CENTRO PER LA FORMAZIONE IN ECONOMIA E POLITICA DELLO SVILUPPO RURALE
UNIVERSITÀ DI NAPOLI FEDERICO II - DIPARTIMENTO DI SCIENZE STATISTICHE
UNIVERSITÀ DI SALERNO - DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE

Quaderni di STATISTICA

VOLUME 9 - 2007

LIGUORI EDITORE

Volume 9, anno 2007

ISSN 1594-3739 (edizione a stampa)

Registrazione al n. 5264 del 6/12/2001 presso il Tribunale di Napoli
ISBN-13 **978 - 88 - 207 - 4209 - 6**

Direttore responsabile: Gennaro Piccolo

Prima edizione italiana Dicembre 2007
Finito di stampare in Italia nel mese di Dicembre 2007 da OGL - Napoli

Il C.F.E.P.S.R. si avvale per la stampa dei Quaderni di Statistica del contributo dell'Istituto Banco di Napoli - Fondazione.

La carta utilizzata per la stampa di questo volume è inalterabile, priva di acidi, a PH neutro, conforme alle norme UNI EN Iso 9706 ×, realizzata con materie prime fibrose vergini provenienti da piantagioni rinnovabili e prodotti ausiliari assolutamente naturali, non inquinanti e totalmente biodegradabili.

Indice

FORUM

Quaderni di Statistica
Vol. 9, 2007

A statistical approach for modelling Urban Audit Perception Surveys

Maria Iannario
Dipartimento di Scienze Statistiche, Università di Napoli Federico II
E-mail: maria.iannario@unina.it

Summary: This work presents a statistical approach to study, measure and evaluate the perception concerning the most serious problems which arise in urban areas. It could also be generalised to understand the perception of citizens after the introduction of new policies or the activation of different local or global practices. For this reason, we have introduced CUB models to analyse ordinal data resulting from a rank procedure of several items expressed by a group of raters. Specifically, we classify some issues as emergencies and ask a sample of people to rank them with respect to their relevance in terms of personal concern. The paper discusses the logic of the approach and some interpretative issues arising from the estimated CUB models with special reference to the environmental questions related to urban territory.

Keywords: Urban Audit, Ordinal data, CUB models.

1. Introduction

In the cognitive sciences, perception is the process of acquiring, interpreting, selecting and organizing sensory information. Its measurement is one of the primary elements of consciousness. In order to assess human perception, some researchers often rely on surveys concerning evaluation or preference.

In the classical setup, respondents are asked to select the option they prefer the most out of a discrete set of alternatives. The evaluations for several items are (generally) strongly correlated, for they express a common (positive/negative) judgement about the problem. Information usually may be obtained if respondents are asked to rank the set of alternatives instead. In this case, the answer is the result of a paired/sequential selection process which gives a more efficient estimation of the preferences.

Systematic approaches to the study of ranks data tend to be based on categorizations and on the location of an item in a given ordered list. Ranks expressed for each item are related to the subject's level of a given perception. Obviously, psychological, sociological and environmental causes influence (and bias) the responses. For instance, the extreme answers are more reliable than the middle ones. The respondents sometimes tend to find the central items less important and rank them with a reduced accuracy. One of the possible reasons is that the person has no experience of some items, and hence, he/she is not able to indicate a proper ranking order (Chapman and Staelin, 1982).

In general, ranks data can be found in several situations; in this paper ranking is used for studying how the urban environmental issues of Naples, a city of the Campania region, are experienced by citizens and how this perception is influenced by personal attitudes, and social and territorial variables.

This approach could be applied to gain an understanding of the citizens perception after the introduction of new policy or the activation of different local or global practices.

The purpose is to study the determinants of the degree of concern (judgment) of individuals towards a discrete set of items connected to serious problems which arise in a large urban area. Thus, a sample of people living in the Campania region was asked to rank a list of items, according to how relevant they felt the emergency.

This evaluation is a very complex task; the measure is referred to a person's behaviour and activity connected to emotional and cognitive effects. Thus, the objectives of the paper are testing whether some individuals are able to rank consistently all alternatives, analysing what is their perception as to *emergencies* and observing if there is a relationship of ranks with subject's covariates.

The paper is organized as follows: in the next section, we establish the notation for ordinal data and in section 3 we define CUB models and discuss their statistical relevance. Special emphasis will be devoted to the analysis and testing the covariates significance. Then, section 4 presents some empirical evidences of these models when applied to Urban Audit Perception Survey. Some final remarks conclude the paper.

2. Ordinal data and statistical models

Models are abstract and simplified representations of reality which involve variability due to unknown random factors (Lindsey, 1997). The main focus of these statistical models is to generate suitable data dependent structures in order to interpret, fit and forecast real data sets.

Models based on ordered responses represent a complex probability structure, that, in general, is included in the domain of qualitative and multinomial models. For this reasons, they require specific methods to avoid difficulties in the interpretation and/or loss of efficiency in the analysis of real data.

In the literature one of the most important examples of ordinal data models is given by Generalized Linear Model (GLM) introduced by Nelder and Wedderburn (1972), and McCullagh and Nelder (1989). This model has two main features: it involves a variety of distributions selected from the exponential family and it is related to transformations of the mean value through a *link function* which relates expectation to covariates.

The key point is that these approaches are able to model the log-odds of $P_r\,(R \leq r)$, the distribution function of the ranks, as a linear function of the subject's covariates. In this respect, a simple and parsimonious probability structure, that uses cumulative logits to relate ordinal data to subjects' covariates, is the *proportional odds model* (McCullagh, 1980; Agresti, 2002). Interest in this technique is often derived from the common need to report odds ratios of application.

This model presents some characteristics. First of all, the effects of the rank (choice) and of the subject are clearly separated; then, the interpretation of the parameters is related to the cumulative odds ratios (and their logarithms); finally, the role of the number of items (m) is only related to the normalization constraint.

The direct investigation of the psychological process that generates the choice mechanism among a discrete set of m alternatives, motivates the need for a new class of models which could overcome some limits of ordered logit models. This theory has led to a series of results based on a class of probability distributions, the MUB random variable (D'Elia and Piccolo, 2005a). Then, it has been generalized by including subjects' and objects' covariates by defining CUB models (Piccolo, 2006; Piccolo and D'Elia, 2007), and a marketing oriented study has been pursued by Iannario and Piccolo (2007).

The probability distribution of these models does not belong to the exponential class. They express directly the probability of an ordinal choice, and relate the parameters to the subjects' covariates without any reference to expectations, as in King *et al.* (2000). Thus, they fulfill the objective to explain, estimate and forecast in a simple way the probability $Pr\,(R = r)$ for an ordinal variable R assuming values in $\{1, 2, \dots, m\}$, for a given integer $m > 3$.

3. CUB Models: description and inference

Mc Fadden (1974) stated: "Application of the model should be limited to situations where the alternatives can plausibly be assumed to be distinct and weighed independently in the eyes of each decision-maker". Thus, it is important to exploit the background setting and the psychology of the choice for implementing a new statistical model.

Each choice results from a *paired* or *sequential* comparison of the items. It can be expressed as the result of two hierarchical steps: a *general* and immediate evaluation of the feeling (agree/disagree or indifferent), and a *specific* and reflective setting, within the global assessment, for expressing the final rank. Usually, when individuals are asked to rank alternatives or only to choose the most preferred option, the parameters of the choice model can be connected with two components: *feeling* and *uncertainty*. These are *continuous* and *latent* random variables that manifest themselves as discrete responses.

The first component, *feeling*, is the result of a continuous random variable that becomes a discrete one. Following a *latent variable approach* we assume that the observations are generated by an unobserved normally distributed random variable (say R^*), and we define a correspondence with a discrete ordinal random variable R by means of ordered threshold parameters to be estimated. Considering this idea, a suitable model for achieving the mapping of the unobserved continuous variable R^* into a discrete random variable R may be the *shifted Binomial* distribution with a probability mass function defined by:

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

The second component, *uncertainty*, depends on the specific values (knowledge, ignorance, personal interest, engagement, time spent to decide) concerning people. If the subject shows complete indifference towards a given item, it seems appropriate to model ranks by means of a discrete Uniform random variable U with a probability mass function defined by:

$$Pr\,(U = r) = \frac{1}{m}, \qquad r = 1, 2, \ldots, m\,. \tag{2}$$

Of course, in real cases, it is necessary to weight such an extreme situation to take account of the real expressed uncertainty.

The final result for interpreting the responses of the raters is a mixture model for ordered data in which we assume that the rank r is the realization of a random variable R, that is a mixture of an Uniform and a shifted Binomial random variable, defined on the support $r = 1, 2, \ldots, m$, with a probability distribution:

$$Pr(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} + (1-\pi)\,\frac{1}{m}\,, \tag{3}$$

and with $\pi \in [0, 1]$ and $\xi \in [0, 1]$. Since $m > 3$ is fixed and known, we will denote this random variable as $R \sim MUB(\pi, \xi)$.

We observe that the π parameter is inversely related to the weight of the uncertainty component: thus, $(1-\pi)/m$ is a *measure of the uncertainty* which spreads uniformly over all the support. The exact meaning of ξ, instead, changes with the setting of the analysis and, being the MUB random variable reversible, it depends on how the responses have been codified (the first position represents the higher feeling/concern and the last one the lower, or vice versa).

An important characterization of this random variable is that we can map a set of expressed rankings into an estimated model via (π, ξ) parameters. Thus, an observed complex situation of preferences/choices may be simply related to a unique point in the parametric space.

In this context, better solutions are obtained when we introduce the *covariates* for relating both the feeling and the uncertainty to the subject's features. Generally, covariates improve the model fitting and allow for better discrimination among different sub-populations and more accurate predictions. Moreover, the interpretation of the parameters estimates opens the discussion of different possible scenarios.

In fact, it is reasonable to assume that the main components of the choice mechanism change with the subjects' characteristics (covariates). Thus, CUB models are able to include explanatory variables which influence the position of different response choices. It is also interesting to analyze the values of the corresponding parameters conditioned to the covariate values.

Following a general paradigm (King *et al.*, 2000; Piccolo, 2006), we relate π and ξ parameters to the subjects' covariates through a logistic function, that is:

$$(\pi \mid \boldsymbol{y}_i) = \frac{1}{1+e^{-\boldsymbol{y}_i\boldsymbol{\beta}}} = \left[1 + e^{-\sum_{s=0}^{p} \beta_s\, y_{is}}\right]^{-1}; \tag{4}$$

$$(\xi \mid \boldsymbol{w}_i) = \frac{1}{1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}} = \left[1 + e^{-\sum_{t=0}^{q} \gamma_t\, w_{it}}\right]^{-1}. \tag{5}$$

The chosen mapping is the simplest ones among the many transformations of real variables into the unit interval and *a posteriori* it helps interpretation.

In the following, we refer to MUB random variables to denote the mixture probability distribution and to CUB models to the same structure when one or both parameters are explained by covariates[1].

[1] Of course, a $CUB(0, 0)$ model is just a MUB random variable.

In fact, when one or both parameters of the mixture probability distribution of a MUB random variable is explained by means of covariates, the model is denoted as $CUB(p, q)$, with parameter vector $\boldsymbol{\theta}$.

Specifically, we define:

CUB(0, 0)*: *without covariates and parameters, $\boldsymbol{\theta} = (\pi,\, \xi)'$;

CUB(p, 0)*: *with p covariates for π, $\boldsymbol{\theta} = (\boldsymbol{\beta}',\, \xi)'$;

CUB(0, q)*: *with q covariates for ξ, $\boldsymbol{\theta} = (\pi,\, \boldsymbol{\gamma}')'$;

CUB(p, q)*: *with (p,q) covariates for (π, ξ), $\boldsymbol{\theta} = (\boldsymbol{\beta}',\, \boldsymbol{\gamma}')'$.

Inferential issues for CUB models are tackled by maximum likelihood (ML) methods, using –as it is common for mixture models– the E-M algorithm (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000). The related asymptotic inference may be applied using the approximate variance and covariance matrix of the ML estimators (Piccolo, 2006).

Generally, in order to test the significance of the covariates in the model, we compare the log-likelihood of the $CUB(0, 0)$ model (without covariates) with the log-likelihoods of different CUB models with covariates. Thus, we will apply a Likelhood Raio Test considering $-2 \log \lambda \sim \chi^2_{(g)}$ where (g) represents the number of restriction.

Comparisons	g
$CUB(p, 0)$ *versus* $CUB(0, 0)$	p
$CUB(0, q)$ *versus* $CUB(0, 0)$	q
$CUB(p, q)$ *versus* $CUB(0, 0)$	$p + q$

It is worth to say that while the sequences:

$$CUB(0, 0) \leftarrow CUB(p, 0) \leftarrow CUB(p, q);$$
$$CUB(0, 0) \leftarrow CUB(0, q) \leftarrow CUB(p, q);$$

are nested ones, the models $CUB(p, 0)$ and $CUB(0, q)$ are not nested each other.

Model validation is a multifaceted activity, based on the interpretative content, the estimated coefficients, an effective information reduction and a sensible fitting. Then, in this regard, we prefer a normalized *dissimilarity index* defined by:

$$Diss = \frac{1}{2} \sum_{r=1}^{m} \mid f_r - p_r(\hat{\pi}, \hat{\xi}) \mid . \qquad (6)$$

where f_r and $p_r(\hat{\pi}, \hat{\xi})$ are the observed relative frequencies and the estimated probabilities from the CUB model, respectively[2].

The models we have introduced are able to fit and explain the behaviour of a univariate rank variable. Instead, we realize that the expression of a complete ranking list of m objects/items/services by n subjects should require a multivariate setting. Thus, the analysis that will be pursued in this paper must be interpreted as a marginal one since we will study the ranks distribution of a single item without reference to the ranks expressed towards the remaining ones. Although the ranks cannot be strictly independent[3], when the number of the objects is not extremely limited (in our applications, $m = 9$), we may consider the rank towards an item as conditionally independent from the others.

4. Urban Audit Perception Survey: some empirical evidence

In this case study we analyse the degree of concern[4] of individuals over a discrete set of 9 items (*political patronage and corruption*; *organized crime*; *unemployment*; *environmental pollution*; *public health short-*

[2] The index $Diss \in [0, 1]$ possesses a simple interpretation: it measures the proportion of subjects which should be moved among the cells of the frequency distribution in order to achieve a perfect fit. Moreover, according to the experience (based on several data sets and some preliminary bootstrap studies), when $Diss < 0.09$, we can be adequately satisfied from the fitting point of view (Piccolo, 2006).

[3] All the realizations of the multivariate random variable $(R_1, R_2, \ldots, R_m)$ are permutations of the first m integers and, thus, they sum $m(m+1)/2$.

[4] Notice that, in our case study, the *feeling* parameter is a direct measure of the degree of *concern* about a problem since we are asking respondents to rank the most serious problems in decreasing order of worry.

comings; *petty crimes*; *immigration*; *streets cleanness and waste disposal*; *traffic and local transport*).

We classify these issues as “emergencies” and we ask people to rank them in a decreasing order with respect to the worry they generate. This kind of audit survey, submitted for the first time in December 2004, was repeated in December 2006 to an homogeneous sample. We refer to the last data set for the modelling approach[5].

Several information related to the subjects have been collected: gender, age, diploma, residence, working condition, etc. At the end, $n = 419$ complete questionnaires form the basis of the following analyses.

4.1. Exploratory Data Analysis

In Table 1 we list the expressed frequency distributions of ranks for each item (bold marks are the maximum frequency for each row) and we plot each of them in Figure 1, in order to show the different shapes.

To synthesize the ranks for the items, we present some location measures in Table 2. Of course, since the rank is just a value for a qualitative judgment, the arithmetic mean should not be computed; however, its interpretation may be still accepted if we consider the ranks as a *proxy* of the (latent) continuous variable which subjects express towards the single item. Moreover, some doubt may be generated by comparing only mode and/median of the expressed ranks.

For instance, we notice that *Traffic and local transport* and *Streets cleanness and waste disposal* have the same mode (7) but different means and medians. The first (*Streets cleanness and waste disposal*) is perceived as less dangerous than the second one (*Traffic and local transport*).

[5] The survey has been submitted in December 2006 to students attending a lecture in the Faculty of Political Science, University of Naples Federico II. Thus, it can not be considered as a random sample of the population living in the area; however, the peculiar socio-demographic structure of the respondents gives some relevance to survey. In fact, the sample is made by 182 (43.44%) males and 237 (56.56%) females, with an age between 18 and 57 years. They are students mostly with scientific or technical diploma (36.99% and 33.5%, respectively), originated from the metropolitan area (38.68%). About 40.09% are full-time students, thus most of the respondents are part-time or full-time workers.

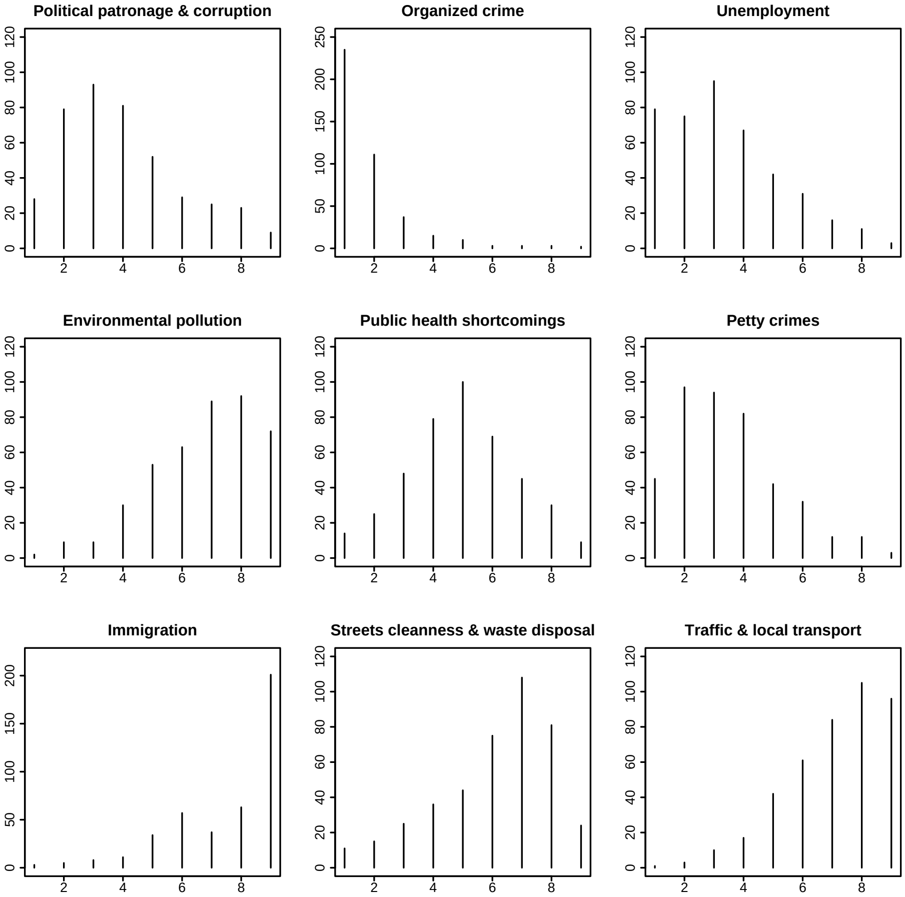

Figure 1. Frequency distributions of expressed ranks

Table 1. Main problems frequency distributions

Main problems	1	2	3	4	5	6	7	8	9
Political patronage	28	79	**93**	81	52	29	25	23	9
Organized crime	**235**	111	37	15	10	3	3	3	2
Unemployment	79	75	**95**	67	42	31	16	11	3
Pollution	2	9	9	30	53	63	89	**92**	72
Public health	14	25	48	79	**100**	69	45	30	9
Petty crimes	45	**97**	94	82	42	32	12	12	3
Immigration	3	5	8	11	34	57	37	63	**201**
Streets and waste	11	15	25	36	44	75	**108**	81	24
Traffic-transport	1	3	10	17	42	61	84	**105**	96

Table 2. Location Indexes

Main problems	*Mode*	*Median*	*Average*
Political patronage	4	3	3.96
Organized crime	1	1	1.81
Unemployment	3	3	3.35
Pollution	7	8	6.72
Public health	5	5	4.95
Petty crimes	3	2	3.48
Immigration	8	9	7.55
Streets and waste	7	7	6.08
Traffic-transport	7	8	7.09

A second measure for ordinal data should be related to their variability, and we prefer[6] the index of Laakso and Taagepera (1979) in its normalized version in $[0, 1]$:

$$\mathcal{A} = \frac{1}{m-1}\left\{\left(\sum_{i=1}^{m} f_i^2\right)^{-1} - 1\right\}. \tag{7}$$

[6] A discussion about the usefulness of this measure for CUB models with reference to preliminary estimation and parameters interpretation is reported in D'Elia and Piccolo (2005b).

Table 3 summarizes this measure for both 2004 and 2006 surveys. It seems evident that variability has homogeneously increased in the second year, and in significant amount for several instances. Instead, we observe that *Organized crime* shows a limited heterogeneity denoting a distribution of responses strongly concentrated on few values: Table 1 and Figure 1 confirm these results. Specifically, we register that 83% of respondents ranked this item as 1 or 2. Thus, people reacted in a strongly similar way towards this emergency and we should expect low uncertainty in the CUB model for their responses.

Table 3. Index of Laakso and Taagepera

Main Problems	$\mathcal{A}(2004)$	$\mathcal{A}(2006)$	% *increase*
Political patronage and corruption	0.6400	0.6873	6.88
Organized crime	0.1616	0.1918	15.70
Unemployment	0.5078	0.6452	21.30
Environmental pollution	0.5757	0.6211	7.30
Public health shortcomings	0.5625	0.6837	17.74
Petty crimes	0.5183	0.6045	14.26
Immigration	0.2270	0.3107	26.93
Streets cleanness and waste disposal	0.5170	0.6415	19.40
Traffic and local transport	0.5337	0.5363	0.48

4.2. CUB models for concern towards urban problems

In this subsection, we propose the application of CUB models to our data set. In Table 4, we list ML estimates (with asymptotic standard errors es) and related log-likelihoods and dissimilarity indexes. Notice that log-likelihoods are inverse measures of goodness of fit, given that all models are estimated with the same sample size The previous models are generally satisfying, as confirmed by the significance of the estimated parameters and the low values of *Diss*; some problems might be generated by *Pollution*, *Traffic and local transport* and *Immigration*, where *Diss* ≥ 0.13.

Table 4. Estimation of CUB models for main urban problems

Main problems	$\hat{\pi}$	$es(\hat{\pi})$	$\hat{\xi}$	$es(\hat{\xi})$	*log-lik*	*Diss*
Political patronage	0.629	*0.045*	0.704	*0.013*	-843.215	0.045
Organized crime	0.898	*0.021*	0.936	*0.005*	-522.663	0.055
Unemployment	0.674	*0.045*	0.774	*0.012*	-830.845	0.115
Pollution	0.751	*0.044*	0.236	*0.011*	-815.021	0.130
Public health	0.719	*0.047*	0.506	*0.013*	-838.779	0.036
Petty crimes	0.768	*0.040*	0.736	*0.011*	-804.718	0.061
Immigration	0.567	*0.035*	0.037	*0.007*	-719.248	0.154
Streets and waste	0.665	*0.040*	0.293	*0.013*	-831.655	0.075
Traffic-transport	0.820	*0.038*	0.197	*0.010*	-773.402	0.136

In fact, these results might be caused by two *sub-populations* of respondents.

To deepen the analysis and the comparison of the estimated CUB models, we represent them in a parametric space, as in Figure 2. In this way, the measures of concern and uncertainty on different variables[7] and their relative position are clearly enhanced.

Specifically, Figure 2 shows that respondents classify the issues in four clusters, where the first (CRI) and the last (IMM) expresses the most and the least serious concern; instead, the second group (DIS, MIC, CLI, MAL) is referred to personal care while the third (PUL, INQ, TRA) includes topics related to environment.

Moreover, on the same space, we may consider the interaction of these problems with gender and age. In fact, gender is a dichotomous variable and age has been made dichotomous by discriminating respondents under and above 30 years.

Figure 3 helps in the interpretation of the responses with respect to the gender both for the concern and the uncertainty. Thus, for instance, it seems that women are more concerned than men with respect to health

[7] In Figg.2-4, the emergencies has been labelled in the following way: *Political patronage and corruption*=CLI; *Organized crime*=CRI; *Unemployment*=DIS *Environmental pollution*=INQ; *Public health shortcomings*=MAL; *Petty crimes*=MIC; *Immigration*=IMM; *Streets cleanness and waste disposal*=PUL; *Traffic and local transport*=TRA.

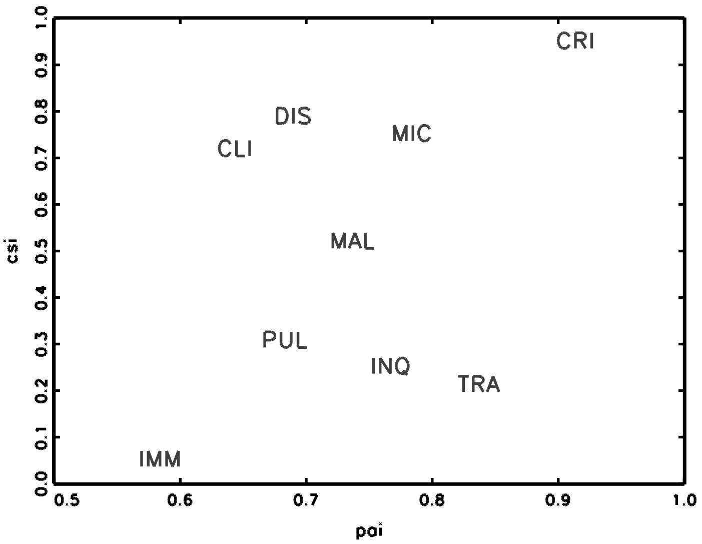

Figure 2. Parametric representation of the estimated CUB models

problems. Similar consideration might be applied to uncertainty, where large differences are registered for *Political patronage and corruption* and *Streets cleanness and waste disposal* (men give more uncertain answers) and for *Environmental pollution* (women are more uncertainty). In Figure 4 we have similar results by differentiating respondents whose age is under and over 30 years (labelled as 0 and 1, respectively). Thus, it turns out that concern is almost the same, while uncertainty is the main difference caused by the age. Specifically, we observe that uncertainty decreases in elderly people for problems of general interest (as *Organized crime*, *Public health shortcomings*, *Traffic and local transport*, for instance).

5. CUB models with covariates

The introduction of subjects' characteristics for explaining both feeling and the uncertainty in the previous models improves the results and the interpretation. More specifically, we will focus our attention on the variables which are connected to environmental problems in relation to

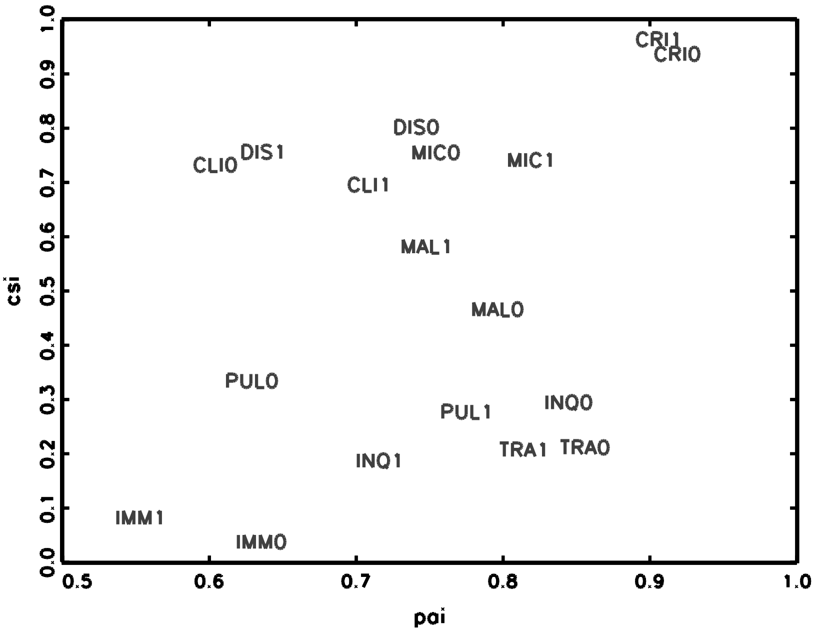

Figure 3. Estimated CUB parameters for men (label=0) and women (label=1)

urban territory (*Environmental pollution, Streets cleanness and waste disposal, Traffic and local transport*). They are related to urban policy and, generally, are not considered among the first worries in the students' responses.

- ***Environmental pollution***

For this item, the significant covariates are *regular job* to improve the model for π (degree of uncertainty) and *gender* and *age*[8] to explain ξ (the degree of concern). Table 5 shows the hierarchical estimated models from $CUB(0,0)$ (without covariates) until to $CUB(0,2)$.

The parameters are all significant and consistent with the corresponding interpretations. Then, we use the difference between the deviances

[8] For all the models which we will estimate, we prefer to use $\ln(age)$ instead of *age* since, in general, the logarithmic transformation reduces the variability of the covariates.

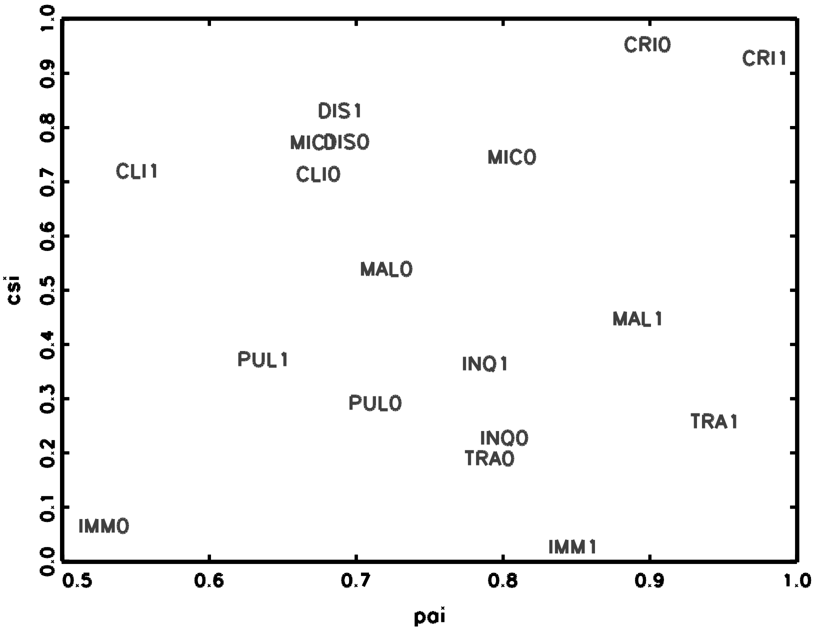

Figure 4. Estimated CUB parameters for people under (label=0) and over 30 years (label=1)

as a test for preferring the last model, and we compare this difference with the critical levels of a $\chi^2_{(g)}$, where g is the difference between the parameters of the models. Specifically, we get:

$$2(\ell_{02} - \ell_{00}) = 2\,(-795.540 - (-815.021)) = 38.962\,;$$

this value is highly significant if compared with the quantiles of a $\chi^2_{(g)}$ random variable with $g = 2$.

Thus, we retain the last model for it gives a significant and convincing interpretation of our data. It results that the expected rank of *Environmental pollution* is modified by the covariates gender and age, according to

Table 5. CUB models for Environmental pollution

Models	$\hat{\pi} = \pi(regjob)$	$\hat{\xi} = \xi(gender; ln(age))$	*log-lik*
CUB(0, 0)	$\hat{\pi} = \ \ 0.751$ *(0.045)*	$\hat{\xi} = \ \ 0.236$ *(0.012)*	-815.021
CUB(1, 0)	$\hat{\beta}_0 = \ \ 1.547$ *(0.334)* $\hat{\beta}_1 = -1.141$ *(0.464)*	$\hat{\xi} = \ \ 0.233$ *(0.012)*	-811.930
CUB(0, 2)	$\hat{\pi} = \ \ 0.792$ *(0.040)*	$\hat{\gamma}_0 = -3.357$ *(0.538)* $\hat{\gamma}_1 = -0.378$ *(0.124)* $\hat{\gamma}_2 = \ \ 0.745$ *(0.163)*	-795.540

the formula[9]:

$$\mathbb{E}\left(R \mid gender, age\right) = 8.168 - \frac{6.336}{1 + e^{3.357 + 0.378\, gender - 0.745\, \ln(age)}}\,.$$

Given the dichotomous character of the covariate gender, it is possible to express this expectation in a more direct way:

$$\mathbb{E}\left(R \mid age\right) = \begin{cases} 8.168 - \dfrac{6.336}{1 + 28.703\,(age)^{-0.745}}, & \text{if gender} = 0 \text{ (men);} \\ 8.168 - \dfrac{6.336}{1 + 41.388\,(age)^{-0.745}}, & \text{if gender} = 1 \text{ (women);} \end{cases}$$

Thus, this formula confirms that men are more concerned than women (Figure 5). Moreover, the expected rank reduces for increasing ages, that is elderly people are more concerned than young with regard to *Environmental pollution*.

- ***Streets cleanness and waste disposal***

We consider the relationship among the perception of this item and job, age and gender. The main result is that regular job is a significant covariate for explaining π and gender and (logarithm of) age for explaining ξ, as shown in Table 6.

Then, in the last model we observe a barely significant level for the covariate $\ln(age)$ of the ξ parameter; also, the increase in log-likelihoods

[9] We are letting $m = 9$ in the general formula of the expectation of a CUB model computed with the estimated parameters.

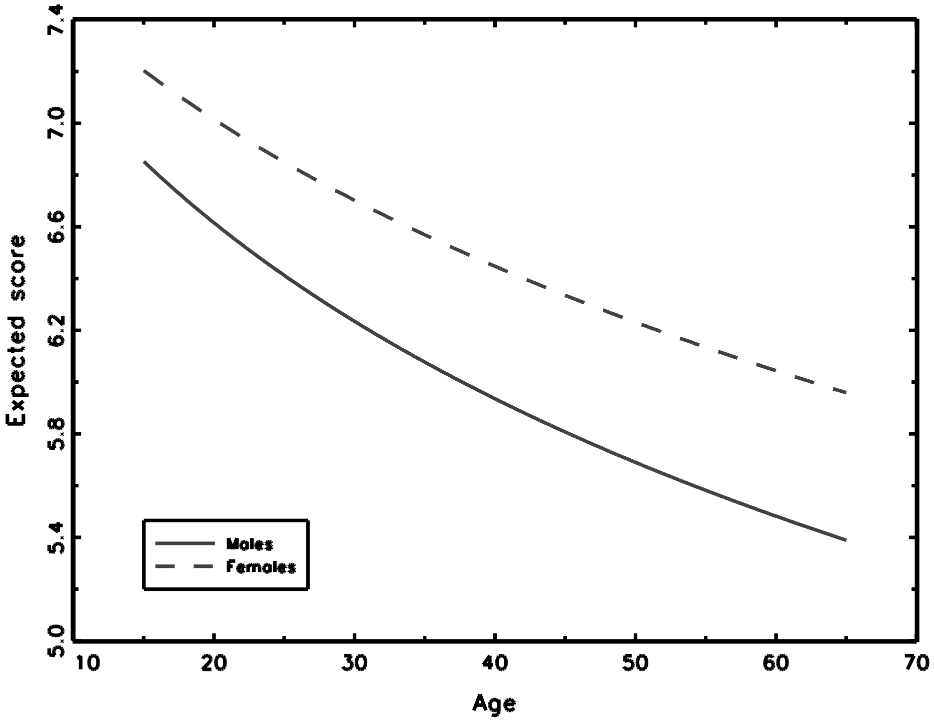

Figure 5. Environmental pollution: expected ranks as function of gender and age

from $CUB(1,1)$ to $CUB(1,2)$ is not satisfactory. In fact, if we plot the expected rank for varying ages, we observe (Figure 6) that most of variation is accounted for the presence/absence of a regular job while the age effect does not seem relevant.

As a consequence, for this item, we choose a $CUB(1,1)$ model where regular job acts as a covariate for π and only gender is a significant covariate for ξ. It improves the fitting with respect to the previous models, all the parameters are significant and the increase in log-likelihoods from $CUB(0,0)$ to $CUB(1,1)$ model is relevant if we compare:

$$2(\ell_{11} - \ell_{00}) = 2\,(-822.733 - (-831.655)) = 17.844$$

with the quantiles of a $\chi^2_{(g)}$ random variable with $g = 2$.

Table 6. CUB models for Streets cleanness and waste disposal

Models	$\hat{\pi} = \pi(regjob)$	$\hat{\xi} = \xi(gender, ln(age))$	*log-lik*
CUB(0, 0)	$\hat{\pi} =$ 0.665 *(0.044)*	$\hat{\xi} =$ 0.293 *(0.013)*	-831.655
CUB(1, 0)	$\hat{\beta}_0 =$ 1.265 *(0.278)* $\hat{\beta}_1 = -1.439$ *(0.421)*	$\hat{\xi} =$ 0.288 *(0.013)*	-825.251
CUB(0, 2)	$\hat{\pi} =$ 0.696 *(0.044)*	$\hat{\gamma}_0 = -1.939$ *(0.565)* $\hat{\gamma}_1 =$ 0.387 *(0.176)* $\hat{\gamma}_2 = -0.307$ *(0.118)*	-825.732
CUB(1, 1)	$\hat{\beta}_0 =$ 1.291 *(0.278)* $\hat{\beta}_1 = -1.353$ *(0.428)*	$\hat{\gamma}_0 = -0.764$ *(0.085)* $\hat{\gamma}_1 =$ 0.266 *(0.118)*	-822.733
CUB(1, 2)	$\hat{\beta}_0 =$ 1.294 *(0.276)* $\hat{\beta}_1 = -1.272$ *(0.433)*	$\hat{\gamma}_0 = -1.859$ *(0.622)* $\hat{\gamma}_1 =$ 0.353 *(0.198)* $\hat{\gamma}_2 = -0.259$ *(0.117)*	-821.190

In the preferred model both covariates are dichotomous and to see their effect on the expected rank we refer to the scheme indicated in Table 7.

Table 7. Expected ranks, gender and regular job

Gender	*Regular Job*	*Expected rank*
Males	Yes	5.706
Females	Yes	5.918
Males	No	6.143
Females	No	6.487

From this model, we deduce that people without regular job consider this item less relevant than the others; similarly, men are more concerned towards this problem than women. Moreover, from the previous table, the most prominent effect on the graduation of the worry is due to the presence/absence of a regular job.

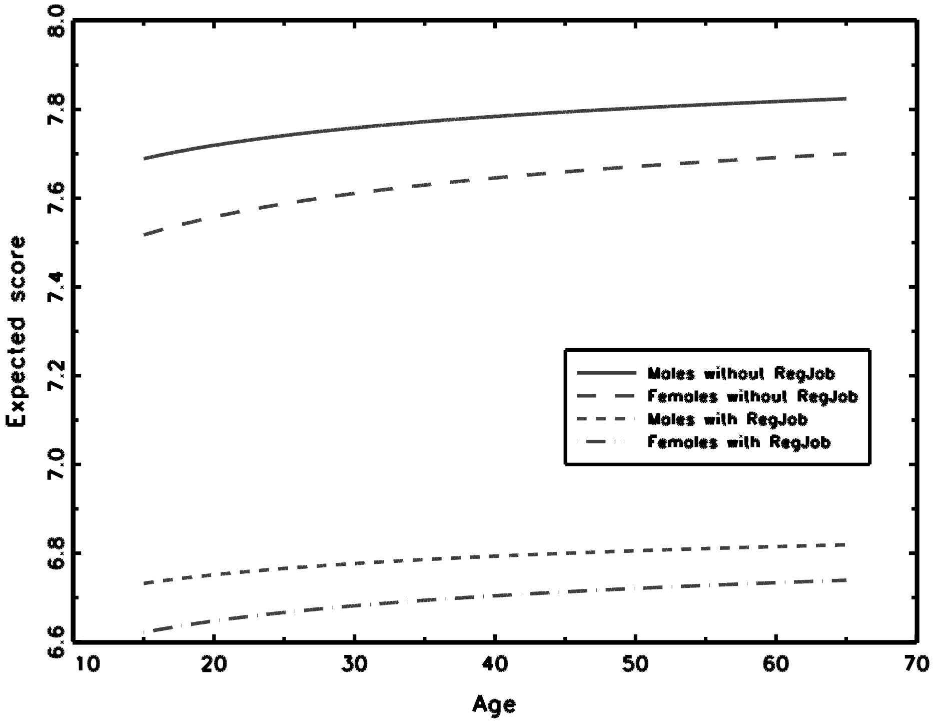

Figure 6. Streets cleanness and waste disposal: expected ranks from model $CUB(1,2)$

- ***Traffic and local transport***

The last item which we will consider is *Traffic and local transport* where the best fitting was obtained by a $CUB(0,2)$ model with (logarithm of) age and neomelodic music[10] as explanatory covariates for the ξ parameter. Then, the estimated CUB models are summarized in Table 7.

The model explains that the concern towards *Traffic and local transport* increases with age but it is not so high among people which prefers neomelodic music. We may confirm this aspects if we plot (Figure 7) the

[10] This is a dummy variable that assumes value 1 if respondents declared to love neomelodic music, a kind of music quite popular in the Campania region. It seems a *proxy* variable related to the socio-cultural context of the respondents.

Table 8. CUB models for Traffic and local transport

Models	$\hat{\pi}$	$\hat{\xi} = \xi(ln(age), neomelodic)$	*log-lik*
CUB(0, 0)	$\hat{\pi} = 0.820$ *(0.038)*	$\hat{\xi} = \quad 0.197$ *(0.010)*	-773.402
CUB(0, 2)	$\hat{\pi} = 0.830$ *(0.037)*	$\hat{\gamma}_0 = -2.369$ *(0.505)* $\hat{\gamma}_1 = -0.394$ *(0.130)* $\hat{\gamma}_2 = -0.341$ *(0.154)*	-766.391

expected rank for this item as a function of these covariates according to the formula:

$$\mathbb{E}\left(R \mid age, neomelodic\right) = 8.320 - \frac{6.640}{1 + e^{2.369 + 0.394\,\ln(age) - 0.341\,neomelodic}}\,.$$

The chosen variables have a sensible effect on the maximization of the likelihood function, and the difference between the deviances amounts to:

$$2(\ell_{02} - \ell_{00}) = 2\left(-766.391 - (-773.403)\right) = 14.024$$

which is significant if compared with the quantiles of a $\chi^2(g)$ random variable with $g = 2$.

6. Concluding remarks

In this paper, we obtained some results about direct inference on the expressed ranks by the introduction of feeling and uncertainty components that drive the choice as explicit parameters in CUB models, with and without subjects' covariates. The experiments have confirmed that this innovative approach gave a different perspective to evaluate the psychological process that drives the choice, the selection of an item and the ranking procedure.

Then, the introduction of covariates is an added value for interpreting, clustering and discriminating sub-populations with respect to a fixed item, and this may open new perspectives for the evaluation of territorial data. By using this class of models, we may understand and measure the effects of new policies and practices, and to differentiate them with respect to (local/global) area. Moreover, it permits to classify the subjects' position and their psychological behaviour with respect to a given stimulus. This conceptualization may help policy makers to improve economic and political strategy of good governance and to measure the perception of their effect.

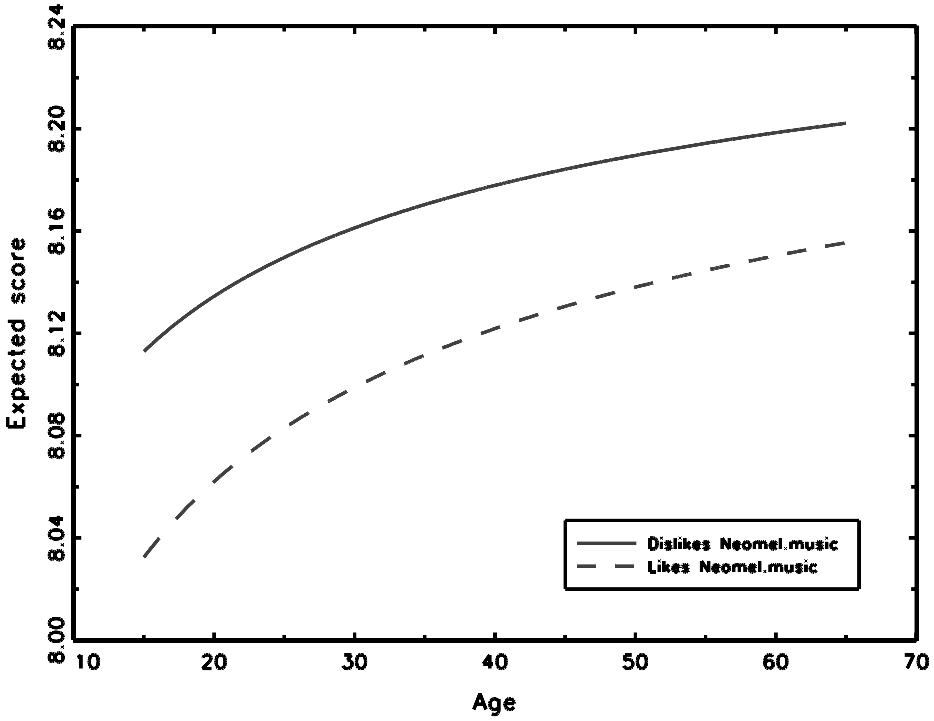

Figure 7. Traffic and local transport: expected ranks from model $CUB(0,2)$

The empirical evidence supported by a specific survey aimed at measuring the worry about urban problems in a large metropolitan area has confirmed its usefulness on a real data set.

Acknowledgements: This work has been written while the Author received a research grant by Dipartimento di Scienze Statistiche, Università di Napoli Federico II and has been also supported by the 2006 PRIN-MIUR research project: “Stima e verifica di modelli statistici per l'analisi della soddisfazione degli studenti universitari". Helpful suggestions by the Editor and referees are gratefully acknowledged.

References

Agresti A. (2002), *Categorical Data Analysis*, 2^{nd} edition, J. Wiley & Sons, New York.

Chapman R., Staelin R. (1982), Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, *Journal of Marketing Research*, 19, 288-301.

D'Elia A., Piccolo D. (2005a), A mixture model for preference data analysis, *Computational Statistics & Data Analysis*, 49, 3, 917-934.

D'Elia A., Piccolo D. (2005b), Uno studio sulla percezione delle emergenze metropolitane: un approccio modellistico, *Quaderni di Statistica*, 7, 121-161.

Iannario M., Piccolo D. (2007), A new statistical model for the analysis of customer satisfaction, *Quality Technology and Quantitative Management*, submitted for publication.

King G., Tomz M., Wittenberg J. (2000), Making the most of statistical analyses: improving interpretation and presentation, *American Journal of Political Science*, 44, 341-355.

Laakso M., Taagepera R. (1979), Effective number of Parties: a measure with application to West Europe, *Comparative Political Studies*, 12, 3-27.

Lindsey K.L. (1997), *Applying Generalized Linear Models*, Springer-Verlag, New York.

McCullagh P. (1980), Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society, Series B*, 42, 109-142.

McCullagh P., Nelder J. A. (1989), *Generalized Linear Models*, 2^{nd} edition, Chapman and Hall, London.

McFadden D. (1974), Conditional logit analysis of qualitative choice behavior, in: P. Zarembka ed., *Frontiers in Econometrics*, Academic Press, New York, 105-142.

McLachlan G., Krishnan G. J. (1997), *The EM Algorithm and Extensions*, J. Wiley & Sons, New York.

McLachlan G., Peel G. J. (2000), *Finite Mixture Models*, J. Wiley & Sons, New York.

Nelder J. A., Wedderburn R. W. M. (1972), Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384.

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

Piccolo D., D'Elia A. (2007), A new approach for modelling consumers' preferences, *Journal of Food Quality and Preferences*, doi:10.1016 / j.foodqual. 2007.07.002.