Testing distributional assumptions in CUB models for the analysis of rating data

Francesca Di Iorio[1] · Riccardo Lucchetti[2] · Rosaria Simone[1]

Abstract

In this paper, we propose a *portmanteau* test for misspecification in combination of uniform and binomial (CUB) models for the analysis of ordered rating data. Specifically, the test we build belongs to the class of information matrix (IM) tests that are based on the information matrix equality. Monte Carlo evidence indicates that the test has excellent properties in finite samples in terms of actual size and power versus several alternatives. Differently from other tests of the IM family, finite-sample adjustments based on the bootstrap seem to be unnecessary. An empirical application is also provided to illustrate how the IM test can be used to supplement model validation and selection.

Keywords CUB model · Information matrix test · Ordered data · Misspecification

1 Introduction

The array of statistical models for the analysis of rating data is truly enormous. Among the many alternatives that have been proposed, the CUB mixture model, introduced by Piccolo (2003), D'Elia and Piccolo (2005), offers a unique approach to the problem; its most distinctive feature lies in its parameterization, which lends itself to an interpretation of the data generating process as a combination of perceptual and decisional aspects of the choice. A recent comprehensive discussion of the state of the art about the thread of research triggered by the seminal paper (Piccolo

---

✉ Rosaria Simone
   rosaria.simone@unina.it

   Francesca Di Iorio
   fdiiorio@unina.it

   Riccardo Lucchetti
   r.lucchetti@univpm.it

1 Dipartimento di Scienze Politiche, Università degli Studi di Napoli Federico II, Via Rodinò 22/a, 80133 Naples, NA, Italy

2 Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, P.le Martelli 8, 60121 Ancona, AN, Italy

2003) can be found in Piccolo and Simone (2019a, b), whereas a comparative analysis with the class of cumulative link models is performed in Piccolo et al. (2019). The main advantage of a modelling paradigm based on finite discrete mixtures is that it allows for a very versatile interpretation of the response distributions.

It is for this reason that the CUB model has been used in a wide range of applications, including sensory studies (Capecchi et al. 2016; Piccolo and D'Elia 2008; Corduas et al. 2013), consumers' preferences, satisfaction and opinions (D'Elia and Piccolo 2005; Capecchi et al. 2019a, b; Ribecco et al. 2022; Tovar et al. 2023; Xu and Zhang 2020; Finch and Hernández Finch 2020), subjects' perceptions on variety (Manisera et al. 2020), discrimination (Capecchi and Curtarelli 2020; Capecchi and Simone 2019), pain (D'Elia 2008), and health (Venson et al. 2023), to quote a few. In all applications, the explicit account of the uncertainty surrounding the rating process has provided effective visualization of results and added values to the characterization and interpretation of response profiles. From the methodological point of view, several extensions have enriched the literature, as in Manisera and Zuccolotto (2014), Corduas (2015), Cappelli et al. (2019), Di Nardo and Simone (2019), Biasetton et al. (2023), Corduas (2022), Simone et al. (2023), for instance.

Software implementations are available for the R environment (Iannario et al. 2018), as well as for Gretl (Simone et al. 2019) and STATA (Cerulli et al. 2022). Maximum likelihood inference is then based on the observed information matrix (Piccolo 2006): recently, the Louis' identity was used to derive the Information matrix as part of the EM algorithm and to implement an acceleration procedure which allows the best-subset variable selection become more feasible from the computational point of view (Simone 2020, 2021).

Given the peculiar interpretation of the parameters of the CUB model, the issue of model misspecification is clearly one of great importance. Yet, despite its widespread adoption in empirical applications, surprisingly few efforts have been so far devoted to model diagnostics and validation.

In order to make inference robust to potential misspecification, two main avenues are possible: either the CUB model is taken to represent an approximation to an underlying unknown data generating process, or its usage must be validated *ex post* by appropriate diagnostic tests: see for instance (Agresti et al. 2022; Colombi and Giordano 2019) for the case of categorical data models. In the former case, the appropriate inference procedures lie in the realm of Quasi-Maximum Likelihood and associated concepts (see for example Lv and Liu (2014) for the issue of model selection), which may be somewhat out of the reach of the average practitioner. Therefore, the present paper offers a contribution in the latter direction: we describe a testing procedure to check for the correct specification of CUB models via the information matrix test, introduced in White (1982) and developed henceforth for a wide array of statistical models (see for example Lucchetti and Pigini 2014).

The paper is organized as follows: Sect. 2.1 is devoted to a concise presentation of the modelling framework we move within; similarly, Sect. 2.2 establishes the necessary background for the implementation for the information matrix test for the class of CUB models. The proposal is discussed and validated via extensive Monte Carlo experiments presented in Sect. 3, whereas Sect. 4 provides two examples on

real data: we show how the proposed procedure can supplement model selection within the class of CUB models. A conclusion section ends the paper.

2 Definitions and preliminaries

2.1 The CUB model

For rating data such as those collected in survey studies to assess subjective evaluations and opinions, the class of CUB models employs a mixture of two distributions in its baseline specification. Suppose that R_i measures the response of i-th subject over m ordered categories, with $m > 3$. The data generating process is described as the combination of a *feeling* component and an *uncertainty* component. The former component is assumed a shifted binomial distribution:

$$b_r(\xi_i) = \binom{m-1}{r-1} \xi_i^{m-r}(1 - \xi_i)^{r-1} \quad r = 1, \ldots, m. \tag{1}$$

The latter accounts for heterogeneity of the distribution and is modelled via a discrete uniform distribution over the m categories.

As a result, a CUB model for R_i is specified via the following mixture:

$$Pr(R_i = r | \pi_i, \xi_i) = \pi_i b_r(\xi_i) + (1 - \pi_i)\frac{1}{m}, \quad r = 1, \ldots, m, \tag{2}$$

with $\pi_i \in (0, 1]$ and $\xi_i \in (0, 1)$. Note, however, that for $\pi = 0$ the parameter ξ in under-identified. As a consequence, in all the experiments when we generate CUB data (for example, in the simulation analysis of the size of the test) we will not consider the $\pi = 0$ case.

As for the interpretation of the parameters, the larger ξ_i is, the more the distribution is concentrated along the lowest scores. Thus, if the scale gives larger scores to positive evaluations, a low level of ξ_i indicates a positive tendency in the i-th observation with respect to the topic under investigation. For instance, if the respondent is asked to express his/her accordance to a given statement, then $1 - \xi_i$ can be viewed as a measure of agreement, or an indicator of satisfaction if he/she is asked to assess the quality of a service/product. After this interpretation, ξ_i is referred to as the *feeling* parameter. The mixing weight π_i of the feeling component in (2) is, instead, referred to as *uncertainty* parameter.

A richer CUB model can be obtained by including explanatory variables so that the feeling and/or uncertainty components directly depend on respondents' profiles: if $\mathbf{y}_i$ is the row vectors of p covariates $\mathbf{y}_i$ for the i-th subject that drives his/her uncertainty, whereas $\mathbf{w}_i$ is the row vectors of q covariates driving his/her feeling, then a logit link is customarily employed[1]:

---

[1] Clearly, the choice of the logit link could be generalized, but it is generally preferred on the grounds of simplicity of implementation and interpretation.

$$\lambda(\pi_i) = \beta_0 + \boldsymbol{y}_i \boldsymbol{\beta}_1^{'} \tag{3}$$

$$\lambda(\xi_i) = \eta_0 + \boldsymbol{w}_i \boldsymbol{\eta}_1^{'} \tag{4}$$

where $\lambda(x) \equiv \log\left(\frac{x}{1-x}\right)$. This generalization is referred to as a CUB(p, q) model, with estimable parameters $\boldsymbol{\beta}' = (\beta_0, \boldsymbol{\beta}_1)$ and $\boldsymbol{\eta}' = (\eta_0, \boldsymbol{\eta}_1)$. A simple CUB model with no covariates is indicated with CUB(0, 0) model (in this case, $\pi_i = \pi$ and $\xi_i = \xi$ are constant among subjects).

Finally, the CUB model can be inflated to take into account the presence of a "shelter" category (see Iannario 2012). A shelter category $c \in \{1, \dots, m\}$ is an item in the support of R_i that receives an upward bias of preference with respect to the expected response. The shelter effect can be accommodated in the CUB model by introducing a further mixture element, that is a degenerate distribution $D_r^{(c)} = I(R = c)$, whose probability mass is concentrated at $r = c$. Thus, the model becomes:

$$Pr(R_i = r) = \delta_i D_r^{(c)} + (1 - \delta_i)\left[\pi_i b_r(\xi_i) + (1 - \pi_i)\frac{1}{m}\right] \quad r = 1, \dots, m, \tag{5}$$

where the weight δ_i measures the shelter effect. The shelter coefficient may be constant across individuals ($\delta_i = \delta$), or it can, in turn, depend on a set of s covariates $\boldsymbol{x}_i$:

$$\lambda(\delta_i) = \omega_0 + \boldsymbol{x}_i \boldsymbol{\omega}_1^{'} \tag{6}$$

Given the previous parameterization, the matrices $\boldsymbol{Y}$, $\boldsymbol{W}$ and $\boldsymbol{X}$ may or may not possess an arbitrary number of common columns, that is the same covariates can be used as explanatory variables for feeling, uncertainty and shelter at the same time.

Estimation of CUB models is typically performed by maximum likelihood, where the log-likelihood is as follows:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[Pr(R_i = r|\boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{x}_i; \boldsymbol{\theta})\right] = \sum_{i=1}^{n} \log\left(p_i; \boldsymbol{\theta}\right) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) \tag{7}$$

with $\boldsymbol{\theta}$ is the vector of parameters. Maximization of the likelihood may be performed via the EM algorithm (McLachlan and Krishnan 1997; Dempster et al. 1977) as in Iannario et al. (2018), or by gradient-based methods as in Simone et al. (2019).

2.2 The Information Matrix (IM) test

The test we propose builds on a conditional moment approach and uses the fact that, under correct specification, the information matrix equality implies that the score variance and the expected Hessian should sum to zero. This result provides a set of moment conditions that can be empirically tested. The original idea was put forward in White (1982).

The information matrix test is therefore a test for $E(C_i) = 0$, where E is the expectation operator and

$$C_i = \text{vech}\left[\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\theta\partial\theta'} + G_i G_i'\right], \tag{8}$$

where ℓ_i is the log-likelihood for the i-th observation ($i = 1, \ldots, n$), $\boldsymbol{\theta}$ is the k-vector of parameters and $G_i \equiv \frac{\partial\ell_i}{\partial\theta}$; all quantities are evaluated at the "true" vector $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Clearly, C_i is a vector with $\tilde{k} = \frac{k(k+1)}{2}$ elements. In the rest of the paper, we will adopt the notational convention to indicate individual elements of the moment condition vector C_i by superscripting the two elements of the vector $\boldsymbol{\theta}$ with respect to which the derivatives are taken: for example, $C_i^{\beta,\sigma}$ indicates $\frac{\partial^2\ell_i}{\partial\beta\partial\sigma} + \frac{\partial\ell_i}{\partial\beta} \cdot \frac{\partial\ell_i}{\partial\sigma}$.

Under a set of regularity conditions (see White 1982, pages 2–10) that ensure asymptotic normality of the relevant quantities and the existence of an appropriate covariance matrix, a Wald-type test for $C_i = 0$ is asymptotically χ^2 distributed. Note that in some cases the degrees of freedom of the limit distribution may be smaller than $\tilde{k}$. More generally, the number of moment conditions to use in practice is open to choice. Such choice must be based on a mix of different considerations: small sample performance, ease of computation and scope of the alternative hypothesis. Tests based on a subset of the available moment conditions are sometimes termed "directional". See e.g. Golden et al. (2016), Lucchetti and Pigini (2013) for an extended discussion.

This strategy leads to many well-known and established test procedure: for example, it can be proven that the Jarque-Bera test for normality (Jarque and Bera 1980) is a simple special case of the IM test. In order to compute the test statistic, the covariance matrix of C_i is needed. In White's original formulation, this involves the third derivatives of the log-likelihood, which can make computation awkward in some cases. However, as pointed out in Chesher (1983) and Lancaster (1984), the test can be interpreted as a score test, which leads to a simplified formulation, in which the asymptotic version of the test is calculated via an Outer Product of the Gradient (OPG) "artificial regression" (see also Davidson and MacKinnon 2001): the test statistic equals nR^2 of the regression of a vector of ones on a matrix M, with typical row $M_i' = [G_i', C_i']$, that is a pseudo-model of the form

$$\boldsymbol{\iota} = G\boldsymbol{\beta} + C\boldsymbol{\eta} + \text{residual} \tag{9}$$

It can be proven that in the cases hinted at above, where some of the moment conditions are redundant, the artificial regression approach makes this problem evident because some of the columns of C may be collinear with G and the column rank of the matrix M is not full. Under the null, the test statistic has an asymptotic χ^2 distribution with degrees of freedom given by $df = \text{rank}(M) - k$.

In this paper, like in most applications, the "score form" of the test is adopted on account of its computational advantages, although for certain models its finite-sample performance can be inferior to other methods, as argued in Orme (1990). It must be noted in this regard that the problems are linked to the inefficient estimation of higher-order moments, and that an increasingly common alternative to analytical methods to correct the score form of the IM test has been the recourse to bootstrap methods, as suggested by Horowitz (1994). This technique has been used, among others in Lucchetti

and Pigini ([2013](), [2014]()), who tested the bivariate normality assumption in the bivariate probit and sample selection models.

That said, these problems do not arise in the present case, as will be shown in the rest of the paper: the CUB model is used for analyzing variables whose support is discrete, finite, and as a rule very small. Therefore, the issues typically encountered with higher-order moments in the general case are not a particularly serious problem here. This is arguably the reason why the experiments presented in Sect. 3.1 show that a bootstrap correction is, by and large, unnecessary.

2.2.1 The IM test for the CUB(0, 0) model

To give a practical example of the way the IM test can be applied to CUB models, consider the CUB(0, 0) model, where

$$Pr(R_i = r | \pi, \xi) = p_i = \pi b_r(\xi) + (1 - \pi) \frac{1}{m} \tag{10}$$

and adapt the notation of Appendix A to this special case as

$$\frac{\partial b_r(\xi)}{\partial \xi} = b_r(\xi) v_r(\xi)$$

$$v_r(\xi) = \frac{m - r}{\xi} - \frac{r - 1}{1 - \xi}$$

$$q_r(\xi) = -\frac{\partial v_r(\xi)}{\partial \xi} = \frac{m - r}{\xi^2} + \frac{r - 1}{(1 - \xi)^2}.$$

Considering that $\frac{\partial p_i}{\partial \pi} = b_r(\xi)$ and $\frac{\partial p_i}{\partial \xi} = \pi b_r(\xi) v_r(\xi)$, it is straightforward to compute the score $G_i = \left[G_i^\pi, G_i^\xi \right]'$ with respect to π and ξ as

$$G_i^\pi = \frac{\partial \ell_i}{\partial \pi} = s_i^\pi = \frac{b_r(\xi)}{p_i} \tag{11}$$

$$G_i^\xi = \frac{\partial \ell_i}{\partial \xi} = s_i^\xi = \frac{\pi b_r(\xi) v_r(\xi)}{p_i} = \pi s_i^\pi v_r(\xi) \tag{12}$$

The first two elements of the Hessian are also easy to calculate as

$$\frac{\partial^2 \ell_i}{\partial \pi^2} = h_i^\pi = -\left[s_i^\pi \right]^2 \tag{13}$$

$$\frac{\partial^2 \ell_i}{\partial \pi \partial \xi} = h_i^{\pi, \xi} = v_r(\xi) s_i^\pi \left(1 - \pi s_i^\pi \right) \tag{14}$$

$$\frac{\partial^2 \ell_i}{\partial \xi^2} = h_i^\xi = \pi \left[h_i^{\pi, \xi} v_r(\xi) - s_i^\pi q_r(\xi) \right] \tag{15}$$

The moment conditions can now be computed as

$$C_i^{\pi,\pi} = \left[s_i^\pi\right]^2 + h_i^\pi = \left[s_i^\pi\right]^2 - \left[s_i^\pi\right]^2 = 0 \tag{16}$$

$$C_i^{\pi,\xi} = \left[s_i^\pi s_i^\xi\right] + h_i^{\pi,\xi} = s_i^\pi v_r(\xi) \tag{17}$$

$$C_i^{\xi,\xi} = \left[s_i^\xi\right]^2 + h_i^\xi = \pi s_i^\pi \left[v_r(\xi)^2 + q_r(\xi)\right] \tag{18}$$

In this case, the regularity conditions needed for the test to be asymptotically χ^2-distributed are trivially satisfied under the null, since all the derivatives of the expressions above exist and are continuous in the interior of the parameter space, and moments of all orders exist for expressions (16)–(18) since the support of R is positive and finite (assuming a numerical coding for categories).

As for the degrees of freedom of the limit distribution of the test, note that in the artificial regression (9), only the third moment condition can be used: $C_i^{\pi,\pi}$ is identically 0, and $C_i^{\pi,\xi}$ is a scalar multiple of G_i^ξ (see Eq. 12). As a consequence, the IM test for the CUB(0, 0) model has an asymptotic χ_1^2 distribution.

3 Monte Carlo evidence

In this section, we analyze the features of the IM specification test via a series of simulation experiments, investigating its empirical size and its power against a range of alternatives.

3.1 Empirical size

In order to focus on the many aspects of interest, we begin by considering the special case of the simple CUB(0,0) model and gradually generalize the experiments to more complex specifications. To assess the empirical size of the IM test for the simple CUB(0,0) model, we simulated artificial data from eight different Data Generating Processes, corresponding to the points in the parameter space shown in Fig. 1.

The reason for choosing these points can be motivated as follows: apart from the obvious relevance of the point at the center of the parameter space (F), we consider the performance of the test for values of the parameter ξ close to the boundary of the parameter space, that is 0.1 and 0.9. Since the parameter space for this model is $(0, 1) \times (0, 1)$, all points are clearly in the interior of the parameter space, thus satisfying one of the basic regularity conditions. However, since for $\pi \to 0_+$ the Hessian tends to a singular matrix and the parameter ξ is under-identified in the limit, it is interesting to consider the performance of the test for moderate ($\pi = 0.25$) and serious ($\pi = 0.1$) cases of weak identification; in the latter case, we also consider the intermediate case (B) to get a clearer picture of the test performance.

Fig. 1 Selected points in the CUB(0,0) parameter space for the Monte Carlo experiment
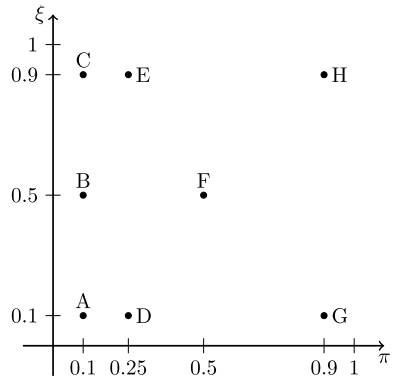
Each of these DGPs was simulated $J = 1000$ times, for a varying number of categories $m = 5$ and $m = 7$ and varying sample sizes $n = 128, 512, 1024$, that we consider as representative of typical empirical applications. Since the IM test is known to be liable to severe size bias in finite samples (see Horowitz 1994), we also examined the performance of bootstrap-corrected version of the IM test along the lines of Lucchetti and Pigini (2014). The number of bootstrap replications B is set to 999.

In practice, our experiment can be described as follows:

1. for each $j = 1, \ldots, J$

    (a) generate a sample $\boldsymbol{R}^{(j)}$ of n ordinal observations from a CUB(0,0) model over m categories, with parameters $\boldsymbol{\theta} = (\pi, \xi)$;

    (b) estimate the CUB(0, 0) model by ML, and compute the IM test statistics T_j using the estimated parameters $\hat{\boldsymbol{\theta}}$;

    (c) for $b = 1 \ldots B$
    (i) generate a bootstrap sample $\boldsymbol{R}^{(j,b)}$ by sampling from a CUB(0,0) model with parameters $\hat{\boldsymbol{\theta}}$;
    (ii) compute the corresponding IM test statistics $T_{j,b}$; in case of failure, regenerate $\boldsymbol{R}^{(j,b)}$ and repeat.
    (iii) upon successful computation of $T_{j,b}$, determine the quantile of order $1 - \alpha$ for the empirical distribution of $\{T_{j,b} : b = 1, \ldots, B\}$, say $q_{1-\alpha}^{(j)}$

2. Estimate the empirical size $\hat{\alpha}$ of the uncorrected IM test as:

$$\hat{\alpha} = \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}\{T_j > p_{1-\alpha}\},$$

    that is, by counting the frequency of the IM statistic T_j exceeding the χ_1^2 critical value $p_{1-\alpha}$.

3. Estimate the empirical size $\tilde{\alpha}$ of the bootstrap-corrected IM test as:

Table 1 Empirical size of the IM Test for CUB models if $n = 128$

Parameters			Size	$m = 5$				$m = 7$			
Label	π	ξ		0.01	0.05	0.10	Failed	0.01	0.05	0.10	Failed
A	0.1	0.1	$\hat{\alpha}$	0.02	0.06	0.11	1.4	**0.02**	**0.07**	**0.12**	0.8
			$\tilde{\alpha}$	0.01	0.04	0.09		0.01	0.06	0.10	
B	0.1	0.5	$\hat{\alpha}$	**0.02**	**0.07**	0.11	0.6	0.01	0.06	0.10	0.3
			$\tilde{\alpha}$	**0.02**	0.06	0.10		0.01	0.05	0.10	
C	0.1	0.9	$\hat{\alpha}$	**0.02**	**0.08**	**0.13**	1.9	0.01	0.06	0.10	1.0
			$\tilde{\alpha}$	0.01	**0.07**	**0.12**		0.01	0.05	0.09	
D	0.25	0.1	$\hat{\alpha}$	0.01	0.05	0.11	0.5	0.01	0.05	0.11	0.2
			$\tilde{\alpha}$	0.01	0.05	0.10		0.01	0.05	0.10	
E	0.25	0.9	$\hat{\alpha}$	0.02	0.05	0.09	0.6	0.01	0.05	0.11	0.2
			$\tilde{\alpha}$	0.01	0.04	0.09		0.01	0.04	0.11	
F	0.5	0.5	$\hat{\alpha}$	0.01	0.05	0.12	0.0	0.01	0.04	0.10	0.0
			$\tilde{\alpha}$	0.01	0.05	0.11		0.01	0.04	0.09	
G	0.9	0.1	$\hat{\alpha}$	**0.02**	**0.06**	0.10	0.0	**0.02**	**0.08**	**0.13**	0.0
			$\tilde{\alpha}$	0.01	**0.02**	**0.05**		**0.00**	0.04	0.09	
H	0.9	0.9	$\hat{\alpha}$	**0.02**	**0.07**	0.11	0.1	0.01	0.06	**0.12**	0.0
			$\tilde{\alpha}$	0.01	0.05	0.10		0.01	0.05	0.11	

failed = percentage of failed IM test evaluations (see text)

boldface: empirical size significantly different from nominal size (5% level)

$$\tilde{\alpha} = \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}\left\{T_j > q_{1-\alpha}^{(j)}\right\},$$

that is, by counting the frequency of the IM statistic T_j exceeding the bootstrap critical value $q_{1-\alpha}^{(j)}$.

Note that, at step 1(c)(ii), the computation of the test may fail because the generated data make the model under-identified; typically, this happens when no observations are generated with $R_i = m$.

Tables 1, 2, 3 report the empirical rejection rates over 1000 MC replications; entries in boldface indicate cases when the empirical rejection rate $\bar{\alpha}$ was significantly different form the nominal rejection rate α, that is when

$$\sqrt{n}\left|\frac{\bar{\alpha} - \alpha}{\sqrt{\alpha(1 - \alpha)}}\right| > 1.96$$

where $\bar{\alpha}$ is $\hat{\alpha}$ or $\tilde{\alpha}$, according to cases.

It appears that the size of the IM test for the CUB model without covariates seems to be rather satisfactory in most of the cases considered, even without a bootstrap correction. As is common with conditional moment tests, size bias is largely a small sample phenomenon. However, for a large region of the parameter space

Table 2 Empirical size of the IM Test for CUB models if $n = 512$

Parameters			Size	$m = 5$				$m = 7$			
Label	π	ξ		0.01	0.05	0.10	Failed	0.01	0.05	0.10	Failed
A	0.1	0.1	$\hat{a}$	**0.02**	0.06	0.11	1.0	**0.02**	**0.06**	0.12	0.4
			$\tilde{a}$	0.01	0.06	0.10		**0.02**	0.05	0.11	
B	0.1	0.5	$\hat{a}$	0.01	0.05	0.11	0.3	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.06	0.11		0.01	0.05	0.10	
C	0.1	0.9	$\hat{a}$	0.01	0.06	0.12	1.0	0.01	0.06	0.11	0.3
			$\tilde{a}$	0.01	0.05	0.10		0.01	0.05	0.10	
D	0.25	0.1	$\hat{a}$	0.01	0.05	0.09	0.0	0.01	**0.06**	0.12	0.0
			$\tilde{a}$	0.01	0.05	0.09		0.01	**0.07**	0.12	
E	0.25	0.9	$\hat{a}$	0.01	0.05	0.10	0.0	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.05	0.11		0.01	0.05	0.10	
F	0.5	0.5	$\hat{a}$	0.01	0.06	0.10	0.0	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.05	0.11		0.01	0.05	0.10	
G	0.9	0.1	$\hat{a}$	0.01	0.05	0.11	0.0	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.05	0.10		0.01	0.05	0.10	
H	0.9	0.9	$\hat{a}$	0.01	0.05	0.11	0.0	0.01	0.05	0.09	0.0
			$\tilde{a}$	0.01	0.05	0.11		0.01	0.04	0.09	

failed = percentage of failed IM test evaluations (see text)

boldface: empirical size significantly different from nominal size (5% level)

Table 3 Empirical size of the IM Test for CUB models if $n = 1024$

Parameters			Size	$m = 5$				$m = 7$			
Label	π	ξ		0.01	0.05	0.10	Failed	0.01	0.05	0.10	Failed
A	0.1	0.1	$\hat{a}$	0.01	0.05	0.10	0.4	0.01	0.04	**0.08**	0.0
			$\tilde{a}$	0.01	0.05	0.10		0.01	*0.04*	**0.08**	
B	0.1	0.5	$\hat{a}$	0.01	0.05	0.10	0.1	0.01	0.04	0.09	0.0
			$\tilde{a}$	0.01	0.05	0.10		0.01	0.04	0.09	
C	0.1	0.9	$\hat{a}$	0.01	0.06	0.09	0.5	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.05	0.09		0.01	0.04	0.09	
D	0.25	0.1	$\hat{a}$	0.01	0.06	0.11	0.0	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.05	0.11		0.01	0.05	0.10	
E	0.25	0.9	$\hat{a}$	0.01	0.06	0.13	0.0	0.01	0.05	0.10	0.0
			$\tilde{a}$	0.01	0.06	0.12		0.01	0.05	0.10	
F	0.5	0.5	$\hat{a}$	0.01	0.06	0.12	0.0	0.01	0.04	0.09	0.0
			$\tilde{a}$	0.01	0.06	0.12		0.01	0.05	0.09	
G	0.9	0.1	$\hat{a}$	0.01	0.06	0.11	0.0	0.01	0.05	0.09	0.0
			$\tilde{a}$	0.01	0.06	0.11		0.01	0.04	0.08	
H	0.9	0.9	$\hat{a}$	0.02	0.06	0.11	0.0	0.01	0.06	0.10	0.0
			$\tilde{a}$	0.01	0.06	0.10		0.01	0.05	0.09	

failed = percentage of failed IM test evaluations (see text)

boldface: empirical size significantly different from nominal size (5% level)

Table 4 Selected CUB with shelter for DGP

Experiment	m	π	ξ	δ	Shelter at
1	9	0.6	0.4	0.1	1
2	10	0.7	0.5	0.05	10
3	7	0.3	0.1	0.1	5

Table 5 Empirical size of the IM Test for CUB models with shelter $n = 128$ and $n = 256$

	$n = 128$			$n = 256$		
	0.01	0.05	0.10	0.01	0.05	0.10
Exp. 1	0.01	0.06	0.11	0.01	0.04	0.09
Exp. 2	0.01	0.05	0.11	0.01	0.05	0.10
Exp. 3	0.01	0.06	0.10	0.01	0.06	0.11

Table 6 Empirical size of the IM Test for CUB models with shelter $n = 512$ and $n = 1024$

	$n = 512$			$n = 1024$		
	0.01	0.05	0.10	0.01	0.05	0.10
Exp. 1	0.01	0.05	0.08	0.01	0.04	0.09
Exp. 2	0.01	0.06	0.10	0.02	0.06	0.11
Exp. 3	0.01	0.05	0.09	0.01	0.05	0.10

the actual rejection rate is not significantly different from the nominal size even for moderately sized samples ($n = 128$): entries in boldface indicate that the hypothesis that the empirical size was equal to the nominal one was rejected at a 5% level.

The regions of the parameter space where the sample size bias is most severe are the ones closer to the lower border of the parameter space for the uncertainty parameter (points A, B and C); it is safe to explain this result by the fact that in that parameter region, especially when ξ is small, the model is only weakly identified, so a procedure such as the IM test, which is essentially based on the curvature of the log-likelihood, can be expected to perform rather poorly in small samples.

3.1.1 Empirical size of the IM test for CUB models with shelter

As a final Monte Carlo exercise on the empirical size, we report the main results for the IM test for CUB models with a shelter specification (see Sect. 2.1). Table 4 lists the DGP configurations we considered, on sample sizes $n = 128, 256, 512, 1024$. The shelter category is usually chosen *a priori* or selected on the basis of goodness of fit criteria; in the following, we will be mainly concerned with testing the correct specification of the shelter against alternative shelter choices.

Tables 5 and 6 report the empirical size obtained in a Monte Carlo exercise with 1000 replications to verify the performance of the IM test for CUB with shelter.

In this context, it is important to note that the position of the shelter is not irrelevant, especially in small samples, because different choices for parameters π and

ξ have a dramatic impact on the shape of the distribution. As a result, if the shelter is not distant from the mode, the estimation procedure of the parameters obtained through gradient-based ML methods may encounter difficulties and the procedure may not converge.[2]

3.2 Empirical power without covariates

In this section, we provide and discuss Monte Carlo evidence related to the study of the power of the Information Matrix test for CUB models.

A general remark on the power of the IM test is in order here: as White (1982) argued, the QMLE $\hat{\theta}$ can be thought of as a consistent estimator of θ_*, the point in the parameter space that minimizes the Kullback–Leibler divergence to the true probability distribution. If the model is correctly specified, θ_* is the "true" parameter vector, and inference is standard; otherwise, $\hat{\theta}$ may not have usual desirable asymptotic properties if θ_* is not in the interior of the parameter space. Since the IM test is based on the QMLE $\hat{\theta}$, it is particularly interesting to perform Monte Carlo experiments under alternatives for which the pseudo-true value θ_* may lie on the boundary (note that the QML estimator will generally be away from the boundary even in these cases).

Having said this, power analysis can be undertaken from different perspectives in order to assess the extent by which the IM test is able to detect violation of the distributional assumptions of CUB models, depending on the data generating process (DGP): a modification of one mixture component, or an entirely different model. With some more details, our power analysis will follow this scheme:

- First, we will check the power performance of the IM test when modifying only the feeling or the uncertainty components, by assuming, respectively, the addition of over-dispersion and the inclusion of shelter effect: see Sects. 3.2.1 and 3.2.2, respectively. In the latter setting, we will also discuss the behavior of the test when the interest lies in verifying the correct location of the shelter, assuming that CUB with shelter is the DGP: see Sect. 3.2.2.
- Then, in Sect. 3.3 we will switch to power analysis when specifying covariates for either CUB components and when assuming the proportional odds version of the cumulative models (POM) as DGP.
- Last, we will discuss power performance in case the DGP does not belong to the class of CUB models: this will be the case of the discretized beta and the beta-binomial distributions: see Sect. 3.2.3.

In all cases, we report results corresponding to a significance level of $\alpha = 0.05$.

---

[2] The IM test procedure developed with `gretl` allows to distinguish between non-convergences of the MLE and cases where the number of degrees of freedom of the IM test is zero.

3.2.1 Testing over-dispersion under the alternative

Over-dispersion in ordinal ratings and count data is a subtle phenomenon to measure: it can be empirically shown that it is related to an excess in variability with respect to the underlying model, where variability has to be meant as mutual differences among frequencies (see Piccolo and Simone 2019a). Within the class of CUB mixture models, accounting for over-dispersion is usually handled by assuming a beta-binomial distribution for the feeling component. For $r = 1, \ldots, m$, this probability model is parameterized as follows:

$$g_r(\xi, \phi) = \binom{m-1}{r-1} \frac{\prod_{k=1}^{r} [1 - \xi + \phi(k-1)] \prod_{k=1}^{m-r+1} [\xi + \phi(k-1)]}{[1 - \xi + \phi(r-1)][\xi + \phi(m-r)] \prod_{k=1}^{m-1} [1 + \phi(k-1)]};$$

(19)

where the parameter ξ has the same interpretation as in the binomial distribution and the additional parameter $\phi \geq 0$ accounts for over-dispersion, so that the binomial distribution arises as the limiting case $\phi = 0$: its mixture with a uniform distribution is referred to as CUBE model.

Clearly, the CUBE model nests CUB as a special case, so in principle a test for discriminating between the two could be based on the hypothesis $H_0 : \phi = 0$. However, we believe the performance of the IM test is worth investigating for several reasons: first, the point $\phi = 0$ is on the boundary of the parameter space, so it is not obvious to ascertain if the usual regularity conditions apply. Moreover, testing for $\phi = 0$ implies assuming that the alternative model is known *a priori*, which is obviously not the case in actual practice; the advantage of the IM test is in this case its being agnostic with respect to the alternative DGP.

In short, in this subsection we investigate the performance of the IM test when the data feature over-dispersion that cannot be accounted for by CUB models. To this goal, we will determine the empirical power function for different scenarios with a CUBE model as DGP, for varying values of ϕ.

It is worth remarking here that the most likely elements that determine the power of the IM test in this case are (a) the value of ϕ, since for $\phi = 0$, the two models coincide; (b) the distance of ξ from 0.5, since for $\xi = 0.5$ both distributions are symmetric and the only distinguishing feature of CUB vs CUBE is over-dispersion, and (c) the value of the parameter π, since for $\pi \to 0_+$, the number of observations coming from the feeling portion of the mixture is small, thereby making the effective sample size much smaller than the actual one.

In the light of the above, our findings are rather unsurprising: we display selected results corresponding to $\alpha = 0.05$ in Fig. 2, with values of (π, ξ) matching the ones used for the analysis of the empirical size; since power can be small for some configurations, we set the sample size n to 2^{12} to make results visually more evident. We find evidence that power performance is poor for large heterogeneity (low values of π), whereas they are satisfactory for symmetric distribution ($\xi = 0.5$) only for a large number of categories. In general, power improves with larger m, arguably
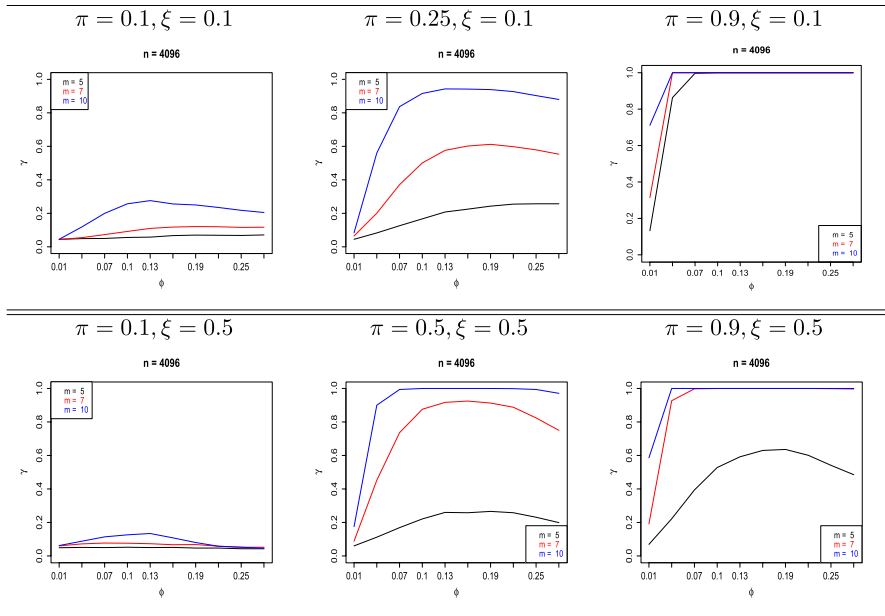
Fig. 2 (Empirical) Power function of the IM test for CUB models if the DGP is a CUBE model

due to a larger variability in higher sample moments. By focusing on symmetric distributions ($\xi = 0.5$), it can be said that a constrained parameter space for ϕ in $(0, 0.2)$ is advisable, since in this case the beta-binomial distribution converges quite fast to a uniform distribution for larger values of ϕ. Indeed, over-dispersion is inversely related to mutual differences among frequencies.

Notice that CUB models are reversible (in the sense that, if $R \sim \text{CUB}(\pi, \xi)$, then $m - R + 1 \sim \text{CUB}(\pi, 1 - \xi)$); for this reason, Fig. 2 does not display the results corresponding to $\xi = 0.9$, since these coincide with those corresponding to $\xi = 0.1$.

3.2.2 Shelter effect

In this subsection, we study the performance of the IM test to detect misspecification within the uncertainty component, with respect to the presence of a shelter effect (for the sake of illustration, this will be assumed at $c = 1$) or its potential misplacement.

From Eq. (5), it is clear that when $\delta = 0$, the CUB with shelter model collapses to a baseline CUB model. In our first experiment, we explore the power of the IM test to detect the presence of a shelter effect when the estimated model has none.

Figure 3 displays the empirical power function for different choices of model parameters. In this setting, it can be seen that the behavior is quite satisfactory, unless the location of the shelter coincides with the modal value of the underlying
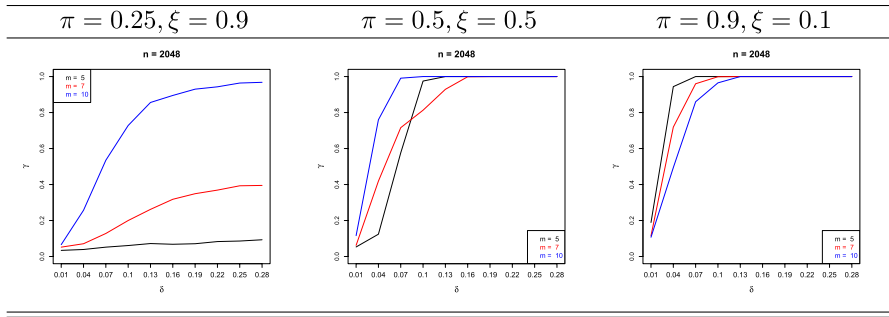
Fig. 3 Empirical power function of the IM test for CUB models if the DGP is a CUB with shelter at $c = 1$
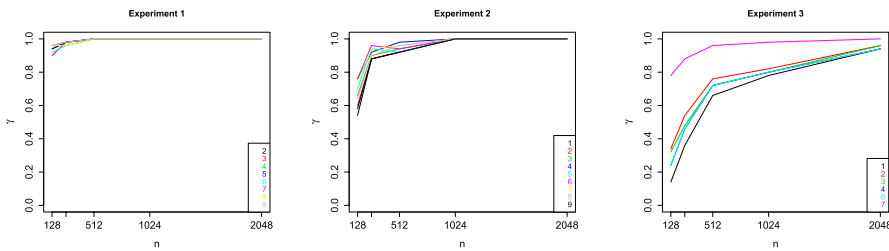
Fig. 4 (Empirical) Power function of the IM test on CUB with shelter at $r \neq c$, if the DGP is a CUB with shelter at c

binomial distribution (if $\xi = 0.9$, for instance), especially for low values of m. Thus, larger values for m are preferable for inferential procedures in this regard.

In the next three experiments, we assume that the DGP is a CUB with shelter, but we assume that the correct location of the shelter category is unknown and test its location via IM test. Thus, in this circumstance model misspecification amounts to the distance between the true shelter category and the one assumed under the null. Figure 4 displays the results for varying n with the following DGPs, showing that the IM test behaves consistently.

Experiment 1: $m = 9; \pi = 0.6; \xi = 0.4; \delta = 0.1$, and shelter at $c = 1$. For the underlying CUB model, the modal value is at $Mo = 6$; Thus, we have run the IM test for a CUB model with shelter assuming that the shelter is at $s = r, r \neq c$;

Experiment 2: $m = 10; \pi = 0.7; \xi = 0.5; \delta = 0.05$, and shelter at $c = m$. For the underlying CUB model, the modal values are at $Mo = 5, 6$; thus, we have run the IM test for CUB with shelter assuming that the shelter is at $s = r, r \neq c$;

Experiment 3: $m = 7; \pi = 0.3; \xi = 0.1; \delta = 0.1$, and shelter at $c = 5$. For the underlying CUB model, the modal value is at $r = 7$; thus, we have run the IM test for CUB with shelter assuming that the shelter is at $s = r, r \neq c$.
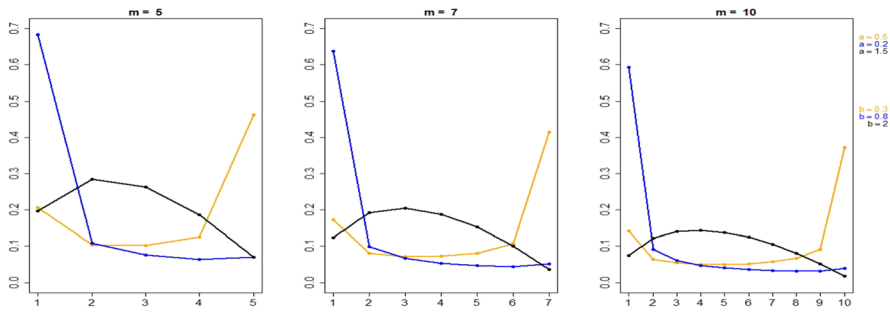
Fig. 5 DB distributions used for power analysis with $m = 5, 7, 10$

Table 7 Parameter values for beta-binomial model used as DGP

	1	2	3	4
ξ	0.1	0.3	0.5	0.8
ϕ	0.2	0.15	0.25	0.3

3.2.3 Power with discretized beta and beta-binomial as DGP

In this section, we perform some Monte Carlo experiments where data are generated using alternative distributions, with samples of size $n = 128, 256, 512, 1024, 2048$, and different number of categories $m = 5, 7, 10$. The chosen data generating processes are as follows:

- Discretized Beta DB(a, b) (Ursino and Gasparini 2018): If $X \sim Beta(a, b)$ is a beta-distributed random variable, a discrete random variable D over the support $\{1, \ldots, m\}$ follows the discretized beta distribution DB(a, b) with parameters $a, b > 0$ if, for $r = 1, \ldots, m$:

$$Pr(D = r|a, b) = Pr\left(\frac{r-1}{m} \leq X < \frac{r}{m} \,\middle|\, a, b \right).$$

  The broad flexibility of this distribution makes it possible to consider a broad range of shapes, as shown in Fig. 5.
- Beta-binomial defined in (19). We analyzed the power of the IM test for CUB models against the four scenarios for the DGP, listed in Table 7, each tested with $m = 5, 7, 10$:
  Figure 6 displays the distributions for all the chosen scenarios.

It should be remarked that for some parameter configurations the probability limit of the CUB estimator may lie on the boundary of the parameter space; specifically, when the true probability is U-shaped, $\xi \to 0$ or $\xi \to 1$, and $\delta \to 0$ if the shelter effect is negligible, so standard asymptotic arguments do not apply and the
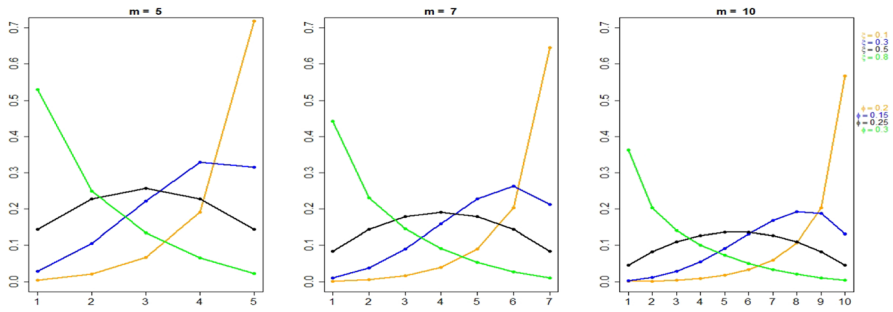
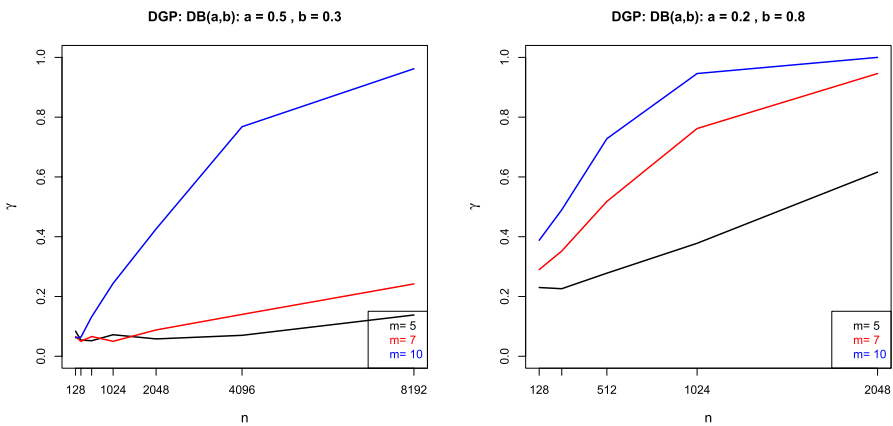**Fig. 6** Beta-binomial distributions used for power analysis with $m = 5, 7, 10$

Fig. 7 (Empirical) Power function of the IM test for CUB model with shelter if the DGP is a U-shaped discretized beta, with shelter at $c = 1$ for $(a = 0.5, b = 0.3)$ and at $c = m$ for $(a = 0.2, b = 0.8)$

distribution of the estimator is unknown. It is especially interesting to analyse the power properties of the IM test via simulation under these scenarios. Section C in the appendix analyses these cases in greater depth.

We start our discussion by assuming the discretized beta model as DGP. If the data exhibit a U-shape distribution, the practitioner may want to use a CUB model with shelter. In this case, the IM test can be used to assess the validity of the choice. Thus, we run the IM test for a CUB with shelter at $c = 1$ or $c = m$ if the largest frequency is at $r = m$ $(a > b)$ or $r = 1$ $(a < b)$, respectively.

Power performance is displayed in Fig. 7 and they generally improve for growing m, but with a slower increase with n when the U-shape of the distribution is more evident for low values of m.[3] It can be shown that a U-shape DB model corresponds to a mixture of a J-shaped DB model with a reverse J-shaped DB model (see the Appendix to Simone (2022)). In order to show our

---

[3] For one scenario, convergence of power to 1 is quite slow, so we show results up to $n = 8192$ to aid visualization.

arguments, without loss of generality assume the case $a = 0.1, b = 0.2$: the corresponding U-shape DB distribution will have modal value at $c = 1$, and a huge excess of frequency at $c = m$, with a flat distribution in between, and it is equivalent to a mixture between a DB(0.1, 1) and a DB(1, 0.2).

If the latter distribution can be approximated by a binomial, accounting for feeling, the former J-shaped distribution can be in turn written as a mixture of a DB model close to the uniform distribution (DB(1, 1.2)) and an almost degenerate DB model ($a = 2, b = 0.2$) with mode at $r = 1$. As a consequence, in case of extremely polarized distributions arising from the DB model as a DGP (low values of both a, b), the power performance of the IM test to check the correct specification of a CUB with shelter become satisfactory slowly, and in general faster for larger m, while the power performance of the IM test to check the correct specification of a baseline CUB without shelter are satisfactory, as expected. It could be surmised that the poor power properties of the IM test in the case ($a = 0.5, b = 0.3$) could be attributable to the pseudo-true parameter θ_* being on the edge of the parameter space. This, however, is also true for the ($a = 0.2, b = 0.8$) model, which yields much better power properties.

Therefore, the power of the IM test for CUB models appears to be larger when the shape of the distribution is untypical of CUB models, which is not surprising. More in detail, in the case of U-shaped distributions, the test is very powerful to detect misspecification if the U-shape is particularly marked: performance worsens as the U-shape becomes flatter, in which case an IM test for CUB with shelter gives more satisfactory performance instead.

These considerations are due to the circumstance that some binomial distributions are well approximated by a DB model (Ursino and Gasparini 2018). This issue explains why, when the DGP is a unimodal DB model, with modal value at an inner category (that is, different from 1 and m), performance of the IM test for CUB models for small samples are satisfactory only for moderate or large number of categories and for distributions with a certain extent of heterogeneity. For the sake of completeness, Fig. 8 displays the empirical power function for a Monte Carlo experiment related to data generated from a DB(a, b), with $a = 1.5, b = 2$, for varying m and n.

We now switch to the analysis of the power performance of the IM test if the DGP is a beta-binomial model. Figure 9 displays the results at a glance. It follows that the performance of the test are very satisfactory, especially for medium and large scales. A larger value of n is instead needed to attain satisfactory performance for symmetric distributions ($\xi = 0.5$), especially for a small number of categories ($m = 5$).

3.3 Empirical power with covariates

In this section, we analyze the power of the IM test with respect to model misspecification in terms of explanatory variables in the model. We begin by
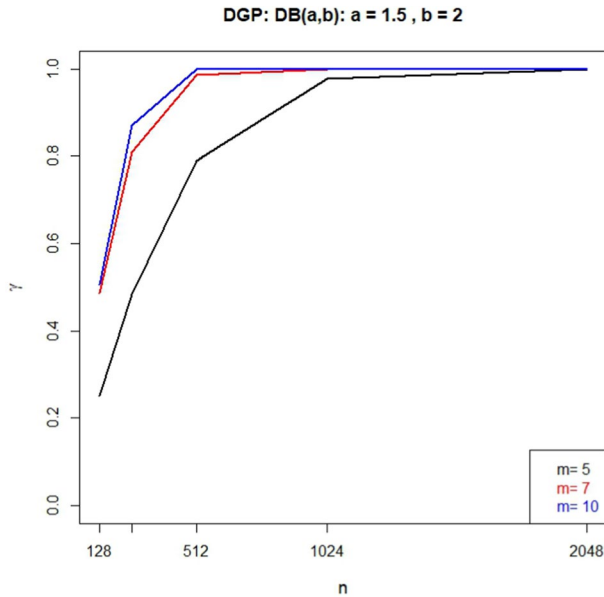
Fig. 8 (Empirical) Power function of the IM test for CUB model if the DGP is a unimodal discretized beta with inner modal value

exploring this aspect by itself and then combine covariate omission with misspecification of the data distribution assuming POM as DGP.

3.3.1 CUB with covariates as DGP

We generate data using the CUB(1,1) model as DGP, with a dummy covariate D entering both feeling and uncertainty Eqs. (3)–(4):

$$\lambda(\pi_i) = \beta_0 + \beta_1 D_i, \qquad \lambda(\xi_i) = \eta_0 + \eta_1 D_i.$$

In this case, the parameter space is $\mathbb{R}^4$ so any possible set of values lie in the interior of the parameter space; for our experiment, the parameters $\beta_0, \beta_1, \eta_0, \eta_1$ were chosen so as to yield:

$$(\pi_i | D_i = 0) = 0.3, \ (\pi_i | D_i = 1) = 0.7$$
$$(\xi_i | D_i = 0) = 0.6, \ (\xi_i | D_i = 1) = 0.2.$$

The dummy variables D_i are iid with $Pr(D_i = 1) = 0.6$. Figure 10 shows the resulting conditional and unconditional distributions for $m = 5, 7, 10$.

We then run the IM test for CUB models with covariates by considering as null hypothesis three separate cases of misspecification, that is CUB(0,0), CUB(1,0) and CUB(0,1). Figure 11 displays the results corresponding to significance level $\alpha = 0.05$.
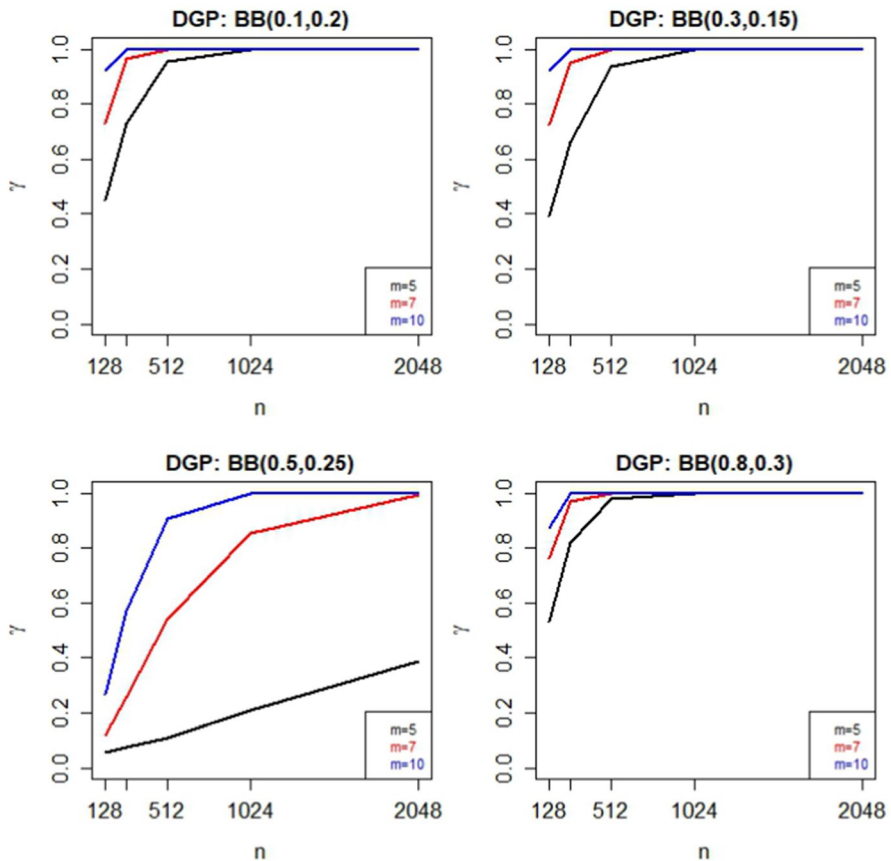
Fig. 9 (Empirical) Power function of the IM test for CUB models if the DGP is a beta-binomial model

Testing the correct specification of CUB(0,0), CUB(1,0) or CUB(0,1) via the IM test if the DGP is a CUB(1,1) model is equivalent to testing if the missed specification of significant effects of D for at least one model components is successfully identified. Results show satisfactory performance of the power of the IM test to check the correct specification of CUB(1,0) and CUB(0,1) models. With respect to the null of a CUB(0,0) model, instead, performances are weakest due to the unimodality of the overall distribution: in particular, the empirical power converges to 1 more slowly as n grows.

3.3.2 Proportional odds model as DGP

In the following paragraph, we show and discuss the performance of the IM test for CUB models if data are sampled according to a proportional odds version of a cumulative link model (POM, McCullagh 1980):
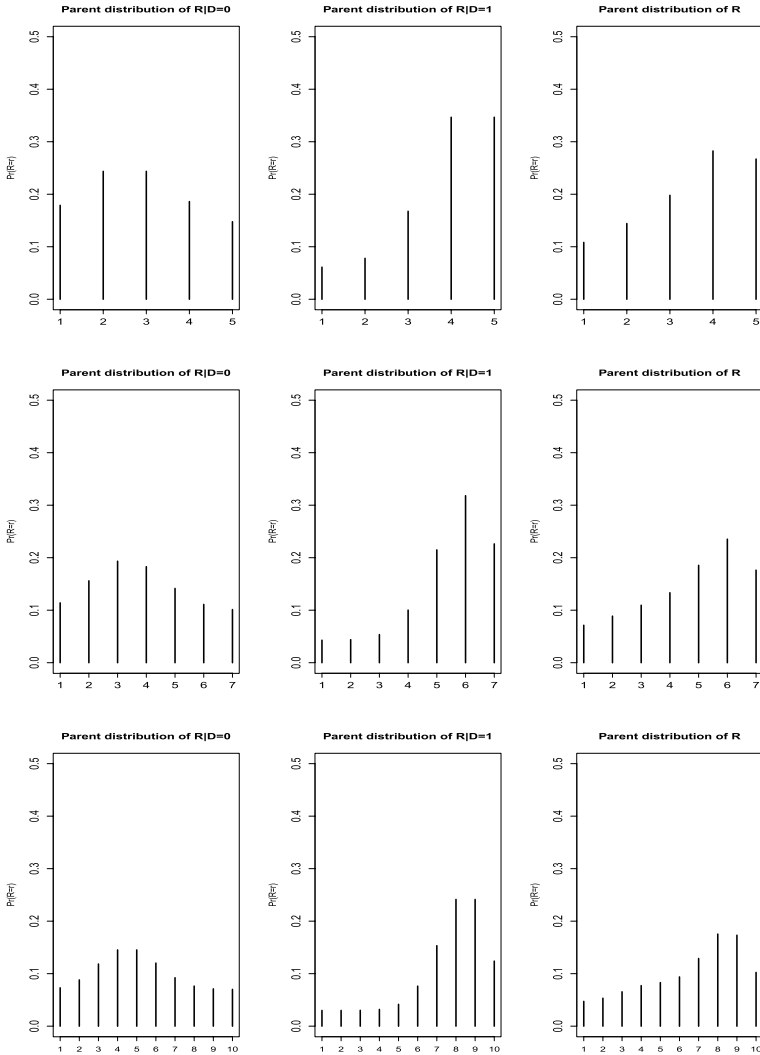
Fig. 10 Power analysis: Generating CUB distributions (conditional to D and unconditional) for fixed values of m (top: $m = 5$, center: $m = 7$, bottom: $m = 10$)

$$\lambda(Pr(R_i \leq r | \tau, \beta, x_i)) = \tau_r - \beta x_i, \qquad r = 1, \ldots, m.$$

In the POM model, the data generating process does not include any uncertainty component. Then, our power analysis aims at showing how the IM test for possible CUB model specifications behaves. In other words, if the data provide evidence against such null hypothesis, one could conclude that mixture models including an uncertainty component are unlikely to be adequate for the data at hand.

As for the variable x, we considered two cases: one where x is continuous, generated from a standard Gaussian distribution, and one when x comes from a Bernoulli
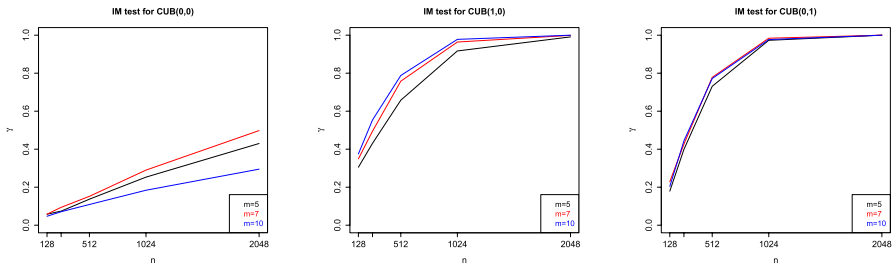
Fig. 11 (Empirical) Power function of the IM test for CUB specification without covariates or with covariates only for one component, assuming CUB(1,1) as DGP
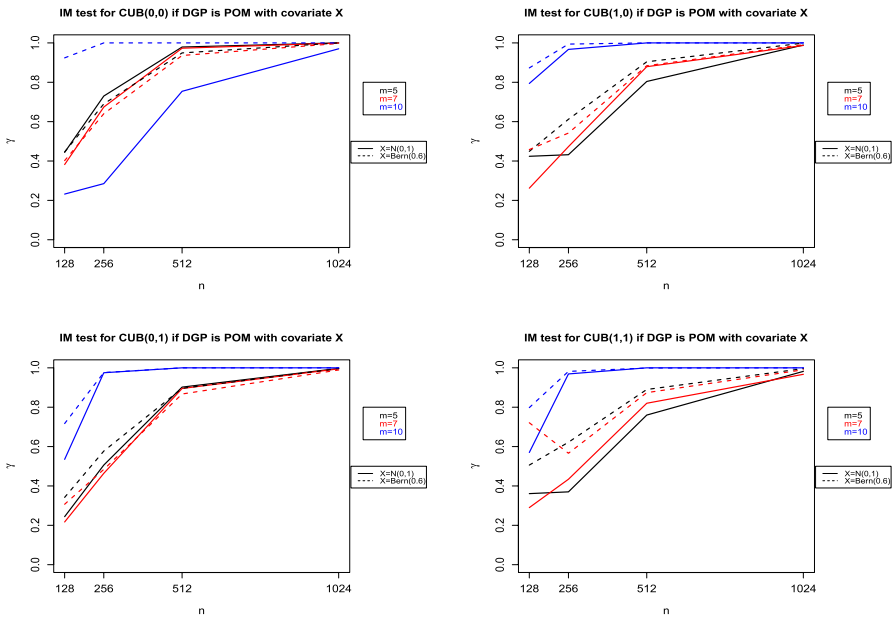

Fig. 12 (Empirical) Power function of the IM test for CUB specifications if the DGP is a POM with $N(0, 1)$ or Bernoulli covariate

distribution with $p = 0.6$. In this case we computed the power of the IM test to check the correct specification of the possible CUB models (CUB(0,0) with no covariate, CUB(1,0) and CUB(0,1) with covariate only for one component, and CUB(1,1)). Figure 12 displays the empirical power as a function of n, for varying m and significance level $\alpha = 0.05$, showing that in this case the test behaves satisfactorily, with slightly superior performance if x is Bernoulli rather than continuous for low and moderate sample sizes, and generally improving with growing m.

Table 8 Statistics p values for the IM test for baseline CUB(0,0) specification

	Informat	Willing	Officeho	Compete	Global
IM statistics	15.512	29.137	57.053	20.434	2.216
p value	0.000	0.000	0.000	0.000	0.137

Table 9 p values for the IM test for CUB(0,0) with shelter

	Shelter category						
	1	2	3	4	5	6	7
Informat	0.00	0.00	0.00	0.00	0.00	0.00	0.05
Willingn	0.00	0.00	0.00	0.00	0.00	0.00	0.04
Officeho	0.00	0.00	0.00	0.00	0.00	0.00	0.14
Compete	0.00	0.00	0.00	0.00	0.00	0.00	0.53
Global	0.07	0.08	0.15	0.03	0.11	0.02	0.43

4 Example applications

In this section, we provide two examples of our proposed test procedure. First, in Sect. 4.1 we illustrate how to check for misspecification of CUB models without covariates, possibly with shelter, on a real dataset that has been traditionally used as a test bed in the CUB literature. Next, in Sect. 4.2 we illustrate the usage of the IM test as a support tool to specification search in the context of a CUB model with covariates.

4.1 Student satisfaction

As an illustrative example, we consider the survey on student satisfaction for the Orientation services provided by the University of Naples Federico II in 2002. The data contain $n = 2179$ questionnaire responses, with ratings collected over $m = 7$ ordered categories[4]. We report the results of the IM test to check for the correct specification of a CUB model without covariates on the ratings expressed for global satisfaction and satisfaction on willingness of the staff, competence of the staff, information provided and office hours.

As can be seen from Table 8, results for nearly all ratings (with the only exception of `global`) indicate that the plain CUB(0,0) model is probably misspecified. In order to improve on the simple CUB(0,0) model, we consider a possible shelter effect: results reported in Table 9 indicate that indeed a CUB with specification of a shelter at category $c = 7$ can be assumed to be correctly specified at a 5% significance level (4% for `willingn`).

---

[4] This dataset is bundled within both the `R` and `gretl` libraries under the name `univer`. For a full description of the dataset, see Iannario et al. (2018).
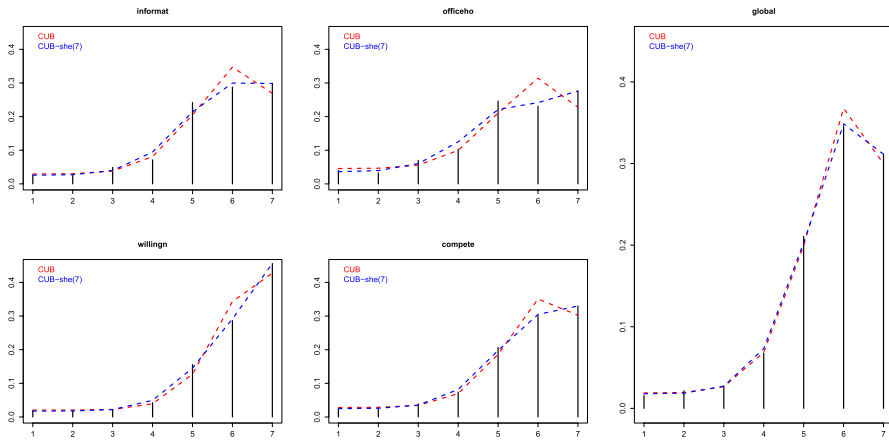
Fig. 13 Rating distributions within `univer` survey: estimated CUB and CUB with shelter at $c = 7$ probability models are superimposed to the bar plot of observed frequencies

For the `global` variable, instead, we use the LR test as a model selection procedure among the specifications that pass the IM test (see Table 8): for this survey item, the CUB(0,0) model without shelter can be accepted as a valid model specification. For the other items, matching results from the IM and LR tests shows that shelter category at $c = 7$ is the unique setting with both evidence for correct specification and goodness of fit (see Fig. 13).

4.2 The IM test as a support tool to model selection

As we argued in the introduction, practitioners almost invariably ignore the potential pitfalls stemming from misspecification and implicitly assume that ordinal data arise from a pre-specified distribution. This is often the case with CUB models, on the grounds of their attractiveness in terms of interpretation of parameters and parsimony.

In this section, we give a practical example on the way the IM test can be used to validate *ex post* a CUB model when its specification is chosen by relying on information criteria, as is common with models with covariates. In these cases, the process of variable selection for the most significant and relevant predictors of the feeling and uncertainty components is a challenging task. Classical backward and forward algorithms are not straightforward to apply since variable selection should proceed jointly for both model components. For this reason, in principle best-subset variable selection is a candidate algorithm to pursue a joint identification of response drivers.[5]

We use the BIC criterion here to conform to the vast majority of empirical applications of CUB models. In fact, recent research (Lv and Liu 2014) has proposed

---

[5] Clearly, an efficient computation strategy is needed. See (e.g. Simone 2021, 2020).

Table 10 Estimated parameters and standard errors for the best model (M_1) with respect to BIC criterion

	Estimates	SE	Wald
Uncertainty			
β_0	− 0.1336	0.1323	− 1.0095
β_1	1.1015	0.1917	5.7467
Feeling			
η_0	− 0.816	0.0439	− 18.581
η_1	0.232	0.0623	3.717
η_2	− 0.157	0.0314	− 5.016

modified versions of the BIC and AIC which take possible misspecification into account. Here, however, we try to stick as closely as possible to common empirical practice and just exemplify the usage of the IM test as a diagnostic procedure.

We consider a survey run by the Italian National Office of Statistics (ISTAT) on the professional placement of PhDs (see https://www.istat.it/en/archivio/87789), and we focus on the overall satisfaction in the doctoral experience.[6] All the ratings were collected on a scale with 11 ordered categories (from 0 to 10): the rating scale has been subsequently modified to a scale with 8 ordered categories because of zero-scores observed in certain categories, with higher scores corresponding to higher satisfaction levels. We consider the 2012 and 2014 surveys, and after omitting missing values for the variables of interest, we have $n = 2053$ and $n = 1777$ observations, respectively.

For this case study, several covariates are available, including subject-specific ones (gender, current employment status, residence, discipline of the PhD study, marital status and others) and PhD-specific ones (participation in research projects, geographical location of the University, a binary variable indicating if the PhD candidate took periods abroad, scholarships, standardized number of published papers during PhD courses, and more).

First, for the selected set of covariates, we estimate a CUB model with full covariate specification on both feeling and uncertainty parameters: the resulting model consists of several non-significant covariate effects. Therefore, we omit the non-significant effects and pursue a best-subset variable specification to select the best model using the BIC criterion. The resulting model is as follows:

$$\lambda(\pi_i) = \beta_0 + \beta_1 \text{STEM}_i$$
$$\lambda(\xi_i) = \eta_0 + \eta_1 \text{Gender}_i + \eta_2 \text{Npaper}_i;$$

parameter estimates are reported in Table 10:

Subsequently, we selected the models estimated within the best-subset search that are *closest* to the best one (M_1), according to the given criterion (in this case, those with low difference in BIC from the best one: $\Delta BIC < 5$), and perform the IM

---

[6] Satisfaction in the PhD experience was rated with reference to several aspects (quality of teaching courses, spaces and tools at disposal, etc): here we concentrate on overall satisfaction.

Table 11 Ranking of best models, with relevant information on IM test results

Model	Delta			
Ranking	BIC	IM statistic	df	p val
M_1	0	13.038	8	0.111
M_2	0.973	15.716	15	0.401
M_3	1.864	15.555	11	0.158
M_4	3.332	16.192	19	0.644
M_5	4.814	27.228	12	0.007

Table 12 Best fitting CUB models for which correct specification can be assumed at 5% significance level according to the IM test

	Estimates	SE	Wald
M_2 (*Loglik* $= -3951.77$)			
Uncertainty			
β_0	-0.518	0.210	-2.470
β: teaching	0.555	0.219	2.529
β: stem	1.073	0.193	5.533
Feeling			
η_0	-0.813	0.044	-18.621
η: gender	0.229	0.062	3.687
η: Npaper	-0.1587	0.0315	-5.040
M_3 (*Loglik* $= -3952.217$)			
Uncertainty			
β_0	-0.402	0.180	-2.229
β: gender	0.482	0.204	2.363
β: stem	1.202	0.201	5.976
Feeling			
η_0	-0.827	0.046	-17.893
η: gender	0.252	0.063	4.001
η: Npaper	-0.161	0.032	-5.080
M_4 : (*Loglik* $= -3949.137$)			
Uncertainty			
β_0	-0.759	0.239	-3.178
β: gender	0.463	0.205	2.264
β: teaching	0.538	0.220	2.447
β: stem	1.161	0.202	5.748
Feeling			
η_0	-0.825	0.0463	-17.969
η: gender	0.249	0.063	3.972
η: Npaper	-0.163	0.032	-5.092

test for all of them: Table 11 reports the relevant information. It turns out that the hypothesis of correct specification can be rejected only for model 5 (see Tables 11 and 12 for details on estimated models). In general, the IM test can be used as a

supplement procedure when performing model selection, being a tool to further investigate the appropriateness of a model beyond its fitting and predictive abilities.

5 Conclusions

In this paper, we study the application of the information matrix test to perform diagnostics of the correct specification of statistical models for rating data, with focus on the class of CUB mixture models to account for heterogeneity. Our approach is very general and makes it also possible to determine groups of models which are homogeneous with respect to fitting performances, by inspecting simultaneously which ones can be considered correctly specified according to the proposed testing procedure. Then, all attempts to perform multi-model inference should be preferably based on the subset of models that pass this check.

The code to perform the proposed testing procedure to check for correct specification of CUB models has been programmed for both R and Gretl environments: the procedures are available upon request from Authors, and they will be released soon on the official repositories.

Appendix A: Analytical derivatives: CUB models

Given the observed sample $(r_1, \dots, r_n)$, for notational convenience consider the following notation for $i = 1, \dots, n$:

$$p_i = Pr(R_i = r_i | \theta, y_i, w_i)$$

$$\tau_i = \frac{\pi_i \, b_{r_i}(\xi_i)}{p_i}$$

$$v_i = \frac{m - r_i}{\xi_i} - \frac{r_i - 1}{1 - \xi_i}$$

$$q_i = \frac{m - r_i}{\xi_i^2} + \frac{r_i - 1}{(1 - \xi_i)^2}$$

$$\widetilde{\xi}_i = \xi_i(1 - \xi_i)$$

$$\widetilde{\pi}_i = \pi_i(1 - \pi_i)$$

$$h_i = v_i(1 - 2\xi_i) - q_i \widetilde{\xi}_i + v_i^2 \widetilde{\xi}_i \frac{1 - \pi_i}{m \, p_i}$$

A.1 Score vector

- For $j = 0, \dots, p$, if $\widetilde{y}_i = (1, y_i)$, then:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{1}{p_i} \widetilde{y}_{ij}\, \widetilde{\pi}_i \left(b_{r_i}(\xi_i) - \frac{1}{m} \right)$$

- For $j = 0, \dots, q$, if $\widetilde{w}_i = (1, w_i)$, then:

$$\frac{\partial \ell_i}{\partial \eta_j} = \tau_i\, v_i\, \widetilde{w}_{ij}\, \widetilde{\xi}_i$$

A.2 Hessian matrix

- For $j, h = 0, \dots, p$, if $\widetilde{y}_i = (1, y_i)$, then:

$$\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_h} = \widetilde{y}_{ij}\, \widetilde{y}_{ih}\, \widetilde{\pi}_i \frac{b_{r_i}(\xi_i) - \frac{1}{m}}{p_i} \left[(1 - 2\pi_i) - \widetilde{\pi}_i \frac{\left(b_{r_i}(\xi_i) - \frac{1}{m} \right)}{p_i} \right]$$

- For $j, h, 0, \dots, q$, if $\widetilde{w}_i = (1, w_i)$, then:

$$\frac{\partial^2 \ell_i}{\partial \eta_j \partial \eta_h} = \widetilde{w}_{ij}\, \widetilde{w}_{ih}\, \tau_i\, \widetilde{\xi}_i\, h_i$$

- For $j = 0, \dots, q$, for $h, 0, \dots, p$, if $\widetilde{w}_i = (1, w_i)$ and $\widetilde{y}_i = (1, y_i)$, then:

$$\frac{\partial^2 \ell_i}{\partial \eta_j \partial \beta_h} = \frac{\widetilde{w}_{ij}\, \widetilde{y}_{ih}\, \widetilde{\xi}_i\, \widetilde{\pi}_i\, v_i\, b_{r_i}(\xi_i)}{m\, p_i^2}$$

Appendix B: Analytical score and Hessian for CUB with shelter

For the CUB with shelter specification, let:

$$\lambda(\delta_i) = w_i\, \alpha \tag{B1}$$

$$\lambda(\pi_i) = x_i\, \beta \tag{B2}$$

$$\lambda(\xi_i) = y_i\, \eta \tag{B3}$$

where y_i, x_i, w_i are row-vectors of by-subject covariates in which the first position is 1, so as to accommodate for intercepts $(\alpha_0, \beta_0, \eta_0)$.

Letting $v_i = (m - r_i)(1 - \xi_i) - (r_i - 1)\xi_i$, so that $\frac{\partial v_i}{\partial \eta_j} = -(m - 1)\, y_{ij}\, \xi_i(1 - \xi_i)$, then the score vector equals

$$\frac{\partial \ell_i}{\partial \alpha_k} = \frac{1}{\tilde{p}_i} \left(D_{r_i}^{(c)} - p_i \right) \delta_i (1 - \delta_i) w_{ik}$$

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{1}{\tilde{p}_i} (1 - \delta_i) \left(b_{r_i}(\xi_i) - \frac{1}{m} \right) \pi_i (1 - \pi_i) x_{ij}$$

$$\frac{\partial \ell_i}{\partial \eta_k} = \frac{1}{\tilde{p}_i} \pi_i (1 - \delta_i) b_{r_i}(\xi_i) y_{ik} v_i$$

As for the Hessian, we have

$$\frac{\partial^2 \ell_i}{\partial \alpha_k \partial \alpha_j} = \frac{(D_{r_i}^{(c)} - p_i)}{\tilde{p}_i} w_{ij} w_{ik} \delta_i (1 - \delta_i) \left(1 - 2\delta_i - \frac{\delta_i (1 - \delta_i)}{\tilde{p}_i} (D_{r_i}^{(c)} - p_i) \right)$$

$$\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_k} = \frac{(1 - \delta_i)}{\tilde{p}_i} x_{ij} x_{ik} \left(b_{r_i}(\xi_i) - \frac{1}{m} \right) \pi_i (1 - \pi_i) \times$$

$$\times \left[1 - 2\pi_i - \left(b_{r_i}(\xi_i) - \frac{1}{m} \right) \frac{(1 - \delta_i) \pi_i (1 - \pi_i)}{\tilde{p}_i} \right]$$

$$\frac{\partial^2 \ell_i}{\partial \alpha_k \partial \beta_j} = - \frac{\delta_i (1 - \delta_i)}{\tilde{p}_i} w_{ik} x_{ij} \left(b_{r_i}(\xi_i) - \frac{1}{m} \right) \pi_i (1 - \pi_i) \left[1 + (1 - \delta_i) \frac{(D_{r_i}^{(c)} - p_i)}{\tilde{p}_i} \right]$$

Using $\frac{\partial p_i}{\partial \eta_j} = \pi_i y_{ij} b_{r_i}(\xi_i) v_i$, we have

$$\frac{\partial^2 \ell_i}{\partial \alpha_k \partial \eta_j} = \frac{\delta_i (1 - \delta_i)}{\tilde{p}_i} w_{ik} \left(-(1 - \delta_i) \frac{(D_{r_i}^{(c)} - p_i)}{\tilde{p}_i} - 1 \right) \frac{\partial p_i}{\partial \eta_j}.$$

Finally, using $\frac{\partial b_{r_i}(\xi_i)}{\partial \eta_k} = y_{ik} b_{r_i}(\xi_i) v_i$ yields

$$\frac{\partial^2 \ell_i}{\partial \beta_j \partial \eta_k} = \frac{(1 - \delta_i) \pi_i (1 - \pi_i) x_{ij}}{\tilde{p}_i} \frac{\partial b_{r_i}(\xi_i)}{\partial \eta_k} \left[1 - \frac{(1 - \delta_i)}{\tilde{p}_i} \pi_i \left(b_{r_i}(\xi_i) - \frac{1}{m} \right) \right].$$

Appendix C: The CUB estimator as QMLE for the discretized Beta distribution

In this section, we offer a more detailed treatment of the situation analyzed in Sect. 3.2.3, where the asymptotic properties of the test statistic may break down because the pseudo-true parameter θ_* is on the boundary of the parameter space.

Therefore, we assume that the true model (DGP) is a discretized beta model DB(a, b), and the pseudo-true model (assumed under the null) is a CUB model when the DB model has a unique inner mode and a CUB with shelter when the DB(a, b) has a U-shaped distribution: see Simone (2022) for details on the DB model and its mixtures.
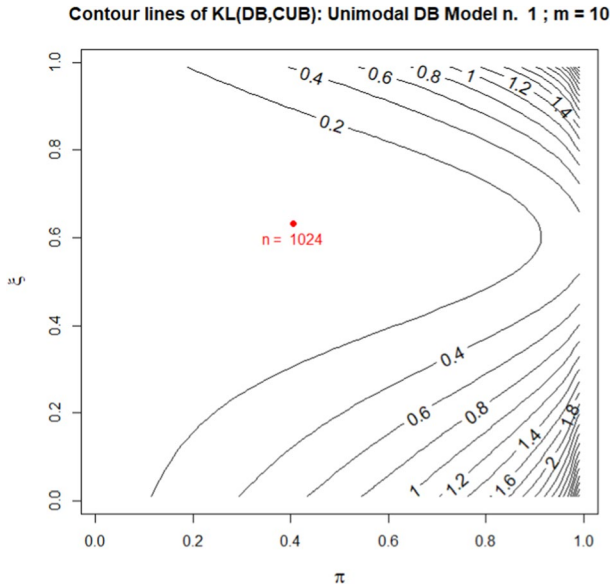
Contour lines of KL(DB,CUB): Unimodal DB Model n. 1 ; m = 10

Fig. 14 Contour lines of theoretical KL(DB, CUB) divergence; average sample estimates of (π, ξ) for $n = 1024$ highlighted in red

Since the QMLE $\hat{\theta}$ can be thought of as the minimizer of a criterion function whose probability limit is the KL divergence between the two distributions, we plot the contour lines of the Kullback–Leibler divergence

$$KL(\mathbf{p}, \mathbf{q}) = \sum_{r=1}^{m} p_r \log\left(\frac{p_r}{q_r}\right).$$

where $\mathbf{p}$ is the "true" vector of probabilities generated by the DGP (DB(a, b)) while $\mathbf{q}$ is the corresponding CUB model with parameters (π, ξ) ranging over the parameter space. This plot is meant to put in evidence the location of the pseudo-true parameter θ_*.

Assume first that $a = 1.5$ and $b = 2$: Fig. 14 displays contour lines of theoretical values of KL[DB(a, b), CUB(π, ξ)] as a function of (π, ξ) ranging over the parameter space. Red points indicate average estimated CUB parameters in the Monte Carlo experiment for selected sample sizes n. Results do not change for varying m and n: thus, for illustrative purposes, only $m = 10$ and $n = 1024$ are considered. As can be seen, the pseudo-true vector θ_* is in the interior of the parameter space and $\hat{\theta}$ is a proper QML estimator.

Next, assume that the true model $\mathbf{p}$ is a U-shaped DB model. In this case, the feeling parameter of a pseudo-true CUB model with shelter will be necessarily close to boundaries of the unit interval to model one of the two extreme modal values (the shelter category is $c = 1$ for Model 1, where $a = 0.5, b = 0.3$ and $c = m$ for Model 2, where $a = 0.2, b = 0.8$).
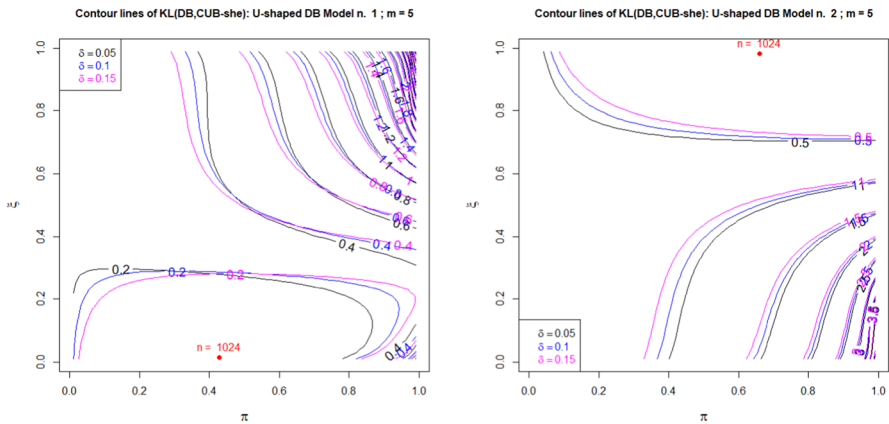
Fig. 15 Contour lines of theoretical divergence KL(DB, CUB − she)

Figures 15 display contour lines of the KL(DB, CUB − she) divergence, with model 1 in the left pane and model 2 in the right pane[7], for various levels of the shelter parameter δ. As can be seen in both cases the minimum is attained on the border of the parameter space.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

---

[7] We only display results for $m = 5$; with $m = 10$ the plots are very similar. Like in Fig. 14, estimated mean values of sampling estimators of (π, ξ) are displayed as red points for a selected sample size.

References

Agresti, A., Giordano, S., Gottard, A.: A review of score-test-based inference for categorical data. J. Quant. Econ. **20**, 31–48 (2022)

Biasetton, N., Disegna, M., Barzizza, E., Salmaso, L.: A new adaptive membership function with CUB uncertainty with application to cluster analysis of Likert-type data. Expert Syst. Appl. **213**, 118893 (2023)

Capecchi, S., Curtarelli, M.: A mixture model to assess perception of discrimination on grounds of sexual orientation for policy considerations. J. Appl. Stat. **47**(3), 554–567 (2020)

Capecchi, S., Simone, R.: A proposal for a model-based composite indicator: experience on perceived discrimination in Europe. Soc. Indic. Res. **141**, 95–110 (2019)

Capecchi, S., Endrizzi, I., Gasperi, F., Piccolo, D.: A multi-product approach for detecting subjects' and objects' covariates in consumer preferences. Br. Food J. **118**(3), 515–526 (2016)

Capecchi, S., Simone, R., Ghiselli, S.: Drivers and uncertainty for job satisfaction of the Italian graduates. Ital. J. Appl. Stat. **31**(2), 227–250 (2019a)

Capecchi, S., Meleddu, M., Pulina, M.: Quality evaluation and preferences of healthcare services: the case of telemedicine in Sardinia. Qual. Quant. **53**(5), 2339–2351 (2019b)

Cappelli, C., Simone, R., Di Iorio, F.: CUBREMOT: a tool for building model-based trees for ordinal responses. Expert Syst. Appl. **124**, 39–49 (2019)

Cerulli, G., Simone, R., Di Iorio, F., Piccolo, D., Baum, C.F.: The CUB STATA module: mixture models for feeling and uncertainty of rating data. Stand. Genom. Sci. **22**(1), 195–223 (2022). https://doi.org/10.1177/1536867X221083927

Chesher, A.: The information matrix test: simplified calculation via a score test interpretation. Econ. Lett. **13**(1), 45–48 (1983)

Colombi, R., Giordano, S.: Likelihood-based tests for a class of misspecified finite mixture models for ordinal categorical data. TEST **28**, 1175–1202 (2019)

Corduas, M.: Analyzing bivariate ordinal data with CUB margins. Stat. Model. **15**(5), 411–432 (2015)

Corduas, M.: Gender differences in the perception of inflation. J. Econ. Psychol. **90**, 102522 (2022)

Corduas, M., Cinquanta, L., Ievoli, C.: The importance of wine attributes for purchase decisions: a study of Italian consumers' perception. Food Qual. Prefer. **28**, 407–418 (2013)

Davidson, R., MacKinnon, J.G.: Artificial regressions. In: Baltagi, B. (ed.) A Companion to Theoretical Econometrics, pp. 16–37. Blackwell, Hoboken (2001)

D'Elia, A.: A statistical modelling approach for the analysis of TMD chronic pain data. Stat. Methods Med. Res. **17**(4), 389–403 (2008)

D'Elia, A., Piccolo, D.: A mixture model for preferences data analysis. Comput. Stat. Data Anal. **49**(3), 917–934 (2005)

Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B **39**(1), 1–38 (1977)

Di Nardo, E., Simone, R.: A model-based fuzzy analysis of questionnaires. Stat. Methods Appl. **28**, 187–215 (2019)

Finch, W.H., Hernández Finch, M.E.: Modeling of self-report behavior data using the generalized covariates in a uniform and shifted binomial mixture model. Psychol. Methods **25**, 113–127 (2020)

Golden, R., Henley, S., White, H., Kashner, T.: Generalized information matrix tests for detecting model misspecification. Econometrics **4**(4), 1–24 (2016). https://doi.org/10.3390/econometrics4040046

Horowitz, J.L.: Bootstrap-based critical values for the information matrix test. J. Econom. **61**, 365–411 (1994)

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

Iannario, M., Piccolo, D., Simone, R.: The R Package CUB: a Class of Mixture Models for Ordinal Rating data (2018). https://cran.r-project.org/web/packages/CUB/vignettes/CUBvignette-knitr.pdf

Jarque, C.M., Bera, A.K.: Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ. Lett. **6**(3), 255–259 (1980)

Lancaster, T.: The covariance matrix of the information matrix test. Econometrica **52**, 1051–1053 (1984)

Lucchetti, R., Pigini, C.: A test for bivariate normality with applications in microeconometric models. Stat. Methods Appl. **22**(4), 535–572 (2013)

Lucchetti, R., Pigini, C.: A simple and effective misspecification test for the double-hurdle model. Econ. Lett. **123**(1), 75–78 (2014)

Lv, J., Liu, J.S.: Model selection principles in misspecified models. J. R. Stat. Soc. Ser. B Stat. Methodol. **76**, 141–167 (2014)

Manisera, M., Zuccolotto, P.: Modeling rating data with nonlinear CUB models. Comput. Stat. Data Anal. **78**, 100–118 (2014)

Manisera, M., Zuccolotto, P., Brentari, E.: How perceived variety impacts on choice satisfaction: a two-step approach using the CUB class of models and best-subset variable selection. Electron. J. Appl. Stat. Anal. **13**(2), 519–535 (2020)

McCullagh, P.: Regression models for ordinal data. J. Roy. Stat. Soc.: Ser. B (Methodol.) **42**(2), 109–127 (1980)

McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. Wiley Series in Probability and Statistics, vol. 656, 2nd edn. Wiley, New York (1997)

Orme, C.: The small-sample performance of the information-matrix test. J. Econom. **46**(3), 309–331 (1990)

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

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

Piccolo, D., D'Elia, A.: A new approach for modelling consumers? Preferences. Food Qual. Prefer. **19**, 247–259 (2008)

Piccolo, D., Simone, R.: The class of CUB models: statistical foundations, inferential issues and empirical evidence. Stat. Methods Appl. **28**, 389–435 (2019a)

Piccolo, D., Simone, R.: Rejoinder to the discussion of "The class of CUB models: statistical foundations, inferential issues and empirical evidence. Stat. Methods Appl. **28**, 477–493 (2019b)

Piccolo, D., Simone, R., Iannario, M.: Cumulative and CUB models for rating data: a comparative analysis. Int. Stat. Rev. **87**(2), 207–236 (2019)

Ribecco, N., D'Uggento, A.M., Labarile, A.: What influences the perception of immigration in Italian adolescents? An analysis with CUB models for rating data. Socioecon. Plann. Sci. **82**, 101295 (2022)

Simone, R., Di Iorio, F., Lucchetti, R.: CUB for Gretl. In: Di Iorio, F., Lucchetti, R. (eds.) Gretl 2019: Proceedings of the International Conference on the GNU Regression, Econometrics and Time Series Library, pp. 147–166. feDOA University Press, Naples (2019)

Simone, R.: FastCUB: Fast EM and Best-Subset Selection for CUB Models for Rating Data. R Package Version 0.0.2 (2020). https://CRAN.R-project.org/package=FastCUB

Simone, R.: An accelerated EM algorithm for mixture models with uncertainty for rating data. Comput. Stat. **36**, 691–714 (2021)

Simone, R.: On finite mixtures of discretized Beta model for ordered responses. TEST **31**, 828–855 (2022)

Simone, R., Corduas, M., Piccolo, D.: Dynamic modelling of price expectations and judgments. Metron **81**, 323–342 (2023)

Tovar, B., Boto-Garcìa, D., Pino, J.F.: Meeting externalities: the effects of educational training on support for tourism activities. Tour. Econ. (2023). https://doi.org/10.1177/13548166231185897

Ursino, M., Gasparini, M.: A new parsimonious model for ordinal longitudinal data with application to subjective evaluations of a gastrointestinal disease. Stat. Methods Med. Res. **27**(5), 1376–1396 (2018)

Venson, A.H., Jacinto, P.A., Sbicca, A.: Cognitive dissonance in the self-assessed health in Brazil: a CUB model analysis using 2013 National Health Survey Data. Integr. Psychol. Behav. Sci. **57**, 1284–1311 (2023)

White, H.: Maximum likelihood estimation of misspecified models. Econometrica **50**(1), 1–25 (1982)

Xu, H., Zhang, N.: From contextualizing to context-theorizing: assessing context effects in privacy research. Manag. Sci. (2020). https://doi.org/10.2139/ssrn.3624056