Statistical Papers
https://doi.org/10.1007/s00362-024-01560-2

REGULAR ARTICLE

A mixture distribution for modelling bivariate ordinal data

Ryan H. L. Ip[1,2] · **K. Y. K. Wu**[3]

Received: 26 April 2023 / Revised: 15 February 2024

Abstract

Ordinal responses often arise from surveys which require respondents to rate items on a Likert scale. Since most surveys contain more than one question, the data collected are multivariate in nature, and the associations between different survey items are usually of considerable interest. In this paper, we focus on a mixture distribution, called the combination of uniform and binomial (CUB), under which each response is assumed to originate from either the respondent's uncertainty or the actual feeling towards the survey item. We extend the CUB model to the bivariate case for modelling two correlated ordinal data without using copula-based approaches. The proposed model allows the associations between the unobserved uncertainty and feeling components of the variables to be estimated, a distinctive feature compared to previous attempts. This article describes the underlying logic and deals with both theoretical and practical aspects of the proposed model. In particular, we will show that the model is identifiable under a wide range of conditions. Practical inferential aspects such as parameter estimation, standard error calculations and hypothesis tests will be discussed through simulations and a real case study.

Keywords CUB · Finite mixture · Identifiability · Likert scale · Survey responses · Uncertainty

1 Introduction

Ordinal data are frequently encountered in various disciplines. As mentioned in Anderson (1984), ordinal data often arise in two situations: (1) thresholding an underlying

R. H. L. Ip and K. Y. K. Wu have contributed equally to this work.

✉ Ryan H. L. Ip
ryan.ip@aut.ac.nz

1 Department of Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand

2 School of Computing, Mathematics and Engineering, Charles Sturt University, Boorooma Street, Wagga Wagga, NSW 2650, Australia

3 The School of Business, Singapore University of Social Sciences, Clementi Road, Singapore 599494, Singapore

Published online: 20 May 2024

continuous variable, and (2) ranking provided by an assessor after processing unspecified amount of available information. An example of the first type could be the abundance of species based on percentage cover on the ground, which can be defined as 0 (absence), 1 (> 0–5% cover), 2 (> 5–12% cover), and so on (Guisan and Harrell 2000). When it is reasonable to assume the existence of a latent continuous variable, logit- or probit-type regression models are commonly employed to analyse the data (McCullagh 1980; Agresti 2010), see also a recent review for a detailed account of various ordinal regression models (Tutz 2022).

The second type of ordinal data is usually recorded in terms of a Likert scale, which has become a widely used tool in researches that involve surveys and questionnaires (Joshi et al. 2015). For example, in visual grading experiments for medical images, assessors are often requested to classify an image using one of several possible options such as "Definitely it is not clearly visible", "Probably it is not clearly visible", and so on (Al-Humairi et al. 2022). Since this type of data is usually collected from human respondents, there exists response biases which may make the data not truly reflecting the respondent's actual opinion towards the survey item (Baumgartner and Steenkamp 2006). For example, in answering a survey question, some people may choose a satisficing option rather than investing their time to give the optimal answer (Krosnick 1999). Van Vaerenbergh and Thomas (2013) have also reported different response styles where respondents tend to choose an answer regardless of the content. Thus, any serious attempt to analyse survey data should take into account the potential response biases inherent in the data. As argued by Iannario and Piccolo (2016), one of the simplest ways to model these kinds of data is to use a two-component model, which explicitly assumes that the data are generated from two processes as described below.

To this end, this paper focuses on the use of finite mixture models to analyse ordinal data arising from surveys. An advantage of using finite mixture models is that the data can be considered as generated from different underlying processes or heterogeneous populations, allowing for a greater flexibility (McLachlan et al. 2019). A popular mixture model that has gained attention recently is the combination of uniform and binomial (CUB) model. Since introduced by Piccolo (2003) and D'Elia and Piccolo (2005), CUB models and their variants have been widely applied in various disciplines to model ordinal data, especially those arising from surveys which require respondents to choose a response from a Likert scale. For example, CUB models have been applied in modelling survival probabilities (Iannario and Piccolo 2010b), customer preferences on food quality (Piccolo and D'Elia 2008), and job satisfaction (Gambacorta and Iannario 2013), just to name a few.

Under the settings of CUB models, the uniform component represents the indecisiveness or uncertainty of the respondent towards the survey item. In such a case, the respondent is assumed to pick an answer completely at random. The binomial component, on the other hand, is related to the feeling or actual opinion of the respondent towards the survey item. The stronger the feeling, the higher the rating. However, the analyst will not be able to distinguish whether the response is a completely random selection or a reflection of the actual feeling of the respondent. Nonetheless, the estimated parameters could inform the measure of uncertainty and preference for typical respondents. More details regarding the foundations and developments of CUB models can be found in a recent review (Piccolo and Simone 2019).

Most of the CUB models developed so far are univariate in nature. In other words, they focus on merely one survey item or question. Since most surveys contain more than one question, the data collected are multivariate in nature. To capture the dependency structure between the responses from several survey items, multivariate models are required. Some notable attempts to introduce multivariate CUB distributions include Corduas (2011, 2015), Andreis and Ferrari (2013), Colombi and Giordano (2016) and Colombi et al. (2019). Except the last one, all these works use copula-based methods to combine univariate CUB random variables. In particular, Colombi and Giordano (2016) employed the Sarmanov distribution while the others used the Plackett distribution. On a related note, Barbiero (2021) demonstrates how a joint distribution of two CUB margins can be constructed using copulas to match a desired correlation. While copula-based methods are flexible, there are limitations that cannot be overlooked. Firstly, copulas are usually applied to continuous random variables. The dangers and restrictions of applying the same practices to discrete distributions have been outlined by several authors, see Genest and Nešlehová (2007) and Geenens (2020) for example. Specifically, since copulas cannot be uniquely defined for discrete variables (Nelsen 2006), there are identifiability issues, which may cause inconsistency in parameter estimation (Genest and Nešlehová 2007). Secondly, parameter(s) in copula models is (are) usually related to either the rank or Pearson correlation between the two univariate random variables. However, since CUB random variables are a combination of two processes, the copula parameter(s) (assuming consistent) would relate to the overall correlation between the mixtures only, rather than the correlation between the individual uniform or binomial components. This may make the interpretation of the estimated copula parameters difficult.

To avoid the above concerns, this paper aims to construct a joint distribution for (R_1, R_2), which represents a pair of ratings arising from a survey, using bivariate uniform and bivariate binomial distributions. Some important features of the proposed model include (1) both R_1 and R_2 follow a CUB distribution marginally, (2) the joint distribution is not derived through copula-based routines, and (3) the dependency between the uniform and binomial components can be estimated separately, allowing better interpretation of model parameters. Our proposed model is similar to the hierarchical marginal models with latent uncertainty (HMMLU) proposed by Colombi et al. (2019), which will be described more formally in Sect. 3. Briefly, in their work, the uncertainty components can take a more flexible shape while the feeling components and the corresponding associations are modelled directly using marginal logits and log odds ratios, in the spirit of marginal models (Molenberghs and Lesaffre 1994; Bartolucci et al. 2007). One drawback of HMMLU is that the uncertainty components are assumed to be independent. Our proposed model overcomes this by having a parameter that directly measures the correlation between the uncertainty components. Another drawback of HMMLU lies in the large number of parameters, which characterise the marginal logits and log odds ratios, especially in the absence of covariates. In our proposed model, the feelings are modelled using a bivariate binomial distribution which contains only three parameters, making it more parsimonious.

The rest of the paper is organised as follows. Section 2 provides a brief account of the CUB, bivariate uniform and bivariate binomial distributions. Section 3 demonstrates how these distributions can be combined to form a new class of bivariate CUB models. A comparison between the proposed model and HMMLU is provided as well. Section 4 deals with various inferential issues including identifiability, parameter estimation, calculation of standard errors and hypothesis tests. Simulation and application results are reported in Sects. 5 and 6, respectively. Finally, Sect. 7 provides a conclusion and discussions.

2 Preliminaries

Formally, a random variable R is said to follow the CUB distribution with parameters π and ξ, denoted by $R \sim CUB(\pi, \xi)$, if the probability mass function (pmf) is a mixture of the discrete uniform distribution and a binomial distribution. Suppose R takes one of the $(m + 1)$ values from $\{0, 1, 2, \ldots, m\}$, the pmf admits the form

$$P(R = r) = \frac{1 - \pi}{1 + m} + \pi C_r^m (1 - \xi)^r \xi^{m-r}.$$

Here, the mixing weight $(1 - \pi)$ measures the degree of uncertainty while $(1 - \xi)$ measures the degree of feeling. As $(1 - \xi)$ increases, there is a higher chance of observing a higher rating.

In the literature, CUB random variables are assumed to range from 1 to m instead of starting from zero, represented by a shifted binomial distribution (Iannario and Piccolo 2010a). However, in this paper, we use the ordinary binomial distribution for several reasons. Firstly, real survey choice sets are often textual and arbitrary in numbering, making numerical interpretation less meaningful. Secondly, we treat the binomial component as a sum of independent Bernoulli variables, which naturally starts from zero. Lastly, ordinary binomial distribution results are more accessible and less confusing for readers unfamiliar with the history of CUB models.

To construct a bivariate model for R_1 and R_2 (which may represent rating responses from two survey questions), we first provide some details on a bivariate discrete uniform distribution and a bivariate binomial distribution which we have chosen to work on. A main feature of these distributions is that the marginal distributions belong to the same class.

2.1 Bivariate discrete uniform distribution

Let U_1 and U_2 be two random variables where the pmf for U_1 admits the form

$$P(U_1 = u_1) = \frac{1}{m + 1}, \qquad u_1 = 0, 1, 2, \ldots, m. \tag{1}$$

We further assume the following form for the conditional distribution of U_2 given U_1:

$$P(U_2 = u_2 | U_1 = u_1) = \begin{cases} \frac{1+\alpha_U}{m+1}, & \text{if } u_2 = u_1; \\ \frac{m-\alpha_U}{m(m+1)}, & \text{otherwise.} \end{cases} \tag{2}$$

In other words, the conditional distribution of $U_2|U_1$ is not uniform but a categorical distribution. The parameter α_U characterises the dependence between U_1 and U_2. Depending on the value of α_U, the probability of choosing the same answer in Question 2, given the response in Question 1, can be higher, lower, or unchanged. The admissible range of α_U is $[-1, m]$, with $\alpha_U = 0$ representing the case of independence. Marginally, U_2 follows the discrete uniform distribution, since

$$\begin{aligned} P(U_2 = u_2) &= \sum_{u_1=0}^{m} P(U_2|U_1)P(U_1) \\ &= \frac{1}{m+1}\left(\frac{1+\alpha_U}{1+m}\right) + m\frac{1}{m+1}\left(\frac{m-\alpha_U}{m(m+1)}\right) = \frac{1}{m+1}. \end{aligned}$$

The joint distribution of U_1 and U_2 can be written as

$$P(U_1 = u_1, U_2 = u_2) = \frac{m + m\alpha_U 1_{u_2=u_1} - \alpha_U 1_{u_2 \neq u_1}}{m(m+1)^2} \equiv U_{12}(u_1, u_2, \alpha_U),$$

where 1_A is the indicator variable which takes a value of 1 if condition A is satisfied; and 0 otherwise. Notice that U_1 and U_2 are independent only if $\alpha_U = 0$.

Studies on psychological aspects of survey responses have revealed the tendency for respondents to select the same category regardless of the question, thus we expect α_U to be positive in practice. For example, three of the common response styles reported by Baumgartner and Steenkamp (2001) and Van Vaerenbergh and Thomas (2013) are acquiescence response style, extreme response style and midpoint responding. These response styles refer to the tendency to agree with the item regardless of content, to select the most extreme category regardless of content, and to choose the middle scale category regardless of content. All these tendencies would make the probability of having two identical responses higher than expected under the independence assumption. The first two moments of U_1 and U_2 are summarised below. The derivations can be found in Appendix A.

$$\begin{aligned} E(U_i) &= \frac{m}{2}, \\ \mathrm{Var}(U_i) &= \frac{m(m+2)}{12}, \\ \mathrm{Cov}(U_1, U_2) &= \frac{\alpha_U(m+2)}{12}, \\ r_U = \mathrm{Corr}(U_1, U_2) &= \frac{\alpha_U}{m}. \end{aligned}$$

2.2 Bivariate binomial distribution

We will make use of the bivariate binomial distribution introduced in Biswas and Hwang (2002), where further details can be found. Let T_1 and T_2 be two random variables. Following Biswas and Hwang (2002), we consider T_1 as a sum of m independent Bernoulli variables, i.e., $T_1 = \sum_{i=1}^{m} T_{1i}$ where $T_{1i} \overset{\text{i.i.d.}}{\sim} Ber(1-\xi_1)$. Given T_{1i}, another Bernoulli variable T_{2i} is generated such that

$$P(T_{2i} = 1|T_{1i}) = \frac{1 - \xi_2 + \alpha_B(\xi_1 - \xi_2) + \alpha_B T_{1i}}{1 + \alpha_B}, \tag{3}$$

where α_B measures the dependency between T_{1i} and T_{2i}, with the admissible ranges

$$\alpha_B \in \begin{cases} \left(\max\left\{-\frac{\xi_2}{1-\xi_1+\xi_2}, \frac{\xi_2-1}{1+\xi_1-\xi_2}\right\}, \frac{1-\xi_2}{\xi_2-\xi_1}\right), & 1-\xi_1 > 1-\xi_2; \\ \left(\max\left\{-\frac{\xi_2}{1-\xi_1+\xi_2}, \frac{\xi_2-1}{1+\xi_1-\xi_2}\right\}, \frac{\xi_2}{\xi_1-\xi_2}\right), & 1-\xi_2 > 1-\xi_1; \\ (\max\{-\xi, -(1-\xi)\}, \infty), & 1-\xi_2 = 1-\xi_1 = 1-\xi. \end{cases} \tag{4}$$

Remark 1 The admissible ranges given in (4) ensure $P(T_{2i} = 0|T_{1i})$ and $P(T_{2i} = 1|T_{1i})$ are between 0 and 1. The above ranges correct the ones provided in Biswas and Hwang (2002).

When $\alpha_B = 0$, T_{1i} and T_{2i} are independent. Furthermore, T_{1i} and T_{2j} are assumed to be independent for all $i \neq j$. Marginally, it can be checked that $T_{2i} \sim Ber(1-\xi_2)$ since

$$\begin{aligned} P(T_{2i} = 1) &= (1-\xi_1)\left(\frac{1-\xi_2+\alpha_B(\xi_1-\xi_2)+\alpha_B}{1+\alpha_B}\right) + \xi_1\left(\frac{1-\xi_2+\alpha_B(\xi_1-\xi_2)}{1+\alpha_B}\right) \\ &= 1-\xi_2. \end{aligned}$$

We further define $T_2 = \sum_{i=1}^{m} T_{2i}$. In other words, both T_1 and T_2 follow the binomial distribution with parameters $(m, 1-\xi_1)$ and $(m, 1-\xi_2)$, respectively. The conditional distribution of T_2 given T_1 is given as

$$P(T_2 = t_2|T_1 = t_1) = (1+\alpha_B)^{-m} \times \sum_{j=0}^{t_1} C_j^{t_1} C_{t_2-j}^{m-t_1} w_1^j w_2^{t_1-j} w_3^{t_2-j} w_4^{m-t_1-t_2+j},$$

where

$$\begin{aligned} w_1 &= 1-\xi_2+\alpha_B(\xi_1-\xi_2)+\alpha_B, \\ w_2 &= \xi_2-\alpha_B(\xi_1-\xi_2), \\ w_3 &= 1-\xi_2+\alpha_B(\xi_1-\xi_2), \quad \text{and} \\ w_4 &= \xi_2-\alpha_B(\xi_1-\xi_2)+\alpha_B. \end{aligned}$$

Hence, the joint distribution of T_1 and T_2 can be written as

$$
\begin{aligned}
P(T_1 = t_1, T_2 = t_2) &= B_1(t_1) \times (1+\alpha_B)^{-m} \\
&\qquad \times \sum_{j=0}^{t_1} C_j^{t_1} C_{t_2-j}^{m-t_1} w_1^j w_2^{t_1-j} w_3^{t_2-j} w_4^{m-t_1-t_2+j} \\
&\equiv B_{12}(t_1, t_2; \xi_1, \xi_2, \alpha_B),
\end{aligned}
$$

where $B_1(t_1) = C_{t_1}^m (1-\xi_1)^{t_1} \xi_1^{m-t_1}$. The covariance and correlation of T_1 and T_2 are given below [see also Biswas and Hwang (2002) for a more general class of the bivariate binomial distribution]:

$$
\begin{aligned}
\mathrm{Cov}(T_1, T_2) &= \frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1), \\
r_T = \mathrm{Corr}(T_1, T_2) &= \frac{\alpha_B}{1+\alpha_B}\sqrt{\frac{\xi_1(1-\xi_1)}{\xi_2(1-\xi_2)}}.
\end{aligned} \tag{5}
$$

Mathematical derivations are provided in Appendix A. When two survey questions inquire about similar aspects, it is reasonable to expect a positive correlation in the responses ($\alpha_B > 0$). In the opposite, if the two questions are probing for conflicting aspects (for example, satisfaction of salary and tendency to leave the company), one may anticipate a negative α_B. Studies of survey response have also revealed that prior questions often influence later responses (Krosnick and Alwin 1987; Tourangeau et al. 2000), thus it is important to capture the correlation between T_1 and T_2.

3 A new class of bivariate CUB distributions

Suppose R_1 and R_2 represent the ordinal responses from two survey questions answered by the same respondent. Although there is no requirement for R_1 and R_2 to be the responses from two consecutive questions, it may be easier to understand the process considering that way. We assume the following generating process.

The respondent first decides if s/he is uncertain or certain about his/her feeling towards Question 1. If s/he is uncertain, the rating is given randomly according to a discrete uniform distribution. If s/he is certain, the rating is given by a binomial distribution reflecting her/his feeling. Hence, R_1 resembles the generating process of a univariate CUB variable. The same process is repeated Question 2. However, this time the rating may depend on the rating provided in the previous question.

Since the decision process is repeated two times, there are four scenarios: (uncertain, uncertain), (uncertain, certain), (certain, uncertain) and (certain, certain), with respective probabilities $(1-\pi_1)(1-\pi_2)$, $(1-\pi_1)\pi_2$, $\pi_1(1-\pi_2)$ and $\pi_1\pi_2$. Symbolically, let D_1 and D_2 be two independent Bernoulli variables with $P(D_i = 1) = \pi_i$. The four scenarios can be written as $(D_1 = 0, D_2 = 0)$, $(D_1 = 0, D_2 = 1)$, $(D_1 = 1, D_2 = 0)$, and $(D_1 = 1, D_2 = 1)$. We also assume that if the 'regime' goes from uncertain to certain (or vice versa), the ratings given in the two questions are independent. Such

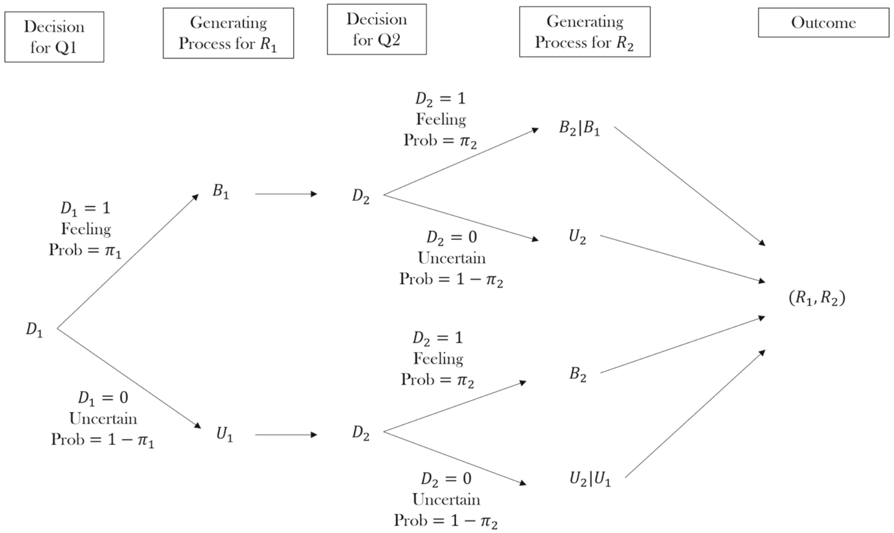

Fig. 1 Schematic flowchart showing the generating process of (R_1, R_2)

a process is represented schematically in Fig. 1. Note that all the stages except the outcome are unobservable and therefore unobserved. The above described process would result in the following joint distribution:

$$\begin{aligned} &P(R_1 = r_1, R_2 = r_2) \\ &\quad = (1-\pi_1)(1-\pi_2)U_{12}(r_1, r_2, \alpha_U) + \frac{(1-\pi_1)\pi_2 B_2(r_2)}{m+1} + \frac{\pi_1(1-\pi_2)B_1(r_1)}{m+1} \\ &\quad\quad + \pi_1\pi_2 B_{12}(r_1, r_2; \xi_1, \xi_2, \alpha_B) \end{aligned} \tag{6}$$

From the joint distribution, it can be checked that, marginally, both R_1 and R_2 follow a univariate CUB distribution with parameters (π_1, ξ_1) and (π_2, ξ_2), respectively. Also, R_1 and R_2 are independent if and only if $\alpha_B = \alpha_U = 0$. In that case,

$$P(R_1 = r_1, R_2 = r_2) = \left[\frac{1-\pi_1}{m+1} + \pi_1 B_1(r_1)\right]\left[\frac{1-\pi_2}{m+1} + \pi_2 B_2(r_2)\right].$$

The first two moments of the proposed bivariate CUB distribution are given by

$$\begin{aligned} E(R_i) &= (1-\pi_i)\frac{m}{2} + \pi_i m(1-\xi_i), \quad i = 1, 2; \\ \mathrm{Var}(R_i) &= (1-\pi_i)m\left[\frac{2m+1}{6} - \frac{(1-\pi_i)m}{4}\right] \\ &\quad + \pi_i m(1-\xi_i)\xi_i[1 - m(1-\pi_i)], \quad i = 1, 2; \\ \mathrm{Cov}(R_1, R_2) &= (1-\pi_1)(1-\pi_2)\frac{\alpha_U(m+2)}{12} + \pi_1\pi_2\frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1). \end{aligned} \tag{7}$$

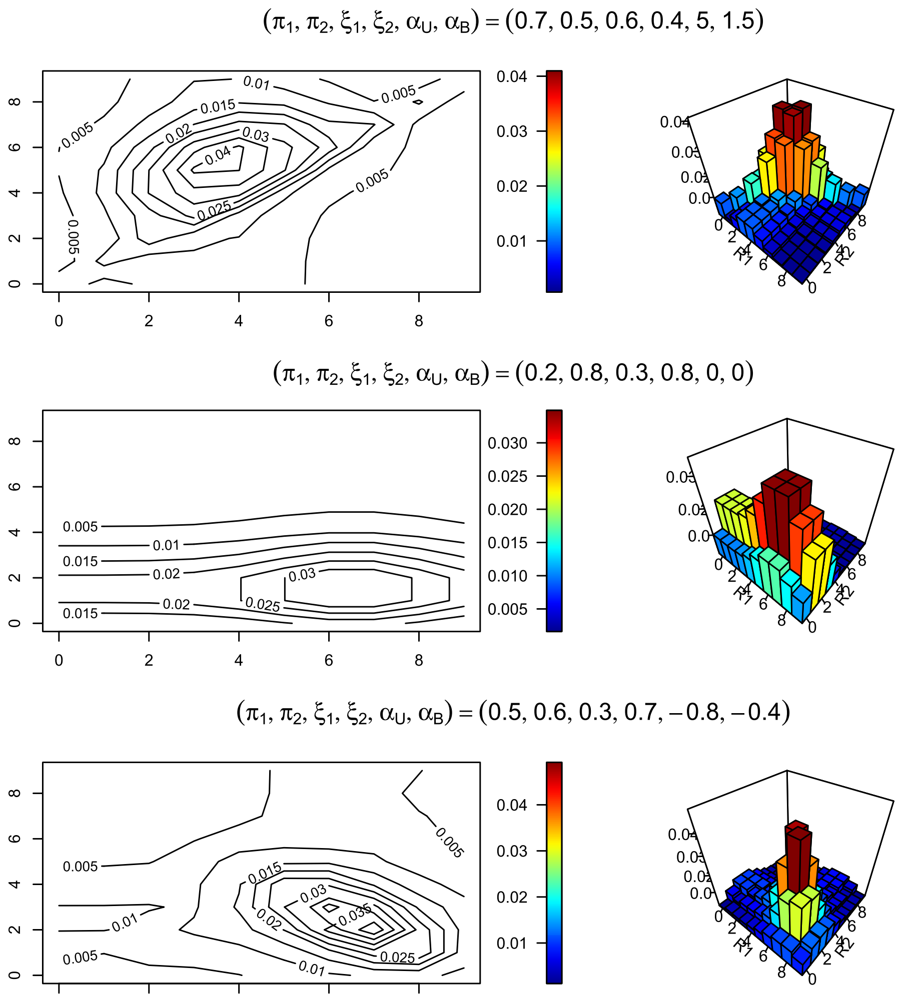

Fig. 2 Contour plots and 3D histograms of some bivariate CUB models under three sets of parameters with $m = 9$

Derivation details are provided in Appendix A. The correlation, r_R, can then be derived from the covariance and the variances. Figure 2 shows the contour plots and 3D histograms for the joint probability mass functions under three sets of parameters. From top to bottom panels, the figure demonstrates the cases where R_1 and R_2 are positively correlated, independent, and negatively correlated, respectively.

From (7), it can be deduced that the correlation between R_1 and R_2 is zero if $\alpha_B = \alpha_U = 0$, or when

$$\alpha_U = \frac{-12\pi_1\pi_2}{(1-\pi_1)(1-\pi_2)} \frac{m\xi_1(1-\xi_1)}{m+2} \frac{\alpha_B}{1+\alpha_B},$$

as long as the right-hand-side (RHS) of the above equation is within the admissible range provided in (4). In other words, the dependency within the uniform components may sometimes cancel out that due to the binomial components.

The correlation r_R between the two responses are governed by not only α_U and α_B, but also all other parameters. For this reason, r_R may sometimes be misleading, or at least undermining the dependency between the respondent's feelings towards the two items. For instance, if $\pi_1 = \pi_2 = 0.5$, then approximately half of the pairs (R_1, R_2) will be generated independently, which may shrink the overall correlation r_R, even when r_U and r_T are reasonably large. Yet, in practice, r_U and r_T may be of higher interest. The former represents the tendency of choosing the same category when the respondent was uncertain towards both questions, while the latter represents the correlation between the liking of the two survey items. Once the model parameters were estimated, r_T and r_U can be found correspondingly. The separation of the overall dependency into different components cannot be accomplished in any previously proposed copula-based methods, as these methods tend to estimate the overall correlation between the two margins.

3.1 Comparison with HMMLU

A model that is similar to the bivariate CUB model proposed is the aforementioned HMMLU (Colombi et al. 2019). Similar to our approach, the data generating process of HMMLU assumes the existence of latent states that represent if the respondent's answer was based on feeling or uncertainty. In the bivariate case, the four scenarios $(D_1 = 0, D_2 = 0)$, $(D_1 = 0, D_2 = 1)$, $(D_1 = 1, D_2 = 0)$, and $(D_1 = 1, D_2 = 1)$ would still apply. The major difference between HMMLU and our proposal lies in the distributions of the responses under each of the four scenarios.

When an the answer is given with uncertainty $(D = 0)$, HMMLU assumes a distribution $h_i(r_i)$, $i = 1, 2$, which can take different shapes such as U-shape and bell shape. The uniform distribution is one of the special cases. When both answers are given with uncertainty, R_1 and R_2 are assumed to be independent under HMMLU. In the opposite, when an answer is given with certainty $(D = 1)$, HMMLU does not impose any specific distribution for the responses. Rather, the marginal distributions and the joint distribution are parameterised through marginal logits and log odds ratios, respectively. Such an approach stems from the general framework of marginal models for categorical data (Bergsma and Rudas 2002; Bartolucci et al. 2007). The joint distribution of R_1 and R_2 under HMMLU can be written as

$$
\begin{aligned}
P(R_1 = r_1, R_2 = r_2) &= \sum_{i,j=0,1} \pi_{ij} P(R_1 = r_1, R_2 = r_2 | D_1 = i, D_2 = j) \\
&= \pi_{00} h_1(r_1) h_2(r_2) + \pi_{01} h_1(r_1) P(R_2 = r_2 | D_2 = 1) \\
&\quad + \pi_{10} h_2(r_2) P(R_1 = r_1 | D_1 = 1) \\
&\quad + \pi_{11} P(R_1 = r_1, R_2 = r_2 | D_1 = D_2 = 1)
\end{aligned} \tag{8}
$$

Comparing Eqs. (6) and (8), some differences between HMMLU and the proposed bivariate CUB model are notable. Firstly, HMMLU does not allow correlation between uncertain responses. In the bivariate CUB model, such a correlation is captured through α_U in U_{12}. Of course, when $\alpha_U = 0$, the uncertain responses under the bivariate CUB model are generated in the same manner as HMMLU when both h_i take the uniform distribution. Secondly, the mixing weights (π's) are generated differently. Implicitly, our approach assumes that D_1 and D_2 are independent while HMMLU allows them to be dependent.

Lastly, the distributions of the certain responses under HMMLU need not be the binomial distributions, and are hence more flexible. However, a consequence of which is that HMMLU contains way more parameters. The situation is more obvious in the absence of covariates. For example, with $m+1$ categories, HMMLU would require m parameters for the marginal logits for each of R_1 and R_2, and $(m-1)^2$ log odds ratios to parameterise the joint distribution. As mentioned in Colombi et al. (2019, p. 599), the large number of parameters will usually lead to identifiability issues, and constraints are therefore required. In the opposite, the distributions of the certain responses under the proposed bivariate CUB model can be characterised using three parameters ξ_1, ξ_2 and α_B.

4 Inferential issues

Next, we discuss various issues related to the inferential processes. We start with the identifiability since the estimation of the parameters is only meaningful if the model is identifiable. Next, we discuss the strategy of estimating the parameters. Before closing this section, we provide details for the standard error calculations and hypothesis tests for some of the parameters.

4.1 Identifiability

The following theorem specifies the conditions under which the bivariate CUB model is identifiable.

Theorem 1 *Given that* $0 < \pi_1, \pi_2, \xi_1, \xi_2 < 1$, $\xi_1 \neq \xi_2$, *and* $m \geq 3$, *the bivariate CUB model given in* (6) *is identifiable.*

Before we provide a proof for the theorem above, we first place a couple of remarks. Firstly, the condition $m \geq 3$ (i.e., the number of categories is at least 4) is equivalent to the condition required in the univariate case (Iannario 2010). Similar to the identifiability condition for HMMLU, restrictions on the number of categories are necessary to make sure that the number of parameters is less than the number of free frequencies (Colombi et al. 2019). The univariate CUB model is still identifiable when $\pi = 1$ (Iannario 2010) because the discrete uniform distribution does not contain any parameters (ξ can be identified even when $\pi = 1$). Thus, the identifiability is ensured for $\pi > 0$. However, in the bivariate case, as specified in Theorem 1, while it is still required that $\pi_1, \pi_2 > 0$, either π_1 or $\pi_2 = 1$ would make the model non-identifiable, since there are infinite number of possible α_U that would yield the same joint distribution (6) under such case. Similarly, when ξ_1 or ξ_2 takes either values of 0 or 1, α_B cannot be identified as well. The additional requirement of $\xi_1 \neq \xi_2$ may seem restrictive. In practice, however, since ξ_1 and ξ_2 correspond to the feeling of a respondent towards two survey questions, the values rarely coincide, unless the two questions are probing for exactly the same aspect (in that case, a single question would be sufficient).

Proof Let $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)' \in \boldsymbol{\Theta} = (0, 1)^4 \times [-1, m] \times \mathcal{A}_B$ where $\mathcal{A}_B$ is the parameter space for α_B governed by (4), with the exception that ξ_1 cannot be equal to ξ_2. Further, denote by $P_{r_1,r_2}(\boldsymbol{\theta}) = P(R_1 = r_1, R_2 = r_2; \boldsymbol{\theta})$, $P_{\bullet r_2} = \sum_{r_1=0}^{m} P_{r_1,r_2}$ and $P_{r_1 \bullet} = \sum_{r_2=0}^{m} P_{r_1,r_2}$. The bivariate CUB model is identifiable if and only if, for any parameter vector $\boldsymbol{\theta}^*$, the system of equations in $\boldsymbol{\theta}$:

$$P_{r_1,r_2}(\boldsymbol{\theta}) = P_{r_1,r_2}(\boldsymbol{\theta}^*), \qquad r_1, r_2 = 0, 1, \ldots, m, \tag{9}$$

admits only one solution in the parameter space (Manisera and Zuccolotto 2015). With $(m+1)$ categories, there are altogether $(m+1)^2$ equations in (9). Fortunately, results in Manisera and Zuccolotto (2015) also demonstrate that it is possible to reduce the number of equations in the system by constructing some equations that allow the parameters to be specified sequentially.

For the bivariate CUB model on hand, we consider the following system of equations:

$$\begin{cases} \frac{P_{m\bullet}(\boldsymbol{\theta}) - P_{0\bullet}(\boldsymbol{\theta})}{P_{0\bullet}(\boldsymbol{\theta}) - 1/(m+1)} = \frac{P_{m\bullet}(\boldsymbol{\theta}^*) - P_{0\bullet}(\boldsymbol{\theta}^*)}{P_{0\bullet}(\boldsymbol{\theta}^*) - 1/(m+1)} \\ \pi_1 = \frac{P_{0\bullet}(\boldsymbol{\theta}^*) - 1/(m+1)}{\xi_1^m - 1/(m+1)} \\ \frac{P_{\bullet m}(\boldsymbol{\theta}) - P_{\bullet 0}(\boldsymbol{\theta})}{P_{\bullet 0}(\boldsymbol{\theta}) - 1/(m+1)} = \frac{P_{\bullet 0}(\boldsymbol{\theta}^*) - P_{\bullet 0}(\boldsymbol{\theta}^*)}{P_{\bullet 0}(\boldsymbol{\theta}^*) - 1/(m+1)} \\ \pi_2 = \frac{P_{\bullet 0}(\boldsymbol{\theta}^*) - 1/(m+1)}{\xi_2^m - 1/(m+1)} \\ \frac{P_{01}(\boldsymbol{\theta}) - P_{10}(\boldsymbol{\theta})}{\xi_1^{m-1}(\xi_1 - \xi_2)} = \frac{P_{01}(\boldsymbol{\theta}^*) - P_{10}(\boldsymbol{\theta}^*)}{\xi_1^{m-1}(\xi_1 - \xi_2)} \\ \alpha_U = \left(P_{00}(\boldsymbol{\theta}^*) - \frac{(1-\pi_1)\pi_2 B_2(0)}{m+1} - \frac{\pi_1(1-\pi_2)B_1(0)}{m+1} - \frac{\pi_1\pi_2 B_1(0) w_4^m}{(1+\alpha_B)^m} \right) \frac{(m+1)^2}{(1-\pi_1)(1-\pi_2)} - 1 \end{cases} \tag{10}$$

The selection of the above system was merely due to the simplicity of algebra involved, as shown below. In the first equation, both $P_{m\bullet}$ and $P_{0\bullet}$ represent marginal probabilities which are free of α_U and α_B. According to Iannario (2010), the first two equations in (10) allow π_1 and ξ_1 to be uniquely specified. Similarly, the second two allow π_2 and ξ_2 to be uniquely specified. If α_B can be uniquely specified, the last equation will only yield one α_U (hence unique). Thus, it remains to prove the uniqueness of α_B. For this purpose, we consider in details the fifth equation in (10).

Since

$$
\begin{aligned}
P_{01}(\boldsymbol{\theta}) &= (1-\pi_1)(1-\pi_2)\frac{m-\alpha_U}{m(m+1)^2} + \frac{(1-\pi_1)\pi_2 B_2(1)}{m+1} + \frac{\pi_1(1-\pi_2)B_1(0)}{m+1} \\
&\quad + \frac{\pi_1\pi_2 B_1(0) m w_3 w_4^{m-1}}{(1+\alpha_B)^m}, \qquad \text{and} \\
P_{10}(\boldsymbol{\theta}) &= (1-\pi_1)(1-\pi_2)\frac{m-\alpha_U}{m(m+1)^2} + \frac{(1-\pi_1)\pi_2 B_2(0)}{m+1} + \frac{\pi_1(1-\pi_2)B_1(1)}{m+1} \\
&\quad + \frac{\pi_1\pi_2 B_1(1) w_2 w_4^{m-1}}{(1+\alpha_B)^m},
\end{aligned}
$$

we have

$$
\begin{aligned}
\frac{P_{01}(\boldsymbol{\theta}) - P_{10}(\boldsymbol{\theta})}{\xi_1^{m-1}(\xi_1-\xi_2)} &= \frac{(1-\pi_1)\pi_2(B_2(1)-B_2(0))}{(m+1)\xi_1^{m-1}(\xi_1-\xi_2)} + \frac{\pi_1(1-\pi_2)(B_1(0)-B_1(1))}{(m+1)\xi_1^{m-1}(\xi_1-\xi_2)} \\
&\quad + \pi_1\pi_2\frac{m w_4^{m-1}}{(1+\alpha_B)^{m-1}}
\end{aligned} \tag{11}
$$

which is a function in α_B, and free of α_U, provided all other specified parameters π_1, π_2, ξ_1 and ξ_2. Furthermore, this function is continuous in α_U. To see this, we simply need to show that $\alpha_B > -1$. With $\xi_1 \neq \xi_2$, the lower bound of α_B is always greater than -1 since

$$
\begin{aligned}
-\frac{\xi_2}{1-\xi_1+\xi_2} - (-1) &= \frac{1-\xi_1}{1-\xi_1+\xi_2} > 0, \qquad \text{and} \\
\frac{\xi_2-1}{1+\xi_1-\xi_2} - (-1) &= \frac{\xi_1}{1+\xi_1-\xi_2} > 0.
\end{aligned}
$$

Now, we will show that the above function is monotonically increasing in α_B. Differentiating (11) with respect to α_B yields

$$
\frac{\pi_1\pi_2 m(m-1)(1-\xi_1)[\xi_2 - \alpha_B(\xi_1-\xi_2) + \alpha_B]^{m-2}}{(1+\alpha_B)^m}.
$$

Since $\alpha_B > -1$, the denominator is always positive. Now, consider $\xi_2 - \alpha_B(\xi_1 - \xi_2) + \alpha_B$. The lower bound of α_B is given by

$$
\alpha_B > \max\left\{-\frac{\xi_2}{1-\xi_1+\xi_2}, \frac{\xi_2-1}{1+\xi_1-\xi_2}\right\}.
$$

Since

$$
\frac{\xi_2-1}{1+\xi_1-\xi_2} - \frac{-\xi_2}{1-\xi_1+\xi_2} = \frac{\xi_1+\xi_2-1}{1-(\xi_1-\xi_2)^2},
$$

we can deduce that

$$\max\left\{-\frac{\xi_2}{1-\xi_1+\xi_2}, \frac{\xi_2-1}{1+\xi_1-\xi_2}\right\} = \begin{cases} \frac{\xi_2-1}{1+\xi_1-\xi_2}, & \text{if } \xi_1+\xi_2-1 \geq 0; \\ -\frac{\xi_2}{1-\xi_1+\xi_2}, & \text{if } \xi_1+\xi_2-1 < 0. \end{cases}$$

If $\xi_1 + \xi_2 - 1 \geq 0$,

$$\begin{aligned} \xi_2 + \alpha_B(1-\xi_1+\xi_2) &> \xi_2 + \left(\frac{\xi_2-1}{1+\xi_1-\xi_2}\right)(1-\xi_1+\xi_2) \\ &= \frac{\xi_1+\xi_2-1}{1-\xi_1+\xi_2} \geq 0 \end{aligned}$$

In the opposite, if $\xi_1 + \xi_2 - 1 < 0$,

$$\begin{aligned} \alpha_B &> -\frac{\xi_2}{1-\xi_1+\xi_2} \\ \alpha_B(1-\xi_1+\xi_2) &> -\xi_2 \\ \xi_2 + \alpha_B(1-\xi_1+\xi_2) &> 0. \end{aligned}$$

Hence, $\xi_2 + \alpha_B(1 - \xi_1 + \xi_2)$ is always positive. Since Eq. (11) is continuous and monotonically increasing, one and only one α_B will be specified. This completes the proof. □

4.2 Parameter estimation

The parameter estimation can be carried out using the EM algorithm (Dempster et al. 1977). Although the chief focus of Dempster et al. (1977) was on handling incomplete data, the EM algorithm has been proven to work well for mixture distributions, including CUB models (Piccolo 2006). Further details on this topic can be found in Everitt and Hand (1981), Redner and Walker (1984), McLachlan and Peel (2000) and Arcidiacono and Jones (2003), among many others. The details of the algorithm for the proposed bivariate CUB model are provided in Appendix B.

4.3 Standard errors

The variance-covariance matrix of the estimated parameters can be obtained by inverting the observed information matrix:

$$\text{Var}(\hat{\boldsymbol{\theta}}) = I(\hat{\boldsymbol{\theta}})^{-1} = [-\Delta^2 \log L(\boldsymbol{\theta})]^{-1}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

The standard errors of the parameters are the square root of the diagonal elements of $\text{Var}(\hat{\boldsymbol{\theta}})$. The use of observed information matrix, instead of expected information matrix, has been justified in Efron and Hinkley (1978). Explicit expressions of the elements in $\text{Var}(\hat{\boldsymbol{\theta}})$ are provided in Appendix C.

4.4 Model selection

For a particular dataset on hand, when selecting between non-nested models such as the bivariate CUB and HMMLU, common measures such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be employed. In the context of the proposed bivariate CUB model, when comparing between nested models, it can be done via hypothesis tests by means of the likelihood ratio test (Hoel 1962). Here, we list some of the tests can be done regarding the dependency parameters:

- $H_0^1 : \alpha_B = c_1, \alpha_U = c_2$,
- $H_0^2 : \alpha_B = c$, and
- $H_0^3 : \alpha_U = c$,

for some constants c, c_1 and c_2, against the alternative hypothesis that H_0 is not true. In particular, testing if any or both of α_U and α_B is/ are zero would be of high interest. Under H_0^1, if $\alpha_B = \alpha_U = 0$, R_1 and R_2 are completely independent. Under H_0^2, if $\alpha_B = 0$, provided that the respondent chose to express his/ her opinions on both questions, the feelings towards the two questions are independent. Under H_0^3, if $\alpha_U = 0$, provided that the respondent was uncertain to both questions, his/ her choices of the categories are independent (both completely random). The test statistic is

$$-2\log[L(\boldsymbol{r}_k; \hat{\boldsymbol{\theta}}_0)/L(\boldsymbol{r}_k; \hat{\boldsymbol{\theta}})],$$

where $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ evaluated under the restrictions specified in H_0. The test statistic follows a χ^2 distribution approximately, with a degrees of freedom of 2 for the H_0^1, and 1 for both H_0^2 and H_0^3.

5 Simulation

Simulations were conducted to investigate the accuracy of the estimates based on the procedure described in Sect. 4.2 under two cases: (1) large sample with many categories, and (2) small sample with relatively fewer categories. As the number of categories is typically between 2 and 11, with 5 to 10 categories being the easiest to rate (Wakita et al. 2012), we have purposely chosen 5, 7 and 10 categories in the simulation studies below.

5.1 Large sample with 10 categories

In this simulation study, we set $m = 9$ (which means a total of 10 categories) and used two sets of parameters as given below.

- Set 1: $(\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)' = (0.7, 0.5, 0.6, 0.4, 5.0, 1.5)'$
- Set 2: $(\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)' = (0.5, 0.6, 0.6, 0.4, 3.0, -0.3)'$

For each set of parameters, we first simulated two Bernoulli variables D_1 and D_2 using π_1 and π_2 as the respective parameters. If $D_1 = D_2 = 0$, R_1 and R_2 were simulated using (1) and (2), respectively. If $D_1 = 0$ and $D_2 = 1$, R_1 was simulated using (1) and

Table 1 Mean and coefficient of variation (CV) of the estimated parameters under different simulation scenarios for the large sample sizes with ten categories

	Parameter	π_1	π_2	ξ_1	ξ_2	α_U	α_B
Set 1	True	0.7	0.5	0.6	0.4	5.0	1.5
$n = 1000$	Mean	0.7016	0.5018	0.5997	0.3999	5.1898	1.5141
	CV	0.0385	0.0625	0.0124	0.0260	0.2264	0.2192
$n = 2000$	Mean	0.7010	0.5014	0.6001	0.3999	5.0646	1.5040
	CV	0.0258	0.0442	0.0093	0.0180	0.1452	0.1682
$n = 3000$	Mean	0.7007	0.5006	0.5998	0.3996	5.0616	1.5058
	CV	0.0218	0.0338	0.0079	0.0146	0.1231	0.1522
Set 2	True	0.5	0.6	0.6	0.4	3.0	−0.3
$n = 1000$	Mean	0.5013	0.6013	0.6000	0.4006	3.0782	− 0.2918
	CV	0.0644	0.0523	0.0183	0.0227	0.2411	− 0.1823
$n = 2000$	Mean	0.5004	0.6001	0.5997	0.4000	3.0543	− 0.2969
	CV	0.0457	0.0353	0.0126	0.0161	0.1541	− 0.1222
$n = 3000$	Mean	0.5001	0.6006	0.6000	0.4001	3.0330	− 0.2992
	CV	0.0369	0.0296	0.0103	0.0131	0.1333	− 0.0957

R_2 was simulated using a binomial distribution with parameter $(1 - \xi_2)$. If $D_1 = 1$ and $D_2 = 0$, R_1 was simulated using a binomial distribution with parameter $(1 - \xi_1)$ and R_2 was simulated using (1). If $D_1 = D_2 = 1$, then m Bernoulli variables were simulated using $(1 - \xi_1)$ as the parameter. These m Bernoulli variables were summed up to yield R_1. Conditional on each value of these m Bernoulli variables, another m Bernoulli variables were generated with a parameter specified in (3). The sum of the latter m Bernoulli variables resulted in R_2. Three sample sizes $n = \{1000, 2000, 3000\}$ were used. For each sample size, 1000 replicates were simulated. The convergence threshold for the EM algorithm was set to be 1×10^{-5}.

Under the parameters specified in Set 1, $r_U = 0.56$, $r_T = 0.60$, and R_1 and R_2 are positively correlated, with a theoretical correlation of 0.24. Under those specified in Set 2, $r_U = 0.33$, $r_T = -0.43$, and R_1 and R_2 are only weakly positively correlated, with a theoretical correlation of 0.05. Table 1 summarises the estimation results across all simulation replicates.

For both sets of parameters, the biases of all estimated parameters were very small, with a generally decreasing trend with sample sizes. Meanwhile, the coefficients of variation decrease with the sample size as well, as one would expect. Not surprisingly, the variabilities of α_U and α_B were greater than the other parameters. This is probably due to the fact that these parameters can only be estimated when $D_1 = D_2 = 0$ and $D_1 = D_2 = 1$, respectively, hence requiring a larger sample size than the marginal parameters in achieving a lower variability. Overall, we conclude that the EM algorithm proposed in Sect. 4.2 worked well and is therefore an appropriate method for fitting the bivariate CUB model when both the sample size and the number of categories are large.

Table 2 Mean and CV of the estimated parameters for 1000 simulation replicates with $n = 100$, 200 and 300, and $m = 4$ and 6

	Parameter	π_1	π_2	ξ_1	ξ_2	α_U	α_B
$m = 4$	TRUE	0.7	0.3	0.8	0.4	3.0	0.2
$n = 100$	Mean	0.7035	0.3408	0.8012	0.4000	3.1515	0.3919
	CV	0.1199	0.3789	0.0401	0.3059	0.3053	2.4392
$n = 200$	Mean	0.7034	0.3148	0.8001	0.3995	3.1187	0.3005
	CV	0.0831	0.3161	0.0281	0.1817	0.2589	1.8987
$n = 300$	Mean	0.7017	0.3169	0.8000	0.4015	3.1556	0.2864
	CV	0.0700	0.2553	0.0231	0.1345	0.2276	1.7073
$m = 6$	TRUE	0.7	0.3	0.8	0.4	3.0	0.2
$n = 100$	Mean	0.7074	0.3281	0.7995	0.4053	3.5520	0.3433
	CV	0.1116	0.3440	0.0304	0.2487	0.4753	2.3310
$n = 200$	Mean	0.7024	0.3151	0.8001	0.3985	3.4175	0.3040
	CV	0.0815	0.2651	0.0220	0.1557	0.3907	1.7871
$n = 300$	Mean	0.7019	0.3071	0.8002	0.4012	3.2567	0.2853
	CV	0.0656	0.2205	0.0183	0.1163	0.3294	1.6311

5.2 Small sample with 5 or 7 categories

The data generating process was the same as those reported in Sect. 5.1, except the number of categories and sample sizes are smaller. Specifically, the cases where $m = 4$ and 6 were considered. For each value of m, sample sizes of 100, 200 and 300 were used. The parameters used were $(\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)' = (0.7, 0.3, 0.8, 0.4, 3.0, 0.2)'$. Compared to the previous two sets of parameters used, this set of parameters would make the data more sparse as ξ_1 is closer to 1, meaning that the values for R_1 are more concentrated in the lower end. The simulation results are provided in Table 2. From the results, as the sample size increases, a generally decreasing trend in the biases of the estimates can be observed. The marginal parameters can be accurately estimated even with the lower sample size considered, although larger biases can be observed compared to the large sample cases reported in Table 1. Consistent with the large sample case, the estimation of the dependency parameters α_U and α_B is less accurate than the marginal parameters. The number of categories does not seem to have a huge impact on the estimation of the parameters.

6 Application

The proposed bivariate CUB model was applied to the “relgoods” dataset, available within the CUB package (Iannario et al. 2020) in R (R Core Team 2022). The dataset contains results from a survey conducted in Naples, Italy, in 2014. Respondents of the survey were asked to evaluate their scores for various relational goods (for example, time dedicated to friends and family) and related issues such as safety of surroundings and their feeling of happiness. We focused on two of the questions related to the following aspects:

- Environment: the level of comfort with the surrounding environment, and
- Safety: the level of safety in the streets.

In the original survey, for both questions, respondents provided a score on a 10-point Likert scale, ranging from 1 = "never, at all" to 10 = "always, a lot". For our purpose, we have re-scaled the responses to 0 to 9 by subtracting 1 from each response (meaning that $m = 9$). The dataset contains many other variables. Univariate analysis results on some of the variables can be found in, for example, Iannario and Simone (2017) and Capecchi et al. (2018). Further details regarding the dataset can be found on https://rdrr.io/cran/CUB/man/relgoods.html. The R code used to obtain the results in this section is available as Supplementary Information online.

As one can naturally expect some association between the level of comfort with the surrounding environment and the level of safety in the surrounding areas, a bivariate model would be appropriate. Originally, there were a total of 2,459 responses. Upon removing 9 observations that contained missing values, the proposed bivariate CUB model was fitted on the remaining 2,450 observations. Here we label "Environment" as R_1 and "Safety" as R_2. The procedures described in Sects. 4.2 to 4.4 were employed to gain insights from the dataset.

Table 3 depicts the estimated parameters based on the proposed bivariate CUB model and separate univariate CUB models. The parameters under the univariate case were obtained using the functionalities within the CUB package (Iannario et al. 2020). Overall, the bivariate model resulted in a higher log-likelihood as well as a lower AIC and BIC, indicating a better goodness-of-fit (GOF). The better performance can also be checked visually by assessing the contour plots and 3D histograms provided in Fig. 3. In particular, the separate model was not able to capture the positive correlation between the two ratings.

Based on the estimated parameters in the bivariate model, we have $\hat{r}_U = 0.191$ and $\hat{r}_T = 0.316$, while the empirical correlation between R_1 and R_2 was $r_R = 0.229$. Thus, the correlation between the feelings of the two questions was larger than that suggested by r_R. Results of hypothesis tests in Table 4 also show that both α_U and α_B are significantly different from zero.

Suppose the respondent was uncertain towards both questions, the estimated value of $\hat{\alpha}_U = 1.723$ suggests that the estimated probability of choosing the same category, given the first response, was $1.723/10 = 0.1723$, a 72.3% increase compared to a model assuming independence among the responses. Moreover, suppose the respondent chose to express his/ her feeling towards the two questions, the model found a moderate positive correlation ($\hat{r}_T = 0.316$) among the two responses, indicating that the two responses tended to go in the same direction. That is, respondents who are satisfied with the level of comfort with the surrounding environment tended to be satisfied with the level of safety in the streets as well. These kinds of insights regarding the associations between the two survey items were not obtainable if the two variables were fitted separately.

The same dataset was also analysed using HMMLU with $h_i(r_i)$ taking the form of discrete uniform distribution. In total, 22 parameters were used: three for the mixing

Table 3 Estimated parameters under the proposed bivariate CUB model and separate univariate CUB models

	Parameter	π_1	π_2	ξ_1	ξ_2	α_U	α_B	$\log L(\boldsymbol{r}_k; \hat{\boldsymbol{\Theta}})$	AIC	BIC
Joint	Estimate	0.618	0.242	0.256	0.460	1.723	0.565	– 10699.00	21,410.0	21,444.8
	Standard error	0.019	0.022	0.005	0.015	0.221	0.260	–	–	–
Separate	Estimate	0.640	0.226	0.256	0.460	–	–	– 10781.4	21,570.8	21,609.6
	Standard error	0.019	0.022	0.005	0.014	–	–	–	–	–

The univariate estimates were obtained from the CUB package in R (Iannario et al. 2020)

Table 4 Results of likelihood ratio tests under $H_0^1 : \alpha_B = \alpha_U = 0$, $H_0^2 : \alpha_B = 0$ and $H_0^3 : \alpha_U = 0$

	$\log L(\boldsymbol{r}_k; \hat{\boldsymbol{\Theta}}_0)$	χ^2	df	p-value
H_0^1	− 10764.06	130.11	2	< 0.0001
H_0^2	− 10705.12	12.24	1	0.0004
H_0^3	− 10754.63	111.27	1	< 0.0001

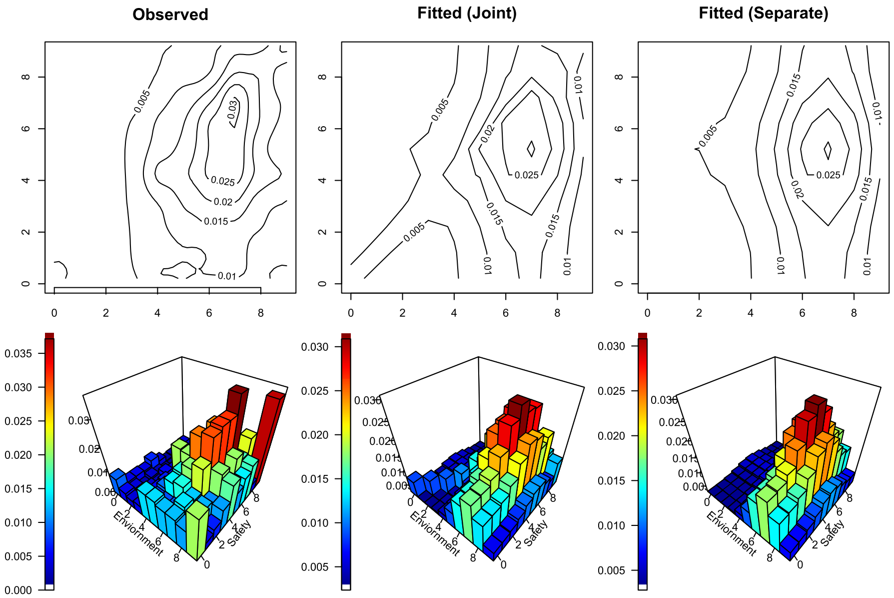

Fig. 3 Contour plots and 3D histograms for the observed data (left) and fitted models (middle: using bivariate CUB model; right: using univariate CUB models fitted separately)

weights π_{00}, π_{01} and π_{10} (π_{11} can be derived from these three), nine for the marginal logits for each of R_1 and R_2 and one for the log odds ratio. In particular, local logits in the form of $\eta_r^j = \log\left[P(R_j = r+1|D_j = 1)/P(R_j = r|D_j = 1)\right]$ for $j = 1, 2$ and $r = 0, 2, \ldots, 8$ were used, and a global odds ratio (Dale 1986)

$$\psi = \frac{P(R_1 \le i, R_2 \le j)P(R_1 > i, R_2 > j)}{P(R_1 > i, R_2 \le j)P(R_1 \le i, R_2 > j)}$$

that is identical for all i and j was used. The use of only one log odds ratio was to ensure model identifiability (Colombi et al. 2019, p. 599). Table 5 shows the estimated values of the parameters and the overall GOF of the model. Not surprisingly, HMMLU provided a better fit in terms of all measures used since it contained substantially more parameters. The relative advantage of the bivariate CUB model lies in parsimony and interpretability.

Table 5 Estimated parameters under HMMLU with $h_i(r_i)$ taking the form of discrete uniform distribution, and the overall GOF of the model

Mixing weights	π_{00}	π_{01}	π_{10}	π_{11}					
	5.31×10^{-6}	6.70×10^{-6}	0.446	0.594					
Logits for R_1	η_0^1	η_1^1	η_2^1	η_3^1	η_4^1	η_5^1	η_6^1	η_7^1	η_8^1
	−0.823	0.439	0.454	0.915	0.242	0.283	0.280	−0.452	−0.013
Logits for R_2	η_0^2	η_1^2	η_2^2	η_3^2	η_4^2	η_5^2	η_6^2	η_7^2	η_8^2
	−0.756	0.488	0.336	0.657	−0.328	0.111	−0.123	−0.906	0.658
Odds ratio	$\log(\psi)$								
	4.357								
GOF measures	$\log L$	AIC	BIC						
	−10572.4	21188.8	21316.5						

7 Discussions and conclusion

In this research work, we have proposed a novel bivariate CUB model for modelling correlated ordinal variables, especially those arising from surveys that require people to rate or express their opinions on a Likert scale. The joint distribution belongs to a general class of mixture distributions while the marginal variables belong to the CUB distribution. Combining the two CUB variables facilitates further insights, such as the association between the two variables, to be drawn from the dataset. Identifiability and other inferential issues around the proposed model have been discussed throughout the paper. The estimation procedure has been found to work satisfactorily through simulation studies. Additional simulation studies under varied scenarios would enhance comprehension of the model's performance. Upon applying the proposed model to a set of publicly available data, we have demonstrated the capability of the model in analysing two variables jointly instead of separately, and how further insights on the associations of the survey items could be discovered.

Since responses from surveys involve psychological behaviours of the respondents, it is important to take into account the potential biases that may have been introduced. Apart from indecision or uncertainty, the uncertainty component of the CUB model can also be used to account for other elements such as difficulty in expressing an actual feeling, limited knowledge, fatigue or willingness to satisfy the interviewer (Iannario and Piccolo 2016; Iannario and Tarantola 2023). As shown by Colombi et al. (2019), the ignorance of the uncertainty component during the modelling stage would lead to substantial biases in the estimation results.

One distinctive feature of the proposed model is the ability to estimate the associations within the uncertainty and feeling components separately. Previous attempts to generalise CUB models to the multivariate setting typically rely on copula-based methods, in particular the Plackett distribution (Corduas 2011; Andreis and Ferrari 2013; Corduas 2015). Another notable work by Colombi and Giordano (2016) used Sarmanov distribution to bind the univariate margins. Both the Plackett and Sarmanov distributions have a parameter that is related to either the rank or Pearson correlation of the two marginal variables. However, it is not possible to tell whether the correlation results from the uncertainty or the feeling component of the underlying CUB variables. Our proposal, on the other hand, allows the decomposition of the overall correlation into two separate elements. In particular, the estimated correlation between the respondents' feelings/preferences would be considered an important measure in many applications. Although Colombi et al. (2019) do not use copula, it assumes independence between the uncertain responses.

One of the reasons why CUB models have become popular is the ability to include respondents' covariates in the model, enabling analysts to explore the relationship between the CUB parameters and the subjects' covariates for better interpretation. Under the proposed bivariate CUB model, we conjecture that it would be straightforward to include covariates for the uncertainty parameters π. However, it may be challenging to include covariates for the feeling parameters ξ, as the admissible range of α_B [which is a function of ξ_1 and ξ_2 as provided in Eq. (4)] will then be affected by the covariates. Re-parameterising α_B could be a way to overcome this challenge, but it is unclear at this stage how this would affect the likelihood function and the mechanism of the EM algorithm introduced in this paper. Further studies are needed to devise a solution. Nonetheless, we have purposely not considered models with covariates since the identifiability has not been established. In fact, to the best of our knowledge, we are not aware of any work that has fully tackled the identifiability issue even for univariate CUB models with covariates.

Our proposed model can be extended in several ways. For example, inclusion of "shelter"/ "refuge" (Iannario 2012) or "don't know" category (Manisera and Zuccolotto 2014; Iannario et al. 2018) would be a direction for future research. Assuming identifiability is not an issue, other bivariate binomial and discrete uniform distributions would replace those utilised in this work. In the univariate case, Gottard et al. (2016) provide details of some other distributions that could be used to replace the uniform distribution in the uncertainty part. Building a bivariate model using these distributions would potentially lead to models that are more interpretable under certain contexts.

In this work, we have focused on the bivariate case. The model developed will serve as a building block for higher dimensional models. As the dependency structure becomes more complicated, the number of parameters will inevitably increase as well. Our proposed bivariate model would be useful if some pairwise dependence or Markov assumptions are to be imposed. These assumptions are particularly suitable for time series (Varin and Vidoni 2006) or spatial ordinal data (Feng et al. 2014; Ip and Wu 2024). More parameters will also mean a higher complexity of the observed information matrix. In that case, the empirical information matrix (Meilijson 1989; McLachlan and Peel 2000; Scott 2002), which requires only the first derivatives, can be used to ease the laborious burden in obtaining the second derivatives.

Appendix A: Moments of *U*, *T* and *R*

The covariance between U_1 and U_2 is given as

$$
\begin{aligned}
\mathrm{Cov}(U_1, U_2) &= E(U_1U_2) - E(U_1)E(U_2)\\
&= \sum_{u_1,u_2} u_1u_2P(U_1 = u_1, U_2 = u_2) - \left(\frac{m}{2}\right)^2\\
&= \frac{1+\alpha_u}{(m+1)^2}\sum_{u=0}^{m} u^2 + \frac{m-\alpha_U}{m(m+1)^2}\sum_{u_1\neq u_2} u_1u_2 - \left(\frac{m}{2}\right)^2\\
&= \frac{1+\alpha_u}{(m+1)^2}\frac{m(m+1)(2m+1)}{6}\\
&\quad + \frac{m-\alpha_U}{m(m+1)^2}\left[\frac{m^2(m+1)^2}{4} - \frac{m(m+1)(2m+1)}{6}\right] - \left(\frac{m}{2}\right)^2\\
&= \frac{\alpha_U(m+2)}{12}.
\end{aligned}
$$

Since

$$\operatorname{Var}(U_1) = \operatorname{Var}(U_2) = \frac{m(m+2)}{12},$$

we have

$$\operatorname{Corr}(U_1, U_2) = \frac{\alpha_U}{m}.$$

The covariance and correlation between T_1 and T_2 provided below are special cases of those presented in Biswas and Hwang (2002). To obtain the covariance between T_1 and T_2, observe that

$$\begin{aligned} E(T_{1i}T_{2i}) &= E\left[T_{1i}E(T_{2i}|T_{1i})\right] = E\left[T_{1i}\left(\frac{1-\xi_2+\alpha_B(\xi_1-\xi_2)+\alpha_B T_{1i}}{1+\alpha_B}\right)\right] \\ &= \left(\frac{1-\xi_2+\alpha_B(\xi_1-\xi_2)+\alpha_B}{1+\alpha_B}\right)(1-\xi_1), \end{aligned}$$

meaning that

$$\begin{aligned} \operatorname{Cov}(T_{1i}, T_{2i}) &= \left(\frac{1-\xi_2+\alpha_B(\xi_1-\xi_2)+\alpha_B}{1+\alpha_B}\right)(1-\xi_1) - (1-\xi_1)(1-\xi_2) \\ &= \frac{\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1). \end{aligned}$$

Thus, the covariance between T_1 and T_2 is

$$\begin{aligned} \operatorname{Cov}(T_1, T_2) &= \operatorname{Cov}\left(\sum_{i=1}^m T_{1i}, \sum_{i=1}^m T_{2i}\right) = \sum_{i=1}^m \operatorname{Cov}(T_{1i}, T_{2i}) \\ &= \frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1), \end{aligned}$$

leading to a correlation of

$$\operatorname{Corr}(T_1, T_2) = \frac{\alpha_B}{1+\alpha_B}\sqrt{\frac{\xi_1(1-\xi_1)}{\xi_2(1-\xi_2)}}.$$

The mean and variance of R_i, $i = 1, 2$, can be derived as follows.

$$
\begin{aligned}
E(R_i) &= \sum_{r=0}^{m} r[(1-\pi_i)\frac{1}{m+1} + \pi C_r^m (1-\xi_i)^r \xi_i^{m-r}] \\
&= (1-\pi_i)\frac{m}{2} + \pi_i m(1-\xi_i) \\
E(R_i^2) &= \sum_{r=0}^{m} r^2[(1-\pi_i)\frac{1}{m+1} + \pi_i C_r^m (1-\xi_i)^r \xi_i^{m-r}] \\
&= (1-\pi_i)\frac{m(2m+1)}{6} + \pi_i m(1-\xi_i)[m(1-\xi_i)+\xi_i] \\
\mathrm{Var}(R_i) &= (1-\pi_i)\frac{m(2m+1)}{6} + \pi_i[m^2(1-\xi_i)^2 + m(1-\xi_i)\xi_i] \\
&\quad - \left[(1-\pi_i)\frac{m}{2} + \pi_i m(1-\xi_i)\right]^2 \\
&= (1-\pi_i)m\left[\frac{2m+1}{6} - \frac{(1-\pi_i)m}{4}\right] + \pi_i m(1-\xi_i)\xi_i[1-m(1-\pi_i)]
\end{aligned}
$$

The covariance between R_1 and R_2 can be derived as follows. Since

$$
\begin{aligned}
E(R_1R_2) = {} & (1-\pi_1)(1-\pi_2)\frac{\alpha_U(m+2)+3m^2}{12} + (1-\pi_1)\pi_2\left(\frac{m}{2}\right)m(1-\xi_2) \\
& +\pi_1(1-\pi_2)\left(\frac{m}{2}\right)m(1-\xi_1) \\
& +\pi_1\pi_2\left[\frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1) + m^2(1-\xi_1)(1-\xi_2)\right],
\end{aligned}
$$

we have

$$
\begin{aligned}
\mathrm{Cov}(R_1, R_2) = {} & (1-\pi_1)(1-\pi_2)\frac{\alpha_U(m+2)+3m^2}{12} + (1-\pi_1)\pi_2\left(\frac{m}{2}\right)m(1-\xi_2) \\
& +\pi_1(1-\pi_2)\left(\frac{m}{2}\right)m(1-\xi_1) \\
& +\pi_1\pi_2\left[\frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1) + m^2(1-\xi_1)(1-\xi_2)\right] \\
& -\left[(1-\pi_1)\frac{m}{2} + \pi_1 m(1-\xi_1)\right]\left[(1-\pi_2)\frac{m}{2} + \pi_2 m(1-\xi_2)\right] \\
= {} & (1-\pi_1)(1-\pi_2)\frac{\alpha_U(m+2)}{12} + \pi_1\pi_2\frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1).
\end{aligned}
$$

Finally, the correlation between R_1 and R_2 can be found using

$$
\mathrm{Corr}(R_1, R_2) = \frac{(1-\pi_1)(1-\pi_2)\frac{\alpha_U(m+2)}{12} + \pi_1\pi_2\frac{m\alpha_B}{1+\alpha_B}\xi_1(1-\xi_1)}{\sqrt{Var(R_1)}\sqrt{Var(R_2)}}.
$$

Appendix B: EM algorithm

Assume n pairs of (r_{1k}, r_{2k}) are observed. For the sake of notational simplicity, let $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)'$, $\boldsymbol{r}_k = (r_{1k}, r_{2k})$, $k = 1, 2, \ldots, n$, $q_1 = 1 - \pi_1$, $q_2 = 1 - \pi_2$, $g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}) = U_{12}(\boldsymbol{r}_k, \alpha_U)$, $g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta}) = B_2(r_{2k})/(m+1)$, $g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}) = B_1(r_{1k})/(m+1)$, and $g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}) = B_{12}(\boldsymbol{r}_k, \xi_1, \xi_2, \alpha_B)$. The aim is to maximise

$$\begin{aligned} L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) = {} & q_1 q_2 g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) + q_1 \pi_2 g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) + \pi_1 q_2 g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\ & + \pi_1 \pi_2 g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}), \end{aligned}$$

which could be done using the iterative procedure described below.

1. Get initial values of $\pi_1^{(0)}, \pi_2^{(0)}, \xi_1^{(0)}, \xi_2^{(0)}$ by considering the margins separately. This step can be done using the CUB package in R (Iannario et al. 2020). Also, set $\alpha_U^{(0)} = \alpha_B^{(0)}$=0.
2. Get the posterior probabilities $\hat{p}_{ij}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) = P(D_1 = i, D_2 = j | R_1 = r_{1k}, R_2 = r_{2k}; \boldsymbol{\theta}^{(0)})$, $i, j = 0, 1$, $k = 1, 2, \ldots, n$, based on each individual paired observation $\boldsymbol{r}_k$:

$$\begin{aligned} \hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) &= (1 - \pi_1^{(0)})(1 - \pi_2^{(0)}) g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) / L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\ \hat{p}_{01}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) &= (1 - \pi_1^{(0)}) \pi_2^{(0)} g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) / L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\ \hat{p}_{10}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) &= \pi_1^{(0)} (1 - \pi_2^{(0)}) g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) / L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\ \hat{p}_{11}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) &= \pi_1^{(0)} \pi_2^{(0)} g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) / L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}), \end{aligned}$$

   which give the overall estimates of

$$\hat{p}_{ij}^{(0)} = \frac{1}{n} \sum_{k=1}^{n} \hat{p}_{ij}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}), \qquad i, j = 0, 1.$$

   From the second iterations ($t > 1$) onward, the above estimates can be obtained from

$$\begin{aligned} \hat{p}_{ij}^{(t+1)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t)}) &= \hat{p}_{ij}^{(t)} g_{ij}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t)}) / L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t)}), \quad \text{and} \\ \hat{p}_{ij}^{(t+1)} &= \frac{1}{n} \sum_{k=1}^{n} \hat{p}_{ij}^{(t+1)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t)}), \quad t > 1. \end{aligned}$$

3. Update π_1 and π_2 through

$$\pi_1^{(1)} = \hat{p}_{10}^{(0)} + \hat{p}_{11}^{(0)}, \tag{B1}$$

$$\pi_2^{(1)} = \hat{p}_{01}^{(0)} + \hat{p}_{11}^{(0)}. \tag{B2}$$

It can be shown that Eqs. (B1) and (B2) provide the best estimate of π_1 and π_2, respectively. The steps below largely follow those provided in Everitt and Hand (1981), except that our mixing probabilities are constrained in a different manner. Our objective is to maximise

$$\ell(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log[L(\boldsymbol{r}_k; \boldsymbol{\theta})] - \lambda_1(\pi_1 + q_1 - 1) - \lambda_2(\pi_2 + q_2 - 1),$$

where λ_1 and λ_2 are Lagrange multipliers corresponding to the constraints $q_1 + \pi_1 = 1$ and $q_2 + \pi_2 = 1$, respectively. Differentiating $\ell(\boldsymbol{\theta})$ with respect to π_1, and setting the equation to 0 yields

$$\frac{\partial}{\partial \pi_1}\ell(\boldsymbol{\theta}) = \sum_{k=1}^{n} \frac{q_2 g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta})}{L(\boldsymbol{r}_k; \boldsymbol{\theta})} + \sum_{k=1}^{n} \frac{\pi_2 g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta})}{L(\boldsymbol{r}_k; \boldsymbol{\theta})} - \lambda_1 = 0. \tag{B3}$$

Similarly, we have

$$\frac{\partial}{\partial q_1}\ell(\boldsymbol{\theta}) = \sum_{k=1}^{n} \frac{q_2 g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta})}{L(\boldsymbol{r}_k; \boldsymbol{\theta})} + \sum_{k=1}^{n} \frac{\pi_2 g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta})}{L(\boldsymbol{r}_k; \boldsymbol{\theta})} - \lambda_1 = 0. \tag{B4}$$

Multiplying (B3) by π_1 and (B4) by q_1, and adding them up:

$$\begin{aligned}
&\sum_{k=1}^{n} \frac{\pi_1 q_2 g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}) + \pi_1 \pi_2 g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}) + q_1 q_2 g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}) + q_1 \pi_2 g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta})}{L(\boldsymbol{r}_k; \boldsymbol{\theta})} \\
&\quad -\lambda_1(\pi_1 + q_1) = 0 \\
&\hat{\lambda}_1 = n.
\end{aligned}$$

Suppose one has $\boldsymbol{\theta}^{(0)}$, using $\hat{\lambda}_1 = n$ and multiplying (B3) by π_1 yields

$$\sum_{k=1}^{n} \frac{\pi_1 q_2^{(0)} g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) + \pi_1 \pi_2^{(0)} g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})}{L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})} - n\pi_1 = 0.$$

Thus,

$$\hat{\pi}_1 = \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\pi_1^{(0)}(1 - \pi_2^{(0)}) g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})}{L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})} + \frac{\pi_1^{(0)} \pi_2^{(0)} g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})}{L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)})} \right],$$

where the RHS is equivalent to the posterior probability $P(D_1 = 1 | R_1 = r_1, R_2 = r_2; \boldsymbol{\theta}^{(0)})$. The derivation of π_2 follows in a similar fashion.

4. Update ξ_1 and ξ_2 using

$$
\begin{aligned}
\xi_1^{(1)} &= \underset{\xi_1 \in \Xi_1}{\operatorname{argmax}} \sum_{k=1}^{n} \Big[\hat{p}_{10}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{10}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\
&\quad + \hat{p}_{11}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \Big], \\
\xi_2^{(1)} &= \underset{\xi_2 \in \Xi_2}{\operatorname{argmax}} \sum_{k=1}^{n} \Big[\hat{p}_{01}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{01}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \\
&\quad + \hat{p}_{11}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \Big],
\end{aligned}
$$

where Ξ_1 and Ξ_2 are the ranges for $\xi_1^{(1)}$ and $\xi_2^{(1)}$, respectively, to ensure w_1, w_2, w_3 and w_4 are all positive, provided $\boldsymbol{\theta}^{(0)}$. Explicitly, the corresponding ranges are

$$
\Xi_1 = \begin{cases}
\Bigg(\max\left\{0, \frac{-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)} - \alpha_B^{(0)}}{-\alpha_B^{(0)}}\right\}, \\
\quad \min\left\{\xi_2^{(0)}, \frac{1-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)}}{-\alpha_B^{(0)}}, \frac{(1+\alpha_B^{(0)})(\xi_2^{(0)}-1)}{\alpha_B^{(0)}}\right\}\Bigg), & \text{if } \xi_1^{(0)} < \xi_2^{(0)} \,\&\, \alpha_B^{(0)} < 0; \\
\Bigg(\max\left\{0, \frac{1-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)}}{-\alpha_B^{(0)}}, \frac{(1+\alpha_B^{(0)})(\xi_2^{(0)}-1)}{\alpha_B^{(0)}}\right\}, \\
\quad \min\left\{\xi_2^{(0)}, \frac{-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)} - \alpha_B^{(0)}}{-\alpha_B^{(0)}}\right\}\Bigg), & \text{if } \xi_1^{(0)} < \xi_2^{(0)} \,\&\, \alpha_B^{(0)} > 0; \\
\Bigg(\max\left\{\frac{-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)} - \alpha_B^{(0)}}{-\alpha_B^{(0)}}, \frac{\xi_2^{(0)} + \alpha_B^{(0)}\xi_2^{(0)}}{\alpha_B^{(0)}}, \xi_2^{(0)}\right\}, \\
\quad \min\left\{\frac{(1+\alpha_B^{(0)})(\xi_2^{(0)}-1)}{\alpha_B^{(0)}}, 1\right\}\Bigg), & \text{if } \xi_1^{(0)} > \xi_2^{(0)} \,\&\, \alpha_B^{(0)} < 0; \\
\Bigg(\max\left\{\xi_2^{(0)}, \frac{(1+\alpha_B^{(0)})(\xi_2^{(0)}-1)}{\alpha_B^{(0)}}\right\}, \\
\quad \min\left\{\frac{-\xi_2^{(0)} - \alpha_B^{(0)}\xi_2^{(0)} - \alpha_B^{(0)}}{-\alpha_B^{(0)}}, \frac{\xi_2^{(0)} + \alpha_B^{(0)}\xi_2^{(0)}}{\alpha_B^{(0)}}, 1\right\}\Bigg), & \text{if } \xi_1^{(0)} > \xi_2^{(0)} \,\&\, \alpha_B^{(0)} > 0; \\
(0, 1), & \text{if } \alpha_B^{(0)} = 0, \text{ and}
\end{cases}
$$

$$
\Xi_2 = \begin{cases}
\Bigg(\max\left\{\frac{-(1-\xi_1^{(0)})\alpha_B^{(0)}}{1+\alpha_B^{(0)}}, \xi_1^{(0)}\right\}, \\
\quad \min\left\{\frac{1+\xi_1^{(0)}\alpha_B^{(0)}}{1+\alpha_B^{(0)}}, \frac{1+\alpha_B^{(0)}+\alpha_B^{(0)}\xi_1^{(0)}}{1+\alpha_B^{(0)}}, 1\right\}\Bigg), & \text{if } \xi_1^{(0)} < \xi_2^{(0)} \,\&\, \alpha_B^{(0)} \neq 0; \\
\Bigg(\max\left\{0, \frac{-(1-\xi_1^{(0)})\alpha_B^{(0)}}{1+\alpha_B^{(0)}}, \frac{\xi_1^{(0)}\alpha_B^{(0)}}{1+\alpha_B^{(0)}}\right\}, \\
\quad \min\left\{\xi_1^{(0)}, \frac{1+\alpha_B^{(0)}+\alpha_B^{(0)}\xi_1^{(0)}}{1+\alpha_B^{(0)}}\right\}\Bigg), & \text{if } \xi_1^{(0)} > \xi_2^{(0)} \,\&\, \alpha_B^{(0)} \neq 0; \\
(0, 1), & \text{if } \alpha_B^{(0)} = 0.
\end{cases}
$$

5. Update α_B using

$$\alpha_B^{(1)} = \underset{\alpha_B \in \mathcal{A}_B}{\operatorname{argmax}} \sum_{k=1}^{n} \hat{p}_{11}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{11}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}),$$

where $\mathcal{A}_B$ is the admissible range of α_B based on $\xi_1^{(0)}$ and $\xi_2^{(0)}$ as provided in (4).

6. Update α_U using

$$\alpha_U^{(1)} = \frac{mS_1 - S_2}{S_1 + S_2}, \tag{B5}$$

where

$$S_1 = \sum_{k=1}^{n} \hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) 1_{r_{1k}=r_{2k}} \quad \text{and} \quad S_2 = \sum_{k=1}^{n} \hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) 1_{r_{1k}\neq r_{2k}}.$$

To see that (B5) gives the best estimate for α_U, notice that our aim is to find

$$\alpha_U^{(1)} = \underset{\alpha_U \in [-1,m]}{\operatorname{argmax}} \sum_{k=1}^{n} \left[\hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \right].$$

Expanding $\sum_{k=1}^{n} \left[\hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log g_{00}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \right]$ gives

$$\sum_{k=1, r_{1k}=r_{2k}}^{n} \left[\hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log \left(\frac{m + m\alpha_U}{m(m+1)^2} \right) \right] + \sum_{k=1, r_{1k}\neq r_{2k}}^{n} \left[\hat{p}_{00}^{(0)}(\boldsymbol{r}_k; \boldsymbol{\theta}^{(0)}) \log \left(\frac{m - \alpha_U}{m(m+1)^2} \right) \right]. \tag{B6}$$

Upon differentiating (B6) with respect to α_U, we have

$$\frac{S_1}{1 + \alpha_U} - \frac{S_2}{m - \alpha_U} = 0,$$

which gives (B5).

7. Calculate $\log L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(1)})$. Repeat Steps 2 to 7 until L converges, that is, $\log L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t+1)}) - \log L(\boldsymbol{r}_k; \boldsymbol{\theta}^{(t)}) < \varepsilon$ for some threshold ε.

Appendix C: Detailed expressions for the information matrix

In this appendix, to simplify the notation, define $f^{\theta} = \frac{\partial}{\partial \theta} f$ and $f^{\theta_i \theta_j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} f$ for some expressions f. Whenever there is no chance of causing confusion, the dependency of the functions on the parameters and/or data are often omitted. For instance, we write g_{00} instead of $g_{00}(\boldsymbol{\theta}, \boldsymbol{r}_k)$. In addition, the index j in $\sum_j$ runs from 0 to r_1, and the index k in $\sum_k$ runs from 1 to n. We first list some recurrent expressions.

$$
\begin{aligned}
L_k &= L(\boldsymbol{\theta}, \boldsymbol{r}_k) = (1-\pi_1)(1-\pi_2)g_{00} + (1-\pi_1)\pi_2 g_{01} \\
&\quad +\pi_1(1-\pi_2)g_{10} + \pi_1\pi_2 g_{11} \\
B_i &= C^m_{r_i}(1-\xi_i)^{r_i}\xi_i^{m-r_i}, \qquad i = 1, 2 \\
B_i^{\xi_i} &= B_i\left(\frac{m-r_i}{\xi_i} - \frac{r_i}{1-\xi_i}\right) \\
B_i^{\xi_i\xi_i} &= B_i\left[\frac{-(m-r_i)}{\xi_i^2} - \frac{r_i}{(1-\xi_i)^2} + \left(\frac{m-r_i}{\xi_i} - \frac{r_i}{1-\xi_i}\right)^2\right] \\
W_j &= w_1^j w_2^{r_1-j} w_3^{r_2-j} w_4^{m-r_1-r_2+j} \\
W_j^{\xi_1} &= W_j\left[\frac{j\alpha_B}{w_1} + \frac{-(r_1-j)\alpha_B}{w_2} + \frac{(r_2-j)\alpha_B}{w_3} + \frac{-(m-r_1-r_2+j)\alpha_B}{w_4}\right] \\
W_j^{\xi_2} &= W_j\left[\frac{-j(1+\alpha_B)}{w_1} + \frac{(r_1-j)(1+\alpha_B)}{w_2} + \frac{-(r_2-j)(1+\alpha_B)}{w_3}\right. \\
&\quad \left. + \frac{(m-r_1-r_2+j)(1+\alpha_B)}{w_4}\right] \\
W_j^{\alpha_B} &= W_j\left[\frac{j(1+\xi_1-\xi_2)}{w_1} + \frac{-(r_1-j)(\xi_1-\xi_2)}{w_2} + \frac{(r_2-j)(\xi_1-\xi_2)}{w_3}\right. \\
&\quad \left. + \frac{(m-r_1-r_2+j)(1-\xi_1+\xi_2)}{w_4}\right] \\
W_j^{\xi_1\xi_1} &= W_j\left[\frac{-j\alpha_B^2}{w_1^2} + \frac{-(r_1-j)\alpha_B^2}{w_2^2} + \frac{-(r_2-j)\alpha_B^2}{w_3^2} + \frac{-(m-r_1-r_2+j)\alpha_B^2}{w_4^2}\right] \\
&\quad + \frac{W_j^{\xi_1} W_j^{\xi_1}}{W_j} \\
W_j^{\xi_1\xi_2} &= W_j\left[\frac{j\alpha_B(1+\alpha_B)}{w_1^2} + \frac{(r_1-j)\alpha_B(1+\alpha_B)}{w_2^2} + \frac{(r_2-j)\alpha_B(1+\alpha_B)}{w_3^2}\right. \\
&\quad \left. + \frac{(m-r_1-r_2+j)\alpha_B(1+\alpha_B)}{w_4^2}\right] \\
&\quad + \frac{W_j^{\xi_2} W_j^{\xi_1}}{W_j} \\
W_j^{\xi_1\alpha_B} &= W_j\left[\frac{j(w_1-\alpha_B(1+\xi_1-\xi_2))}{w_1^2} + \frac{-(r_1-j)(w_2+\alpha_B(\xi_1-\xi_2))}{w_2^2}\right. \\
&\quad + \frac{(r_2-j)(w_3-\alpha_B(\xi_1-\xi_2))}{w_3^2} \\
&\quad \left. + \frac{(m-r_1-r_2+j)(-w_4+\alpha_B(1-\xi_1+\xi_2))}{w_4^2}\right] + \frac{W_j^{\alpha_B} W_j^{\xi_1}}{W_j}
\end{aligned}
$$

$$\begin{aligned}
W_j^{\xi_2\xi_2} &= W_j\left[\frac{-j(1+\alpha_B)^2}{w_1^2}+\frac{-(r_1-j)(1+\alpha_B)^2}{w_2^2}+\frac{-(r_2-j)(1+\alpha_B)^2}{w_3^2}\right.\\
&\quad\left.+\frac{-(m-r_1-r_2+j)(1+\alpha_B)^2}{w_4^2}\right]+\frac{W_j^{\xi_2}W_j^{\xi_2}}{W_j}\\
W_j^{\xi_2\alpha_B} &= W_j\left[\frac{j(-w_1+(1+\alpha_B)(1+\xi_1-\xi_2))}{w_1^2}+\frac{(r_1-j)(w_2+(1+\alpha_B)(\xi_1-\xi_2))}{w_2^2}\right.\\
&\quad+\frac{(r_2-j)(-w_3+(1+\alpha_B)(\xi_1-\xi_2))}{w_3^2}\\
&\quad\left.+\frac{(m-r_1-r_2+j)(w_4-(1+\alpha_B)(1-\xi_1+\xi_2))}{w_4^2}\right]\\
&\quad+\frac{W_j^{\alpha_B}W_j^{\xi_2}}{W_j}\\
W_j^{\alpha_B\alpha_B} &= W_j\left[\frac{-j(1+\xi_1-\xi_2)^2}{w_1^2}+\frac{-(r_1-j)(\xi_1-\xi_2)^2}{w_2^2}+\frac{-(r_2-j)(\xi_1-\xi_2)^2}{w_3^2}+\right.\\
&\quad\left.\frac{-(m-r_1-r_2+j)(1-\xi_1+\xi_2)^2}{w_4^2}\right]+\frac{W_j^{\alpha_B}W_j^{\alpha_B}}{W_j}\\
g_{00}^{\alpha_U} &= \frac{m1_{r_1=r_2}-1_{r_1\neq r_2}}{m(m+1)^2}\\
g_{10}^{\xi_1} &= \frac{B_1^{\xi_1}}{m+1}\\
g_{10}^{\xi_1\xi_1} &= \frac{B_1^{\xi_1\xi_1}}{m+1}\\
g_{01}^{\xi_2} &= \frac{B_2^{\xi_2}}{m+1}\\
g_{01}^{\xi_2\xi_2} &= \frac{B_2^{\xi_2\xi_2}}{m+1}\\
g_{11}^{\xi_1} &= (1+\alpha_B)^{-m}\left[B_1^{\xi_1}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j+B_1\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_1}\right]\\
g_{11}^{\xi_2} &= B_1(1+\alpha_B)^{-m}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_2}\\
g_{11}^{\alpha_B} &= B_1(1+\alpha_B)^{-m-1}\left[-m\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j+(1+\alpha_B)\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\alpha_B}\right]\\
g_{11}^{\xi_1\xi_1} &= (1+\alpha_B)^{-m}\left[B_1\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_1\xi_1}\right.\\
&\quad\left.+2B_1^{\xi_1}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_1}+B_1^{\xi_1\xi_1}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j\right]
\end{aligned}$$

$$
\begin{aligned}
g_{11}^{\xi_1\xi_2} &= (1+\alpha_B)^{-m}\left[B_1^{\xi_1}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_2} + B_1\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_1\xi_2}\right]\\
g_{11}^{\xi_1\alpha_B} &= (1+\alpha_B)^{-m}\left[B_1^{\xi_1}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\alpha_B} + B_1\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_1\alpha_B}\right]\\
&\quad -m(1+\alpha_B)^{-1}g_{11}^{\xi_1}\\
g_{11}^{\xi_2\xi_2} &= B_1(1+\alpha_B)^{-m}\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_2\xi_2}\\
g_{11}^{\xi_2\alpha_B} &= B_1(1+\alpha_B)^{-m-1}\\
&\quad\left[-m\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_2} + (1+\alpha_B)\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\xi_2\alpha_B}\right]\\
g_{11}^{\alpha_B\alpha_B} &= B_1(1+\alpha_B)^{-m-2}\left[m(m+1)\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j\right.\\
&\quad -2m(1+\alpha_B)\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\alpha_B}\\
&\quad\left.+(1+\alpha_B)^2\sum_j C_j^{r_1}C_{r_2-j}^{m-r_1}W_j^{\alpha_B\alpha_B}\right]\\
L_k^{\pi_1} &= -(1-\pi_2)g_{00} - \pi_2 g_{01} + (1-\pi_2)g_{10} + \pi_2 g_{11}\\
L_k^{\pi_2} &= -(1-\pi_1)g_{00} + (1-\pi_1)g_{01} - \pi_1 g_{10} + \pi_1 g_{11}\\
L_k^{\xi_1} &= \pi_1(1-\pi_2)g_{10}^{\xi_1} + \pi_1\pi_2 g_{11}^{\xi_1}\\
L_k^{\xi_2} &= (1-\pi_1)\pi_2 g_{01}^{\xi_2} + \pi_1\pi_2 g_{11}^{\xi_2}\\
L_k^{\alpha_U} &= (1-\pi_1)(1-\pi_2)g_{00}^{\alpha_U}\\
L_k^{\alpha_B} &= \pi_1\pi_2 g_{11}^{\alpha_B}
\end{aligned}
$$

The elements of the negative information matrix, $J(\boldsymbol{\theta}) = \Delta^2 \log L(\boldsymbol{\theta})$, is shown below, where the order of the parameters follows: $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi_1, \xi_2, \alpha_U, \alpha_B)'$.

$$
\begin{aligned}
\{J\}_{11} &= \sum_k \frac{-L_k^{\pi_1}L_k^{\pi_1}}{L_k L_k}\\
\{J\}_{12} &= \sum_k \frac{L_k(g_{00}-g_{01}-g_{10}+g_{11}) - L_k^{\pi_1}L_k^{\pi_2}}{L_k L_k}\\
\{J\}_{13} &= \sum_k \frac{L_k((1-\pi_2)g_{10}^{\xi_1} + \pi_2 g_{11}^{\xi_1}) - L_k^{\pi_1}L_k^{\xi_1}}{L_k L_k}\\
\{J\}_{14} &= \sum_k \frac{L_k(-\pi_2 g_{01}^{\xi_2} + \pi_2 g_{11}^{\xi_2}) - L_k^{\pi_1}L_k^{\xi_2}}{L_k L_k}
\end{aligned}
$$

$$\{J\}_{15} = \sum_k \frac{L_k(-(1-\pi_2)g_{00}^{\alpha_U}) - L_k^{\pi_1} L_k^{\alpha_U}}{L_k L_k}$$

$$\{J\}_{16} = \sum_k \frac{L_k \pi_2 g_{11}^{\alpha_B} - L_k^{\pi_1} L_k^{\alpha_B}}{L_k L_k}$$

$$\{J\}_{22} = \sum_k \frac{-L_k^{\pi_2} L_k^{\pi_2}}{L_k L_k}$$

$$\{J\}_{23} = \sum_k \frac{L_k(-\pi_1 g_{10}^{\xi_1} + \pi_1 g_{11}^{\xi_1}) - L_k^{\pi_2} L_k^{\xi_1}}{L_k L_k}$$

$$\{J\}_{24} = \sum_k \frac{L_k((1-\pi_1)g_{01}^{\xi_2} + \pi_1 g_{11}^{\xi_2}) - L_k^{\pi_2} L_k^{\xi_2}}{L_k L_k}$$

$$\{J\}_{25} = \sum_k \frac{L_k(-(1-\pi_1)g_{00}^{\alpha_U}) - L_k^{\pi_2} L_k^{\alpha_U}}{L_k L_k}$$

$$\{J\}_{26} = \sum_k \frac{L_k \pi_1 g_{11}^{\alpha_B} - L_k^{\pi_2} L_k^{\alpha_B}}{L_k L_k}$$

$$\{J\}_{33} = \sum_k \frac{L_k(\pi_1(1-\pi_2)g_{10}^{\xi_1\xi_1} + \pi_1\pi_2 g_{11}^{\xi_1\xi_1}) - L_k^{\xi_1} L_k^{\xi_1}}{L_k L_k}$$

$$\{J\}_{34} = \sum_k \frac{L_k \pi_1\pi_2 g_{11}^{\xi_1\xi_2} - L_k^{\xi_1} L_k^{\xi_2}}{L_k L_k}$$

$$\{J\}_{35} = \sum_k \frac{-L_k^{\xi_1} L_k^{\alpha_U}}{L_k L_k}$$

$$\{J\}_{36} = \sum_k \frac{L_k \pi_1\pi_2 g_{11}^{\xi_1\alpha_B} - L_k^{\xi_1} L_k^{\alpha_B}}{L_k L_k}$$

$$\{J\}_{44} = \sum_k \frac{L_k((1-\pi_1)\pi_2 g_{01}^{\xi_2\xi_2} + \pi_1\pi_2 g_{11}^{\xi_2\xi_2}) - L_k^{\xi_2} L_k^{\xi_2}}{L_k L_k}$$

$$\{J\}_{45} = \sum_k \frac{-L_k^{\xi_2} L_k^{\alpha_U}}{L_k L_k}$$

$$\{J\}_{46} = \sum_k \frac{L_k \pi_1\pi_2 g_{11}^{\xi_2\alpha_B} - L_k^{\xi_2} L_k^{\alpha_B}}{L_k L_k}$$

$$\{J\}_{55} = \sum_k \frac{-L_k^{\alpha_U} L_k^{\alpha_U}}{L_k L_k}$$

$$\{J\}_{56} = \sum_k \frac{-L_k^{\alpha_U} L_k^{\alpha_B}}{L_k L_k}$$

$$\{J\}_{66} = \sum_k \frac{L_k \pi_1 \pi_2 g_{11}^{\alpha_B \alpha_B} - L_k^{\alpha_B} L_k^{\alpha_B}}{L_k L_k}$$

The lower triangular elements are the same as the upper triangular ones, and are thus omitted.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00362-024-01560-2.

Author contributions All authors contributed to the study conception and design. The first draft was of the manuscript was written by Ryan H. L. Ip and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions No funds, grants, or other support was received.

Data availability The dataset analysed during the current study are available within the CUB package in R (Iannario et al. 2020).

Code availability The code used to produce part of the results in the Application section is available as Supplementary Information.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Ethical approval Ethics approval is not required for this work.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

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

Al-Humairi A, Ip RHL, Spuur K, Zheng X, Huang B (2022) Visual grading experiments and optimization in CBCT dental implantology imaging: preliminary application of integrated visual grading regression. Radiat Enviorn Biophys 61:133–145. https://doi.org/10.1007/s00411-021-00959-x

Anderson JA (1984) Regression and ordered categorical variables. J R Stat Soc B Met 46(1):1–22. https://doi.org/10.1111/j.2517-6161.1984.tb01270.x

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

Arcidiacono P, Jones JB (2003) Finite mixture distributions, sequential likelihood and the EM algorithm. Econometrica 71(3):933–946. https://doi.org/10.1111/1468-0262.00431

Barbiero A (2021) Inducing a desired value of correlation between two point-scale variables: a two-step procedure using copulas. Adv Stat Anal 105:307–334. https://doi.org/10.1007/s10182-021-00405-9

Bartolucci F, Colombi R, Forcina A (2007) An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. Stat Sinica 17(2):691–711

Baumgartner H, Steenkamp JEM (2001) Response styles in marketing research: a cross-national investigation. J Mark Res 38(2):143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Baumgartner H, Steenkamp JEM (2006) Response biases in marketing research. In: Grover R, Vriens M (eds) The handbook of marketing research: uses, misuses, and future advances. SAGE, London

Bergsma WP, Rudas T (2002) Marginal models for categorical data. Ann Stat 30(1):140–159. https://doi.org/10.1214/aos/1015362188

Biswas A, Hwang JS (2002) A new bivariate binomial distribution. Stat Probab Lett 60(2):231–240. https://doi.org/10.1016/S0167-7152(02)00323-1

Capecchi S, Iannario M, Simone R (2018) Well-being and relational goods: a model-based approach to detect significant relationships. Soc Indic Res 135:729–750. https://doi.org/10.1007/s11205-016-1519-7

Colombi R, Giordano S (2016) A class of mixture models for multidimensional ordinal data. Stat Model 16(4):322–340. https://doi.org/10.1177/1471082X16649730

Colombi R, Giordano S, Gottard A, Iannario M (2019) Hierarchical marginal models with latent uncertainty. Scand J Stat 46(2):595–620. https://doi.org/10.1111/sjos.12366

Corduas M (2011) Modelling correlated bivariate ordinal data with CUB margins. Quad Stat 13:109–119

Corduas M (2015) Analyzing bivariate ordinal data with CUB margins. Stat Model 15(5):411–432. https://doi.org/10.1177/1471082X14558770

Dale JR (1986) Global cross-ratio mdoels for bivariate, discrete, ordered responses. Biometrics 42(4):909–917. https://doi.org/10.2307/2530704

D'Elia A, Piccolo D (2005) A mixture model for preferences data analysis. Comput Stat Data Anal 49:917–934. https://doi.org/10.1016/j.csda.2004.06.012

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B Met 39(1):1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Efron B, Hinkley DV (1978) Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. Biometrika 65(3):457–483. https://doi.org/10.1093/biomet/65.3.457

Everitt BS, Hand DJ (1981) Finite mixture distributions. Chapman and Hall, London

Feng X, Zhu J, Lin P, Steen-Adams MM (2014) Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values. Environmetrics 25(8):571–583. https://doi.org/10.1002/env.2306

Gambacorta R, Iannario M (2013) Measuring job satisfaction with CUB models. Labour 27(2):198–224. https://doi.org/10.1111/labr.12008

Geenens G (2020) Copula modeling for discrete random vectors. Depend Model 8:417–440. https://doi.org/10.1515/demo-2020-0022

Genest C, Nešlehová J (2007) A primer on copulas for count data. Astin Bull 37(2):475–515. https://doi.org/10.2143/AST.37.2.2024077

Gottard A, Iannario M, Piccolo D (2016) Varying uncertainty in CUB models. Adv Data Anal Classif 10:225–244. https://doi.org/10.1007/s11634-016-0235-0

Guisan A, Harrell FE (2000) Ordinal response regression models in ecology. J Veg Sci 11(5):617–626. https://doi.org/10.2307/3236568

Hoel PG (1962) Introduction to mathematical statistics. Wiley, New York

Iannario M (2010) On the identifiability of a mixture model for ordinal data. Metron 68(1):87–94. https://doi.org/10.1007/BF03263526

Iannario M (2012) Modelling shelter choices in a class of mixture models for ordinal response. Stat Method Appl 21:1–22. https://doi.org/10.1007/s10260-011-0176-x

Iannario M, Manisera M, Piccolo D, Zuccolotto P (2018) Ordinal data models for no-opinion responses in attitude surveys. Sociol Method Res 49(1):250–276. https://doi.org/10.1177/0049124118769081

Iannario M, Piccolo D (2010) A new statistical model for the analysis of customer satisfaction. Qual Technol Quant Manage 7(2):149–168. https://doi.org/10.1080/16843703.2010.11673225

Iannario M, Piccolo D (2010) Statistical modelling of subjective survival probabilities. Genus 66(2):17–42

Iannario M, Piccolo D (2016) A comprehensive framework of regression models for ordinal data. Metron 74:233–252. https://doi.org/10.1007/s40300-016-0091-x

Iannario M, Piccolo D, Simone R (2020) CUB: a class of mixture models for ordinal data. R package version 1.1.4. https://CRAN.R-project.org/package=CUB

Iannario M, Simone R (2017) Mixture models for rating data: the method of moments via Gröbner basis. J Algebr Stat 8(2):1–28. https://doi.org/10.18409/JAS.V8I2.60
Iannario M, Tarantola C (2023) How to interpret the effect of covariates on the extreme categories in ordinal data models. Sociol Method Res 52(1):231–267. https://doi.org/10.1177/0049124120986179
Ip RHL, Wu KYK (2024) A Markov random field model with cumulative logistic functions for spatially dependent ordinal data. J Appl Stat 51(1):70–86. https://doi.org/10.1080/02664763.2022.2115985
Joshi A, Kale S, Chandel S, Pal DK (2015) Likert scale: explored and explained. Brit J Appl Sci Technol 7(4):157. https://doi.org/10.9734/BJAST/2015/14975
Krosnick JA (1999) Survey research. Annu Rev Psychol 50:537–567. https://doi.org/10.1146/annurev.psych.50.1.537
Krosnick JA, Alwin DF (1987) An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opin Q 51(2):201–219. https://doi.org/10.1086/269029
Manisera M, Zuccolotto P (2014) Modeling "don't know" responses in rating scales. Pattern Recogn Lett 45:226–234. https://doi.org/10.1016/j.patrec.2014.04.012
Manisera M, Zuccolotto P (2015) Identifiability of a model for discrete frequency distributions with a multi-dimensional parameter space. J Multivar Anal 140:302–316. https://doi.org/10.1016/j.jmva.2015.05.011
McCullagh P (1980) Regression models for ordinal data. J R Stat Soc B Met 42:109–142. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x
McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York
McLachlan GJ, Lee SX, Rathnayake SI (2019) Finite mixture models. Annu Rev Stat Appl 6:355–378. https://doi.org/10.1146/annurev-statistics-031017-100325
Meilijson I (1989) A fast improvement to the EM algorithm on its own terms. J R Stat Soc B Met 51(1):127–138. https://doi.org/10.1111/j.2517-6161.1989.tb01754.x
Molenberghs G, Lesaffre E (1994) Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. J Am Stat Assoc 89:633–644. https://doi.org/10.1080/01621459.1994.10476788
Nelsen RB (2006) An introduction to copulas. Springer, New York
Piccolo D (2003) On the moments of a mixture of uniform and shifted binomial random variables. Quad Stat 5:85–104
Piccolo D (2006) Observed information matrix for MUB models. Quad Stat 8:33–78
Piccolo D, D'Elia A (2008) A new approach for modelling consumers' preferences. Food Qual Prefer 19(3):247–259. https://doi.org/10.1016/j.foodqual.2007.07.002
Piccolo D, Simone R (2019) The class of CUB models: statistical foundations, inferential issues and empirical evidence. Stat Method Appl 28:389–435. https://doi.org/10.1007/s10260-019-00461-1
R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Redner RA, Walker HF (1984) Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 26(2):195–239. https://doi.org/10.1137/1026034
Scott WA (2002) Maximum likelihood estimation using the empirical Fisher information matrix. J Stat Comput Simul 72(8):599–611. https://doi.org/10.1080/00949650213744
Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press, Cambridge
Tutz G (2022) Ordinal regression: a review and a taxonomy of models. WIRES Comput Stat 14(2):e1545. https://doi.org/10.1002/wics.1545
Van Vaerenbergh Y, Thomas TD (2013) Response styles in survey research: a literature review of antecedents, consequences, and remedies. Int J Public Opin Res 25(2):195–217. https://doi.org/10.1093/ijpor/eds021
Varin C, Vidoni P (2006) Pariwise likelihood inference for ordinal categorical time series. Comput Stat Data Anal 51(4):2365–2373. https://doi.org/10.1016/j.csda.2006.09.009
Wakita T, Ueshima N, Noguchi H (2012) Psychological distance between categories in the Likert scale: comparing different number of options. Educ Psychol Meas 72(4):533–546. https://doi.org/10.1177/0013164411431162