Chapter 7
A class of statistical models for evaluating services and performances

Marcella Corduas, Maria Iannario and Domenico Piccolo

7.1 Introduction

Evaluation can be described as the psychological process which a subject has to perform when a subject is requested to give a determination of merit regarding an item (the attributes of a service, a product or in general, any tangible or intangible object) using a certain ordinal scale. This process is rooted in the subject's perception of the value/quality/performance of the object under evaluation.

The mechanism governing individual choices between a set of possible alternative options has been widely studied by the latent variables theory. From a statistical point of view, however, the focus is concentrated on modelling empirical observations from sample surveys and on the investigation of the stochastic mechanism generating the ordinal data.

Sample surveys gather measures of satisfaction which are a manifest expression of respondents' constructs. For instance, measuring the satisfaction with a given service, the agreement with a specific statement, the strength of consensus on a certain rule, the perceived experience of a system's performance represent situations where a continuum latent variable (representing the profound belief of the respondent) has to be transformed by a mental process into a discrete state in order to assign an evaluation referred to the graded scale proposed by the interviewer.

Marcella Corduas
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, e-mail: marcella.corduas@unina.it

Maria Iannario
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, e-mail: maria.iannario@unina.it

Domenico Piccolo
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, e-mail: domenico.piccolo@unina.it

M. Bini et al. (eds.), *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, Contributions to Statistics,
DOI 10.1007/978-3-7908-2385-1_7,

The general pattern of responses to a questionnaire aimed at evaluating a service surely presents common features originating from a few latent traits (constructs, variables, factors). This condition, of course, is not immediately recognizable from the observed ratings. Empirical evidences confirm that similarities, differences and contrasts among responses are very common. However, although a remarkable number of hypothetical patterns can be conjectured for rating distributions, only a small subset of them are observed in practice with noticeable frequency.

In the previous Chapters, various approaches widely discussed and applied in the literature have been examined. Attention has been focused on generalized linear latent models, Item Response Theory, unobserved variable approach, and several methodological developments and tools for real applications have been discussed. The main merit of such approaches relies on the possibility of dealing with manifest and latent variables starting from a unique paradigm.

In this Chapter a mixture distribution for ordinal data is introduced. This proposal, as with any innovative tool, is not aimed at replacing existing modelling which are surely based on theories widely investigated and experimented. Instead, it is intended as an additional tool which may be of help in order to better understand real data providing an alternative point of view.

The Chapter is organized as follows: firstly, a simplified description of the evaluation process is presented in order to specify the final result originated from such a process as the combined effect of two unobservable components, one related to the individual feeling for the object under evaluation and the other related to the intrinsic uncertainty which affects any human decision. Later, in Sect. 7.3, a class of models (named CUB) is logically derived from these assumptions, and properties and extensions are illustrated. In Sect. 7.4, inferential issues and numerical procedures for maximum likelihood parameter estimation and related asymptotic inference are discussed; in addition, the main steps of the EM estimation algorithm is provided for a specific CUB model. Sections 7.5 and 7.6 deal with possible applications of this class of models for ordinal data analysis. In particular, a data set concerning students' satisfaction with university "orientation" services is examined. Finally, some remarks on further generalizations and extensions conclude this contribution.

7.2 Unobserved components in the evaluation process

Perception is a cognitive process by which a subject attains awareness or understanding of sensory information and translates them into a form that is meaningful for his/her conscience. In real applications, where statistical tools are needed to analyze evaluation data a simplified archetype of such a process may be of help. In this respect, we can start by considering a simple example concerning university teaching assessment. When a student is asked to answer a specific question about the quality of teaching, he/she has to bring his/her perception of the problem into focus and then he/she has to summarize this perception into a well-defined category using a finite set of ordinal values.

Thus, the final evaluation is the effect of complex causes. It is influenced by considerations fully related to the object of evaluation, but also by the inherent uncertainty that accompanies any human decisions. Moreover, individual behaviours may significantly differ depending on a specific subject's characteristics. Consequently, judgements can be considered as the realization of a stochastic phenomenon which needs to be modelled by taking into account the impact of individual covariates on the expression of the perception. Specifically, with respect to the assessment of university teaching and services, a sensible approach should study how expressed evaluations change with students' profiles.

For this purpose, final judgements, originating from a mental process of selecting among a discrete number of options, can be described as the compounding of two elements:

- a *primary* component, generated by the respondent's sound impression related to awareness and a full understanding of problems, personal or previous experience, group partnership, etc.;
- a *secondary* component, generated by the intrinsic uncertainty affecting the final choice. This may be due, for instance, to the amount of time spent elaborating the answer, the limited range of available information, a partial understanding of the question or to subject's laziness.

Then, the psychological mechanism, by which the choice is made, is the result of a personal *feeling* for the object under judgement and an inherent *uncertainty* associated with the selection of the ordinal value of the response.

7.2.1 Rationale for a new class of models

The interpretation of the respondents' final choice as a weighted combination of individual *feeling (agreement)* and some intrinsic *uncertainty (fuzziness)* leads to the definition of a mixture distribution that will be formally introduced in the following pages. Here, we briefly discuss the rationale behind this new probabilistic model.

Feeling is usually related to subjects' motivations, whereas *uncertainty* mostly depends on circumstances that surround the process of judging. Consequently, the first component is related to the several causes leading to a certain choice, whereas the second is simply related to the confidence/firmness/resolution of such a choice.

In order to model the first component, a *shifted Binomial* random variable is introduced. This is motivated by two arguments.

From a statistical point of view, a standard Binomial distribution is generated by adding several independent and identically distributed Bernoulli choices. Then, we may think that when a subject chooses a rating (among m possible categories) he/she excludes the others by a pairwise comparison (D'Elia, 2000, 2001). For instance, assuming a m-points graded scale (where 1 is related to the best rate), assigning the third grade to an item means that this rate is worse than the first two and better than the other $(m-3)$ ones. Generally, one chooses $(Y = y)$ when the selected choice y is not preferred to the previous $(y-1)$ but it is instead preferable to the remaining $(m-y)$ alternatives. If $(1-\xi)$ and ξ are the probability that each comparison is lost and won, respectively, a given sequence of "failure/success" has a probability of $(1-\xi)^{y-1}\,\xi^{m-y}$. A combinatorial argument proves immediately that the probability of a given choice is: $\binom{m-1}{y-1}(1-\xi)^{y-1}\xi^{m-y}$, for $y=1,2,\ldots,m$. Of course, this reasoning assumes that the random variables describing comparisons are both independent and identically distributed, and, as often happens for a statistical model, this provides a crude approximation of the respondents' effective behaviour.

From a heuristic point of view, the shifted Binomial distribution is able to map a continuous latent variable (characterized by a single mode distribution: Normal, Student-t, logistic, etc.) into a discrete set of values $\{1,2,\ldots,m\}$. The shape of the resulting distribution depends on the way the cut-points are originally chosen. This fact adds further flexibility in modelling the observations since it allows for very different mode location and skewness.

The second component, describing *uncertainty*, is given by a *discrete Uniform* random variable over the support $\{1,2,\ldots,m\}$. This probability distribution is intended as an extreme solution to represent the evaluation process. In this regard, we are not stating that people answer questions in a purely random manner, instead we are saying that the uncertainty affecting any choice can, at worst, be constituted by a situation where no category prevails over the others, and that is the case of a uniform distribution. In fact, the latter maximizes entropy with respect to any other distribution which shares the same finite discrete support.

The random variables related to *feeling* and *uncertainty*, are then combined in a mixture distribution with different weights (π) and $(1-\pi)$ respectively, which denote *propensities* of the subject for one or the other way of constructing his/her choice. In addition, the interpretation for the two unobserved components implies an immediate meaning for the two involved parameters: $(1-\xi)$ will be considered a measure of agreement/feeling for the item of interest whereas $(1-\pi)$ will provide a measure of fuzziness/uncertainty that accompanies the choice.

Some further remarks on the rationale behind the proposed mixture distribution may be useful at this stage. Firstly, it is important to make clear that we are not conjecturing that the population is composed of two subgroups of respondents, each behaving according to one of the two above-mentioned probability distributions.

Secondly, it is worth noticing that *uncertainty*, the component related to choosing, is completely different from *randomness*, which is instead a concept related to sampling variability of surveys.

7.3 Specification and properties of CUB models

Formally, CUB models are specified by considering the ordinal response y as a realization of a discrete random variable Y defined on the support $\{y=1,2,\ldots,m\}$. For given $m>3$, the random variable Y is a mixture of Uniform and shifted Binomial random variables and its probability mass function is given by:

$$Pr(Y=y) = \pi \binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} + (1-\pi)\frac{1}{m}, \quad y = 1,2,\ldots,m,$$

where $\pi \in (0,1]$ and $\xi \in [0,1]$ (Piccolo, 2003; D'Elia and Piccolo, 2005). Thus, the parametric space is:

$$\Omega(\pi,\xi) = \{(\pi,\xi):\ 0 < \pi \le 1;\ 0 \le \xi \le 1\}.$$

From a theoretical point of view, Iannario (2009c) proved that CUB models are fully identifiable for any $m > 3$. Moreover, the proposed mixture distribution is rather flexible and, depending on the parameters, it is able to assume very different shapes: symmetric or extremely skewed, rather flat or with definite mode, and this fact makes it a very useful tool for describing observed ordinal data.

As mentioned in the previous section, $(1-\pi)$ is a *measure of uncertainty* whereas $(1-\xi)$ may be interpreted as a *measure of performance*. Considering the whole random variable support, $(1-\pi)/m$ is a measure of the related *uncertainty share*.

The interpretation of ξ needs some caution because it depends on the initial coding of the responses (as a matter of fact, the graded scale may represent the strongest feeling/concern either by the highest value or by the lowest value). In particular, in several studies conducted in various fields, the parameter ξ has been related to the degree of perception, the strength of selectiveness/awareness, the measure of concern and the threshold of pain.

The parameter values help to locate CUB models in the parametric space defined by the unit square. This is a convenient way of giving an interpretation to results since it allows immediate comparisons among probability structures describing observed ratings. Thus, since $1-\pi$ quantifies the *propensity* of respondents to behave in accordance to a completely random choice, the more π is located to the right side of the unit square, the more respondents give definite answers (uncertainty is low). Similarly, since $1-\xi$ measures the strength of feeling of the subjects for a direct and positive evaluation of the object, the closer ξ is located to the border of the upper region of the unit square the less the item has been preferred.

Fitting to observed ordinal data usually improves when the subjects' *covariates* are introduced in order to relate both the feeling and the uncertainty to the respondents' features. Besides the presence of significant covariates helps the model interpretation and the discrimination among different sub-populations. The latter aim is accomplished by using dummy covariates (Iannario, 2008b) or by clustering methods (Corduas, 2008c,b). In addition, objects' covariates may be introduced (Piccolo and D'Elia, 2008) and thus, similarly to other contexts, CUB models may include *choices' covariates* and *chooser's covariates*: Agresti (2002).

In this regard, we should observe that the expected value of Y is given by:

$$\mathbb{E}(Y) = \pi\,(m-1)\left(\frac{1}{2}-\xi\right) + \frac{(m+1)}{2}.$$

Consequently, different parameter vectors $\boldsymbol{\theta} = (\pi,\,\xi)'$ may generate the same mean value. In such a context, it would not be therefore correct to introduce a link among expectation and covariates (as usually happens in GLM framework). In fact, CUB distributions can be rather different even if they have the same mean value. For this reason, we prefer a more general framework (advocated by King et al., 2000) where parameters describing the probability distribution are directly related to covariates.

Then, the general formulation of a CUB (p,q) model (with p covariates to explain uncertainty and q covariates to explain feeling) is expressed by the *stochastic component*:

$$Pr(Y=y \mid \boldsymbol{x}_i; \boldsymbol{w}_i) = \pi_i \binom{m-1}{y-1} \xi_i^{m-y} (1-\xi_i)^{y-1} + (1-\pi_i)\left(\frac{1}{m}\right), \quad y = 1,2,\dots,m,$$

and two *systematic components*:

$$\pi_i = \frac{1}{1+e^{-\boldsymbol{x}_i \boldsymbol{\beta}}}; \quad \xi_i = \frac{1}{1+e^{-\boldsymbol{w}_i \boldsymbol{\gamma}}}; \quad i = 1,2,\dots,n;$$

where $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ are the subjects' covariates for explaining π_i e ξ_i, respectively (Table 7.1).

Table 7.1 Notation of CUB (p,q) models, without and with covariates

Models	Covariates	Parameters	Parameter spaces
CUB $(0,0)$	No covariates	$\boldsymbol{\theta} = (\pi, \xi)'$	$(0,1] \times [0,1]$
CUB $(p,0)$	Only for π	$\boldsymbol{\theta} = (\boldsymbol{\beta}', \xi)'$	$\mathbb{R}^{p+1} \times [0,1]$
CUB $(0,q)$	Only for ξ	$\boldsymbol{\theta} = (\pi, \boldsymbol{\gamma}')'$	$(0,1] \times \mathbb{R}^{q+1}$
CUB (p,q)	For π and ξ	$\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$	$\mathbb{R}^{p+q+2}$

Notice that this formalization allows that the two sets of covariates may present some overlapping.

The nature of the probability distributions (Uniform and shifted Binomial) included in the mixture and the presence of Covariates justify the acronym CUB (the acronym MUB was used in some initial contributions).

With respect to the classical GLM approach (where proportional, adjacent or continuation ratio probabilities are introduced for ordinal data), CUB models offer a straightforward relationship between a probability statement for ordinal answers and subjects' covariates by means of a monotone function (logistic function, in most cases). Moreover, although latent variables are conceptually necessary in order to specify the nature of the mixture components, the inferential procedures are not based upon the knowledge (or estimation) of cut-points. As a consequence, when the CUB model turns out to be adequate in fitting data, it is usually more parsimonious with respect to models derived by the GLM approach.

CUB models have been further generalized for taking the possible effect of atypical situations into account. Sometimes, these are derived by *shelter choices*, which represent categories frequently selected by respondents in order to avoid more elaborate decisions.

Specifically, an *extended* CUB *model* is defined by:

$$p_y(\boldsymbol{\theta}) = \pi_1 \binom{m-1}{y-1} \xi^{m-y}(1-\xi)^{y-1} + \pi_2 \frac{1}{m} + (1-\pi_1-\pi_2)\, D_y^{(c)}, \quad y = 1,2,\ldots,m,$$

where $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi)'$ is the parameter vector characterizing the distribution of this new mixture random variable and $D_y^{(c)}$ is a degenerate random variable whose probability mass is concentrated at $y = c$, that is:

$$D_y^{(c)} = \begin{cases} 1, \text{ if } y = c; \\ 0, \text{ otherwise.} \end{cases}$$

We observe that extended CUB models are identifiable only for $m > 4$.

Of course, if $\pi_1 + \pi_2 = 1$ the extended CUB model collapses to the standard one. Instead, if $\pi_2 = 0$ we are just considering a mixture of a shifted Binomial distribution and a degenerate probability with mass at $(Y = c)$. Moreover, if $\pi_1 = \pi_2 = 0$ the extended model is able to account also for the (rare) situation where most of respondents' choices are concentrated at a single intermediate category.

A remarkable feature of the extended model is that parameter $\delta = 1 - \pi_1 - \pi_2$ measures the added relative contribution of the *shelter choice* at $y = c$ with respect to the standard version of the model. Since its significance may be tested via standard asymptotic inference, extended CUB models may check the effective relevance of the presence of a *shelter choice*. Furthermore, it should be noted that in some circumstances – if one avoids considering this component – parameter estimates are biased and inefficient, and fitting and predictions are not satisfactory.

Among others, this effect has been found in the evaluations of a data set collected among students attending courses at the University of Naples Federico II. The main objective of the survey was to measure several aspects of students' satisfaction with the teaching, lecture halls, time scheduling, services, etc. The survey was conducted using a questionnaire where the assessment of each item was based on the following 7 points scale: "extremely unsatisfied" $(= 1)$, "very unsatisfied" $(= 2)$, "unsatisfied" $(= 3)$, "indifferent" $(= 4)$, "satisfied" $(= 5)$, "very satisfied" $(= 6)$, "extremely satisfied" $(= 7)$. Thus, the assessment of a given item generates a rating Y with $m = 7$. In general, it has been observed that the distributions for most of the items under investigation present a very marked mode at $Y = 5$ (corresponding to the "satisfied" category).

Since respondents were a selected subset of enrolled students (those who regularly attend lectures are more likely to be satisfied with University life), a consistent part of them preferred to select the first positive judgement available on the proposed graded scale in order to avoid a more thoughtful assessment. In these cases, one should test the hypothesis $H_0 : \delta = 0$ in the extended CUB model with $c = 5$. In the examined data set, the parameter estimate $\hat{\delta} = 0.223$ (with a standard error of *0.004*) confirms a substantial effect of the *shelter choice* with respect to the expected one. Moreover, the model fitting and the prediction of expected responses are improved.

A final remark concerns the possible presence of bimodal (multimodal) distributions which, at a first sight, may suggest adding further Binomial components to the mixture distribution in order to model the presence of various modes. In our opinion, adding random variables of the same family in order to explain the different behavior of respondents should be avoided since problems concerning model identifiability may arise. Instead, for this purpose, the introduction of subjects covariates should be seriously considered so that clustered responses might be taken into account. For instance, when people are asked to give a rate to a politician, the bimodal distributions of responses may be easily modelled if the ideological position (left/right) of the respondents are surveyed. In such a case, dichotomous or polytomous variables will be introduced as explanatory variables in the CUB model in order to explain the opposite expressed feeling.

7.4 Inferential issues and numerical procedures

Given a sample of observed ordinal data and covariates $(y_i, \boldsymbol{x}_i, \boldsymbol{w}_i)'$, for $i = 1, 2, \ldots, n$, the log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ in a general CUB (p, q) model is defined by:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \left[\frac{1}{1+e^{-\boldsymbol{x}_i\boldsymbol{\beta}}} \left\{ \binom{m-1}{y_i-1} \frac{e^{(-\boldsymbol{w}_i\boldsymbol{\gamma})(y_i-1)}}{(1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}})^{m-1}} - \frac{1}{m} \right\} + \frac{1}{m} \right].$$

Inferential issues for the joint efficient estimation of the parameters are discussed in details by Piccolo (2006) who derived the EM algorithm for maximum likelihood (ML) estimation. The procedure is effective but convergence to maximum can be rather slow; then, several proposals for improving preliminary parameter estimates have been suggested in order to improve the rate of convergence (Iannario, 2009a). In this regard, moment estimators provide useful initial values but some problems arise for models with covariates. These aspects are currently under investigation.

In the following section, a brief illustration of the EM estimation algorithm is presented with special reference to the extended model without covariates.

7.4.1 The EM algorithm

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ be the sample of ordinal data generated by a survey where n respondents are asked to choose an integer in the support $\{1, 2, \ldots, m\}$, for a given $m > 4$. We suppose the location $c \in \{1, 2, \ldots, m\}$ of the *shelter choice* is known.

For the extended CUB model, the log-likelihood function $\ell(\boldsymbol{\theta})$ for the sample $\boldsymbol{y}$, with $\boldsymbol{\theta} = (\pi_1, \pi_2, \xi)'$, is

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log\left[Pr(Y=y_i \mid \boldsymbol{\theta})\right] \\ &= \sum_{i=1}^{n} \log\left[\pi_1\, b_{y_i}(\xi) + \pi_2\, U_{y_i}(m) + (1-\pi_1-\pi_2) D_{y_i}^{(c)}\right],\end{aligned}$$

where the components of the mixture are specified, for $i = 1,2,\ldots,n$, by:

$$p_g(y_i;\boldsymbol{\theta}_g) = \begin{cases} b_{y_i}(\xi) = \binom{m-1}{y_i-1}\xi^{m-y_i}(1-\xi)^{y_i-1}, & g=1; \\ U_{y_i}(m) = \frac{1}{m}, & g=2; \\ D_{y_i}^{(c)}, & g=3. \end{cases}$$

We introduce the unobservable vector $\boldsymbol{z} = (\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_n)'$ where each $\boldsymbol{z}_i = (z_{1i}, z_{2i}, z_{3i})'$ is a three-dimensional vector such that, for $g = 1,2,3$:

$$z_{gi} = \begin{cases} 1, \text{ if the } i\text{-th subject belongs to the } g \text{ group}; \\ 0, \text{ otherwise.} \end{cases}$$

Simplifying the notation, we let:

$$\pi_g = \begin{cases} \pi_1, & g=1; \\ \pi_2, & g=2; \\ 1-\pi_1-\pi_2, & g=3. \end{cases} \qquad \boldsymbol{\theta}_g = \begin{cases} \boldsymbol{\theta}_1 = (\pi_1,\xi)', & g=1; \\ \boldsymbol{\theta}_2 = \pi_2, & g=2; \\ \boldsymbol{\theta}_3 = 1-\pi_1-\pi_2, & g=3. \end{cases}$$

Then, the likelihood function of the complete-data vector $(\boldsymbol{y}', \boldsymbol{z}')'$ is given by:

$$L_c(\boldsymbol{\theta}) = \prod_{g=1}^{3}\prod_{i=1}^{n}\left[\pi_g p_g(y_i;\boldsymbol{\theta}_g)\right]^{z_{gi}},$$

and the complete-data log-likelihood function is:

$$\ell_c(\boldsymbol{\theta}) = \sum_{g=1}^{3}\sum_{i=1}^{n} z_{gi}\left[\log(\pi_g) + \log(p_g(y_i;\boldsymbol{\theta}_g))\right].$$

The $(k+1)$-th iteration of the EM algorithm consists of the following steps:

• *E-step*:

The conditional expectation of the indicator random variable Z_{gi}, given the observed sample $\boldsymbol{y}$, is:

$$\mathbb{E}\left(Z_{gi} \mid \boldsymbol{y}, \boldsymbol{\theta}^{(k)}\right) = Pr\left(Z_{gi} = 1 \mid \boldsymbol{y}, \boldsymbol{\theta}^{(k)}\right) = \frac{\pi_g^{(k)}\, p_g\left(\boldsymbol{y};\boldsymbol{\theta}_g^{(k)}\right)}{\sum_{j=1}^{3}\pi_j^{(k)}\, p_j\left(\boldsymbol{y};\boldsymbol{\theta}_j^{(k)}\right)} = \tau_g(y_i;\boldsymbol{\theta}^{(k)}),$$

for any $g = 1,2,3$. This quantity is the posterior probability that the i-th subject of the sample with the observed y_i belongs to the g-th component of the mixture.

Then, the expected log-likelihood of complete-data vector is obtained as:

$$\mathbb{E}\left(\ell_c(\boldsymbol{\theta})\right) = \sum_{g=1}^{3}\sum_{i=1}^{n} \tau_g(y_i;\boldsymbol{\theta}^{(k)}) \left[\log\left(\pi_g^{(k)}\right) + \log\left(p_g\left(y_i;\,\boldsymbol{\theta}_g^{(k)}\right)\right)\right].$$

• *M-step*:

At the $(k+1)$-th iteration, the function $Q(\boldsymbol{\theta}^{(k)}) = \mathbb{E}\left(\ell_c(\boldsymbol{\theta})\right)$ has to be maximized with respect to the parameters (π_1, π_2) and ξ. If the parameters of the components are specified, this quantity may be expressed as follows:

$$\begin{aligned}
Q(\boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^{n}\left[\tau_1\left(y_i;\boldsymbol{\theta}_1^{(k)}\right)\log\left(\pi_1^{(k)}\right) + \tau_2\left(y_i;\boldsymbol{\theta}_2^{(k)}\right)\log\left(\pi_2^{(k)}\right) + \tau_3\left(y_i;\boldsymbol{\theta}_3^{(k)}\right)\log\left(\pi_3^{(k)}\right)\right] \\
&+ \sum_{i=1}^{n}\sum_{g=1}^{3}\left[\tau_g\left(y_i;\boldsymbol{\theta}_g^{(k)}\right)\log\left(p_g\left(y_i;\,\boldsymbol{\theta}_g^{(k)}\right)\right)\right] \\
&= S_1\log\left(\pi_1^{(k)}\right) + S_2\log\left(\pi_2^{(k)}\right) + (n - S_1 - S_2)\log\left(1 - \pi_1^{(k)} - \pi_2^{(k)}\right) + Q^*
\end{aligned}$$

where Q^* is independent from $\pi_g^{(k)}$ parameters, and

$$S_g = \sum_{i=1}^{n} \tau_g(y_i;\boldsymbol{\theta}^{(k)}), \quad g = 1,2; \qquad S_3 = n - S_1 - S_2.$$

Then, by solving the system: $\frac{\partial\, Q(\boldsymbol{\theta}^{(k)})}{\partial\, \pi_g} = 0$, for $g = 1,2$, we get:

$$\pi_1^{(k+1)} = \frac{S_1}{n} = \frac{1}{n}\sum_{i=1}^{n}\tau_1(y_i;\boldsymbol{\theta}^{(k)}); \qquad \pi_2^{(k+1)} = \frac{S_2}{n} = \frac{1}{n}\sum_{i=1}^{n}\tau_2(y_i;\boldsymbol{\theta}^{(k)}).$$

Instead, the estimate of ξ, for a given k, is obtained from:

$$\sum_{i=1}^{n}\tau_1\left(y_i;\boldsymbol{\theta}_1^{(k)}\right)\frac{\partial\,\log(p_1(y_i;\,\xi))}{\partial\,\xi} = 0.$$

A simple algebra produces the solution:

$$\xi^{(k+1)} = \frac{m - \overline{Y}_n(p)}{m-1}; \qquad \overline{Y}_n(p) = \frac{\sum_{i=1}^{n} y_i\,\tau_1\left(y_i;\boldsymbol{\theta}_1^{(k)}\right)}{\sum_{i=1}^{n}\tau_1\left(y_i;\boldsymbol{\theta}_1^{(k)}\right)}.$$

Here, $\overline{Y}_n(p)$ is the average of the observed sampled values weighted with the posterior probability that y_i is a realization of the first component of the mixture (that is a shifted Binomial distribution).

Then, E- and M- steps are repeated with new parameters $\left(\pi_1^{(k+1)}, \pi_2^{(k+1)}, \xi^{(k+1)}\right)'$ until a convergence criterion is satisfied. For instance, this could be given by:

$\mid \ell(\boldsymbol{\theta}^{(k+1)}) - \ell(\boldsymbol{\theta}^{(k)}) \mid < \varepsilon$, for a small $\varepsilon > 0$.

Notice that, as far as ML estimation is concerned, sample data $(y_1, y_2, \dots, y_n)'$ is equivalently represented by the vector of absolute frequencies $(n_1, n_2, \dots, n_m)'$. For computational efficiency, it is therefore convenient to use in previous steps the log-likelihood function for grouped data. To this end, we will compute:

$$S_g = \sum_{y=1}^{m} n_y \, \tau_g(y; \boldsymbol{\theta}^{(k)}), \quad g = 1, 2 .$$

The step-by-step formulation of the EM algorithm may be easily programmed in formal languages (such as *GAUSS*©, *Matlab*© or *R*).

Maximum likelihood inference has been developed by using standard approaches (Piccolo, 2006). Specifically, the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of ML estimators $\hat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ of CUB model is based on the *observed information matrix* $\mathscr{I}(\boldsymbol{\theta})$, that is the negative of the Hessian computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$; it shares the same asymptotic properties of the expected information matrix (as argued by Pawitan (2001), 244–247 among others).

Then, the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of the ML estimators of $\boldsymbol{\theta}$, computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} = (\hat{\pi}, \hat{\xi})'$, is obtained as:

$$\boldsymbol{V}(\boldsymbol{\theta}) = \left[\mathscr{I}(\hat{\boldsymbol{\theta}})\right]^{-1} = - \begin{pmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi \partial \xi} \\ \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \pi \partial \xi} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \xi^2} \end{pmatrix}^{-1}_{(\boldsymbol{\theta} = \hat{\boldsymbol{\theta}})} .$$

The computational details for implementing these results are discussed by Piccolo (2006) and a related software in *R* is currently available for estimation and inference about CUB models with (or without) covariates (Iannario and Piccolo 2009).

7.4.2 *Fitting measures*

The adequacy of models may be checked by means of several measures (significance of parameters, sensible increase in log-likelihood, and so on). However, the sample size of evaluation data sets is generally large and thus, in order to verify how estimated CUB models fit empirical data, we prefer to introduce a descriptive measure for models without covariates and refer to likelihood-based indexes for more general comparisons.

Specifically, from a descriptive point of view, we consider the normed *dissimilarity index* Diss $\in [0, 1]$ defined by:

$$\text{Diss} = \frac{1}{2}\sum_{y=1}^{m}\left|Pr\left(Y=y \mid \hat{\boldsymbol{\theta}}\right) - \frac{n_y}{n}\right| .$$

This index has an appealing interpretation since it measures the proportion of subjects that should modify their choices in order to reach a perfect fit between observed and theoretical distributions (Leti, 1979; Simonoff, 2003). Unfortunately, it cannot be immediately extended to the case of CUB models with covariates.

For this aim, using an obvious notation, log-likelihoods of CUB models can be compared as follows:

Comparisons	Deviances difference	Degrees of freedom
CUB $(p,0)$ versus CUB $(0,0)$	$2\,(\ell_{10}-\ell_{00})$	p
CUB $(0,q)$ versus CUB $(0,0)$	$2\,(\ell_{01}-\ell_{00})$	q
CUB (p,q) versus CUB $(0,0)$	$2\,(\ell_{11}-\ell_{00})$	$p+q$

The difference between deviances should be compared with the quantile of the χ^2 distribution with degrees of freedom as reported in the table above.

In this regard, the log-likelihood for the *saturated* CUB model can provide a useful benchmark:

$$\ell_{\text{sat}} = -n\log(n) + \sum_{y=1}^{m} n_y \log(n_y) .$$

The fitting measure may be obtained by defining a pseudo-R^2, that is named ICON (=*I*nformation CONtent), which compares the log-likelihood of the estimated model with the log-likelihood of a discrete Uniform random variable fitted to data (this is in fact the uninformative model). Thus, the ICON index is:

$$\text{ICON} = 1 + \frac{\ell(\hat{\boldsymbol{\theta}})/n}{\log(m)} .$$

It measures the improvement achieved by a CUB model, without or with covariates, with respect to a completely uninformative distribution (such as the Uniform distribution). In other words, this index is related to the displacement of the log-likelihood of the estimated model with respect to an extreme situation.

7.5 Fields of application

In opinion surveys people are often requested to arrange a list of m items in order of preferences or, alternatively, they are asked to express judgements or evaluations using a given m-point ordinal scale. In this respect, we need to distinguish clearly between two situations: the *rating* where the subject's answer is a single score for each item, and the *ranking* where the answer is a permutation of the first m integers, that is a vector of numbers specifying sequentially the degree of preferences for the m objects.

For the correct understanding of the usage of CUB models, it is important to underline that our approach suggests a mixture distribution useful for modelling the random variable generated by the assessment of a single item (*rating*) or by the positions of a single object in the ordering (*ranking*). However, notice that while in the first case the CUB model is applied to study the univariate response of a group of subjects, in the second case the model is used for the marginal analysis of the discrete multivariate random variable generated by the observed preferences for m objects. In the latter case, it is evident that adopting this strategy in turn for all the m marginal distributions of the ranks leads to non independent random variables.

In previous studies, various applications of the proposed approach have been elaborated in order to fit and interpret univariate *rating data*, especially in relation to evaluations of attributes of goods and services (Corduas, 2008c) and other fields of analysis such as social analysis (Iannario, 2007, 2008a), medicine (D'Elia, 2008), sensometric studies (Piccolo and D'Elia, 2008) and linguistics (Balirano and Corduas, 2008). In such contexts, the paradigm based on modelling the feeling and uncertainty components has turned out to be very useful for its interpretative content.

A further kind of application stems from *categorical data* that are qualitative in their nature although they are actually measured by means of a quantitative scale. In these cases, a genuine ordinal approach proves to be more fruitful for the interpretation and assessment of original data. This approach has been pursued in order to investigate how final grades achieved by university students are related to gender and time spent to complete the university program of studies. As a matter of fact, although the final grade is expressed on a quantitative scale it should be regarded as a qualitative assessment of the examining committee about candidates. This case study has confirmed the better performance of qualitative models with respect to standard quantitative models in relation to the tails of the distribution (given the robustness property of ordinal values) and to the prediction of extreme data.

Finally, the transformation of subjective survival probability (expressed by a percentage on $[0,100]$ scale) into an ordinal score described by a standard 7-point scale has provided another interesting data base for further modelling. This is, in fact, a typical case where the numerical value, that a subject gives in reply to a question, is clearly generated by a qualitative consideration about the perception he/she has of "high" or "low" probability. Again, CUB models has proved to be effective.

7.6 Further developments: a clustering approach

In order to compare rating distributions related to a number of items or to different groups of respondents, a clustering procedure for ordinal data based on estimated CUB models has been introduced by Corduas (2008d).

The search for a special approach is motivated by the risk of misleading interpretations of data arising from the representation of CUB models in the parameter space, because in such a situation the user tends to assess the closeness of two (estimated)

CUB distributions in terms of the Euclidean distance between the corresponding estimated parameters. As a matter of fact, the variability of π and ξ estimators are different and, in addition, the role of CUB coefficients in determining the shape of the estimated distribution is very dissimilar (Piccolo, 2003).

For this reason, Kullback-Liebler (KL) divergence can be used for testing similarities among distributions. Consider two discrete populations each characterized by a probability distribution function having the same functional form $p(y,\boldsymbol{\theta}_i)$ with unspecified parameters $\boldsymbol{\theta}_i$, $i = 1,2$. Also assume that $p(y,\boldsymbol{\theta}_i) > 0$, $\forall y$. Suppose that two samples of n_1 and n_2 observations have been randomly drawn from each population, respectively. In order to test the hypothesis $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, the KL divergence statistic is defined by:

$$\widehat{J} = \frac{n_1\,n_2}{n_1+n_2}\left[\sum_y \Big(p(y;\boldsymbol{\theta}_1) - p(y;\boldsymbol{\theta}_2)\Big)\,\log\frac{p(y;\boldsymbol{\theta}_1)}{p(y;\boldsymbol{\theta}_2)}\right]_{(\boldsymbol{\theta}_1=\hat{\boldsymbol{\theta}}_1,\,\boldsymbol{\theta}_2=\hat{\boldsymbol{\theta}}_2)}$$

where the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ have been replaced by the ML estimators. Under the null hypothesis, it can be shown that $\widehat{J}$ is asymptotically distributed as a $\chi^2_{(g)}$ random variable (Kullback, 1959), being g the common dimension of the parameter vector; in the special case under investigation $g = 2$.

The strategy for grouping a set of CUB models combines hypotheses testing with a clustering algorithm. Firstly, $\widehat{J}$ for each couple of models is evaluated. Secondly, a binary matrix is built by setting the (i, j)th entry equal to 0, when the hypothesis of homogeneity of the i-th and j-th models is rejected, and 1 otherwise. Finally, by means of convenient algorithms (such as BEA: McCormick et al., 1972; Arabie and Hubert, 1990), this matrix is rearranged into an approximate block diagonal form. A clearly defined (unit) triangle immediately under the diagonal will indicate a cluster of items for which the judgements expressed by respondents (and summarized by CUB distributions) are similar. The presence of any zero value in such a triangle indicates that the cluster may be elongated or constituted by other well separated small clusters.

The proposed technique is able to discriminate the different patterns of the rating distributions with respect to skewness, kurtosis, mode and it is very effective and selective as has been proved by various empirical studies (for instance, see Corduas, 2008b,c,d for a study concerning university students' opinions about teaching quality) .

7.7 Case study

In the years 2002–2004 and 2007–2008, the University of Naples Federico II carried out an extensive survey of students' opinions concerning the Orientation services which operated in the 13 Faculties. In this section, the study will focus on the data sets gathered in the last 2-year period.

A questionnaire was submitted to a sample of users and each student was asked to give a score for expressing his/her satisfaction with various aspects of the Orientation service. Eight items have been investigated: staff willingness (=WILL) and competence (=COMPE), clearness of information (=INFO), suitable opening hours (=TIME), adequate equipment and structure (=STRU), advertisement of the service (=ADVE), usefulness of information for decisions (=DECI), and a final overall evaluation (=GLOBA). Judgments were expressed using the ordinal scale ranging from 1 (="completely unsatisfied") to 7 (="completely satisfied").

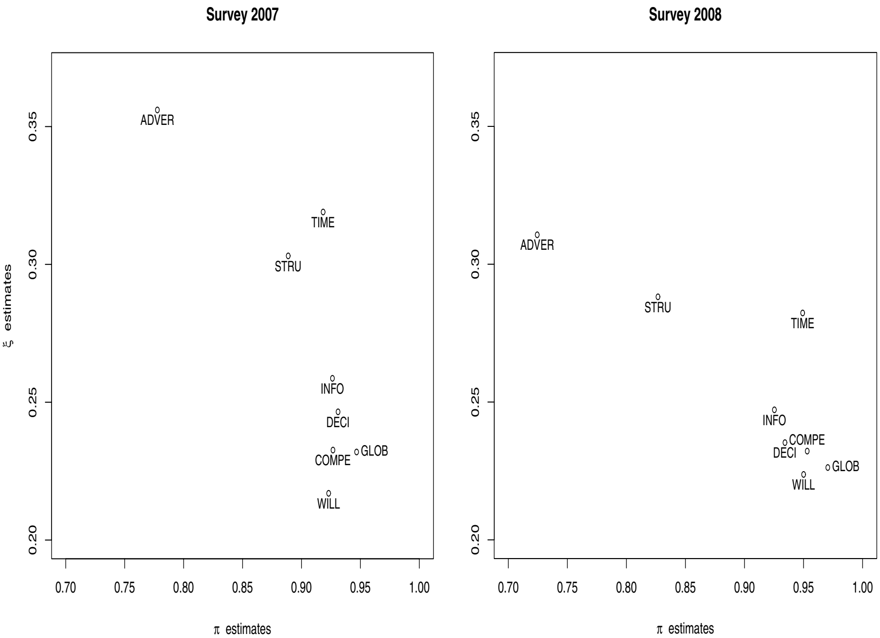

Fig. 7.1 CUB models of students' satisfaction with University Orientation services.

In Fig. 7.1, the estimated parameters of CUB models built for the eight items for 2007 and 2008 surveys are plotted in the parameter space. The results refer to $3,511$ and $4,042$ validated questionnaires for the first and second survey, respectively.

Respondents show a different attitude towards the activities performed by the staff and the aspects related to office organization and equipment since the former type of items systematically receive higher evaluations than the others.

Moreover, comparing the results from the first and the second survey, the expressed satisfaction with items concerning office organization and equipment seems to improve. In Fig. 7.1, the corresponding estimated CUB models, in fact, moves from the top part of the graph to the lower one. Noticeably, the opinion for the lack of adequate advertisements of the service is very critical. Finally, items related to staff evaluations receive more resolute assessment in the second year; the estimated value of $(1-\xi)$ for this type of items is higher than for any other.

Furthermore, the plot suggests that at least two latent variables may govern the responses since the models appear to be grouped in two separate clusters.

The merit of the previous examples is that CUB models are able to summarize and visualize rating distributions originating from thousands of opinions given in different periods of time.

Afterwards, attention is concentrated on 2007 data sets which have been partly examined by Iannario and Piccolo (2008). We conjecture that the CUB model related to students' satisfaction with the office opening hours (=TIME) may improve by introducing significant covariates. Then, available covariates have been added to the CUB model by a stepwise strategy. Specifically, the covariate that mostly improves the log-likelihood function, compared to the others, has been preferred. The resulting parameter estimates (in parentheses their standard errors) and the corresponding log-likelihood values are presented in Table 7.2. Comparison of deviances (not reported here) confirms that the fitted models are all significant and better than the nested ones.

Table 7.2 CUB (p,q) models of students' evaluation for *opening hours*

Models	$\hat{\pi}$	$\hat{\xi}(\boldsymbol{w})$	log-likelihood
► CUB $(0,0)$	0.918 (0.011)	$\hat{\xi} =$ 0.319 (0.004)	$\ell_{00} = -5714.8$
► CUB $(0,1)$	0.920 (0.011)	$\hat{\gamma}_0 =$ 1.464 (0.347)	$\ell_{01} = -5693.6$
log(Age)		$\hat{\gamma}_1 = -0.722$ (0.113)	
► CUB $(0,2)$	0.921 (0.010)	$\hat{\gamma}_0 =$ 1.505 (0.348)	$\ell_{02} = -5687.6$
log(Age)		$\hat{\gamma}_1 = -0.756$ (0.114)	
Gender		$\hat{\gamma}_2 =$ 0.116 (0.034)	
► CUB $(0,3)$	0.921 (0.010)	$\hat{\gamma}_0 =$ 1.601 (0.349)	$\ell_{03} = -5681.6$
log(Age)		$\hat{\gamma}_1 = -0.793$ (0.114)	
Gender		$\hat{\gamma}_2 =$ 0.114 (0.034)	
Change		$\hat{\gamma}_3 =$ 0.190 (0.054)	
► CUB $(0,4)$	0.922 (0.010)	$\hat{\gamma}_0 =$ 1.879 (0.375)	$\ell_{04} = -5679.4$
log(Age)		$\hat{\gamma}_1 = -0.866$ (0.120)	
Gender		$\hat{\gamma}_2 =$ 0.116 (0.034)	
Change		$\hat{\gamma}_3 =$ 0.182 (0.054)	
Full-time (FT)		$\hat{\gamma}_4 = -0.078$ (0.038)	

We denote the covariates for the i-th subject as:

$$\boldsymbol{w}_i = (\log(\text{Age}_i), \text{Gender}_i, \text{Change}_i, FT_i)'.$$

Then, given $m = 7$, the best CUB $(0,4)$ model implies the following probability distributions for the expressed evaluations:

$$Pr(Y = y \mid \boldsymbol{w}_i) = 0.011 + 0.922 \binom{6}{y-1} (1-\xi_i)^{y-1} \xi_i^{7-y}, \; y = 1, 2, \ldots, 7,$$

where the parameters $\xi_i = \xi_i \mid \mathbf{w}_i$, $i = 1, 2, \ldots, n$, are specified by:

$$\frac{1}{1+\exp\{-1.879+0.866\log(\text{Age}_i)-0.116\,\text{Gender}_i-0.182\,\text{Change}_i+0.078\,FT_i\}}$$

Since $(1-\xi)$ is a measure of satisfaction, the estimated model shows that evaluation increases with *Age* and for full-time students $(FT = 1)$ whereas women $(\text{Gender} = 1)$ and students who change their original enrollment and move from one Faculty to another $(\text{Change} = 1)$ lower their preferences, and thus they are more critical about Opening Hours.

Table 7.3 Comparison of different students' profiles and corresponding parameters

Profiles	Age	Gender	Change	Full-time	$\mathbf{w}_i$	$\xi_i \mid \mathbf{w}_i$	$Pr(Y \geq 5)$
A	20	Woman	No	Yes	$(20,1,0,1)'$	0.337	0.654
B	40	Woman	No	Yes	$(40,1,0,1)'$	0.218	0.843
C	20	Man	Yes	Yes	$(20,0,1,1)'$	0.352	0.627
D	40	Man	Yes	Yes	$(40,0,1,1)'$	0.229	0.828
E	20	Woman	Yes	Yes	$(40,1,1,1)'$	0.379	0.576
F	40	Woman	Yes	Yes	$(40,1,1,1)'$	0.251	0.798

The model allows immediate comparison of different profiles; some of them are proposed in Table 7.3. Notice that the implied coefficient $\pi = 0.922$ is constant for all profiles since there are no significant covariates for the uncertainty component in the best estimated model.

It is evident from Fig. 7.2 how the age of the student is the relevant covariate forcing rating distribution into higher values. The last column in Table 7.3 shows that the probability of a positive evaluation mostly changes with age. Marginal changes in the distribution shape are determined by job position and by changing the original university enrollment to enter a new Faculty. Because of the large sample size, these covariates are significant although they achieve a modest impact.

The examination of expected evaluation for given profiles of respondents allows further considerations about the use of CUB models in empirical studies. As far as ordinal variables are concerned, expected values should only be considered for comparative purposes rather than being used as an index which is meaningful in itself.

In the present work, ordinal variables are intended as a monotone transformation of a latent variable Y^* then the study of the expected value of the random variable Y is worthy of interest whenever it is referred to groups of respondents with the same profile.

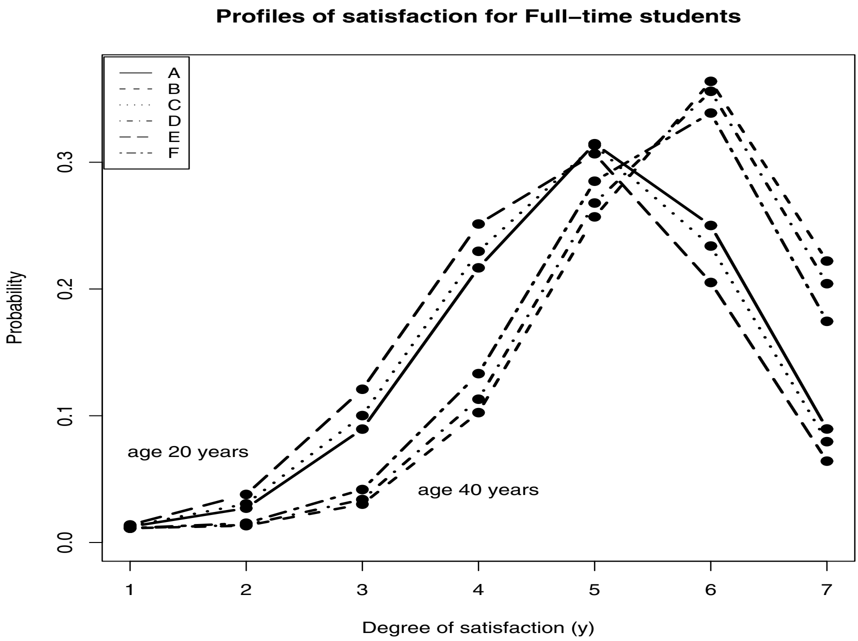

Fig. 7.2 CUB models of students' satisfaction with university orientation services.

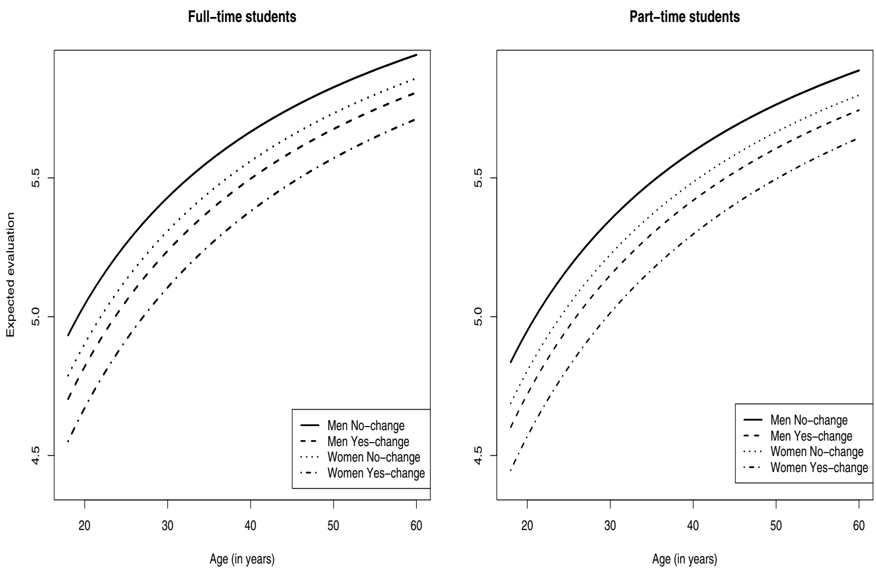

Fig. 7.3 Expected student satisfaction with opening hours for given covariates.

Figure 7.3 exemplifies this approach. In particular, the expected satisfaction is shown for the varying age of the students and their significant covariates. The plots confirm that satisfaction improves with age in a systematic way, a small increase may be observed for full-time students and those that did not change Faculty and a more severe judgement is formulated by women compared to men.

7.8 Concluding remarks

Although CUB models only describe univariate distributions of judgements, their use seems to be effective for investigating sound relationships among ordinal responses and covariates and, in addition, for enhancing unobserved traits in the data. In particular, the role of covariates is made manifest in the model and result in a useful device for the analysis of profiles.

Some unexplored issues that deserve further research are worth mentioning:

- Evaluation data and performances measures are collected in stratified subgroups both for economic reasons and research needs. Then, the introduction of multilevel CUB models is a relevant issue for further developments in this area.
- It is well-known that the range of multivariate distributions implied by the given marginal CUB models is limited. Thus, the efforts for generalizing the approach to a multivariate framework should help to retain the effectiveness of this parametrization and improve current interpretations.
- Fitting measures should be examined closely in order to exploit information carried by likelihood functions for sampled data.
- Differences and areas of complementary usage with other well-established approaches, such as Item Response Theory, are currently under investigation.
- Since large data sets with a great quantity of information about subjects are commonly available from surveys, further studies are needed in order to improve the criteria for the selection of significant covariates and the preliminary parameter estimation by considering both numerical algorithms and data mining procedures.