Quaderni di Statistica
Vol. 10, 2008

Modelling University students' final grades by ordinal variables

Domenico Piccolo
Dipartimento di Scienze Statistiche, Università di Napoli Federico II
E-mail: domenico.piccolo@unina.it

Summary: The final assessment received by graduates in Italy is expressed on a 66-110 scale points and this quantity is determined by the marks average of all exams and an extra value depending on the judgement of the Commission with regard to the thesis defense. Although the final grade is expressed on a quantitative scale, what it is really important for the future career of a graduated student is the qualitative ordinal evaluation of the final grade (from a very low assessment up to excellent, and first class honors). Empirical evidence and academic rules confirm that finale grades are related both to subjects' covariates and latent variables connected to students' background and career. In this paper, we will apply models for ordinal data as this framework seems more reliable with respect to the final judgment. Specifically, we introduce a class of ordinal models where a direct evaluation of the probability of a qualitative result is fitted and then modelled by means of gender, duration of studies and secondary school diploma marks. An empirical case study have been analyzed for a large data set of graduated in Political Sciences, University of Naples Federico II, and the main results are discussed.

Keywords: Ordinal data modelling, Qualitative assessment, CUB Models.

1. Introduction

The statistical analysis of ordinal data is a growing area for studies of methodological and applied interests. In current literature a large number of works are related to exploratory analysis, testing hypothesis and efficient estimation algorithms. Several contributions deal with classical approaches as correlation analysis and multivariate methods. However, the turning point of this kind of research has been the inclusion of the topic within the logic of Generalized Linear Models (GLM: McCullagh and Nelder, 1989).

In fact, researchers have looked for adequate transformations of the original qualitative information in order to obtain numerical values and to apply standard statistical methods, as regression models for instance. Specifically, several studies imply models that explain the probability that responses are not greater that a given category, and this statement may be usefully maintained even for qualitative ordinal data. Thus, the class of ordinal logistic and probit models has been applied, as discussed by McCullagh (1980), Agresti (2002) and Dobson and Barnett (2008); for an updated survey see: Bock and Moustaki (2007). Moreover, a new perspective has been introduced by D'Elia and Piccolo (2005) and Piccolo (2006) who proposed a direct formulation of the probability distribution for a discrete ordinal choice, mainly based on the psychological mechanism of human decisions.

In this paper we examine a different problem: we have a sample of numerical values, which are the results of a qualitative assessment about the value of a performance or the evaluation of an object/service/sentence, and so on. In fact, the final judgement is expressed as a quantitative value within a prefixed range. However, this number is just a proxy of a qualitative ordinal evaluation expressed by raters; then, although classical regression methods are quite diffuse for these data, specific analyses should be more effective. In this case, we argue that it may be convenient to come back to a qualitative approach since the explicit expression of ordinal values may be modelled in an effective and direct way.

These general considerations will be pursued by means of a detailed discussion of a real case study consisting of a large collection of marks obtained by graduates in Political Sciences at University of Naples Federico II. Moreover, we will relate the final educational results to gender, diploma marks and duration of permanence at University; this will confirm that a qualitative model can achieve useful results and more interesting interpretation.

The paper is organized as follows: in the next section, we will establish formal background and notations for data, variables and models; then, sections 3-4 will present the data set and some issues related to the problem of the final evaluation of graduates. Furthermore, we will discuss both standard regression analysis (section 5) and ordinal models (section 6) of the same data set and, in section 7, we will compare the corresponding results via expected grades given subjects' covariates. Some final considerations end the paper.

2. Formal background and notations

In this section, we will introduce some notations for our study. The common background is that a set of raters/judges expresses subjective evaluations about people, sentences, objects, questions, and so on. A consistent hypothesis is that this evaluation may depend on both items or raters. However, in order to simplify the analysis, we will assume that results are only related to subjects' measurements and behaviours.

Notice that this background is different from the standard framework where an object has to be evaluated by several judges, and our interest is in the raters' covariates: a common situation in marketing studies and clinical experiments, for instance. Instead, we are saying that a collective rater (that is a jury, committee, commission, and so on) is evaluating several outputs (each of them characterized by some covariates) and we wish to relate the final qualitative assessment to the objects' characteristics. Then, in the real case studies we will discuss about in the next sections, "objects" will be students to be graduated and the assessment is the final grade they received after the thesis defence.

Of course, the scheme we will assume is that a collective jury acts on several objects with constant rules of evaluation within some well defined criteria and rules. This assertion may be consistently accepted although the jury changes continuously its composition from a session to another. The circumstance adds some uncertainty to the subsequent study but it is a common prerequisite for any sociological analysis where collective agents are supposed to act in a consistent way although they manifest themselves in different time, space and condition. Actually, the adhesion to this postulate legitimates all studies concerning the behaviour of Parliaments, Justice Courts, public opinions, etc. in relation to everyday decisions.

Then, we will indicate by $(r_1, r_2, \ldots, r_n)'$ the sample data where each observed evaluation r_i is an integer expressing intensity, level, amount of an ordinal variable R for some real phenomenon. We will assume that R is a discrete random variable defined on the support $\{1, 2, \ldots, m\}$, where m is a known fixed integer value.

Then, with reference to the evaluation of an object, the jury transforms a continuous latent variable into m discrete ordered bins according to some perception of values induced by quality, expression, belief, adhesion, etc. We call this component *feeling* and it represents the primary component in the final judgment. The secondary component is a variable that represents *uncertainty* and it always accompanies human choices.

As a consequence, for modelling ordinal discrete data, D'Elia and Piccolo (2005) introduced a mixture distribution defined by:

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

It is possible to check that, for a given $m > 3$, this mixture random variable is identifiable (Iannario, 2008b) and its parameters belong to the unit square: $\pi \in (0,1]\,,\ \xi \in [0,1]$.

In previous works, empirical evidence and psychological considerations motivate us to introduce for these components a *shifted Binomial* random variable as an adequate probability model for representing the discrete version of a latent judgement process, able to measure the subjective *feeling* in expressing evaluations. Similarly, the *discrete Uniform* random variable has been introduced to take into account of the inherent *uncertainty* of a discrete choice process. In fact, we are not saying that raters introduce a completely random selection mechanism among the m categories, but that any observed uncertainty on a finite discrete support may be weighted with respect to this extreme case.

Thus, $(1-\pi)/m$ is a *measure of the uncertainty* because this quantity is the constant amount of probability which spreads uniformly over all the support. Instead, the interpretation of $(1-\xi)$ is mainly related to a direct and positive evaluation of the object. A more precise definition would depend on the context, and for our data set we have found convenient to interpret $(1-\xi)$ as a *measure of performance*[1].

In previous analysis of several evaluation data (in different contexts, as marketing, psychological studies, sociological surveys, political polls, and so on) a class of ordinal models, called CUB, has been applied. Then, in order to adopt this paradigma to our specific context we have to modify some interpretation of this basic model[2].

Since, for a given m, the expected value of R is given by: $E(R) = \pi\ (m-1)\left(\frac{1}{2}-\xi\right)+\frac{(m+1)}{2}$, it is evident that different parameter vectors $\boldsymbol{\theta} = (\pi,\, \xi)'$ may generate the same mean value; thus, it is not adequate to introduce a link among expectation and covariates.

Instead, in CUB models we assume that feeling and uncertainty parameters are related to covariates by a logistic function (or any normed one-to-one function), that is by means of two *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 subjects' covariates for explaining π_i e ξ_i, respectively. If necessary, the model may be generalized to include both objects' and subjects' covariates (Piccolo and D'Elia, 2008). In these cases, we will use the notation $CUB(p,q)$ for denoting the number of covariates entering in the model for explaining feeling and uncertainty components, respectively.

The sampling experiment consists of the collection of evaluations and covariates $(r_i,\, \boldsymbol{y}_i,\, \boldsymbol{w}_i)'$, for $i = 1,2,\ldots,n$. This information (for moderate and large size) is sufficient to generate reliable inference on the estimated parameters $\hat{\boldsymbol{\theta}} = (\hat{\pi},\, \hat{\xi})'$ via the log-likelihood function $\ell(\boldsymbol{\theta})$ and related asymptotic results, as detailed by Piccolo (2006). Finally, a software

[1] The circumstance that the key concepts of our study are inversely related to the parameters of the probability distribution is due to the fact that this random variable has been previously studied in a ranking analysis, where ξ is positively related to likeness (Iannario, 2007; 2008). Here, we have preferred to maintain a consistent notation with previous works. For different interpretations, see: Iannario and Piccolo (2009).

[2] These models have been called MUB or CUB, respectively, according to the absence or presence of covariates in the mixture distribution: D'Elia and Piccolo (2005); Piccolo and D'Elia (2008). Hereafter, we will denote both structures as CUB models.

for estimating and fitting CUB models is currently available (Piccolo and Iannario, 2008a).

In this context, some criterion is necessary to select among different models, and classical Chi-square tests are not effective for moderate and large sample sizes. Thus, we prefer a dissimilarity index $Diss$ defined as the absolute distance among the observed relative frequencies f_r, $r = 1, 2, \ldots, m$ and the probabilities computed from the estimated model:

$$Diss = \frac{1}{2} \sum_{r=1}^{m} \left| f_r - Pr\left(R = r|\hat{\boldsymbol{\theta}}\right) \right| .$$

The measure is normalized, since $Diss \in [0, 1]$, and generally estimated models such that $Diss < 0.10$ may be considered as acceptable.

Another measure we adopt is $ICON$ (=*I*nformation *CON*tent), that is a pseudo-R^2; it compares the log-likelihood of the estimated model with the log-likelihood of a discrete Uniform random variable fitted to data (this is the worst uninformative model, given ordinal data and m categories). The *ICON* index is defined by:

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

It measures the improvement that we obtain when we move from a completely uninformative distribution (as the Uniform one) to a well structured random variable (as CUB models), without or with covariates. This index is related to the displacement of the estimated log-likelihood with respect to an extreme situation.

Notice that this class of models is able to take into account the discrete nature of the answers and to explain the evaluation process without referring to log-odds, adjacent and continuation probabilities (as usual in the GLM approach). Instead, CUB models offer a straightforward probability statement between the ordinal answer and the covariates. In addition, it should be noted that, although latent variables are conceptually necessary in order to specify the nature of the mixture components, we never rely the inferential procedures upon the knowledge (or estimation) of cutpoints. As a consequence, given the model, this simplification turns into a more parsimonious parametric structure.

3. Qualitative assessment of final grades

The data set of this study consists of final assessment grades received by graduates[3] in Political Sciences at University of Naples Federico II after a 4-years course (data have been collected before the start up of the current University reform where two levels of studies are considered). In Italy, a Commission expresses these evaluations on a $66 - 110$ scale points (66 is the minimum to be graduated) and the final grade is determined both on average marks of exams and an overall judgment of the Commission with regard to the thesis defense. Moreover, in positive circumstances, a first class honors (=*summa cum laude*) is given to the candidate: this is a very important result for the graduates' career, both in public and private job interviews[4].

Thus, although the received mark is a expressed on a quantitative scale (that is a number $V \in [66, 112]$), what it is really important for a graduate is to receive a substantiative evaluation belonging to a certain class of merit. Table 1 offers an acceptable correspondence[5] among quantitative and qualitative assessments by introducing an ordinal variable R with $m = 7$ categories for each value (or interval values) of V.

The observed frequency distribution of grades (Figure 1) received by $n = 2324$ graduates enhances a strong atypical value at $R = 7$ and confirms the importance of taking into account this modal value in order to improve the fitting of the observed data, as we will discuss in section 6.

To be consistent with our modelling interpretation, we adopt the following paradigma. The final qualitative assessment R is assumed to be the result of several evaluation steps, some of them related to the University *performance of the subjects* and some others related to the *inherent uncertainty* that it is intrinsic in any competition (if examined from

[3] The sample includes also graduates that received a previous University degree; this circumstance a very short duration of studies (less than 4 years) for a limited subset of subjects $(= 2.54\%)$.

[4] In the following, we adopt the convention to attribute the numeric value of 112 to first class honors.

[5] Of course, any correspondence includes some degree of arbitrariness. Thus, the intervals shown in Table 1 have been defined after several discussions with people usually involved in University Commissions.

Table 1. Quantitative and qualitative evaluations of graduates marks.

Final grade	*Qualitative evaluation*	*Class (Rating)*
$V = 110\ cum\ laude$	First class honors	$A \quad (R = 7)$
$V = 110$	Excellent	$B \quad (R = 6)$
$105 \leq V < 110$	Very good	$C \quad (R = 5)$
$100 \leq V < 105$	Good	$D \quad (R = 4)$
$90 \leq V < 100$	Sufficient	$E \quad (R = 3)$
$80 \leq V < 90$	Low	$F \quad (R = 2)$
$66 \leq V < 80$	Very low	$G \quad (R = 1)$

Figure 1. Observed distribution of grades

the student's point of view) and judgement (if examined from the Commission's point of view). Then, both components may be fitted into the standard framework and maintained hypotheses of CUB models.

More specifically, we will discuss about the performance that is related to the personal history and ability of the subject (this determines the average marks received during University training) and a general aptitude towards study (in fact, gender and duration of studies are relevant covariates). In our data set we have full information about these covariates, and thus we will include them in the subsequent models as sound proxies of latent variables that explain the final grades.

Although the proposed models are able to quantify also the weight of uncertainty, we will not deepen here the role of the secondary component.

4. Exploratory analyses

Data set includes personal information on graduates (gender, diploma marks, duration of the University studies) and, in this section, we will check if these information are sensible explanatory variables for grades.

Our sample of graduates consists of 55.3% women and 45.7% men, with an average duration of studies of 7.31 years after the enrollment (median value is 6.67, and there are even extreme cases: 26.5 years, for instance).

Marks received at the end of secondary school (expressed with respect to the maximum 60) are generally not quoted as significant covariates for predicting the future performance of a University student; however, this covariate is a good indication of general background and attitude towards studies and quite often it is strongly related with regularity of the attendance. This variable is almost uniformly distributed over the range with a mean of 46.5, a first mode at the minimum of 36 (8.7% of students) and a secondary modal value at the maximum of 60 (8.5% of students).

For checking possible relationships among these covariates, we will examine some plots and robust locally weighted regression lines (Cleveland, 1981). Specifically, Figure 2 shows boxplots of Laurea marks (V) and Duration of studies (Dur) with respect to Gender.

It seems evident that women received on average Laurea marks higher than men and they finish their training in shorter times. Notice that a large number of outliers characterizes both women marks and duration of studies of both genders.

Then, the relationship between Laurea and Diploma marks is shown in Figure 3 (both variables are jittered in order to enhance the general pattern). The robust fitted line confirms a positive linear association among

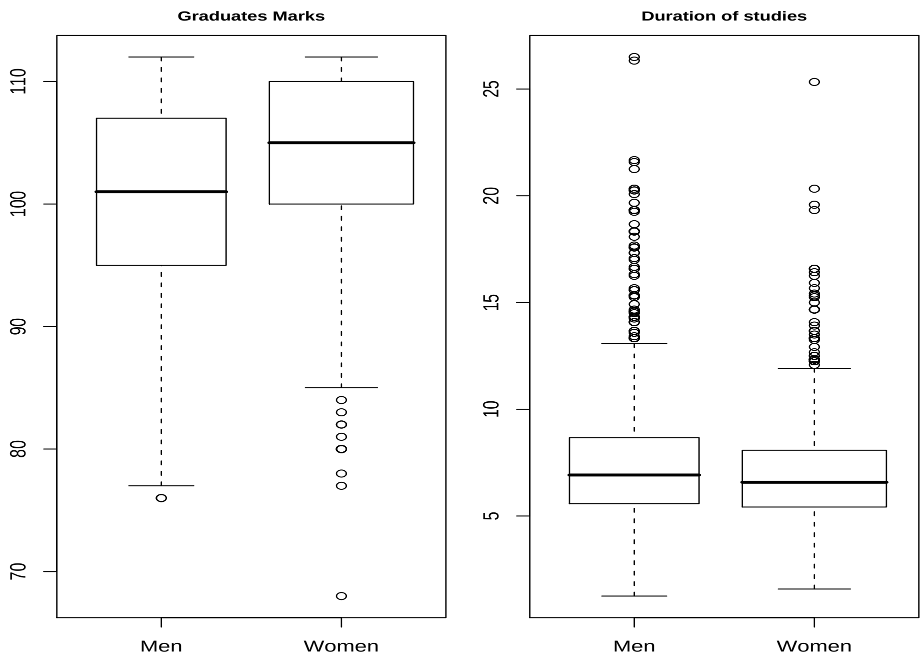

Figure 2. Box-plots of Laurea marks and Duration of studies vs. Gender

these marks although a large variability is present in the data set; thus, we may detect a significant trend between covariates but a prediction of Laurea marks based on Diploma marks is a difficult task, mainly for students that received low and intermediate marks.

Figure 4 shows that the relationship among Laurea marks and Duration of studies is strictly non-linear, and it increases up to 5 years (a modal value for getting the best performance of a graduate) and decreases regularly and slowly down to a stable value. This pattern is the consequence of a small subgroup (people that get a second grade in a duration shorter than the standard one) and of a larger subgroup of regular students (for which the final result deteriorates if the duration of studies lengthens as a consequence of intrinsic personal difficulties). Notice that a robust fitting highlights this complex pattern but we cannot rely too much on the shape of the relationship since extreme durations are limited in number.

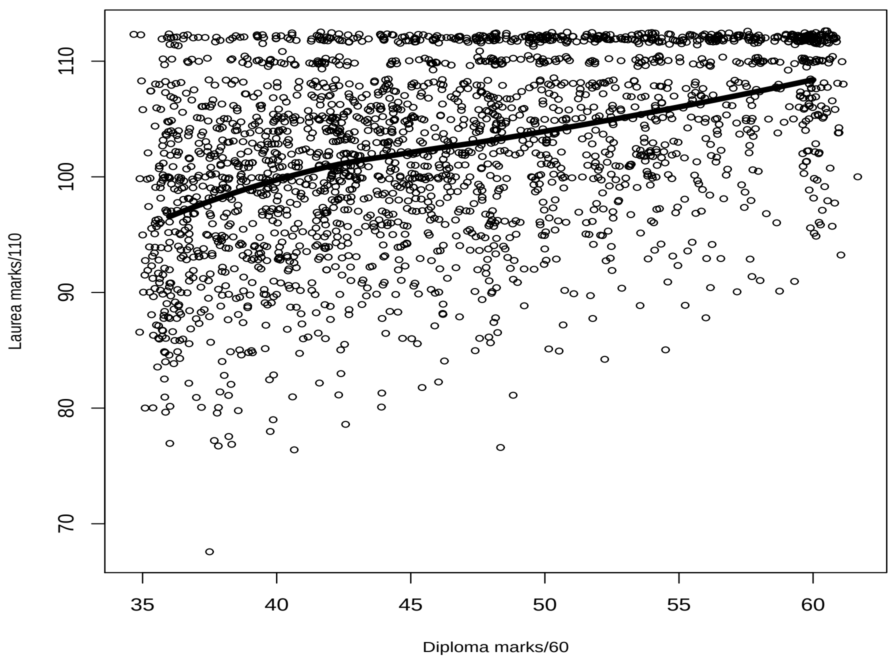

Figure 3. Relationship between Laurea and Diploma marks (jittered)

However, this result anticipates that an effective regression model should take into account some non-linear relationship among Laurea marks and Duration of studies.

If we code Gender ($= 0, 1$ for Men, Women, respectively), the correlation coefficients matrix of all considered variables (Table 2) shows that linear relationships are all significant but not so strong, given that data exhibits a large variability.

Explorative analyses confirmed that Gender, Duration of studies and Diploma marks turned out to be significant covariates for possible models aimed at explaining the final graduated marks for this data set. As final remarks, we firstly observe that the large variability of data will produce models with a low degree of forecast ability. Secondly, regression lines

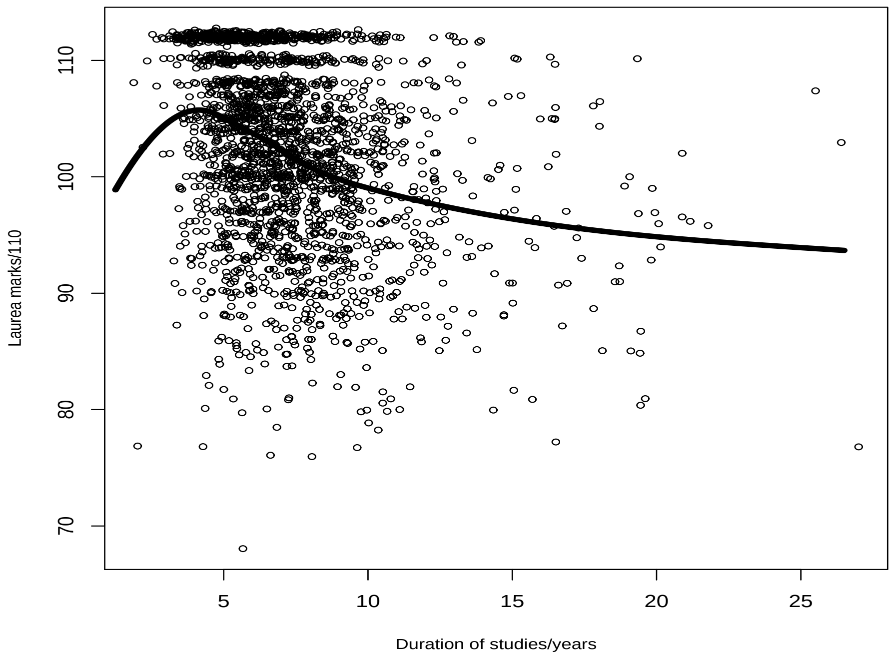

Figure 4. Relationship between Laurea marks and Duration (jittered)

Table 2. Correlation coefficients matrix

	Laurea marks	Diploma marks	Duration	Gender
Laurea marks	1.000	0.433	−0.296	0.222
Diploma marks	0.433	1.000	−0.206	0.201
Duration	−0.296	−0.206	1.000	−0.121
Gender	0.222	0.201	−0.121	1.000

will express average patterns with respect to the covariates for moderate and high levels of marks; thus, we may expect their inability to cope with low values of the dependent variable.

5. Regression models of final grades

We exploited several solutions for estimating regression models, and we report here only the final result of such steps. Table 3 summarizes estimated coefficients and fitting measures for models of increasing complexity. Notice that, following a stepwise procedure, each covariate enters in the model according to a decreasing order of explanatory power of the dependent variable[6].

Table 3. Regression models for graduation marks

Models	Constant	Markdip	Dur	Gender	$(\text{Dur})^2$	R^2	BIC
$\mathcal{M}_1$	81.123	0.457				0.187	9086.0
	(0.929)	*(0.020)*					
$\mathcal{M}_2$	87.750	0.410	−0.608			0.232	8963.1
	(1.069)	*(0.020)*	*(0.052)*				
$\mathcal{M}_3$	87.582	0.386	−0.579	1.940		0.246	8926.4
	(1.060)	*(0.020)*	*(0.052)*	*(0.290)*			
$\mathcal{M}_4$	90.643	0.379	−1.226	2.001	0.032	0.251	8920.7
	(1.345)	*(0.020)*	*(0.183)*	*(0.290)*	*(0.009)*		

We see that the best model $(\mathcal{M}_4)$ explains just a fourth of the total variability of graduates marks. The most important explanatory variable is Diploma marks (*Markdip*) whereas a significant negative impact is due to the Duration of studies (variables *Dur* and Dur^2 with opposite signs of the coefficients capture the non-monotonic impact implied by Figure 4); then, the model estimates that *ceteris paribus* women graduates receive 2 marks on average more than men. In addition, we observe that the impact of Diploma marks on the final grades is moderate (although this is the most significant covariate) as the average contribution of this variables ranges from 13.644 to 22.740.

[6] We do not report here the results of a further regression model that includes an interaction term between Duration and Gender variables. Actually, this addition is significant but it worsens both the significance of others parameters and BIC criterion.

6. Ordinal models of final grades

In a different perspective, we fit CUB models to the same data set for explaining the probability to receive a qualitative grade (as defined in Table 1) given the subjects' covariates. The model building procedure is greatly simplified by the previous correlation analysis although there is no necessary relationship between a significant estimate in linear regression and a corresponding parameter of CUB models.

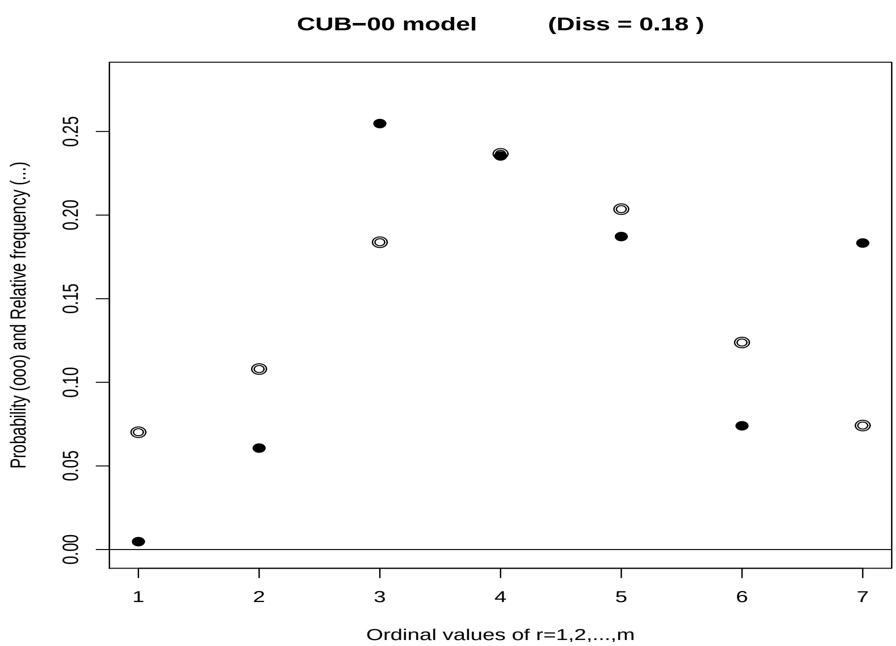

Figure 5. $CUB(0,0)$ *model fitted to grades*

The benchmark for comparing structures with covariates is a $CUB(0,0)$ model and Figure 5 shows observed relative frequencies and corresponding estimated probabilities for each ordinal grades. We find a not very good fitting mainly because of a modal behaviour at $R = 7$. This behavior is not consistent with a simple $CUB(0,0)$ model, caused by a "shelter effect" for the first class A (12.5% for men and 20.6% for women). However, we will not discuss here the implication of this further characteristic[7] and limit ourselves to consider the performance of standard CUB models without and with covariates. In fact, we present the sequence of fitted models for the class $CUB(0, q)$, for $q = 1, 2, 3, 4$; thus, we will include explanatory variables only for the parameter (ξ).

Then, with a stepwise procedure, we add covariates for explaining the ξ parameter following the logic of a decreasing contribution to log-likelihoods. Estimated $CUB(0, q)$ models of increasing complexity are listed in Table 4 with a fitting measure ($ICON$ index).

Table 4. CUB models for grades

$\hat{\pi}$	$\hat{\xi}$ or $\hat{\gamma}_0$	Markdip	log(Dur)	Gender	Gender × log(Dur)	$ICON$
0.557	0.481					0.05038
(0.025)	*(0.008)*					
0.806	3.601	−0.084				0.10794
(0.021)	*(0.166)*	*(0.004)*				
0.854	1.132	−0.072	0.963			0.12648
(0.019)	*(0.227)*	*(0.003)*	*(0.075)*			
0.867	1.112	−0.067	0.933	−0.251		0.12972
(0.019)	*(0.222)*	*(0.003)*	*(0.074)*	*(0.046)*		
0.863	1.806	−0.066	0.559	−2.037	0.907	0.13394
(0.019)	*(0.250)*	*(0.003)*	*(0.093)*	*(0.297)*	*(0.149)*	

All models are nested and we check their significance by means of likelihood asymptotic tests (deviance differences). The last one improves the information content more than 13%, and this is an important result for overdispersed data. Notice that the uncertainty share regularly decreases (from 0.063 down to 0.020) by including significant covariates. Then, the

[7] Actually, Iannario and Piccolo (2008) cope with this situation by means of *extended CUB models* with just an extra parameter. In this regard, it is noticeable to observe the related fitting measure ($Diss = 0.043$) which achieves a 76% reduction with respect to the previous one ($Diss = 0.180$).

sign of estimated parameters are all consistent with the expected impact of the covariates[8] on the final grades.

From the last model, we derive the estimated relationships of the parameters ξ_i with the covariates, both for men and women. Specifically, for any $i = 1, 2, \dots, n$, we get:

$$\xi_i^{(Men)} = \frac{1}{1 + 0.164\,(1.068)^{Markdip_i}\,(Dur_i)^{-0.559}}\,;$$
$$\xi_i^{(Women)} = \frac{1}{1 + 1.650\,(1.068)^{Markdip_i}\,(Dur_i)^{-1.466}}\,.$$

We remember that final grades are inversely related to the ξ parameter, and thus the students' *performance* is in direct relationship with $(1 - \xi)$. Then, the estimated CUB model predicts that the probability of high grades decreases with Duration of studies whereas it increases with Diploma marks and for Women. An interesting feature of this model is the presence of a significant interaction between the variable Gender and Duration; this aspect will be remarked in the next section.

As further aid to the interpretation of the estimated model, in Table 5 we present how the estimated $(1 - \hat{\xi})$ performance parameter changes with the student's profile; similarly, we show the probability to receive a grade in class A (=first class honors) or G (=very low), respectively. Then, some typical patterns are chosen and we examine several combinations of low (L), medium (M) and high (H) profiles specified by Diploma marks ($= 36, 48, 60$), Duration of studies ($= 5, 7, 10$ years) and Gender ($= 0, 1$ for Men and Women), respectively.

The scheme confirms the effect of covariates as previously discussed and also the sensible difference between genders, mainly for low Diploma marks.

[8] We prefer to introduce the logarithm transformation for the variable Duration since this improves the speed of convergence of the EM algorithm and the condition number of the observed variance-covariance matrix of estimators.

Table 5. Estimated probabilities for given students' profiles

Profiles	Markdip	Dur	Gender	$1-\hat{\xi}$	$Pr\,(R=1)$	$Pr\,(R=7)$
L1	36	5	0	0.417	0.053	0.024
L2	36	7	0	0.372	0.073	0.022
L3	36	10	0	0.327	0.100	0.021
L1	36	5	1	0.559	0.026	0.046
L2	36	7	1	0.437	0.047	0.026
L3	36	10	1	0.315	0.109	0.020
M1	48	5	0	0.612	0.023	0.065
M2	48	7	0	0.566	0.025	0.048
M3	48	10	0	0.517	0.031	0.036
M1	48	5	1	0.737	0.020	0.157
M2	48	7	1	0.631	0.022	0.074
M3	48	10	1	0.503	0.033	0.033
H1	60	5	0	0.776	0.020	0.208
H2	60	7	0	0.742	0.020	0.164
H3	60	10	0	0.702	0.020	0.123
H1	60	5	1	0.860	0.020	0.369
H2	60	7	1	0.790	0.020	0.230
H3	60	10	1	0.690	0.020	0.113

7. Comparative assessment of expected grades

Regression and CUB models are generated by different objectives. Thus, they can not be strictly compared since the first approach looks for a linear relationship among a dependent variable and covariates whereas CUB models are able to explain the probability distribution of ordinal values as determined by covariates affecting uncertainty and performance parameters. However, in both cases it is possible to compute the expectation of the dependent variable given a model and a set of values of covariates.

Then, we will estimate the expected marks $\mathbb{E}\,(V \mid \boldsymbol{w}_i)$ of a graduate implied by the regression model and the corresponding expected grade $\mathbb{E}\,(R \mid \boldsymbol{w}_i)$ implied by the CUB model, given the same covariates $\boldsymbol{w}_i$ for the i-th subject. Finally, in Figures 6 and 7, we plot expectations for both models, respectively, as functions of Duration, given Gender and some prefixed Diploma marks (we chose 36, 48, 60 as low, medium, high profiles).

From a formal point of view, the expected marks of a *regression model* ($\mathcal{M}_4$ in Table 3)are simply obtained by inserting the estimated parameters of section 5 in the linear relationship:

$$\begin{aligned} \mathbb{E}\,(V \mid \boldsymbol{w}_i) &= 90.643 + 0.379\, Markdip_i - 1.226\, Dur_i \\ &+ 2.001\, Gender_i + 0.032\,(Dur_i)^2\,. \end{aligned}$$

Then, from Figure 6, it is immediate to observe the constant displacement induced by Gender and Diploma marks and the parabolic behaviour of the mean value with respect to the Duration of studies: in fact, all parabolas have a minimum at $Dur = 19.259$ years and we have no empirical or logical evidence to support such assertion.

We also notice that the regression model can achieve an expected minimum of Laurea marks of $V = 92.487$ for a man with Duration of studies of 19.259 years and with Diploma marks of 36. To get an idea of this constraint for interpreting and forecasting data, we notice that 274 ($= 12\%$) graduates received a Laurea mark lower than this minimum. As a consequence, too many observations (graduates in classes F, G and most of E) might be considered as extreme according to the regression model.

The expectation implied by the $CUB(0, 4)$ *model* (last line in Table 4) is obtained by inserting the estimated parameters of section 6 in the formula of section 2. Since $m = 7$, we get:

$$\mathbb{E}\,(R \mid \boldsymbol{w}_i) = 4 + 5.178\, \left(\frac{1}{2} - \xi_i\right) = 6.589 - 5.178\,\xi_i\,,$$

where $\xi_i = [1 + \exp\,(g_i)]^{-1}$ for any $i = 1, 2, \ldots, n$, and we put:

$$\begin{aligned} g_i &= -1.806 + 0.066\, Markdip_i - 0.559\, \log(Dur_i) + 2.037\, Gender_i \\ &\quad -0.907\, Gender_i \times \log(Dur_i)\,. \end{aligned}$$

Figure 7 confirms the previous considerations and the common pattern of expectations for both models; however, some advantage of the qualitative models should be enhanced.

For instance, by a simple algebra, it is possible to estimate that the turning point for a different behaviour between genders happens when

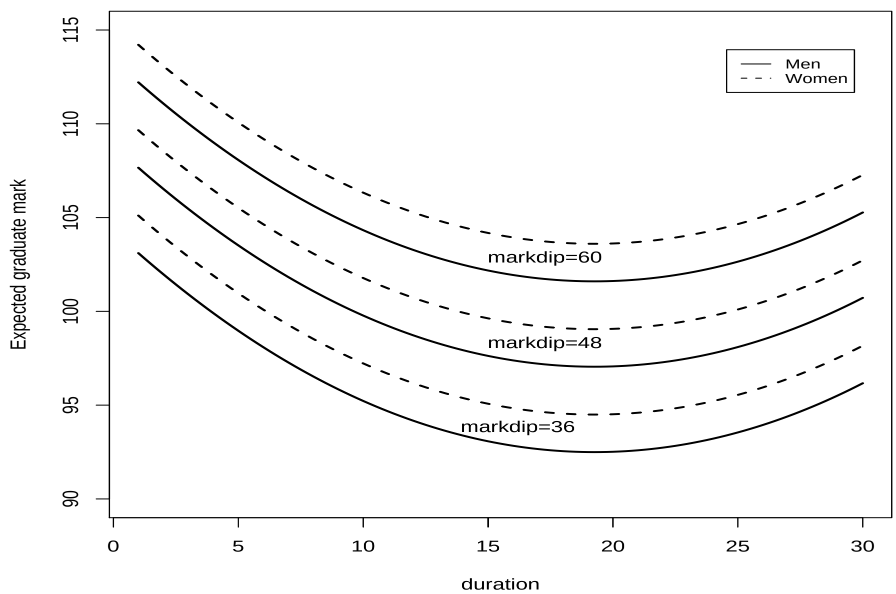

Figure 6. Expected marks implied by the estimated regression model

$Dur = \exp(-\hat{\gamma_3}/\hat{\gamma_4}) = 9.409$ years, for any Diploma mark. This estimate is more consistent with the real behavior of students when time spent at University becomes too long.

A further characteristic of the estimated CUB models is that they are able to capture even low grades, belonging to class F. Finally, thanks to the significant interaction effect, the models predict a general decay as a function of Duration but, after 9.4 years of enrollment, the acceleration for women is more and more evident. This last finding is empirically ascertainable and deserves consideration.

8. Concluding remarks

The case study discussed in this paper suggests a better performance of qualitative models in the tails of the distribution caused by the robust-

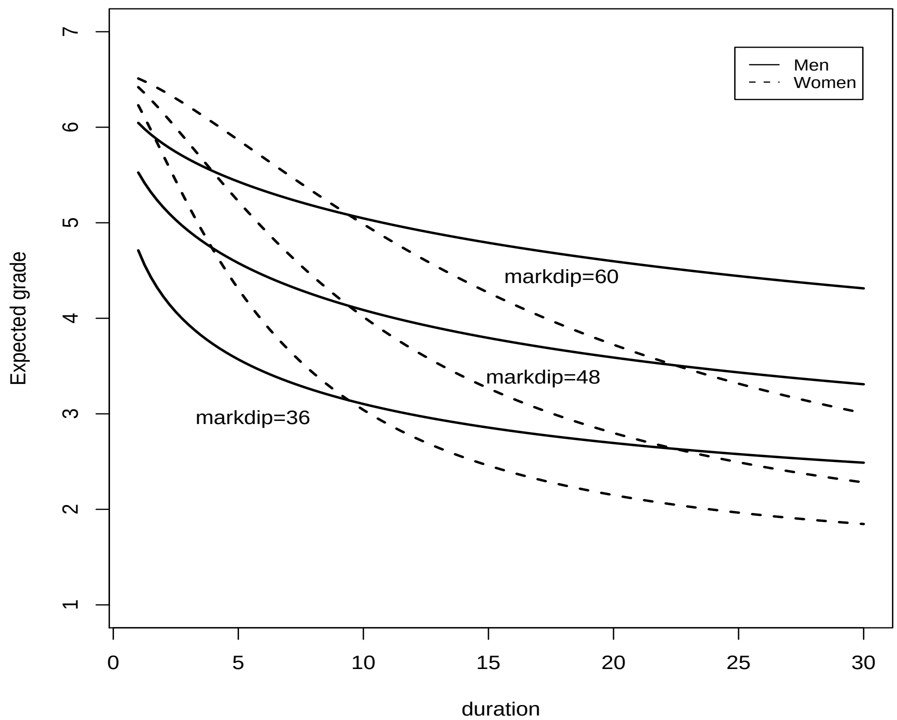

Figure 7. Expected grades implied by the estimated $CUB(0,4)$ model

ness property of ordinal values. This is a general issue that can be easily maintained also in different contexts.

From a practical point of view, we found that a qualitative data modelling should be considered as a sensible alternative approach for this kind of data since the quantitative determination of graduates marks is derived from a qualitative assessment. In such cases, any choice stems from a composite procedure, where several latent variables are to be combined for producing the final evaluation. Then, the proposed mixture solution turns out to be adequate for interpreting and fitting the observed data.

Finally, in further studies, we plan to deepen the comparative performance of qualitative and quantitative models by means of measures of predictive ability.

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 main result have been presented at the 7th International Conference on Social Sciences Methodology, RC33, Naples, 1-5 September 2008, and a preliminary version has been published in Piccolo and Iannario (2008b). Critical suggestions by Editor and a referee are gratefully acknowledged.

References

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

Bock R. D., Moustaki I. (2007), *Item response theory in a general framework*. In Psychometrics, C.R.Rao and S.Sinharay, eds., Handbook of Statistics, vol.26, 469–513.

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

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.

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

Iannario M. (2008a), A class of models for ordinal variables with covariates effects, *Quaderni di Statistica*, this issue.

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

Iannario M., Piccolo D. (2008), Modelling *shelter* choices in ordinal surveys, *preliminary report*.

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

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.

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

Piccolo D., Iannario M. (2008b), Qualitative and quantitative models for ordinal data analysis, *Proceedings of MTISD 2008, Methods, Models and Information Technologies for Decision Support Systems*, Università del Salento, Lecce, 140–143.