JOURNAL OF ALGEBRAIC STATISTICS
Vol. 8, No. 2, 2017, 1-28
ISSN 1309-3452 – www.jalgstat.com

Mixture models for rating data: the method of moments via Gröbner basis

Maria Iannario[1], Rosaria Simone[1,*]

[1] *Department of Political Sciences, University of Naples Federico II, Naples, Italy*

Abstract. A recent thread of research in ordinal data analysis involves a class of mixture models that designs the responses as the combination of the two main aspects driving the decision process: a *feeling* and an *uncertainty* components. This novel paradigm has been proven flexible to account also for overdispersion. In this context, Gröbner bases are exploited to estimate model parameters by implementing the method of moments. In order to strengthen the validity of the moment procedure so derived, alternatives parameter estimates are tested by means of a simulation experiment. Results show that the moment estimators are satisfactory *per se*, and that they significantly reduce the bias and perform more efficiently than others when they are set as starting values for the Expectation-Maximization algorithm.

2000 Mathematics Subject Classifications: 62F10, 13P10, 13P25

Key Words and Phrases: Ordinal data, Gröbner basis, Method of moments, Uncertainty, Overdispersion

1. Introduction

When administering a survey, respondents are asked to express a preference in a list of ordinal categories, corresponding to levels of agreement/pleasantness/involvement with respect to a given item. In this regard, CUB models' paradigm [33, 13, 23] prescribes that the data generating process yielding to the evaluation from an underlying latent trait is basically structured as the combination of two elements: the actual *feeling* towards the item and an inherent *uncertainty* concerning the final rating [43]. The decision-making process results in a two-component mixture distribution, where the meditated behavior is shaped by a shifted Binomial random variable and the uncertainty is modelled by a discrete Uniform distribution, thus the acronym CUB standing for *C*ombination of a *U*niform and a shifted *B*inomal random variable. In this sense, CUB models directly shape the latent components of the decision process, offering an alternative setting to classical methods for the analysis of ordinal data (see, for instance, [1, 30]) in which an added value is the

*Corresponding author.

Email addresses: maria.iannario@unina.it (M. Iannario), rosaria.simone@unina.it (R. Simone)

explicit specification of uncertainty.

CUB models approach has been extended in several directions to perform more refined analysis, for instance to account for inflated categories [19], and for overdispersion, which leads us to the main focus of the present paper. CUBE models [20, 22] have been introduced to deal with the *overdispersion effect*, occurring when an excess of variability is observed among subjects: as a consequence, the feeling component is adjusted and then modelled via a shifted Beta-Binomial distribution. Several attempts have been pursued in the statistical literature to understand and model the overdispersion phenomenon by means of the Beta-Binomial distribution [8, 15, 28, 42]. In addition, [42] focuses on estimation issues for location and overdispersion parameters of the Beta distribution, with comparisons between the Maximum Likelihood method (ML) and an unbiased version of the method of moments. For estimation of parameters of the Beta-Binomial distribution, see also [41]. In this paper an original methodology to tackle moment estimation for CUB and CUBE models is proposed. In order to solve the system of moment equations, a classical choice would be to implement a Newton-Raphson algorithm in a suitable variant. However, such a procedure does not guarantee convergence within the boundaries of the parameter space and, additionally, it is very sensitive to the choice of starting values. For this reason and since the system of the moment equations simply lends itself to be algebraically manipulated and reduced to polynomial form, a Gröbner bases approach can be pursued [9, 37]. This technique yields to a system in a simplified form that can be solved with a forward or backward substitution procedure, allowing in principle to derive an explicit expression for the moment estimates.

As for any finite mixture model, inference about CUB and CUBE models is currently based on the maximum likelihood methods and relies on the Expectation-Maximization (EM) algorithm [31, 32], whose slowness can be sometimes discouraging. Indeed, both experience and literature suggest that, for general mixtures, it is usually preferable to run preliminary iterations of the EM algorithm starting with different initial values in order to exclude possible local maxima. See [4] for a focus on initialization procedures and [25] for a discussion of possible improvements of the algorithm. For CUB and CUBE models, devoted studies have been addressed, also when subjects' characteristics are taken into account [34, 35].

The ultimate idea of the paper is to derive the moment estimators of parameters for CUB and CUBE models in an algebraic setting -see [3] for a study on moment varieties for mixtures of multivariate Gaussian distributions grounded on tools from computational algebra. Then, in the more involved setting afforded by CUBE models, the core of the presentation is to test their performances when compared to other methods. Classical techniques for selecting preliminary estimates for a statistical model include a grid search across the parameter space, random choices, as well as short preliminary runs of the EM algorithm for finite mixture models, or compound methods. In the present work and specifically for CUBE models, these standard proposals are compared with both moment and *naive* estimators, the latter consisting in model-based parameter estimates, computed after an exploratory data analysis. After a brief theoretical discussion on advantages and disad-

vantages for each alternative, this task is pursued by running a Monte Carlo experiment in which all methods are applied on simulated data sampled from selected CUBE distributions. In this respect, the simulation results support that the moment estimators significantly reduce the bias and improve efficiency as compared with other proposals; these conclusions are supported by [29] distance and Kullback-Leibler divergence with respect to true parameter vectors. Then, to further validate the moment proposal, ML estimation is performed by running the EM algorithm initialized by using the different preliminary estimators. Thus, as a by-product of this study, the advantages of the derived moment proposal can be assessed also in terms of acceleration of the ML procedure, as indicated by the mean number of iterations needed to achieve convergence, the bias and root mean square errors of final ML estimates. It is worth to mention that similar comparative studies have been carried out for other mixture models, as for the Normal and Poisson cases [5, 26, 16], as well as for CUB models in [18]. Initial values for the overdispersion parameter can be derived by exploiting consistent estimators of the Beta-Binomial parameters as those proposed by [44], [45] and [46]: see [2] for estimation of overdispersion based on likelihood methods in the framework of generalized linear models.

The paper is organized as follows: in Section 2 CUBE models are shortly described, whereas the proposed method of moments is discussed in Section 3. Section 4 is devoted to the joint preliminary estimation of model parameters based on grid, naive, mixing and random estimators. The design and the results of the planned simulation study are discussed in Sections 5 and 6: then, final considerations, a real application and hints on future developments end the paper (the Appendix addresses a short overview of the theory of Gröbner basis). The methodology discussed in the paper has required both Maple Software (Version 15) and the `R` programming environment: in particular, the simulation study has been performed by means of the `R` package `CUB` available on CRAN [24], devoted to fit and test ordinal data within CUB models and their extensions.

2. Mixture models for uncertainty and overdispersion

CUB models [13] were devised to mimic the data generating process letting respondents produce an evaluation over an ordinal scale, say with $J > 3$ categories. The discrete choice process disentangling the latent perception is designed as a mixture distribution with parameters $\boldsymbol{\theta}' = (\pi, \xi)'$ of a shifted Binomial distribution with parameters J and ξ:

$$b_r(\xi) = \binom{J-1}{r-1} \xi^{J-r} (1-\xi)^{r-1}\,,\ r = 1, 2, \ldots, J\,, \tag{1}$$

and a discrete Uniform distribution over the support $\{1, 2, \ldots, J\}$. Then, a CUB random variable $R_{\text{CUB}} \sim \text{CUB}\ (\pi, \xi)$ has probability distribution given by:

$$p_r(\boldsymbol{\theta}) = \Pr(R_{\text{CUB}} = r \mid \boldsymbol{\theta}) = \pi\, b_r(\xi) + (1-\pi)\,\frac{1}{J}\,,\ r = 1, 2, \ldots, J\,. \tag{2}$$

The parameter ξ is referred to as the *feeling* parameter: indeed, $1 - \xi$ can be interpreted as the probability of each category being preferred over the previous ones in a pairwise

comparison procedure [11], while the uncertainty parameter π, acting as reverse weight to the Uniform distribution, is responsible for observed heterogeneity. Both empirical evidence and methodological studies indicate that the model specification benefits from the inclusion of the uncertainty component: indeed, the Uniform distribution accounts for the unavoidable fuzziness produced in the discretization of a latent perception into categories, as well as for the respondents' indecision, response styles, ignorance of the subject, laziness of respondents and other nuisances.

Starting from CUB models, the specification of CUBE models [20] was motivated by the necessity of directly shaping the *overdispersion effect*: this arises when an excess of inter-subjects variability occurs with respect to that accounted by the shifted Binomial component (for a discussion on the genesis of the overdispersion effect in CUBE models, see [21]). More specifically, overdispersion and uncertainty are different sources of variability in the data: confusing the two effects might result in mis-specified models, thus both components deserve an *ad-hoc* specification. Indeed, uncertainty is a measure of the heterogeneity of the distribution, while overdispersion is mainly related to the mutual variability in the sense of the Gini mean difference (as discussed in Subsection 4.2). The importance of isolating location and dispersion in the analysis of ordinal data is highlighted also in [42], where the Authors' motivation supporting the choice of the Beta distribution as a model for the latent response in overdispersed ordinal data meets the rationale of CUBE models.

For $J > 4$, a CUBE random variable R with parameters $\boldsymbol{\theta}' = (\pi, \xi, \phi)'$ is obtained from (2) by considering a random variable X with the shifted Beta-Binomial distribution of parameters (ξ, ϕ) for the feeling component:

$$p_r(\boldsymbol{\theta}) = \Pr(R = r \mid \boldsymbol{\theta}) = \pi \Pr(X = r|\xi, \phi) + (1 - \pi)\frac{1}{J}, \tag{3}$$

where $\Pr(X = r|\xi, \phi)$, for $r = 1, \ldots, J$, is parameterized as:

$$\Pr(X = r|\xi, \phi) = \binom{J-1}{r-1} \frac{\prod_{k=1}^{r} [1 - \xi + \phi(k-1)] \prod_{k=1}^{J-r+1} [\xi + \phi(k-1)]}{[1 - \xi + \phi(r-1)]\,[\xi + \phi(J-r)] \prod_{k=1}^{J-1} [1 + \phi(k-1)]}, r = 1, \ldots, J. \tag{4}$$

For a general presentation of the Beta-Binomial model, see [41, 42]. The CUBE distribution (3) is well defined for $\boldsymbol{\theta}' = (\pi, \xi, \phi)'$ belonging to the parameter space:

$$\Omega(\boldsymbol{\theta}) = \{(\pi, \xi, \phi) : \ 0 < \pi \leq 1; \ 0 \leq \xi \leq 1; \ 0 \leq \phi < \infty\}. \tag{5}$$

The parameterization chosen in (4) highlights that CUB models are nested into CUBE, that is, when $\phi = 0$ the probability distribution (3) collapses to (2). Notice that the expectation of the CUBE random variable R is unaffected by the overdispersion parameter

ϕ:

$$\mathbb{E}\,(R) = \mathbb{E}(R_{CUB}) = \frac{J+1}{2} + \pi(J-1)\left(\frac{1}{2} - \xi\right). \tag{6}$$

Instead, the overdispersion strongly affects the variance:

$$\mathrm{Var}(R) = \mathrm{Var}(R_{CUB}) \; + \; \phi_J(\boldsymbol{\theta})\,, \tag{7}$$

where

$$\mathrm{Var}(R_{CUB}) = (J-1)\left[\pi\,\xi\,(1-\xi)\left[\pi\,(J-1)-(J-2)\right] + (1-\pi)\,\frac{3\,\pi\,(J-1)+(J+1)}{12}\right] \tag{8}$$

is the variance of a CUB model with the same $(\pi, \xi)'$ parameters and $\phi_J(\boldsymbol{\theta})$ is the *overdispersion effect*:

$$\phi_J(\boldsymbol{\theta}) = \pi\,\xi\,(1-\xi)\,(J-1)\,(J-2)\,\frac{\phi}{1+\phi}\,. \tag{9}$$

It follows that the variance of R increases with ϕ (and π) by a quantity which, *ceteris paribus*, is maximized when $\xi = \frac{1}{2}$. When ξ tends to 0 or to 1, or when π tends to 0, one fails to capture the overdispersion effect. This circumstance indicates that, in case of predominance of extreme feelings or maximum heterogeneity, the final result is not so affected by any inter-variability among the subjective selection of the categories (as confirmed by the simulation experiment presented in Sections 5 and 6).

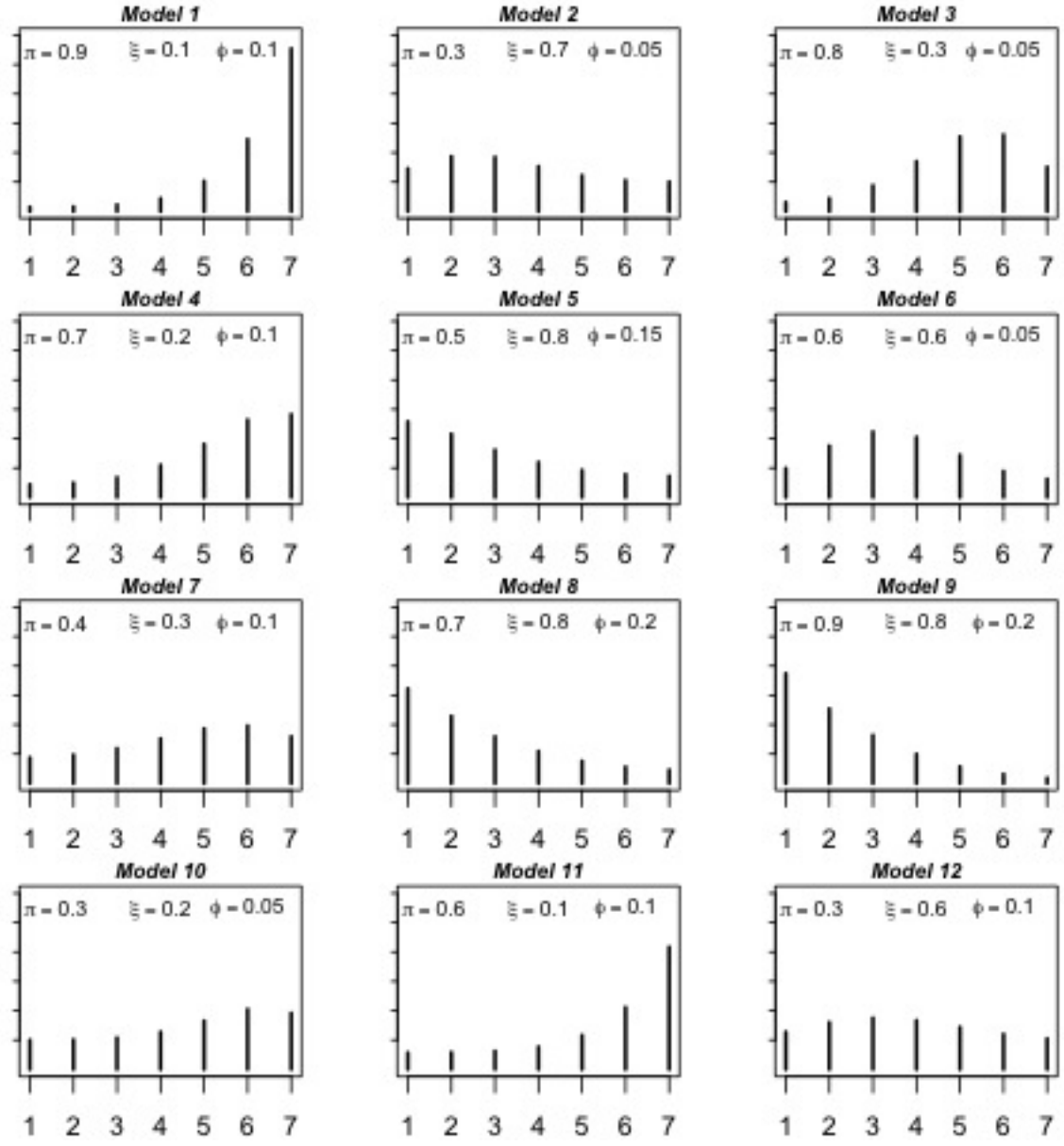
Hereafter, the focus will be on unimodal distributions: thus, we have to require $\phi < 0.5$ [21] in order to have CUBE models with mode at an intermediate category. Moreover, it is reasonable to assume a measure of overdispersion that does not exceed the threshold $\phi \leq 0.3$: this choice prevents from confounding effect with the latent uncertainty. To deal with multimodal distributions, specification of subjects' covariates is more appropriate, as discussed in [35].

In Figure 1, some unimodal CUBE distributions over $J = 7$ categories are shown with varying parameters, highlighting the wide flexibility of CUBE models in fitting observed distributions with different shapes and features. Parameter values have been chosen to range the whole parameter space with several combinations (small, medium, large) of uncertainty, feeling and overdispersion components. These models will be the object of the simulation study discussed in Sections 5 and 6.

3. Moment estimators

As a matter of fact, for CUB models it is quite straightforward to obtain the moment estimates from the first two moment equations (by substitution) corresponding to given data. The method of moments for CUB models has been directly applied in [33] to obtain estimates of uncertainty and feeling parameters π and ξ respectively; in addition, a comparative analysis with the ML approach has been run in a real case study. For

Figure 1: CUBE probability distributions for varying π, ξ, ϕ ($J = 7$)



CUBE models, instead, a direct approach yields to cumbersome computation since high order polynomials are involved. In what follows, a strategy to tackle moment estimation is proposed by considering the theory of Gröbner basis. In order to work in a homogeneous framework, our strategy for the forthcoming simulation study is to exploit the Gröbner basis proposal also to derive moment estimates for CUB mdoels: see the Appendix for related computational details.

3.1. CUBE distribution: theoretical moments

Let $BB \sim BetaBin(\alpha, \beta, J-1)$ be a Beta-Binomial random variable with parameters $\alpha, \beta > 0$, over the support $\{0, 1, \ldots, J-1\}$, and probability mass function:

$$\Pr(BB = r) \;=\; \binom{J-1}{r} \frac{Be(r+\alpha, J-1-r+\beta)}{Be(\alpha,\beta)}\,, \qquad r = 0, 1, \ldots, J-1\,,$$

where $Be(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ denotes the Euler Beta function [36].
If $(\alpha+\beta)^{(k)}$ denotes the rising factorial:

$$(\alpha+\beta)^{(k)} = (\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+k-1), \tag{10}$$

the first three moments of BB are given by:

$$\begin{aligned}
\mathbb{E}\,(BB) &= (J-1)\,\alpha\,\frac{1}{\alpha+\beta}\,,\\
\mathbb{E}\left(BB^2\right) &= (J-1)\,\alpha\,\frac{\left[(J-1)\,(1+\alpha)+\beta\right]}{(\alpha+\beta)^{(2)}}\,,\\
\mathbb{E}\left(BB^3\right) &= (J-1)\,\alpha\,\frac{(J-1)(1+\alpha)\left[(J-1)(2+\alpha)+3\beta\right]+\beta(\beta-\alpha)}{(\alpha+\beta)^{(3)}}.
\end{aligned}$$

The selected parameterization (4) is preferred in the framework of CUBE models, and it arises from the one-to-one correspondence [20]:

$$\begin{cases} \xi = \dfrac{\beta}{\alpha+\beta} \\ \phi = \dfrac{1}{\alpha+\beta} \end{cases} \qquad \Longleftrightarrow \qquad \begin{cases} \alpha = \dfrac{1-\xi}{\phi} \\ \beta = \dfrac{\xi}{\phi}\,. \end{cases}$$

Given the latter parameterization, the first three moments of BB can be written as:

$$\begin{aligned}
\mathbb{E}\,(BB) &= (J-1)(1-\xi)\\
\mathbb{E}\left(BB^2\right) &= (J-1)(1-\xi)\frac{(J-1)(\phi+1)+\xi(J+2)}{1+\phi}\\
\mathbb{E}\left(BB^3\right) &= (J-1)(1-\xi)\frac{\left\{(J-1)(\phi+1-\xi)[(J-1)(2\phi+1-\xi)+3\xi]+\xi(2\xi-1)\right\}}{(1+\phi)(1+2\phi)}.
\end{aligned}$$

When shifting the support from $\{0, 1, \ldots, J-1\}$ to $\{1, \ldots, J\}$, the first three moments $\mu_1^{(X)}, \mu_2^{(X)}, \mu_3^{(X)}$ of the shifted Beta-Binomial random variable $X = BB+1$ can be written as:

$$\begin{aligned}
\mu_1^{(X)} &= A_0 - (J-1)A_1(\phi)\,\xi\,;\\
\mu_2^{(X)} &= B_0 - (J-1)B_1(\phi)\,\xi + (J-1)(J-2)B_2(\phi)\,\xi^2\,;
\end{aligned}$$

$$\mu_3^{(X)} \;=\; C_0 - (J-1)C_1(\phi)\,\xi + (J-1)(J-2)C_2(\phi)\,\xi^2 - (J-1)(J-2)(J-3)C_3(\phi)\,\xi^3\,;$$

where:

$$\begin{array}{lll} A_0 = J; & B_0 = J^2; & C_0 = J^3; \\ A_1(\phi) = 1\,; & B_1(\phi) = \dfrac{(J+1)\phi + (2J-1)}{1+\phi}\,; & C_1(\phi) = \dfrac{2(J^2+J+1)\phi^2 + 3(2J^2-1)\phi + (3J^2-3J+1)}{(1+\phi)(1+2\phi)}\,; \\ B_2(\phi) = \dfrac{1}{1+\phi}\,; & C_2(\phi) = 3\,\dfrac{(J+1)\phi + (J-1)}{(1+\phi)(1+2\phi)}\,; & C_3(\phi) = \dfrac{1}{(1+\phi)(1+2\phi)}\,. \end{array}$$

Let U denote a random variable with the discrete Uniform distribution over $\{1,\ldots,J\}$. Since:

$$\mu_1^{(U)} = \mathbb{E}\,(U) = \frac{J+1}{2}, \tag{11}$$

$$\mu_2^{(U)} = \mathbb{E}\left(U^2\right) = \frac{(J+1)(2J+1)}{6}, \tag{12}$$

$$\mu_3^{(U)} = \mathbb{E}\left(U^3\right) = \frac{J(J+1)^2}{4}, \tag{13}$$

the first three moments $\mu_1^{(R)}, \mu_2^{(R)}, \mu_3^{(R)}$ of a CUBE random variable R (3) are given by:

$$\mu_k^{(R)} = \pi\,\mathbb{E}\left(X^k\right) + (1-\pi)\,\mathbb{E}\left(U^k\right) \;=\; \pi\,\left[\mu_k^{(X)} - \mu_k^{(U)}\right] + \mu_k^{(U)}\,, \qquad k = 1,2,3,$$

namely:

$$\begin{array}{ll} \mu_1^{(R)} & = \pi\left[(A_0 - \mu_1^{(U)}) - (J-1)A_1(\phi)\,\xi\right] + \mu_1^{(U)}\,; \\ \mu_2^{(R)} & = \pi\left[(B_0 - \mu_2^{(U)}) - (J-1)B_1(\phi)\,\xi + (J-1)(J-2)B_2(\phi)\,\xi^2\right] + \mu_2^{(U)}\,; \\ \mu_3^{(R)} & = \pi\left[(C_0 - \mu_3^{(U)}) - (J-1)C_1(\phi)\,\xi + \pi\,(J-1)(J-2)C_2(\phi)\,\xi^2 + (J-1)(J-2)(J-3)C_3(\phi)\,\xi^3\right] + \mu_3^{(U)}\,. \end{array} \tag{14}$$

In the end, given a sample of observations $(r_1,\ldots,r_n)'$ drawn from a CUBE distribution with parameters $\boldsymbol{\theta}' = (\pi,\xi,\phi)'$, and the corresponding first three sample moments m_1, m_2, m_3:

$$m_j = \frac{1}{n}\sum_{k=1}^{n}(r_k)^j\,, \quad j = 1,2,3, \tag{15}$$

the solution $(\tilde{\pi},\,\tilde{\xi},\,\tilde{\phi})$ of the three non-linear equations:

$$\left\{\begin{array}{l} E_1(\pi,\xi,\phi) \;=\; \mu_1^{(R)}(\pi,\xi,\phi) - m_1 \;=\; 0 \\ E_2(\pi,\xi,\phi) \;=\; \mu_2^{(R)}(\pi,\xi,\phi) - m_2 \;=\; 0 \\ E_3(\pi,\xi,\phi) \;=\; \mu_3^{(R)}(\pi,\xi,\phi) - m_3 \;=\; 0 \end{array}\right. \tag{16}$$

gives the moment estimates for the parameter vector $\boldsymbol{\theta}$ [38, pp. 351]. When computing moment estimates one should keep in mind that they usually underperform the ML

estimates in terms of efficiency, but the method of moments yields always consistent estimates. A part from a relative lack of efficiency, one of the main drawbacks of the method of moment is that it may happen that estimates lay out of the parameter space, in which case they are meaningless. Furthermore, they are unique only under invertibility of the defining moment equations [10, 39]. The ML method, however, also implies several disadvantages: estimates can be biased for small samples and the procedure is sensitive to the choice of starting values for the optimization runs (sometimes generated with moment estimators). On the other hand, moment estimates are usually easier to determine. Our investigation will lead to the empirical conclusion that, for CUBE models, moment estimators computed from a suitable Gröbner basis are instead a fairly satisfactory choice. Most importantly, from the simulation study it can be advanced that a moment estimate can be always uniquely determined with the proposed methodology within the parameter space, thus supporting the desired regularity assumptions.

3.2. Moment estimates and Gröbner bases

Due to the regularity of $\mu_1^{(R)}, \mu_2^{(R)}, \mu_3^{(R)}$ on the parameter space as functions of π, ξ, ϕ, one can equivalently consider the nonlinear algebraic system obtained from (16) after reducing the moment equations to polynomial form, say:

$$\begin{cases} \tilde{E}_1(\pi, \xi, \phi) = 0 \\ \tilde{E}_2(\pi, \xi, \phi) = 0 \\ \tilde{E}_3(\pi, \xi, \phi) = 0 \end{cases} \tag{17}$$

whose coefficients are functions of J and of the sample moments m_1, m_2, m_3. The Gröbner bases approach is a symbolic technique to solve an algebraic system of equations: it consists in providing an algebraic system equivalent to (17) but in a simpler form [6, 40]: in few words, solving a system of nonlinear algebraic equations in the ring of polynomials $\mathbb{K}[x_1, \ldots, x_t]$:

$$E_i(x_1, \ldots, x_t) = 0, \quad i = 1, \ldots, s, \tag{18}$$

is equivalent to determine, if they exists, the solutions of the system:

$$G_j(x_1, \ldots, x_t) = 0, \quad j = 1, \ldots, q, \tag{19}$$

where $G_1, \ldots, G_q$ form the reduced Gröbner basis of the polynomial ideal generated by $E_1, \ldots, E_s$. Provided that one chooses a so-called *elimination ordering*, the resulting system could be solved easily by forward (or backward) substitution (see the Appendix for a concise overview).

3.2.1. Computational details

For a fixed number J of ordinal categories, the Gröbner basis associated with (17) has been computed via Maple Software (Version 15), library *Groebner*. For our purposes, the

best performances are obtained by considering the lexicographic order: $\pi > \xi > \phi$. In this case, the first parameter to be solved for is the overdispersion: this ordering is advisable because it recognizes a major importance to the estimation of ϕ and limits the propagation error. Indeed, mis-specification of ϕ yields biased estimates of both π and ξ; in addition, this choice performs more efficiently from a computational view if compared with other choices. Further comments on the selection of the ordering are postponed to the Appendix. As a matter of fact, the randomness of the sampling procedure induces a particularly high variability in the second and the third sample moments: for this reason, it may occur that the Gröbner basis associated with (17) is trivial, and consequently no moment solution exists. Indeed, the proposed strategy always uniquely determines the moment solution when considering theoretical moments in place of the sample moments, but it might fail at the simulation level. Empirical evidence suggests that this circumstance is mostly associated to a population ϕ lying close to the lower bound of its admissible range, that is for negligible overdispersion effect, thus indicating that a CUB model fit could be used instead. This problem could be overcome by looking for a different parameterization of the overdispersion: nevertheless, the chosen one is preferred due to easiness in interpretation. Thus, as an overall rule, at each step of the implementation, whenever the moment solution cannot be provided within the admissible range, the chosen strategy is to minimize the overdispersion effect by setting $\phi = 0.01$, and search for the moment solutions for π and ξ switching to the nested CUB models (see the Section 7 in the Appendix). Specifically, the percentage of times the procedure had to switch to the nested CUB models as indicated is about 5% for all models, expect for models corresponding to $\phi < 0.1$, for which such percentage is greater.

The symbolic approach afforded by the Maple library allows to claim that, when non-trivial, the reduced Gröbner basis for polynomials in (17) is always composed of four generators:

$$G_1(\phi) = a_0 + a_1\phi + a_2\phi^2 + \phi^3 \tag{20}$$

$$G_2(\xi, \phi) = b_0 + b_1\phi + b_2\phi^2 + \xi + \xi\phi \tag{21}$$

$$G_3(\xi, \phi) = c_0 + c_1\phi + c_2\phi^2 - \xi + \xi^2 \tag{22}$$

$$G_4(\pi, \xi, \phi) = d_0 + d_1\phi + d_2\phi^2 + d_3\xi + \pi \tag{23}$$

whose coefficients are functions of J and of the sample moments. Notice that this statement is the result of our extensive numerical experiment, run by means the Maple library and by selecting the Buchberger algorithm as option [†]. The procedure here implemented highlights possible interactions between theory and experiments in Gröbner basis use to overcome moment estimation pitfalls. Secondly, considering the system (20) - (23), a forward substitution strategy can be implemented. First, consider (20): it turns out that $G_1(\phi)$ always admits a root in $\phi = -1$; then, for $\tilde{G}_1(\phi) = G_1(\phi)(1 + \phi)^{-1}$, it is straightforward to determine if it admits solutions and, in this case, the relative values. At this point, a proper stepwise procedure has to be developed to account for all the possible occurrences:

[†]Such structure has been double-checked and validated also by means of the Mathematica software.

1. if no solution to $\tilde{G}_1(\phi) = 0$ exists, or if they exists but they are not admissible, switch to the nested CUB model.

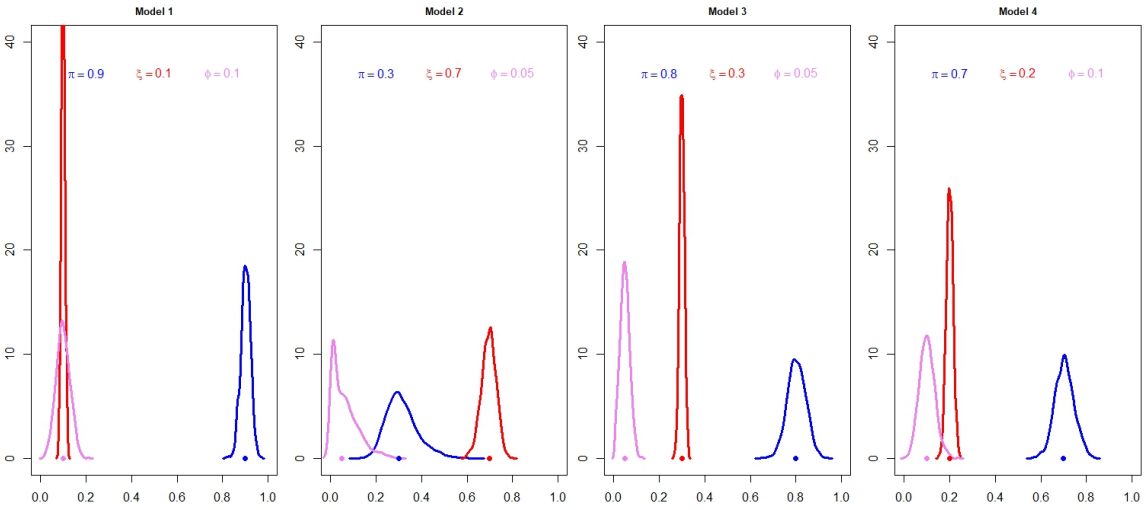
2. When there is exactly one admissible solution $\tilde{\phi}$ to $\tilde{G}_1(\phi) = 0$, then seek for the solutions of $G_2(\xi, \tilde{\phi}) = 0$ w.r.t. ξ within the parameter space.

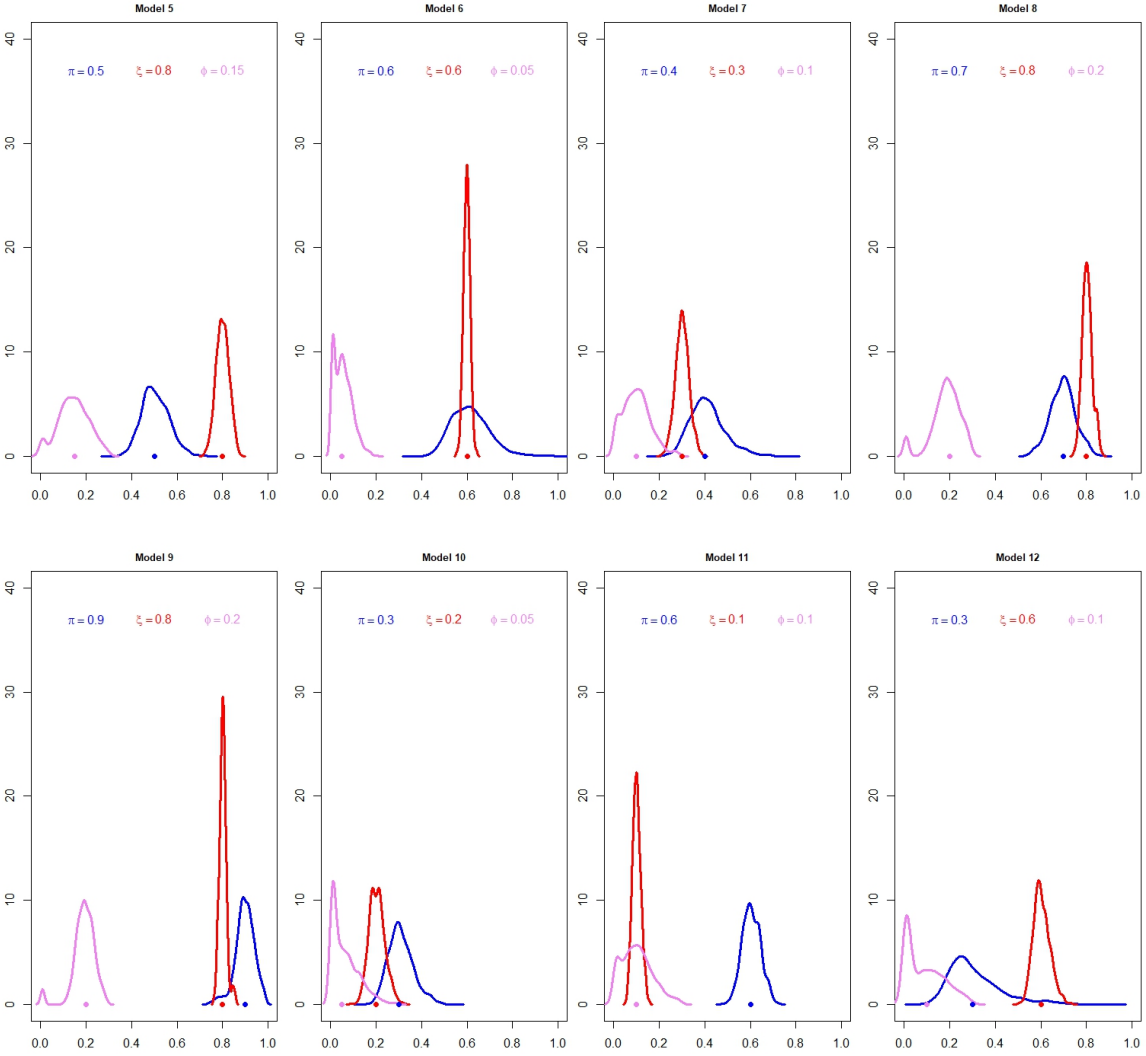
3. If an admissible moment estimate $\tilde{\xi}$ can be derived, then switch to the last equation $G_4(\pi, \tilde{\xi}, \tilde{\phi}) = 0$ to compute the moment estimate $\tilde{\pi}$.

4. If two solutions to $\tilde{G}_1(\phi) = 0$ exist, both of which are admissible, then proceed as explained in the previous steps for both of them. Thus, one would have two final solutions to the system: in this regard, the planned simulation study strongly supports the claim that exactly one of them would be admissible, yielding to a uniquely determined moment estimate.

As a final remark about the adopted solving procedure are needed, notice that it is not required to directly involve $G_3(\xi, \phi)$ into the solving procedure. Specifically, when an admissible solution $\tilde{\xi}$ is derived from (21), it verifies also $G_3(\tilde{\xi}, \tilde{\phi}) = 0$. Figure 2 displays the plotted kernel densities of the moment estimators corresponding to `nsimul` $= 1000$ simulations of samples, each of `n` $= 1000$ observations, generated from CUBE model on a $J = 7$ point ordered scale, illustrated in Figure 1. We acknowledge that, at this preliminary stage, moment estimation for π is smoother as its true value increases and ϕ decreases, that is the lower the variability in the data.

Figure 2: Moment estimator for CUBE models obtained via Gröbner basis





4. Preliminary estimators

The moment estimates derived in the previous section have an independent interest. Nevertheless we propose to challenge them comparatively with alternative estimation procedures, and then by further testing their performances when set as starting values for the EM algorithm to run maximum likelihood estimation.

Among the possible alternatives to obtain preliminary estimators of parameters of a CUBE model, we consider grid, *naive*, mixing and initial guesses derived by short preliminary runs of the EM algorithm. For all proposals, their motivations and their soundness are discussed, especially with reference to the *naive* choice. Random starting values are also considered for the sake of completeness.

4.1. Grid estimators

For a sample of ratings with absolute frequencies $n_1, \ldots, n_J$, the grid approach consists in obtaining an initial guess of parameter values by computing the log-likelihood function:

$$\ell(\boldsymbol{\theta}) = \sum_{r=1}^{J} n_r \log\left[\Pr(R = r|\boldsymbol{\theta})\right], \tag{24}$$

over a grid of few admissible values for the parameters belonging to $\Omega(\boldsymbol{\theta})$ (5), and then to choose those maximizing it. The grid approach has been advocated by [27] for finite mixtures and it has become very popular because it is easy to implement. Moreover, it is particularly suitable in frameworks where the parameter space has bounded range, and it is useful when questioning about possible local maxima issues. When such occurrence can be excluded, the denser the grid the more reliable are the resulting estimates.

In order to range all the parameter space and consider all possible combinations 'low/high' for uncertainty, feeling and overdispersion, the log-likelihood function (24) will be computed over the grid:

$$\Omega(\boldsymbol{\theta}_0) = N_\pi \times N_\xi \times N_\phi, \tag{25}$$

where N_π, N_ξ, N_ϕ are the chosen grid ranges for π, ξ and ϕ, respectively:

$$N_\pi = \{0.30, 0.50, 0.70\}; \tag{26}$$

$$N_\xi = \{0.20, 0.40, 0.80\}; \tag{27}$$

$$N_\phi = \{0.05, 0.15, 0.30\}. \tag{28}$$

Then, the *grid* estimator $\boldsymbol{\theta}^* = (\pi^*, \xi^*, \phi^*)'$ of the true parameter vector $\boldsymbol{\theta}$ is defined as:

$$(\pi^*, \xi^*, \phi^*) = \underset{\boldsymbol{\theta}_0 \in \Omega(\boldsymbol{\theta}_0)}{argmax}\ \ell(\boldsymbol{\theta}_0)\,. \tag{29}$$

4.2. Naive and mixing estimators

This subsection aims at designing preliminary estimators for parameters in CUBE models that can be directly computed starting from an exploratory analysis of available data and that are connected to the parameters features of the distribution: the *naive* estimators. In particular, the naive proposal relies on [18], where the Author discusses preliminary estimators for parameters in CUB models. For what concerns the feeling parameter, it can be derived on the basis of a location property of the sample mode M_n:

$$\xi_{\text{CUB}} = 1 + \frac{0.5 - M_n}{J}. \tag{30}$$

Let $\mathcal{G}$ denote the normalized Gini heterogeneity index for a given probability distribution $(p_1, \ldots, p_J)$:

$$\mathcal{G} = \frac{J}{J-1}\left(1 - \sum_{r=1}^{J} p_r^2\right), \tag{31}$$

attaining its maximum for the Uniform distribution: $\mathcal{G}_U = 1$. Then, the uncertainty parameter can be effectively estimated by:

$$\pi_{\text{CUB}} = \min\left(\sqrt{\frac{\mathcal{G}_U - \mathcal{G}_{Obs}}{\mathcal{G}_U - \mathcal{G}_{SB}(\xi_{\text{CUB}})}}, 1\right) \tag{32}$$

where $\mathcal{G}_U$, $\mathcal{G}_{Obs}$, $\mathcal{G}_{SB}(\xi_{\text{CUB}})$ denote the index (31) for the Uniform, the observed and the shifted Binomial distribution of parameter ξ_{CUB}, respectively.

Since CUB models are nested into CUBE, estimators (30) and (32) may be considered as naive choices for feeling and uncertainty parameters also for CUBE distributions. For deriving a naive estimate of the overdispersion parameter ϕ, empirical evidence [21] suggests to relate the overdispersion parameter to the mutual variability of responses among categories according to the following regression model:

$$\Delta(\pi, \xi, \phi) = a(\pi, \xi) + b(\pi, \xi)\sqrt{\phi}, \tag{33}$$

where $a(\pi, \xi)$, $b(\pi, \xi)$ are functions of the uncertainty and the feeling parameters, and $\Delta(\pi, \xi, \phi)$ denotes the mean difference index, expressed according to [14] formulation as:

$$\Delta = \sum_{r=1}^{J-1} F_r(1 - F_r), \tag{34}$$

with $F_r = \sum_{j=1}^{r} f_r$ is the empirical cumulative distribution function. The regression model (33) has been tested with a short simulation plan and it has shown nice performances in terms of goodness of fit when sampling from a theoretical Beta-Binomial distribution.

Back to CUBE models, the naive strategy requires to regress the Mean Difference computed for the CUBE probability distributions with feeling and uncertainty parameters (30) and (32), respectively, and let ϕ varying in the established admissible range. Finally, for the estimated intercept and slope, a naive estimate ϕ_Δ for ϕ is obtained by plugging the observed Mean Difference into model (33), in such a way that the naive estimator of true parameter vector $\boldsymbol{\theta}$ is given by:

$$\boldsymbol{\theta}'_{\text{naive}} = (\pi_{\text{CUB}}, \xi_{\text{CUB}}, \phi_\Delta)'. \tag{35}$$

4.2.1. Mixing estimators

A choice that takes advantages from both naive and grid estimation is to merge the two techniques into a mixing procedure in which the following *grid* for the true parameter vector $\boldsymbol{\theta}' = (\pi, \xi, \phi)'$ is considered:

$$\Omega(\boldsymbol{\theta}_{mix}) = N_\pi(\epsilon) \times N_\xi(\epsilon) \times N_\phi(\epsilon), \tag{36}$$

where ϵ is a small number (say $\epsilon = 0.1$) such that the grid ranges for parameters are set to:

$$N_\pi(\epsilon) = \{\pi_{\text{CUB}} - \epsilon, \pi_{\text{CUB}}, \pi_{\text{CUB}} + \epsilon\}; \tag{37}$$

$$N_\xi(\epsilon) = \{\xi_{\text{CUB}} - \epsilon, \xi_{\text{CUB}}, \xi_{\text{CUB}} + \epsilon\}; \quad (38)$$

$$N_\phi(\epsilon) = \{0.05, 0.10, 0.30\}. \quad (39)$$

After computing the log-likelihood function on the given grid, the *mixing* estimator of $\boldsymbol{\theta}$ is chosen as the parameter vector maximizing the log-likelihood $\ell(\boldsymbol{\theta})$ in (24):

$$\bar{\boldsymbol{\theta}} = (\bar{\pi}, \bar{\xi}, \bar{\phi}) = \underset{\boldsymbol{\theta}_0 \in \Omega(\boldsymbol{\theta}_{mix})}{argmax} \ \ell(\boldsymbol{\theta}_0) \ . \quad (40)$$

4.3. Random starting estimators

The most popular device for initializing the EM algorithm is to consider random starting values. For CUBE distributions, they are based on a continuous Uniform distribution for π and for ξ over the unit interval $(0, 1)$, whereas for the overdispersion parameter ϕ a continuous Uniform distribution $U(0, 0.3)$ over the interval $(0, 0.3)$ is assigned. Thus, given $\check{\pi} \sim U(0, 1)$, $\check{\xi} \sim U(0, 1)$, $\check{\phi} \sim U(0, 0.3)$, the proposed *random starting estimators* are $\check{\boldsymbol{\theta}} = (\check{\pi}, \check{\xi}, \check{\phi})'$.

4.4. Short runs of EM algorithm

A common and recognized practice used to initialize the EM algorithm is to consider ML estimates obtained from short preliminary runs of the algorithm itself [5], obtained by allowing a low maximum number of iterations and a high error tolerance to stop iterations when comparing the log-likelihood values in subsequent steps [‡]. Considering that ML convergence can be quite slow and that the choice of preliminary values should not to be too demanding at this stage, within the planned simulation study the initial value of each parameter will be set to the midpoint of its admissible range:

$$\mathring{\pi} = 0.50, \quad \mathring{\xi} = 0.50, \quad \mathring{\phi} = 0.15. \quad (41)$$

5. Comparative analysis

After having replicated the experiment for different numbers of categories ($J = 5, 7, 9, 10$), it results that the performances of the different methods do not substantially depend upon the value J: thus, the analysis here reported concerns 1000 simulations of samples of size $n = 1000$ when $J = 7$ (other simulation results are available from Authors upon request).

The convenience of the moment proposal over the other alternatives appears evident also at a preliminary stage, as indicated by the bias and the root mean square errors (RMSE) reported in Tables 1 and 2. For each method and each model of the 12 models illustrated in Figure 1, the best results are highlighted in bold. Note that the mixing and the naive alternatives provide satisfactory performances for several models, as well as the grid approach. In particular, when some components of the true parameter vector fall

[‡]In the simulation plan to follow, `maxiter = 3`, `toler = 1e-2`.

Table 1: Bias of preliminary estimators

Model	Moment (π,ξ,ϕ)	Naive (π,ξ,ϕ)	Grid (π,ξ,ϕ)	Mixing (π,ξ,ϕ)	Random (π,ξ,ϕ)	Short run EM (π,ξ,ϕ)
1	(**0.001**, **0.000**, **0.000**)	(−0.084, −0.029, 0.084)	(−0.200, 0.100, 0.200)	(−0.004, −0.001, 0.004)	(−0.410, 0.416, 0.054)	(−0.212, −0.032, −0.056)
2	(0.016, **-0.005**, 0.014)	(−0.037, 0.024, 0.019)	(**0.000**, 0.098, **0.001**)	(0.033, −0.018, 0.040)	(0.200, −0.209, 0.102)	(0.183, −0.096, 0.180)
3	(**0.003**, **0.001**, 0.001)	(−0.131, −0.032, −0.020)	(−0.127, −0.023, 0.009)	(−0.057, 0.005, **0.000**)	(−0.298, 0.201, 0.107)	(−0.207, −0.006, −0.037)
4	(0.004, 0.001, **0.002**)	(−0.228, −0.096, 0.140)	(**0.000**, **0.000**, 0.008)	(−0.139, −0.019, −0.037)	(−0.202, 0.319, 0.053)	(−0.142, −0.002, 0.009)
5	(0.010, −0.003, **0.003**)	(−0.199, 0.126, 0.146)	(**0.000**, **0.000**, 0.004)	(−0.101, 0.028, −0.048)	(0.002, −0.305, 0.009)	(**0.000**, −0.055, 0.121)
6	(0.015, **-0.001**, 0.006)	(−0.095, 0.019, −0.042)	(−0.300, 0.022, **0.000**)	(**-0.006**, 0.014, **0.000**)	(−0.106, −0.102, 0.106)	(−0.059, −0.011, −0.001)
7	(**0.016**, **0.003**, **0.006**)	(−0.095, −0.045, −0.042)	(0.195, 0.056, 0.139)	(**-0.016**, 0.004, 0.007)	(0.083, 0.208, 0.056)	(0.091, 0.066, 0.108)
8	(0.007, −0.001, **0.000**)	(−0.280, 0.129, 0.100)	(**0.000**, **0.000**, −0.008)	(−0.180, 0.029, −0.061)	(−0.206, −0.305, −0.045)	(−0.157, 0.012, −0.001)
9	(**0.001**, **0.000**, **-0.003**)	(−0.361, 0.129, 0.100)	(−0.200, **0.000**, −0.045)	(−0.261, 0.031, −0.067)	(−0.393, −0.294, −0.043)	(−0.294, 0.052, −0.094)
10	(0.011, 0.006, 0.016)	(−0.025, −0.015, 0.074)	(0.002, **0.001**, 0.012)	(**-0.001**, 0.006, **0.010**)	(0.194, 0.297, 0.101)	(0.175, 0.141, 0.242)
11	(0.006, **0.003**, **0.008**)	(−0.056, −0.029, 0.038)	(0.100, 0.100, 0.200)	(**0.002**, −0.009, −0.014)	(−0.102, 0.401, 0.052)	(−0.055, 0.027, 0.149)
12	(**0.028**, **0.000**, **0.001**)	(−0.074, 0.026, −0.047)	(**0.000**, −0.112, 0.050)	(−0.036, 0.016, −0.014)	(0.211, −0.086, 0.056)	(0.184, −0.048, 0.119)

Table 2: RMSE of preliminary estimators

Model	Moment (π,ξ,ϕ)	Naive (π,ξ,ϕ)	Grid (π,ξ,ϕ)	Mixing (π,ξ,ϕ)	Random (π,ξ,ϕ)	Short run EM (π,ξ,ϕ)
1	(0.021, 0.008, **0.030**)	(0.087, 0.029, 0.139)	(0.200, 0.100, 0.200)	(**0.013**, **0.005**, 0.050)	(0.497, 0.506, 0.100)	(0.212, 0.033, 0.057)
2	(0.074, **0.033**, 0.059)	(**0.057**, 0.079, 0.054)	(0.000, 0.102, **0.016**)	(0.064, 0.040, 0.064)	(0.347, 0.347, 0.131)	(0.183, 0.097, 0.183)
3	(**0.040**, **0.011**, 0.021)	(0.149, 0.076, 0.024)	(0.145, 0.100, 0.030)	(0.076, 0.032, **0.007**)	(0.412, 0.344, 0.136)	(0.208, **0.011**, 0.037)
4	(0.042, 0.015, 0.034)	(0.251, 0.113, 0.178)	(**0.006**, **0.000**, 0.050)	(0.165, 0.026, 0.050)	(0.345, 0.422, 0.100)	(0.142, 0.010, **0.024**)
5	(0.064, 0.029, 0.068)	(0.201, 0.127, 0.149)	(**0.009**, **0.000**, **0.041**)	(0.103, 0.028, 0.070)	(0.279, 0.419, 0.085)	(0.008, 0.056, 0.124)
6	(0.085, **0.014**, 0.039)	(0.103, 0.058, 0.044)	(0.300, 0.200, **0.003**)	(**0.049**, 0.043, **0.003**)	(0.299, 0.302, 0.135)	(0.060, 0.016, 0.016)
7	(0.086, **0.030**, 0.063)	(0.105, 0.082, 0.064)	(0.257, 0.100, 0.174)	(**0.051**, 0.039, **0.051**)	(0.292, 0.352, 0.099)	(0.091, 0.068, 0.112)
8	(0.058, 0.022, 0.057)	(0.281, 0.129, 0.100)	(**0.009**, **0.000**, 0.068)	(0.181, 0.030, 0.070)	(0.350, 0.416, 0.095)	(0.158, 0.016, **0.031**)
9	(**0.041**, 0.015, **0.044**)	(0.362, 0.129, 0.100)	(0.200, **0.000**, 0.053)	(0.262, 0.033, 0.077)	(0.488, 0.407, 0.093)	(0.294, 0.053, 0.096)
10	(0.057, 0.036, 0.063)	(0.062, 0.065, 0.132)	(0.025, 0.013, 0.037)	(**0.032**, **0.023**, **0.032**)	(0.344, 0.409, 0.131)	(0.175, 0.142, 0.242)
11	(**0.041**, 0.019, 0.071)	(0.061, 0.029, 0.121)	(0.100, 0.100, 0.200)	(0.050, **0.016**, **0.050**)	(0.297, 0.487, 0.096)	(0.056, 0.028, 0.151)
12	(0.142, **0.036**, 0.090)	(0.085, 0.095, 0.098)	(**0.000**, 0.200, 0.094)	(0.075, 0.053, **0.051**)	(0.351, 0.294, 0.103)	(0.184, 0.051, 0.124)

on the fixed grid, obviously the resulting grid estimator is unbiased. Nevertheless, the corresponding moment estimates behaves likewise.

An efficient indicator of the goodness of the estimated parameter vector $\hat{\boldsymbol{\theta}}$ is the Mahalanobis distance to the true one $\boldsymbol{\theta}$:

$$\text{dist}_M(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sqrt{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{V}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}. \tag{42}$$

where $\mathbf{V}$ denotes the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ [29]. The Mahalanobis distance is suitable for taking into account the usually non-negligible correlation between parameters and for its asymptotical distributional properties. The results reported in Table 3 show the relative convenience derived by introducing the moment estimators with respect to the other methods. The `NA` values corresponds to non-positive definite variance-covariance matrix: in this respect, note that the proposal behaves smoothly and prevents the occurrence of non-admissible values.

The symmetrized Kullback-Leibler divergence has been also computed in order to get an overall measure of the discrepancies between the estimated probability distribution and the theoretical one. Recall that, given two (discrete) probability distributions p, q over the

same support (here, $\{1, \ldots, J\}$), the Kullback-Leibler divergence is defined by:

$$D_{KL}(p|q) = \sum_{r=1}^{J} log\left(\frac{p_r}{q_r}\right) p_r.$$

$D_{KL}(p|q)$ is not a metric: in particular it is not symmetric, and hence one usually symmetrizes it by considering:

$$\tilde{D}_{KL}(p, q) = \frac{1}{2} D_{KL}(p|q) + \frac{1}{2} D_{KL}(q|p).$$

Table 3: Mahalanobis Distance and Kullback-Leibler divergence

	Mahalanobis distance						Kullback-Leibler divergence					
Model	Moment	Naive	Grid	Mixing	Random	Short run EM	Moment	Naive	Grid	Mixing	Random	Short run EM
1	**3.002**	NA	NA	3.163	7.623	2035.342	0.**0017**	0.0139	0.1733	0.0035	0.9636	0.0475
2	**3.057**	3.838	NA	3.753	5.579	6615.974	**0.0014**	0.0047	0.0116	0.0020	0.2003	0.0041
3	**3.003**	27.899	6.759	4.509	6.152	1025.594	**0.0016**	0.0326	0.0616	0.0094	0.4016	0.0158
4	**3.007**	22.953	NA	7.320	5.384	647.523	0.**0016**	0.0495	0.0026	0.0109	0.4201	0.0124
5	**3.032**	79.140	NA	29.030	4.129	107.261	0.0017	0.0203	**0.0006**	0.0036	0.2824	0.0055
6	**3.031**	18.933	NA	3.129	4.888	142.768	0.**0015**	0.0092	0.0755	0.0062	0.2622	0.0021
7	**3.046**	10.517	9.313	3.352	4.065	1420.891	**0.0015**	0.0061	0.0054	0.0023	0.2047	0.0033
8	**3.028**	NA	NA	112.183	4.938	1286.176	**0.0017**	0.0328	0.0023	0.0105	0.3915	0.0108
9	**3.019**	NA	NA	225.734	6.382	4413.853	**0.0019**	0.0710	0.0275	0.0333	0.5829	0.0355
10	**3.071**	3.741	3.134	3.169	5.982	2887.882	0.0015	0.0052	**0.0004**	0.0011	0.2162	0.0063
11	**3.021**	NA	NA	4.118	5.809	47.824	**0.0020**	0.0039	0.0160	0.0036	0.4580	0.0103
12	**3.168**	8.260	NA	3.332	4.082	7168.244	0.0014	0.0043	0.0239	0.0019	0.1938	0.0015

6. Simulation experiment

For a fixed $J > 4$, the simulation experiment has been carried out according to the following steps:

1. the CUBE models shown in Figure 1 have been chosen as a benchmark since they are sufficiently scattered over the parameter space;

2. for varying sample size $n = 250, 500, 1000, 3000$, `nsimul` $= 500$, 1000 random samples of ordinal values have been drawn by a CUBE distribution for each given parameter vector $\boldsymbol{\theta}' = (\pi, \xi, \phi)'$;

3. for each random sample, the EM algorithm has been run once for every choice of preliminary estimators described in Section 4, set as starting values, as well as for the *moment* estimators. For each method, the mean number over runs of iterations needed to reach convergence to the ML solution, the mean bias and the RMSE are computed.

Table 4 provides the mean number of iterations required for convergence to the ML solutions for each group of simulated samples. In these terms, the moment proposal yields consistently faster convergence. The mixing and the naive proposals are also rather satisfactory, especially for well-behaved distributions. It appears instead evident that the strong bias of the preliminary estimates obtained by ML short runs of the EM algorithm affects the performance of the convergence. It is often argued that the prevailing advantage of the ML estimation procedure over the method of moments is that the former yields the most efficient of the unbiased estimators, while the latter does not generally ensure efficiency ([38]). In this framework, the two approaches give pretty close results in terms of efficiency. Let us denote with $\boldsymbol{\theta}'_{GB} = (\pi_{GB}, \xi_{GB}, \phi_{GB})'$ the moment estimators derived with the Gröbner basis approach, and by the hat notation $\hat{\boldsymbol{\theta}}'_{GB} = (\hat{\pi}_{GB}, \hat{\xi}_{GB}, \hat{\phi}_{GB})'$ the ML estimates achieved after EM iterations started with moment estimates themselves. Table 4 reports (on the right) the relative efficiency of $\hat{\boldsymbol{\theta}}_{GB}$ with respect to $\boldsymbol{\theta}_{GB}$, showing that the distance to ML efficiency is rather negligible; thus, the moment proposal here developed can be considered as a safe, convenient and fairly easy technique to obtain good estimates of parameters for CUBE models.

Model	Method					
	Moment	Naive	Grid	Mixing	Random	Short run EM
1	**28.295**	42.527	41.459	33.477	60.450	46.150
2	**100.786**	146.774	143.636	137.527	201.828	207.876
3	**41.721**	73.790	74.276	67.334	83.673	78.366
4	**44.901**	77.769	59.059	73.888	86.396	70.449
5	**87.241**	148.552	116.082	131.940	154.333	122.769
6	**130.618**	157.840	171.894	137.008	178.026	144.307
7	**98.429**	175.767	246.098	153.507	207.339	187.791
8	**66.955**	143.152	108.633	133.155	145.844	128.808
9	**79.447**	157.265	148.768	152.075	157.757	153.634
10	**86.008**	105.814	96.783	97.110	147.198	160.346
11	59.960	68.839	95.157	**59.411**	90.314	64.206
12	**272.627**	306.812	317.305	292.527	365.308	392.952

Model	$Eff(\hat{\boldsymbol{\theta}}_{GB}\|\boldsymbol{\theta}_{GB})$		
	π	ξ	ϕ
1	1.2216	1.3061	1.2346
2	1.0563	1.0635	0.9669
3	1.0000	1.2100	1.1025
4	1.0494	1.1480	1.0615
5	1.0656	1.0727	1.1289
6	1.3194	1.0000	1.3157
7	1.0736	1.0000	1.0325
8	1.1121	1.0975	1.2016
9	1.1052	1.1480	1.3407
10	1.2016	1.4400	1.3121
11	1.2279	1.4102	1.7287
12	1.0436	1.0000	0.9570

Table 4: Mean number of iterations for ML convergence (left table). Relative efficiency (right table).

The impressive performances of the moment solution are motivated by the fact that the corresponding log-likelihood values are perfectly comparable with those achieved after the EM algorithm started with the other preliminary estimates. Table 5 summarizes the mean values over simulations attained by the estimated CUBE log-likelihood both at the different initial preliminary values and at the final ML estimates.

In addition, the convenience of the moment proposal is supported by the fact that the final ML estimates obtained by the different starting values for the EM algorithm are characterized by substantially the same final bias and root mean square errors, as detailed in the Tables 6 and 7.

Table 5: Log-Likelihood of preliminary estimates and final ML estimates

	Moment		Naive		Mixing		Short runs EM	
	LL Initial	LL Final	LL Initial	LL Final	LL Initial	LL Final	LL Initial	LL Final
1	−1253.879	−1253.694	−1295.036	−1253.694	−1256.227	−1253.694	−1294.720	−1253.694
2	−1918.672	−1918.546	−1923.923	−1918.546	−1919.737	−1918.546	−1922.358	−1918.546
3	−1743.660	−1743.627	−1776.688	−1743.627	−1752.653	−1743.627	−1758.427	−1743.627
4	−1728.412	−1728.327	−1810.780	−1728.327	−1739.301	−1728.327	−1740.071	−1728.327
5	−1839.518	−1839.184	−1880.065	−1839.184	−1842.538	−1839.184	−1844.603	−1839.184
6	−1862.259	−1862.031	−1871.020	−1862.031	−1866.613	−1862.031	−1863.893	−1862.031
7	−1905.543	−1905.488	−1912.128	−1905.489	−1907.066	−1905.489	−1908.562	−1905.488
8	−1736.376	−1736.018	−1798.710	−1736.018	−1746.277	−1736.018	−1746.272	−1736.018
9	−1588.194	−1587.446	−1699.769	−1587.446	−1616.688	−1587.446	−1618.111	−1587.446
10	−1904.188	−1903.890	−1910.412	−1903.890	−1904.870	−1903.890	−1910.393	−1903.890
11	−1642.941	−1642.285	−1656.147	−1642.285	−1644.414	−1642.285	−1652.495	−1642.285
12	−1930.139	−1929.681	−1934.469	−1929.680	−1930.878	−1929.679	−1930.970	−1929.682

Table 6: Bias of final ML estimators, given different starting values for EM

Model	Moment (π, ξ, ϕ)	Naive (π, ξ, ϕ)	Grid (π, ξ, ϕ)	Mixing (π, ξ, ϕ)	Random (π, ξ, ϕ)	Short run EM (π, ξ, ϕ)
1	(**0.002**, **0.000**, **0.001**)	(**0.002**, **0.000**, **0.001**)	(**0.002**, **0.000**, **0.001**)	(**0.002**, **0.000**, **0.001**)	(−0.019, **0.000**, **0.001**)	(**0.002**, **0.000**, **0.001**)
2	(**0.019**, −0.006, 0.014)	(**0.019**, **-0.005**, **0.013**)	(**0.019**, **-0.005**, **0.013**)	(**0.019**, −0.006, 0.014)	(**0.019**, −0.006, 0.014)	(0.020, −0.006, 0.014)
3	(**0.003**, **0.001**, **0.001**)	(**0.003**, **0.001**, **0.001**)	(**0.003**, **0.001**, **0.001**)	(**0.003**, **0.001**, **0.001**)	(**0.003**, **0.001**, **0.001**)	(**0.003**, **0.001**, **0.001**)
4	(0.004, **0.001**, **0.002**)	(0.004, **0.001**, **0.002**)	(0.004, **0.001**, **0.002**)	(0.004, **0.001**, **0.002**)	(**-0.002**, **0.001**, 0.003)	(0.004, **0.001**, **0.002**)
5	(0.011, **-0.004**, 0.008)	(0.011, **-0.004**, **0.007**)	(0.011, **-0.004**, 0.008)	(0.011, **-0.004**, **0.007**)	(**0.007**, −0.005, 0.008)	(0.011, **-0.004**, 0.008)
6	(0.015, **-0.001**, **0.005**)	(**0.014**, **-0.001**, **0.005**)	(**0.014**, **-0.001**, **0.005**)	(0.015, **-0.001**, **0.005**)	(0.015, **-0.001**, **0.005**)	(0.015, **-0.001**, **0.005**)
7	(0.017, **0.003**, **0.006**)	(**0.016**, **0.003**, **0.006**)	(0.017, **0.003**, **0.006**)	(0.016, **0.003**, **0.006**)	(**0.016**, **0.003**, **0.006**)	(0.017, **0.003**, **0.006**)
8	(0.008, **-0.002**, **0.002**)	(0.007, **-0.002**, **0.002**)	(0.007, **-0.002**, **0.002**)	(0.007, **-0.002**, **0.002**)	(**0.001**, **-0.002**, **0.002**)	(0.007, **-0.002**, **0.002**)
9	(0.004, **-0.001**, **0.001**)	(0.004, **-0.001**, **0.001**)	(0.004, **-0.001**, **0.001**)	(0.004,**-0.001**, **0.001**)	(**0.001**, **-0.001**, **0.001**)	(0.004, **-0.001**, **0.001**)
10	(**0.011**, **0.006**, **0.010**)	(**0.011**, **0.006**, **0.010**)	(**0.011**, **0.006**, **0.010**)	(**0.011**, **0.006**, **0.010**)	(0.009, 0.007, 0.011)	(0.011, **0.006**, 0.011)
11	(**0.005**, **0.002**, 0.007)	(**0.005**, **0.002**, **0.006**)	(**0.005**, **0.002**, 0.007)	(**0.005**, **0.002**, 0.007)	(**-0.005**, 0.004, 0.010)	(0.005, 0.002, 0.007)
12	(0.044, −0.003, 0.015)	(**0.034**, **-0.002**, **0.010**)	(0.037, −0.003, 0.012)	(0.036, **-0.002**, 0.011)	(0.057, −0.006, 0.024)	(0.050, −0.005, 0.020)

7. Discussion and conclusions

The original motivation for the present work was to emphasize that modelling overdispersion in ordinal data is a non-trivial task and deserves accurate analysis also at a preliminary stage. For CUB and CUBE models, this issue is tackled by developing a method of moments grounded on Gröbner basis theory: the methodological efforts spent in this direction are related mainly to the choice of the best ordering among parameters when treated as polynomial indeterminates. Then, once and for all computed the theoretical Gröbner basis in full generality, the technique is not consuming, in that only a step-wise procedure has to be implemented to directly solve the corresponding system. Globally, the moment proposal derived by the Gröbner basis approach outperforms other preliminary estimation solutions since it uniformly and substantially reduces the average number of required iterations for ML convergence and it is characterized by the lowest (initial) bias, Mahalanobis distance and Kullback-Leibler divergence. Most importantly, the log-likelihood at the moment estimates is strikingly close to the maximized value attained via ML procedure, and moment and final ML estimates of parameters are perfectly comparable in terms of efficiency.

The simulation experiment confirms the relevant role of both uncertainty and overdispersion parameters in determining the speed of convergence of the numerical routines and

Table 7: RMSE of final ML estimators, given different starting values for EM

Model	Moment (π, ξ, ϕ)	Naive (π, ξ, ϕ)	Grid (π, ξ, ϕ)	Mixing (π, ξ, ϕ)	Random (π, ξ, ϕ)	Short run EM (π, ξ, ϕ)
1	(**0.019**, **0.007**, **0.027**)	(**0.019**, **0.007**, **0.027**)	(**0.019**, **0.007**, **0.027**)	((**0.019**, **0.007**, **0.027**)	(0.138, 0.009, **0.027**)	(**0.019**, **0.007**, **0.027**)
2	(**0.072**, **0.032**, 0.060)	(**0.072**, **0.032**, **0.059**)	(**0.072**, **0.032**, **0.059**)	(**0.072**, **0.032**, 0.060)	(0.075, 0.033, 0.061)	(**0.072**, **0.032**, 0.060)
3	(**0.040**, **0.010**, **0.020**)	(**0.040**, **0.010**, **0.020**)	(**0.040**, **0.010**, **0.020**)	(**0.040**, **0.010**, **0.020**)	(**0.040**, **0.010**, **0.020**)	(**0.040**, **0.010**, **0.020**)
4	(**0.041**, **0.014**, **0.033**)	(**0.041**, **0.014**, **0.033**)	(**0.041**, **0.014**, **0.033**)	(**0.041**, **0.014**, **0.033**)	(0.078, 0.015, **0.033**)	(**0.041**, **0.014**, **0.033**)
5	(0.062, **0.028**, 0.064)	(0.062, **0.028**, **0.063**)	(**0.061**, **0.028**, **0.063**)	(**0.061**, **0.028**, **0.063**)	(0.078, 0.029, 0.064)	(**0.061**, **0.028**, **0.063**)
6	(0.074, **0.014**, **0.034**)	(**0.073**, **0.014**, **0.034**)	(**0.073**, **0.014**, **0.034**)	(**0.073**, **0.014**, **0.034**)	(0.074, **0.014**, **0.034**)	(**0.073**, **0.014**, **0.034**)
7	(0.083, **0.030**, **0.062**)	(**0.082**, **0.030**, **0.062**)	(0.084, **0.030**, **0.062**)	(**0.082**, **0.030**, **0.062**)	(0.083, **0.030**, **0.062**)	(**0.082**, **0.030**, **0.062**)
8	(**0.055**, **0.021**, **0.052**)	(**0.055**, **0.021**, **0.052**)	(**0.055**, **0.021**, **0.052**)	(**0.055**, **0.021**, **0.052**)	(0.086, 0.022, **0.052**)	(**0.055**, **0.021**, **0.052**)
9	(**0.039**, **0.014**, **0.038**)	(**0.039**, **0.014**, **0.038**)	(**0.039**, **0.014**, **0.038**)	(**0.039**, **0.014**, **0.038**)	(0.063, **0.014**, **0.038**)	(**0.039**, **0.014**, **0.038**)
10	(**0.052**, **0.030**, 0.055)	(**0.052**, **0.030**, **0.054**)	(**0.052**, **0.030**, **0.054**)	(**0.052**, **0.030**, **0.054**)	(0.057, 0.041, 0.057)	(**0.052**, **0.030**, 0.055)
11	(**0.037**, **0.016**, **0.054**)	(**0.037**, **0.016**, **0.054**)	(**0.037**, **0.016**, **0.054**)	(**0.037**, **0.016**, **0.054**)	(0.086, 0.019, 0.059)	((**0.037**, **0.016**, **0.054**)
12	(0.139, 0.036, 0.092)	(**0.122**, **0.035**, **0.086**)	(0.124, **0.035**, 0.087)	(0.124, **0.035**, 0.087)	(0.148, 0.037, 0.097)	(0.134, 0.036, 0.094)

the efficiency of ML estimation. As expected, random starting is comparatively the worst method, thus it should not be chosen as a criterion for initial values. Finally, the mixing and the naive proposals are generally intermediate but they require a substantially longer time to achieve convergence with respect to the moment proposal for most of the selected models. The dependence of the EM convergence on the initial estimates appears evident: with a low number of iterations, if the starting values are far enough from the true parameter vector, the resulting short run EM estimates are poor. Moreover, the presence of extreme feeling seriously undermines the ability of detecting the overdispersion effect. In this regard, a plausible solution to deal with borderline feeling parameter and almost negligible uncertainty component can be hinted: in this case, one could derive parameter estimates by fitting an Inverse Hypergeometric Model (for short, IHG) [11, 12], since it corresponds to a shifted Beta-Binomial model unaffected by uncertainty: indeed, IHG models are nested into CUBE ones [17].

In conclusion, we briefly shows the performances of the proposed method of moments on a real application. In December 2014, an observational survey was run at University of Naples Federico II to investigate relational goods, perceived happiness and leisure time habits[§]. Questionnaires were collected according to a snowball sampling scheme starting from students of the Department of Political Sciences. All the ratings were measured on a $J = 10$ point ordinal scale (ranging from 1 = "Never", "Not at all good", to 10 = "Always", "A lot", "Absolutely good"). Here we shall focus on a variable measuring the quality of time spent with relatives ($n = 2458$), for which both the classical ML estimation procedure and the Gröbner basis approach to moment estimation have been tested. CUBE models allows an impressive fit for this variable: indeed, ML estimates for CUB parameters are $(\hat{\pi}, \hat{\xi})' = (0.302, 0.364)'$, with a maximized log-likelihood of $\ell(\hat{\boldsymbol{\theta}}) = -5564.84$ (the saturated log-likelihood equals $\ell = -5499.13$), whereas the maximized log-likelihood for a CUBE fit attains the value $\ell(\hat{\boldsymbol{\theta}}) = -5530.60$.

In order to enhance comparisons, the (normalized) Dissimilarity index between the observed (relative) frequency distribution and the estimated CUBE probability distribution

[§]The data set is freely available at http://www.labstat.it/home/wp-content/uploads/2015/09/relgoods.txt and it is further bundled within the R package CUB.

has been computed in addition to the Kullback-Leibler divergence. This index is a fitting measure that allows effective interpretation in ordinal data analysis as it corresponds to the proportion of respondents whose answers should be modified to achieve a perfect fit. For a (discrete) probability model $(p_1(\boldsymbol{\theta}), \ldots, p_J(\boldsymbol{\theta}))$ over $\{1, \ldots, J\}$ and given a random sample with relative frequency distribution $(f_1, \ldots, f_J)$, the dissimilarity index evaluates the goodness of fit of an estimated probability model $(p_1(\hat{\boldsymbol{\theta}}), \ldots, p_J(\hat{\boldsymbol{\theta}}))$:

$$Diss = \frac{1}{2} \sum_{r=1}^{J} |f_r - p_r(\hat{\boldsymbol{\theta}})|. \tag{43}$$

Results for the application here discussed are reported in Table 8. Thus, empirical evidence supports the efficacy of the proposal in that the model estimated via the method of moments here implemented is really close to the one estimated via ML procedure. In this regard, we underline that the Kullback-Leibler divergence between the two estimated probability distributions attains a very low value $KL = 0.0007$, and that, in order to reach convergence with the ML method with a tolerance equal to $1e-6$, 457 `iterations` are needed), whereas the moment estimation is straightforward. The Delta method can be used to estimate the variance of moment estimates.

	$\hat{\pi}$	$\hat{\xi}$	$\hat{\phi}$	$\ell(\hat{\boldsymbol{\theta}})$	Dissimilarity	K-L Divergence
ML	0.713	0.388	0.244	-5530.60	0.059	0.0131
Moment	0.913	0.413	0.298	-5532.67	0.066	0.0139

Table 8: Comparisons on real data between the Method of Moment and the ML estimation

Acknowledgements

The Authors are grateful to Editors and Reviewers for their comments and remarks that substantially contributed to improve the paper. This work has been partially supported by the 'CUB Regression Models Trees' project (code RBFR12SHVV) at University of Naples Federico II.

Appendix

Gröbner bases

Gröbner bases are a widely used mathematical tool within algebraic statistics, which have mostly applications in experimental design. The following presentation aims at providing the necessary insights for the understanding of the proposed approach to the method of moments for CUBE models. For any unspecified detail and further discussion about Gröbner bases, see [9, 37] and references therein.

If $\mathbb{K}$ is an algebraic closed field, let $\mathbb{K}[x_1, \ldots, x_t]$ denote the ring of multivariate polynomials with coefficients in $\mathbb{K}$ in the independent indeterminates $x_1, \ldots, x_t$, provided with

an *initial order*, say $x_1 \succ \cdots \succ x_t$. The second step in this context is the choice of a *monomial order*: indeed, while in the univariate setting the degree of a polynomial is uniquely determined as the highest order exponent of the variable, in the multivariate case several choices are available, and therefore an ordering has to be set for developing a proper extension of the univariate setting. In more precise terms, a *monomial order* is an order relation on the set of monomials in the fixed indeterminates that allows one to compare them, according to their multidegree and the fixed initial order. Specifically, an order that is compatible with cancellation between monomials is required.

Throughout the analysis here presented, a lexicographic order on monomials has been chosen: for $\alpha_1, \ldots, \alpha_t, \beta_1, \ldots, \beta_t$ non-negative integers, $x_1^{\alpha_1} \cdots x_t^{\alpha_t} \succ x_1^{\beta_1} \cdots x_t^{\beta_t}$ if and only if, for $p = \min\big(i : \alpha_i \neq \beta_i\big)$, $\alpha_p \succ \beta_p$. The lexicographic order is a particular instance of what is called an *elimination order*, needed to apply Gröbner basis within elimination theory and solving systems of nonlinear algebraic equations. For a fixed monomial order on $\mathbb{K}[x_1, \ldots, x_t]$, the first fundamental concept to introduce are the definitions of *leading term* $\mathcal{LT}(f)$ and *leading coefficient* $\mathcal{LC}(f)$ of a polynomial $q(x_1, \ldots, x_t)$, namely, the highest-degree monomial with respect to the fixed ordering and its coefficient, respectively. Given the dependence of these concepts on the monomial and on the initial order, division between polynomials does not produce uniquely determined quotient and remainder: more precisely, given a polynomial f and polynomials $h_1, \ldots, h_p$, it is always possible to find polynomials $q_1, \ldots, q_p, r$, such that:

$$f = \sum_{i=1}^{p} q_i h_i + r, \tag{44}$$

with r being the remainder, and such that there is no index i with $\mathcal{LT}(h_i)$ dividing $\mathcal{LT}(r)$. For a generic set of polynomials $h_1, \ldots, h_p$, the remainder is not unique. In order to develop a unique Euclid's division procedure in the multivariable setting, Gröbner basis needs to be introduced, which requires, in turn, the notion of polynomial ideal. A subset $\mathcal{I}$ of $\mathbb{K}[x_1, \ldots, x_t]$ is called an *ideal* if the following properties are satisfied:

- $\mathcal{I}$ is a subring, that is $0 \in \mathcal{I}$ (where 0 denotes the identity element with respect to summation of polynomials), and $p_1 + p_2 \in \mathcal{I}$ $\forall p_1, p_2 \in \mathcal{I}$;
- $\forall p \in \mathcal{I}$, $\forall q \in \mathbb{K}[x_1, \ldots, x_t]$, $p\,q \in \mathcal{I}$.

According to the Hilbert's Basis Theorem, every ideal $\mathcal{I}$ of $\mathbb{K}[x_1, \ldots, x_t]$ is finitely generated, namely, there exists a finite number of polynomials $p_1, \ldots, p_h \in \mathcal{I}$ such that every other polynomial $p \in \mathcal{I}$ can be written as:

$$p = \sum_{i=1}^{h} a_i\, p_i, \quad a_i \in \mathbb{K}[x_1, \ldots, x_t]. \tag{45}$$

In this case, $p_1, \ldots, p_h$ are called generators of $\mathcal{I}$, and this circumstance is denoted as $\mathcal{I} =< p_1, \ldots, p_h >$.

A Gröbner basis $\mathcal{G}$ for an ideal $\mathcal{I}$ is a set of generators $\mathcal{G} = \{g_1, \ldots, g_k\}$ of $\mathcal{I}$ such that, according to the given order, the leading term of any polynomial in $\mathcal{I}$ can be divided uniquely in the set of the leading terms of $\mathcal{G}$, producing no remainder, for short:

$$< \mathcal{LT}(\mathcal{I}) >=< \mathcal{LT}(g_1), \ldots, \mathcal{LT}(g_k) > . \tag{46}$$

A Gröbner basis $\mathcal{G}$ is said *minimal* if the following conditions apply:

- $\mathcal{LC}(g) = 1$ for every $g \in \mathcal{G}$,
- $\forall g \in \mathcal{G}, \mathcal{LT}(g) \notin < \mathcal{LT}(\mathcal{G} \setminus g) >$,

while it is called *reduced* if:

- $\mathcal{LC}(g) = 1$ for every $g \in \mathcal{G}$;
- $\forall g \in \mathcal{G}$, there is no monomial of g belonging to $< \mathcal{LT}(\mathcal{G} \setminus g) >$.

Every non-trivial ideal of $\mathbb{K}[x_1, \ldots, x_t]$ admits a unique reduced Gröbner basis for the given monomial order, and reduced Gröbner basis solves the *ideal membership problem*. As far as the object of the present paper is concerned, the most interesting application of Gröbner basis fits in the so-called *elimination theory*, that is, the solutions of a system of K algebraic equations:

$$F_i(x_1, \ldots, x_t) = 0, \quad i = 1, \ldots, K \tag{47}$$

can be equivalently computed by considering the system of algebraic equations:

$$G_j(x_1, \ldots, x_t) = 0, \quad j = 1, \ldots, H, \tag{48}$$

where $\mathcal{G} = \{G_1, \ldots, G_H\}$ is the reduced Gröbner basis for the ideal generated by $F_1, \ldots, F_K$. The choice of an elimination ordering guarantees that the latter system can be solved by forward/backward substitution, simplifying hence the solving procedure. In particular, when $\mathcal{G}$ is trivial, that is $\mathcal{G} = \{1\}$ (with 1 denoting the identity element with respect to product of polynomials), then $F_1, \ldots, F_K$ have no root in common.
Although there is no explicit formula for computing the number of polynomials composing the Gröbner basis (for the given monomial ordering), Buchberger algorithm provides the steps required to compute the reduced Gröbner basis in polynomial time [7].

Gröbner basis for CUB models

The Maple library *Gröbner* includes a routine that suggests the variable ordering that yields to the simplest possible equivalent system. For a CUB (π, ξ) model over J ordinal categories, the suggested ordering is $\pi > \xi$: thus, for a given sample of observations, the (reduced) Gröbner basis corresponding to the first two moment equations – obtained by selecting the Buchberger-Möller algorithm as option for computation – is composed of:

$$G_1(\xi) = A(J, m_1, m_2)\,\xi^2 + B(J, m_1, m_2)\,\xi + C(J, m_1, m_2) \tag{49}$$

$$G_2(\pi, \xi) = D(J, m_1, m_2)\,\pi + E(J, m_1, m_2)\,\xi + F(J, m_1, m_2) \tag{50}$$

where, if m_1, m_2 denote the first two sample moments, we have set:

1. $A(J, m_1, m_2) = -3(J^2 - J\,(1 + 2\,m_1) - 2(1 - 2\,m_1))$;
2. $B(J, m_1, m_2) = 4\,J^2 - 12\,m_1\,J + 6\,m_2 - 4$;
3. $C(J, m_1, m_2) = -J^2 + J\,(4\,m_1 - 1) + m_1 - 3\,m_2$;
4. $D(J, m_1, m_2) = J^2 - 3\,J + 2$ ¶;
5. $E(J, m_1, m_2) = -6(J^2 - J\,(1 + 2\,m_1) - 2(1 - 2\,m_1))$;
6. $F(J, m_1, m_2) = 5J^2 + 3J\,(1 - 6\,m_1) + 2(6\,m_2 - 1)$.

From (49)-(50) it is immediate to study and derive the moment solutions. From empirical evidence, it turns out that the only admissible moment solution for ξ corresponds to:

$$\xi_{\text{cub}} = \frac{-B(J, m_1, m_2) - \sqrt{B(J, m_1, m_2)^2 - A(J, m_1, m_2)C(J, m_1, m_2)}}{2\,A(J, m_1, m_2)}, \text{ if } p_2 \neq 0.$$

Notice that when $m_1 = \frac{J+1}{2}$, $A(J, \frac{J+1}{2}, m_2) = 0$ for all values of J and m_2: this situation occurs whenever we sample from a perfectly symmetric CUB distribution (having true value of ξ equal to 0.5). In this case, (49) collapses to a one-degree equation and, accordingly, the resulting ξ estimate is given by:

$$\xi_{\text{cub}} = -\frac{C(J, m_1, m_2)}{B(J, m_1, m_2)} = 0.5 \tag{51}$$

for all values of J and m_2.
Notice also that, when $m_1 = \frac{J+1}{2}$, the identifiability of π from (50) is always ensured since $F(J, m_1, m_2) \neq 0$.

Gröbner basis for CUBE models

For CUBE models, the suggested ordering would be the lexicographical order $\pi > \phi > \xi$, for which the Gröbner basis corresponding to (17) consists of 3 polynomials of the type:

1. $H_1(\xi) = \xi(\xi - 1)(k_0 + k_1\xi + k_2\xi^2)$
2. $H_2(\xi, \phi) = j_0 + j_1\xi + j_2\xi^2 + l_1\phi$
3. $H_3(\xi, \phi) = s_0 + s_1\xi + s_2\xi^2 + l_2\pi$,

whose coefficients are functions of J and of the first three sample moments. Thus, given the solutions to $k_0 + k_1\xi + k_2\xi^2 = 0$, the corresponding solutions for ϕ and π would be easily obtained. However, an imputation choice should be thought in case there is no (admissible) solution to $H_1(\xi) = 0$ other than the trivial ones $\xi = 0$ or $\xi = 1$. In the setting here developed, based on the method of moments, the most natural strategy would

¶(which is always non-zero under our assumption $J > 3$)

be to impute the moment estimate for ξ of the nested CUB model and then solve for π and ϕ the remaining equations by conditioning on the estimated ξ value. Nevertheless, Gröbner basis theory involves symbolic variables, whereas the approach here advanced has to consider parameters constraints and their mutual interactions. Then, the choice of the ordering $\pi > \xi > \phi$ assumed for our analysis is better-suited for CUBE models estimation since it allows to first solve for ϕ, which is the main feature of the model, and then, if it is not possible to determine its moment estimate within the given boundaries, one could exploit the nesting of CUB into CUBE to estimate π and ξ.

References

[1] Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, 2^{nd} edition. J.Wiley & Sons, Hoboken.

[2] Aitkin, M. (1996), A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, **6**(3), 251–262.

[3] Améndola, C., Faugère, J.C. and Sturmfels, B. (2016), Moment Varieties of Gaussian Mixtures, *Journal of Algebraic Statistics*, **7**(1), 14–28.

[4] Baudry, J.P. and Celeux, G. (2015), EM for mixtures - Initialization requires special care. *Statistics and Computing*, **25**(4), 713–726.

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

[6] Borisevich, V.D., Potemkin, V.G. and Strunkov, S.P. (2000), Global methods for solving systems of nonlinear algebraic equations. *Computer and mathematics with applications*, **40**(8-9), 1015–1025.

[7] Buchberger, B. (1976), *A theoretical basis for the reduction of polynomials to canonical forms*. In ACM SIGSAM Bulletin 39, pp. 19-29.

[8] Cox, D.R. (1983), Some remarks on overdispersion. *Biometrika*, **70**, 269–274.

[9] Cox, D., Little, J., O'Shea, D. (1997), *Ideals, Varieties, and Algorithms*. Springer Verlag, New York, Second Edition.

[10] Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton Mathematical Series, Princeton University Press-Princeton.

[11] D'Elia, A. (2000), A shifted binomial model for rankings, *Proceedings of the 15th International Workshop on Statistical Modelling. New Trends in Statistical Modelling (IWSM)* - Bilbao, Spain, 412–416.

[12] D'Elia, A. (2003), Modelling ranks using the Inverse Hypergeometric distribution. *Statistical Modelling: an International Journal*, **3**(1), 65–78.

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

[14] de Finetti, B. and Paciello, U. (1930), Calcolo della differenza media, *Metron*, **8**(3), 89–94.

[15] Fitzmaurice, G.M., Heath, A.F. and Cox, D.R. (1997), Detecting overdispersion in large scale surveys: Application to a study of education and social class in Britain. *Applied Statistics*, **46**(4), 415–432.

[16] Furman, W.D. and Lindsay, B.G. (1994), Measuring the relative effectiveness of moment estimators as starting values in maximising likelihoods, *Computational Statistics & Data Analysis*, **17**(5), 493–507.

[17] Iannario, M. (2012a), CUBE models for interpreting ordered categorical data with overdispersion, *Quaderni di Statistica*, **14**(1), 137–140.

[18] Iannario, M. (2012b), Preliminary estimators for a mixture model of ordinal data. *Advances in Data Analysis and Classification*, **6**(3), 163–184.

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

[20] Iannario, M. (2014), Modelling uncertainty and overdispersion in ordinal data. *Communications in Statistics. Theory and Methods*, **43**(4), 771–786.

[21] Iannario, M. (2015), Detecting latent components in ordinal data with overdispersion by means of a mixture distribution. *Quality & Quantity*, **49**(3), 977–987.

[22] Iannario, M. (2016), Testing the overdispersion parameter in CUBE models. *Communications in Statistics: Simulation and Computation*, **45**(5), 1621–1635.

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

[24] Iannario, M., Piccolo, D. and Simone, R. (2016), *CUB: A Class of Mixture Models for Ordinal Data* (`R` package version 1.0), http://CRAN.R-project.org/package=CUB.

[25] Karlis, D. and Xekalaki, E. (1999), Improving the EM algorithm for mixtures. *Statistics and Computing*, **9**(4), 303–307.

[26] Karlis, D. and Xekalaki, E. (2003), Choosing initial values for the EM algorithm for finite mixtures, *Computational Statistics & Data Analysis*, **41**(3-4), 577–590.

[27] Laird, N. (1978), Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of the American Statistical Association*, **73**(364), 805–811.

[28] Ligget, R.E. and Delwiche, J.F. (2005), The Beta-Binomial model: variability in overdispersion across methods and over time, *Journal of Sensory Studies*, **20**(1), 48–61.

[29] Mahalanobis, P.C. (1936), On the generalised distance in statistics, *Proceedings of the National Institute of Sciences of India*, **2**(1), 49–55.

[30] McCullagh, P. (1980), Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, **42**(2), 109–142.

[31] McLachlan, G. and Krishnan, T. (1997), *The EM algorithm and extensions*. New York: J.Wiley & Sons.

[32] McLachlan, G. and Peel, G.J. (2000), *Finite mixture models*. New York: J.Wiley & Sons.

[33] Piccolo, D. (2003), On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, **5**(1), 85–104.

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

[35] Piccolo, D. (2015), Inferential issues on CUBE models with covariates. *Communications in Statistics. Theory and Methods*, **44**(23), 771–786.

[36] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2007), *Numerical Recipes: The Art of Scientific Computing* (3rd ed.), New York: Cambridge University Press.

[37] Pistone, G., Riccomagno, E. and Wynn, H.P. (2000), *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman & Hall/CRC, Monographs on Statistics and Applied Probability.

[38] Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, 2nd edition. New York: J.Wiley & Sons.

[39] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, J.Wiley & Sons.

[40] Shacham, M., Brauner, N. and Pozin, M. (1998), Comparing software for interactive solution of systems of nonlinear algebraic equations, *Computers and Chemical Engineering*, **22**(1-2), 323–331.

[41] Skellam, J.G. (1948), A probability distribution derived from the Binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society, Series B*, **10**(2), 257–261.

[42] Tamhane, A., Ankemanman, B. and Yang, Y. (2002), The Beta distribution as a latent response model for ordinal data (I): Estimation of location and dispersion parameters. *Journal of Statistical Computation and Simulation*, **72**(6), 473–494.

[43] Tourangeau, R., Rips, L. J., Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

[44] Tripathi, R.C., Gupta, R.C. and Gurland, J. (1994), Estimation of parameters in Beta Binomial models. *Annals of the Institute of Statistical Mathematics*, **46**(2), 317–331.

[45] Wilcox, R.R. (1979), Estimating the parameters of the Beta-binomial distribution, *Educational and Psychological Measurement*, **39**(3), 527–535.

[46] Yamamoto, E., Yanagimoto, T. (1992), Moment estimators for the beta-binomial distribution, *Journal of Applied Statistics*, **19**(2), 273–283.