On finite mixtures of Discretized Beta model for ordered responses

Simone, Rosaria

ORIGINAL PAPER

On finite mixtures of Discretized Beta model for ordered responses

Rosaria Simone[1]

Abstract

The paper discusses the specification of finite mixture models based on the Discretized Beta distribution for the analysis of ordered discrete responses, as ratings and count data. The ultimate goal of the paper is to parameterize clusters of opposite and intermediate response outcomes. After a thorough discussion on model interpretation, identifiability and estimation, the proposal is illustrated on the wake of a case study on the probability to vote for German Political Parties and with a comparative discussion with the state of the art.

Keywords Ordered data · Finite mixture models · Discretized Beta distribution

Mathematics Subject Classification 62F99 · 62P25 · 62915 · 62E99

1 Motivation

With the upcoming of any electoral competition, parties' share of the electoral consensus can be measured by pollsters if voting intentions on nominal scales are surveyed. A more innovative approach consists in gauging probability to vote for each candidate as ratings on ordered scales in order to assess the extent by which respondents' opinions hold. Similarly, marketing stakeholders prefer to survey intention to take a certain decision in the future, rather than asking questions with yes/no answers about respondents' likings and habits. Thus, suitable statistical modelling of ordered evaluations is advocated to characterize clusters of both extreme and intermediate response choices.

Polarization is hereafter meant as the process by which evaluations about an item converge towards one of two opposing poles of the response spectrum, in the spirit of

✉ Rosaria Simone
   rosaria.simone@unina.it

[1] University of Naples Federico II: Universita degli Studi di Napoli Federico II, Naples, Italy

(Apouey [2007])[1]. Possibly, a further cluster may be expected as a result of un-polarized respondents, corresponding to a concentration of responses away from the extremes: the term *floatation* is hereafter used to indicate this circumstance as complementary to polarization.

A candidate model allowing to directly parameterize polarization towards the extremes is the two-component mixture of Inverse Hypergeometric distributions (MIHG, (Simone and Iannario [2018])), whereas a mixture of Binomial and Discretized Beta models can be considered to analyse overall response feeling and certain symmetric response styles (CAUB, (Simone and Tutz [2018])). For count data, bimodality (not necessarily at the extremes of the response support) can be tackled via suitable adaptation of the (shifted) Poisson distribution (Gómez-Déniz et al. [2020]) or by resorting to a two-component mixture of Conway–Maxwell–Poisson models (Sur et al. [2015]).

With respect to the state of the art, the paper discusses the specification of mixture models based on the Discretized Beta distribution (Ursino [2014]; Ursino and Gasparini [2018]) as a flexible class of statistical models to parameterize polarization and floatation of ordered evaluations. The proposal is designed to attain broad and straightforward interpretation for marketing, psychology and socio-economic studies, as it allows to characterize opposite and intermediate response clusters. Further relevant applications include self-reported wealth or health, or Net Promoter Score type evaluations (NPS, (Reichheld [2003])) to assess the extent by which *attractors* outclass *detractors* (Capecchi and Piccolo [2017]).

The paper is organized as follows: Sect. 2 recalls the baseline framework of the Discretized Beta model. The core of the paper is Sect. 3, with a detailed discussion on mixtures based on the Discretized Beta distribution to jointly model polarization and floatation of ordered evaluations; goodness-of-fit criteria and inferential aspects are described in Sects. 3.2–3.5, whereas a comparative discussion of the state of the art is delivered in Sect. 3.6. A case study is pursued in Sect. 4 to support the proposal with empirical evidence. Concluding remarks are addressed in Sect. 5. A devoted appendix supplements the presentation with a discussion on the optimal number of components for Discretized Beta mixtures and of the parameter constraints needed to prevent identifiability issues.

2 Discretized Beta mixtures for polarization and floatation of ordered data

Let R be a rating variable collected on a response scale with m ordered categories, say $c_1 \prec c_2 \prec \cdots \prec c_m$: the numeric scoring $c_r = r$ will be made merely for notational convenience. Without loss of generality, assume that the scale has a positive orientation with the trait being examined.

Definition 1 For $\alpha, \beta \in \mathbb{R}^+$, let $X \sim Beta(\alpha, \beta)$ be a Beta distributed random variable over the real interval [0, 1]. For a given $m > 3$, a discrete variable R, with

---

[1] It is worth to emphasize that the term *polarization* for ordinal data analysis is not new: in general, it is meant as the extent by which the response distributions of a-priori groups (typically, identified by covariates) are concentrated around certain locations of the response support. A nonparametric approach to assess polarization in this sense is presented in (Mussini [2018]), where a detailed summary of different concepts and measures of polarization for ordinal evaluations is provided.

support $\{1, 2, \ldots, m\}$, is said to be distributed according to a Discretized Beta model of parameters α, β ($R \sim \mathrm{DB}(\alpha, \beta)$, for short) if:

$$Pr(R = r | \alpha, \beta) \, = \, Pr\left(\frac{r-1}{m} \leq X \leq \frac{r}{m} \Big| \alpha, \beta\right), \quad r = 1, \ldots, m. \quad (1)$$

For notational convenience, set $db(r; \alpha, \beta) := Pr(R = r | \alpha, \beta)$. This model has been already acknowledged in the literature on ordinal data analysis in view of the flexibility inherited from the underlying Beta distribution, which does not impose a predetermined shape for the latent continuous trait (Ursino 2014; Fasola and Sciandra 2015; Ursino and Gasparini 2018; Simone and Tutz 2018). Similar arguments can be advanced for the Beta-Binomial model (Morrison 1979), yet the Discretized Beta is more versatile as it can be either overdispersed or underdispersed (Ursino 2014). The uniform distribution arises as a limit case when $\alpha = \beta = 1$. Location and shape properties of the latent Beta model imply the following features of the DB distribution (Abramowitz and Stegun 1972; Forbes et al. 2011). Given that the discretization of the latent Beta model occurs at equi-spaced intervals for a fixed m, the modal value $Mo(R)$ of $R \sim \mathrm{DB}(\alpha, \beta)$ satisfies:

– $Mo(R) = 1$ if $\alpha < 1$ and $\beta \geq 1$ or, in case $\min(\alpha, \beta) > 1$, if $\frac{\alpha-1}{\alpha+\beta-2} < \frac{1}{m}$;
– $Mo(R) = m$ if $\alpha \geq 1$ and $\beta < 1$ or, in case $\min(\alpha, \beta) > 1$, if $\frac{\alpha-1}{\alpha+\beta-2} > 1 - \frac{1}{m}$;
– $Mo(R) = r \in \{2, \ldots, m-1\}$ if and only if $\min(\alpha, \beta) > 1$ and $\frac{\alpha-1}{\alpha+\beta-2} \in (\frac{r-1}{m}, \frac{r}{m}]$.
  Thus, the following condition implies an inner mode:

$$\frac{1}{m} < \frac{\alpha-1}{\alpha+\beta-2} < 1 - \frac{1}{m}; \quad (2)$$

– The distribution is U-shaped with two modal values at the first and at the last categories if $\max(\alpha, \beta) < 1$ and if, for the given m, parameters satisfy the following system of inequalities[2] based on the incomplete Beta function $I_x(\alpha, \beta)$:

$$\begin{cases} 1 + I_{\frac{m-2}{m}}(\alpha, \beta) & > \ 2\, I_{\frac{m-1}{m}}(\alpha, \beta); \\ 2\, I_{\frac{1}{m}}(\alpha, \beta) & > \ I_{\frac{2}{m}}(\alpha, \beta). \end{cases} \quad (3)$$

As a consequence, a necessary condition for a Discretized Beta model to be applied for polarization of either favourable or unfavourable responses is the constraint $\min(\alpha, \beta) < 1$. Under this circumstance, parameter α governs the polarization of the unfavourable responses: hereafter, this cluster will be referred to as *opponents' pole*. If $\beta = \max(\alpha, \beta) \geq 1$, the closer α is to 0, the stronger is the polarization of the opponents, with positive asymmetry increasing with growing β. Conversely, β governs the polarization of the favourable responses (say, the *supporters' pole*). If $\alpha = \max(\alpha, \beta) \geq 1$, the closer β is to 0, the higher is the probability assigned to the last category and thus the stronger is the polarization of the supporters, with

---

[2] Notice that these conditions, depending on m, are eventually needed only for values of $\max(\alpha, \beta)$ very close to 1 from below.

negative asymmetry strengthening with growing α. A Discretized Beta model with $\max(\alpha, \beta) < 1$, instead, can be specified to account for polarization towards both the extremes (provided that (3) holds), whereas floatation between the two response endpoints can be modelled by assuming a DB(α, β) distribution with $\min(\alpha, \beta) > 1$, such that (2) holds true, given the number of categories[3]. Asymmetry and intensity of floatation can be measured in terms of skewness $\gamma_1(\alpha, \beta)$ and excess kurtosis $\gamma_2(\alpha, \beta)$ of the underlying Beta distribution[4]:

$$\gamma_1(\alpha, \beta) = 2 \frac{\beta - \alpha}{\alpha + \beta + 2} \sqrt{\frac{\alpha + \beta + 1}{\alpha \beta}}; \qquad (4)$$

$$\gamma_2(\alpha, \beta) = \frac{6\big(\alpha^3 + \alpha^2(1 - 2\beta) + \beta^2(1 + \beta) - 2\alpha\beta(2 + \beta)\big)}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}; \qquad (5)$$

such that $\gamma_1(\alpha, \beta) = -\gamma_1(\beta, \alpha)$ and $\gamma_2(\alpha, \beta) = \gamma_2(\beta, \alpha)$. However, interpretation of excess kurtosis is not straightforward for asymmetric distributions: the measure of kurtosis adjusted for skewness introduced in (Blest 2003) can be considered to overcome this issue (see (15) in Appendix 1 for details).

3 Finite mixtures of Discretized Beta model

Given the flexibility in both shape and interpretation of the DB model, polarization and floatation in ordered data can be jointly parameterized by specifying suitable mixture distributions.

By virtue of the comments delivered in Sect. 2, if floatation can be shaped via a DB model with parameters $\alpha_2, \beta_2 > 1$ satisfying (2) for given m, alternative specifications are possible for the polarization effect:

1. a unique component DB(α_1, β_1) with $\min(\alpha_1, \beta_1) < 1$, and with $\max(\alpha_1, \beta_1) < 1$ satisfying (3) if two opposing clusters at the extremes are present, or with $\max(\alpha_1, \beta_1) \geq 1$ if only one pole of supporters or opponents is found, yielding the two-component mixture:

$$Pr\big(R = r \mid \boldsymbol{\theta}\big) = (1 - \delta)\, db(r; \alpha_1, \beta_1) + \delta\, db(r; \alpha_2, \beta_2), \quad r = 1, \ldots, m\,; (6)$$

2. a mixture of two J-shaped DB models, DB(α_1, β_1) and DB(α_3, β_3), yielding the 3-component mixture specification:

$$Pr\big(R = r \mid \boldsymbol{\theta}\big) = \delta_1\, db(r; \alpha_1, \beta_1) + \delta_2\, db(r; \alpha_2, \beta_2) \\ + \delta_3\, db(r; \alpha_3, \beta_3), \qquad (7)$$

---

[3] The conditions $\frac{\alpha - 1}{\alpha + \beta - 2} < \frac{1}{m}$ and $\frac{\alpha - 1}{\alpha + \beta - 2} > 1 - \frac{1}{m}$ needed to achieve $Mo(R) = 1$ or $Mo(R) = m$, respectively, if $\min(\alpha, \beta) > 1$, will not be considered for the sake of simplicity.

[4] For a symmetric distribution ($\alpha = \beta$), the excess kurtosis $\gamma_2(\alpha, \beta) = -\frac{6}{2\alpha + 3}$ tends to 0 as α grows to infinity, and the distribution becomes degenerate with mass concentrated at the middle of the support. Notice that a symmetric DB distribution has two modal values at the central categories if m is even.

so that $\delta_1 + \delta_2 + \delta_3 = 1$, and:

- $\alpha_1 \in (0, 1)$ and $\beta_1 \geq 1$ to shape the opponents' pole;
- $\alpha_2, \beta_2 > 1$ satisfy (2) to shape floatation;
- $\alpha_3 \geq 1$, $\beta_3 \in (0, 1)$ to shape the supporters' pole.

Some identifiability issues may arise for the polarization components in both (6) and (7), due to a Beta approximation of the latent Beta models. Appendix 2 collects all the relevant discussion and results pertaining to these topics: the present section will focus on the proposed class of mixtures, stemming from (7) under suitable parameter constraints.

3.1 The OFS mixture for polarization and floatation of ordered evaluations

In order to overcome possible identifiability issues for mixtures of DB models, the proposed strategy is to constrain $\beta_1 = 1$ and $\alpha_3 = 1$ for the mixture specification (7).

Hereafter, the acronym OFS will stand for Opponent-Floatation-Supporter, and three 0-1 subscripts will indicate if each component is specified in the mixture (1) or not (0). Thus, models $DB(\alpha_1, 1)$, with $\alpha_1 \in (0, 1)$, and $DB(1, \beta_3)$, with $\beta_3 \in (0, 1)$, will be referred to as OFS_{100} and OFS_{001} to indicate a DB distribution to model polarization towards the opponents' and the supporters' pole, respectively. Consequently, as a benchmark for bi-polarization towards the end-points, the proposal is to assume the following mixture specification.

Definition 2 If $\alpha_1, \beta_3, \delta \in (0, 1)$, the OFS_{101} model is defined by the mixture:

$$Pr(R = r|\boldsymbol{\theta}) = \delta\, db(r; \alpha_1, 1) + (1 - \delta)\, db(r; 1, \beta_3), \qquad r = 1, \ldots, m. \quad (8)$$

The mixture of OFS_{101} for polarization with an OFS_{010} distribution for floatation (so that (2) holds) can be safely considered to jointly model polarization towards either one or both the extremes and possible floatation in between.

Definition 3 If the above notation prevails, the OFS_{111} model is defined by:

$$Pr(R = r|\boldsymbol{\theta}) = \delta_1\, db(r; \alpha_1, 1) + \delta_2\, db(r; \alpha_2, \beta_2) + \delta_3\, db(r; 1, \beta_3). \quad (9)$$

Remark 1 With reference to the procedures outlined in Appendix 2 and unlike for (6) and (7), the Beta approximation of the latent polarization components in (9), and its combination with the latent floatation, does not correspond to an OFS_{111} specification. The same arguments apply if either OFS_{100} or OFS_{001} are assumed for polarization. Thus, identifiability of parameters can be assumed for OFS mixture models.

Both asymmetric and symmetric floatation are encompassed by the OFS_{111} model (under the constraint $\alpha_2 = \beta_2$). In case the floatation component is symmetric, the superscript (s) will be used. If m is odd, a degenerate floatation component corresponds to neutrality (in case $\alpha_2 = \beta_2$ tends to infinity), resulting in inflation in the middle of the response scale: in this case, the superscript (i) will replace (s), and the resulting $OFS_{111}^{(i)}$ model will denote a mixture of an OFS_{101} model with a degenerate distribution

$\mathbb{1}_{c=r}$ with mass concentrated at $c = \frac{m+1}{2}$ (so that $\mathbb{1}_{c=r} = 0$ if $r \neq c$, and $\mathbb{1}_{c=r} = 1$ if $r = c$).

Remark 2 OFS models encompass also inflated responses at the extremes of the response support. Consider, for instance, the OFS_{110} model: the $\text{DB}(\alpha_1, 1)$ component identifies the opponents' cluster, which is characterized by a mode at the first category and decreasing probabilities as scores increase, thus allowing to account also for scale usage diversity among opponents and for different strengths of opposition. As a limit case, the OFS_{110} tends to an inflated DB model with inflation at the first category if $\alpha_1 \to 0$. The dual remark applies for the OFS_{011} model[5]. Thus, the smoothed switch between extreme modal values and inner categories implied by the OFS approach is more general than DB models with inflation at either one of the end-points (see the example discussed in Sect. 3.6).

Remark 3 Covariate effects on model parameters can be investigated via suitable link functions. If x_i, y_i, u_i, z_i, t_i are selected subjects' characteristics, a logarithmic link can be set for individual floatation parameters α_{2i}, $\beta_{2i} > 1$:

$$\log(\alpha_{2i}) = z_i \, \gamma_2; \quad \log(\beta_{2i}) = u_i \, \eta_2,$$

provided that the constraint (2) is taken into account also conditional to covariates, whereas a logit link can be set for polarization parameters $\alpha_1, \beta_3, \delta_1, \delta_3 \in (0, 1)$:

$$\text{logit}(\alpha_{1i}) = y_i \, \gamma_1; \;\; \text{logit}(\beta_{3i}) = w_i \, \eta_3; \;\; \text{logit}(\delta_{1i}) = x_i \, \omega_1; \;\; \text{logit}(\delta_{3i}) = t_i \, \omega_3.$$

3.2 Fitting performances and model selection

Model selection within the OFS class can be performed in terms of likelihood ratio test for pairs of nested models (to compare the symmetric and asymmetric specification for floatation, for instance). More generally, fitting performance of an OFS model against competing alternatives can be assessed by resorting to information criteria: in the following, the BIC index will be considered to account also for model complexity. Standard goodness-of-fit tests relying on Pearson X^2 statistics could be performed provided that $m - 1 - k > 0$, if k is the number of estimable parameters. For instance, $m > 7$ is needed to apply this test for OFS_{111} models.

The normalized Leti's dissimilarity index (Leti 1983):

$$Diss(f, p) = \frac{1}{2} \sum_{r=1}^{m} |f_r - p_r|, \quad Diss(f, p) \in [0, 1], \tag{10}$$

will be considered to measure the goodness of fit of an estimated model $p = p(\theta) = (p_1, \ldots, p_m)$ to the observed relative frequency distribution $f = (f_1, \ldots, f_m)$.

---

[5] It is worth to remark that the OFS model is reversible with respect to the scale, in the sense that if $R \sim \text{OFS}_{111}(\delta_1, \delta_3, \alpha_1, \alpha_2, \beta_2, \beta_3)$, then $m - R + 1 \sim \text{OFS}_{111}(\delta_3, \delta_1, \beta_3, \beta_2, \alpha_2, \alpha_1)$.

With respect to more traditional indicators, as the Hellinger distance $H(\boldsymbol{p}, \boldsymbol{q})$ (Gibbs and Su 2002), so that:

$$H^2(\boldsymbol{p}, \boldsymbol{q}) \leq Diss(\boldsymbol{p}, \boldsymbol{q}) \leq \sqrt{2}\, H(\boldsymbol{p}, \boldsymbol{q}), \quad H^2(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2}\sum_{r=1}^{m}\left(\sqrt{p_r} - \sqrt{q_r}\right)^2, \tag{11}$$

the Dissimilarity value is interpretable as the percentage of responses that are missed by the model[6]. For this reason, it can be also exploited to check the ability of a model $\boldsymbol{p}$, estimated on a training set, to predict the test set distribution $\boldsymbol{f}$. With the same goal and for comparative purposes, the Kullback–Leibler Divergence $KL(\boldsymbol{f}||\boldsymbol{p}) = \sum_{r=1}^{m} f_r \log(\frac{f_r}{p_r})$ will be also computed.

3.3 Inferential issues for the OFS model

Hereafter, the main steps of the expectation–maximization algorithm for mixtures (EM, (McLachlan and Krishnan 1997)) to perform maximum likelihood estimation of parameters are outlined for the general OFS_{111} specification.

For a sample of ratings $\boldsymbol{r} = (r_1, \ldots, r_n)$, the complete log-likelihood of the OFS_{111} model, with parameter vector $\boldsymbol{\theta} = (\delta_1, \delta_3, \alpha_1, \alpha_2, \beta_2, \beta_3)$, is given by:

$$l_c(\boldsymbol{\theta}; \boldsymbol{r}) = \log(\delta_1)\sum_{i=1}^{n} Z_{1i} + \log(\delta_3)\sum_{i=1}^{n} Z_{3i} + \log(1 - \delta_1 - \delta_3)\sum_{i=1}^{n} Z_{2i} \tag{12}$$

$$+ \sum_{i=1}^{n} Z_{1i}\, \log\big(db(r_i; \alpha_1, 1)\big) + \sum_{i=1}^{n} Z_{2i}\, \log\big(db(r_i; \alpha_2, \beta_2)\big)$$

$$+ \sum_{i=1}^{n} Z_{3i}\, \log\big(db(r_i; 1, \beta_3)\big) \tag{13}$$

where Z_{ji} is a random variable with $Z_{ji} = 1$ if the i-th rating is drawn from the j-th component in the mixture, and $Z_{ji} = 0$ otherwise (so $Z_{2i} = 1 - Z_{1i} - Z_{3i}$). Thus, if $\boldsymbol{\theta}^{(k)}$ is the current estimate at the k-th iteration, the posterior probabilities of the i-th rating being drawn from the opponents' component $DB(\alpha_1, 1)$ and the supporters' component $DB(1, \beta_3)$ are computed within the E-step as:

$$\mathbb{E}[Z_{1i}|\boldsymbol{\theta}^{(k)}] = \tau_{1i}^{(k)} = \frac{\delta_1^{(k)}\, db(r_i; \alpha_1^{(k)}, 1)}{Pr(R_i = r_i|\boldsymbol{\theta}^{(k)})};$$

$$\mathbb{E}[Z_{3i}|\boldsymbol{\theta}^{(k)}] = \tau_{3i}^{(k)} = \frac{\delta_3^{(k)}\, db(r_i; 1, \beta_3^{(k)})}{Pr(R_i = r_i|\boldsymbol{\theta}^{(k)})},$$

---

[6] Indeed, from the identity $\min(a, b) = \frac{1}{2}\big((a + b) - |a - b|\big)$ holding for $a, b \in \mathbb{R}^+$, one can write $Diss(\boldsymbol{f}, \boldsymbol{p}) = 1 - \sum_{r=1}^{m} \min(f_r, p_r)$.

so that $\tau_{2i}^{(k)} = 1 - \tau_{1i}^{(k)} - \tau_{3i}^{(k)}$. In case covariates effects are not specified in the model, then one can write $\tau_{ji}^{(k)} = \tau_{jr}^{(k)}$ if $r_i = r$, $r = 1, \ldots, m$, $j = 1, 2, 3$, and the expected complete log-likelihood to be maximized at the M-step can be rewritten as:

$$\mathbb{E}[l_c(\boldsymbol{\theta})|\boldsymbol{\theta}^{(k)}] = Q_1^{(k)}(\delta_1, \delta_3) + Q_2^{(k)}(\alpha_1) + Q_3^{(k)}(\alpha_2, \beta_2) + Q_4^{(k)}(\beta_3),$$

where $(n_1, n_2, \ldots, n_m)$ denotes the frequency distribution of the sample, and one sets:

- $Q_1^{(k)}(\delta_1, \delta_3) = \log(\delta_1) \sum_{r=1}^{m} n_r \tau_{1r}^{(k)} + \log(\delta_3) \sum_{r=1}^{m} n_r \tau_{3r}^{(k)}$
  $+ \log(1 - \delta_1 - \delta_3) \sum_{r=1}^{m} n_r \tau_{2r}^{(k)},$
  yielding, after differentiation, the updated estimates:

$$\delta_1^{(k+1)} = \frac{1}{n} \sum_{r=1}^{m} n_r \, \tau_{1r}^{(k)}; \quad \delta_3^{(k+1)} = \frac{1}{n} \sum_{r=1}^{m} n_r \tau_{3r}^{(k)}; \quad \delta_2^{(k+1)} = 1 - \delta_1^{(k+1)} - \delta_3^{(k+1)};$$

- $Q_2^{(k)}(\alpha_1) = \sum_{r=1}^{m} n_r \, \tau_{1r}^{(k)} \, \log(db(r; \alpha_1, 1));$
  $Q_4^{(k)}(\beta_3) = \sum_{r=1}^{m} n_r \, \tau_{3r}^{(k)} \, \log(db(r; 1, \beta_3));$
- $Q_3^{(k)}(\alpha_2, \beta_2) = \sum_{r=1}^{m} n_r \, \tau_{2r}^{(k)} \, \log(db(r; \alpha_2, \beta_2)).$

At each step, the updated estimates of $\alpha_1, \alpha_2, \beta_2, \beta_3$ have to be obtained from numerical optimization of the corresponding functions, under the required bound constraints[7].

3.4 Small simulation experiment

In order to show the performance of the estimation procedure, a small simulation experiment has been carried out: for each scenario, $B = 200$ samples of size n were generated. Table 1 reports the mean squared error (MSE) of the sampling distribution of parameter estimators obtained over the simulation runs. The average dissimilarity between generating model $\boldsymbol{p}$ and estimated distribution $\hat{\boldsymbol{p}}$ ($\widehat{Diss}(\boldsymbol{p}, \hat{\boldsymbol{p}})$) and between frequency distribution of the sample $\boldsymbol{f}$ and estimated distribution ($\widehat{Diss}(\boldsymbol{f}, \hat{\boldsymbol{p}})$) is reported. Analogous simulation experiments are pursued also for OFS$_{101}$ and OFS$_{110}$ for the sake of completeness (see Tables 2 and 3). Results are satisfactory and indicate that the model is correctly specified and estimated, with efficiency improving with sample size.

---

[7] The optimization of $Q_3^{(k)}(\alpha_2, \beta_2)$ under the inequality constraints (2) can be implemented via the R package nloptr (Ypma 2018). The optimization procedure requires to set a finite upper bound for the floatation parameters α_2, β_2, estimated on logarithmic scale.

Table 1 MSE of the sampling estimators of OFS$_{111}$ parameters

	δ_1	δ_3	α_1	$\log(\alpha_2)$	$\log(\beta_2)$	β_3	$\widehat{Diss}(\boldsymbol{p}, \hat{\boldsymbol{p}})$	$\widehat{Diss}(\boldsymbol{f}, \hat{\boldsymbol{p}})$
OFS$_{111}$: $m = 10$, $\delta_1 = 0.3$; $\delta_3 = 0.2$; $\alpha_1 = 0.4$; $\beta_3 = 0.2$; $\alpha_2 = 3$; $\beta_2 = 2.5$								
$n = 500$	0.0204	0.0255	0.0484	0.1130	0.1140	0.0451	0.0378	0.0284
$n = 2500$	0.0061	0.0111	0.0145	0.0531	0.0555	0.0250	0.0182	0.0119
$n = 10000$	0.0015	0.0034	0.0027	0.0171	0.0193	0.0106	0.0092	0.0057
OFS$_{111}^{(s)}$: $m = 9$, $\delta_1 = 0.2$; $\delta_3 = 0.4$; $\alpha_1 = 0.3$; $\beta_3 = 0.6$; $\alpha_2 = 3$; $\beta_2 = 3$								
$n = 500$	0.0115	0.0236	0.0436	0.1295	0.1295	0.0418	0.0316	0.0293
$n = 2500$	0.0042	0.0193	0.0208	0.0862	0.0862	0.0282	0.0159	0.0124
$n = 10000$	0.0011	0.0069	0.0055	0.0328	0.0328	0.0081	0.0083	0.0060
OFS$_{111}$: $m = 11$, $\delta_1 = 0.3$; $\delta_3 = 0.5$; $\alpha_1 = 0.6$; $\beta_3 = 0.7$; $\alpha_2 = 2.5$; $\beta_2 = 4$								
$n = 500$	0.0172	0.0407	0.0730	0.2345	0.3609	0.0475	0.0361	0.0371
$n = 2500$	0.0113	0.0184	0.0369	0.1130	0.1583	0.0122	0.0174	0.0158
$n = 10000$	0.0052	0.0049	0.0136	0.0271	0.0375	0.0021	0.0089	0.0078

Table 2 MSE of the sampling estimators of OFS_{101} parameters

	δ	α_1	β_3	$\widehat{Diss}(\boldsymbol{p}, \hat{\boldsymbol{p}})$	$\widehat{Diss}(\boldsymbol{f}, \hat{\boldsymbol{p}})$
	$OFS_{101}: m = 7, \delta_1 = 0.6; \alpha_1 = 0.4; \beta_3 = 0.7$				
$n = 500$	0.0327	0.016	0.0529	0.0287	0.0242
$n = 2500$	0.0089	0.0037	0.0093	0.0134	0.0105
$n = 10000$	0.0023	$9 * 10^{-4}$	0.0021	0.0066	0.0054

Table 3 MSE of the sampling estimators of OFS_{110} parameters

	δ	α_1	$\log(\alpha_2)$	$\log(\beta_2)$	$\widehat{Diss}(\boldsymbol{p}, \hat{\boldsymbol{p}})$	$\widehat{Diss}(\boldsymbol{f}, \hat{\boldsymbol{p}})$
	$OFS_{110}: m = 9, \delta_1 = 0.6; \alpha_1 = 0.2; \alpha_2 = 3; \beta_2 = 2$					
$n = 500$	0.0076	0.0044	0.0930	0.0436	0.0311	0.0266
$n = 2500$	0.0015	0.0008	0.0218	0.0101	0.0149	0.0124
$n = 10000$	0.0006	0.0003	0.0080	0.0034	0.0076	0.0060

3.5 Standard errors for OFS parameters

Uncertainty evaluation of parameters estimates could be performed by resorting to asymptotic information theory on the basis of the observed information matrix (see Appendix 1 for details). Potential drawbacks of this procedure may arise due to possible occurrence of numerical overflow in the approximation of the involved integrals. In this respect, numerical derivatives of the log-likelihood can be computed directly with Richardson's extrapolation method, as suggested in (Ursino and Gasparini 2018)[8]. By considering that information theory results apply only asymptotically under regularity conditions, re-sampling methods as the bootstrap (Efron 1981) can be assumed as a general practice for OFS models, allowing to obtain stable accuracy evaluations on parameter estimates even for small sample sizes.

A small Monte-Carlo experiment has been pursued to compare the asymptotic performance of the different methods: for selected OFS models, n observations were sampled. For the general OFS_{111} model, Table 4 reports standard errors' estimates obtained on the basis of the observed information matrix (*Inf.*), numerical approximation of the derivatives of the log-likelihood function with the Richardson's extrapolation method (*Num.*), and nonparametric bootstrap with $B = 500$ replicates (*Boot.*). The three methods are asymptotically equivalent, but for small and moderate sample sizes, the data-driven procedure *Boot* entails more accurate results[9]. For instance, numerical divergence for some of the integrals involved in the computation of the observed information matrix occurred for $n = 500$.

The same check limited to numerical and bootstrap methods is pursued for instances of OFS_{110} and OFS_{101} models (see Tables 5 and 6).

---

[8] The R package `numDeriv` has been considered for this task (Gilbert and Varadhan 2019).

[9] The circumstance that standard errors obtained from bootstrap methods outperform the ones obtained from observed information matrix occurred also in (Basford et al. 1997).

Table 4 Comparison of standard errors: OFS_{111} model with $m = 11, \delta_1 = 0.25; \delta_3 = 0.4; \alpha_1 = 0.2; \beta_3 = 0.6; \alpha_2 = 3; \beta_2 = 4$

	δ_1			δ_3			α_1		
	Inf.	Num.	Boot.	Inf.	Num.	Boot.	Inf.	Num.	Boot.
$n = 500$	NA	0.106	0.073	NA	0.117	0.107	NA	0.257	0.118
$n = 5000$	0.081	0.081	0.058	0.053	0.053	0.057	0.131	0.131	0.086
$n = 50000$	0.022	0.022	0.020	0.016	0.016	0.017	0.030	0.030	0.028
$n = 100000$	0.016	0.016	0.015	0.012	0.012	0.012	0.024	0.024	0.023

	$\log(\alpha_2)$			$\log(\beta_2)$			β_3		
	Inf.	Num.	Boot.	Inf.	Num.	Boot.	Inf.	Num	Boot.
$n = 500$	NA	0.472	0.296	NA	0.454	0.336	NA	0.114	0.114
$n = 5000$	0.223	0.223	0.160	0.197	0.197	0.160	0.061	0.061	0.065
$n = 50000$	0.066	0.066	0.060	0.061	0.061	0.057	0.017	0.017	0.018
$n = 100000$	0.050	0.050	0.047	0.045	0.045	0.043	0.012	0.012	0.013

Table 5 Comparison of standard errors obtained by numerical differentiation of log-likelihood to obtain the Hessian matrix (Num.) and nonparametric bootstrap: OFS_{101} model with $m = 7, \delta_1 = 0.6; \alpha_1 = 0.4; \beta_3 = 0.7$

	δ		α_1		β_3	
	Num.	Boot.	Num.	Boot.	Num.	Boot.
$n = 500$	0.188	0.154	0.123	0.111	0.174	0.181
$n = 5000$	0.075	0.069	0.043	0.041	0.073	0.068
$n = 50000$	0.023	0.022	0.013	0.013	0.022	0.022
$n = 100000$	0.016	0.016	0.010	0.010	0.014	0.014

Table 6 Comparison of standard errors obtained by numerical differentiation of log-likelihood to obtain the Hessian matrix (Num.) and nonparametric bootstrap: OFS_{110} model with $m = 9, \delta_1 = 0.6; \alpha_1 = 0.3; \alpha_2 = 4; \beta_2 = 1.5$

	δ_1		α_1		$\log(\alpha_2)$		$\log(\beta_2)$	
	Num.	Boot.	Num.	Boot.	Num.	Boot.	Num.	Boot.
$n = 500$	0.049	0.054	0.044	0.053	0.268	0.276	0.187	0.179
$n = 5000$	0.018	0.017	0.016	0.015	0.084	0.079	0.054	0.051
$n = 50000$	0.006	0.006	0.005	0.005	0.027	0.029	0.018	0.018
$n = 100000$	0.004	0.004	0.003	0.004	0.020	0.021	0.013	0.013

3.6 A comparative discussion with the state of the art

Like the OFS family, MIHG (Simone and Iannario 2018) and CAUB (Simone and Tutz 2018) mixture models pursue a direct parameterization of the features of interest of the distribution, with easy interpretation and explicit location of modal values (yet the MIHG does not consider floatation). In this context, a 3-component mixture of Binomial distributions could be also considered if suitable constraints are put on Binomial parameters to model polarization and floatation: its specification will not be

discussed hereafter, since the Binomial model can be approximated by the DB model (Ursino 2014): see (Grilli et al. 2015) for further applications of Binomial mixtures to discrete data.

The proposal of the bimodal discrete shifted Poisson model (Bi-Poiss) advanced in (Gómez-Déniz et al. 2020), instead, deals with a construction to encompass bimodal count data starting from the Poisson model, with addition of an extra dispersion parameter θ responsible for bimodality (not necessarily at the extremes of support)[10]. After truncation at m, the main drawback of the Bi-Poiss model is the lack of an explicit link between parameter values and polarization and floatation of the response: for instance, theoretical values for the modes can be obtained in terms of parameters by solving numerically nonlinear equations. In addition, the Bi-Poiss does not encompass the scenario of three response clusters as the OFS$_{111}$ model. Conversely, the Bi-Poiss model is directly applicable in case of bimodality at inner categories, whereas specification of mixtures of DB models in this case should be designed carefully for identifiability issues (see Appendix 2).

For bimodal discrete data, a two-component mixture of (truncated) Conway–Maxwell–Poisson models can be considered as well (Mix-CMP, (Sur et al. 2015))[11]. With respect to computational aspects, the M-step within the EM algorithm needs to be performed with a computationally demanding grid search since the ML solution for Mix-CMP is highly dependent of initial values. With respect to the problem under examine, the main drawback about Mix-CMP concerns identifiability, which causes several limitations on interpretation of the response location and dispersion. Specifically, parameters are not straightforwardly interpretable in terms of polarization and floatation, as for the OFS family. As to fitting performances, a tentative approach to pursue a comparative analysis with the OFS family requires to set suitable parameter constraints to mitigate identifiability issues for the Mix-CMP, at the cost of lack of flexibility. For instance, the supporters' pole can be shaped by restricting to a $CMP(\lambda_S, \nu_S)$ with $\lambda_S \in (m-2, m)$, $\nu_S \in (0, 1)$, whereas a $CMP(\lambda_O, \nu_O)$ model with $\lambda_O, \nu_O \in (0, 1)$ can be considered for the opponents' pole. Floatation could be possibly considered explicitly if a component $CMP(\lambda_F, \nu_F)$, $\lambda_F, \nu_F > 1$, is specified in the mixture. For count data, each component should be truncated from below at the minimum observed count, and from above at the largest observed count or at the censoring threshold, whereas it should be truncated from above at $m-1$ and then shifted upward by 1 in case of ratings on Likert-type scales, as argued for the Binomial component in CUB mixtures (Piccolo and Simone 2019).

---

[10] If $g(x; \lambda)$ denotes the (shifted) Poisson probability function, $x \in \mathbb{N}$, the bimodal discrete shifted Poisson model is defined by the probability mass function:

$$f(x; \lambda, \theta) = w(x; \lambda, \theta)\, g(x; \lambda), \ x \geq 1$$

where $w(x; \lambda, \theta) = \dfrac{2\lambda + \theta(1 + \lambda - x)(2\sqrt{\lambda} + \theta(1 + \lambda - t))}{\lambda(2 + \theta^2)}$. In particular, $\theta > 0$ or $\theta < 0$ for overdispersed or underdispersed distribution with respect to the Poisson.

[11] A discrete random variable $X \sim CMP(\lambda, \nu)$ has the Conway–Maxwell–Poisson distribution of parameters $\lambda, \nu > 0$ if $Pr(X = x) = \dfrac{\lambda^x}{(x!)^\nu} \dfrac{1}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\lambda}}$, $x \geq 0$. If $\nu = 1$, the Poisson model is recovered, whereas $\nu < 1$ or $\nu > 1$ implies overdispersion or underdispersion, respectively.

Table 7 Models comparison for the Number of Days in Hospital dataset discussed in (Sur et al. 2015) (the two-component Mix-CMP has been fitted with the discussed constraints: $\hat{p} = 0.95$, $\hat{\lambda}_1 = 0.64$; $\hat{v}_1 = 0.05$; $\hat{\lambda}_2 = 13$; $\hat{v}_2 = 0.88$)

	Log-lik	BIC	Diss	Mean p value X^2
Mix-CMP (constrained)	−41521.19	83052.38	0.061	0.007
Bi-Poiss	−47691.41	95402.85	0.247	0
MIHG	−41448.58	82915.22	0.056	0.0257
CAUB	−42964.14	85958.34	0.157	0
DB	−43051.03	86122.10	0.128	0
OFS$_{100}$	−44147.53	88305.09	0.194	0
OFS$_{110}$	−41472.07	82984.22	0.023	0.016
OFS$_{101}$	−44147.53	88325.13	0.194	0
OFS$_{111}$	−41235.64	82531.40	0.019	0.098
CUB	−42964.47	85948.43	0.157	0
CUB + she(15)	−42941.99	85914.05	0.155	0

It is worth to remark that for data exhibiting bi-polarization and floatation, a 3-component mixture of CMP would have a higher model complexity than the OFS$_{111}$ model; similarly, the Mix-CMP would be less parsimonious than MIHG and OFS$_{101}$ for U-shaped distributions and than OFS$_{110}$ or OFS$_{011}$ for bimodal data with one mode at one of the extremes.

In order to show that the OFS family is successfully applicable also in case of (truncated) distributions of count data, Table 7 reports some performance indicators of alternative models for the Health Heritage Competition data discussed in (Sur et al. 2015)[12]. Fitting results of a unique DB(α, β) model with no parameter constraints, and of the CUB mixture (Piccolo and Simone 2019), possibly allowing for inflation at the last category (CUB with shelter), are also reported. The last column reports the average of the p values for the Pearson X^2 goodness-of-fit statistics, applied on each test set of a $K = 30$-fold cross-validation[13] based on the model estimated on the remaining $K - 1$ folds[14]: it follows that the OFS$_{111}$ entails very satisfactory performance.

Thus, OFS mixtures could be successfully applied to assess the efficiency of health care structures, for instance, as well as for other count data, thanks to good flexibility in both fitting and interpretation. For instance, in this case floatation covers the intermediate stays, whereas polarization should be interpreted as the predominance of short and long hospitalizations, with parameters α_1, β_3 describing the concentration of brief and lengthy stays towards the lowest and largest count, respectively. Finally, OFS mixing weights quantify how frequent short, intermediate and long hospitalizations are overall. For the example, results indicate that intermediate hospitalizations tend to

---

[12] After omitting zero counts, the observed distribution has been censored at 15, so that the observed scores $\{1, \ldots, 15+\}$ have frequencies (9299, 4548, 2882, 1819, 1093, 660, 474, 316, 263, 209, 145, 135, 111, 65, 479), with saturated log-likelihood $l_{sat} = \sum_{r=1}^{15} n_r \log(\frac{n_r}{n}) = -41180.28$.

[13] The R package caret has been exploited to split the data (Khun 2020).

[14] The choice of setting $K = 30$ allows that that each fold has a moderate sample size of 750 observations.

be as shorter as possible since the floatation component is right-skewed with modal value at the second category[15].

For the subsequent case studies, fitting results of both the Bi-Poiss and the (constrained) Mix-CMP models will be reported for the sake of comparisons.

Remark 4 Noticeably, the latent Beta polarization components $f(x; \alpha_1, 1)$ and $f(x; 1, \beta_3)$ of the OFS family are particular cases of the Kumaraswami distribution (Jones 2009), with density $g(x; \alpha, \beta) = \alpha \beta x^{\alpha-1}(1 - x^{\alpha})^{\beta-1}$ for $x \in [0, 1]$ that is similar to the Beta distribution for several aspects, yet more tractable from the mathematical point of view. Preliminary investigations seem to indicate that mixture specification within this family would not imply identifiability issues as for the Beta mixtures discussed in Appendix 2. Thus, a mixture of two discretized Kumaraswami distributions, one with parameters (α_1, β_1) such that $\min(\alpha_1, \beta_1) < 1$ for polarization, and one component with parameters (α_2, β_2) with $\min(\alpha_2, \beta_2) > 1$ for floatation, could be an alternative model for the problem under examine, yet with lack of straightforward and symmetrical interpretation of parameters with respect to polarization and floatation; further, non-uniform symmetric shapes would not be encompassed.

4 A case study on the probability to vote for German Political Parties

The data analysed in the present section are taken from the GESIS ALLBUS German Social Survey (Gesis 2016). On a rating scale ranging from $1 = $ "*very unlikely*", $10 = $ "*very likely*", respondents were asked to rate: "*How likely it is that you would ever vote for this German party?*". Hereafter, ratings for the four main parties (CDU, SPD, FDP, The Greens) collected in 2002 and 2008 will be considered. The last two categories have been collapsed to yield rating measurements on a scale with $m = 9$ categories. After list-wise omission of missing values, samples of $n = 2738$ and $n = 3056$ observations are analysed for 2002 and 2008 data, respectively. Within the OFS framework, polarization is meant as resoluteness of the opinion of opponents and supporters, whereas floatation can be also interpreted as indecision.

Table 8 reports the best model for each rating variable, selected on the basis of a joint analysis of multiple criteria, including X^2 Statistics, likelihood ratio tests for nested models and BIC values. As a general rule, the most parsimonious specification has been preferred in case of weakly significant evidence for a more complex model, if comparable satisfactory results hold for the other criteria (see Appendix 3 for details).

It follows that:

– For the CDU, the structural components of the probability to vote have not changed neither in size nor in intensity from 2002 to 2008;
– For the SPD, the neutrality component in 2002 has transformed to a more general yet symmetric indecision component;

---

[15] For the OFS$_{111}$, $\hat{\delta}_1 = \underset{(0.023)}{0.452}$, $\hat{\delta}_3 = \underset{(0.0032)}{0.0162}$, $\hat{\alpha}_1 = \underset{(0.007)}{0.175}$, $\widehat{\log(\alpha_2)} = \underset{(0.071)}{0.478}$, $\widehat{\log(\beta_2)} = \underset{(0.060)}{2.258}$, $\hat{\beta}_3 = \underset{(0.077)}{0.010}$, so that OFS$_{111}$ is substantially an OFS$_{110}$ with inflation at $m = 15$ (the *supporters*' pole is degenerate: see Remark 2).

Table 8 Best OFS mixture (see Table 12 in Appendix 3 for details)

	CDU	SPD	FDP	The Greens
2002	$OFS_{111}^{(i)}$	$OFS_{111}^{(i)}$	OFS_{110}	OFS_{110}
2008	$OFS_{111}^{(i)}$	$OFS_{111}^{(s)}$	$OFS_{111}^{(s)}$	$OFS_{111}^{(s)}$
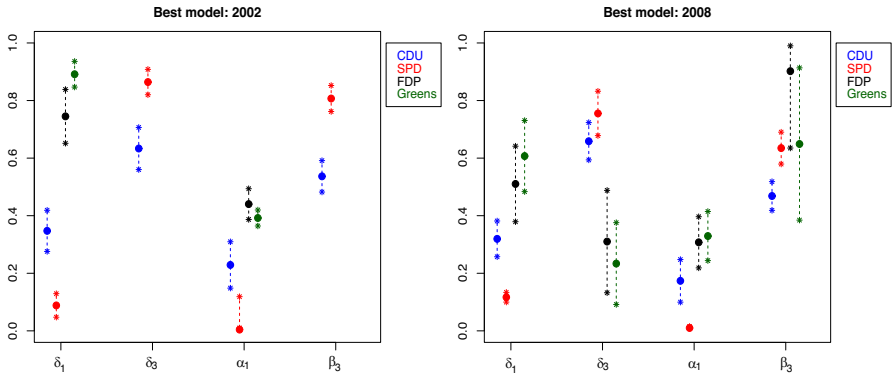
Fig. 1 Polarization parameters for the best OFS mixture (see Table 8)

– For the Greens and the FDP, instead, evidence for the supporter pole was found only in 2008: given the positive asymmetry of the floatation component in 2002 (see Table 9) and its symmetry in 2008, it can be concluded that there has been a movement of the undecided opinions towards the supporter pole from 2002 and 2008.

The parameterization of polarization and indecision accomplished via OFS mixtures allows to identify if and to what extent changes have occurred in the probability to vote for German Parties. Figure 1 shows estimated polarization parameters $\hat{\delta}_1, \hat{\delta}_3, \hat{\alpha}_1, \hat{\beta}_3 \in (0, 1)$ for all parties in 2002 (left panel) and in 2008 (right panel). Lower and upper bounds of 95%-bootstrap confidence intervals are displayed with star symbols at the edge of the whiskers departing from the point estimates. It follows that:

– Polarization and floatation components of the voting probabilities for the CDU are overall stable from 2002 to 2008, in both intensity and size;
– For the SPD, a significant decrease is observed for both δ_3 and β_3: thus, given that no relevant variation is observed for δ_1, it can be inferred that indecision has increased, but positive evaluations have further polarized.
– For the Green and the FDP parties, a significant decrease is observed in both δ_1 and α_1, indicating that the opposition pole grew in intensity but decreased in size. As a result, it can be inferred that some negative yet un-polarized evaluations have floated towards a symmetric indecision (see also Table 9).

Figure 2 provides a joint representation of estimation results for the sizes of polarization and floatation with a ternary plot of mixing weights (left), whereas a scatter plot of polarization parameters α_1, β_3 in the unit square is displayed to compare the strengths of unfavourable and favourable opinions over time (right).

Table 9 Asymmetry and Adjusted kurtosis of the floatation component for the best OFS mixture (see Table 8)

	γ_1				$\gamma_2^\star$			
	CDU	SPD	FDP	The Greens	CDU	SPD	FDP	The Greens
2002	0	0	0.0756	0.324	–	–	2.513	2.649
2008	0	0	0	0	–	2.679	2.656	2.671
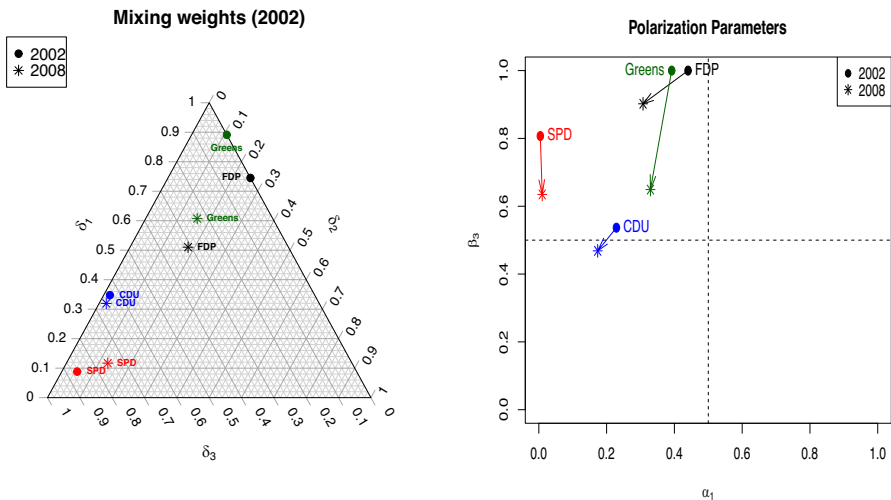
Fig. 2 Visualization of estimation results for polarization parameters: Probability to vote for German parties

Finally, Table 9 reports the chosen asymmetry measure γ_1 defined in (4) and the adjusted kurtosis value $\gamma_2^\star$ (15) for the estimated indecision component for those parties and time points where it is not degenerate. The extent of floatation of negative opinions towards neutrality is then quantified, as is the extent by which un-polarized opinions became more homogeneous from 2002 to 2008 for both the FPD and the Greens (more for the FDP than for the Greens). The reverse circumstance is observed for SPD, for which the neutrality component in 2002 left the place to a general yet symmetric indecision. The analysis and the proposed visualization tools for the results could be replicated conditional to covariates values (as gender, geographical residence, etc) to give local assessments of the polarization and floatation dimensions.

Finally, a 10-fold cross-validation is performed to check the ability of the selected best model (Table 8) to predict the rating distribution. Table 10 reports some summarizing indicators: average and 9th decile over folds of the dissimilarity index between the best model $p_{\text{train}}^\star$, estimated on the training set, and the response distribution on the test set (f_{test}), are proposed as a proxy of prediction errors for the test set distribution. With the same goal and strategy, the average over folds of the Kullback–Leibler divergence is reported for candidate OFS models. Results indicate that, beyond fitting ability, the flexibility of OFS models allows to attain satisfactory predictive performance.

Table 10 Summarizing results for goodness of fit and predicting ability of the best model, over 10-fold cross-validation

	2002				2008			
	CDU	SPD	FDP	The Greens	CDU	SPD	FDP	The Greens
Average p value for X^2	0.186	0.369	0.077	0.141	0.333	0.088	0.121	0.239
Mean $Diss(f_{test}, \boldsymbol{p}_{\text{train}}^{\star})$	0.061	0.048	0.074	0.056	0.047	0.071	0.065	0.054
9th Decile: $Diss(f_{test}, \boldsymbol{p}_{\text{train}}^{\star})$	0.085	0.069	0.097	0.077	0.068	0.087	0.088	0.079
OFS_{111}: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.017	0.008	0.019	0.015	0.013	0.016	0.016	0.011
$OFS_{111}^{(s)}$: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.017	0.008	0.019	0.018	0.013	0.018	0.016	0.012
OFS_{101}: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.016	0.018	0.042	0.021	0.013	0.026	0.033	0.022
$OFS_{111}^{(i)}$: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.015	0.009	0.029	0.018	0.010	0.019	0.021	0.014
OFS_{110}: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.052	0.020	0.019	0.015	0.075	0.029	0.022	0.019
$OFS_{110}^{(s)}$: Mean $KL(f_{test} \Vert \boldsymbol{p}_{\text{train}})$	0.089	0.041	0.019	0.018	0.134	0.079	0.022	0.020

5 Final considerations

The paper has discussed mixture specification of Discretized Beta models to explicitly parameterize polarization and floatation of discrete ordered evaluations, as ratings and (truncated) count data. The proposal is more flexible than other alternative models in both fitting performance and interpretation: for instance, the method presented in (Gómez-Déniz et al. 2020) to induce bimodality in a distribution could be applied also to the DB model, at the cost of losing the direct a-priori parameterization of polarization features afforded by the OFS models. A devoted R package for OFS implementation is under development.

Further research will be tailored to the analysis of tail dependencies of polarization and floatation of different survey items with suitable copula modelling, as well as to the implementation of model-based trees to derive response profiles in terms of covariates entailing a significant effect in at least one model's features (see (Cappelli et al. 2019; Simone et al. 2019) for the case of CUB models for rating data). A comparative analysis with mixtures of discretized Kumaraswami distributions also deserve in-depth investigation in future research.

Appendix 1: Miscellanea about the Discretized Beta model

For a generic random variable X, let μ and σ be the mean value and the standard deviation, respectively. If m_3 and m_4 denote the standardized central moments around the mean of order three and four, respectively, then the *meson* f is the location index proposed in (Blest 2003), defined by:

$$f = \left(\sqrt{\frac{1}{4}m_3^2 + 1} + \frac{1}{2}m_3\right)^{\frac{1}{3}} - \left(\sqrt{\frac{1}{4}m_3^2 + 1} - \frac{1}{2}m_3\right)^{\frac{1}{3}} \tag{14}$$

and verifying $\mathbb{E}[(X - \xi)^3] = 0$, if $\xi = \mu + f\sigma$, so that ξ minimizes the fourth moment of X around the mean μ. Then, the measure of kurtosis adjusted for skewness proposed in (Blest 2003) allows to compare distributions with the same degree of flatness, regardless of the location of their tails. It is defined as the standardized fourth moment around the *meson* f, namely:

$$\gamma_2^\star = m_4 - 3\big((1 + f^2)^2 - 1\big). \tag{15}$$
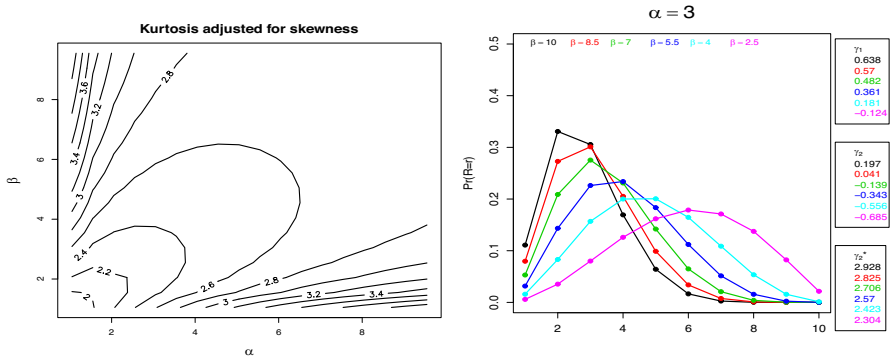
Fig. 3 Left: contour lines for the adjusted measure of kurtosis for a DB model with $\min(\alpha, \beta) > 1$. Right: Continuous lines matching probability masses for selected DB models (skewness, kurtosis and adjusted kurtosis are reported in legend)

For the DB model, Fig. 3 displays contour lines for the adjusted kurtosis index (left) and selected DB distributions with corresponding values of $\gamma_1(\alpha, \beta)$, $\gamma_2^\star(\alpha, \beta)$ (right). As a benchmark, notice that $\gamma_2^\star = 1.8$ for the uniform distribution, whereas $\gamma_2^\star \to 3$ if $\alpha = \beta \to \infty$.

Appendix 1.2: On the observed information matrix for DB and OFS models

If $\Gamma(x)$ denotes the Euler Gamma function, let $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ be the *digamma* function, and $\psi_1(x) = \frac{d}{dx}\psi(x)$ be the *trigamma* function. To shorten the notation, for $k = 1, \ldots, m$, let $db_k = Pr(R = k|a, b) = \dfrac{1}{B(a, b)} I_k(a, b)$, where $I_k(a, b)$

$= \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}dx$, and consider the identity $\dfrac{\partial B(x, y)}{\partial x} = B(x, y)\big(\psi(x) - \psi(x$

$+ y)\big)$. By virtue of the symmetry of the Beta function ($B(x, y) = B(y, x)$), the first- and second-order derivatives of the logarithm of DB probabilities-according to the chosen parameterization-are:

- $\dfrac{\partial \log(db_k)}{\partial a} = \psi(a+b) - \psi(a) + \frac{1}{I_k(a,b)}\frac{\partial I_k(a,b)}{\partial a};$
- $\dfrac{\partial \log(db_k)}{\partial b} = \psi(a+b) - \psi(b) + \frac{1}{I_k(a,b)}\frac{\partial I_k(a,b)}{\partial b};$
- $\dfrac{\partial^2 \log(db_k)}{\partial^2 a} = \psi_1(a+b) - \psi_1(a) - \frac{1}{I_k(a,b)^2}\left(\frac{\partial I_k(a,b)}{\partial a}\right)^2 + \frac{1}{I_k(a,b)}\frac{\partial^2 I_k(a,b)}{\partial^2 a};$
- $\dfrac{\partial^2 \log(db_k)}{\partial^2 b} = \psi_1(a+b) - \psi_1(b) - \frac{1}{I_k(a,b)^2}\left(\frac{\partial I_k(a,b)}{\partial b}\right)^2 + \frac{1}{I_k(a,b)}\frac{\partial^2 I_k(a,b)}{\partial^2 b};$
- $\dfrac{\partial^2 \log(db_k)}{\partial a\,\partial b} = \psi_1(a+b) - \frac{1}{I_k(a,b)^2}\left(\frac{\partial I_k(a,b)}{\partial a}\right)\left(\frac{\partial I_k(a,b)}{\partial b}\right) + \frac{1}{I_k(a,b)}\frac{\partial^2 I_k(a,b)}{\partial a\partial b};$

where:

- $\dfrac{\partial I_k(a,b)}{\partial a} = \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}(\log(x))dx;$
- $\dfrac{\partial^2 I_k(a,b)}{\partial^2 a} = \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}(\log(x))^2 dx;$
- $\dfrac{\partial I_k(a,b)}{\partial b} = \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}(\log(1-x))dx;$
- $\dfrac{\partial^2 I_k(a,b)}{\partial^2 b} = \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}(\log(1-x))^2 dx;$
- $\dfrac{\partial^2 I_k(a,b)}{\partial a\partial b} = \int_{\frac{k-1}{m}}^{\frac{k}{m}} x^{a-1}(1-x)^{b-1}(\log(1-x)\log(x))dx.$

Then, from the results above, the observed information matrix for OFS models can be promptly derived (the chain rule has to be suitably applied if the log transform is considered for the parameters of a given DB component). Similar steps are needed to compute first- and second-order derivatives of the complete log-likelihood function in order to obtain the observed information matrix within the EM algorithm with Louis' identity and to speed up convergence, as proposed in (Simone 2021) for CUB models.

Appendix 2: On identifiability of Discretized Beta mixture models

On the basis of a characterization proven in (Yakowitz and Spragins 1968), mixtures of Beta distributions are not identifiable, in general (Ahmad and Al-Hussaini 1982). Thus, suitable constraints should be put on parameters of mixture components when

combining DB models for polarization with a DB model for floatation. The core of the discussion is the following theorem, whose proof follows straightforwardly.

Theorem 1 *Let $f(x; \alpha, \beta)$ be the Beta probability density function over $[0, 1]$, with $\alpha, \beta > 0$. Then:*

$$f(x; \alpha, \beta) = c_1 f(x; \alpha, \beta + 1) + (1 - c_1) f(x; \alpha + 1, \beta) \qquad (16)$$

where $c_1 = \dfrac{\beta}{\alpha + \beta}$. In particular, if $\max(\alpha, \beta) < 1$, the U-shaped Beta density function $f(x; \alpha, \beta)$ can be written as a mixture of a J-shaped Beta $f(x; \alpha + 1, \beta)$ and a reverse J-shaped Beta $f(x; \alpha, \beta + 1)$.

Mixture specification within the DB family should consider also the following arguments leading to a Beta approximation of Beta mixtures[16].

For a fixed $k \geq 2$, consider a mixture $g(x) = \sum_{i=1}^{k} d_i f(x; \alpha_i, \beta_i)$ of Beta densities. If μ_{i1} and μ_{i2} denote the first and second moments of the i-th mixture component, let $\mu_1 = \sum_{i=1}^{k} d_i \mu_{i1}$, $\mu_2 = \sum_{i=1}^{k} d_i \mu_{i2}$ be the first and second moment of the mixture, respectively. If $s = \mu_2 - \mu_1^2$ is the variance of the mixture, and $h = \frac{\mu_1}{1 - \mu_1}$, the following approximation can be derived:

$$g(x) \approx f(x; \alpha, \beta), \quad \text{with} \beta = \frac{h}{s(1 + h)^3} - \frac{1}{1 + h}, \quad \alpha = h\beta. \qquad (17)$$

For instance, assume that $k = 2$ and that $X_1 \sim f(\alpha_1, \beta_1)$ is J-shaped, whereas $X_2 \sim f(\alpha_2, \beta_2)$ is reversed J-shaped. Their mixture $g(x)$ (with weights d_1 and $1 - d_1$) can be approximated by a U-shaped Beta density $f(x; \alpha, \beta)$ with parameters (α, β) obtained as in (17). Table 11 reports some instances. The last 4 columns report the Dissimilarity index between the discretized versions of $g(x)$ and its approximation $f(x; \alpha, \beta)$ given in (17), for varying number of categories m. Results indicate that this approximation is satisfactory: thus, an OFS_{101} model could be approximated by a $DB(\alpha, \beta)$ model, which in turn can be written as a further mixture of two DB models after discretization of the representation in (16). Then, specifying DB mixtures is a challenging task, especially for small m.

Remark 5 As to explicative performances, it is worth to notice that the OFS_{101} model allows to assess the prevalence of opponents with respect to supporters directly in terms of δ_1, whereas this can be assessed only indirectly from the mixture decomposition established in Theorem 1 for a $DB(\alpha_1, \beta_3)$ model.

Appendix 2.1: Two-component mixture of DB models for polarization and floatation

With reference to model specification (6), assume that $\max(\alpha_1, \beta_1) > 1$: without loss of generality, assume that $\alpha_1 = \min(\alpha_1, \beta_1) < 1$ and that $\beta_1 = \max(\alpha_1, \beta_1) > 1$. By

---

[16] The approximation (17) stems from a straightforward adaptation to mixtures of the method of moments applied in (Jóhannesson and Giri 1995) to derive a Beta approximation to linear combinations of independent Beta random variables.

Table 11 Instances of Beta approximation of 2-component mixtures of Beta distributions according to (17)

	d_1	α_1	β_1	α_2	β_2	α	β	$Diss$ $(m=5)$	$Diss$ $(m=7)$	$Diss$ $(m=9)$	$Diss$ $(m=11)$
Case 1	0.4	0.2	1	1	0.1	0.190	0.120	0.004	0.006	0.008	0.010
Case 2	0.8	0.7	1	1	0.3	0.638	0.682	0.014	0.020	0.026	0.030
Case 3	0.5	0.3	1	1	0.4	0.371	0.415	0.006	0.007	0.010	0.012
Case 4	0.2	0.3	0.6	0.5	0.2	0.395	0.224	0.002	0.003	0.003	0.003

applying Theorem 1 on the latent polarization component with parameters (α_1, β_1), it follows that:

$$
Pr(R=r|\boldsymbol{\theta}) = \frac{(1-\delta)\beta_1}{\alpha_1+\beta_1}db(r;\alpha_1,\beta_1+1)
$$

$$
+\left(\frac{(1-\delta)\alpha_1}{\alpha_1+\beta_1}db(r;\alpha_1+1,\beta_1) + \delta\, db(r;\alpha_2,\beta_2)\right) \quad (18)
$$

$$
\approx (1-\delta^\star)db(r;\alpha_1^\star,\beta_1^\star) + \delta^\star db(r;\alpha_2^\star,\beta_2^\star) = Pr(R=r|\boldsymbol{\theta}^\star), \quad (19)
$$

where:

- $\alpha_1^\star = \alpha_1 \in (0,1)$, $\beta_1^\star = \beta_1 + 1 > 1$ (so that DB$(\alpha_1^\star, \beta_1^\star)$ is reverse J-shaped as DB(α_1, β_1));
- $\delta^\star = \delta + \frac{(1-\delta)\alpha_1}{\alpha_1+\beta_1}$;
- $\alpha_2^\star, \beta_2^\star$ are obtained numerically following (17) and thus satisfy:

$$
\delta^\star db(r;\alpha_2^\star,\beta_2^\star) \approx \frac{(1-\delta)\alpha_1}{\alpha_1+\beta_1}db(r;\alpha_1+1,\beta_1) + \delta\, db(r;\alpha_2,\beta_2).
$$

If the constraints $\min(\alpha_2^\star, \beta_2^\star) > 1$ and (2) are satisfied, identifiability concerns may arise depending on the goodness of the approximation $Pr(\cdot;\boldsymbol{\theta}) \approx Pr(\cdot;\boldsymbol{\theta}^\star)$ and on the distance between parameter vectors.

Similar issues may arise if $\max(\alpha_1,\beta_1) \leq 1$. Applying Theorem 1 iteratively to the underlying Beta mixture, it follows that:

$$
Pr(R=r|\boldsymbol{\theta}) = (1-\delta)\left(\frac{\beta_1}{\alpha_1+\beta_1}db(r;\alpha_1,\beta_1+1) + \frac{\alpha_1}{\alpha_1+\beta_1}db(r;\alpha_1+1,\beta_1)\right) + \delta db(r;\alpha_2,\beta_2)
$$

$$
\approx (1-\delta^\star)db(r;\alpha_1^\star,\beta_1^\star) + \delta^\star db(r;\alpha_2^\star,\beta_2^\star) = Pr(R=r|\boldsymbol{\theta}^\star),
$$

where (17) implies that:

-

$$
f(x;\alpha_1^\star,\beta_1^\star) \approx \frac{d_1}{d_1+d_2}f(x;\alpha_1,\beta_1+2) + \frac{d_2}{d_1+d_2}f(x;\alpha_1+2,\beta_1); \quad (20)
$$

- $d_1 = (1-\delta)\frac{\beta_1}{\alpha_1+\beta_1}\frac{\beta_1+1}{\alpha_1+\beta_1+1}$ and $d_2 = (1-\delta)\frac{\alpha_1}{\alpha_1+\beta_1}\frac{\alpha_1+1}{\alpha_1+\beta_1+1}$;

–

$$\delta^\star f(x; \alpha_2^\star, \beta_2^\star)$$
$$\approx \left(\frac{2(1-\delta)\alpha_1\beta_1}{(\alpha_1+\beta_1)(\alpha_1+\beta_1+1)} f(x; \alpha_1+1, \beta_1+1) + \delta f(x; \alpha_2, \beta_2) \right).$$
(21)

Thus, despite immediate interpretation of parameters, the mixture specification in (6) does not ensure identifiability.

Appendix 2.2: Three-component mixture of DB models for polarization and floatation

From the specification point of view, model (7) has a higher model complexity than model (6), yet the two mixture specifications have the same explicative power. Theorem 1 implies that, if $\max(\alpha_1, \beta_1) < 1$, one can write:

$$R \sim (1-\delta)\mathrm{DB}(\alpha_1, \beta_1) + \delta \mathrm{DB}(\alpha_2, \beta_2)$$
$$\sim \delta_1 \mathrm{DB}(\alpha_1^\star, \beta_1^\star) + \delta_2 \mathrm{DB}(\alpha_2^\star, \beta_2^\star) + \delta_3 \mathrm{DB}(\alpha_3^\star, \beta_3^\star)$$

with $\delta_1 = (1-\delta)\frac{\beta_1}{\alpha_1+\beta_1}, \alpha_1^\star = \alpha_1, \beta_1^\star = 1+\beta_1 > 1$; $\delta_2 = \delta, \alpha_2^\star = \alpha_2, \beta_2^\star = \beta_2$; $\delta_3 = (1-\delta)\frac{\alpha_1}{\alpha_1+\beta_1}, \alpha_3^\star = \alpha_3+1, \beta_3^\star = \beta_3$.

Vice versa, the convex combination of the two polarization components within the mixture (7) can be approximated by a unique U-shaped polarization component. Indeed, if a continuous random variable X has density:

$$g_X(x) = \frac{1}{\delta_1 + \delta_3} \big(\delta_1 f(x; \alpha_1, \beta_1) + \delta_3 f(x; \alpha_3, \beta_3) \big),$$

then $g_X(x)$ will be U-shaped and can be numerically approximated by a Beta model according to (17). Thus, for the underlying Beta specification of the 3-component mixture (7), it follows that:

$$\delta_1 f(x; \alpha_1, \beta_1) + \delta_3 f(x; \alpha_3, \beta_3) \approx (\delta_1 + \delta_3) f(x; \alpha, \beta).$$
(22)

If parameters (α, β) are obtained numerically as in (17), then the 3-component mixture (7) can be approximated by a 2-component mixture as in (6):

$$Pr(R = r|\boldsymbol{\theta}) \approx (\delta_1 + \delta_3)\, db(r; \alpha, \beta) + \delta_2\, db(r; \alpha_2, \beta_2).$$

Then, identifiability issues may arise for the 3-component mixture (7) if $\boldsymbol{\theta}^\star$ and $\boldsymbol{\theta}$ are not close enough while the approximation is good: indeed, by applying Theorem 1

to the approximating Beta(α, β) in (22), since $\alpha, \beta \in (0, 1)$, one could write:

$$Pr(R = r|\boldsymbol{\theta}) \approx (\delta_1 + \delta_3) \left(\frac{\beta}{\alpha + \beta} db(r; \alpha, \beta + 1) + \frac{\alpha}{\alpha + \beta} db(r; \alpha + 1, \beta) \right)$$
$$+ \delta_2 db(r; \alpha_2, \beta_2)$$
$$= \delta_1^{\star} db(r; \alpha_1^{\star}, \beta_1^{\star}) + \delta_2^{\star} db(r; \alpha_2^{\star}, \beta_2^{\star}) + \delta_3^{\star} db(r; \alpha_3^{\star}, \beta_3^{\star})$$

with $\delta_1^{\star} = (\delta_1 + \delta_3) \frac{\beta}{\alpha+\beta}$, $\alpha_1^{\star} = \alpha \in (0, 1)$, $\beta_1^{\star} = \beta + 1 > 1$, $\delta_2^{\star} = \delta_2, \alpha_2^{\star} = \alpha_2, \beta_2^{\star} = \beta_2, \delta_3^{\star} = (\delta_1 + \delta_3) \frac{\alpha}{\alpha+\beta}$, $\alpha_3^{\star} = \alpha + 1 > 1$, $\beta_3^{\star} = \beta_3 \in (0, 1)$.

Appendix 3: Fitting results for the case study

Table 12 reports the main fitting indicators for the competing models estimated on the data used in Sect. 4: for each criterion, the best performance are highlighted in bold font.

Table 12 Summary of fitting results for the case study presented in Sect. 4

	OFS$_{111}$	OFS$_{111}^{(s)}$	OFS$_{101}$	OFS$_{111}^{(i)}$	OFS$_{110}$	OFS$_{110}^{(s)}$	DB	MIHG	CAUB	Bl-Poiss	Mix-CMP
2002											
CDU: Loglik	−5645.27	−5645.27	−5645.26	**−5639.60**	−5733.70	−5840.61	−5655.62	−5651.06	−5653.73	−6427.07	−5713.40
CDU: BIC	11338.02	11330.11	11314.27	**11310.86**	11499.07	11699.87	11327.08	11325.86	11331.20	12869.97	11457.88
CDU: X^2	5.67	5.67	5.67	**3.62**	39.89	94.65	9.40	8.01	8.69	356.97	30.29
CDU: $Diss$	0.04	0.04	0.04	**0.03**	0.11	0.13	0.05	0.04	0.04	0.32	0.10
SPD: Loglik	**−5921.02**	−5921.53	−5949.64	−5924.46	−5953.24	−6010.09	−5967.87	−5949.69	−5965.53	−8208.67	−5984.14
SPD: BIC	11889.54	11882.64	11923.02	**11880.58**	11938.13	12038.83	11951.57	11923.13	11954.81	16433.17	11999.36
SPD: X^2	**0.16**	0.34	11.50	1.45	12.35	32.78	17.84	11.51	16.86	5437.91	23.64
SPD: $Diss$	**0.01**	0.01	0.05	0.02	0.06	0.11	0.07	0.05	0.06	0.37	0.09
FDP: Loglik	−5562.94	−5565.88	−5631.46	−5594.03	**−5562.94**	−5565.88	−5595.27	−5642.50	−5595.02	−6019.39	−5638.31
FDP: BIC	11173.38	11171.34	11286.66	11219.73	11157.54	**11150.41**	11206.37	11308.75	11213.78	12054.60	11307.68
FDP: X^2	2.57	3.62	27.74	13.33	**2.57**	3.62	14.82	31.52	14.03	234.73	30.90
FDP: $Diss$	0.03	0.04	0.08	0.05	**0.03**	0.04	0.06	0.12	0.06	0.22	0.10
Greens: Loglik	**−5196.44**	−5208.80	−5215.22	−5206.19	−5199.03	−5208.80	−5209.47	−5249.57	−5260.69	−6548.37	−5249.88
Greens: BIC	10440.37	10457.17	10454.18	10444.03	**10429.72**	10436.24	10434.77	10522.89	10545.12	13112.56	10530.84
Greens: X^2	**3.21**	8.04	10.35	7.07	4.18	8.04	8.12	22.59	27.48	1208.42	22.13
Greens: $Diss$	**0.02**	0.04	0.05	0.03	0.03	0.04	0.05	0.09	0.08	0.38	0.09

Table 12 continued

	OFS$_{111}$	OFS$_{111}^{(s)}$	OFS$_{101}$	OFS$_{111}^{(i)}$	OFS$_{110}$	OFS$_{110}^{(s)}$	DB	MIHG	CAUB	BI-Poiss	Mix–CMP
2008											
CDU: Loglik	−6149.81	−6150.28	−6152.50	**−6143.68**	−6335.85	−6513.82	−6169.43	−6163.64	−6169.46	−7225.23	−6284.03
CDU: BIC	12347.76	12340.69	12329.07	**12319.45**	12703.80	13046.29	12354.91	12351.36	12363.00	14466.52	12599.13
CDU: X^2	3.82	4.02	4.92	**1.90**	69.52	164.66	9.99	8.66	9.99	436.43	46.43
CDU: $Diss$	0.02	0.03	0.03	**0.02**	0.14	0.16	0.05	0.05	0.04	0.36	0.13
SPD: Loglik	**−6523.06**	−6526.12	−6555.88	−6534.48	−6563.49	−6714.69	−6596.71	−6553.49	−6607.26	−9233.81	−6618.57
SPD: BIC	13094.27	**13092.37**	13135.84	13101.05	13159.09	13448.03	13209.46	13131.05	13238.59	18483.67	13268.21
SPD: X^2	**1.90**	2.85	12.66	5.38	15.86	64.56	24.18	12.04	26.75	5785.76	31.89
SPD: $Diss$	**0.02**	0.03	0.05	0.03	0.07	0.14	0.08	0.05	0.08	0.38	0.11
FDP: Loglik	−6287.54	**−6287.54**	−6345.14	−6307.25	−6307.28	−6308.94	−6349.59	−6347.56	−6312.95	−6842.33	−6391.41
FDP: BIC	12623.22	**12615.20**	12714.35	12646.59	12646.67	12636.52	12715.24	12719.19	12649.97	13700.71	12813.90
FDP: X^2	2.64	**2.64**	22.62	9.21	9.32	9.81	23.15	22.46	11.01	232.24	35.90
FDP: $Diss$	0.02	**0.02**	0.07	0.05	0.05	0.05	0.08	0.07	0.05	0.21	0.11
Greens: Loglik	**−6230.50**	−6232.46	−6267.03	−6241.05	−6256.94	−6259.01	−6281.29	−6268.20	−6265.17	−6874.32	−6338.57
Greens: BIC	12509.15	**12505.04**	12558.13	12514.19	12545.97	12536.66	12578.63	12560.47	12554.42	13764.69	12708.20
Greens: X^2	**1.76**	2.44	14.36	5.17	10.56	11.22	18.61	14.23	13.10	280.00	35.57
Greens: $Diss$	0.02	**0.02**	0.06	0.04	0.05	0.05	0.07	0.05	0.06	0.22	0.11

References

Abramowitz M, Stegun IA (eds) (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th printing. Dover, New York, p 930

Ahmad KE, Al-Hussaini EK (1982) Remarks on the non-identifiability of mixtures of distributions. Ann Inst Stat Math 34:543–544

Apouey B (2007) Measuring health polarization with self-assessed health data. Health Econ 16(9):875–894

Basford KE, Greenway DR, McLachlan GJ, Peel D (1997) Standard errors of fitted component means of normal mixtures. Comput Stat 12(1):1–17

Blest DC (2003) A new measure of kurtosis adjusted for skewness. Aust N Z J Stat 45(2):175–179

Capecchi S, Piccolo D (2017) The distribution of Net Promoter Score in socio-economic surveys. In: SIS 2017. Statistics and Data Science: new challenges, new generations. Proceedings of the Conference of the Italian Statistical Society. Petrucci A., Verde R. Editors. 247–252. ISBN: 978-88-6453-521-0

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

Efron B (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 68(3):589–599

Fasola S, Sciandra M (2015) New flexible probability distributions for ranking data. In: Morlini I, Minerva T, Vichi M (eds) Advances in statistical models for data analysis. Springer, Berlin, pp 117–124

Forbes C, Evans M, Hastings N, Peacock B (2011) Statistical distributions, 4th edn. Wiley, Hoboken

GESIS Leibniz Institute for the Social Sciences (2016) German General Social Survey (ALLBUS) - Cumulation 1980-2014, GESIS Data Archive, Cologne. ZA4584 Data file version 1.0.0. https://doi.org/10.4232/1.12574

Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. Int Stat Rev 70:419–435

Gilbert P, Varadhan R (2019) numDeriv: accurate numerical derivatives. R package version 2016.8-1.1. https://CRAN.R-project.org/package=numDeriv

Gómez-Déniz E, Pérez-Rodríguez JV, Reyes J, Gómez HW (2020) A bimodal discrete shifted Poisson distribution. A case study of tourists' length of stay. Symmetry 12(3):442

Grilli L, Rampichini C, Varriale R (2015) Binomial mixture modelling of University Credits. Commun Stat Theory Methods 44(22):4866–4879

Jóhannesson B, Giri N (1995) On approximations involving the Beta distribution. Commun Stat Simul Comput 24(2):489–503

Jones MC (2009) Kumaraswamy's distribution: a beta-type distribution with some tractability advantages. Stat Methodol 6(1):70–81

Leti G (1983) Statistica descrittiva. Il Mulino, Bologna

Kuhn M (2020) caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret

McLachlan GJ, Krishnan T (1997) The EM algorithm and extensions, 2nd edn. Wiley Series in Probability and Statistics

Morrison DG (1979) Purchase intentions and Purchase Behavior. J Mark 43(2):65–74

Mussini M (2018) On measuring polarization for ordinal data: an approach based on the decomposition of the Leti index. Stat Transit New Ser 19(2):277–296

Piccolo D, Simone R (2019) The class of cub models: statistical foundations, inferential issues and empirical evidence. Stat Method Appl 28(3):389–435 (with discussions and rejoinder)

Reichheld FF (2003) The one number you need to grow. Harv Bus Rev 81:46–54

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

Simone R, Iannario M (2018) Analysing sport data with clusters of opposite preferences. Stat Model 18(5–6):505–524

Simone R, Tutz G (2018) Modelling uncertainty and response styles in ordinal data. Stat Neerl 72:224–245

Simone R, Cappelli C, Di Iorio F (2019) Modelling marginal ranking distributions: the uncertainty tree. Pattern Recogn Lett 125:278–288

Sur P, Shmueli G, Bose S, Dubey P (2015) Modeling bimodal discrete data using Conway–Maxwell–Poisson mixture models. J Bus Econ Stat 33(3):352–365

Ursino M (2014) Ordinal Data: a new model with applications, PhD Thesis, http://porto.polito.it/2535701/, Politecnico di Torino, Italy

Ursino M, Gasparini M (2018) A new parsimonious model for ordinal longitudinal data with application to subjective evaluation of a gastrointestinal disease. Stat Methods Med Res 27(5):1376–1393

Yakowitz SJ, Spragins JD (1968) On the identifiability of finite mixtures. Ann Math Stat 39:209–214

Ypma J, with contributions by Borchers HW, Eddelbuettel D (2018) https://CRAN.R-project.org/package=nloptr