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/i20705948v8n3p312

On the interpretation of the uncertainty parameter in CUB models

By Golia

Published: 20 December 2015

Electronic Journal of Applied Statistical Analysis
Vol. 08, Issue 03, December 2015, 312-328
DOI: 10.1285/i20705948v8n3p312

On the interpretation of the uncertainty parameter in CUB models

Silvia Golia*

Department of Economics and Management, University of Brescia
C.da S.Chiara 50, 25122 Brescia, Italy

Published: 20 December 2015

The rationale for the CUB models is based on the fact that the response of a person to an item is a weighted combination of two factors: a subjective agreement towards the item and some intrinsic fuzziness in the final response. These two components are parametrized as a Shifted Binomial and a discrete Uniform random variable respectively. The final model is a mixture of these two random variables with weights π and $(1 - \pi)$. The last quantity is currently interprets as a measure of the uncertainty that accompanies the choice of a response category made by the subjects who form the population under study. This paper wants to be a warning for this interpretation, when the data of interest derive from a questionnaire designed to measure an overall latent trait. Through a simulation study one will show that in this context there are situations in which a high value of $(1 - \pi)$ is connected to the distribution of the overall latent trait among the subjects in the population and to the facet of the overall latent trait that the item wants to represent and not to an high level of fuzziness in the final response.

keywords: CUB model, uncertainty parameter, Partial Credit Model.

1 Introduction

In the last decades the study of models able to treat ordinal data has acquired an increasing importance. Generally, such kind of data comes from surveys in which respondents are asked to answer to questions using a Likert scale. A survey is in general composed by a number of questionnaires created to investigate different aspects of the main object

*Corresponding author: silvia.golia@unibs.it

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

of interest of the study for which the survey has been made. Some questionnaires are designed to measure specific latent traits and the corresponding items are strictly connected to the latent traits of reference identified from now on as the overall latent trait to distinguish it from its components associated to the features that describe it (for example, the passenger satisfaction for punctuality of the urban public transport is a feature of the overall passenger satisfaction for the urban public transport). In this context, that is, when there is a questionnaire associate to an overall latent trait to be measured, it is possible to hypothesize that the level of overall latent trait owned by a subject, the type of item considered and an amount of fuzziness, are the main elements that determine the choice of a specific response category by a subject. There are different paradigms that take into account this kind of psychological mechanism in responding to items, and among them the CUB model is the one considered in this paper. The rationale for CUB models (Piccolo, 2003; D'Elia and Piccolo, 2005) is based on the fact that the response of a person to an item is a weighted combination of two factors, that is a subjective agreement or feeling towards the item and some intrinsic fuzziness or uncertainty in the final response. The first factor is related to awareness of the topic, previous experience, group membership, and so on, whereas the second component results from different facts such as the amount of time available to respond, the use of limited set of information, partial understanding, laziness and so on. In the mathematical formulation of CUB models the weights are denoted by π and $(1 - \pi)$ and the standard interpretation of $(1 - \pi)$ states that it is a measure of the uncertainty that is present in the final response (Iannario and Piccolo, 2011). As will be shown in Section 2, the CUB model proposed by Piccolo and co-authors provides the probability of response to an item and the parameters involved in this probability are related only to the item, as a consequence the rationale for the CUB model explained above is applied to the population as a single entity and not to the single persons in the population. In contexts like the one considered in this paper and recalled at the beginning of this introduction, it is hard to think that the value assumed by $(1-\pi)$ can be interpret as a measure of uncertainty for the subjects in the population under study; an high value of $(1 - \pi)$ should mean that all the subjects belonging to that population are uncertain in the choice of their right answer to the item, and this is unrealistic. The aim of this paper is to show that in contexts in which it is reasonable to assume that there is a latent variable which determines the answers of the subjects to an item, there are situations in which a high value of $(1 - \pi)$ is connected to the distribution of the overall latent trait among the subjects in the population and to the facet of the latent trait that the item wants to represent. Therefore, this paper wants to be a warning for the current interpretation of $(1 - \pi)$ as a measure of the uncertainty that accompanies the choice of a response category made by the subjects who form the population under study.

This paper deals with the original (global) CUB model, for which no covariates are considered. Piccolo (2006) and Piccolo and D'Elia (2008) extended the CUB model in order to include in the model the information deriving from covariates describing subjects' characteristics. The extended model relates the parameters π and ξ to the covariates; in this way, it is possible, in presence of at least one continuous covariate, to have a measure of the uncertainty $(1 - \pi_n)$ for each subject n. It has to be noted that $(1-\pi_n)$ represents a measure of the uncertainty of any individual with the personal characteristics recorded by the considered covariates and owned by the subject n. In this context, the current interpretation of the parameter $(1-\pi)$ is convincing because it is related to a single subject.

In order to achieve the aim declared previously, a simulation study will be performed. Artificial datasets will be simulated according to the Partial Credit Model (PCM) (Masters, 1982), which is a model belonging to the item response theory (IRT) approach to ordinal data. The models belonging to this approach, give a functional form that relates the probability of a person responding to an item in a specific way to the position of that person on the trait that the item is measuring. As will be shown in Section 2, PCM allows one to generate, for each person in a population, his/her response to an item which depends on the level of overall latent trait owned by the subject and the feature of the latent trait that the item wants to represent. Moreover, there will be shown that the findings derived from the simulation study can find a counterpart in real applications like the ones described in the second part of the paper.

The rest of the paper is organized as follows. Section 2 contains an introduction to CUB and Rasch modeling whereas Section 3 describes the simulation design and shows the obtained results. Section 4 reports the analysis of three real datasets, the first is related to the flash Eurobarometer survey on rail and urban transport passenger satisfaction whereas the last twos are part of the survey on nurse-patient relationship. Conclusions are presented in Section 5.

2 CUB and Partial Credit models

This section contains the formal definition of the models used in the paper. With respect of PCM, this is one of the models for polytomous data that belong to the IRT approach. IRT models can be distinguished into two classes, in the first there are the Rasch-type models, such as the PCM, that attempt to conform to fundamental measurement theory, whereas in the other class there are the models that do not conform to that theory, such as the Graded Response Model or the Generalized Partial Credit Model (Ostini and Nering, 2006). The choice to use, as a data generating model, an IRT model aims at reproducing the context of the paper; IRT models are be able to make explicit the main elements that determine the choice of a specific response category by a given subject, that is the level of overall latent trait owned by the subject and the type of item considered. So, PCM is used in the paper as an instrumental model, able to generate data for which it is known the level of the overall latent trait of each individual as well as some item characteristics.

The CUB model, which is an acronym for Combination of discrete Uniform and Shifted Binomial random variables, hypothesizes that the choice of a response category is determined by a mixture of agreement or feeling towards the item and fuzziness or uncertainty. The agreement or feeling towards the item is captured by a Shifted Binomial distribution, the second component by a discrete Uniform distribution across response categories. These two components parameterized in such manner allow CUB models to be extremely flexible for interpreting the different shapes of ordinal data distributions. Therefore, the CUB model expresses the response to the item i with $m+1$ response categories ($x = 1, 2, \cdots, m+1$), as a realization of a random variable X_i with the following probability distribution, coming from the mixture of Shifted Binomial and discrete Uniform distributions:

$$P(X_i = x) = p_{ix} = \pi_i \binom{m}{x-1} \xi_i^{(m+1)-x} (1-\xi_i)^{x-1} + (1-\pi_i)\frac{1}{m+1} \tag{1}$$

with $\pi_i \in (0, 1]$ and $\xi_i \in [0, 1]$. The parameters π_i and $(1-\pi_i)$ are the weights of the mixture, where $(1-\pi_i)$ is a measure of the uncertainty present in the final response to item i, whereas ξ_i is one of the parameters of Shifted Binomial distribution and $(1-\xi_i)$ is a measure of the agreement or feeling towards the item i. Iannario (2010) demonstrated that CUB models are identifiable for any $m > 2$.

The two parameters π and ξ play different roles in determining the shape of the response probability. The parameter ξ is strongly influenced by the skewness of responses; in fact it increases when subjects choose mostly the low categories, and vice-versa. The parameter π, adding dispersion to the Shifted Binomial distribution, increases the frequencies of each category, modifying the heterogeneity of the distribution. Iannario (2012) has demonstrated that the normalized Gini heterogeneity index is inversely related to π and increases with uncertainty, measured by $(1-\pi)$. In fact, defined $G = (1 - \sum_{x=1}^{k} p_x^2)k/(k-1)$ as the normalized Gini heterogeneity index, where p_x represents the probability distribution of any discrete random variable with k possible outcomes, for the CUB model G is given by $G_{CUB} = 1 - \pi^2(1 - G_{SB})$, where G_{SB} is the normalized Gini index for the Shifted Binomial random variable. So, in this contest $(1-\pi)$ can be qualified as a measure of heterogeneity.

The PCM belongs to the family of Rasch models, which is a family of measurement models that convert raw scores into linear and reproducible measurements. The distinguishing characteristics are: separable person and item parameters, sufficient statistics for the parameters and conjoint additivity. Moreover, models belonging to this family require unidimensionality and local independence. If the data fit the model, then the produced measures are objective and expressed in logits (Wright and Masters, 1982). Following the PCM, given an item i with $m+1$ response categories ($x = 0, 1, \cdots, m$), the probability of the subject n with level of overall latent trait or ability β_n to respond in category x is given by:

$$P(X_{ni} = x) = p_{nix} = \frac{exp\left\{\sum_{j=0}^{x}(\beta_n - \delta_{ij})\right\}}{\sum_{s=0}^{m} exp\left\{\sum_{j=0}^{s}(\beta_n - \delta_{ij})\right\}} \tag{2}$$

where δ_{ij} is a parameter associated with the transition between response category $j-1$ and j and it is equal to the point of equal probability of categories $j-1$ and j. δ_{ij} can be decompose into two components $\delta_{ij} = \delta_i + \tau_{ij}$, $j = 1, 2, \ldots, m$, where δ_i is the mean difficulty of item i and τ_{ij} is called threshold ($\tau_{i0} \equiv 0$ and $\sum_{j=1}^{m} \tau_{ij} = 0$).

For a given value of m and given values of the item parameters $\delta_{i1}, \delta_{i2}, \ldots, \delta_{im}$, from (2) it is possible to draw the Item Characteristic Curves (ICCs), which are graphical

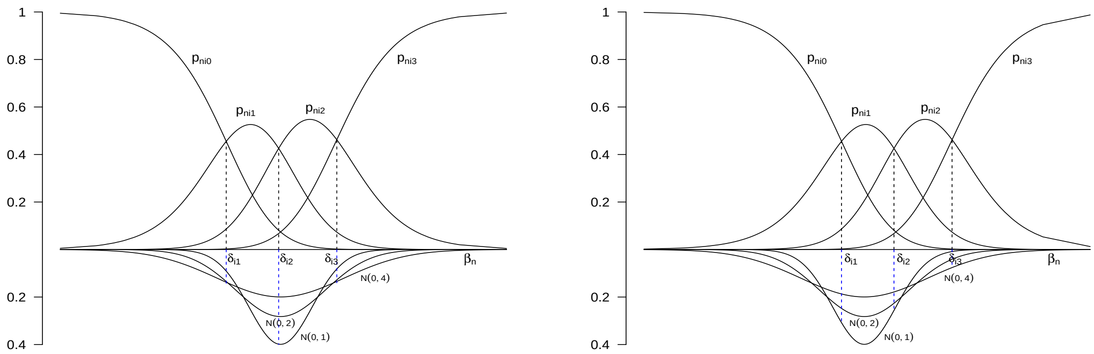

Figure 1: ICCs for an item with m=3 and on the left: $\delta_{i1} = -1.72, \delta_{i2} = -0.06, \delta_{i3} = 1.78$, on the right: $\delta_{i1} = -0.72, \delta_{i2} = 0.94, \delta_{i3} = 2.78$

functions that represent the probabilities p_{nix} as a function of the subjects' ability β_n. Figure 1 shows an example of these curves for two items with 4 response categories ($m = 3$), the first one with parameters $\delta_{i1} = -1.72, \delta_{i2} = -0.06, \delta_{i3} = 1.78$ that imply $\delta_i = 0, \tau_{i1} = -1.72, \tau_{i2} = -0.06, \tau_{i3} = 1.78$ (left panel) and the second one with parameters $\delta_{i1} = -0.72, \delta_{i2} = 0.94, \delta_{i3} = 2.78$ that imply $\delta_i = 1, \tau_{i1} = -1.72, \tau_{i2} = -0.06, \tau_{i3} = 1.78$ (right panel). In both cases the three parameters δ_{i1}, δ_{i2} and δ_{i3} delimit the regions of most probable response to the item i, so once the person's level of latent trait β is fixed, it is possible to read off her/his most probable response to that item. For example, if β_n has value between δ_{i2} and δ_{i3}, the most probable response for the subject n to the item i is $x = 2$. Figure 1 displays also three different distributions for the latent trait β (three normal distributions with zero mean and variance equal to one, two and four) and it shows that the expected frequency of membership to each region of most probable response varies with the subjects' distribution. For example, the response frequency for the last category, interpreted as the number of subjects for whom the last response category is the most probable, expected from a population distributed as N(0,1), is lower than the one expected from a population distributed as N(0,4), and this can be seen regardless of the set of δ_{ij} considered. So, the higher the variance of β, the more similar to each other are these expected response frequencies. The two panels of Figure 1 describe two different situations; on the left panel the mean difficulty is equal to the mean ability ($\delta_i = 0$ and $E(\beta) = 0$), whereas on the right panel the mean difficulty is bigger than the mean ability ($\delta_i = 1$ and $E(\beta) = 0$) and the distance between δ_i and $E(\beta)$ is bigger in the right panel ($\delta - E(\beta) = 1$). It is possible to observe, following the same paradigm used previously, that, considering the same ability distribution, the expected response frequencies for each response category are more similar to each other when the distance between δ_i and $E(\beta)$ is smaller. Summarizing what observed above, the distribution of the responses is more heterogeneous if the ability distribution has higher variance and if the expected ability is closer to the mean difficulty δ_i.

3 Simulation study and results

This section contains the description of the simulation design used in the paper and the results obtained. One sample of 1000 subjects was drawn from different distributions that are three normal distributions with zero mean and variance respectively equal to one, two and four and one Student t distribution with six degrees of freedom ($variance = 1.5$ and $Kurtosis = 5$)[1]. These abilities were referred to as the levels of the overall latent trait or abilities β owned by the subjects. Thirty one difficulty parameters δ, ranging between -1.5 to 1.5, were considered, with values set in the following way: $\delta_i = -1.5 + 0.1 * i$, with $i = 0, 1, \ldots, 30$. They are able to reproduce the variability in responses' distribution observed in the real word. Four, five and six-level response scales were considered and the set of the corresponding threshold parameters τ_j and the size of the interval between the first and the last threshold were displayed in tables 1, 2 and 3. The sets of thresholds values used in this paper were inspired by thresholds sets observed in real data and they were chosen so that they were symmetric (for example, -1.72, -0.06, 1.78) as well as asymmetric (for example -1.87, 0.74, 1.13 or -1.47, -1.06, 2.53) with respect to the central threshold and with increasing width of thresholds interval.

Table 1: Thresholds sets for a four-level response scale and the corresponding sizes of thresholds interval.

	τ_1	τ_2	τ_3	Size		τ_1	τ_2	τ_3	Size
C1	-1.01	-0.48	1.49	2.5	C2	-1.87	0.74	1.13	3.0
C3	-1.72	-0.06	1.78	3.5	C4	-1.47	-1.06	2.53	4.0
C5	-1.94	-0.54	2.52	4.5	C6	-2.17	-0.66	2.83	5.0

For each combination of mean difficulty δ_i, thresholds $\{\tau_{ij}\}_{j=1}^m$ and subject's ability β_n, data were sampled from the distribution defined in (2) in order to simulate the responses that were been used in the analysis performed with the CUB model. The data simulation was carried out using the software R 3.1.1 (R Core Team, 2014).

For each combination of difficulty δ, thresholds set $\{\tau_1, \ldots, \tau_m\}$ and ability distribution, 100 data sets were simulated so that 100 sets of estimated uncertainty measures $(1-\pi)$ were computed and their mean value was used in the analysis. The estimated

[1] There were considered also six skew-normal distributions with skewness 0.576 or 0.851, zero mean and variance equal to one, two or four, as for the normal case. The behavior of $(1-\pi)$ did not significantly differ from what observed when the distribution of the abilities was the normal one with the same variance.

Table 2: Thresholds sets for a five-level response scale and the corresponding sizes of thresholds interval.

	τ_1	τ_2	τ_3	τ_4	Size		τ_1	τ_2	τ_3	τ_4	Size
C1	-1.01	-0.58	0.10	1.49	2.5	C2	-1.13	-0.94	0.20	1.87	3.0
C3	-1.78	-0.56	0.62	1.72	3.5	C4	-2.53	-0.29	1.35	1.47	4.0
C5	-1.88	-1.18	0.44	2.62	4.5	C6	-2.17	-1.26	0.60	2.83	5.0

Table 3: Thresholds sets for a six-level response scale and the corresponding sizes of thresholds interval.

	τ_1	τ_2	τ_3	τ_4	τ_5	Size		τ_1	τ_2	τ_3	τ_4	τ_5	Size
C1	-1.71	-0.58	-0.10	1.10	1.29	3.0	C2	-1.78	-0.57	-0.10	0.73	1.72	3.5
C3	-1.47	-1.35	-0.70	0.99	2.53	4.0	C4	-2.30	-0.83	0.03	0.90	2.20	4.5
C5	-3.10	-0.76	0.60	1.36	1.90	5.0	C6	-2.85	-0.57	-0.12	0.89	2.65	5.5

parameters of CUB model were obtained by Maximum Likelihood exploiting the EM algorithm implemented in R by Iannario and Piccolo (2014).

Figures 2, 3 and 4 show the average value of the estimated uncertainty $(1 - \pi)$ as function of the distance between the mean difficulty δ and the expected value of the ability distribution $E(\beta)$, the ability distribution (black empty circle for N(0,1), red full circle for N(0,2), blue full diamond for N(0,4) and green empty diamond for t_6) and the thresholds set, when the number of allowed response categories is four (Figure 2), five (Figure 3) or six (Figure 4).

In all the cases, it is evident a parabolic relationship between the estimated uncertainty measure $(1 - \pi)$ and the distance $\delta - E(\beta)$. The parabolas are concave down and the steepness of the parabolas is directly proportional to the variance of the ability distribution and inversely proportional to the extent of the thresholds set. When few response categories are available, the impact of the increasing of the size of the interval between the first and last threshold on the uncertainty measure $(1 - \pi)$ is more evident than in the case of more available response categories, regardless of the ability distribution.

In the case of a four-response scale, width of the thresholds interval equal to 2.5 and high variability in the ability distribution, the uncertainty measure $(1 - \pi)$ vary from a medium to an high level for most part of the items considered, as shown in Figure 2. This behavior shows up again weakened when the size of the thresholds interval is equal to 3 and remains evident only if the ability distribution is N(0,4) and until the width of the thresholds interval is equal to 4.

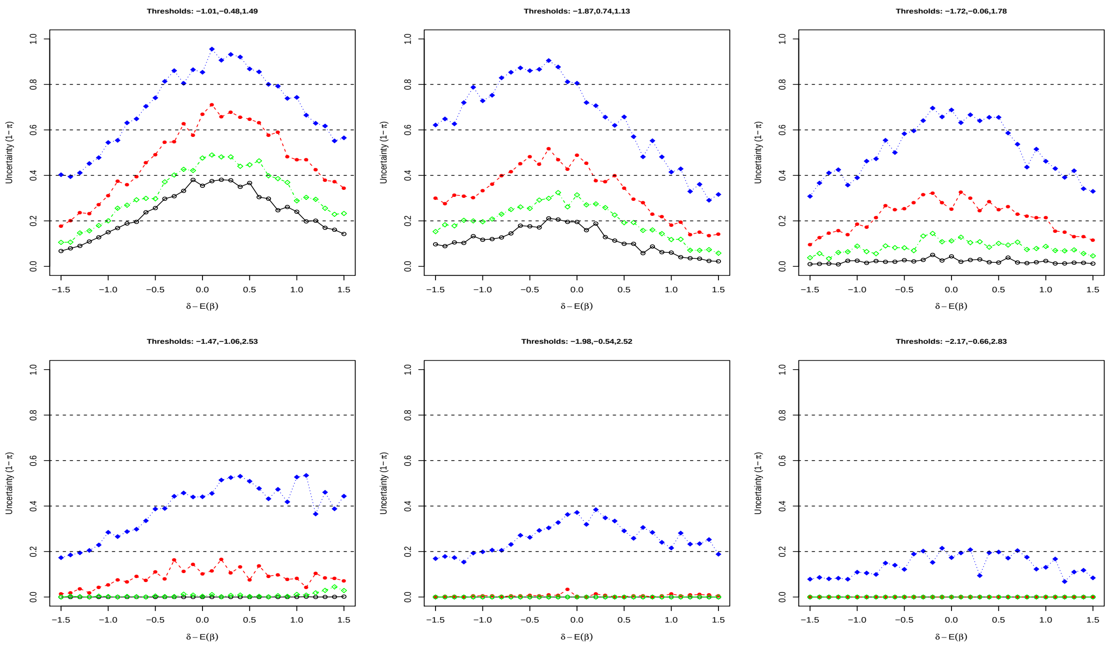

Figure 2: Four response categories. Estimated uncertainty $(1-\pi)$ as function of $\delta - E(\beta)$, thresholds set and ability distribution: N(0,1) (black empty circle), N(0,2) (red full circle), N(0,4) (blue full diamond) and t_6 (green empty diamond)

When a five-level response scale is considered and the size of the thresholds interval is equal to 2.5, $(1-\pi)$ assumes medium-high values for most part of the items considered regardless of the variability of the ability distribution, as shown in Figure 3. With the increasing of the width of the thresholds interval, this behavior remains evident in the presence of the N(0,4) ability distribution and weakens if the ability distribution is N(0,2).

In the case of a six-level response scale, the behavior observed when the response categories were four or five shows up more strongly, as seen in Figure 4, highlighting the presence of many cases in which the uncertainty measure $(1-\pi)$ assumes medium or high values.

Therefore, it was shown that under the conditions of this simulation design, there are situations in which the measure of uncertainty assumes medium-high values in absence of a real uncertainty in the responses, given that the data used come from a simulation.

4 Key Studies

In this section three key studies are describe and analyzed. For each of them there is a unique latent trait measured by the items included in the corresponding questionnaire.

A preliminary list-wise deletion was performed in order to delete subjects with missing values in their response record. Then, a Rasch analysis was performed in order to identify

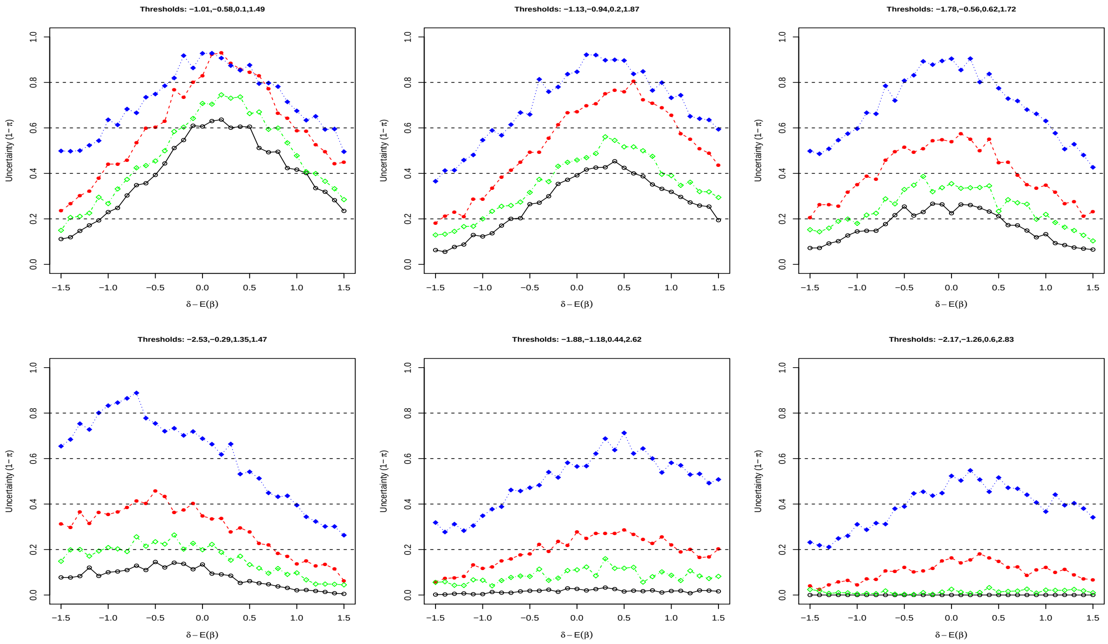

Figure 3: Five response categories. Estimated uncertainty $(1-\pi)$ as function of $\delta - E(\beta)$, thresholds set and ability distribution: N(0,1) (black empty circle), N(0,2) (red full circle), N(0,4) (blue full diamond) and t_6 (green empty diamond)

the appropriate scale and number of categories able to produce an objective measure of the latent trait under study. If data fit the model, the categories and thresholds estimates are ordered; if the categories and thresholds are disordered, merging categories may improve item fit and the overall scale and may reveal the effective number and ordering of categories (Andrich et al., 1997). When the estimated thresholds related to the original Likert scale were found disordered, they were merged together properly.

Making use of the PCM, the estimates of the level of latent trait owned by each subject β, the mean difficulty of the items δ and the thresholds associated to the response categories $\{\tau_j\}_{j=1}^{m}$, were obtained making use of the joint maximum likelihood estimation method implemented in Winsteps 3.75 (Linacre, 2012). Then the CUB model was applied to the data used in the final Rasch analysis and the estimate of the uncertainty parameter for each item was produced.

As it will be shown in the following subsections, the items with a medium or high value of uncertainty $(1-\pi)$ have a counterpart in the simulation study developed in the previous section. This confirm the thesis of this paper that the interpretation of $(1-\pi)$ as a measure of population uncertainty towards the item can be misleading.

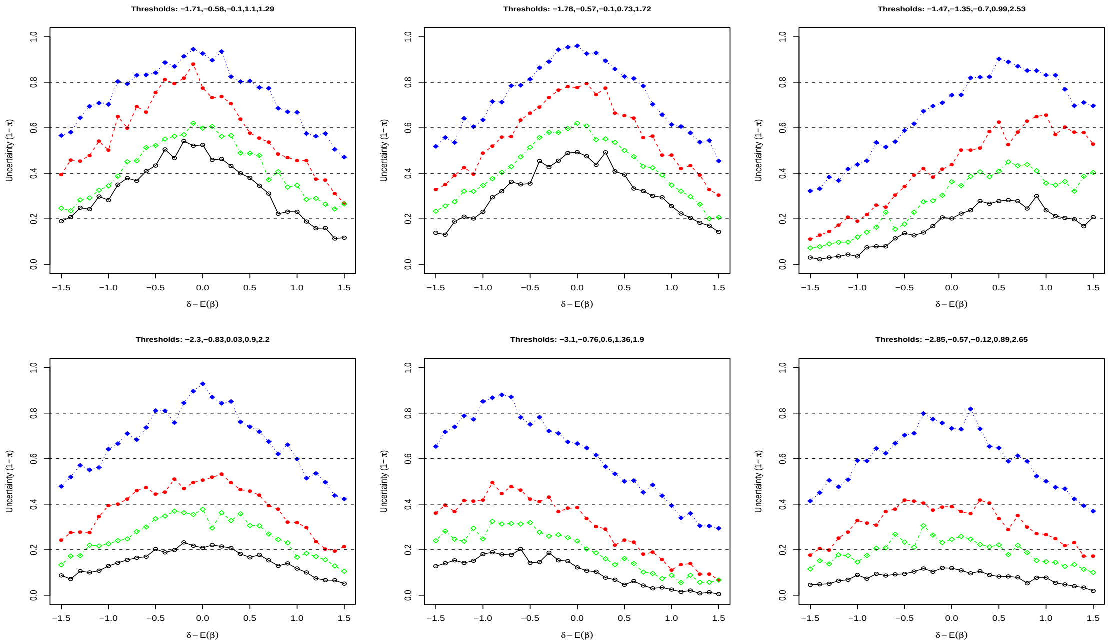

Figure 4: Six response categories. Estimated uncertainty $(1-\pi)$ as function of $\delta - E(\beta)$, thresholds set and ability distribution: N(0,1) (black empty circle), N(0,2) (red full circle), N(0,4) (blue full diamond) and t_6 (green empty diamond)

4.1 The flash Eurobarometer survey on rail and urban transport passenger satisfaction

The aim of the flash Eurobarometer survey on rail and urban transport passenger satisfaction is to analyze European citizens satisfaction with rail services in their country. It was carried out by TNS Political & Social network in the 28 member states of the European Union between 9h and 11th September 2013. The respondents, aged 15 years old or more from different social and demographic groups, were interviewed via telephone in their mother tongue (European Commission, 2014).

The data analyzed in this subsection concern the responses of the Italian citizens to the items composing question Q5b "Are you satisfied or not with the following features of travel by urban public transport (bus, metro, tram etc.) in Italy?". Participants were asked to indicated their satisfaction using a 4-point Likert scale, ranging from 1 (Very dissatisfied) to 4 (Very satisfied); no merging categories was necessary. The sample size was 599.

Table 4 shows the estimates of the PCM e CUB parameters for the items considered. PCM provides also the estimate, for each passenger, of her/his urban transport satisfaction β. The mean satisfaction is -0.035, the variance is 4.045, the skewness is 0.142 and the kurtosis is 4.063.

The analysis of the data using the CUB model highlights an item with a medium level

Table 4: Estimates of PCM and CUB parameters for question Q5b of the Eurobarometer survey

	PCM parameters					CUB parameters	
Item	δ_i	τ_1	τ_2	τ_3	Size	$1-\pi$	$1-\xi$
Frequency	0.15	-1.76	-0.87	2.63	4.39	0.29	0.52
Punctuality	0.12	-1.44	-1.03	2.48	3.92	0.43	0.55
Information	0.11	-2.29	-0.83	3.12	5.41	0.01	0.51
Cleanliness	0.48	-2.16	-0.50	2.66	4.82	0.15	0.45
Routes	-0.75	-2.22	-0.67	2.88	5.10	0.08	0.61
Security	-0.11	-1.99	-0.69	2.68	4.67	0.20	0.54

of uncertainty, that is "Punctuality and reliability" (*Punctuality*). The distance between its mean difficulty and the mean satisfaction is 0.12 and the size of the thresholds interval is 3.92. An analogous behavior can be seen in Figure 2, second row, left panel and line with blue full diamonds, which corresponds to the case of a size of the thresholds interval equal to 4 and a N(0,4) ability distribution.

4.2 Survey on nurse-patient relationship

The aim of the survey on nurse-patient relationship is to study the nurse-patient relationship taking into account the nurses personal feelings of comfort with touch in their daily activities, their capability to provide care-giving, their working well-being, including attachment, self-efficacy to negative and positive emotions and burnout, and their affective commitment with the hospital. The survey was carried out in 2014 involving nurses working in North East Italy (Pedrazza et al., 2015). In this paper two questionnaires of the survey are taking into account, that is the questionnaire that includes the items that measure the anxious attachment and the questionnaire that includes the items that measure the perceived self-efficacy to negative emotions.

4.2.1 Anxious attachment

The attachment theory (Bowlby, 1969, 1975) is the theoretical framework within which the degree and quality of responsiveness of significant others to one's own needs and the own re-action to their behavior can be conceptualized. According to Bartholomew and Horowitz (1991), there are four main adult attachment styles such as secure, preoccupied, dismissive and fearful. The one considered in this subsection is the anxious (preoccupied) attachment style.

Nurses were asked to indicate their degree of agreement to some statements related to thoughts and feelings that usually they feel with respect to their close or intimate relationships, using a 7-point Likert scale, assigning a score from 1 (Complete disagreement) to 7 (Complete agreement). Merging categories was necessary, ending with a 5-point Likert scale. The sample size was 502.

Table 5: Estimates of PCM and CUB parameters for anxious attachment

	PCM parameters						CUB parameters	
Item	δ_i	τ_1	τ_2	τ_3	τ_4	Size	$1-\pi$	$1-\xi$
Be abandoned	0.21	-1.00	-0.57	0.04	1.53	2.53	0.65	0.36
Care	0.06	-1.14	-0.93	0.22	1.86	3.00	0.47	0.49
Strength of feelings	-0.46	-1.55	-0.84	0.21	2.18	3.73	0.24	0.60
Close relationships	0.76	-1.82	-0.58	0.60	1.79	3.61	0.16	0.30
Be alone	0.10	-1.16	-0.30	-0.25	1.70	2.86	0.68	0.44
Really care	-0.08	-1.31	-0.78	0.12	1.97	3.28	0.42	0.53
Coerce others	0.70	-2.05	-0.84	0.25	2.63	4.68	0.08	0.37
Be anxious	-0.03	-1.87	-0.78	0.07	2.58	4.45	0.17	0.51
Be frustrated	-0.23	-1.50	-1.43	0.21	2.73	4.23	0.13	0.56
Be criticized	-1.04	-1.88	-1.18	0.41	2.65	4.53	0.04	0.65

Table 5 shows the estimates of the PCM e CUB parameters for the items considered. PCM provides also the estimate, for each nurse, of her/his level of anxiety β. The mean level of anxiety is -0.201, the variance is 1.413, the skewness is 0.027 and the kurtosis is 6.697.

The analysis of the data using the CUB model highlights two items with a medium level of uncertainty, that is "I worry that others do not care about me as much as I care about them" (*Care*) and "I need to be reassured that the people close to me really cares about me" (*Really care*) and two items with a high level of uncertainty, that is "I am afraid to be abandoned" (*Be abandoned*) and "I am afraid to be alone" (*Be alone*).

The distances between the mean difficulty of items *Care* and *Really care* and the mean level of anxiety are respectively 0.261 and 0.121 and the sizes of the thresholds intervals are 3 and 3.28 respectively. An analogous behavior for *Care* can be seen in Figure 3, first row, central panel and line with green empty diamonds, which corresponds to the case of a size of the thresholds interval equal to 3 and a t_6 ability distribution. Analyzing the values of the green empty diamonds in correspondence to 0.10 in the second and third panel on the first row of Figure 3, which correspond to a size of the threshold interval equal to 3 and 3.5 respectively and to a t_6 distribution for the ability, it is possible to infer the value of the uncertainty measure observed for the item *Really care*.

The distances between the mean difficulty of items *Be abandoned* and *Be alone* and the mean level of anxiety are respectively 0.441 and 0.305 and the sizes of the thresholds intervals are 2.53 and 2.86 respectively. An analogous behavior for *Be abandoned* can be seen in Figure 3, first row, left panel and line with green empty diamonds, which corresponds to the case of a size of the thresholds interval equal to 2.5 and a t_6 ability distribution. Analyzing the values of the green empty diamonds in correspondence to 0.30 in the first two panels on the first row of Figure 3, which correspond to a size of the threshold interval equal to 2.5 and 3 respectively and to a t_6 distribution for the ability, it is possible to infer the value of the uncertainty measure observed for the item *Be alone*.

4.2.2 Emotional self-efficacy: negative emotions

Emotional self-efficacy is concerned with peoples belief in their ability to regulate their positive and negative affective reactions, in response to different situations (Caprara et al., 1999). The survey contained two scales developed by Caprara et al. (1999) able to measure the perceived self-efficacy to negative and positive emotions and the one analyzed here regards the ability to cope with negative emotions.

Nurses were asked to indicate to what extent they felt able to cope with each situation using a 7-point Likert scale, assigning a score from 1 (Not at all) to 7 (Completely). Merging categories was necessary, ending with a 6-point Likert scale and the sample size was 533.

Table 6: Estimates of PCM and CUB parameters for the perceived self-efficacy to negative emotions

	PCM parameters							CUB parameters	
Item	δ_i	τ_1	τ_2	τ_3	τ_4	τ_5	Size	$1-\pi$	$1-\xi$
Lack of appreciation	-0.19	-2.90	-0.92	0.02	1.42	2.37	5.27	0.20	0.60
Heavy criticism	0.16	-2.49	-0.73	0.03	0.99	2.20	4.69	0.40	0.56
Distance from beloved	-0.08	-2.92	-0.73	-0.07	0.98	2.74	5.66	0.23	0.60
Adversity	-0.11	-2.36	-0.87	0.00	0.81	2.42	4.78	0.31	0.64
Stress	-0.08	-2.80	-0.52	-0.19	0.88	2.63	5.43	0.30	0.62
Suffered injustices	0.30	-2.44	-0.81	-0.09	0.96	2.38	4.82	0.37	0.54

Table 6 shows the estimates of the PCM e CUB parameters for the items considered. PCM provides also the estimate, for each nurse, of her/his perceived self-efficacy to negative emotions β. The mean level of perceived self-efficacy is 0.562, the variance is 2.226, the skewness is 0.576 and the kurtosis is 5.757.

The analysis of the data using the CUB model highlights two items with a medium level of uncertainty, that is "Do not get discouraged after a heavy criticism" (*Heavy criticism*) and "Overcome irritation caused by suffered injustices" (*Suffered injustices*).

The distances between the item mean difficulty and the mean perceived self-efficacy are -0.402 and -0.262 respectively, and the sizes of the thresholds interval are 4.69 and 4.82 respectively. An analogous behavior can be inferred considering the values of the red full circles in correspondence to -0.40 and [-0.20, -0.30] respectively in the first two panels on the second row of Figure 4 which correspond to the cases of a size of the thresholds interval equal to 4.5 and 5 and a N(0,2) ability distribution.

5 Conclusions

The aim of this paper is to show that in contexts in which it is reasonable to assume that there is an overall latent variable which determined the answers of the subjects to items, the current interpretation of $(1-\pi)$ parameter in the CUB model as a measure of the uncertainty component in the decision processes could be misleading. In fact, a high value of $(1-\pi)$ should mean that all the subjects belonging to the population are uncertain in the choice of their right answer to the item, and this is unrealistic.

A simulation study was built in order to show that there are situations in which a high value of $(1-\pi)$ does not imply uncertainty in the choice of the response made by the subjects belonging to a population of interest. The PCM[2] has been used as an instrumental model able to generate data for which it is known the level of overall latent trait of each individual as well as some item characteristics, reproducing in this way the specific context of the paper.

The obtained results identified situations in which the measure of uncertainty assumes medium-high values in absence of a real uncertainty in the responses, given that the data used came from a simulation. There is an evident parabolic relationship between the estimated uncertainty measure $(1-\pi)$ and the distance between the mean difficulty of the item and the average ability. The parabolas are concave down and the steepness of the parabolas is directly proportional to the variance of the ability distribution and inversely proportional to the extent of the thresholds set.

Then, three real key studies were analyzed, regarding three different latent traits that are the passengers satisfaction for urban transports, the nurses level of anxious attachment and the nurses perceived self-efficacy to negative emotions. The items considered in each of these key studies were designed to measure the corresponding unique latent trait. It was shown that for the items with a medium or high value of uncertainty $(1-\pi)$ there was a counterpart in the simulation study developed in the first part of the paper, confirming the thesis of the study that the interpretation of $(1-\pi)$ as a measure of population uncertainty towards the item can be misleading.

Therefore, this paper wants to be a warning for the current interpretation of the $(1-\pi)$ as a measure of the uncertainty that accompanies the choice of a response category made by the subjects who form the population under study.

[2]The use of the Generalized Partial Credit Model as instrumental model to the study, does not change the final considerations of this paper. The difference between PCM and the Generalized Partial Credit Model is the introduction in the last one of a discrimination parameter that can differ from item to item.

When the latent trait do not vary much in the population, there is an inverse linear relationship between the CUB feeling towards the item i $(1-\xi_i)$ and the PCM mean difficulty parameter δ_i, as shown in Figure 5, where the δ_i is plotted against $(1-\xi_i)$. In fact, a high (low) level of feeling implies that the higher (lower) response categories are the most chosen which means, in PCM paradigm, that the item is felt easy (difficult) to endorse. So, if the feeling $(1-\xi)$ of an item is very high or very low, which in general

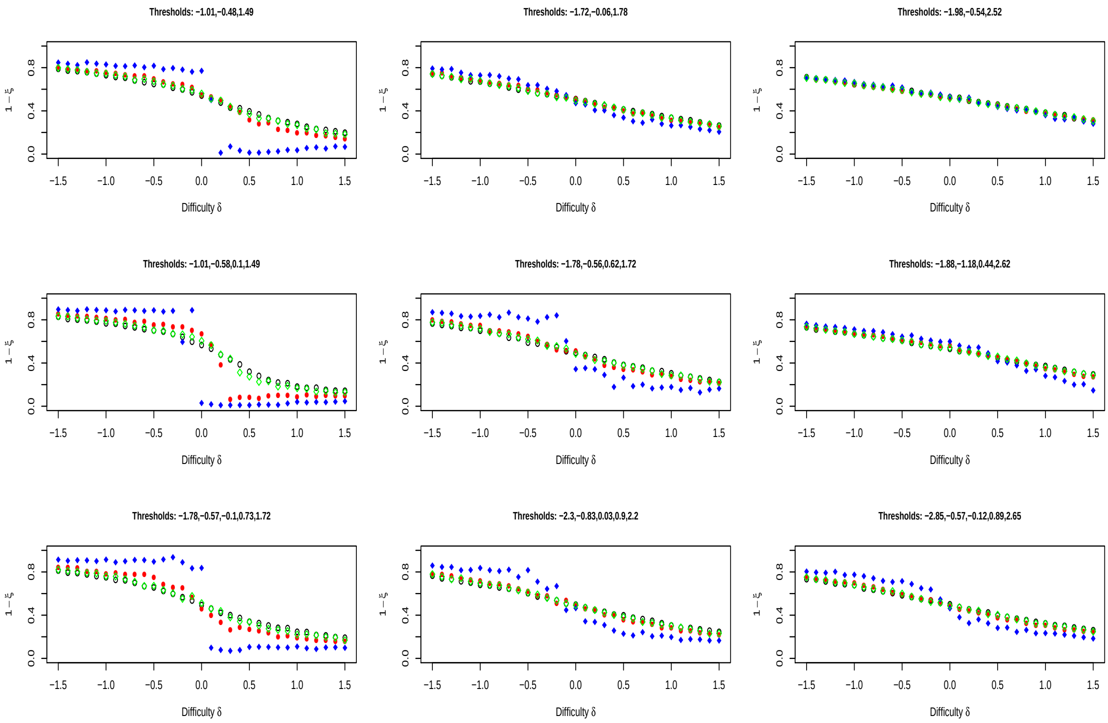

Figure 5: PCM mean difficulty parameter δ versus CUB feeling $(1-\xi)$ for different widths of the threshold interval, 4 (first row), 5 (second row) and 6 (third row) response categories and different ability distribution: N(0,1) (black empty circle), N(0,2) (red full circle), N(0,4) (blue full diamond) and t_6 (green empty diamond)

corresponds to an easy or difficult item, the quantity $(1-\pi)$ could be interpret as a measure of the variability of the overall latent trait over the population; a high value of $(1-\pi)$ suggests a high variance of the distribution of the overall latent trait. Following the PCM paradigm, the items located in the middle of the overall latent trait range provide a great amount of information regarding the latent trait under study, given that they are useful in differentiating respondents who have low to high levels of latent trait and in terms of feeling parameter, these items assume a value of $(1-\xi)$ around 0.5. For these kind of items, at least a medium level of $(1-\pi)$ can be expected if the variability of the latent trait is sufficiently high; therefore again $(1-\pi)$ could be interpret as a measure of the variability of the overall latent trait over the population. A low value of $(1-\pi)$, when $(1-\xi)$ is around 0.5, suggests a low variance of the distribution of the overall latent trait.

This paper dealt with the original global CUB model, for which no covariates were considered. When covariates describing subjects characteristics are included in the model (Piccolo, 2006; Piccolo and D'Elia, 2008) and there is at least a continuous covariate, it is possible to have a uncertainty parameter π for each subject. In this context, the current interpretation of the parameter $(1-\pi)$ is convincing because it is related to the single subjects.

Acknowledgement

The author wishes to thank M. Carpita, M. Pedrazza, D. Piccolo and the two anonymous reviewers for their useful comments.

References

Andrich, D., de Jong, J.H.A.L. and Sheridan, B.E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost and R. Langeheine (Eds.) *Application of latent trait and latent class models in the social science*. Munster: Waxmann.

Bartholomew, K. and Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology* ,61, 226–244.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* New York: Baic Books.

Bowlby, J. (1975). *Attachment and loss: Vol. 2. Separation: Anxiety and Anger.* New York: Baic Books.

Caprara, G. V., Scabini, E., Barbaranelli, C., Pastorelli, C., Regalia, C., and Bandura, A. (1999). Perceived emotional and interpersonal self-efficacy and good social functioning. *Giornale Italiano di Psicologia*, 26,769–789.

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

European Commission, Brussels DG Communication COMM A1 Strategy, Corporate Communication Actions and Eurobarometer Unit (2014). *Flash Eurobarometer 382 (Rail and Urban Transport Passenger Satisfaction).* Version: 1.0.0. GESIS Data Archive. Dataset. http://doi.org/10.4232/1.12027

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

Iannario, M. (2012). Preliminary estimators for a mixture model of ordinal data. *Advances in Data Analysis and Classification*, 6, 163-184.

Iannario, M. and Piccolo, D. (2011). University teaching and students perception: models of the evaluation process. In M. Attanasio and V. Capursi (Eds.) *Statistical Methods for the Evaluation of University Systems*, pages 93–112. Springer-Verlag Berlin Heidelberg.

Iannario, M. and Piccolo, D. (2014). Inference for CUB models: a program in R (Version 4.0). *Statistica & Applicazioni*, XII(2), 177-204.

Linacre, J.M. (2012). Winsteps (Version 3.75.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2012. Available from http://www.winsteps.com/

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.

Masters, G.N. (1988). Measurement models for ordered response categories. In J. Rost and R. Langeheine (Eds.) *Latent trait and latent class model*, pages 11–29. Springer.

Ostini, R. and Nering, M.L. (2006). *Polytomous item response theory models.* Sage Publications

Pedrazza, M., Trifiletti, E., Berlanda, S., Minuzzo, S. and Motteran, A. (2015). Development and initial validation of the nurses' comfort with touch scale. *Journal of Nursing Measurement*, 23(3), 364-378.

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

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

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

R Core Team (2014). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing

Wright, B.D. and Masters, G.N. (1982). *Rating scale analysis.* MESA Press