Quaderni di Statistica
Vol. 11, 2009

Fitting measures for ordinal data models

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

Summary: A relevant issue for validating models is the assessment of goodness-of-fit and related measures of predictive ability. When data are nominal, and specifically ordinal, the main problem is the absence of a standard paradigm as in the regression framework for residual variability; in fact, several measures have been proposed. In this contribution we explore fitting measures for ordinal data when these are modelled by a mixture distribution. Some new indexes are evaluated and a comparison with previous proposals is performed by means of simulated and real data sets.

Keywords: Fitting measures, Ordinal data, Mixture model

1. Introduction

In several contexts sampling data are collected for acquiring knowledge about perception, feeling, evaluation, concern, ability with regard to people, sentences, objects, rules, services, and so on. Such information are conveyed by choosing comparative qualitative assessments, generally graduated on some Likert-type scale and expressed as ordinal values. Conventionally, we rename these categories as the first m integers: this parameterization acts as a convenient tool for both simplifying discussion and introducing random variable theory in a qualitative framework.

Exploratory analysis is pursued with ordinal data by several approaches but the main interest of researcher is generally devoted to build efficient models given sampled data in order to relate expressed responses to subjects' covariates.

For such data, statisticians introduced the Generalized Linear Models

(GLM) framework as fully discussed by McCullagh (1980), McCullagh and Nelder (1989). A different approach has been proposed by Piccolo (2003), D'Elia and Piccolo (2005a), Iannario and Piccolo (2009a) leading to a class of models of increasing complexity, denoted as CUB. In both cases, the estimation and validation steps are discussed and explored by exploiting maximum likelihood (ML) theory.

However, more consideration is necessary when we deal with fitting and prediction issues given the *nature* of ordinal data. This stems from the circumstance that "while the models predict probabilities, they must be tested on observed events" (Hauser, 1978, 407); that is, we validate a whole probability distribution whereas, for each sampled subject, we only observe a single realization of his/her choice and not the whole assessment of his/her probability distribution.

In this context, from the vast literature (both statistical and econometric one) we select only measures which seem more relevant for a simple and effective use in a class of models for ordinal data. Specifically, we only mention fitting measures which we consider as fundamental for successive approaches. We quote, for example, the probit and logit proposals by Haglie and Mitchell (1992), Veall and Zimmermann (1990a) and the indexes for Tobit models as in Veall and Zimmermann (1990b) and Laitila (1993). Similarly, we refer to the literature for fitting measures when data are generated by counts, and Poisson distributions are involved, as in Cameron and Windmeijer (1996). In any case, in this work we do not discuss generalized residuals, as in Franses and Paap, (2001), 123 and Hübler (1997), and the prospect of comparison of model-based predictions and realizations by association tables and related indexes, as listed by Bishop *et al.* (1977), Veall and Zimmermann (1992, 1996), Meinel (2009), for instance.

This paper is aimed at exploring fitting measures by first principles, statistical indexes and empirical analysis within a class of model for ordinal data. It is organized as follows: after establishing basic notation, we discuss in section 3 some preliminary statements on the joint problems of fitting and prediction for ordinal data. Then, in sections 4-5 we consider the main indexes proposed in the literature by selecting those more relevant to our purposes, while section 6 is devoted to a comparative assessment among them. In sections 7 and 8, respectively, we synthesize empirical evidence by extensive simulations and a real case study. Some concluding remarks end the paper.

2. Basic notation for ordinal models

Responses given to rating and evaluation surveys are generally qualitative ranging from "extremely satisfied" to "extremely dissatisfied", or "complete agreement" to "complete disagreement", and so on. The number of admissible categories - denoted by m - varies from as low as 3 to more than 10. The choice of an optimal value of m is a statistical problem *per se* which deserves attention but, in this paper, we assume that m is given and known in advance. In this respect, we subscribe the requirement advocated by McCullagh and Nelder (1989), 151: "the nature of the conclusions should not be affected by the number or choice of response categories" and the pragmatic consideration that the selected paradigm should "work well in practice".

In order to define a statistical experiment for rating surveys it is convenient to map the (ordinal) manifest expressions of the interviewees into the first integers number set. As a consequence, we assume that the response R of a subject is a discrete random variable fully specified by a well defined probability mass function $p_r(\boldsymbol{\theta})$ on the support $\mathbb{S}_m = \{1, 2, \dots, m\}$, characterized by a vector of parameters $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta})$.

Then, survey data are an observed sample $\boldsymbol{r} = (r_1, r_2, \dots, r_n)'$, which is a realization of the random sample $(R_1, R_2, \dots, R_n)$. We denote by $\boldsymbol{X}$ a $n \times (k+1)$ matrix of observed k covariates related to n subjects; then, the row $\boldsymbol{x}_i = (x_{0i}, x_{1i}, \dots, x_{ki})$ contains the measurements of k variables on the i-th subjects, for $i = 1, 2, \dots, n$. The convention: $x_{0i} \equiv 1$, $\forall i = 1, 2, \dots, n$ simplifies the notation. Afterwards, this matrix may be splitted as $\boldsymbol{Y}$ and/or $\boldsymbol{W}$ matrices, if convenient.

In the following, $L(\boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$ will denote likelihood and log-likehood functions, respectively. Then, we introduce a dummy variable related to the i-th respondent when he/she chooses the r-th category:

$$d_{ir} = \begin{cases} 1, & \text{if } y_i = r\,; \\ 0, & \text{otherwise}\,. \end{cases}$$

Thus, given the sample data $\boldsymbol{r}$, the general log-likelihood function is:

$$\ell(\boldsymbol{\theta};\boldsymbol{r}) = \ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{r=1}^{m} d_{ir}\,\log(Pr\,(R_i = r \mid \boldsymbol{x}_i))\,.$$

When covariates are present, the corresponding parameter vectors will be denoted by $\boldsymbol{\beta}$ and/or $\boldsymbol{\gamma}$, as convenient.

This basic notation may be converted when we define ordinal structures (as GLM or CUB models, for example).

3. Fitting, prediction and information content

Fitting concerns the validation of observed data with reference to an hypothesis (a postulated model) whereas prediction concerns the validation of data postulated by the model with respect to future or different situations (assuming the truth of an estimated model). These aspects are logically connected if one adheres to the *axiom of model persistency*: by exploiting within-sample (out-of-sample) data we check the validity of the model to reproduce observed (new) data. We assume in any case a model stability over space, time and occasions, although we do not (and never will) know the data generating process. Thus, previous statements should be considered in terms of informational content of both model and data:

- *structural viewpoint*: it may happen that phenomenon at hand is a totally random experiment and data are genuine realizations of an unpredictable random variable. In this situation the *best model* which a statistician should fit is correctly the *worst predictive tool* both within- and out-of-sample horizons.
- *statistical viewpoint*: it may happen that phenomenon at hand is generated by a very complex mechanism and thus the entertained

model is just a first step that requires further improvements. The achieved model should be checked by fitting and prediction measures and a statistician should look for better modelling schemes by maintaining parsimony and efficiency criteria.

- *interpretative viewpoint*: it may happen that phenomenon at hand is to be interpreted as generated by a formal mechanism involving exogenous covariates whose relevance has to be tested within the fitting and predictive paradigm.

If we translate these viewpoints in terms of ordinal data, the totally random experiment is a discrete Uniform random variable defined on the support of m categories. In a sense, the Uniform distribution is the equivalent of a regression model with only intercept since it has no parameters (m is given). Thus, any validation test should take this benchmark into account as the extreme unsatisfactory situation when performing a statistical analysis on categorical ordinal data.

As a consequence, in proposing indexes one should consider three sequential steps:

- *Uniformness*: it measures how far the estimated model is from a completely unpredictable situation. If sample size is adequate, this measure is a consistent proxy for the information content of the model as supported by data.

- *Standard Features*: it measures how a parametric model within a predefined class (without covariates) is able to improve fitting to data by including parameters in a parsimonious manner.

- *Significant Covariates*: it measures how large is the additional contribution of covariates with respect to previous models. In the regression context, this aspect is achieved by comparing estimated models with covariates to models with only intercept.

In the following, we will mostly deepen the first two steps and we briefly mention some proposal for the third.

4. Overview of fitting measures for ordinal data models

In this section, we briefly overview the main approaches suggested for introducing adequate measures of fitting: generally, they are first proposed for binary responses and then generalized to polytomous data, especially in ordinal contexts. We observe that R^2 is a standard benchmark as a measure of "proportion of the total variability explained by the model"; thus, most proposals mimic its definition, reproduce properties and allow similar interpretations.

Although there is great overlapping among the different approaches, we will distinguish among testing, latent variables, likelihood frameworks and AIC-type measures.

4.1. Testing approach

The first attempt to recover definition and properties of fitting measures in general models stems by Dhrymes (1986) from the relationship between R^2 and F-test. Indeed, in classical regression model the statistic:

$$F = \frac{R^2/k}{R^2/(n-k-1)}$$

allows testing the null hypothesis of no joint relevance of included covariates. Thus, the starting point for generating R^2 measures from a testing perspective is the inverse relationship:

$$R^2 = \frac{k\,F}{n-1-k(1-F)}\,.$$

This line of reasoning uses likelihood ratio test (LRT) as a test measure; indeed, a plethora of proposals with several variants have been released. This strong relationship has been emphasized by Pudney (1989), 109, whereas Vandaele (1981) and Magee (1990) analyse the connections among R^2, F and Wald tests.

4.2. Latent variables approach

As already noted, in ordinal data analysis, we derive an estimated probability distribution for each i-th subject whereas we possess only an observed values y_i to be compared with. As a consequence, in GLM literature, to avoid such situation and emulate regression model characteristics several measures have been proposed by introducing standard indexes for latent variables.

For instance, McKelvey and Zavoina (1975) consider regression models on latent variables y_i^* estimated by data and introduce a pseudo-R^2 measure defined by:

$$R^2_{MZ} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i^* - \overline{y}^*\right)^2}{\sum_{i=1}^{n} \left(\hat{y}_i^* - \overline{y}^*\right)^2 + n\,\sigma^2} = \left[1 + \frac{\sigma^2}{\sum_{i=1}^{n} \left(\hat{y}_i^* - \overline{y}^*\right)^2 / n}\right]^{-1},$$

where $\sigma^2 = 1$ or $\sigma^2 = \pi^2/3$ according to standard Gaussian or logistic distribution for the random errors ϵ_i, respectively. This index derives from the standard decomposition $y_i = \hat{y}_i^* + \epsilon_i$, where the latent variables $\hat{y}_i^* = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$ are estimated by using observed covariates (Windmeijer, 1995).

Simulation experiments support the conclusion that R^2_{MZ} is a good proxy for an R^2-type index if one knew the (unobservable) latent variables: Hagle and Mitchell (1992), Veall and Zimmermann (1996). In addition, the measure has been generalized to multivariate models by Spiess and Keller (1999) and further analyzed by Spiess and Tutz (2004) and Meinel (2009).

For dichotomous data, Lave (1970) compared observations with their distribution function by the measure:

$$R^2_{LA} = 1 - \frac{\sum_{i=1}^{n} (y_i - F(y_i))^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2},$$

and similar measures have been advocated by Efron (1978) in the context of standard regression and analysis of variance.

4.3. Likelihood-based measures

Since likelihood function is a sufficient statistic for sampled data given a postulated model, it is reasonable to relate data and model by means of this quantity (and corresponding log-likelihood function). A common feature is the property that by adding covariates likelihood functions never decrease. Sometimes, these measures are called pseudo-R^2 and we mention several of them, although they are strongly related each other. Specifically, a relevant issue is the derivation of log-likelihood functions of *null, estimated* and *saturated models* obtained when only a constant, parsimonious structures and as many parameters as data are fitted, respectively.

A simple starting point is the quantity $LRT = 2(\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_0))$ whose maximum $LRT_{max} = 2(0 - \ell(\hat{\boldsymbol{\theta}}_0)) = -2\,\ell(\hat{\boldsymbol{\theta}}_0)$ may be introduced for a normalized version, as in the next McFadden's proposal. Similar indexes, as those by Aldrich and Nelson (1984) normalized by Veall and Zimmermann (1992), obey the same logic and are not discussed hereafter.

For qualitative data models, one of the oldest proposal has been credited to McFadden (1974) who introduced a pseudo-R^2 index (denoted also as *likelihood ratio index*):

$$R^2_{MF} = 1 - \frac{\ell(\hat{\boldsymbol{\theta}})}{\ell(\hat{\boldsymbol{\theta}}_0)}\,,$$

where $\ell(\hat{\boldsymbol{\theta}})$ and $\ell(\hat{\boldsymbol{\theta}}_0)$ are the log-likelihood functions evaluated with ML estimates $\hat{\boldsymbol{\theta}}$ with the proposed model and only with a constant (intercept) $\hat{\boldsymbol{\theta}}_0$ model, respectively. Its range is $[0, 1]$ although unity is reached only if at least one explanatory covariate explodes to $\pm\infty$.

McFadden's index is regularly applied for its simplicity although some Authors criticize the absence of any substantive interpretation. Hauser (1978) tackles this problem within information theory framework: how much covariates does reduce entropy (that is, uncertainty) of the system? His paper concludes that R^2_{MF} has an information-theoretic interpretation as the proportion of uncertainty explained by data: this approach has been deepened also in Judge *et al.* (1985), 773-777 by using entropy and divergence concepts[1].

[1] Within the divergence paradigm, we mention the likelihood ratio for testing a multinomial

Cameron and Windmeijer (1997) support R^2 measures derived by considering the Kullback-Leibler (KL) divergence for densities parameterized only by their mean values; thus, their consideration could be not translate in CUB paradigm where different links are eligible. In fact, their study is applied to exponential family regression models (where the scale parameter is known or absent) and it is based upon the following definition of KL divergence (factor 2 is for distributional convenience):

$$KL(\boldsymbol{y}\,,\hat{\boldsymbol{\theta}}) = 2[\ell(\boldsymbol{\theta}_{sat}) - \ell(\hat{\boldsymbol{\theta}})]$$

where $\ell(\boldsymbol{\theta}_{sat})$ is the log-likelihood for a saturated model with as many parameters as observations. Since deviance is "twice the difference between the maximum achievable log-likehood function and that attained under the fitted model" (McCullagh and Nelder, 1989, 33; 118), it turns out that the KL-divergence is nothing else that the deviance (Hastie, 1987).

Now, the order relationship among log-likelihood functions of null, estimated and saturated model is:

$$\ell(\hat{\boldsymbol{\theta}}_0) \leq \ell(\hat{\boldsymbol{\theta}}) \leq \ell(\hat{\boldsymbol{\theta}}_{sat})\,.$$

and thus we may derive the following *deviance decomposition* (suggested by Cameron and Windmeijer, 1993):

$$\underbrace{\ell(\hat{\boldsymbol{\theta}}_{sat}) - \ell(\hat{\boldsymbol{\theta}}_0)}_{\textit{Total Deviance}} = \underbrace{[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_0)]}_{\textit{Explained Deviance}} + \underbrace{[\ell(\hat{\boldsymbol{\theta}}_{sat}) - \ell(\hat{\boldsymbol{\theta}})]}_{\textit{Unexplained Deviance}}$$

where in the brackets we put the explained and unexplained quantities of the log-likelihood gains.

distribution. It is proportional to the corresponding divergence $KL(\boldsymbol{f}\,,\hat{\boldsymbol{p}})$ between the vector of observed relative frequencies $\boldsymbol{f}$ and probabilities $\hat{\boldsymbol{p}} = \boldsymbol{p}_r(\hat{\boldsymbol{\theta}})$ according to the well known G^2 index:

$$G^2 = 2\,n \sum_{r=1}^{m} f_r \,\log \quad \frac{f_r}{p_r(\hat{\boldsymbol{\theta}})}\Bigg) = n\,KL(\boldsymbol{f}\,,\hat{\boldsymbol{p}}) \simeq X^2\,,$$

where we denoted by X^2 the standard Pearson fitting measure. Since both G^2 and X^2 are not normalized and are also criticized for moderate or large n (as their significance is extreme in these cases even for acceptable fitting), the study of such measures will not be pursued hereafter.

Finally, according to the R^2 paradigm, these Authors advanced a proposal for any model specification:

$$R^2_{KL} = \frac{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_0)}{\ell(\hat{\boldsymbol{\theta}}_{sat}) - \ell(\hat{\boldsymbol{\theta}}_0)} \, .$$

A possible interpretation is that: "R^2_{KL} measures the proportionate reduction in recoverable information due to the inclusion of regressors, where information is measured by the estimated Kullback-Leibler divergence" (Cameron and Windmeijer, 1997, 333). Of course, last definition of R^2_{KL} is a sensible one if $\ell(\hat{\boldsymbol{\theta}}_{sat})$ is easily computable for the given model.

Kent (1983) and Maddala (1983) introduced for binary models a measure related to coefficient of determination arguing that if likelihood is the criterion of fitting, its reduction should be a convenient measure; moreover, when the model is a classical linear regression, a consistent proposal should at least coincides with traditional R^2.

Thus, both Cox and Snell (1989, 208-209) and Magee (1990) proposed the following index:

$$R^2 = 1 - \left(\frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} \right)^{2/n}$$

where $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}$ are ML estimates of "null" and estimated model, respectively. Its rationale stems from the circumstance that in a linear regression model with Gaussian errors previous definition is just the formula connecting R^2 to likelihood ratio.

In discrete random variables, this quantity should be normalized since the maximum is by far less than 1 (it happens when all observations collapse at a single category) and it is $1 - L(\hat{\boldsymbol{\theta}}_0)^{2/n}$. Then, a better measure referred to Cragg and Uhler (1970), 400 and Maddala (1983), 39 is:

$$R^2_{CU} = \frac{1 - \left(\frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} \right)^{2/n}}{1 - L(\hat{\boldsymbol{\theta}}_0)^{2/n}} \, .$$

Several properties of this measure are discussed by Nagelkerke (1991) who shows its consistency with standard properties of R^2 measures.

A further proposal has been advanced by Estrella (1998) for models with dichotomous dependent variables by adding to standard requirements of a fitting measure the need of adequate interpretation of intermediate values. After solving a differential equation, it turns out that such index is defined by:

$$\phi_0 = 1 - \left(\frac{\ell(\hat{\boldsymbol{\theta}})}{\ell(\hat{\boldsymbol{\theta}}_0)} \right)^{-(\frac{2}{n})\,\ell(\hat{\boldsymbol{\theta}}_0)} = 1 - \left(1 - R^2_{MF}\right)^{-(\frac{2}{n})\,\ell(\hat{\boldsymbol{\theta}}_0)} ,$$

where the second formulation relates ϕ_0 to McFadden's proposal. As noticed by Estrella (1998), 200, the applicability of ϕ_0 extends beyond the dichotomous case since the requirements upon which it has been derived are also valid for polytomous models; then, we will use it in the comparative analysis performed in section 6.

4.4. AIC-type indexes

A more general approach derived from likelihood function (not restricted to the comparison of nested models) leads to AIC-type measures, that is:

$$AIC = -2\,\ell(\hat{\boldsymbol{\theta}}) + 2(k+1)$$

where $(k+1)$ is the number of parameters for estimating $\boldsymbol{\beta}$. In the likelihood estimation of different models on the same sampled data, AIC acts as a mixing between a not-decreasing function (that is: $-2\,\ell(\hat{\boldsymbol{\theta}})$) and an increasing linear function of the number of parameters (that is: $2(k+1)$). Thus, a model with a minimum AIC is derived from a compromise between efficiency and parsimony.

Since AIC tends to overparameterize the preferred model, some corrections have been introduced for reducing this effect. Generally, it is better to use the Schwarz information criterion defined as BIC (=*Bayesian Information Criterion*), that is:

$$BIC = -2\,\ell(\hat{\boldsymbol{\theta}}) + (k+1)\,\log(n)$$

Of course, $BIC = AIC + (k+1)\,(\log(n) - 2)$ and $BIC > AIC$ as long as $n > 7$; then, in any relevant applications, it limits the overparameterization effect of AIC.

With obvious notations, it is immediate to derive:

$$R^2_{MF} = 1 - \frac{AIC(\hat{\boldsymbol{\theta}}) - 2(k+1)}{AIC(\hat{\boldsymbol{\theta}}_0) - 2}\,.$$

Finally, in this area, we mention the AIC-type proposal by Estrella (1998), 203, aimed at modifying ϕ_0 in order to take also the number of parameters into account.

5. Fitting measures for CUB models

We remember that parameters of CUB models are usually denoted by $\boldsymbol{\theta} = (\pi, \xi)'$ and belong to the open-left parametric space:

$$\boldsymbol{\theta} \in \Omega(\pi, \xi) = \{(\pi, \xi) : \; 0 < \pi \leq 1, \; 0 \leq \xi \leq 1\}\,.$$

Moreover, identifiability requires $m > 3$, as shown by Iannario (2009a).

In this regard, we will discuss how previous proposals may be translated in the CUB model framework and introduce some new ones as convenient. The first point to notice is the circumstance that uncertainty is a starting point in statistical modelling of ordinal models, and thus the first benchmark is to measure how estimated models are able to improve it. In addition, one should consider that CUB models logically assume the existence of latent variables for both components but they never consider their estimates; as a consequence, we could not apply the measures of subsection 4.2.

5.1. Uniformness measures

The amount of uncertainty in data and model is mostly related to heterogeneity and does not coincide with the common concept of variability. As a matter of fact, while a discrete Uniform random variable maximizes entropy on a finite discrete support, the maximum variability over a discrete support, given expectation, is achieved by a random variable whose probability mass is halved at categories 1 and m, respectively (for any $m > 2$).

Among the heterogeneity measures computed for the estimated probability $\hat{p}_r = p_r(\hat{\boldsymbol{\theta}})$, we remember Gini (1912), Frosini (1981, 2003) and Laakso and Taagepera (1979) indexes. In the normalized formulations, they are defined, respectively, by:

$$\mathcal{G}^* = \frac{m}{m-1}\left(1 - \sum_{r=1}^{m} \hat{p}_r^2\right); \;\; \mathcal{F}^* = 1 - \sqrt{1-\mathcal{G}^*}; \;\; \mathcal{A}^* = \frac{\mathcal{G}^*}{m - \mathcal{G}^*(m-1)}\,.$$

For more discussion on this topic and related concepts, see: Haberman (1982), Patil and Taillie (1982), Leti (1983), Grilli and Rampichini (2002) and quoted references.

As heterogeneity is a complementary concept with respect to uncertainty, we will introduce direct (and normalized) *measures of uncertainty*:

$$\mathcal{G} = \frac{m\sum_{r=1}^{m} \hat{p}_r^2 - 1}{m-1}; \qquad \mathcal{F} = \sqrt{\mathcal{G}}; \qquad \mathcal{A} = \frac{m\,\mathcal{G}}{1+(m-1)\,\mathcal{G}}\,,$$

firstly adopted for CUB models by D'Elia and Piccolo (2005a). All of them depend on the Gini index (indeed, they are direct function of the sum of squared probabilities) and thus, given its simplicity, we will work with $\mathcal{G}$ as a starting point for further investigations.

Instead, a likelihood-based measure is the *ICON* index (=*I*nformation *CON*tent) introduced for CUB models (Iannario, 2008; Piccolo, 2008) and currently computed in the implemented software; it compares the log-likelihood function at the maximum given the data and at the null model $\ell_0 = -n\,\log(m)$, that is for a discrete Uniform distribution (which possesses maximum entropy). Thus, we get the measure:

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

Notice that $ICON$ belongs to McFadden's family and it is a normalized index varying between 0 (when data support a Uniform distribution) and 1 (when data collapse to a degenerate distribution at $R = 0$ or $R = 1$, in standard CUB model, or at $R = c \in \{1, 2, \dots, m\}$ in extended models with a *shelter* effect; see: Iannario, 2009b). It should be added that a long experience on this index confirms that for real data sets it generally assumes quite small values; thus, it is of limited help for effective discrimination among models.

5.2. Standard features measures

For a given $m > 3$, information contained in the sample $(r_1, r_2, \dots, r_n)'$ are strictly equivalent to that contained in the frequencies $(n_1, n_2, \dots, n_m)'$ of ordered categories or the relative frequencies $\boldsymbol{f} = (f_1, f_2, \dots, f_m)'$, where $f_r = n_r/n$, $r = 1, 2, \dots, m$.

Thus, the log-likelihood for the *saturated* CUB model (without covariates) is obtained as:

$$\ell_{sat} = -n\, \log(n) + \sum_{r=1}^{m} n_r\, \log(n_r) = -n\, \mathcal{E}(\boldsymbol{f})\,.$$

where

$$\mathcal{E}(\boldsymbol{f}) = -\sum_{r=1}^{m} f_r\, \log(f_r)\,,$$

is the empirical entropy computed on the frequency distribution. This consideration is correct if and only if no further information (covariates) are available about subjects and thus we are using the maximum number of admissible parameters, given the data.

Then, a normalized measure of fitting (analogous to R^2_{KL} of section 4.3) is:

$$\mathcal{I} = \frac{\ell(\hat{\boldsymbol{\theta}}) - \ell_0}{\ell_{sat} - \ell_0} = \frac{\displaystyle\sum_{r=1}^{m} f_r\, \log(\hat{p}_r) + \log(m)}{\displaystyle\sum_{r=1}^{m} f_r\, \log(f_r) + \log(m)}\,,$$

The last expression confirms that $\mathcal{I}$ compares entropies of (observed) relative frequencies and (estimated) CUB probabilities with respect to the extreme one (that is, discrete Uniform). This index tends to 0 if and only if the estimated probability converges to the Uniform distribution and it is 1 if and only if a perfect fitting is achieved.

5.3. Significant covariates measures

We will briefly discuss the modifications induced in the previous measures when we are interested in the improvement of model gained by introducing covariates. Shortly, this is equivalent to consider a CUB model without covariates as the new "null" model to compare with estimated and saturated.

If we denote as $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ the parameter vector of covariates in a CUB model, a measure of the fitting achieved by adding these covariates is based on:

$$\mathcal{I}(\boldsymbol{\delta}) = \frac{\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) - \ell(\hat{\boldsymbol{\theta}})}{\ell_{sat} - \ell(\hat{\boldsymbol{\theta}})},$$

where the log-likelihood functions are properly defined.

Its rationale stems from an extended deviance decomposition (similar to that discussed in subsection 4.3). The precise definition of ℓ_{sat} depends on the desired comparison. The latter specification seems difficult to assess in general unless covariates are discrete as we elaborate in the next subsection.

5.4. Significant covariates measures in discrete subgroups

Suppose that a significant covariate, denoted by Y, is splitted into J categories by its very nature (as it happens for: gender, education, region, $\dots$) or by convention (as it happens for: income, age, distance, $\dots$) and suppose that we cluster the ordinal responses distributions in such a way that relevant quantities are shown in Table 1. Specifically, we consider the observed (absolute and relative) frequency distributions of the j-th cluster (consisting of n_j subjects, for $j = 1, 2, \dots, J$).

Then, we denote by $\ell_{sat(j)}$ the log-likelihood function of the saturated

Table 1. Observed (relative) frequency distributions of a subgroup.

Categories	1	2	$\dots$	r	$\dots$	m	Total
Absolute Frequencies	n_{1j}	n_{2j}	$\dots$	n_{rj}	$\dots$	n_{mj}	n_j
Relative Frequencies	f_{1j}	f_{2j}	$\dots$	f_{rj}	$\dots$	f_{mj}	1

model for the j-th cluster:

$$\ell_{sat(j)} = -n_j \, \log(n_j) + \sum_{r=1}^{m} n_{rj} \, \log(n_{rj}) = -n_j \, \mathcal{E}(\boldsymbol{f}_j) \, ,$$

where $\boldsymbol{f}_j = (f_{1j},\, f_{2j},\, \dots,\, f_{mj})'$ is the vector of relative frequencies of the j-th subgroup.

Since clusters are independent with respect to responses given the covariates, the log-likelihood function of the saturated model will be:

$$\ell_{sat^*} = \sum_{j=1}^{J} \ell_{sat(j)} = -\sum_{j=1}^{J} n_j \, \mathcal{E}(\boldsymbol{f}_j) \, .$$

In this way, previous indexes may be easily computed and referred to a saturated model when a covariate is intrinsically dichotomous or polytomous.

For continuous covariates, we first suggest to check a coarse subdivision of their range and then to split data into finer and finer intervals in order to verify if, by increasing the number of subdivisions, there is some convergence in ℓ_{sat} measures. Of course, this proposal requires samples of adequate sizes in order to obtain sensible clusters.

5.5. A normalized fitting index

Given the inability of common X^2 index to detect adequacy of the fitted models in several case studies (D'Elia and Piccolo, 2005a), new measures based on the comparison among observed relative frequencies and estimated probabilities have been introduced for CUB models without covariates.

Then, a normalized fitting measure $\mathcal{F}^2$ is defined by:

$$\mathcal{F}^2 = 1 - Diss = 1 - \frac{1}{2} \sum_{r=1}^{m} \mid f_r - p_r(\hat{\boldsymbol{\theta}}) \mid \ .$$

where the dissimilarity index $Diss$ measures the fraction of respondents that should change selection in order to achieve a perfect fit (Leti, 1983; Simonoff, 2003) and it is often computed as a benchmark for judging the adequacy of the model: values of $\mathcal{F}^2 \geq 0.90$ are considered as compatible with an acceptable fitting.

In other contexts, $Diss$ has been used for measuring the relevance of a *shelter effect* in extended CUB model, as advocated by Iannario (2009b) and Corduas *et al.* (2009). Some modifications are necessary for adopting this measure when significant covariates are present.

5.6. A new likelihood-based proposal

A model-dependent measure of dissimilarity is currently computed in the released version 2.0 of the software for CUB model inference (Iannario and Piccolo, 2009a). It relies on the ML estimates and it is strictly dependent upon the correctness of this formalization.

Specifically, the log-likelihood function $\ell(\boldsymbol{\theta}) = \ell(\pi, \xi)$ for a CUB model without covariates, is given by:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n} r_i \, \log(p_r(\boldsymbol{\theta})) = n \sum_{r=1}^{m} f_r \, \log(p_r(\boldsymbol{\theta})) \\ &= n \sum_{r=1}^{m} f_r \, \log\left[\pi \left(b_r(\xi) - \frac{1}{m}\right) + \frac{1}{m}\right] . \end{aligned}$$

Then, it is easily found that:

$$0 = \frac{\partial \ell(\pi, \xi)}{\partial \pi} = n \sum_{r=1}^{m} f_r \, \frac{b_r(\xi) - \frac{1}{m}}{p_r(\pi, \xi)} = n \, \frac{1}{\pi} \sum_{r=1}^{m} f_r \, \left(1 - \frac{1}{m \, p_r(\pi, \xi)}\right) .$$

As $\pi > 0$, from the last constraint, we deduce that ML estimates of CUB model parameters must obey the relationship:

$$\frac{1}{m}\sum_{r=1}^{m}\frac{f_r}{p_r(\hat{\pi},\hat{\xi})} = 1\,.$$

This is a noticeable result as it assesses that the average of ratio between observed relative frequencies and ML estimated probabilities fluctuates around 1: this property is analogous to the OLS constraint on residuals when intercept is present in a regression model.

Of course, large fluctuations imply bad fitting whereas a perfect fit requires $f_r \equiv p_r(\hat{\pi},\hat{\xi})$, $\forall r = 1, 2, \ldots, m$. Then, the variance of the ratios $\left(\dfrac{f_r}{p_r(\hat{\pi},\hat{\xi})}\right)$ is a measure of dissimilarity and we denote it as:

$$\begin{aligned}\tilde{D}^2 &= \frac{1}{m}\sum_{r=1}^{m}\left(\frac{f_r}{p_r(\hat{\pi},\hat{\xi})}-1\right)^2 = \frac{1}{m}\sum_{r=1}^{m}\frac{\left(f_r - p_r(\hat{\pi},\hat{\xi})\right)^2}{p_r^2(\hat{\pi},\hat{\xi})}\\ &= \frac{1}{m}\sum_{r=1}^{m}\left(\frac{f_r}{p_r(\hat{\pi},\hat{\xi})}\right)^2 - 1\,.\end{aligned}$$

The last formula is correct *if and only if* probabilities are exactly computed with ML estimates at convergence point; thus, we prefer the first expression to avoid numerical inconsistencies.

Notice the close relationship between $\tilde{D}^2$ and X^2 Pearson's index:

$$X^2 = n\sum_{r=1}^{m}\frac{\left(f_r - p_r(\hat{\pi},\hat{\xi})\right)^2}{p_r(\hat{\pi},\hat{\xi})}\,.$$

In fact, both of them are generated by a unique class of indexes aimed at measuring a Euclidean distance among relative frequencies and estimated probabilities described by:

$$\sum_{r=1}^{m} w_r\,\left(f_r - p_r(\hat{\pi},\hat{\xi})\right)^2\,,$$

where the weights are $w_r = n/p_r$ and $w_r = 1/(m\, p_r^2)$ for X^2 and $\tilde{D}^2$, respectively (see: von Mises (1994), 447-452, and also Cramér (1946), Read and Cressie (1988), Cressie and Read (1989), Greenwood and Nikulin (1996) for the statistical motivations leading to the class of X^2 measures). These considerations shows that $\tilde{D}^2$ is more sensible to small probabilities with respect to X^2.

Finally, since $0 \le \tilde{D}^2 \le D^2_{max} \to \infty$, a direct normalized fitting measure is:

$$\mathcal{L}^2 = 1 - \frac{\tilde{D}^2}{1+\tilde{D}^2} = \left[1 + \frac{1}{m} \sum_{r=1}^{m} \left(\frac{f_r}{p_r(\hat{\pi}, \hat{\xi})} - 1 \right)^2 \right]^{-1} .$$

It may be shown some connection of $\mathcal{L}^2$ with Laakso and Taagepera (1979) index of heterogeneity, introduced in ordinal models by D'Elia and Piccolo (2005b).

6. A comparative discussion of fitting measures

Previous discussion may be summarized by listing the measures we think useful for ascertaining goodness of fit for CUB models. Some of them have been originated as dissimilarity and discrepancy measures and then transformed into direct measures of goodness of fit. Table 2 summarizes such results in a compact formulation.

As a consequence, the benchmark of *uniformness* implies that $\boldsymbol{\theta}_0$ arises from a discrete Uniform distribution. Thus, we get:

$$L(\hat{\boldsymbol{\theta}}_0) = L(\boldsymbol{\theta}_0) = m^{-n}; \quad \ell(\hat{\boldsymbol{\theta}}_0) = \ell(\boldsymbol{\theta}_0) = -n\, \log(m)\, .$$

A simple screening shows that some indexes discussed in section 5 are strongly interrelated, in some cases by algebraic formula and in some other by logical arguments; in fact, all of them are functions of $\boldsymbol{f}$, $\hat{\boldsymbol{p}}$ and $\ell(\hat{\boldsymbol{\theta}})$.

Table 2. Information content and goodness-of-fit indexes.

•	*Information content indexes*
$\mathcal{G} = \dfrac{m \sum_{r=1}^{m} \hat{p}_r^2 - 1}{m-1}$	
$ICON2 = 1 - \left(\dfrac{\ell(\hat{\boldsymbol{\theta}})/n}{\log(m)} \right)^{2\,\log(m)}$	
•	*Goodness-of-fit indexes*
$\mathcal{F}^2 = 1 - \dfrac{1}{2} \sum_{r=1}^{m} \mid f_r - p_r(\hat{\boldsymbol{\theta}}) \mid$	
$\mathcal{L}^2 = \left[1 + \dfrac{1}{m} \sum_{r=1}^{m} \left(\dfrac{f_r}{p_r(\hat{\pi}, \hat{\xi})} - 1 \right)^2 \right]^{-1}$	
$\mathcal{I} = \dfrac{\sum_{r=1}^{m} f_r\, \log(\hat{p}_r) + \log(m)}{\sum_{r=1}^{m} f_r\, \log(f_r) + \log(m)}$,	

For instance, we remark that Estrella's measure ϕ_0 when computed for *uniformness* becomes a monotone transformation of $ICON$ index, and we will denote it as $ICON2$; in fact, after some algebra:

$$ICON2 = 1 - \left(\frac{\ell(\hat{\boldsymbol{\theta}})/n}{\log(m)} \right)^{2\,\log(m)} = 1 - (1 - ICON)^{2\,\log(m)} .$$

In addition, if $ICON < 1 - (2\,\log(m))^{-1/[2\,\log(m)-1]}$,small variations in $ICON$ are more evident in $ICON2$ (for this reason we prefer to consider the latter).

For medium and large sample size, asymptotic expansions would dis-

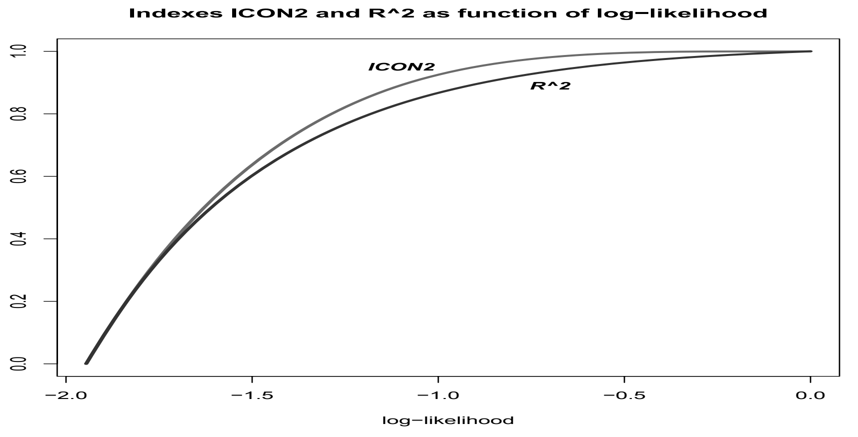

Figure 1. Plot of $ICON2$ and $\mathcal{R}^2$ as functions of $\ell(\hat{\boldsymbol{\theta}})$, when $m = 7$

cover further results for $ICON2$ index:

$$\begin{aligned} ICON2 &= 1 - (1 - ICON)^{2\,\log(m)} \\ &\simeq 2\,log(m)\left[1 + \frac{\ell(\hat{\boldsymbol{\theta}})/n}{\log(m)}\right] = n^{-1}\,2\,\left(\ell(\hat{\boldsymbol{\theta}}) - \ell(0)\right)\,, \end{aligned}$$

which is n^{-1} times the likelihood ratio test for rejecting the estimated model against the discrete Uniform hypothesis. Then, $n\,ICON2$ may be used as an approximate test of significance of *uniformness*.

Finally, Cragg and Uhler's R^2_{CU} index (introduced in 4.3) will be denoted by $\mathcal{R}^2$ when referred to *uniformness* and it simplifies to:

$$\mathcal{R}^2 = 1 + \frac{1 - \exp\left(-\frac{2}{n}\,\ell(\hat{\boldsymbol{\theta}})\right)}{m^2 - 1}$$

Now, both $ICON2$ and $\mathcal{R}^2$ are monotone functions of the average log-likelihood $n^{-1}\,\ell(\hat{\boldsymbol{\theta}}) \in [-\log(m),\,0]$, and Figure 1 shows how much these indexes are virtually coincident. Hereafter, we will refer only to $ICON2$.

7. Simulation experiment

The objective of the simulation experiment consists in simulating data from CUB models, estimating parameters by ML methods and studying empirical distributions of indexes for goodness of fit indexes listed in Figure 2.

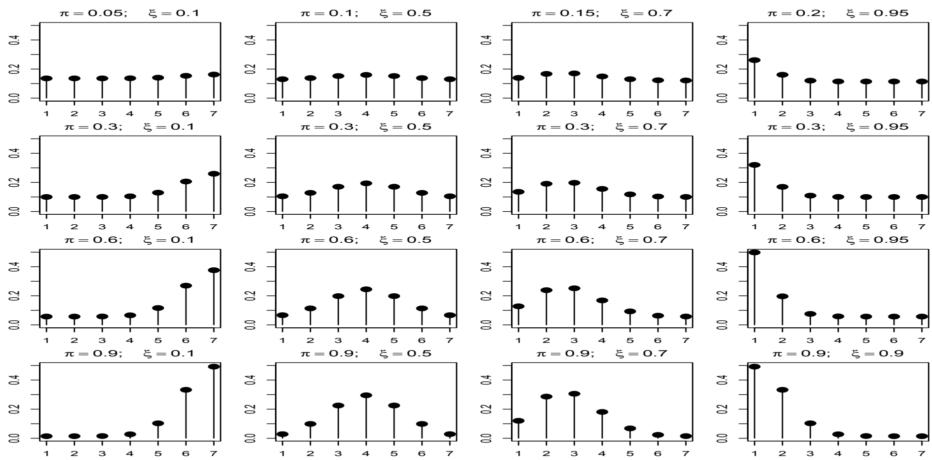

Figure 2. Probability distributions of the selected CUB models

We preliminarily choose $m = 7$, $n = 200$ and we select 16 CUB models by uniformly spanning the parametric space in order to cover both positive and negative skewed distributions and also with low and high proportion of uncertainty. These models are shown in Figure 2 for reference.

Each run of generating data and ML estimation has been performed for $nsimul = 1000$ times; as an example, Figure 3 reproduces indexes for a CUB model when $\pi = 0.9$, $\xi = 0.7$ (that is, the 15^{th} model listed in Figure 2).

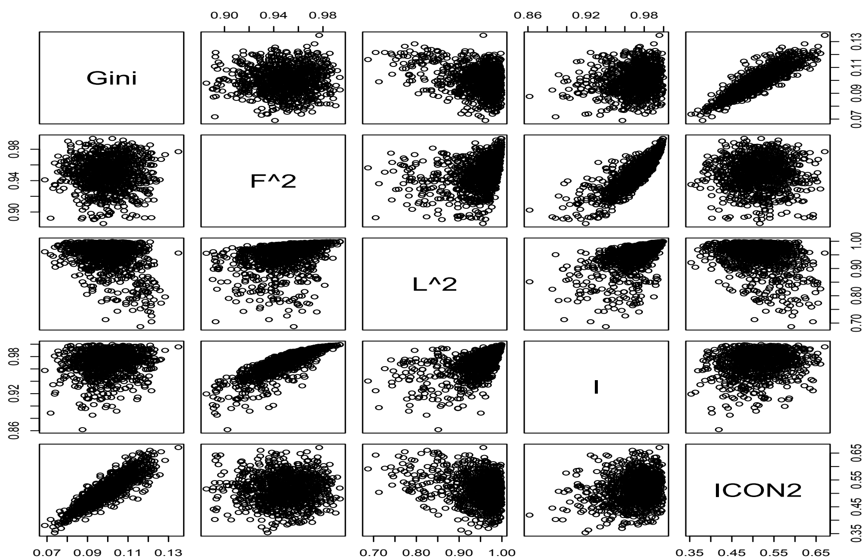

Figure 3. Scatter plots of simulated indexes for a CUB model

Simulations confirms the usefulness of a substantial reduction of several indexes and the choice to retain indexes with a joint low correlation and interesting interpretation as $\mathcal{F}^2$ and $\mathcal{L}^2$, for instance. Among other features, we notice the positive correlation of $\mathcal{G}$ and $ICON2$ indexes.

Instead, we found that sampling distributions of these indexes change their shape in proportion with the relevance of uncertainty in the model. This effect is generated by the greater variability of ML estimators (that increases as long as $\pi \to 0$) which induces similar variability in the computed indexes.

Moreover, to provide evidence of the simulated distributions, we performed 5000 simulations of random samples of size $n = 200$. Firstly, Figure 4 shows histograms and estimated kernel densities for a CUB model with $\pi = 0.15$ and $\xi = 0.70$ (the 3^{rd} among those listed in Figure 2), that is a structure with high uncertainty.

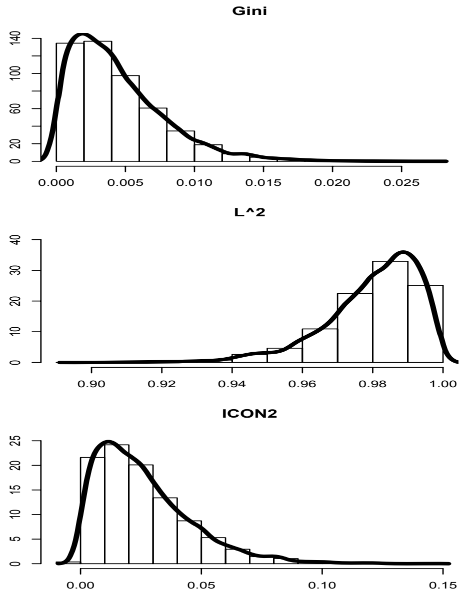

Figure 4. Simulated indexes distributions of a CUB model: $(\pi = 0.15, \xi = 0.70)$

With similar sizes, we generate the sampling distributions for a CUB model with $\pi = 0.90$ and $\xi = 0.70$ (the 15^{th} model of the list), that is a structure with a quite small uncertainty (Figure 5).

More extended simulations, here not reported for brevity, where both sample sizes and number of categories changes, support this initial screening. Thus, we register how the distribution of $\mathcal{I}$ varies from a substantial flatness to a definite skewness; similarly, we found a strong negative skewness for $\mathcal{L}^2$ and $\mathcal{I}$, and a moderate Gaussianity of $\mathcal{F}^2$. Instead, $\mathcal{G}$ and $ICON2$ have parallel behaviour and support positive skewness which reduces with symmetry of CUB model and reduction of uncertainty.

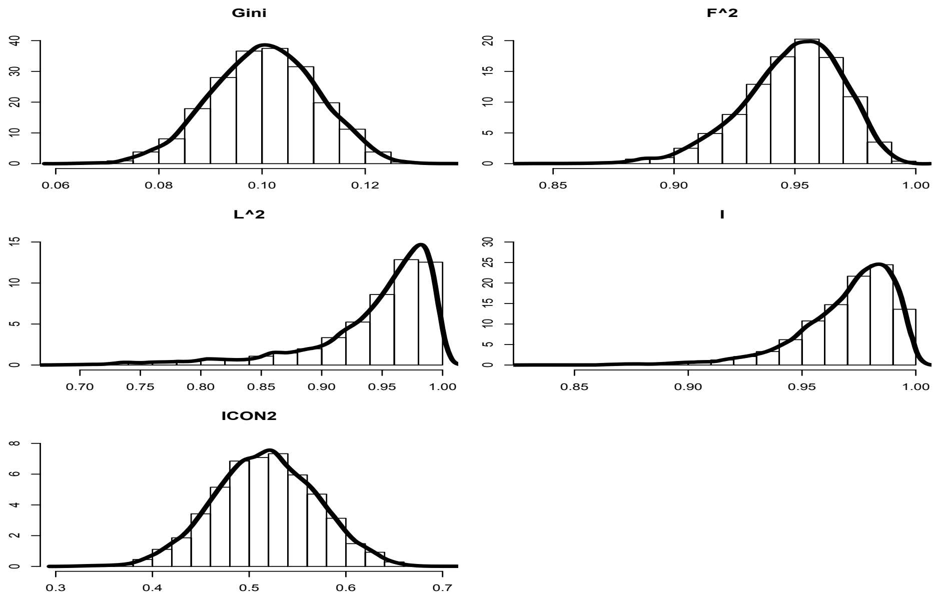

Figure 5. Simulated indexes distributions of a CUB model: $(\pi = 0.90, \xi = 0.70)$

7.1. A case study with dummy covariates

We examine faked data generated by simulating a CUB model with $m = 7$ and a dummy covariate $D_i = 0, 1$ for explaining uncertainty. Sample data consist of $n = 2000$ observations by assuming that $\xi = 0.3$ and half of them are generated by $\pi = \pi_0 = 0.2$, the others by $\pi = \pi_1 = 0.9$. In the left panel of Figure 6 we show the maintained model (circles and dots refer to $D_i = 0$ and $D_i = 1$, respectively) while in the right panel we present the relative frequency distributions of the sampled subgroups (gray and black bars refer to $D_i = 0$ and $D_i = 1$, respectively).

The main reason we are discussing this artificial data set is that the detection of different behaviour in subgroups it is not so evident if one examines the whole sample without the information of subgroups existence. Indeed, as shown in Figure 7, we obtain a very good fitting with a

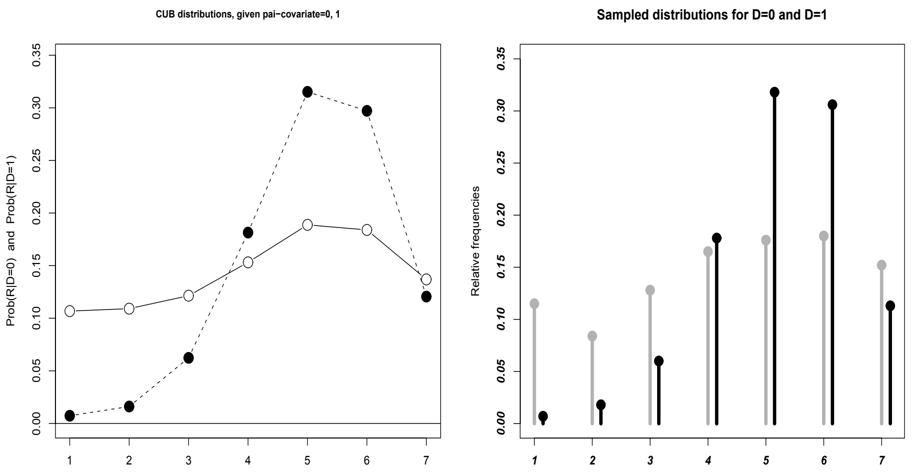

Figure 6. Probability and observed distributions for faked data

CUB model without covariates; specifically, the ML parameter estimates $\hat{\pi} = 0.615$ and $\hat{\xi} = 0.297$ are both highly significant. In addition, the noticeable performance of the model is confirmed by $Diss = 0.017$ and $X^2 = 5.441$, with a p-value$= 0.245$.

Moving to the measures discussed in sections 5-6, for this estimated model we get:

$$\ell(\boldsymbol{\theta}_0) = -3891.8; \quad \ell(\hat{\boldsymbol{\theta}}) = -3610.8; \quad \ell_{sat} = -3608.0;$$

and thus

$$\mathcal{G} = 0.214; \;\; ICON2 = 0.253; \;\; \mathcal{F}^2 = 0.983; \;\; \mathcal{L}^2 = 0.994; \;\; \mathcal{I} = 0.990\,.$$

These values confirm that $\mathcal{F}^2$, $\mathcal{L}^2$ and $\mathcal{I}$ express the exceptional fitting between data and postulated model (even with reference to the maximum achievable, given data); instead, the indexes $\mathcal{G}$ and $ICON2$ measure how

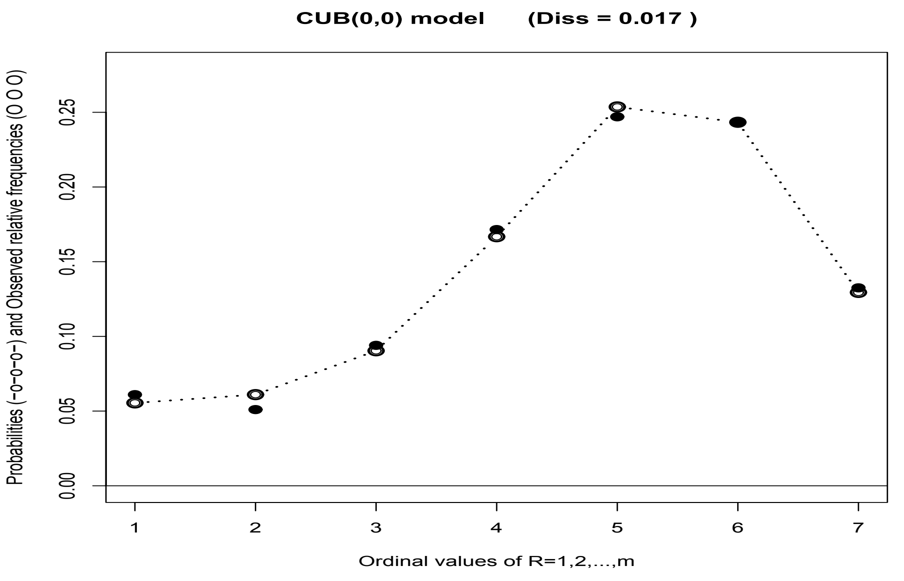

Figure 7. Observed and estimated CUB distributions of aggregated sample data

far is the distance from *uniformness* compared to a situation of degeneracy, and thus they are mostly related to the predictive ability of the estimated model given data. Notice the strong similarity between the values of $\mathcal{G}$ and $ICON2$.

Instead, when a dummy covariate for π is included, discrimination performs in a sharp manner: significant ML estimates of $\hat{\xi} = 0.298$; $\beta_0 = -1.078$; $\beta_1 = 4.103$, are obtained and thus:

$$\hat{\pi}_0 = \frac{1}{1+\exp(1.078-4.103\times 0)} = 0.254;$$
$$\hat{\pi}_1 = \frac{1}{1+\exp(1.078-4.103\times 1)} = 0.954\,.$$

These data confirm that the presence of dummy covariates for the uncertainty component is generally detected with high power.

Then, the measures of subsection 5.3 are shown in Table 3 for a sample consisting of two subgroups of 1000 observations.

Table 3. Sampling information given a dichotomous dummy.

Subgroups	*Relative frequencies*	*Saturated* $\ell(.)$
$(D_i = 0)$	$(0.115, 0.084, 0.128, 0.165, 0.176, 0.180, 0.152)'$	-1918.0
$(D_i = 1)$	$(0.007, 0.018, 0.060, 0.178, 0.318, 0.306, 0.113)'$	-1556.1
All	$(0.061, 0.051, 0.094, 0.171, 0.247, 0.243, 0.133)'$	-3608.0

Since $\ell_{sat^*} = \ell_{sat(0)} + \ell_{sat(1)} = -3474.1 > -3608.0 = \ell_{sat}$, there is evidence that the consideration of subgroups is worth of interest.

Specifically, if we consider that log-likelihood functions for models without, with covariates and saturated are:

$$\ell(\hat{\pi},\, \hat{\xi}) = -3610.8; \quad \ell(\hat{\xi},\, \hat{\boldsymbol{\beta}}) = \ell(\hat{\boldsymbol{\delta}}) = -3480.1; \quad \ell_{sat^*} = -3474.1;$$

respectively, we get:

$$\mathcal{I}(\boldsymbol{\delta}) = \frac{-3480.1 - (-3610.8)}{-3474.1 - (-3610.8)} = 0.956\,.$$

Table 4 summarizes the results obtained on this data set in terms of improvement with respect to *uniformness* (and these are related to the ability to predict) and data fitting with a dummy covariate (and these are related to the ability of a more elaborate model to reproduce patterns in observations).

Table 4. Improvements obtained by fitting nested CUB models.

Estimated Models	Log-likelihood functions
Uniformness	$\ell(0) = -3891.8$
$CUB(0,0)$	$\ell(\hat{\boldsymbol{\theta}}) = -3610.8$
Saturated (no-covariates)	$\ell_{sat} = -3608.0$
$CUB(1,0)$*-Dummy*	$\ell(\hat{\boldsymbol{\delta}}) = -3480.1$
Saturated (Dummy subgroups)	$\ell_{sat^*} = -3474.1$

As a final consideration, it should be added that although the percentage reduction of log-likelihood functions may be considered as moderate (between 3% and 11%), nevertheless these variations are significant. For instance, the introduction of a dummy covariate in the previous

$CUB(0,0)$ model (which causes a reduction in log-likelihood of only 3.62%) produces a likelihood ratio test of: $2(\ell(\hat{\boldsymbol{\delta}}) - \ell(\hat{\boldsymbol{\theta}})) = 261.4$, which is highly significant when compared with the percentile $\chi^2_{(0.05;g=1)} = 3.841$.

8. A real case study

In this section, we discuss a real data set concerning an ordinal response (ranged in a 7-point Likert-type scale) and registered on a sample of $n = 20184$ subjects. To be specific, we fitted CUB models to the perception of subjective survival probabilities to age 75; then, we checked the age of respondents as a relevant covariate for explaining the ordinal response (it is related to the cohort of the subject; further details in: Iannario and Piccolo, 2009b). It turns out that age of respondents is significant for explaining both uncertainty and perception parameters; in addition, statistical considerations suggest to transform it by logging, considering deviations from the average and squaring (for taking reversion effects into account).

With this data set, a model without covariates produces the following measures:

$$\mathcal{G} = 0.323; \;\; ICON2 = 0.632; \;\; \mathcal{F}^2 = 0.914; \;\; \mathcal{L}^2 = 0.927; \;\; \mathcal{I} = 0.954\,.$$

and, moreover, $\ell_0 = -39276$ and $\ell(\hat{\boldsymbol{\theta}}) = -30383$;

Table 5 shows the log-likelihood functions of (nested) fitted CUB models in order to assess the step-by-step improvements obtained by enlarging the set of explanatory covariates and introducing age (transformed). For reference, $\ell_{sat} = -29952$ for a model without covariates.

In this case study, we find that the range between *uniformness* and *saturation* is largely covered by a CUB model without covariates (this denotes a good fitting of the mixture) but the addition of the age covariate (for explaining both uncertainty and feeling) is helpful since this inclusion moves log-likelihood functions towards the upper bound.

Table 5. Improvements obtained by fitting more elaborated CUB models.

Models	*# parameters*	*Log-likelihood functions*	BIC
Uniformness	0	$\ell(0) = -39276$	78552
$CUB(0,0)$	2	$\ell(\hat{\boldsymbol{\theta}}) = -30383$	60786
$CUB(2,0)$	4	$\ell(\hat{\boldsymbol{\beta}}, \xi) = -30289$	60618
$CUB(0,2)$	4	$\ell(\pi, \hat{\boldsymbol{\gamma}}) = -30310$	60660
$CUB(2,2)$	6	$\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = -30221$	60501

9. Concluding remarks

In this paper we discussed the role of several measures of fitting and emphasized the need of characterizing them as indexes of predictive ability or effective fitting measures. Concepts related to *uniformness* of data and information content have deepened and some measures have been suggested when the objective is to fit a specific class of models.

Among the open issues, we quote the problem of developping some distribution theory for the indexes previously discussed; the topic is relevant for decision making and generally it may be convenient to relate them to likelihood ratio theory. In this regard, the introduction of generalized residuals may be of some interest, as in Di Iorio and Piccolo (2009).

More research seems necessary when covariates are present, mainly for exploiting the several contributions of the literature where the association among ordinal and nominal/ordinal covariates are measured: Agresti (1981, 1986), Goodman (1984), Agresti and Natarajan (2001), Piccarreta (2001), Rampichini *et al.* (2004), among the others.

A further problem under scrutiny is the study of measures related to the realization/prediction table and specifically aimed at predicting ability of data as generated by estimated models.

Acknowledgements: Author thanks referees for constructive comments to a preliminary version of the paper. The research has been partly supported by PRIN-2008: "Modelli per variabili latenti basati su dati ordinali: metodi statistici ed evidenze empiriche". Subjective survival probabilities information are subset of ISFOL survey data that has been used under the agreement ISFOL/PLUS 2006/430. The availability of CFEPSR structure of research, Portici, is gratefully acknowledged.

References

Aldrich J.H., Nelson F.D. (1984), *Linear probability, logit, and probit models*, Sage University Press, Beverly Hills.

Agresti A. (1981), Measures of nominal-ordinal association, *Journal of the American Statistical Association*, 76, 524–529.

Agresti A. (1986), Applying R^2-type measures to ordered categorical data, *Technometrics*, 28, 133–138.

Agresti A., Natarajan R. (2001), Modeling clustered ordered categorical data: a survey, *International Statistical Review*, 69, 354–371.

Bishop Y.M.M., Fienberg S.E., Holland P.W. (1977), *Discrete multivariate analysis*, The MIT Press, Cambdridge.

Cameron A.C., Windmeijer F.A.G. (1993), Deviance based R-squared measures of goodness of fit for generalized linear models, *Working Papers in Economics and Econometrics*, University of California at Davis.

Cameron A.C., Windmeijer F.A.G. (1996), R-squared measures for count data regression models with application to health-care utilization, *Journal of Business & Economic Statistics*, 14, 209–220.

Cameron A.C., Windmeijer F.A.G. (1997), An R-squared measure of goodness of fit for some common nonlinear regression models, *Journal of Econometrics*, 77, 329–342.

Corduas M., Iannario M., Piccolo D. (2009), A class of statistical models for evaluating services and performances, in: M.Bini, P.Monari, D.Piccolo, L.Salmaso (eds.): *Statistical methods for the evaluation of educational services and quality of products*, Contribution to Statistics, Springer, 99–117.

Cox D.R., Snell E.J. (1989), *The analysis of binary data*, 2^{nd} edition, Chapman and Hall, London.

Cragg J.G., Uhler R. (1970), The demand for automobiles, *Canadian Journal of Economics*, 3, 386–406.

Cramér H. (1946), *Mathematical methods of statistics*, Princeton University Press, Princeton, NJ.

Cressie N., Read T.R.C. (1989), Pearson's χ^2 and the loglikelihood ratio statistic G^2: a comparative review, *International Statistical Review*, 57, 19–43.

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

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

Dhrymes P.J. (1986), Limited dependent variables, in Griliches Z. and Intriligator M.D. (eds.), *Handbook of Econometrics*, vol.II, North-Holland, Amsterdam, 1576–1631.

Di Iorio F., Piccolo D. (2009), Generalized residuals in CUB models: definition and applications, *Quaderni di Statistica*, 11, this issue.

Efron B. (1978), Regression and ANOVA with Zero-One data: measures of residual variation, Journal of the American Statistical Association, 73, 113–121.

Franses P.H., Paap R. (2001), *Quantitative models in marketing research*, Cambridge University Press, Cambridge.

Frosini B. V. (1981), Heterogeneity indices and distances between distributions, *Metron*, 34, 95–108.

Frosini B. V. (2003), Descriptive measures of ecological diversity, in: *Environmetrics*, Serie E.P. n.119, Istituto di Statistica, Università Cattolica del S. Cuore, Milano.

Gini C. (1912), *Variabilità e mutabilità*, Studi economico-giuridici della Facoltà di Economia dell'Università di Cagliari, III, parte 2.

Goodman L.A. (1984), *Analysis of cross-classified data having ordered categories*, Harvard University Press, Cambridge, MA.

Greenwood P.E., Nikulin M.S. (1996), *A guide to Chi-squared testing*, J.Wiley & Sons, New York.

Grilli L., Rampichini C. (2002), Scomposizione della dispersione per variabili statistiche ordinali, *STATISTICA*, LXII, 111–116.

Haberman, S.J. (1982), Analysis od dispersion in multinomial responses, *Journal of the American Statistical Association*, 77, 568–580.

Hagle T.M., Mitchell G.E. (1992), Goodness-of-fit measures for probit and logit, *American Journal of Political Science*, 36, 762–784.

Hastie T. (1987), A closer look at deviance, *The American Statistician*, 41, 16–20.

Hauser J.R. (1978), Testing the accuracy, usefulness, and significance of probabilistic choice models: an information-theoretic approach, *Operation Research*, 26, 406–421.

Hübler O. (1977), Pseudo latent models: goodness of fit measures, residuals, estimation, testing, and simulation, *Statistical Papers*, 38, 271–285.

Iannario M. (2008), A class of models for ordinal variables with covariates effects, *Quaderni di Statistica*, 10, 53–72.

Iannario M. (2009a), A note on the identifiability of a mixture model for ordinal data, submitted.

Iannario M. (2009b), Modelling *shelter* choices in ordinal data surveys, submitted.

Iannario M., Piccolo D. (2009a), A program in R for CUB models inference, Version 2.0, available at *http://www.dipstat.unina.it*

Iannario M., Piccolo D. (2009b), Statistical modelling of subjective survival probabilities, submitted.

Judge G.G., Griffiths W.E., Hill R.E., Lütkepohl H., Lee T.C. (1985), *The theory and practice of Econometrics*, 2^{nd} edition, J.Wiley & Sons, New York.

Kent J.T. (1983), Information gain and a general measure of correlation, *Biometrika*, 70, 163–173.

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

Laitila T. (1993), A pseudo-R^2 measure for limited and qualitative dependent variable models, *Journal of Econometrics*, 56, 341–356.

Lave C.A. (1970), The demand for urban mass transportation, *Review of Economics and Statistics*, 52, 320-323.

Leti G. (1983), *Statistica descrittiva*, il Mulino, Bologna.

Maddala G.S. (1983), *Limited-dependent and qualitative variables in Econometrics*, Cambridge University Press, Cambridge.

Magee L. (1990), R^2 measures based on Wald and likelihood ratio joint significance tests, *The American Statistician*, 44, 250–253.

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*, vol.2, ch.18, North-Holland, Amsterdam, 1395–1457.

McKelvey R.D., Zavoina W. (1975), A statistical models for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology*, 4, 103–120.

Meinel N. (2009), Comparison of performance measures for multivariate discrete models, *Advances in Statistical Analysis*, 93, 159–174.

Nagelkerke N.J.D. (1991), A note on the general definition of the coefficient of determination, *Biometrika*, 78, 691–692.

Patil G.P., Taillie C. (1982), Diversity as a concept and its measurement (with discussion), *Journal of the American Statistical Association*, 77, 548–567.

Piccarreta R. (2001), A new measure of ordinal-ordinal association, *Journal of Applied Statistics*, 28, 107–120.

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

Piccolo D. (2008), Modelling University students' final grades by ordinal variables, *Quaderni di Statistica*, 10, 205–226.

Pudney S. (1989), *Modelling individual choice. The Econometrics of corners, kinks, and holes*, Basil Blackwell, Oxford.

Rampichini C., Grilli L., Petrucci A. (2004), Analysis of university course evaluations: form descriptive measures to multilevel models, *Statistical Methods & Applications*, 13, 357–373.

Read T.R.C., Cressie N.A.C. (1988), *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer-Verlag, New York.

Simonoff J.S. (2003), *Analyzing categorical data*, Springer, New York.

Spiess M., Keller F. (1999), A mixed approach and a distribution free multiple imputation technique for the estimation of multivariate probit models with missing value, *British Journal of Mathematical and Statistical Psychology*, 52, 1–17.

Spiess M., Tutz G. (2004), Alternative measures of the explanatory power of general multivariate regression models, *Journal of Mathematical Society*, 28, 125–146.

Vandaele W. (1981), Wald, likelihood ratio, and Lagrange multiplier tests as an F test, *Economic Letters*, 8, 361–365.

Veall M.R., Zimmermann K.F. (1990a), Pseudo R^2 in the ordinal probit model, Discussion Paper No.90–15, University of Munich.

Veall M.R., Zimmermann K.F. (1990b), *Goodness of fit measures in the Tobit model*, Discussion Paper No.90–23, University of Munich.

Veall M.R., Zimmermann K.F. (1992), Performance measures from prediction-realization tables, *Economic Letters*, 39, 129–134.

Veall M.R., Zimmermann K.F. (1996), Pseudo R^2 measures for some common limited dependent variables models, *Journal of Economic Survey*, 10, 241–259.

von Mises R. (1994), *Mathematical theory of probability and statistics*, Academic Press, New York and London.

Windmeijer F.A.G. (1995), Goodness-of-fit measures in binary response models, *Econometrics Reviews*, 14, 101–116.