Quaderni di Statistica
Vol. 10, 2008

A class of models for ordinal variables with covariates effects

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

Summary: This paper introduces statistical models aimed at synthesising personal evaluations concerning the main problems of an urban area. They reflect the changing nature of environmental inquiry in the field of urban issues based on dwellers' perception. In fact, they offer integrated approaches to understanding how the final response on urban problems has been generated by the subject's intrinsic awareness (*feeling*) and several external circumstances (*uncertainty*). This analysis is pursued by means of a new different approach to ordinal data which looks for cultural, socio-economic and psychological determinants of responses through the introduction of subjects' covariates. Thus, we model expressed ranks and discuss their interrelationships; then, by using the modelling structures we estimate probabilities and expectations given the characteristics of the respondents. In this way, we are able to perform inferences on the choice mechanism from the observed results, and we gain experience for predicting future behaviours.

Keywords: Ordinal data, Perception evaluation, CUB Models.

1. Introduction

The perception of the main problems of an urban area is an important issue for understanding and questioning inter-relationships between dynamic factors. Differences in evaluation of problems must be understood at the level of social identities. The basic knowledge of this aspect includes cultural, political, socio-economic, strategic and composite aspects which provide a reliable foundation for interpreting current and future developments.

We refer to the perception of risk society thesis (Beck, 1992) which

succeeds in describing the emergence of a risk ethos, the development of a collective risk identity and the formation of communities hold together by an increasing vulnerability to risk. If we consider that there has been a reconfiguration in the way risk/danger is identified, evaluated, communicated and governed, we can expands the traditional concept of risk (interpreted as the product of the probability of an adverse event and the magnitude of the consequences) to include subjective perception, inter-subjective communication and social experience of living in a risk/dangerous environment (Loewenstein *et al.*, 2001).

These considerations stimulate attention to how the very nature of risk in an urban destination has been transformed and how the origin and impact of risk have been reassessed.

The study we propose in this context would examine discussions on the origin and impact of risk. Specifically, we present data on different surveys aimed at measures the perception of urban issues in a specific context by means of discrete choice models. During the months of December 2004, 2006 and 2007 we asked to sampled dwellers to rank several items (*political patronage and corruption; organized crime; unemployment; environmental pollution; public health shortcomings; petty crimes; immigration; streets cleanness and waste disposal; traffic and local transport*), concerning the urban area in which they live, in a decreasing order with respect to the worry/anxiety they generate.

The questionnaire has been submitted in December 2004 and 2006 to homogeneous samples, consisting of students attending University lectures in the Faculty of Political Science, University of Naples Federico II. As a consequence, it can not be considered as a random sample of the population living in the area; however, our analysis can be exploited as a paradigma for similar studies based on a larger audience and on a stratified sampling scheme. Above all, it is a benchmark for raising interpretative and methodological problems on modelling ordinal data in different area (Iannario, 2007a).

Subsequently, in December 2007 the starting point has always been the students attending the same Faculty but the survey has been submitted to a larger audience including relatives and friends, with a sort of a *snowball sampling* scheme. As a consequence, we may assess correct

comparisons only for first two samples. Moreover, it is important to realize that the last sampling has been planned just before the well known upsurge (January 2008) of the crisis concerning *waste disposal* in Naples, as reported by media.

Although the questionnaire involves nine issues, we deepen here two aspects of the risk perception: *organized crime* and *waste disposal*, since the behaviour of respondents with regard to them seems completely different. In fact, public opinion and mass media quite often relate these issues for the implication of *organized crime* in current environmental regulations, laws, strategies and policies about *waste disposal*. The starting hypothesis is that *organized crime* is entrepreneurial in nature and that the dynamics of the market space connected to waste provide one of the main environment and explanation for *organized crime*. Thus, it is interesting to build models able to highlight the significance and the changes in perception conditioned by different profiles of subjects.

In this paper, we will present statistical structures that are able to face with these kind of issues. Moreover, one of the main point we will deal with is that a ranking approach may be considered as an indirect and conditioned evaluation if we analyse a single component of the set by univariate statistical methods.

This aspect is a critical issue and thus we devote to this discussion some space in section 2. Then, we will briefly introduce a recent class of models (defined CUB) whose main features are enhanced from both interpretative and statistical frameworks; specifically, we relate the preferred option of the sampled respondents to relevant covariates by testing their significance by asymptotic results (section 4). Empirical evidence related to *Organized crime* and *Waste disposal* are discussed in section 5; they support the usefulness of the approach. Further considerations and some concluding remarks end the paper.

2. Statistical models for ordinal data

There are several contexts where people are asked to express their judgements or to make a selection in a definite list of known m objects/items/services. In fact, although both schemes produce ordered an-

swers in the set $\{1, 2, \ldots, m\}$, we have to distinguish between the assignment of a well defined position in a list (*ranking*) and the expression of an evaluation about some fixed item (*rating*). To establish notations, we assume that rank 1 denotes the first choice, and thus it may be the preferred issue, the worst result, the extreme worry, and so on, according to the question submitted to sampled people[1].

Specifically, in the *ranking approach* the answer of a subject is a permutation of the first m integers, that is a vector of numbers, according to the degree of preference of the m objects. The procedure for assessing a rank to a given item in a finite discrete set of similar alternatives requires an elicitation strategy, based on either sequential choice of objects or pairwise comparison of items. In this context, classical statistical analyses look for adequate models of permutations or latent variables that motivates the stated arrangement (Fligner and Verducci, 1999; Marden, 1995; Joreskog and Moustaki, 2001; Moustaki, 2003).

Instead, in the *rating approach* the answer of the subject to a fixed item is a single number. The procedure is the output of a personal judgement aimed to quantify the received "stimulus" with reference to the item. Several situations encompass this case and manifest themselves with different features: marks, evaluation scores, threshold levels, hedonic scales, degree of adhesion or awareness, and so on. The standard approach includes several variants of Generalized Linear Models (GLM, see: McCullagh, 1980; Agresti, 2002; Dobson and Barnett, 2008) and it relates the log-odds of cumulative probabilities to linear models for covariates.

Formally, in *ranking analyses* we have a discrete multivariate random variable whose components explain the stated preferences towards m fixed objects; instead, in *rating analyses* we study a univariate random variable with support $\{1, 2, \ldots, m\}$ which expresses the level of consensus of several subjects towards a given item.

A fundamental issue is that the observed components of a *ranking* study are not independent since any admissible vector is strictly a permutation of the first integers; on the contrary, any single answer of a *rating* study expresses the subject's evaluation and it can assume any value on

[1] This assumption is not restrictive since different options may be dealt with a reverse ordering; in fact, the models of next section satisfy a reversibility property.

the given support.

The point is that we may consider the distribution of the ranks given to a fixed object as the realizations of a marginal random variable. It can assume any value on the support $\{1, 2, \ldots, m\}$ depending on the location that sampled respondents attribute to this object. In a sense, we are maintaining that a low (high) rank denotes high (low) confidence with the object; then, the marginal distribution of the ranks given to the chosen object is *de facto* an *indirect, ordered* and *conditioned* evaluation towards the object. Explicitly, we are saying that a marginal ranking analysis produces an *indirect* evaluation since people are not immediately expressing a score for the object; then, it is an *ordered* evaluation as it conveys the answer of the subject on a numeric scale related to the intensity of the perceived evaluation; finally, it is a *conditioned* evaluation as the result is obviously limited by the assignments given to the others objects. Notice that the independence of the sampled values is preserved anyway.

As a consequence of this approach, we will denote by R the univariate random variable generated as a marginal distribution of the multivariate rank distribution. In a sample, we observe $(r_1, r_2, \ldots, r_n)'$, where each $r_i,\ i = 1, 2, \ldots, n$ expresses the position of the object in the list given by the n respondents. Thus, $P_r\,(R = r)$ is the probability that, for a given object, a respondent denotes the integer $r \in \{1, 2, \ldots, m\}$.

In this context, we introduce a class of random variables in order to take into account the discrete nature of the answers and to relate them to subjects' characteristics without referring to a transformation of probability distributions (as log-odds, adjacent and continuation probabilities, generally accepted in the GLM framework: McCullagh and Nelder, 1989). In our approach, assuming that the generated process leading to the proposed models be consistent, there is a direct probability statement about the answers and an immediate link with the covariates; this fact should simplify the interpretation, improve the fitting and lead to parsimonious models.

3. Features of CUB Models

Let us consider situations where people are asked to express their perceived feeling (worry, liking, agreement and so on) toward a fixed object/item/problem by putting it in a list of m similar issues. As we stated in section 2, we interpret this rank as a conditioned evaluation by studying the observed marginal answer $r \in [1, m]$.

This indirect evaluation can be thought as the final outcome of a psychological process of judgement, where the investigated trait is intrinsically continuous but -for convenience- it is expressed in a discrete way. Then, we conjecture that this process is made up of two main components: the assessment of the questioned *feeling* and the fuzziness/*uncertainty* that accompanies any human choice.

Previous studies and several empirical evidence show that the *shifted Binomial* random variable is an adequate probabilistic model for representing the discrete version of a latent judgement process, mapping a continuous and unobserved evaluation into a discrete set of values belonging to $\{1, 2, \ldots, m\}$. An important feature of this correspondence is that it complies with the intrinsic nature of observed choices; moreover, it is extremely flexible. Actually, by varying cutpoints, we are able to fit observed data with marked skewness and peakedness as well as symmetric or flat distributions, with modal values located everywhere on the support.

On the other side, the *discrete Uniform* random variable is a suitable building block for describing the inherent uncertainty of a discrete choice process, for it represents the model with maximum entropy on a finite discrete support. Thus, any observed uncertainty contained in the data may be weighted with respect to this extreme case.

On this basis, D'Elia and Piccolo (2005) and Piccolo (2006) have considered r as a realization of the mixture random variable R of these discrete distributions, that is a mixture of Uniform and Shifted Binomial random variables. These models have been called CUB as they are able to include also significant covariates.

Formally, for a given $m > 3$, the probability mass function of R is

defined by:

$$P_r\left(R=r\right)=\pi\binom{m-1}{r-1}(1-\xi)^{r-1}\xi^{m-r}+(1-\pi)\,\frac{1}{m}\,,\quad r=1,2,\ldots,m\,,$$

with $\pi\in(0,1]$ and $\xi\in[0,1]$. Recently, Iannario (2008) proved that this model is identifiable.

It is immediate to realize that π is a parameter inversely related to the weight of the uncertainty component, and $(1-\pi)/m$ is a *measure of the uncertainty* which spreads uniformly over all the support. Instead, the interpretation of ξ changes with the setting of the analysis since it depends on how the responses have been coded (the first position represents the higher feeling/concern and the last one the lower, or vice versa). Thus, according to the context, we interpret ξ as *degree of risk perception*, *index of selectiveness/awareness*, *measure of worry*, *intensity of pain*, and so on (Iannario and Piccolo, 2009).

Better solutions are usually obtained when we introduce the subjects' *covariates* aimed at relating both the feeling and the uncertainty to the respondents's features. If they are significant, covariates improve model fitting and allow for better discrimination among different sub-populations (for instance, via dummies covariates, as in Iannario, 2007b, or by clustering methods, as in Corduas, 2008a,b).

In addition, CUB models are effective tools for assessing the role of explanatory variables in determining different responses of the subjects. As it will become evident in section 5, the study of expected evaluations conditioned to the covariates values may allow to forecast future behaviors and also to study differential impacts of covariates.

In this regard, we observe that moments of R are not relevant since the sequence $\{1,2,\ldots,m\}$ is just a *proxy* for a qualitative ordering, and no metric property should be attached to these integer values. However, it is sensible to study expectation of these variables to assess time, space and circumstance variations; in fact, we suppose that the observed ordinal value is in a one-to-one correspondence with a continuous latent variable and thus it becomes useful to compute such quantities.

Specifically, the expectation of R is obtained as:

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

Since both parameters apport relevant contributions in determining this quantity, we notice that several models generate the same expectation, and the classical paradigma of GLM (a link function among expectation and covariates) cannot be applied in our case. As a consequence, we prefer to relates directly one or both parameters to subjects' covariates by means of the logistic function (that is, a convenient mapping of the real line into the unit interval).

Then, for a given $m > 3$, the general formulation of a $CUB(p, q)$ model (with p covariates to explain uncertainty and q covariates to explain feeling) is expressed by:

1. *stochastic component*:

$$Pr(R_i = r \mid \boldsymbol{y}_i;\, \boldsymbol{w}_i) = \pi_i \binom{m-1}{r-1} \xi_i^{m-r} (1-\xi_i)^{r-1} + (1-\pi_i)\left(\frac{1}{m}\right)\,;$$

for $r = 1, 2, \ldots, m$ and any i-th subject $i = 1, 2, \ldots, n$.

2. *systematic components*:

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

where $\boldsymbol{y}_i$ and $\boldsymbol{w}_i$ are the observed subjects' covariates for explaining π_i e ξ_i, respectively (Piccolo and D'Elia, 2008).

4. Inferential issues for CUB Models

It is now possible to write the general probability distribution of a $CUB(p, q)$ model as:

$$Pr\left(R = r \mid \boldsymbol{y}_i, \boldsymbol{w}_i; \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \frac{1}{1+e^{-\boldsymbol{y}_i\boldsymbol{\beta}}}\left[\binom{m-1}{r-1}\frac{\left(e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}\right)^{r-1}}{\left(1+e^{-\boldsymbol{w}_i\boldsymbol{\gamma}}\right)^{m-1}} - \frac{1}{m}\right] + \frac{1}{m}$$

for any $r = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, n$.

Then, given a sample of observed values of ordinal and covariates values $(r_i, \boldsymbol{y}_i, \boldsymbol{w}_i)'$, for $i = 1, 2, \ldots, n$, the log-likelihood function is defined as a function of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ by:

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

As it is common for mixture models, maximum likelihood (ML) estimation is pursued by E-M algorithm (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000) and approximate variance and covariance matrix of the ML estimators are derived from asymptotic inferences[2]. Specifically, standard errors of estimated coefficients, log-likelihood comparisons and some fitting measures are available for CUB models (Piccolo, 2006).

Moreover, we use *BIC* model selection criterion, a dissimilarity index $Diss$ (a normalized distance among observed relative frequencies f_r and estimated probabilities) and an $ICON$ measure (a sort of pseudo-R^2) which compares via log-likelihoods the estimated model with the worst one (that is, a discrete Uniform random variable fitted to data):

$$Diss = \frac{1}{2} \sum_{r=1}^{m} \left| f_r - Pr\left(R = r | \hat{\boldsymbol{\theta}}\right) \right|; \qquad ICON = 1 + \frac{\ell(\hat{\boldsymbol{\theta}})/n}{\log(m)}.$$

In the same vein, some alternatives models have been proposed in the past, as the Inverse HyperGeometric (IHG) random variable with covariates generated by a logic of sequential choices (D'Elia, 2003). However, the constraint of an extreme mode for any IHG model limits its use in several applications. Anyway, we have found that for our data set the performance of CUB models has been superior.

[2] An effective procedure has been devised in R code and the software is described in Piccolo and Iannario (2008).

5. Modelling stated worry for organized crime and waste disposal

In this section, we apply previous modelling approach to a rank data set for two of 9 items submitted to a large collection of dwellers in order to quantify the degree of risk perception and concern during recent years in a large urban area. Further related information about subjects' covariates (gender, age, job, residence, education, and so on) have been also collected.

First of all, we concentrate our attention on answers for worry about *Organized crime* and *Waste disposal* obtained in December 2004 and 2006 (subsections 5.1 and 5.2); then, we will discuss the results obtained by a larger sample collected in December 2007 (subsection 5.3).

Table 1 highlights the main characteristics and composition of the samples with regard to gender, age location indexes, quota of residents in the city and percentage of University students which are not involved in any kind of work.

Table 1. Description and composition of the samples

Years	n	*Women* $(\%)$	*Mean Age*	*Residence* $(\%)$	*No Job* $(\%)$
2004	354	41.0	26.1	70.9	62.7
2006	419	43.3	25.5	64.9	59.9
2007	2381	48.2	35.8	82.7	62.1

In Figure 1 the observed frequency distributions of the ranked evaluations for the two problems are shown for both years; they confirm that the shape of the responses is substantially unchanged. Instead, we notice the strong positive skewness of *Organized crime* and the moderate negative asymmetry of *Waste disposal*.

The responses for *Organized crime* are well concentrated on the values 1 and 2 (more than 80% of respondents) although heterogeneity measures increase from 2004 to 2006. We have to expect a strong feeling and limited uncertainty parameters for this issue.

On the contrary, in this period, *Waste disposal* (and also streets cleanness) is usually perceived as less dangerous than the other items connected to environment (e.g. *Traffic and local transport* and *Environmen-*

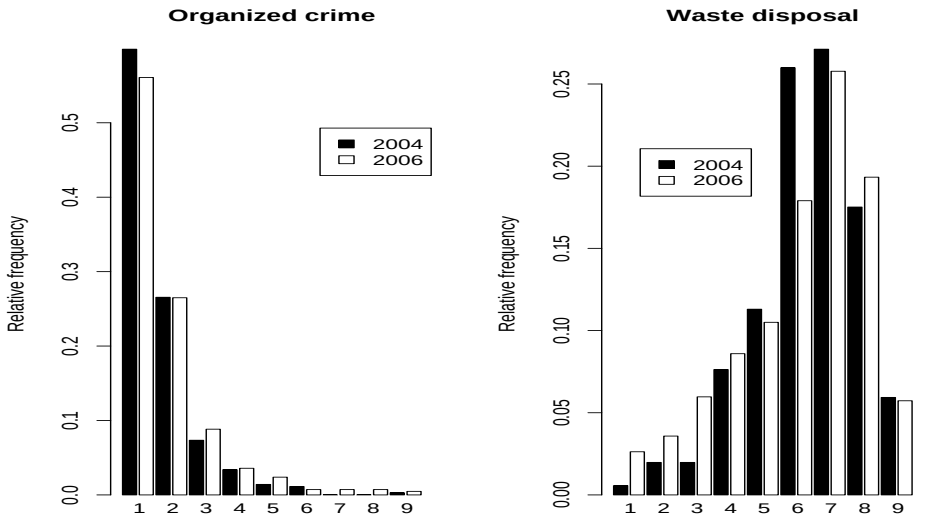

Figure 1. Frequency distributions of Organized crime and Waste disposal

tal pollution). As a matter of fact, the average rank is located in one of the last positions among the items, with mode and median at 7, for both years. Also, for this issue heterogeneity indexes increase over the years, thus we may expect an higher uncertainty in the responses of 2006 survey.

5.1. CUB models without covariates

Table 1 shows that there is a substantial homogeneity between the first two samples (in fact, both of them were collected among students of the Faculty of Political Science, University of Naples Federico II).

The main results for a $CUB(0,0)$ model for both emergencies, that is a probability distribution without covariates, are presented in Table 2. This model acts as a benchmark for measuring the improvement we will obtain when we introduce covariates, but it is also useful for checking if during the years some features have changed.

The estimated models are satisfactory from a statistical point of view

Table 2. Estimation of CUB models (Organized crime and Waste disposal)

Issues	*Years*	π	ξ	*Diss*	*ICON*	*U*
Organized crime	2004	0.937	0.940	0.053	0.489	0.007
		(0.019)	*(0.006)*			
	2006	0.898	0.936	0.056	0.432	0.011
		(0.021)	*(0.005)*			
Waste disposal	2004	0.892	0.310	0.041	0.175	0.012
		(0.033)	*(0.010)*			
	2006	0.665	0.293	0.076	0.097	0.037
		(0.044)	*(0.013)*			

(significance of parameters, small dissimilarity indexes, acceptable values of ICON). The uncertainty share $U = (1 - \pi)/m$ for *Organized crime* ranks is so small that even shifted Binomial and IHG models would give respectable fitting for this data (even though not so good as a CUB model).

It is worthy to notice that both models detect an increasing uncertainty in 2006, and this is more evident for *Waste disposal*. Although the relative importance of the problems does not change over the years, respondents are becoming less and less sharp in their judgments.

It is noticeable that data with different features (skewness is strong and positive for *Organized crime* whereas is moderate and negative for *Waste disposal*) may be well accounted by the same class of probability structures. In Figure 2 we plot the estimated $CUB(0, 0)$ distributions for the models in the two years[3].

Then, in order to relate the stated worry to subject's characteristics we check for a relationship explaining feeling and we look for significant covariates among those collected in the surveys. We found that sensible results are obtained by inserting the covariates *gender* and the logarithm[4]

[3] We are connecting probabilities for enhancing the shape of the probability mass distributions.

[4] In the following models, we will use the covariate $\log(age)$ instead of age since the logarithmic transformation improves slightly the fitting but significantly reduces the variability of

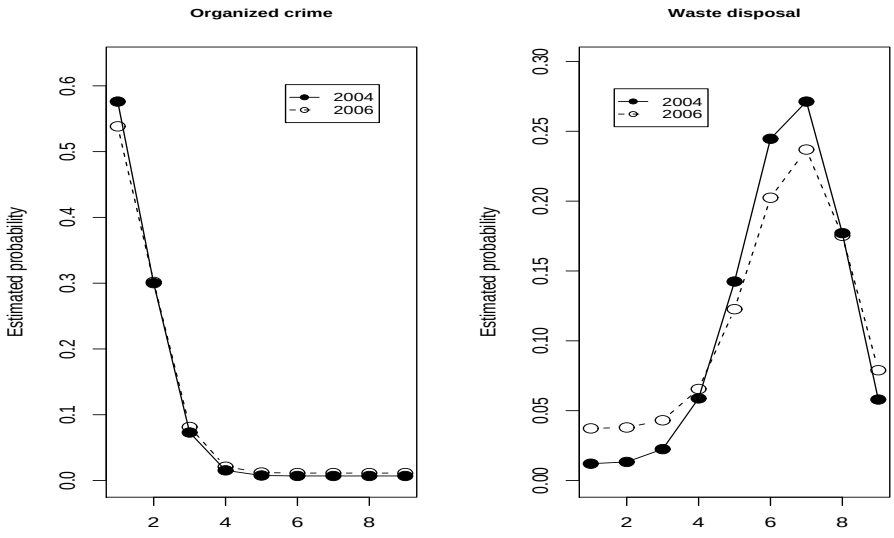

Figure 2. $CUB(0,0)$ *models for Organized crime and Waste disposal*

of the *age* of the respondents. Although some significant relationship has been found also for explaining uncertainty levels, in this paper we prefer to deepen only the impact of covariates on the feeling component.

Then, we will discuss separately the best models obtained for the perception of worry for these emergencies.

5.2. CUB models with covariates for Organized crime

Table 3 refers to the best estimated CUB models with covariates for *Organized crime* with standard notations for parameters. The covariate $gender$ is significant[5] for both years (with an impact estimated by the γ_1 parameter), and the covariate $\log(age)$ is significant only for data of 2006 survey. Instead, we register a sensible modification in the weight of

estimates and the rate of convergence of the E-M algorithm.

[5] In fact, the covariate *gender* is barely significant with a p-value of 0.08, but we prefer to include it in the model as the likelihood ratio test of the extended model is significant.

uncertainty that increases with time.

Table 3. Estimation of CUB(0,q) models for Organized crime

Years	$\hat{\pi}$	$\hat{\xi} = \xi(gender, \log(age))$	$\ell(\theta)/n$
2004	0.948 ***(0.018)***	$\hat{\gamma}_0 = \ \ 4.837$ ***(0.914)***	-1.1057
		$\hat{\gamma}_1 = \ \ 0.348$ ***(0.201)***	
		$\hat{\gamma}_2 = -0.701$ ***(0.276)***	
2006	0.897 ***(0.021)***	$\hat{\gamma}_0 = \ \ 2.505$ ***(0.107)***	-1.2398
		$\hat{\gamma}_1 = \ \ 0.447$ ***(0.183)***	

For expressing the impact of covariates on the feeling parameter ξ, we may explicit the systematic relationships:

$$\xi_i^{(2004)} = \frac{1}{1 + e^{-4.837 - 0.348\, gender_i + 0.701\, \log(age)_i}};$$
$$\xi_i^{(2006)} = \frac{1}{1 + e^{-2.505 - 0.447\, gender_i}}.$$

Thus, remembering that ξ_i is directly related to the degree of worry for a subject with covariates $(gender_i,\, age_i)'$, $i = 1, 2, \ldots, n$, we can deduce that women are more apprehensive than men while elderly are less worried for this problem in both years. However, the concern lowers from the first to the second year.

In order to confirm the previous interpretation[6], we present in Table 4 the expectations implied by the estimated models; of course, as only *gender* is a significant covariate for 2006 data, expectations do not change with the *age* of the respondent.

5.3. CUB models with covariates for Waste disposal

For *Waste disposal* we found similar CUB models with the same covariates but with a different impact. Table 5 summarises the relevant estimates and measures obtained by maximum likelihood inference.

[6] Notice that average rank is low when the concern is very high and the perception of worry towards the item increases when rank diminishes.

Table 4. Estimated expectations for Organized crime for given covariates

	Age (Men)			Age (Women)		
Years	20	30	40	20	30	40
2004	1.669	1.809	1.930	1.539	1.642	1.732
2006	1.953	1.953	1.953	1.767	1.767	1.767

Table 5. CUB models for Waste disposal during 2004 and 2006

Years	$\hat{\pi}$	$\hat{\xi} = \xi(gender, \log(age))$	$\ell(\theta)/n$
2004	0.897 *(0.032)*	$\hat{\gamma}_0 = -0.713$ *(0.060)*	-1.8059
		$\hat{\gamma}_1 = -0.228$ *(0.098)*	
2006	0.696 *(0.044)*	$\hat{\gamma}_0 = -1.939$ *(0.565)*	-1.9707
		$\hat{\gamma}_1 = -0.307$ *(0.118)*	
		$\hat{\gamma}_2 = \ \ 0.387$ *(0.176)*	

The feeling parameters implied by these models are:

$$\xi_i^{(2004)} = \frac{1}{1 + e^{0.713 + 0.228\, gender_i}};$$

$$\xi_i^{(2006)} = \frac{1}{1 + e^{1.939 + 0.307\, gender_i - 0.387\, \log(age)_i}}.$$

The behaviour of models with respect to covariates is now specular if compared with the previous issue. First of all, the covariate *age* is significant only in the more recent year; above all, *age* is positively related to the concern. As a consequence, the models enhance that women are less worried than men while elderly suffer of more concern than young with regard to this issue.

Table 6. Estimated expectations for Waste disposal for given covariates

	Age (Men)			Age (Women)		
Years	20	30	40	20	30	40
2004	6.227	6.227	6.227	6.572	6.572	6.572
2006	6.032	5.839	5.695	6.378	6.207	6.078

A confirmation of these interpretations is obtained if we compute (Table 6) expectations implied by the estimated models for both years. In this case, as only *gender* is a significant covariate for 2004 data, expectations will not change with the *age* of the respondent. Moreover, the concern about *Waste disposal* is uniformly increasing from 2004 to 2006 for genders and varying age.

5.4. Perception of Organized crime and Waste disposal in 2007

As discussed in section 1, the survey submitted in December 2007 has been a larger one and its aim was to reach a wider audience with respect to University students. As shown in Table 1, the related sample is not immediately comparable to the previous ones as far as composition of age, gender and residence are concerned.

Briefly, this extended survey is more balanced with respect to gender with a mean/median age significantly higher than the previous ones and it is made up by a considerable amount (83%) of people that live in the city of Naples. Thus, it is important to check if previous considerations can be again applied, given also the increasing sensitivity towards these two problems at the end of 2007.

Tables 7 and 8 show the main inferential results obtained by fitting to 2007 data set the corresponding CUB models with covariates: again, *gender* (only for *Organized crime*) and *age* are significant covariates for explaining the personal concern towards these items.

Table 7. CUB model for Organized crime in 2007

$\hat{\pi}$	$\hat{\xi} = \xi(gender, \log(age))$	$\ell(\theta)/n$
0.696 ***(0.015)***	$\hat{\gamma}_0 = \quad 4.252$ ***(0.349)***	-1.7068
	$\hat{\gamma}_1 = -0.182$ ***(0.085)***	
	$\hat{\gamma}_2 = -0.505$ ***(0.096)***	

In fact, the impact of covariates on the chosen emergencies is different. For *Organized crime*, the relevance of *age* is preserved but the sign of *gender* is reversed (women are less worried in 2007). Instead, for *Waste*

Table 8. CUB model for Waste disposal in 2007

$\hat{\pi}$	$\hat{\xi} = \xi(\log(age))$	$\ell(\theta)/n$
0.486 ***(0.020)***	$\hat{\gamma}_0 = -1.992$ ***(0.288)***	-2.0845
	$\hat{\gamma}_1 = \quad 0.304$ ***(0.083)***	

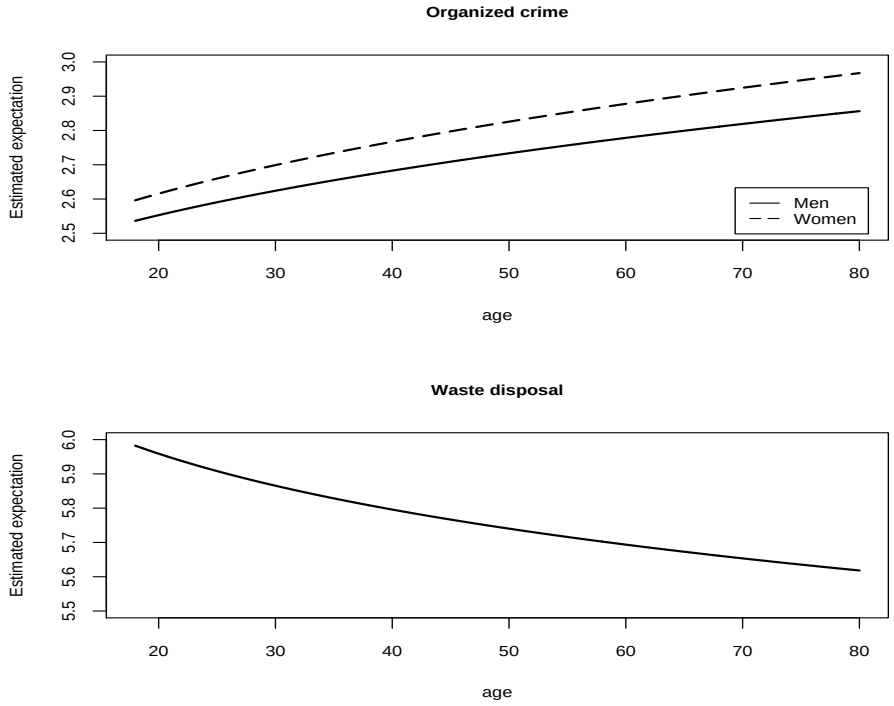

Figure 3. Estimated expectations by CUB models, given covariates

disposal, the only significant impact is registered by *age* and its effect on the concern is now reversed with respect to previous years (elderly are less worried than young in 2007).

To summarise effectively these results, it is useful to plot the expected concern for varying age given the gender for *Organized crime*, and only for varying age for *Waste disposal* (Figure 3).

An important and systematic feature that may be deduced by all the estimated CUB models is the sensible increase in the weight of uncertainty in the responses. This may be interpreted as a common sign of lack of confidence and general confusion among the respondents with regard to urban problems. They register a vaguer sense of generalized worry and the increasing uncertainty they add to answers should be an indication that perception and awareness of emergencies are becoming more and more fuzzy.

As a final comment to the analyses of this section, we should observe that the real impact of the covariates on the responses is not dramatic as we do not observe substantial differences among gender and young and elderly, given a specific emergency. This circumstance is common in sociological studies; however, this enhances the usefulness to introduce a class of models that allow for testing and assessing the significance of even small impacts on human choices.

6. Concluding remarks

In this paper we have shown how to check and fit the observed distribution of the concern expressed by people with reference to some important emergencies of a large city. These kind of problems are common issues in several urban contexts; we have chosen to deepen our study with two of them by analysing distributions and behaviours of the respondents.

The experiment confirmed that the statistical approach expressed by a class of ordinal models is worthy for quantifying the impact of covariates. This may discriminate psychological processes and mechanisms that generate raters' perception and help in interpreting causal relationships for the stated choices. Specifically, a unique parametric family of distributions is able to catch different features of data and significant subjects' covariates.

An open question is the search for efficient methods to select significant covariates from a given data set without testing a huge amount of possible combinations. In this area, we are looking for innovative measures as these models are not simply related to classical correlation analyses. Mostly, one should exploit the ordinal nature of the responses for

selecting appropriate and sensible measures of possible significant covariates.

Acknowledgements: This research has been partly funded by the MIUR-PRIN 2006 grant (Project on: *"Stima e verifica di modelli statistici per l'analisi della soddisfazione degli studenti universitari"*) and supported by the scientific structures of CFEPSR, Portici. The contribution has been presented at the 7th International Conference on Social Sciences Methodology, RC33, Naples, 1-5 September 2008. Critical suggestions by Editor and referees are gratefully acknowledged.

References

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

Beck U. (1992), *Risk society: towards a new modernity*, Sage Publications, London, New Delhi.

Cleveland W. S. (1981), LOWESS: A program for smoothing scatterplots by robust locally weighted regression, *The American Statistician*, 35, 54.

Corduas M. (2008a), Clustering CUB models by Kullback-Liebler divergence, *Proceedings of SCF-CLAFAG Meeting*, ESI, Napoli, 245–248.

Corduas M. (2008b), A study on University students' opinions about teaching quality: a model based approach for clustering ordinal data, *Proceedings of DIVAGO Meeting*, University of Palermo, 10-12 July 2008.

D'Elia A. (2003), Modelling ranks using Inverse Hypergeometric distribution, *Statistical Modelling: an International Journal*, 3, 65–78.

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

Dobson A. J., Barnett A. G. (2008), *An Introduction to generalized linear models*, 3^{nd} edition, Chapman & Hall/CRC, Boca Raton.

Fligner M. A., Verducci J. S. (1999), *Probability models and statistical analysis of ranking data*, Springer-Verlag, New York.

Iannario M. (2007a), A statistical approach for modelling Urban Audit Perception Surveys, *Quaderni di Statistica*, 9, 149–172.

Iannario M. (2007b), Dummy covariates in CUB models. *STATISTICA*, accepted for publication.

Iannario M. (2008), A note on the identifiability of a mixture model for ordinal data, preliminary report.

Iannario M., Piccolo D. (2009), A new statistical model for the analysis of customer satisfaction, *Quality Technology and Quantitative Management*, in press.

Loewenstein F., Hsee C. K., Weber U., Welch N. (2001), Risk as feeling, *Psychological Bulletin*, 127(2), 267–286.

Joreskog K., Moustaki I. (2001), Factor analysis for ordinal variables: a comparison of three approaches, *Multivariate Behavioural Research*, 36, 347–387.

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.

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.

Marden J. I. (1995), *Analyzing and modelling rank data*, Chapman & Hall, London.

Moustaki I. (2003), A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variable, *British Journal of Mathematical and Statistical Psychology*, 56, 337–357.

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

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

Piccolo D, Iannario M (2008) A package in R for CUB models inference, Version 1.1, available at *http://www.dipstat.unina.it*