associated with the uniform variable measures how the uncertainty affects the elicitation mechanism. In case of indifference, it seems appropriate to assume that the item i has the same probability of receiving any possible rating in $\{1, 2, \ldots m_i\}$, therefore the CUB distribution assumes the form:

$$\begin{aligned} f(y_{ijk}|X_{jk}=\ell) &= \pi_{i|\ell}P_{SB|X_{jk}=\ell}(y_{ijk})+(1-\pi_{i|\ell})P_U(y_{ijk}) \\ &= \pi_{i|\ell}\binom{m_i-1}{y_{ijk}-1}(1-\xi_{i|\ell})^{y_{ijk}-1}\xi_{i|\ell}^{m_i-y_{ijk}}+(1-\pi_{i|\ell})\frac{1}{m_i}. \end{aligned} \quad (3)$$

If $\pi_{i|\ell}\longrightarrow 0$ then the CUB distribution tends to behave as a uniform distribution, and the rating assigned to a given item depends only upon the numbers m_i of possible choices, that is, the case of total uncertainty, or equi-preference feeling. On the other hand, if $\pi_{i|\ell}\longrightarrow 1$ then the CUB distribution tends to behave as a shifted binomial distribution, and its features depend only upon the parameter $\xi_{i|\ell}$. The vector of distributional parameters is then $\boldsymbol{\theta}_{i\ell}=\{\pi_{i|\ell},\xi_{i|\ell}\}$.

Because the CUB distribution we use in the multilevel mixture model is itself a mixture of two components with weights $\pi_{i|\ell}$ and $1-\pi_{i|\ell}$, we need to introduce another latent variable Z_{ijk} where $Z_{ijk}=1$ if the preference of the jth rater for the ith item comes from the shifted binomial of parameter $\xi_{i|\ell}$, and $Z_{ijk}=0$ if it comes from the uniform random variable.

Cross-Classified Version A natural extension of the MLC for hierarchical data is the multilevel cross-classified LC (MCCLC), in which, at the higher level of the structure, observations are double-grouped. In this version, modifications arise at the upper level with the introduction of an extra set of membership variables. In the cross-classified version, the jth unit belongs to a unique combination of two second-level groups K ($k=1,\ldots,K$) and Q ($q=1,\ldots,Q$), within each k, q group, there are n_{kq} observations. The response observed on indicator i will then be denoted as y_{ijkq}. Together with X_{jkq} and W_k, we introduce a third latent variable V_q to which will correspond r latent classes. The new model structure is visually described in Fig. 2. The vector of the whole mixing proportions to be estimated is $\{\pi_{\ell|hr}=P(X_{jkq}=\ell|W_k=h,V_q=r), \pi_h=P(W_k=h), \pi_r=P(V_q=r)$. The model formulation is quite similar to the hierarchical one, with two sets of equations to handle the

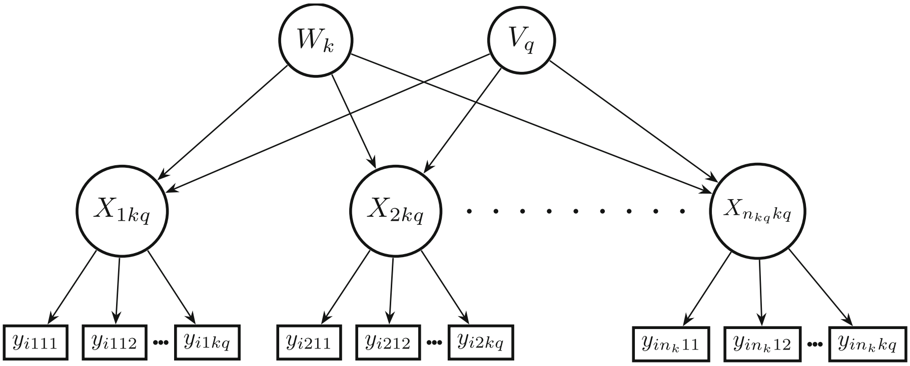

Fig. 2 Graphical model for MCCLC

two levels of the structure. In particular, the first part is modified to take into account the double grouping within higher-level latent classes as:

$$f(\mathbf{y}_{kq}, W_k = h, V_q = r) = \pi_{hr} f(\mathbf{y}_{kq}|W_k = h, V_q = r)$$
$$= \pi_h \pi_r \prod_{j=1}^{n_{kq}} f(\mathbf{y}_{jkq}|W_k = h, V_q = r).$$

We stress that the higher-level latent variables are marginally independent. The second part of the model is almost analogous to (1) within the combination of higher-level membership variables:

$$f(\mathbf{y}_{jkq}|W_k = h, V_q = r) = \sum_{\ell=1}^{L} \pi_{\ell|hr} f(\mathbf{y}_{jkq}|X_{jkq} = \ell) = \sum_{\ell=1}^{L} \pi_{\ell|hr} \prod_{i=1}^{I} f(y_{ijkq}|X_{jkq} = \ell)$$
$$= \sum_{\ell=1}^{L} \pi_{\ell|hr} \prod_{i=1}^{I} f(y_{ijkq}; \boldsymbol{\theta}_{i\ell}).$$

The observations' model is almost identical except that each jth unit is considered within the combination of k and q groups.

3 Model Inference

Estimating model parameters requires maximizing the likelihood function. In particular, for the hierarchical LC model, if $\boldsymbol{\Psi} = \{\pi_h, \pi_{\ell|h}, \boldsymbol{\theta}_{i\ell}\}$ denotes the full vector of parameters, the likelihood function associated takes the form:

$$\mathcal{L}(\boldsymbol{\Psi}; \mathbf{y}) = \sum_{k=1}^{K} f(\mathbf{y}_k) = \sum_{k=1}^{K} \sum_{h=1}^{H} \pi_h \prod_{j=1}^{n_k} \left[\sum_{\ell=1}^{L} \pi_{\ell|h} f(\mathbf{y}_{jk}|X_{jk} = \ell) \right]. \quad (4)$$

By taking X_{jk}, W_k, and Z_{ijk} as unobserved variables, the maximization can be performed, as it is usual for mixture models, through the EM algorithm. The application of the EM algorithm requires the evaluation of the expectation of the complete data log-likelihood, where **x**, **w**, and **z** denote the current values assumed by latent classes in the complete log-likelihood

$$
\begin{aligned}
E(\log \mathcal{L}_c(\boldsymbol{\Psi}; \mathbf{y}, \mathbf{x}, \mathbf{w}, \mathbf{z})) = & \sum_{k=1}^{K} \sum_{h=1}^{H} P(W_k = h | \mathbf{y}_k) \log \pi_h \\
& + \sum_{k=1}^{K} \sum_{h=1}^{H} \sum_{j=1}^{n_k} \sum_{\ell=1}^{L} P(W_k = h, X_{jk} = \ell | \mathbf{y}_{jk}) \log \pi_{\ell|h} \\
& + \sum_{k=1}^{K} \sum_{h=1}^{H} \sum_{j=1}^{n_k} \sum_{l=1}^{L} P(W_k = h, X_{jk} = \ell | \mathbf{y}_{jk}) \\
& \left[\sum_{i=1}^{I} P(Z_{ijk} = 1 | \mathbf{y}_{jk}, X_{jk} = \ell)(\log \pi_{i|\ell} + \log P_{SB}(y_{ijk}, | X_{jk} = \ell)) \right. \\
& \left. + \sum_{i=1}^{I} P(Z_{ijk} = 0 | \mathbf{y}_{jk}, X_{jk} = \ell)(\log(1 - \pi_{i|\ell}) + \log P_U(y_{ijk})) \right]. \quad (5)
\end{aligned}
$$

The E and M steps are iterated until convergence of the log-likelihood, checked by means of a chosen tolerance. The details of the EM steps as proposed in Vermunt (2003) are reported in Appendix 1.

In the case of a cross-classified structure, the likelihood to be maximized gets more complicated. In fact, the presence of a double missing data structure at the higher level, with W_k and V_q unobserved, makes it impossible to express the likelihood as a product of the mixing probabilities. It is required a marginalization over all possible label configurations, and the resulting likelihood takes the form:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Psi}; \mathbf{y}) &= \sum_{h_1=1}^{H} \sum_{h_2=1}^{H} \cdots \sum_{h_K=1}^{H} \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_Q=1}^{R} \prod_{k=1}^{K} \prod_{q=1}^{Q} f(\mathbf{y}_{kq}) \\
&= \sum_{h_1=1}^{H} \sum_{h_2=1}^{H} \cdots \sum_{h_K=1}^{H} \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_Q=1}^{R} \prod_{k=1}^{K} \pi_{h_k} \prod_{q=1}^{Q} \pi_{r_q} \prod_{j=1}^{n_{kq}} \left[\sum_{\ell=1}^{L} \pi_{\ell|h_k, r_q} f(\mathbf{y}_{jkq} | X_{jkq} = \ell) \right]; \quad (6)
\end{aligned}
$$

here the vector of model parameters is given as $\boldsymbol{\Psi} = \{\pi_h, \pi_r, \pi_{\ell|hr}, \boldsymbol{\theta}_{i\ell}\}$. The model's likelihood is intractable, and standard versions of the EM cannot be applied. A possible alternative to handle the complex structure is to consider stochastic variants of the EM (Celeux et al., 1996; Nielsen, 2000). In particular, we consider a solution in which a Gibbs sampler is included between the E and the M step, consisting of consecutive sampling from the marginal posterior distributions of higher and lower-level latent variables in the expected log-likelihood. The algorithm applied returns a Markov chain whose stationary distribution is concentrated around the maximum likelihood. The expected log-likelihood required for the implementation of the estimation algorithm takes the form:

$$
\begin{aligned}
E(\log \mathcal{L}_c(\boldsymbol{\Psi}; \mathbf{y}, \mathbf{x}, \mathbf{w}, \mathbf{v}, \mathbf{z})) = & \sum_{k=1}^{K}\sum_{h=1}^{H} P(W_k = h|\mathbf{y}_{kq}) \log \pi_h + \sum_{q=1}^{Q}\sum_{r=1}^{R} P(V_q = r|\mathbf{y}_{kq}) \log \pi_r \\
& + \sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{h=1}^{H}\sum_{r=1}^{R}\sum_{\ell=1}^{L}\sum_{j=1}^{n_{kq}} P(W_k = h, V_q = r, X_{jkq} = \ell|\mathbf{y}_{kq}) \log \pi_{\ell|hr} \\
& + \sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{h=1}^{H}\sum_{r=1}^{R}\sum_{\ell=1}^{L}\sum_{j=1}^{n_{kq}} P(W_k = h, V_q = r, X_{jkq} = \ell|\mathbf{y}_{kq}) \\
& \Bigg[\sum_{i=1}^{I} P(Z_{ijkq} = 1|\mathbf{y}_{jkq}, X_{jkq} = \ell)(\log \pi_{i|\ell} + \log P_{SB}(y_{ijkq}, |X_{jkq} = \ell)) \\
& + \sum_{i=1}^{I} P(Z_{ijkq} = 0|\mathbf{y}_{jkq}, X_{jkq} = \ell)(\log(1 - \pi_{i|\ell}) + \log P_U(y_{ijkq}))\Bigg]. \quad (7)
\end{aligned}
$$

The stochastic EM defined in Columbu et al. (2025) is adapted to deal with the CUB distribution as shown in Appendix 2.

Initialization Strategy It is known that EM solutions can depend on starting points, leading in some cases to convergence to local maxima Biernacki et al. (2003). To avoid local convergence, a multiple starting strategy is adopted. In particular, the level-2 mixing proportions are randomly initialized to $\left(\frac{1}{H}, \ldots, \frac{1}{H}\right)$, and also to $\left(\frac{1}{R}, \ldots, \frac{1}{R}\right)$ in the CC version, the level-1 conditional mixing proportions $\pi_{\ell|h}$ are initialized to $(\frac{1}{L}, \ldots, \frac{1}{L})$ for all $h = 1, \ldots, H$, the distribution parameters $\{\pi_{i|\ell}, \xi_{i|\ell}\}$ are initialized to the estimates of standard LC model parameters after a multiple starting point selection.

The choice of multiple random initializations, as illustrated in the simulation studies, is an efficient strategy, and the computing time is reasonable (using 10 different initializations with a tolerance of 10^{-8}, simulation 1 took about 8 min, and simulation 2 approximately 5 min). This is due to the fact that the likelihood (4) is not particularly complex, and the computations at the M step are explicit. In the cross-classified version given the more complex likelihood and the stochastic estimation approach, the computing time was higher, although still reasonable (using 16 different initializations with a tolerance of 10^{-8} took approximately 20 min), numerical experiments either in the case of CUB models, that for nominal data (Columbu et al., 2025) confirmed the efficacy of the method adopted. Alternative strategies can also be considered, such as initialization with a CEM or SEM algorithm, short EM runs (Biernacki et al., 2003), data-driven initialization (Iannario, 2012b), or a combination of different initialization strategies (Baudry & Celeux, 2015).

Identifiability In the standard LC model, as shown in Vermunt (2005) and Bennink et al. (2016), it is known that a sufficient condition for local identification is that all the eigenvalues of the information matrix are larger than zero. In general, for multilevel models, both parts of the structure must be identified, which means that there should be a sufficient number of variables y_{ijk} for lower-level identification (at least 3), and a sufficient number of units within groups. In the case of a CC structure, for which the likelihood cannot be computed in closed form, the assessment based on the information matrix is not possible; however, the same reasoning in terms of a sufficient number of observations at both levels of the structure can be applied.

4 Simulation Studies

The estimation performance of the EM algorithms for estimating MLC and MCCLC parameters under the CUB model for ordinal data has been explored through three simulation studies with different data-generating conditions. The first two scenarios assume a hierarchical structure, whereas the third considers a cross-classified structure. Model performances are illustrated in terms of efficiency of parameter estimation, through parameter estimates distribution and average absolute estimate errors, but also in terms of classification accuracy, for which we will compare our proposed model applied to the two multilevel structures with the analogous in case of the assumption of a multinomial model, in which the ordinal nature is disregarded. In the simulation studies presented, the number of latent classes L, H, and R is assumed to be known. The C++ code, together with the R code to generate data and perform simulations, is made available on a GitHub repository at the link https://github.com/NicolaPiras97/MLC_CUB. We observe that the CUB model for the hierarchical structure can be estimated using the LatentGOLD software (Vermunt & Magidson, 2016a). Indeed, following Oberski and Vermunt (2015), the CUB model can be reformulated as a special case of the log-linear latent class model, with the appropriate parameter choices.

4.1 Multilevel Hierarchical Case

For the hierarchical LC, we have designed two experiments under different conditions of data separation at the higher level. Data separation is measured through the entropy-based R-squared, a measure whose values range between 0 and 1, indicating how well one can predict the class memberships based on the observed responses (see Vermunt & Magidson, 2016b). We can compute this measure separately for lower and higher levels of the hierarchy:

$$R^2_{\text{entropy,low}} = 1 - \frac{\displaystyle\sum_{j=1}^{n}\sum_{\ell=1}^{L} -P(X_{jk} = \ell|\mathbf{y}_{jk}) \log(P(X_{jk} = \ell|\mathbf{y}_{jk}))}{\displaystyle n\sum_{\ell=1}^{L} -P(X_{jk} = \ell) \log(P(X_{jk} = \ell))},$$

$$R^2_{\text{entropy,high}} = 1 - \frac{\displaystyle\sum_{k=1}^{K}\sum_{h=1}^{H} -P(W_k = h|\mathbf{y}_k) \log(P(W_k = h|\mathbf{y}_k))}{\displaystyle K\sum_{h=1}^{H} -P(W_k = h) \log(P(W_k = h))}.$$

One hundred datasets were generated for each simulation. Both simulation settings take 5 ordinal variables, each with 7 categories. The number of groups is set to K = 50. We fix $L = 3$ as the number of lower-level latent classes and $H = 2$ as the number of higher-level classes. The mixing proportions for the higher level are $\pi_h = (0.4, 0.6)$, and the mixing for the lower level given higher-level class memberships are

$$\pi_{\ell|h} = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}.$$

The CUB parameters are

$$\pi_{i|\ell} = \begin{pmatrix} 0.7 & 0.6 & 0.8 \\ 0.6 & 0.8 & 0.7 \\ 0.9 & 0.6 & 0.8 \\ 0.5 & 0.7 & 0.6 \\ 0.8 & 0.6 & 0.9 \end{pmatrix}, \quad \xi_{i|\ell} = \begin{pmatrix} 0.7 & 0.6 & 0.2 \\ 0.6 & 0.8 & 0.4 \\ 0.1 & 0.3 & 0.8 \\ 0.3 & 0.4 & 0.6 \\ 0.2 & 0.1 & 0.9 \end{pmatrix}.$$

The level of separation was defined by varying the number of units n_k within each group k. As shown in Lukočiene and Vermunt (2009), the sample size (n_k) and the degree of separation (measured in terms of R^2 entropy) change together (letting the remaining parameters fixed). Having a small/high sample size is the main cause of poor/good separation; therefore, in the two scenarios considered, different sample sizes are used to ensure varying degrees of separation in the simulated datasets.

4.1.1 Results on Parameter Estimation

Simulation 1 The first setting assumes $n_k = 50$ units per group, for a total of $n = 2500$ units. This choice produces, averaged over the 100 simulated datasets, a moderate separation at level-1, $R^2_{\text{entropy,low}} = 0.61$, and almost perfect separation at level-2, $R^2_{\text{entropy,high}} = 0.98$.

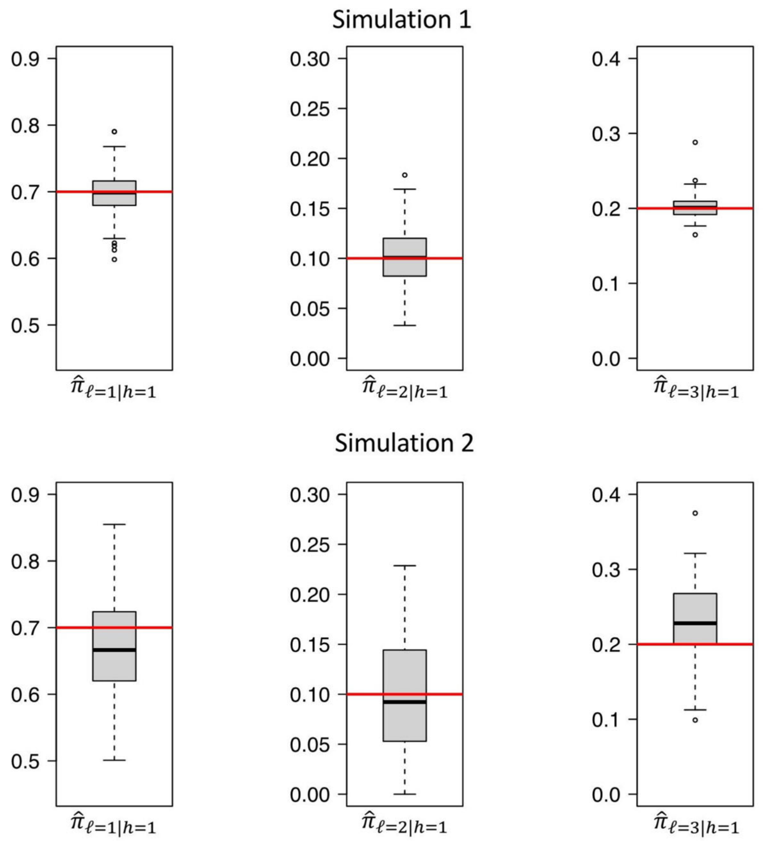

Fig. 3 Estimate distribution of level-1 conditional mixing proportions $\hat{\pi}_{\ell|h=1}$ in both simulation settings. For each boxplot, the true value is highlighted in red as a reference

Figures 3 and 4 report the distribution of estimated conditional mixing proportions of level-1 $\hat{\pi}_{\ell|h=1}$ and $\hat{\pi}_{\ell|h=2}$, respectively, over the 100 datasets, they provide evidence of accuracy in the estimation. Good performances have also been observed when looking at CUB model parameters and high-level mixing proportions; see Table 1, which reports the average error with respect to the true value.

Simulation 2 In the second setting, we set $n_k = 10$ units per group, with $n = 500$ total units. Lowering the number of units within each group reduced separation at level 2. In fact, on average across the 100 simulated datasets, we have $R^2_{\text{entropy,low}} = 0.60$ (moderate separation) at level-1 and $R^2_{\text{entropy,high}} = 0.78$ (moderate separation) at level-2.

Estimate performances are summarized by means of average estimation errors in Table 1, where results are reported for both settings, and it emerges how lowering the degree of separation produces a moderate lowering of performances, which remain satisfactory. In Figs. 3 and 4, the same comparison is also reported in terms of the distribution of estimates around the true value.

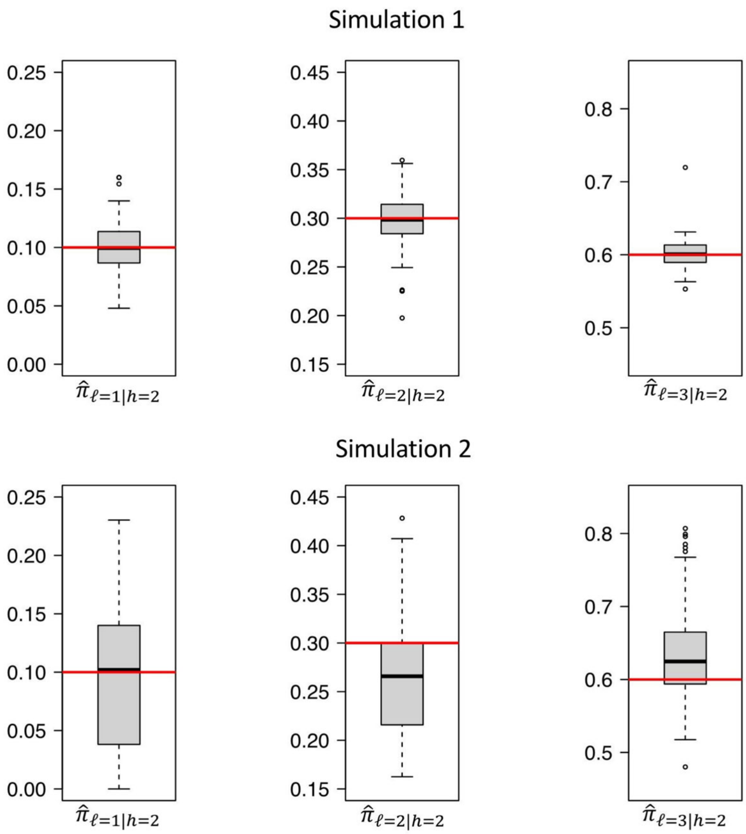

Fig. 4 Estimate distribution of level-1 conditional mixing proportions $\hat{\pi}_{\ell|h=2}$ in both simulation settings. For each boxplot, the true value is highlighted in red as a reference

Table 1 Average absolute errors (AErr) in the estimate of parameters in the two simulation settings. In (a), we report the errors in the estimate of level-2 mixing proportions $\hat{\pi}_h$. In (b), the errors in estimating level-1 conditional proportions $\hat{\pi}_{\ell|h}$. In (c), the errors associated with the estimate of CUB parameters $\hat{\pi}_{1|\ell}$ and $\hat{\xi}_{1|\ell}$

	Simulation 1			Simulation 2		
(a)	$h = 1$	$h = 2$		$h = 1$	$h = 2$	
AErr($\hat{\pi}_h$)	0.00049	0.000485		0.0022	0.002196	
(b)	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 1$	$\ell = 2$	$\ell = 3$
AErr($\hat{\pi}_{\ell\|h=1}$)	0.003	0.001879	0.001121	0.05773	0.00352	0.061248
AErr($\hat{\pi}_{\ell\|h=2}$)	0.0000014	0.00151	0.001507	0.00647	0.04379	0.050259
(c)	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 1$	$\ell = 2$	$\ell = 3$
AErr($\hat{\pi}_{1\|\ell}$)	0.008133	0.00637	0.005242	0.03916	0.03599	0.023614
AErr($\hat{\xi}_{1\|\ell}$)	0.000638	0.003509	0.001384	0.002419	0.01539	0.001600

In agreement with the literature on MLC, we did not encounter, in our experiment, any label switching problem in the estimation process. If a label switching problem appears, a usual solution can be considered (see, for instance, Stephens (2000)).

4.1.2 Results on Classification Accuracy

Experimental results proved that the model proposed is satisfactory also in terms of well-classified lower-level units within the two higher-level classes, as it appears in Table 2. When looking at the higher level, the classification was always correct in agreement with the setting of almost perfect separation in scenario 1, while in scenario 2, we have observed that the proportion of correctly classified units of level 2 is on average 0.94.

The benefit of considering a distributional assumption designed for ordinal data rather than a standard multinomial may be proved in terms of the correct classification rate. With this aim, the multilevel LC model was also fitted, assuming a (misspecified) multinomial distribution, and the final classification was compared with that obtained when considering our proposal. Results are obtained for scenario 1 ($n_k = 50$), scenario 2 ($n_k = 10$), and a new generative scheme with $n_k = 20$ (all remaining parameters and conditions were unchanged). Table 3 shows the proportion of the well-classified lower-level units under the CUB model and the multinomial one over the 100 simulated datasets for each scenario. We observe that the CUB multilevel LC model yields better results in terms of final classification, which becomes evident in situations when lower separation is encountered (i.e., small sample size).

Table 2 Proportion of well-classified lower-level units in both simulations. The proportion is shown separately within each higher-level class ($h = 1$, $h = 2$) and ignoring the nesting (Tot.)

	$h = 1$	$h = 2$	Tot.
Simulation 1	0.799	0.787	0.795
Simulation 2	0.788	0.771	0.778

Table 3 Proportion of well-classified lower-level units over the 100 simulations. Standard deviations are shown in parenthesis

	$n_k = 50$	$n_k = 20$	$n_k = 10$
CUB	0.795 (0.053)	0.786 (0.118)	0.778 (0.165)
Categorical	0.798 (0.085)	0.733 (0.131)	0.681 (0.174)

In particular, when $n_k = 10$, the assumption of a CUB model guarantees an increase of about 10% in the rate of well-classified units.

4.2 Multilevel Cross-Classified Case

An additional experiment was conducted for a CC structure. In this scenario, 50 datasets were generated, taking 6 ordinal variables, three with 7 categories and three with 5 categories. We fix $L = 4$ as the number of lower-level latent classes and $H = 3$ and $R = 2$ as CC level classes. The mixing proportions set were $\pi_h = (0.20, 0.30, 0.50)$ and $\pi_r = (0.40, 0.60)$ and the sample sizes were $K = 50$ and $Q = 15$. At level-1, the total number of units was $n = 6000$, with $n_{kq} = 8$ per combined level-2 unit. The matrix of level-1 mixing proportions $\pi_{\ell|hr}$ was set to

$$\pi_{\ell|hr} = \begin{pmatrix} 0.40 & 0.15 & 0.15 & 0.30 & 0.30 & 0.05 \\ 0.30 & 0.10 & 0.35 & 0.10 & 0.35 & 0.25 \\ 0.20 & 0.35 & 0.20 & 0.40 & 0.15 & 0.40 \\ 0.10 & 0.40 & 0.30 & 0.20 & 0.20 & 0.30 \end{pmatrix}.$$

The CUB parameters are

$$\pi_{i|\ell} = \begin{pmatrix} 0.7 & 0.6 & 0.8 & 0.9 \\ 0.6 & 0.9 & 0.7 & 0.8 \\ 0.9 & 0.6 & 0.8 & 0.7 \\ 0.5 & 0.7 & 0.6 & 0.8 \\ 0.8 & 0.6 & 0.9 & 0.7 \\ 0.7 & 0.6 & 0.8 & 0.9 \end{pmatrix}, \quad \xi_{i|\ell} = \begin{pmatrix} 0.7 & 0.6 & 0.2 & 0.3 \\ 0.6 & 0.8 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.8 & 0.7 \\ 0.3 & 0.4 & 0.6 & 0.8 \\ 0.2 & 0.9 & 0.3 & 0.7 \\ 0.3 & 0.8 & 0.4 & 0.6 \end{pmatrix}.$$

4.2.1 Results on Parameter Estimation

Simulation 3 These choices of parameters and number of units lead to a medium separation at level-1 $R^2_{\text{entropy,low}} \approx 0.64$. The separation at level-2 to around 0.80 in terms of total $R^2_{\text{entropy,CC}}$, with separate $R^2_{\text{entropy},\, W_k} \approx 0.62$, and $R^2_{\text{entropy},\, V_q} \approx 1$ (see Columbu et al. (2025) for the definition of entropy measure in the CC case).

Estimate performances are summarized by means of average estimates errors in Table 4. Good performances have also been observed when looking at CUB model parameters not reported in Tables 4.

With the implementation of stochastic algorithms, label switching may occur; in particular, Columbu et al. (2025) observed that this may occur at higher levels of the structure. To

Table 4 Average absolute errors (AErr) in the estimate of parameters in the simulation setting for multilevel cross-classified data. In (a), we report the errors in the estimate of first-level-2 CC mixing proportions $\hat{\pi}_h$. In (b), we report the errors in the estimate of second-level-2 CC mixing proportions $\hat{\pi}_r$. In (c), the errors in estimating level-1 conditional proportions $\hat{\pi}_{\ell|hr}$. In (d) the errors associated with the estimate of CUB parameters $\hat{\pi}_{1|\ell}$ and $\hat{\xi}_{1|\ell}$

	Simulation 3			
(a)	$h = 1$	$h = 2$	$h = 3$	
AErr($\hat{\pi}_h$)	0.0543	0.0179	0.0425	
(b)	$r = 1$	$r = 2$		
AErr($\hat{\pi}_r$)	0.0463	0.0463		
(c)	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
AErr($\hat{\pi}_{\ell\|h=1,r=1}$)	0.0245	0.0138	0.0287	0.0296
AErr($\hat{\pi}_{\ell\|h=1,r=2}$)	0.0242	0.0576	0.0109	0.0492
AErr($\hat{\pi}_{\ell\|h=2,r=1}$)	0.0235	0.0267	0.0388	0.0230
AErr($\hat{\pi}_{\ell\|h=2,r=2}$)	0.0208	0.0181	0.0287	0.0392
AErr($\hat{\pi}_{\ell\|h=3,r=1}$)	0.0451	0.0195	0.0344	0.0126
AErr($\hat{\pi}_{\ell\|h=3,r=2}$)	0.0285	0.0620	0.0175	0.0322
(d)	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
AErr($\hat{\pi}_{1\|\ell}$)	0.0792	0.105	0.0638	0.112
AErr($\hat{\xi}_{1\|\ell}$)	0.0347	0.0330	0.0565	0.0434

prevent such situations an ordering constraint ex post for the final estimation of level-2 mixing proportions was imposed in the implementation (Marin et al., 2005).

4.2.2 Results on Classification Accuracy

Experimental results showed that the proposed model is also satisfactory in terms of well-classified lower-level units within the two higher-level classes, as shown in Table 5. At the higher level, the classification was always correct for the $R = 2$ level-2 latent classes associated with the latent variable V, while the percentage of well-classified units across the $H = 3$ latent classes associated with the latent variable W was 87%.

The benefit of considering a distributional assumption designed for ordinal data using the CUB distribution rather than a standard multinomial is also proved in this scenario in terms of the correct classification rate. The MCCLC model was also fitted under a (misspecified) multinomial distribution, and the final classification was compared with that obtained using our proposal. Table 5 shows (Tot.) the proportion of the well-classified lower-level units (ignoring the higher-level class grouping) under the CUB model and the multinomial one over the 50 simulated datasets. We observe that the CUB MCCLC model yields better final

Table 5 Proportion of well-classified lower-level units in simulation 3. The proportion is shown separately within each combination of CC higher-level class. The proportion of total (Tot.) well-classified lower-level units under the CUB model, and the multinomial one is also shown. Standard deviations are shown in parenthesis

	$h = 1$	$h = 2$	$h = 3$		Tot.
$r = 1$	0.806	0.794	0.782	CUB	0.796 (0.135)
$r = 2$	0.797	0.789	0.778	Categorical	0.679 (0.156)

classification results; in particular, the assumption of a CUB model increases the rate of well-classified units by about 12%.

5 Real Data Application

The MLC-CUB model has been applied to analyze a dataset coming from a 2002 Italian survey on university evaluation. Data are available from the *R* package CUB (Iannario et al., 2024). The data consist of a sample survey of 2179 students' evaluations of the orientation services, conducted across the 13 faculties of the University of Naples Federico II in 2002. Participants were asked to express their ratings on a 7 point Likert ordinal scale (1 = "Extremely Unsatisfied (EU)," 2 = "Very Unsatisfied (VU)," 3 = "Unsatisfied (U)," 4 = "Indifferent (I)," 5 = "Satisfied (S)," 6 = "Very Satisfied (VS)," 7 = "Extremely Satisfied (ES)") for 5 ordinal variables:

1) Informat: level of satisfaction about the collected information;
2) Willingn: level of satisfaction about the willingness of the staff;
3) Officeho: judgment about the office hours;
4) Competen: judgement about the competence of the staff;
5) Global: global satisfaction.

The results we present are obtained with three latent classes at the lower level ($L = 3$) and two at the higher level ($H = 2$). We selected this number of classes according to a three-step procedure (see Lukočiene & Vermunt, 2009) in which maximum likelihood-based criteria, such as AIC and BIC, are considered across the levels of the data structure. The procedure accounts for the multiple choices at the two levels of the hierarchy and their mutual dependencies. Details on the procedure and on the definition of selection criteria at the higher level of the structure are described in Appendix 3. We had considered removing the Global variable, which represents the overall satisfaction index. However, since the final classification results remain unchanged whether this variable is included or not, we have chosen to present the analysis with it included.

Parameter estimates are reported in Tables 6 and 7. We observe that for the variables Willingn ($i = 2$) and Officeho ($i = 3$) the estimated $\hat{\pi}_{2|1}$ and $\hat{\pi}_{3|1}$ are smaller then 0.5. It means that the evaluation towards these items, for the first level-1 latent class ($\ell = 1$), has a great uncertainty.

Labels can be assigned to level-1 and level-2 classes based on the estimates obtained. In particular, we have assigned a labeling on the basis of the mean value patterns estimated for each item within each level-1 and level-2 latent class. Mean values $E_{i|\ell}$ for item i for each class ℓ were computed considering the CUB distribution (3) and the shifted binomial

Table 6 Class probabilities estimated at each level of the structure. Labels to classes have been assigned based on parameter estimates for the five indicators

	$\hat{\pi}_{\ell\|h}$			$\hat{\pi}_h$
	$\ell = 1$ (U)	$\ell = 2$ (ES)	$\ell = 3$ (S)	
$h = 1$ (S)	0.167	0.158	0.675	0.462
$h = 2$ (VS)	0.015	0.551	0.434	0.538

Table 7 CUB parameters estimated at each level-1 LC

	$\hat{\pi}_{i\|\ell}$			$\hat{\xi}_{i\|\ell}$		
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 1$	$\ell = 2$	$\ell = 3$
$i = 1$	0.849	0.997	0.995	0.789	0.057	0.293
$i = 2$	0.435	0.995	0.998	0.796	0.024	0.195
$i = 3$	0.410	0.846	0.902	0.822	0.062	0.319
$i = 4$	0.907	0.998	0.994	0.766	0.039	0.274
$i = 5$	0.996	0.999	0.999	0.702	0.040	0.258

expected value (2) as:

$$E_{i|\ell} = \hat{\pi}_{i|\ell}(\hat{\xi}_{i|\ell} + m_i(1 - \hat{\xi}_{i|\ell})) + (1 - \hat{\pi}_{i|\ell})\frac{m_i + 1}{2}.$$

Their values within level-1 classes, rounded to the nearest integer, along with the assigned labels, are reported in Table 8. In particular, for the level-1 classes, the labels assigned are "Unsatisfied (U)," "Satisfied (S)," and "Extremely Satisfied (ES)," whereas at the faculty level (level-2), the 2 labels are "Satisfied (S)" and "Very Satisfied (VS)."

Labels for the two higher-level classes are assigned on the basis of the aggregated expected value, computed as:

$$E_{i|h} = \sum_{\ell=1}^{L} E_{i|\ell}\pi_{\ell|h}.$$

A graphical synthesis of the results can be presented in a bivariate barplot showing the proportions of students classified within the faculties; see Fig. 5. The classification obtained is consistent with the labels derived from parameter estimates. At the end, what emerged is that there is a class of unsatisfied students, who reported an average score of 3 for each item in the satisfaction scale, a class of satisfied students, who were satisfied on each of the 5 aspects, and a last class of extremely satisfied ones, who expressed satisfaction on the majority of items. The classification results are consistent across all variables; students who express satisfaction with the services tend to report similar levels of satisfaction with other related dimensions. When moving up to the faculty level, both classes showed a relatively positive perception of orientation services, indicating that all institutions involved were able to implement good orientation practices. In particular, from Fig. 5, it can be seen that in both the "Satisfied" and "Very Satisfied" student classes at the faculty level, the percentage

Table 8 Estimated means for each item i in correspondence of level-1 classes ℓ and level-2 classes h

$E_{i\|\ell}$	$\ell = 1$ (U)	$\ell = 2$ (ES)	$\ell = 3$ (S)	$h = 1$ (S)	$h = 2$ (VS)
$i = 1$	3	7	5	5	6
$i = 2$	3	7	6	6	6
$i = 3$	3	6	5	5	6
$i = 4$	3	7	5	5	6
$i = 5$	3	7	5	5	6

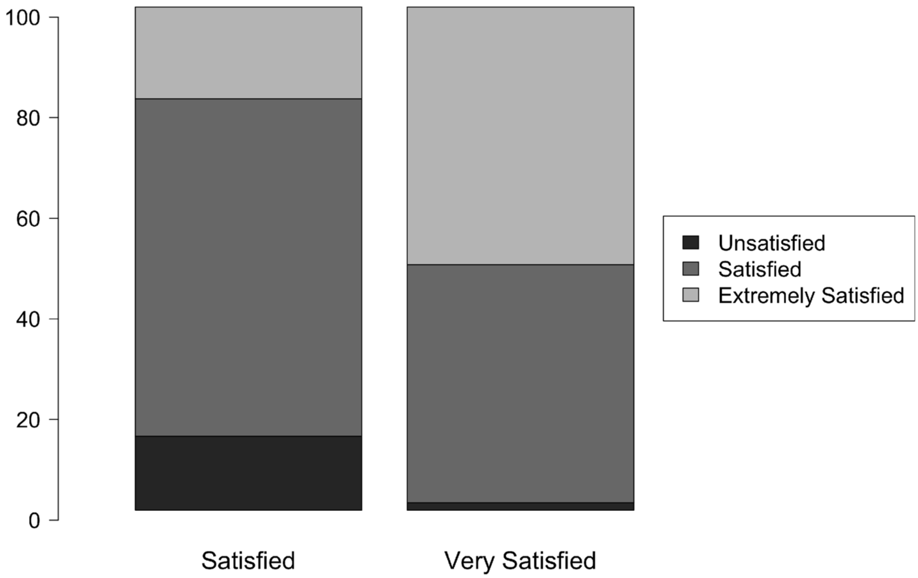

Fig. 5 Classification of students within faculties' clusters

of students classified as unsatisfied is minimal. Furthermore, in the class of "Satisfied," the predominant composition is students classified as "Satisfied," whereas in the faculty class of "Very Satisfied," the largest group is students in the level-1 class of "Extremely Satisfied," and the class of "Unsatisfied" is almost absent.

6 Conclusions

In this paper, we extended the application of the CUB models to handle multilevel latent classes for ordinal data. We apply the EM algorithm introduced in Vermunt (2003) for hierarchical data and the stochastic EM described in Columbu et al. (2025) for cross-classified data, taking into account that the CUB model we use to describe ordinal data requires an additional latent variable. Indeed, this model describes the data-generating process with a mixture model itself, where the two components used describe both feeling and uncertainty about the items. The algorithm was tested in simulation studies in which the generative schemes were defined to provide different levels of data separation at both low and high levels of the multilevel structure. An application to real data for the classification of faculties (group level) and students (individual level) from an Italian survey was also presented. The proposed approach allowed for managing the ordinal nature of the data with a straightforward but efficient model within the multilevel structure.

Future developments involve the inclusion of covariates (see Vermunt, 2010) and the definition of the model to handle mixed-type features, which could not be exclusively ordinal (see Vermunt & Magidson, 2002; Selosse et al., 2020).

Appendix 1: Parameter Estimation in the Case of Hierarchical Multilevel Latent Class

In this section, we report details on the implementation of the EM procedure adapted to the hierarchical structure. The EM algorithm in the case of MLC considers an E step that allows for reducing the computational complexity required to derive the joint conditional probability $P(W_k = h, X_{jk} = \ell|\mathbf{y}_k)$. This will consist of computing in sequence:

$$P(W_k = h|\mathbf{y}_k) = \frac{\pi_h \prod_{j=1}^{n_k} f(\mathbf{y}_{jk}|W_k = h)}{f(\mathbf{y}_k)};$$

$$P(X_{jk} = \ell|\mathbf{y}_{jk}, W_k = h) = \frac{\pi_{\ell|h} f(\mathbf{y}_{jk}|X_{jk} = \ell)}{f(\mathbf{y}_{jk}|W_k = h)};$$

$$\begin{aligned} P(W_k = h, X_{jk} = \ell|\mathbf{y}_k) &= P(W_k = h|\mathbf{y}_k)P(X_{jk} = \ell|\mathbf{y}_k, W_k = h) \\ &= P(W_k = h|\mathbf{y}_k)P(X_{jk} = \ell|\mathbf{y}_{jk}, W_k = h). \end{aligned}$$

The factorization of the joint probability $P(W_k = h, X_{jk} = \ell|\mathbf{y}_k)$ arises from the fact that, given the group's membership (W_k), the class membership of individuals (X_{jk}) is independent of the information about the other group members. This factorization allows the implementation of the E step, which in the standard EM would have required the computation of the joint conditional expectation of the $n_k + 1$ latent variables for higher-level unit k, with $K \times L^{n_k}$ entries.

We also need to compute the posteriors for the membership variables Z_{ijk} related to the mixing $\pi_{i|\ell}$ of the CUB distribution:

$$P(Z_{ijk} = 1|\mathbf{y}_{ik}, X_{jk} = \ell) = \frac{\pi_{i|\ell}\binom{m_i - 1}{y_{ijk} - 1}(1 - \xi_{i|\ell})^{y_{ijk}-1}\xi_{i|\ell}^{m_i - y_{ijk}}}{f(y_{ijk}|X_{jk} = \ell)}.$$

In the M step, the unknown model parameters are updated so that the expected complete data log-likelihood (5) is maximized. The M step updates all parameters $\boldsymbol{\Psi} = \{\pi_h, \pi_{\ell|h}, \boldsymbol{\theta}_{i\ell}\}$ as follows:

$$\pi_h = \frac{\sum_{j=1}^{n} P(W_k = h|\mathbf{y}_{jk})}{Q},$$

where $P(W_k = h|\mathbf{y}_{jk})$ is the expansion of $P(W_k = h|\mathbf{y}_k)$ over the first level units j.

$$\pi_{\ell|h} = \frac{\sum_{j=1}^{n} P(W_k = h, X_{jk} = \ell|\mathbf{y}_{jk})}{\sum_{j=1}^{n} P(W_k = h|\mathbf{y}_{jk})};$$

$$\pi_{i|\ell} = \frac{\sum_{j=1}^{n}\sum_{h=1}^{H} P(X_{jk} = \ell, W_k = h|\mathbf{y}_{jk}) P(Z_{ijk} = 1|\mathbf{y}_{ik}, X_{jk} = \ell)}{\sum_{j=1}^{n}\sum_{h=1}^{H} P(X_{jk} = \ell, W_k = h|\mathbf{y}_{jk})}.$$

The means for the shifted binomial variables are then computed as:

$$M_{i|\ell} = \frac{\sum_{j=1}^{n}\sum_{h=1}^{H} P(X_{jk} = \ell, W_k = h|\mathbf{y}_{jk}) P(Z_{ijk} = 1|\mathbf{y}_{ik}, X_{jk} = \ell) y_{ijk}}{\sum_{j=1}^{n}\sum_{h=1}^{H} P(X_{jk} = \ell, W_k = h|\mathbf{y}_{jk}) P(Z_{ijk} = 1|\mathbf{y}_{ik}, X_{jk} = \ell)},$$

so the feeling parameters $\xi_{i|\ell}$ are update by using (2) as:

$$\xi_{i|\ell} = \frac{m_i - M_{i|\ell}}{m_i - 1}.$$

Appendix 2: Parameter Estimation in the Case of Cross-Classified Multilevel Latent Class

The details on the stochastic EM adapted to the MCCLC-CUB model are described here. The estimation algorithm considers an extra stochastic step between the E and the M step, as described in the following.

E and S Step

After initialization of $\pi_h, \pi_r, \pi_{\ell|hr}, \pi_{i|\ell}, \xi_{i|\ell}$ run a Gibbs sampler

1) Draw $\boldsymbol{w}^{(t)}$ from a Multinomial distribution with probabilities

$$P(W_k = h|\boldsymbol{y}_k, \boldsymbol{v}^{(t-1)}) = \frac{\pi_h P(\boldsymbol{Y}_k = \boldsymbol{m}_k|\boldsymbol{v}^{(t-1)}, W_k = h)}{P(\boldsymbol{Y}_k = \boldsymbol{m}_k|\boldsymbol{v}^{(t-1)})},$$

$$P(\boldsymbol{Y}_k = \boldsymbol{m}_k|\boldsymbol{v}, W_k = h) = \prod_{q_k=1}^{Q_K}\prod_{r=1}^{R}\left[\prod_{j=1}^{n_{kq}} P(\boldsymbol{Y}_{jkq} = \boldsymbol{m}|W_k = h, V_q = r)\right]^{v_q^r}.$$

2) Draw $\boldsymbol{v}^{(t)}$ from a Multinomial distribution with probabilities

$$P(V_q = r|\boldsymbol{y}_q, \boldsymbol{w}^{(t)}) = \frac{\pi_r P(\mathbf{Y}_q = \boldsymbol{m}_q|\boldsymbol{w}^{(t)}, V_q = r)}{P(\boldsymbol{Y}_q = \boldsymbol{m}_q|\boldsymbol{w}^{(t)})},$$

$$P(\boldsymbol{Y}_q = \boldsymbol{m}_q|\boldsymbol{w}, V_q = r) = \prod_{k_q=1}^{K_Q}\prod_{h=1}^{H}\left[\prod_{j=1}^{n_{kq}} P(\mathbf{Y}_{jkq} = \boldsymbol{m}|W_k = h, V_q = r)\right]^{w_k^h}.$$

2) Draw $\boldsymbol{x}^{(t)}$ from a Multinomial distribution with probabilities

$$P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)}) = \frac{\left[\pi_{\ell|hr} P(\boldsymbol{Y}_{jkq} = \boldsymbol{g}|X_{jkq} = \ell)\right]^{w_{jk}^h v_{jq}^r}}{P(\boldsymbol{Y}_{jkq} = \boldsymbol{g})}.$$

3) Compute the posteriors for the membership variables Z_{ijkq} related to the mixing $\pi_{i|\ell}$ of the CUB distribution:

$$P(Z_{ijkq} = 1|\boldsymbol{y}_{ikq}, X_{jkq} = \ell) = \frac{\pi_{i|\ell}\binom{m_i - 1}{y_{ijkq} - 1}(1 - \xi_{i|\ell})^{y_{ijkq}-1}\xi_{i|\ell}^{m_i - y_{ijk}}}{f(y_{ijk}|X_{jkq} = \ell)}.$$

M Step

$$\pi_h = \frac{\sum_{k=1}^{K} w_k^{h(t)}}{K}, \quad \pi_r = \frac{\sum_{q=1}^{Q} v_q^{r(t)}}{Q},$$

$$\pi_{\ell|hr} = \frac{\sum_{j=1}^{n} w_{jk}^{h(t)} v_{jq}^{r(t)} P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)})}{\sum_{j=1}^{n} w_{jk}^{h(t)} v_{jq}^{r(t)}}.$$

The CUB parameters are update as follows:

$$\pi_{i|\ell} = \frac{\sum_{j=1}^{n} P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)}) P(Z_{ijkq} = 1|\boldsymbol{y}_{ikq}, X_{jkq} = \ell)}{\sum_{j=1}^{n} P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)})}.$$

The means for the shifted binomial variables are then computed as:

$$M_{i|\ell} = \frac{\sum_{j=1}^{n} P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)}) P(Z_{ijkq} = 1|\boldsymbol{y}_{ikq}, X_{jkq} = \ell) y_{ijkq}}{\sum_{j=1}^{n} P(X_{jkq} = \ell|\boldsymbol{y}_{jkq}, \boldsymbol{w}^{(t)}, \boldsymbol{v}^{(t)}) P(Z_{ijkq} = 1|\boldsymbol{y}_{ikq}, X_{jkq} = \ell)},$$

so the feeling parameters $\xi_{i|\ell}$ are update by using (2) as:

$$\xi_{i|\ell} = \frac{m_i - M_{i|\ell}}{m_i - 1}.$$

Appendix 3: Choice of the Number of Classes

Concerning the model selection strategy, we have selected the number of classes using the three-step procedure described in Lukočiene and Vermunt (2009). The computation of AIC and BIC is adapted to the multilevel data structure, and the number of higher-level latent classes (H) must also be included. The procedure steps are described below.

Step 1) Determine the number of lower-level classes as in a standard LC model, ignoring the multilevel structure. Level-1 information criteria used are defined as:

$$\text{BIC}_1 = -2\log\mathcal{L} + ((L - 1) + npar)\log(n),$$

$$\text{AIC}_1 = -2\log\mathcal{L} + 2((L - 1) + npar),$$

where $npar = 2L \times I$ is the number of free distribution parameters. The lower the value of an IC, the better the model.

Step 2) Determine the number of higher-level classes, taking fixed the number of level-1 latent classes. Level-2 information criteria used are defined as:

$$\mathrm{BIC}_2 = -2\log\mathcal{L} + ((H-1) + (H(L-1)) + npar)\log(K),$$

$$\mathrm{AIC}_2 = -2\log\mathcal{L} + 2((H-1) + (H(L-1)) + npar).$$

Step 3) Determine again the number of lower-level classes fixing the number of higher-level classes. Level-1 information criteria used are defined as:

$$\mathrm{BIC}_3 = -2\log\mathcal{L} + ((H-1) + (H(L-1)) + npar)\log(n),$$

$$\mathrm{AIC}_3 = -2\log\mathcal{L} + 2((H-1) + (H(L-1)) + npar).$$

We observe that in step 2), in agreement with simulation results discussed in Lukočiene et al. (2010), the computation of BIC takes as sample size the number of groups (K) instead of the total number of individuals (n).

Funding Nicola Piras and Silvia Columbu gratefully acknowledge the partial support to their research by project "GraphNet: models, computation and estimation in networks and graph analysis" funded by D.M. 737/2021 - Linea d'intervento Iniziative di ricerca interdisciplinare su temi di rilievo trasversale per il PNR (grant number F25F21002720001).
Silvia Columbu also acknowledges the partial support of the projects "FIATLUCS - Favorire l'Inclusione e l'Accessibilitá per i Trasferimenti Locali Urbani a Cagliari e Sobborghi" funded by the PNRR RAISE Liguria, Spoke 01, (grant number F23C24000240006) and "MAPS" funded by Fondazione di Sardegna (grant number F73C23001550007).

Data Availability The data used in the application is available at the website: https://cran.r-project.org/package=CUB.

Declarations

Ethical Approval This research does not contain any studies with human participations or animals performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Amato, F., Jacques, J., & Prim-Allaz, I. (2024). Clustering longitudinal ordinal data via finite mixture of matrix-variate distributions. *Statistics and Computing, 34*, 81.

Ananth, C. V., & Kleinbaum, D. (1997). Regression models for ordinal responses: A review of methods and applications. *International journal of Epidemiology, 26*, 1323–1332.

Baudry, J. P., & Celeux, G. (2015). Em for mixtures. *Statistics and Computing, 25*(4), 713–726.

Bennink, M., Croon, M. A., Kroon, B., & Vermunt, J. K. (2016). Micro-macro multilevel latent class models with multiple discrete individual-level variables. *Advances in Data Analysis and Classification, 10*, 139–154.

Biernacki, C., Celeux, G., & Govaert, G. (2003). Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis, 41*(3–4), 561–575.

Biernacki, C., & Jacques, J. (2016). Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm. *Statistics and Computing, 26*(5), 929–943.

Celeux, G., Chauveau, D., & Diebolt, J. (1996). Stochastic versions of the em algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation, 55*(4), 287–314.

Columbu, S., Piras, N., & Vermunt, J. K. (2025). Multilevel latent class models for cross-classified categorical data: Model definition and estimation through stochastic em. *Statistics and Computing, 35*, 50.

D'Elia, A. (2000). A shifted binomial model for rankings. In Nunez-Antón, V. and Ferreira, E., editors, *Proceedings of the 15th International Workshop on Statistical Modelling*, pages 412–416, Bilbao.

D'Elia, A., & Piccolo, D. (2005). A mixture model for preference data analysis. *Computational Statistics and Data Analysis, 49*(3), 917–934.

Dempster, A. P., Laird, N. M., & Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Formann, A. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476–486.

Grilli, L., Iannario, M., Piccolo, D., & Rampichini, C. (2014). Latent class cub models. *Advances in Data analysis and Classification, 8*, 787–800.

Heinen, T. (1996). Latent class and discrete latent trait models: Similarities and differences. *Journal of Educational Measurement, 35*(2), 175–178.

Iannario, M. (2012). Hierarchical cub models for ordinal variables. *Communications in Statistics - Theory and Methods, 41*(16–17), 3110–3125.

Iannario, M. (2012). Preliminary estimators for a mixture model of ordinal data. *Advances in Data analysis and Classification, 6*, 163–184.

Iannario, M., Piccolo, D., & Simone, R. (2024). CUB: A class of mixture models for ordinal data. *R package version, 1*(1), 5.

Lukočiene, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology, 40*(1), 247–283.

Lukočiene, O., & Vermunt, J. (2009). Determining the number of components in mixture models for hierarchical data. In Fink, A., Lausen, B., Seidel, W., and Ultsch, A., editors, *Advances in Data Analysis, Data Handling and Business Intelligence. Studies in Classification, Data Analysis, and Knowledge Organization.*, Berlin.

Marin, J. M., Mengersen, K., & Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. In I. D. K. Dey & C. R., (Eds.), *Handbook of Statistics* (Vol. 25, pp. 459–507). Elsevier.

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. (1989). *Generalized linear models*. London: Chapman & Hall.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

McParland, D., & Gormley, I. C. (2016). Model based clustering for mixed data: clustMD. *Advances in Data Analysis and Classification, 10*(2), 155–169.

Nielsen, S. F. (2000). The stochastic em algorithm: Estimation and asymptotic results. *Bernoulli, 6*(3), 457–489.

Oberski, D., & Vermunt, J. (2015). The relation between cub and loglinear models with latent variables. *Electronic Journal of Applied Statistical Analysis, 8*, 368–377.

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

Piccolo, D., Simone, R., & Iannario, M. (2019). Cumulative and cub models for rating data: A comparative analysis. *International Statistical Review, 87*, 207–236.

Selosse, M., Jacques, J., & Biernacki, C. (2020). Model-based co-clustering for mixed type data. *Computational Statistics and Data Analysis, 144*, Article 106866.

Selosse, M., Jacques, J., Biernacki, C., & Cousson-Gélie, F. (2019). Analyzing health quality survey using constrained co-clustering model for ordinal data and some dynamic implication. *Journal of the Royal Statistical Society, Series C, 65*(5), 1327–1349.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society - Series B, 62*(4), 795–809.

Ventura, M., Jacques, J., & Zuccolotto, P. (2024). Clustering multivariate rating data within the cub framework. In Pollice, A. and Mariani, P., editors, *Methodological and Applied Statistics and Demography IV. SIS 2024. Italian Statistical Society Series on Advances in Statistics.*, Bari.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*, 213–239.

Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed discrete outcome variables. *Multivariate Behavioral Research, 40*(3), 281–301.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research, 17*, 33–51.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450–469.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In I. J. Hagenaars & A. M., (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2016a). *Technical guide for latent GOLD 5.1: Basic, advanced, and syntax.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2016b). *Upgrade manual for latent GOLD 5.1: Basic, advanced, and syntax*. Statistical Innovations Inc., Belmont Massachusetts.