Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v13n2p519

How perceived variety impacts on choice satisfaction: a two-step approach using the CUB class of models and best-subset variable selection

By Manisera, Zuccolotto, Brentari

Published: 14 October 2020

Electronic Journal of Applied Statistical Analysis
Vol. 13, Issue 02, October 2020, 519-535
DOI: 10.1285/i20705948v13n2p519

How perceived variety impacts on choice satisfaction: a two-step approach using the CUB class of models and best-subset variable selection

Marica Manisera*[a], Paola Zuccolotto[b], and Eugenio Brentari[c]

[a,c]*Dipartimento di Economia e Management - Sede Santa Chiara Contrada S. Chiara, 50, 25122, Brescia, Italy*
[b]*Dipartimento di Economia e Management - Sede S. Faustino via S. Faustino, 74/B, 25122, Brescia, Italy*

Published: 14 October 2020

In consumer research, marketing, public policy and other fields, individuals' choice depends on the number of possible alternatives. In addition, according to the literature, the choice satisfaction is influenced not only by the number of options but also by the perceived variety. The aim of the present study is to apply a novel approach to model perceived variety, in order to better understand the perceptions of individuals about the variety of the possible choice options and to model the impact of perceived variety and individuals' characteristics on the choice outcome satisfaction. We resort to the class of CUB (Combination of Uniform and Binomial random variables) models for rating data that model the respondents' decision process as a combination of two latent components, called *feeling* and *uncertainty*, that express, respectively, the level of agreement with the item being evaluated and the human indecision surrounding any discrete choice. The model applied in this paper is an alternative to the most common models used in the studies of human judgments and decisions, whenever attitudes, perceptions and opinions are measured by means of questionnaires having questions with ordered response categories. The chosen approach is composed of two steps: (1) we construct measures of feeling and uncertainty of perceived variety by means of CUB and (2) we investigate their impact (eventually together with personal characteristics) on choice satisfaction. The R FastCUB package is exploited to select the best set of covariates to include in the final model.

*Corresponding author: marica.manisera@unibs.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

keywords: CUB models, feeling, ordinal data, ratings, uncertainty, perceived variety

1 Introduction

According to several studies in the field of human judgement and decision making, the number of options in a choice process influence choice satisfaction (Reutskaja and Hogarth, 2009; Szrek, 2017; Hafner et al., 2018). Some authors support the existence of a phenomenon called "choice overload" or "overchoice", occurring when several - approximately equivalent - options are available and individuals find it difficult to make a decision and report lower choice satisfaction (Schwartz, 2004; Haynes Graeme, 2009; Scheibehenne et al., 2010; Chernev et al., 2015). Satisfaction with choices can be described as an inverted U-shaped function of the number of options: choice satisfaction increases as the number of options increases, but after a certain point starts to decrease (Reutskaja and Hogarth, 2009; Grant and Schwartz, 2011). Actually, choice overload is a debated issue (see Chernev et al., 2015 for a recent review) and the research has been more recently also focused on understanding why choice overload occurs (Dara and Miller, 2015).

This topic is relevant in any field where individuals' choice depends on the number of possible alternatives: for example, in consumer research, when marketing managers must decide how many types to include in a product line, or when retailers must decide how many brands of the same product to place on shelves, or in public policy, when public policy agents must decide how many and which alternatives to offer to citizens, for example health plans (Szrek, 2017). Having more choices is advantageous initially and variety is positive. Nevertheless, too many choices increases complexity, that is often considered negative for choice because it can drive individuals to delay their decision or even be indecisive.

From a psychological point of view, however, the choice satisfaction is influenced not only by the number of options: instead, it is the perceived variety that plays a crucial role (Broniarczyk et al., 1998; Pizzi and Scarpi, 2016; Szrek, 2017). Unlike the number of options, perceived variety is not easy to measure: individuals perceive easily the number of options, while the actual variety, that is the true level of variety of an assortment (for example, assortment of a product category in a supermarket), may be difficult for individuals to perceive correctly. While the literature in marketing e consumer research fields is more interested in revealing the relationships among perceived costs, perceived benefits and choice satisfaction (both choice process satisfaction and choice outcome satisfaction), the aim of the present study is to apply a novel statistical approach to model perceived variety, in order to better understand the perceptions of individuals about the variety of the possible choice options.

Like other individuals' perceptions, perceived variety is a latent variable (not directly measurable) that is often investigated by means of questionnaires, composed of questions (items) with ordered response categories (ratings). The resulting rating data can be modeled by means of several appropriate statistical methods, which must be able to take into account the categorical ordinal nature of the rating variables (for example, Tutz, 2012; Agresti, 2013). Among them, we resort to a very interesting class of models, called CUB (Combination of Uniform and shifted Binomial), that has been proposed by Piccolo (2003) and D'Elia and Piccolo (2005).

In the CUB framework, the respondents' psychological decision process is interpreted as a combination of two latent components, called *feeling* and *uncertainty*, that express, respectively, the level of agreement with the item being evaluated and the human indecision surrounding any discrete choice. CUB models have been applied in a wide range of fields: among others, consumer research (Arboretti and Bordignon, 2016; Capecchi et al., 2016), marketing (Iannario et al., 2012) and sensometrics (Piccolo and D'Elia, 2008). From a methodological point of view, thanks to a very productive research group headed by Domenico Piccolo and their fruitful collaborations with several researchers, the CUB models have been further developed and extended. A recent review with updated references is given by Piccolo and Simone (2019a) and related discussions (Bartolucci and Pennoni, 2019; Colombi et al., 2019; Agresti and Kateri, 2019; Proietti, 2019; Kenett, 2019; Grilli and Rampichini, 2019; Manisera and Zuccolotto, 2019; Tutz, 2019; Piccolo and Simone, 2019b).

In the field of perceived variety, CUB models have been applied in their Nonlinear version (Brentari et al., 2018), aimed at modelling the possible unequal spacing among response categories (Manisera and Zuccolotto, 2014). Interesting insights were obtained, focused on the perception of the response scale in the respondents' mind. In addition, Manisera and Zuccolotto (2020) proposed a novel mixture model to fit data coming from multi-rating semantic differential scales. In studies about perceived variety, such kind of data arise, for example, when respondents are asked to evaluate their perceived variety by answering a question with possible responses given on a multi-rating scale with end-points labelled as "I had too little variety" and "I had too much variety", with mid-point labelled as "I had the right amount of variety".

The aim of this paper is to use CUB models to investigate how perceived variety impacts, together with some individuals' personal characteristics, on the choice outcome satisfaction. The study will be conducted using the data coming from an experiment with prescription drug plans and available as supplementary material of Szrek (2017), downloadable from the website of the journal Judgment and Decision Making. The added value of the present investigation, respect to traditional studies, is related to the possibility (i) to measure, by CUB models, not only the respondents' level of the latent variable under analysis (perceived variety) but also the respondents' uncertainty surrounding their position on the latent trait, also in relation with the available number of options, and (ii) to investigate if and how these two measures, eventually together with personal characteristics, influence the choice outcome satisfaction. In other words, CUB makes it possible to measure how much variety respondents perceive (their feeling towards variety) and how much uncertain they are about it (their uncertainty), in relation with the number of choices they face. In a second step, we will use the two measures of perceived variety (feeling and uncertainty) and some individual features as covariates in another CUB model, in order to study their impact on the choice satisfaction.

The novelty of the proposed application is then the use of this two-step approach based on the estimated latent feeling and uncertainty to measure the impact of perceived variety on the choice satisfaction. In addition, another aspect of novelty refers to the exploitation of the R FastCUB package (Simone, 2020b,a) to select the best set of covariates to include in the final model.

The paper is organized as follows. Section 2, after a brief recall of the characteristics of the experimental study and the participants in Szrek (2017), focuses on the description of CUB models. Section 3 reports the obtained results and Section 4 concludes the paper and highlights the advantages of using CUB models in the analysis of latent variables measured by rating scales.

2 Method

2.1 Participants

Data refer to an experiment with prescription drug plans (Szrek, 2017). In countries like the United States, where the public health system does not provide health care to the entire population, most citizens are covered by a combination of private insurance and various federal and state programs. Citizens usually enroll in health plans, which offer a list of drugs and medical services, equipment and supplies to their members, who choose the best offer on the basis of their health status. The list of covered drugs is commonly called (prescription) drug plan.

Data come from a survey involving 545 individuals, who have been randomized to a set of 2, 5, 10 or 16 drug plan options and asked to select one plan from the set shown to them. In addition, they were asked to rate their perceived variety, answering the question "Do you think that the selection should have included a greater variety of plans?", with responses on a 1-7 scale (1=I had too little variety; 4=I had the right amount of variety; 7=I had too much variety), besides some other information about outcome and process satisfaction, perceived benefits and costs, and individual characteristics. Tables 1 and 2 report the main characteristics of the individuals involved in the study and the frequency distributions of their answers to some questions on subjective perceptions and attitudes. Details are in Szrek (2017) while additional descriptive statistics are available upon request to the Authors.

2.2 Models

CUB models (Piccolo, 2003; D'Elia and Piccolo, 2005) assume that the response of each individual to a given item with a response scale of m ordered categories is the combination of a *feeling* attitude (agreement) towards the item and an intrinsic *uncertainty* component surrounding the discrete choice. The feeling attitude accounts for the subjective feeling towards the object being evaluated and considers any reasoned assessment, as well as the set of emotions, sentiments, and perceptions logically connected with the object. The uncertainty component accounts for indecision in the process of selecting the ordinal response and considers several elements not logically connected with the item being evaluated, such as the unconscious willingness to delight the interviewer, the diffi-

Table 1: Summary statistics concerning the characteristics of the 545 respondents of the survey

Gender
Male (51%), Female (49%)
Education
Lower than high school diploma (10%), High school diploma (32%),
Some college (26%), Bachelor's degree or higher (32%)
Income (in .000 US dollars)
Lower than 7.5 (2%), 7.5 $\vdash$ 10 (1%), 10 $\vdash$12.5 (2%),
12.5 $\vdash$15 (2%), 15 $\vdash$20 (5%), 20 $\vdash$ 25 (8%),
25 $\vdash$ 30 (7%), 30 $\vdash$35 (8%), 35 $\vdash$ 40 (7%),
40 $\vdash$ 50 (13%), 50 $\vdash$ 60 (12%), 60 $\vdash$ 75 (12%),
75 $\vdash$ 85 (6%), 85 $\vdash$ 100 (5%), 100 $\vdash$125 (6%),
125 $\vdash$ 150 (1%), 150 $\vdash$ 175 (1%), 175 or higher (2%),
Age (in years)
min = 65, max = 89, mean =70.72 (standard deviation = 4.9), median = 70
Number of plans (Respondents randomized to 2, 5, 10, 16 drug plans)
2 (23.5%), 5 (27%), 10 (26%), 16 (23.5%)

culty in evaluation using limited information, lack of self-confidence, laziness, boredom, etc.

CUB fits rating data by means of a mixture of two distributions aimed to model the feeling and the uncertainty component. Formally, the observed rating r is the realization of a discrete random variable R defined as a mixture of a shifted Binomial and a Uniform distributions as follows:

$$Pr(R = r; \boldsymbol{\theta}) = \pi b_r(\xi) + (1 - \pi)Pr(U_m = r)$$

with $\boldsymbol{\theta} = (\pi, \xi)'$, with $\pi \in (0, 1]$, $\xi \in [0, 1]$, $r = 1, 2, \ldots, m$, and where $b_r(\xi)$ is the probability distribution of a shifted Binomial random variable, and $P(U_m = r) = 1/m$ is the probability distribution of a discrete Uniform random variable defined over the support $\{1, 2, \ldots, m\}$, for a given m; identifiability of CUB models is guaranteed for $m > 3$ (Iannario, 2010).

The shifted Binomial random variable is introduced to measure the feeling component and $1 - \xi$ is called the feeling parameter; a *shifted* Binomial corresponds to a Binomial random variable with trial parameter $m-1$, defined over the *shifted* domain $\{1, 2, \cdots, m\}$ instead of $\{0, 1, \cdots, m - 1\}$. The discrete Uniform random variable is used to measure the uncertainty component and $1 - \pi$ is called the uncertainty parameter. It is worth noting that uncertainty is not randomness (that is the stochastic component related to sampling selection or measurement errors): rather, it refers to the respondents' subjective

Table 2: Summary statistics concerning subjective perceptions and attitudes of the 545 respondents of the survey

Health status (self-reported)
In general, would you say your health is...
Excellent (10%), Very good (38%), Good (37%), Fair (14%), Poor (2%)
Desire for choice
Would you prefer to choose your own drug plan from a variety of plans or would you rather be automatically enrolled into a single standard plan?
Scale 1-7: (1) I would prefer to not choose my own plan, (4) I am indifferent, (7) I would prefer to choose my own plan
1 (4%), 2 (1%), 3 (5%), 4 (14%), 5 (9%), 6 (23%), 7 (44%)
Perceived variety
Do you think that the selection should have included a greater variety of plans?
Scale 1-7: (1) I had too little variety, (4) I had the right amount of variety, (7) I had too much variety
1 (11%), 2 (8%), 3 (13%), 4 (32%), 5 (13%), 6 (11%), 7 (14%)
Choice outcome satisfaction
How much do you like the plan you decided to pick?
Scale 1-7: (1) Not at all, (7) Extremely
1 (4%), 2 (6%), 3 (15%), 4 (30%), 5 (24%), 6 (18%), 7 (3%)
Choice process satisfaction
How much did you enjoy making the choice?
Scale 1-7: (1) Not at all, (7) Extremely
1 (23%), 2 (17%), 3 (19%), 4 (22%), 5 (8%), 6 (8%), 7 (3%)
Perceived benefits
How different/similar is the plan you chose from the 'ideal' plan you would like to purchase for yourself?
Scale 1-7: (1) Ideal plan would be very different from the plan I chose now, (7) The plan I chose now is the ideal one
1 (10%), 2 (9%), 3 (13%), 4 (22%), 5 (25%), 6 (17%), 7 (4%)
Perceived costs
Did you find it difficult to make your decision?
Scale 1-7: (1) Not at all, (7) Extremely
1 (7%), 2 (8%), 3 (10%), 4 (14%), 5 (21%), 6 (23%), 7 (17%)

indecision. Each CUB model may be represented as a point in the parameter space (that is the unit square), because of the one-to-one relationship between the CUB random variable and the parameter vector $\boldsymbol{\theta} = (\pi, \xi)'$. This representation allows nice visualizations and easy interpretation of the models.

A possible extension of the standard model is the CUB model with "shelter" effect (Iannario, 2012), useful when a proportion of respondents tend to choose one response category as a refuge, for example to avoid a more demanding choice on the response scale. Typical shelter categories are middle indifference categories like "nor satisfied nor dissatisfied". A CUB model with shelter effect at category c can be written as follows:

$$Pr(R = r; \boldsymbol{\theta}) = \pi_1 b_r(\xi) + \pi_2 Pr(U_m = r) + (1 - \pi_1 - \pi_2) D_r^{(c)}$$

with $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi)'$, $r = 1, 2, \ldots, m$ and where $D_r^{(c)}$ is an indicator variable equal to 1 for $R = c$ and 0 otherwise.

The CUB package has been developed in the statistical environment R (Iannario et al., 2018) and is available on CRAN for free. Codes to replicate the proposed analyses are available upon request to the Authors.

Inference for the class of CUB models is achieved by using Maximum Likelihood (ML) methods (Piccolo, 2006). Several fitting measures specific to CUB models (Iannario, 2009) can be considered, based on deviance, BIC and a Dissimilarity index, which compares the relative observed frequencies with the fitted probabilities; it is a normalized index, so it can be expressed as a percentage and when it equals zero, a perfect fit is achieved.

One important extension of CUB models is the introduction of covariates (Piccolo, 2006; Iannario and Piccolo, 2016). The main model is still generated by the combination of feeling and uncertainty components, which can be expressed in function of individual characteristics thanks to the introduction of covariates. For example, it is possible to investigate if and how the perceived variety in a choice process depends on education, income, gender, ... of individuals. Formally, let R_i be the random variable describing the ordinal response of individual i in the random sample $(R_1, R_2, \ldots, R_n)'$, $\boldsymbol{T}$ the matrix of the covariates and $\boldsymbol{t}_i$ the row vector containing the values of the selected covariates for individual i. Then, the CUB mixture is defined for each respondent by:

$$Pr(R_i = r \mid \boldsymbol{t}_i, \boldsymbol{\theta}) = \pi_i \, b_r(\xi_i) + (1 - \pi_i) \, 1/m \tag{1}$$

for $r = 1, 2, \ldots, m$. If we consider the two matrices $\boldsymbol{X}$ and $\boldsymbol{W}$ containing some subjects' covariates extracted from $\boldsymbol{T}$ and a logit link used to preserve the mapping between parameters in $[0, 1]$ and covariates, we have:

$$logit(\pi_i) = \boldsymbol{x}_i \boldsymbol{\beta}; \qquad logit(\xi_i) = \boldsymbol{w}_i \boldsymbol{\gamma} \tag{2}$$

for $i = 1, 2, \ldots, n$. Here, $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ represent the covariates extracted from $\boldsymbol{T}$ to specify the relationship of π_i and ξ_i with the corresponding subjects' covariates $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$, respectively. Covariates in $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ may coincide, overlap or be completely different. The parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ is split into the parameters which measure the impact on uncertainty and feeling components, respectively.

As mentioned before, the aim of this paper is to use CUB models to investigate how perceived variety impacts, together with some individuals' personal characteristics, on the choice outcome satisfaction. The analysis is carried out along the following steps:

1a) We firstly model the variable Perceived variety (see Table 2) by fitting a CUB model without covariates (eventually with a shelter effect).

1b) Secondly, starting from the previous CUB model, we add the variable Number of plans (Table 1) as a covariate for both feeling and uncertainty parameters, in order to investigate if the perceived variety (feeling) and the indecision of respondents in declaring their perceptions about variety of options (uncertainty) are significantly affected by the available number of drug plan options.

2) Finally, in order to examine if and how perceived variety (feeling and uncertainty) as well as individuals' characteristics affect the satisfaction about the selection of the plan in the choice process, we fit a CUB model with covariates on the variable Choice outcome satisfaction, by using the new best-subset variable selection procedure implemented in the R FastCUB package (Simone, 2020b,a).

Steps (1a) and (1b) exploit CUB models in order to measure both the respondents' level of perceived variety and the uncertainty around the respondents' positioning on that latent trait, also in relation with the number of options, by means of an appropriate model for ordinal data. Step (2) allows to investigate if and how the two measures (feeling and uncertainty) obtained by CUB, together with personal characteristics, influence the choice outcome satisfaction.

3 Results

According to the analysis steps described in the end of Subsection 2.2, we firstly fit a CUB model without covariates to the variable Perceived variety (step 1a). The Dissimilarity index equals 16.28%, denoting a poor fit of the model (BIC=2082.026) (Piccolo and Simone, 2019a), as also visible in Figure 1 (left), where the observed relative frequencies (dots) and the fitted probabilities (circles) for each of the response categories (1-7, on x-axis) are jointly represented. The lack of fitting of the estimated model to the observed frequencies is mainly due to the very high frequency of category 4, suggesting a possible shelter effect, which will be investigated later.

In addition, the previous model is also not satisfactory because it is not fair to evaluate the perceived variety for the whole sample of respondents, since we expect a different level of perceived variety according to the number of plans among which respondents had to choose. One possibility is to fit four CUB models to the four groups of individuals differing in the number of plans they have been offered. Results are shown in Figure 2, where for each group of individuals the CUB fitted probabilities and the observed frequencies of each response category are displayed. Dissimilarity indices range from 13.86% (2 plans) to 25.29% (10 plans), showing a very poor fit to the data. In all

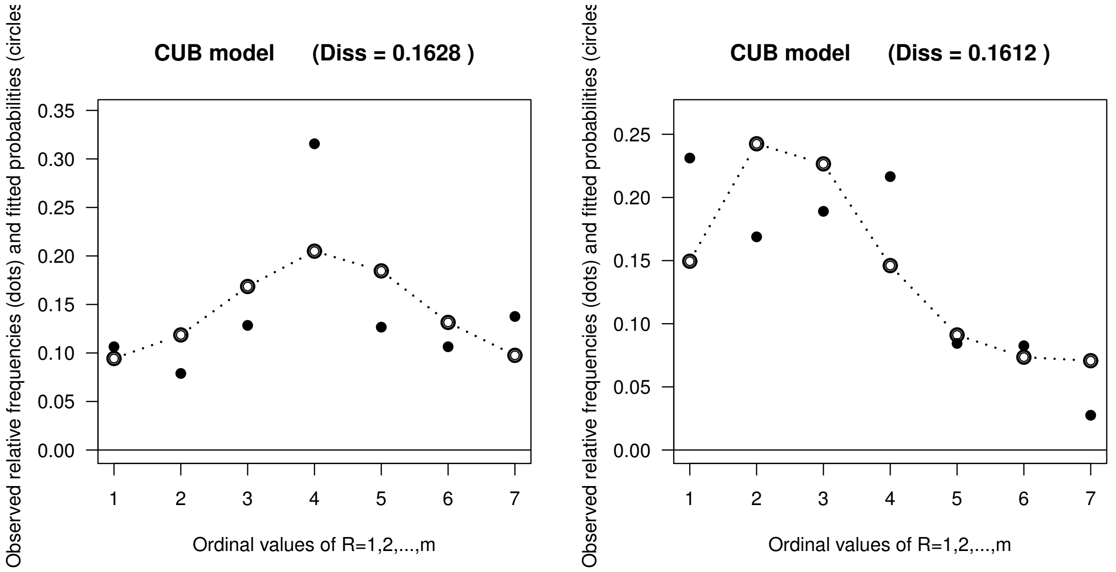

Figure 1: Perceived variety (left) and Choice outcome satisfaction (right): response categories on x-axis, observed relative frequencies (dots) and fitted probabilities (circles) on y-axis obtained with CUB model

.

the four groups, the presence of a shelter category (the mid-point 4) attracting a high percentage of responses reduces the goodness of fit.

A possible improvement can be achieved by fitting a CUB model with shelter, suitable when one response category shows a very high frequency. Here, the middle category 4 in the response scale is very frequently chosen (32% of the 545 respondents, see Table 2). Although the model (with shelter) applied to the whole sample shows a good fit (Dissimilarity index equals 5.05% and BIC is 2022.574), results show that the weight of the Binomial component in the mixture is not statistically different from zero: the model reduces to a very special case of the Generalized CUB (Iannario and Piccolo, 2016), called CUSH (Capecchi and Piccolo, 2017; Capecchi and Iannario, 2016) (Combination of a discrete Uniform random variable with a SHelter effect). We then fitted a CUB model with "shelter" at category 4 conditioning on the number of plans (that is, separately to each subgroup). Results are reported in Figure 3, where for each group of individuals the model fitted probabilities and the observed frequencies of each response category are displayed. Dissimilarity indices range from 4.16% (5 plans) to 9.37% (10 plans), showing a strong improvement on the fit respect the previous model (Figure 2).

Another very nice way to represent the four CUB models fitted for each subgroup is in Figure 4, where each point represents one CUB model in the parameter space (uncertainty, on x-axis and feeling, on y-axis), without (left) and with (right) shelter at category 4 (in the latter case, π_2 is on x-axis). Having in mind that the fit of the CUB models without covariates is not satisfactory, we describe results in the left plot essentially as an example and in comparison with the right plot. The left plot shows that perceived variety (on

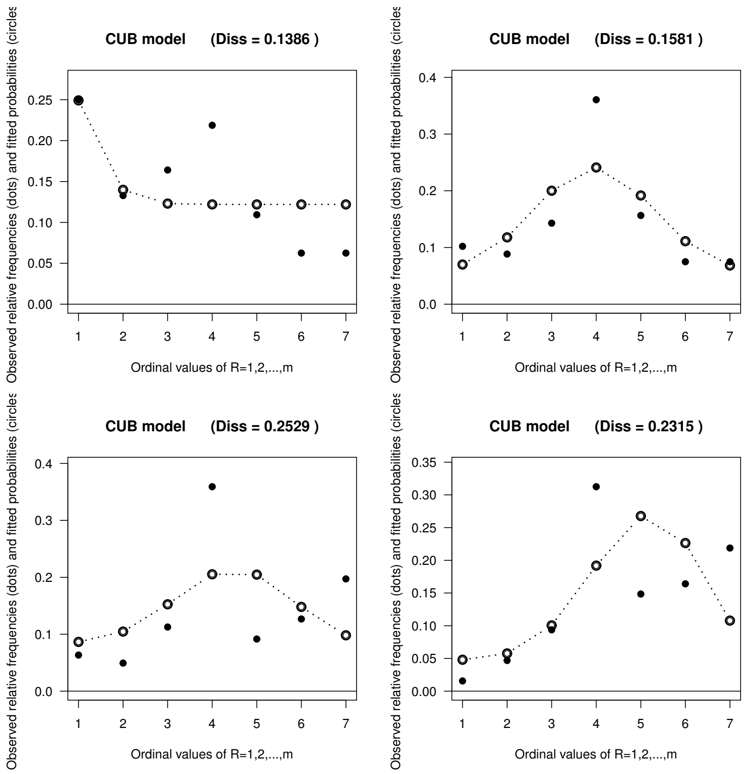

Figure 2: Perceived variety: possible responses on x-axis, observed relative frequencies (dots) and CUB fitted probabilities (circles) on y-axis for the four groups of individuals assigned to 2 (top-left), 5 (top-right), 10 (bottom-left) or 16 (bottom-right) options (prescription drug plans).

y-axis) increases with the number of options, as expected and also highlighted in Szrek (2017). This is substantially confirmed in the right plot, where, however, 10 plans and 16 plans are exchanged between them. The added value of the use of CUB models in place of computing the average of the perceived variety by group (that is not appropriate, moreover, due to the categorical ordinal nature of the variable) is the measure of the respondents' indecision in declaring their perceived variety. Figure 4 (x-axis) shows the different estimates of the uncertainty parameter: in both plots, respondents in the group with 2 plans perceive a very low level of variety, but with a high level of uncertainty. Uncertainty reduces in the other three groups of respondents. Both groups with 10 and 16 plans perceive a high level of variety, but with a different level of uncertainty: in particular, the uncertainty estimate for individuals who have been proposed 16 plans is lower than for individuals with 10 plans: the former agree on their perceived variety much more than the latter.

A more elegant method to study how perceived variety and the uncertainty around it vary according to the number of proposed options to choose consists in including - to the CUB model estimated in step (1a) - the variable Number of plans as a covariate for both

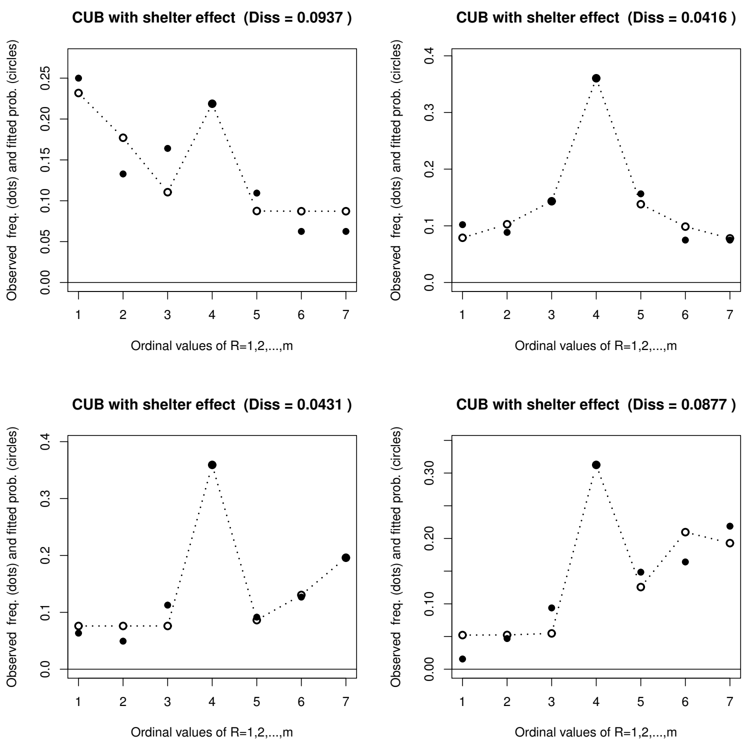

Figure 3: Perceived variety: possible responses on x-axis, observed relative frequencies (dots) and CUB fitted probabilities (circles) on y-axis for the four groups of individuals assigned to 2 (top-left), 5 (top-right), 10 (bottom-left) or 16 (bottom-right) options (prescription drug plans) - CUB model with shelter effect at category 4.

feeling and uncertainty parameters (step 1b). The p-values of the Wald test equal 4.31e-07 and 0.11 for γ_1 and β_1, suggesting that Number of plans is significative only on the feeling parameter: the uncertainty measure is not significantly affected by the number of plans among which to choose. Therefore, the differences among uncertainty estimates noticed in Figure 4 (left) are not statistically significant. Considering the model with shelter effect, the number of plans included as a covariate for the shelter effect does not result significant: it appears that the selection of the middle category as a shelter choice concerns all the respondents, independently on the number of options (plans).

Taking also interpretative reasons into account, results suggest to proceed with the CUB model with Number of plans included as covariate only on the feeling component of the model. This provides a measure of feeling, called “Perceived variety (feeling)” in the following, that differs according to the number of plans and can be used in further analyses as a measure of perceived variety. The related measure of the uncertainty can also be used in further models, but in our case study it does not statistically differ across subjects, as shown before.

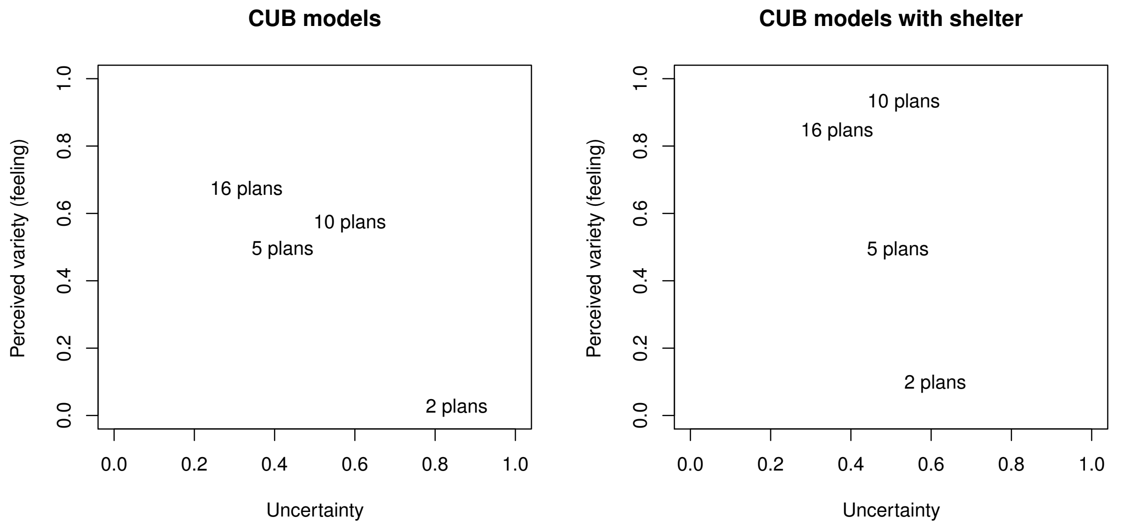

Figure 4: Uncertainty (x-axis) and Perceived variety (y-axis) estimated by CUB for the four groups of individuals assigned to 2, 5, 10 or 16 options (prescription drug plans): CUB model (left) and CUB model with shelter at category 4.

Starting from results obtained in steps (1a) and (1b), step (2) can be performed, with the aim of investigating if and how choice outcome satisfaction is affected by perceived variety (both the level of perceived variety and the uncertainty around the respondent's responses on it) and some individual characteristics. In detail, we fit a CUB model to the variable Choice outcome satisfaction (see Table 2), including the following 7 covariates on both feeling and uncertainty components of the mixture: Perceived variety (feeling) and uncertainty obtained in step (1b); Gender, Education, Income, Age (see Table 1), and (self-reported) Health Status (see Table 2). For simplicity, Education and Health status enter the model as numeric variables, with ordered categories transformed into numeric scores. (For completeness, the right panel in Figure 1 shows the fit of a CUB model without covariates to the variable Choice (outcome) satisfaction).

We selected the best set of covariates to include in the final model by means of a best subset variable selection method. This has been possible thanks to the acceleration of maximum likelihood estimation obtained for the CUB class of mixture models by Simone (2020b,a). This novel approach makes it possible to automatically select the best covariates and then the best model for the data at hand; while this is a common possibility in most of the statistical models (for example, stepwise procedures in linear regression models), in the literature of mixture models for ordinal data, in particular belonging to the CUB class, this was not obvious and the selection of the best model was based in the past on manual detection and somewhat arbitrary choices.

A number of $(2^7-1)^2 = 16.129$ models have been estimated. The best model, selected after a computational time of 271 minutes, is specified by the following model equations

(BIC = 2056.395, standard errors in parentheses):

$$logit(\pi_i) = \underset{(1.68)}{-4.43} + \underset{(0.30)}{0.77}\,\text{Education}_i + \underset{(0.36)}{0.81}\,\text{Health Status}_i$$

$$logit(\xi_i) = \underset{(0.45)}{1.96} + \underset{(0.12)}{0.33}\,\text{Perceived variety (feeling)}_i - \underset{(0.17)}{0.30}\,\text{Health Status}_i$$

Results show that significant covariates for the feeling component, that is the measure of the choice outcome satisfaction, are the Perceived variety (feeling), expressed in function of the number of plans, and the self-reported Health status.

The significant variables affecting the measure of the uncertainty around the choice outcome satisfaction are the self-reported Health status and Education.

Given that the CUB measure (feeling) of the choice outcome satisfaction is $1 - \xi_i$ and the measure of uncertainty is $1 - \pi_i$ (so π_i measures the resoluteness in the responses), and that the levels of Health status are in decreasing order from Excellent to Poor, results can be interpreted as follows. The choice outcome satisfaction increases as the perceived variety decreases; choice outcome satisfaction is higher for people reporting poorer levels of health status. On the other hand, resoluteness in the responses about choice outcome satisfaction is higher for highly-educated persons and for people reporting poorer levels of health.

It is interesting to notice that, as in Szrek (2017), Education does not affect choice outcome satisfaction. However, we are able to show that Education impacts on the uncertainty around the declared choice outcome satisfaction. Perceived variety (feeling), as measured in this study (in function of the number of plans, see step 1b), affects choice outcome satisfaction. This is consistent with results of Model 4 in Szrek (2017). The added value of our proposal is that the use of CUB models and measures of latent constructs derived from CUB (also including uncertainty measures) makes the interpretation of results much more straightforward, because we don't need to include interactions and transformed covariates (for example, squared variables) that are not always easy to understand.

The two-step strategy implemented in this study appears to be even more interesting if we consider that the Number of plans is not significant if tested to explain Choice satisfaction using CUB models, while it is significant for (feeling of) Perceived variety.

4 Concluding remarks

The aim of this paper was to show, starting from an experimental study on perceived variety, how a class of mixture models for ordinal data, called CUB, can be used in the studies of human judgments and decisions, whenever attitudes, perceptions and opinions are measured by means of questionnaires having questions with ordered response categories. The advantages of the CUB model in the analysis of categorical ordinal variables include: an appropriate treatment of the ordinal nature of the data; the possibility to measure both the level of the latent variable under study and the uncertainty surrounding any human choice when forced to choose among a number of response categories; the possibility to investigate if and how individual characteristics, including measures of some latent constructs as well as the associated measures of uncertainty, affect one variable of interest, always taking account of the nominal or ordinal categorical or numerical nature of the variables.

In the proposed analysis, in step (1a) both CUB models and CUB models with shelter have been applied to data. A further analysis could be performed, starting from the models with shelter and linking their results to step (2), in order to investigate the impact of a sort of "structural" shelter effect for perceived variety on the choice satisfaction.

Another possible investigation is to perform a best-subset search in the first step of the procedure, that is for perceived variety without limiting on the number of plans as an explanatory variable (step 1b). Results in terms of feeling and uncertainty as functions of the selected covariates can then be used as explanatory variables in step (2), when the focus is on choice satisfaction. This generalized approach, exploiting the best-subset selection in both steps (1) and (2), can inspire the application of CUB models in several other studies.

The CUB class of models could also be applied to the data at hand to investigate in much detail the relationships of perceived variety with choice and process satisfaction and perceived costs and benefits of the choice. Further and more general studies in the field of judgement and decision making could exploit the advantages of CUB models in order to analyze individuals' latent traits when measured by ordinal categorical variables with appropriate models, able to address the specific nature of the available data.

Acknowledgement

We wish to thank Rosaria Simone for making the R best-subset procedure available and for her useful suggestions.

References

Agresti, A. (2013). *Categorical Data Analysis*. Wiley, New York, 3^{rd} edition.

Agresti, A. and Kateri, M. (2019). The class of CUB models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods & Applications*, 28(3):445–449.

Arboretti, R. and Bordignon, P. (2016). Consumer preferences in food packaging: CUB models and conjoint analysis. *British Food Journal*, 118(3):527–540.

Bartolucci, F. and Pennoni, F. (2019). Comment on: The class of CUB models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods & Applications*, 28(3):437–439.

Brentari, E., Manisera, M., and Zuccolotto, P. (2018). Modelling preceived variety in a choice process with Nonlinear CUB. In Capecchi, S., Di Iorio, F., and Simone, R., editors, *Proceedings of the International Conference ASMOD2018*, pages 69–76. Federico II University Press.

Broniarczyk, S. M., Hoyer, W. D., and McAlister, L. (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of marketing research*, XXXV:166–176.

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

Capecchi, S. and Iannario, M. (2016). Gini heterogeneity index for detecting uncertainty in ordinal data surveys. *Metron*, 74:223–232.

Capecchi, S. and Piccolo, D. (2017). Dealing with heterogeneity in ordinal responses. *Quality & Quantity*, 51(5):2375–2393.

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.

Colombi, R., Giordano, S., and Gottard, A. (2019). Discussion of "the class of CUB models: statistical foundations, inferential issues and empirical evidence". *Statistical Methods & Applications*, 28(3):441–444.

Dara, I. and Miller, E. G. (2015). Competitive forces when choosing from assortments of varying size: How holistic thinking mitigates choice overload. In Diehl, K. and Yoon, C., editors, *Advances in Consumer Research Volume 43*, pages 502–503. Association for Consumer Research.

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

Grant, A. M. and Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, 6(1):61–76.

Grilli, L. and Rampichini, C. (2019). Discussion of 'the class of CUB models: statistical foundations, inferential issues and empirical evidence' by Domenico Piccolo and Rosaria Simone. *Statistical Methods & Applications*, 28(3):459–463.

Hafner, R. J., White, M. P., and Handley, S. J. (2018). The goldilocks placebo effect: Placebo effects are stronger when people select a treatment from an optimal number of choices. *American Journal of Psychology*, 131(2):175–184.

Haynes Graeme, A. (2009). Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology & Marketing*, 26(3):204–212.

Iannario, M. (2009). Fitting measures for ordinal data models. *Quad Stat*, 11:39–72.

Iannario, M. (2010). On the identifiability of a mixture model for ordinal data. *Metron*, LXVIII:87–94.

Iannario, M. (2012). Modelling shelter choices in a class of mixture models for ordinal responses. *Stat Method Appl*, 20:1–22.

Iannario, M., Manisera, M., Piccolo, D., and Zuccolotto, P. (2012). Sensory analysis in the food industry as a tool for marketing decisions. *Adv Data Anal Classif*, 6:303–321.

Iannario, M. and Piccolo, D. (2016). A generalized framework for modelling ordinal data. *Statistical Methods & Applications*, 25(2):163–189.

Iannario, M., Piccolo, D., and Simone, R. (2018). CUB: A class of mixture models for ordinal data. *R package version 1.1.3, available at http://CRAN.R-project.org/package=CUB.*

Kenett, R. (2019). A review of: The class of CUB models: statistical foundations, inferential issues and empirical evidence by Domenico Piccolo and Rosaria Simone. *Statistical Methods & Applications*, 28(3):457–458.

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

Manisera, M. and Zuccolotto, P. (2019). Discussion of "the class of CUB models: statistical foundations, inferential issues and empirical evidence" by Domenico Piccolo and Rosaria Simone. *Statistical Methods & Applications*, 28(3):465–470.

Manisera, M. and Zuccolotto, P. (2020). A mixture model for ordinal variables measured on semantic differential scales. *Submitted.*

Piccolo, D. (2003). On the moments of a mixture of Uniform and shifted Binomial random variables. *Quad Stat*, 5:85–104.

Piccolo, D. (2006). Observed information matrix for MUB models. *Quad Stat*, 8:33–78.

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

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

Piccolo, D. and Simone, R. (2019b). Rejoinder to the discussion of "the class of CUB models: statistical foundations, inferential issues and empirical evidence". *Statistical Methods & Applications*, 28(3):477–493.

Pizzi, G. and Scarpi, D. (2016). The effect of shelf layout on satisfaction and perceived assortment size: An empirical assessment. *Journal of Retailing and Consumer Services*, 28:67–77.

Proietti, T. (2019). Discussion of The class of CUB models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods & Applications*, 28(3):451–456.

Reutskaja, E. and Hogarth, R. M. (2009). Satisfaction in choice as a function of the number of alternatives: When "goods satiate". *Psychology & Marketing*, 26(3):197–203.

Scheibehenne, B., Greifeneder, R., and Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3):409–425.

Schwartz, B. (2004). *The paradox of choice: Why more is less.* Harper Collins, New York.

Simone, R. (2020a). An accelerated EM algorithm for mixture models with uncertainty for rating data. *Computational Statistics.* DOI: 10.1016/j.ecosta.2019.04.002, Online first.

Simone, R. (2020b). FastCUB: Fast EM and best-subset selection for CUB models for rating data. *R package, available on CRAN at https://cran.r-project.org/web/packages/FastCUB/index.html.*

Szrek, H. (2017). How the number of options and perceived variety influence choice satisfaction: An experiment with prescription drug plans. *Judgment and Decision Making*, 12(1):42.

Tutz, G. (2012). *Regression for categorical data.* Cambridge University Press, Cambridge.

Tutz, G. (2019). Comments on the class of CUB models: statistical foundations, inferential issues and empirical evidence by D. Piccolo and R. Simone. *Statistical Methods & Applications*, 28(3):471–475.