

Multilevel factor models for ordinal variables

Leonardo Grilli

Carla Rampichini

Department of Statistics 'Giuseppe Parenti'

University of Florence

e-mail: grilli@ds.unifi.it, carla@ds.unifi.it

Phone: +39 0554237271, +39 0554237246

Final versione printed as:

.
Grilli L. & Rampichini C. (2007) Multilevel factor models for ordinal variables.
Structural Equation Modeling, 14 (1), pp 1-25

.
http://www.leaonline.com/doi/abs/10.1207/s15328007sem1401_1

Multilevel factor models for ordinal variables

The paper tackles several issues involved in specifying, fitting and interpreting the results of multilevel factor models for ordinal variables. First, the problem of model specification and identification is addressed, outlining the parameter interpretation. Special attention is devoted to the consequences on interpretation stemming from the usual choice of not decomposing the specificities into hierarchical components. Then a general strategy of analysis is outlined, highlighting the role of the exploratory steps. The theoretical topics are illustrated through an application to graduates' job satisfaction, where estimation is based on maximum likelihood via an EM algorithm with adaptive quadrature.

Keywords: *factor model, latent variable, multilevel model, ordinal variable*

The factor model is a popular and effective tool of analysis in the social sciences. In its classical formulation (e.g. Anderson, 2003), it concerns a set of continuous variables measured on a set of independent units. However such features may be inadequate in many cases, for example when the response variables are measured on ordinal scales (e.g. Likert scales) or the statistical units are nested in multilevel structures.

In principle, the specification of a factor model when the variables are ordinal (Jöreskog and Moustaki, 2001; Moustaki *et al.*, 2004) does not entail relevant theoretical problems, but estimation is quite complex, so many approximate solutions have been proposed (Bartholomew and Knott, 1999, §5.15): (a) transform the ordinal scale in a continuous scale by assigning a score to each grade, and use methods for continuous data; this procedure works well when there are enough categories, and the frequency distribution is unimodal with an internal mode (Muthén and Kaplan, 1985); (b) estimate the polychoric correlations among the ordinal variables and use these correlations as the input in algorithms for continuous data (Muthén, 1989); this approach, currently used by the LISREL program (Jöreskog, 2004), gives good point estimates, but it does not rely on a complete statistical model for the observed data, so it is not fully efficient; (c) collapse the grades of the scale in two categories, and use methods for binary data; this method may work well in some cases, but it implies arbitrariness in the dichotomization and a loss of information.

Anyway, recent developments in computational statistics have greatly enhanced the feasibility of a maximum likelihood or Bayesian analysis based on the proper model.

In many fields the statistical units are quite often nested in hierarchical, or multilevel, structures. Even if multilevel models are now well developed (Goldstein, 2003), the subclass of multilevel factor models has received relatively few attention, especially in applied work. For the case of continuous variables, two classical references are Goldstein and McDonald (1988) and Longford and Muthén (1992), while the case of binary variables has been treated only recently by Ansari and Jedidi (2000) and Goldstein and Browne (2004).

The present paper will focus on multilevel factor models for ordinal variables, a case where the problems associated with a proper treatment of ordinal variables adds to the difficulties of a multilevel analysis. From a theoretical standpoint such a model is a member of some broad frameworks, such as

the non-linear mixed model framework for IRT analysis of Rijmen *et al.* (2003), and the class of Generalized Linear Latent And Mixed Models (GLLAMM) of Rabe-Hesketh *et al.* (2004a) and Skrondal and Rabe-Hesketh (2004). However, the multilevel factor model for ordinal variables raises several identification and interpretation issues, as well as computational problems, so its implementation is not straightforward and, in fact, its use in applied work is rare.

The paper will discuss how to specify and fit multilevel factor models for ordinal variables, illustrating the theory by means of an application on the job satisfaction of the 1998's graduates of the University of Florence. The graduates responded to a series of items on job satisfaction using a five-point ordinal scale, so the model should treat the responses as ordinal variables. The main aim of the analysis is to describe and summarize the aspects of job satisfaction measured by the considered items, separately for the graduate and degree programme levels, in order to shed light on the effectiveness of the degree programmes. Therefore, the hierarchical nature of the phenomenon, with graduates nested in degree programmes, has a primary role and the use of a multilevel factor model is essential, as it allows to define separate factor structures at the two levels.

All the computations needed in the application are performed using standard software. In particular, the model is fitted with the program *Mplus* (Muthén and Muthén, 2004). The program performs maximum likelihood via an EM algorithm with adaptive Gaussian numerical quadrature.

The paper will be referred to the case where all the response variables are ordinal. The case of binary variables is simply a special instance, while the extension to mixed continuous-ordinal-binary responses is straightforward.

The structure of the paper is as follows. In the second Section the model is defined, writing down the likelihood and outlining the problem of identification; then the interpretation of model parameters is discussed, stressing the implications of not decomposing the specificities in the between and within components. The third Section presents a general strategy of analysis with several exploratory steps, which is made necessary by the computational effort usually needed to fit the model. In the fourth Section an application to the analysis of job satisfaction of the 1998's graduates of the University of Florence illustrates the steps of analysis and the interpretation of the results. The final Section concludes with some remarks.

The two-level factor model for ordinal variables

Description of the model

The analysis of ordinal data requires to choose a set of probabilities to be modelled and a suitable link function (Agresti, 2002). The set of probabilities may be of four main types: reference category, adjacent categories, continuation ratio, and cumulative probabilities. Here we choose the cumulative probabilities because this allows to represent each observed ordinal variable in terms of a continuous latent response endowed with a set of thresholds, a representation which helps the presentation of the model and the interpretation of the results. Note, however, that the latent response approach is only a convenient way to represent an ordinal variable: it does not require that the data have been generated by categorizing latent response variables.

The model can be formulated with any link function. In the following we implicitly adopt the probit link since the latent response is assumed to be Gaussian, a quite natural choice to create a connection with the classical factor model. However, the use of other links entails little modifications and usually leads to negligible differences in the results. In the application, due to software constraints, the fitted model is a logit one.

Let Y_{hij} be the h -th observed ordinal variable (item) ($h = 1, 2, \dots, H$) for the i -th subject ($i = 1, 2, \dots, n_j$) of the j -th cluster ($j = 1, 2, \dots, J$). Even if the items could be seen as lowest-level units, here the hierarchical levels are numbered starting from the subjects, so the subjects will be referred to as ‘level one’ or ‘within’ units, while the clusters will be referred to as ‘level two’ or ‘between’ units. Attention is limited to two hierarchical levels, since the extension to more than two levels is conceptually straightforward. The model allows for item non-response: in such a case some value of the index h is missing. In the application the clusters are the degree programmes, the subjects are the graduates and the ordinal variables are the ratings on 5 items of the questionnaire (i.e. $H = 5$).

A two-level factor model for ordinal variables can be set up by defining two components, namely: (1) a threshold model which relates a set of continuous latent variables \tilde{Y}_{hij} to the observed ordinal counterparts Y_{hij} ; and (2) a two-level factor model for the set of continuous latent variables \tilde{Y}_{hij} .

As for the threshold model, let assume that each of the observed responses Y_{hij} , which takes values

in $\{1, 2, \dots, C_h\}$, is generated by a latent continuous variable \tilde{Y}_{hij} through the following relationship:

$$\{Y_{hij} = c_h\} \Leftrightarrow \left\{ \gamma_{c_h-1,h} < \tilde{Y}_{hij} \leq \gamma_{c_h,h} \right\}, \quad (1)$$

where the thresholds satisfy $-\infty = \gamma_{0,h} \leq \gamma_{1,h} \leq \dots \leq \gamma_{C_h-1,h} \leq \gamma_{C_h,h} = +\infty$.

The factor model can now be defined on the set of latent variables. Ignoring for the moment the hierarchical structure, the standard factor model can be written as

$$\tilde{Y}_{hij} = \mu_h + \sum_{m=1}^M \lambda_{mh} u_{mij} + e_{hij}, \quad h = 1, \dots, H \quad (2)$$

where the u_{mij} 's are the factors, while, for each h , μ_h is the item mean, the λ_{mh} 's are the factor loadings and the e_{hij} 's are the uncorrelated item-specific errors.

A two-level extension of the factor model can be obtained in two different ways (see, for example, Muthén, 1994). The simplest way is to decompose the factors and the item-specific errors in two components, one for each hierarchical level:

$$\begin{aligned} u_{mij} &= u_{mj}^{(2)} + u_{mij}^{(1)}, & m = 1, \dots, M, \\ e_{hij} &= e_{hj}^{(2)} + e_{hij}^{(1)}, & h = 1, \dots, H. \end{aligned}$$

Here we use the superscript (l) to denote random variables defined at level l , along with their parameters and loadings. Therefore, model (2) becomes

$$\begin{aligned} \tilde{Y}_{hij} &= \mu_h + \sum_{m=1}^M \lambda_{mh} \left(u_{mj}^{(2)} + u_{mij}^{(1)} \right) + \left(e_{hj}^{(2)} + e_{hij}^{(1)} \right) \\ &= \mu_h + \left[\sum_{m=1}^M \lambda_{mh} u_{mj}^{(2)} + e_{hj}^{(2)} \right] + \left[\sum_{m=1}^M \lambda_{mh} u_{mij}^{(1)} + e_{hij}^{(1)} \right]. \end{aligned} \quad (3)$$

This formulation is useful if one assumes the existence of certain factors and wishes to study how they vary between and within the clusters. However, in applied work it is common to find completely different factor structures at the two hierarchical levels, so a more general formulation is (Goldstein and McDonald, 1988; Longford and Muthén, 1992):

$$\tilde{Y}_{hij} = \mu_h + \left[\sum_{m=1}^{M_2} \lambda_{mh}^{(2)} u_{mj}^{(2)} + e_{hj}^{(2)} \right] + \left[\sum_{m=1}^{M_1} \lambda_{mh}^{(1)} u_{mij}^{(1)} + e_{hij}^{(1)} \right]. \quad (4)$$

In this model the cluster level has M_2 factors with corresponding loadings $\lambda_{mh}^{(2)}$, while the subject level has M_1 factors with corresponding loadings $\lambda_{mh}^{(1)}$. Note that even if $M_2 = M_1$ the factor loadings in general are different, so the factors may have different interpretations. Obviously, model (3) is a special case of model (4) with $M_2 = M_1$ and $\lambda_{mh}^{(2)} = \lambda_{mh}^{(1)}$.

Now it is convenient to express the general two-level model (4) for the latent responses in matrix notation:

$$\tilde{\mathbf{Y}}_{ij} = \boldsymbol{\mu} + [\boldsymbol{\Lambda}^{(2)} \mathbf{u}_j^{(2)} + \mathbf{e}_j^{(2)}] + [\boldsymbol{\Lambda}^{(1)} \mathbf{u}_{ij}^{(1)} + \mathbf{e}_{ij}^{(1)}], \quad (5)$$

where $\tilde{\mathbf{Y}}_{ij} = (\tilde{Y}_{1ij}, \dots, \tilde{Y}_{Hij})'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_H)'$, $\mathbf{e}_j^{(2)} = (e_{1j}^{(2)}, \dots, e_{Hj}^{(2)})'$, $\mathbf{u}_j^{(2)} = (u_{1j}^{(2)}, \dots, u_{M_2j}^{(2)})'$, $\mathbf{e}_{ij}^{(1)} = (e_{1ij}^{(1)}, \dots, e_{Hij}^{(1)})'$, $\mathbf{u}_{ij}^{(1)} = (u_{1ij}^{(1)}, \dots, u_{M_1ij}^{(1)})'$, while $\boldsymbol{\Lambda}^{(2)}$ is a matrix whose h -th row is $(\lambda_{1h}^{(2)}, \dots, \lambda_{M_2h}^{(2)})$ and $\boldsymbol{\Lambda}^{(1)}$ is a matrix whose h -th row is $(\lambda_{1h}^{(1)}, \dots, \lambda_{M_1h}^{(1)})$.

The standard assumptions on the item-specific errors of model (5) are

$$\mathbf{e}_j^{(2)} \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}^{(2)}), \quad \text{where } \boldsymbol{\Psi}^{(2)} = \text{diag}\{(\psi_h^{(2)})^2\},$$

and

$$\mathbf{e}_{ij}^{(1)} \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}^{(1)}), \quad \text{where } \boldsymbol{\Psi}^{(1)} = \text{diag}\{(\psi_h^{(1)})^2\},$$

while for the factors it is assumed that

$$\mathbf{u}_j^{(2)} \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}^{(2)}),$$

and

$$\mathbf{u}_{ij}^{(1)} \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}^{(1)}),$$

where the covariance matrices $\boldsymbol{\Sigma}^{(2)}$ and $\boldsymbol{\Sigma}^{(1)}$ are, in principle, unconstrained. Moreover, all the errors and factors are assumed to be mutually independent, except for the factors at the same level, so model (5) is equivalent to the following variance decomposition

$$\text{Var}(\tilde{\mathbf{Y}}_{ij}) = [\boldsymbol{\Lambda}^{(2)} \boldsymbol{\Sigma}^{(2)} \boldsymbol{\Lambda}^{(2)'} + \boldsymbol{\Psi}^{(2)}] + [\boldsymbol{\Lambda}^{(1)} \boldsymbol{\Sigma}^{(1)} \boldsymbol{\Lambda}^{(1)'} + \boldsymbol{\Psi}^{(1)}]. \quad (6)$$

This amounts to a couple of factor models, one for the between covariance matrix and the other for the within covariance matrix (Muthén, 1994).

Finally note that the model can be extended by adding a regression component $\mathbf{x}'_{hij}\boldsymbol{\beta}$ in (4). Although it does not alter the essence of the paper, the introduction of a regression component, gives rise to a wide range of models, depending on the nature of the covariates. The covariates can be of various types (Rijmen *et al.*, 2003): item covariates, unit covariates and item by unit covariates, where the unit covariates can be further distinguished into subject-level and cluster-level covariates.

Model likelihood and identification

The full likelihood for the two-level factor model (4) can be derived in the following steps. Denoting with $\boldsymbol{\theta}$ the set of estimable parameters, the conditional likelihood for subject i of cluster j is

$$L_{ij}(\boldsymbol{\theta} | \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)}) = E_{\mathbf{u}^{(1)}} \left[\prod_{\mathbf{c} \in \mathcal{C}} \left(P \left(\bigcap_{h=1}^H \{Y_{hij} = c_h\} \mid \mathbf{u}_{ij}^{(1)}, \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)} \right) \right)^{d_{ij\mathbf{c}}} \right], \quad (7)$$

where \mathcal{C} is the set of all admissible values of the vector $\mathbf{c} = (c_1, \dots, c_H)'$ and $d_{ij\mathbf{c}}$ is the indicator of the observed response pattern $\left\{ \bigcap_{h=1}^H \{Y_{hij} = y_{hij}\} \right\}$.

The overall marginal likelihood is then

$$L(\boldsymbol{\theta}) = \prod_{j=1}^J E_{\mathbf{u}^{(2)}, \mathbf{e}^{(2)}} \left[\prod_{i=1}^{n_j} L_{ij}(\boldsymbol{\theta} | \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)}) \right]. \quad (8)$$

The probabilities that appear in (7), given the relationship (1) between the observed and latent responses and the assumptions on the latent model (4), can be written as

$$\begin{aligned} & P \left(\bigcap_{h=1}^H \{Y_{hij} = c_h\} \mid \mathbf{u}_{ij}^{(1)}, \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)} \right) \\ &= P \left(\bigcap_{h=1}^H \left\{ \gamma_{c_h-1,h} < \tilde{Y}_{hij} \leq \gamma_{c_h,h} \right\} \mid \mathbf{u}_{ij}^{(1)}, \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)} \right) \\ &= \prod_{h=1}^H P \left(\gamma_{c_h-1,h} - \mu_h - \zeta_{hij} < e^{(1)}_{hij} \leq \gamma_{c_h,h} - \mu_h - \zeta_{hij} \mid \mathbf{u}_{ij}^{(1)}, \mathbf{u}_j^{(2)}, \mathbf{e}_j^{(2)} \right) \\ &= \prod_{h=1}^H \left[\Phi \left(\frac{\gamma_{c_h,h} - \mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}} \right) - \Phi \left(\frac{\gamma_{c_h-1,h} - \mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}} \right) \right], \end{aligned} \quad (9)$$

where $\zeta_{hij} = \sum_{m=1}^{M_2} \lambda_{mh}^{(2)} u_{mj}^{(2)} + e_{hj}^{(2)} + \sum_{m=1}^{M_1} \lambda_{mh}^{(1)} u_{mij}^{(1)}$ and Φ is the standard Gaussian distribution function. Note that Φ is the inverse of the link function and stems from the normality assumption on the subject-level item-specific errors $e_{hij}^{(1)}$. Other distributional assumptions lead to different link functions, e.g. the logistic distribution leads to the logit link.

In the light of (7) and (9), the likelihood (8) is equivalent to the likelihood of a three-level model, where the items are the first-level units, the subjects the second-level units and the clusters the third level-units. This correspondence is useful for estimation purposes.

To discuss identification issues it is useful to note that from (9) the model likelihood is based on the quantities

$$\begin{aligned}
& \Phi \left(\frac{\gamma_{c_h, h} - \mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}} \right) \\
= & \Phi \left(\frac{\gamma_{c_h, h} - \mu_h}{\psi_h^{(1)}} - \sum_{m=1}^{M_2} \frac{\lambda_{mh}^{(2)} u_{mj}^{(2)}}{\psi_h^{(1)}} - \frac{e_{hj}^{(2)}}{\psi_h^{(1)}} - \sum_{m=1}^{M_1} \frac{\lambda_{mh}^{(1)} u_{mij}^{(1)}}{\psi_h^{(1)}} \right) \\
= & \Phi \left(\frac{\gamma_{c_h, h} - \mu_h}{\psi_h^{(1)}} - \sum_{m=1}^{M_2} \frac{\lambda_{mh}^{(2)} \sigma_m^{(2)}}{\psi_h^{(1)}} u_{mj}^{(2)*} - \frac{\psi_h^{(2)}}{\psi_h^{(1)}} e_{hj}^{(2)*} - \sum_{m=1}^{M_1} \frac{\lambda_{mh}^{(1)} \sigma_m^{(1)}}{\psi_h^{(1)}} u_{mij}^{(1)*} \right)
\end{aligned} \tag{10}$$

for $h = 1, \dots, H$ and $c = 1, \dots, C_h - 1$, where the asterisk denotes standardized variables. Assuming for simplicity that all the factors are uncorrelated, from (10) the estimable quantities are:

- $\frac{\gamma_{c_h, h} - \mu_h}{\psi_h^{(1)}}$ in number of $\sum_{h=1}^H (C_h - 1)$;
- $\frac{\lambda_{mh}^{(2)} \sigma_m^{(2)}}{\psi_h^{(1)}}$ in number of HM_2 ;
- $\frac{\psi_h^{(2)}}{\psi_h^{(1)}}$ in number of H ;
- $\frac{\lambda_{mh}^{(1)} \sigma_m^{(1)}}{\psi_h^{(1)}}$ in number of HM_1 .

Note that all the estimable quantities are expressed in terms of the item-specific subject-level standard deviations $\psi_h^{(1)}$, but this is not a problem for model interpretation, as will be clear from the following subsection.

The constraints needed for model identification are of two kinds: some are needed for the identification of the distribution of the latent responses determining the ordinal measures (1), and others are needed for the identification of the covariance matrix decomposition (6) associated with the factor model.

The identification of the distribution of the latent responses \tilde{Y}_{hij} yielding the observed responses Y_{hij} does not depend on the hierarchical nature of the model, so the considerations are the same as for single-level factor models for ordinal variables. First recall that in a univariate ordinal model the mean and the standard deviation of the latent response are not identifiable, so it is necessary to put two constraints, for example $\mu_h = 0$ and $\psi_h^{(1)} = 1$. In a multivariate ordinal model (Grilli

and Rampichini, 2003), a possibility is to impose similar constraints on each of the items and freely estimate all the thresholds of all the items. In such a case the threshold model uses all the available degrees of freedom, so the factorial part is not threatened by potentially invalid restrictions on the threshold part. A useful feature of unconstrained thresholds is that all the specificities are equal, so the factor loadings of two items are directly comparable.

However, the number of freely estimable thresholds, $\sum_{h=1}^H (C_h - 1)$, is usually quite large, so the parsimony principle suggests to look for a constrained structure entailing a negligible loss of fit. In general, structuring the thresholds in a sensible way is not straightforward, but when all the items have the same categories ($C_h = C \geq 3$) an appealing option (later called *equal latent thresholds*) is to assume that at every cutpoint c the latent thresholds $\gamma_{c,h}$ are equal, i.e. $\gamma_{c,h'} = \gamma_{c,h''} = \gamma_c$ for each h', h'' and left μ_h and $\psi_h^{(1)}$ free (except for a reference item): in this way the *actual* thresholds are $\tau_{c,h} = (\gamma_c - \mu_h)/\psi_h^{(1)}$, as it is clear from the likelihood contribution (9). Obviously, this kind of restriction makes the threshold structure more interpretable, though it requires some care in the interpretation of the loadings, as each item has its own scale $\psi_h^{(1)}$.

As for the identification of the two-level factor model on the continuous latent variables \tilde{Y}_{hij} 's, the variance-covariance decomposition (6) entails M_2^2 and M_1^2 indeterminacies in $\mathbf{\Lambda}^{(2)}\mathbf{\Sigma}^{(2)}\mathbf{\Lambda}^{(2)'} + \mathbf{\Psi}^{(2)}$ and $\mathbf{\Lambda}^{(1)}\mathbf{\Sigma}^{(1)}\mathbf{\Lambda}^{(1)'} + \mathbf{\Psi}^{(1)}$, respectively. In exploratory factor analysis it is customary to assume uncorrelated (orthogonal) factors with unit variance, putting the remaining constraints on the factor loadings in various forms (see e.g. Anderson, 2003). However, in confirmatory factor analysis it may be useful to relax either or both the assumptions on the factor covariance matrix, i.e. unit variance and uncorrelatedness. Relaxing the unit variance assumption causes a scale indeterminacy which can be solved by fixing to one a loading for each factor, while the uncorrelatedness is usually compensated by an adequate number of zeroes in the matrix of loadings. The main advantage of an unconstrained factor covariance matrix is that the loadings are invariant with regard to certain changes, as in the cases of factor-based unit selection and comparisons among populations (Anderson, 2003). However, it will be clear from the following that correlated factors complicate the interpretation of the results, while unconstrained variances are harmless in this regard.

The analytical approach to identification sketched above can be formalized, in the structural equa-

tion framework, through the definition of identification mappings between the structural parameters and identified reduced form parameters, as in Skrondal and Rabe-Hesketh (2004). Finally, local identification can be empirically checked in the estimation phase by inspection of the rank of the ML information matrix (e.g. computing the condition number), since non-singularity of the information matrix is a sufficient, though not necessary in the case of non-linear models, condition for local identification (Skrondal and Rabe-Hesketh, 2004).

Interpretation of model parameters

The formerly outlined two-level factor model for ordinal variables is based on two components which can be interpreted separately: (1) a threshold model which relates the continuous latent responses \tilde{Y}_{hij} to the observed ordinal counterparts Y_{hij} ; and (2) a two-level factor model for the continuous latent responses \tilde{Y}_{hij} . The following discussion will focus on some issues concerning the second component, which conveys the most important information.

Although the interpretation of the two-level factor model relies on the classical ideas of factor analysis, some clarification may be useful. Note that the following formulae are based on the uncorrelatedness of the factors at both levels, while the factor variances may be fixed or free (in any case it is assumed that the model is identified through adequate constraints on the loadings). It should be noted that the factor variances are not directly interpretable, even when left free, since they simply represent contributions with respect to the arbitrary item which has the loading fixed to one: in general the only interpretable quantity is the variance contribution expressed by the product $(\lambda_{mh}^{(1)}\sigma_m^{(1)})^2$ or $(\lambda_{mh}^{(2)}\sigma_m^{(2)})^2$.

From (6) the total variance of the h -th item is decomposed in

$$Var_T(\tilde{Y}_{hij}) = Var_B(\tilde{Y}_{hij}) + Var_W(\tilde{Y}_{hij}), \quad (11)$$

where

$$Var_B(\tilde{Y}_{hij}) = \sum_{m=1}^{M_2} (\lambda_{mh}^{(2)}\sigma_m^{(2)})^2 + (\psi_h^{(2)})^2$$

Cluster-level communality + Cluster-level specificity,

and

$$\begin{aligned} \text{Var}_W(\tilde{Y}_{hij}) &= \sum_{m=1}^{M_1} (\lambda_{mh}^{(1)} \sigma_m^{(1)})^2 && + (\psi_h^{(1)})^2 \\ &\text{Subject-level communality} && + \text{Subject-level specificity.} \end{aligned}$$

The ratio

$$ICC_h = \frac{\text{Var}_B(\tilde{Y}_{hij})}{\text{Var}_T(\tilde{Y}_{hij})} \quad (12)$$

is the so-called intraclass correlation coefficient, which represents the proportion of variance explained by the clusters.

However, in most applications, in order to save computational resources the cluster-level item-specific errors $e_{hj}^{(2)}$ in equation (4) are omitted, so the variance of the remaining subject-level item-specific errors $e_{hij}^{(1)}$ represents the *total* specificity: in such a case, the variance decomposition (11) is not feasible, so it is important to see the role of such decomposition for interpretation.

Let us first consider the case where the specificities are not disentangled. In general, the interpretation of the factor structures at the two levels does not depend on the decomposition of the specificities and the (*relative*) *communalities* can be computed as well. As in standard factor models, the *communality* is the proportion of the variance of a given response explained by the factors. As usual in the ordinal case, the communalities are referred to the *latent* responses. For example, the total communality of the h -th item is

$$\frac{\sum_{m=1}^{M_1} (\lambda_{mh}^{(1)} \sigma_m^{(1)})^2 + \sum_{m=1}^{M_2} (\lambda_{mh}^{(2)} \sigma_m^{(2)})^2}{\text{Var}_T(\tilde{Y}_{hij})}, \quad (13)$$

while the communality of the h -th item due to the m -th subject-level factor is

$$\frac{(\lambda_{mh}^{(1)} \sigma_m^{(1)})^2}{\text{Var}_T(\tilde{Y}_{hij})}. \quad (14)$$

Moreover, the decomposition of the specificities is not required for the correlation between two latent responses of the same subject, $\tilde{Y}_{h'ij}$ and $\tilde{Y}_{h''ij}$:

$$\frac{\sum_{m=1}^{M_1} \lambda_{mh'}^{(1)} \lambda_{mh''}^{(1)} (\sigma_m^{(1)})^2 + \sum_{m=1}^{M_2} \lambda_{mh'}^{(2)} \lambda_{mh''}^{(2)} (\sigma_m^{(2)})^2}{\sqrt{\text{Var}_T(\tilde{Y}_{h'ij}) \cdot \text{Var}_T(\tilde{Y}_{h''ij})}}.$$

However, there are other interesting quantities which can be computed only if the specificities are disentangled, such as the ICC_h (12) and the communalities at a given level. For example, the total

communality *at subject level* of the h -th item is

$$\frac{\sum_{m=1}^{M_1} \left(\lambda_{mh}^{(1)} \sigma_m^{(1)} \right)^2}{\text{Var}_W \left(\tilde{Y}_{hij} \right)},$$

while the communality *at subject level* of the h -th item due to the m -th subject-level factor is

$$\frac{\left(\lambda_{mh}^{(1)} \sigma_m^{(1)} \right)^2}{\text{Var}_W \left(\tilde{Y}_{hij} \right)}.$$

Moreover, for a given item, the correlation between two distinct subjects belonging to the same cluster is just the ICC_h (12), so it is computable only if the specificities are decomposed.

Finally it is to be stressed that, even if from (10) all the estimable quantities are expressed in terms of the subject-level item-specific standard deviations $\psi_h^{(1)}$, the interpretable quantities just described are unaffected by the item scale, since they are correlations or ratios of parameters within the same item.

Phases of the analysis

The accomplishment of careful exploratory analyses is extremely important in order to achieve a suitable model specification, helping to avoid some of the traps which inevitably characterize the development of a complex model. Moreover, fitting the two-level factor model for ordinal variables, outlined in the previous Section, is computationally intensive. In fact, the marginal likelihood (8) involves multiple integrals with respect to Gaussian densities which cannot be solved analytically. Several estimation methods have been proposed, such as maximum likelihood with adaptive Gaussian quadrature (Rabe-Hesketh *et al.*, 2005) and Bayesian MCMC algorithms (Ansari and Jedidi, 2000; Goldstein and Browne, 2002); other methods can be successfully applied, as discussed in the final Section. Anyway, the computational burden is necessarily heavy, so it is crucial to base model selection on suitable exploratory analyses, which allows to limit the number of fitted models and to supply the algorithms with good starting values.

For the analysis we suggest to adapt to the ordinal case Muthén's strategy for continuous items (Muthén, 1994).

1. *Univariate two-level models.* As a first step, it is advisable to fit a set of univariate ordinal variance component models, one for each item, with the following specification in terms of latent

responses:

$$\tilde{Y}_{hij} = \mu_h + e_{hj}^{(2)} + e_{hij}^{(1)}, \quad h = 1, \dots, H, \quad (15)$$

where $e_{hj}^{(2)}$ are cluster-level errors with standard deviation $\psi_h^{(2)}$ and $e_{hij}^{(1)}$ are subject-level errors with standard deviation $\psi_h^{(1)}$, implying $Var(\tilde{Y}_{hij}) = (\psi_h^{(2)})^2 + (\psi_h^{(1)})^2$. To overcome the usual latent response identification problem, let fix $\mu_h = 0$ and $\psi_h^{(1)} = s$ for each item h , where the constant value s depends on the link, e.g. $s = 1$ for the probit and $s = \pi/\sqrt{3}$ for the logit. The estimable parameters are then the thresholds and $(\psi_h^{(2)})^2$, with the related $ICC_h = (\psi_h^{(2)})^2 / [(\psi_h^{(2)})^2 + s^2]$. The point estimates and significance of the ICC_h allow to evaluate if a two-level analysis is worthwhile, while a comparison of the thresholds among the items should give some hints about possible restrictions to be imposed in the multivariate model.

2. *Exploratory non-hierarchical factor analysis.* In order to shade some light upon the covariance structure of the data, it is useful to estimate the matrix of product-moment correlations among the latent responses, i.e. the polychoric correlation matrix of the items, and to use this matrix to perform an exploratory non-hierarchical (i.e. single-level) factor analysis by means of standard software.
3. *Exploratory between and within factor analyses.* More specific suggestions for the two-level model specification can be obtained from separate exploratory factor analyses on the estimated between and within correlation matrices of the latent responses. The results of this two-stage procedure are expected to be similar to that obtained from the full two-level analysis, as in the continuous case (Longford and Muthén, 1992).

The decomposition of the latent response correlation matrix into the between and within components, can be obtained by means of a multivariate two-level ordinal model with unconstrained covariance structure. For each item the equation for the latent response is just (15), but now the items are jointly modelled with an unconstrained between covariance matrix $Var\{(e_{1j}^{(2)}, \dots, e_{Hj}^{(2)})'\}$ and an unconstrained within covariance matrix $Var\{(e_{1ij}^{(1)}, \dots, e_{Hij}^{(1)})'\}$. Despite the latent nature of the involved variables, the correlation matrices are identified. Note that the number of random effects in this multivariate model is equal to twice the number of considered items, so the

estimation process is computationally demanding. If the computation takes too long, it may be advisable to consider an approximate solution, assigning a score to the item categories and fitting a multivariate two-level model for continuous responses: the resulting correlation matrices will have slightly attenuated values, unless the distributions of the transformed variables are very far from the Gaussian distribution (Muthén and Kaplan, 1985).

4. *Confirmatory two-level factor analysis.* The results of the exploratory two-stage factor analysis outlined in step 3 are used to specify one or more confirmatory two-level ordinal factor models as defined by expression (4). These models can be fitted by means of likelihood or Bayesian methods and compared on the basis of appropriate indicators. The exploratory two-stage factor analysis of point 3 provides fine initial values for the chosen estimation procedure, which may allow a substantial gain in computational time. Note that a large amount of computational time can be saved by omitting the cluster-level item-specific errors $e_{hj}^{(2)}$, so that the variances of the subject errors $e_{hij}^{(1)}$ are in fact the total specificities. As illustrated in *Interpretation of model parameters*, this simplification prevents a full variance decomposition and the computation of the related quantities, but it is expected to be of minor importance as far as the interest of the researcher centers on the factor structure.

Application

The ordinal multilevel factor model has been used to analyze five items on job satisfaction taken from a telephone survey conducted, from one to two years after the degree, on the 1998's graduates of the University of Florence.

The question on job satisfaction was asked to the employed graduates. Altogether the considered data set includes 2432 graduates from 36 degree programmes, with a highly unbalanced structure: the minimum, median and maximum number of employed graduates per degree programme are 3, 31.5 and 495, respectively.

The question: *How much are you satisfied with the following aspects of your present job?* required a response on a five point scale: 1. very much satisfied, 2. very satisfied, 3. satisfied, 4. unsatisfied, 5. very unsatisfied. The five considered items are: 1. *Earning*, 2. *Career* (career's opportunities), 3.

Consistency (consistency with degree programme curriculum), 4. *Professionalism* (acquisition of professionalism), 5. *Interests* (correspondence with own cultural interests). The univariate distributions of the items are reported in Table 1. Note that the number of responses for each item is different, due to item non-response. The multilevel factor model adopted here allows for missing item values: maximum likelihood estimates are unbiased under the usual Missing At Random assumption (Little and Rubin, 2002).

The main aim of the analysis is to describe and summarize the aspects of satisfaction measured by the five considered items, separately for the graduate and degree programme levels. The two-level factor model for ordinal variables, defined by (1) and (4), is a useful tool to achieve this goal. The model is quite complex and, whichever algorithm is used, the fitting process is very time-consuming, so it is useful to follow the exploratory steps outlined in *Phases of the analysis*.

Univariate two-level models

The analysis begins by fitting the univariate ordinal variance component models (15), using the logit link for consistency with the confirmatory factor model. The maximum likelihood estimates, via adaptive quadrature, are reported in Table 2.

The between proportion of variance of the latent responses, ICC_h , is significantly different from zero for all items, as shown by the LRT comparing the models with and without random intercept. Note that when the LRT is testing on the boundary of the parameter space, like in this case, the limiting distribution of the LRT statistic is not the usual χ_1^2 , but instead a 50:50 mixture of a χ_0^2 (i.e. a point mass at zero) and a χ_1^2 . Therefore the p -values reported in Table 2 are halved (Snijders and Bosker, 1999).

Note that the ICC for the first three items is over 5%, which is valuable in a framework with categorical variables.

A comparison among the thresholds gives some ideas on possible constraints on the thresholds in the multivariate model. In particular, the differences between adjacent thresholds among the items should be compared to informally evaluate the plausibility of the *equal latent thresholds* structure outlined in *Model likelihood and identification*. In the present case, these differences are similar for all

the items, except for the third one, which has smaller differences. This suggests that the third item has a higher variability, as also confirmed by the variances calculated after item scoring (see Table 5). In the confirmatory factor analysis we will perform a test comparing the *equal latent thresholds* structure with the unconstrained one.

Exploratory non-hierarchical factor analysis

The second step requires the estimation of the matrix of product-moment correlations among the latent responses, i.e. the polychoric correlation matrix (see Table 3), whose entries are all significantly different from zero. This matrix is used to perform an exploratory maximum likelihood factor analysis via standard software. The results of this analysis (Table 4) suggest the presence of two factors: a *cultural* factor (labelled Factor 1), that explains primarily the *Consistency-Professionalism-Interests* correlations, and a *status* factor (labelled Factor 2), explaining mainly the *Earning-Career* correlation. Given the low proportions of between variance (ICC's of Table 2), this structure is expected to be quite similar to the within structure, though it may be very different from the between structure.

Exploratory between and within factor analyses

The third step of analysis calls for the decomposition of the overall correlation matrix of the latent responses into the between and within components. This task would require to fit a two-level multivariate ordinal model with five random effects for each level, which takes too long to be fitted with numerical integration. Therefore an approximate procedure is adopted, assigning a score to each item category. Various sophisticated scoring systems could be applied (Fielding, 1999), but given the preliminary nature of this step, the simplest scoring system is applied, assigning the rank value to each category. After scoring, the within and between covariance matrices can be estimated by fitting a multivariate two-level model for continuous responses. To this end the MLwiN software with RIGLS algorithm (Rasbash *et al.*, 2004) is used. RIGLS yields restricted ML estimates, which are less biased for variance-covariance parameters than full ML (Goldstein, 2003).

The results are shown in Tables 5 and 6. As for Table 5, note the following points: (i) it is clear from the last row of Table 5 that the third item (*Consistency*) has the higher variability, as yet

noted in the univariate analysis (Table 2); (ii) the between percentages of variance (i.e. the values on the diagonal) are in line with ICC's of Table 2; (iii) the between percentages tend to be higher for covariances than for variances.

As for Table 6 note the following points: (i) the total correlation matrix, which is obtained from the between and within components, is similar to the polychoric correlation matrix (Table 3), with a moderate attenuation; (ii) the structures of the between and within correlation matrices are quite different: for example, the between correlations are always higher than the within correlations; (iii) the within correlation matrix is similar to the total correlation matrix, due to the low proportion of between variances and covariances.

The results of the exploratory maximum likelihood factor analyses performed on the within and between correlation matrices of Table 6 are reported in Table 7.

As for the within structure, Bartlett's test indicates that two factors are sufficient (p -value=0.5082). The factor loadings are similar to those found in the non-hierarchical analysis (Table 4).

As for the between structure, while one factor is not enough, the estimation with two or more factors encounters an Heywood case. We decided to retain two factors, forcing the specificities to be non-negative. The second factor is measured by all items, while the first factor has relevant loadings only for the last three items.

Confirmatory two-level factor analysis with unconstrained thresholds

Finally, in the light of the results of the exploratory analysis, a two-level confirmatory factor analysis is performed using the model defined by equations (1) and (5). Maximum likelihood estimates are obtained with *Mplus* version 3 (Muthén and Muthén, 2004). *Mplus* performs ML estimation via an EM algorithm, solving the integrals with adaptive numerical integration. When the response variables are ordinal, ML estimation can be carried out with the logit link, while the probit link is not available. This implies a little change with respect to the model discussed in *Description of the model*, namely the distribution of the $e_{hi}^{(1)}$ is logistic instead of Gaussian. Note that the logit implies that $\psi_h^{(1)}$ is equal to $\pi/\sqrt{3}$, instead of 1 for the probit, while $\mu_h = 0$ as for the probit.

The within and between structures emerging from the exploratory analyses are not equally reliable:

the within part is estimated on a large number of observations and Bartlett's test clearly indicates the presence of two factors, while the between part is estimated on only 36 degree programmes and the estimation is complicated by the presence of an Heywood case.

Therefore, for the within part of the model the two-factor structure suggested by the exploratory within factor analysis (see Table 7) is retained, constraining to zero the loadings that were close to zero, that is the loading of *Earning* in the first factor and the loadings of *Consistency* and *Interests* in the second factor. As for the between structure, since the hints from the exploratory analysis are less clear, two configurations at this level have been tried: (i) a one-factor unconstrained structure (model *M1*); and (ii) a two-factor structure (model *M2*), with unconstrained loadings in the first factor and two loadings equal to zero in the second factor (*Earning* and *Career*, see Table 7).

Models *M1* and *M2* are fitted without imposing any restriction on the thresholds, in order to preserve the covariance structure from possible misspecifications of the thresholds. Since we are not particularly interested in decomposing the item specificities, in order to reduce the computational effort the cluster-level item-specific errors $e_{hj}^{(2)}$ are omitted, so the specificities are in fact total specificities. The models are fitted using five quadrature points for each factor, for a total of 625 quadrature points. Some limited trials suggest that larger numbers of quadrature points do not improve the estimates in a significant manner.

The likelihood ratio test comparing the models *M1* and *M2* clearly indicates that the second model is better. The preferred model *M2* has 35 estimable parameters: 20 thresholds $\gamma_{c,h}$, 5 factor loadings $\lambda_{mh}^{(1)}$ and 2 factor standard deviations $\sigma_m^{(1)}$ at the graduate level ($m = 1, 2$), 6 factor loadings $\lambda_{mh}^{(2)}$ and 2 factor standard deviations $\sigma_m^{(2)}$ at the degree programme level ($m = 1, 2$). The parameter estimates for model *M2* are reported in Table 8. The interesting part of the model is the covariance structure at both levels, which does not depend on the item means and thresholds and can be summarized by the communalities (see Table 9). These values are obtained as suitable transformations of model parameters: the factor-specific communalities are computed from formulae such as (14), the total communality is obtained by summing the row values FW1, FW2, FB1 and FB2 (see equation (13)), while the last column of the Table is the percentage of total communality due to the between level. The following points should be noted: (i) for the first three items the between component is greater

for the communality (last column of Table 9) than for the total variance (ICC's of Table 2); (ii) the last two items, *Professionalism* and *Interests*, are very poorly explained by the factors at degree programme level; (iii) the first factor at the degree programme level, FB1, is interpretable as a *status* factor, while the second one, FB2, is essentially related to the item *Consistency*.

The factor scores at degree programme level are represented in Figure 1, where the labels are attached to the degree programmes with extreme scores: the points on the right side of the diagram indicate a high satisfaction on *Earning* and *Career*, while the points on the top denote a high satisfaction on *Consistency*; note that there are two degree programmes lying in the left-down corner (Philosophy and Natural Sciences) with low satisfaction on both dimensions.

The analysis could be deepened by adding some covariates, but this is beyond the goal of the paper. Anyway, this extension is straightforward even if it may require a substantial increase in computational time.

Finally, a quick test to evaluate if the model selection is influenced by the omission of the cluster-level specific errors $e_{hj}^{(2)}$ consists in fitting the same models as before except for treating the responses as continuous, i.e. using the item scores. In such a case *Mplus* avoids numerical integration, so estimation takes a few seconds. Two models on item scores are fitted: $M1^*$ and $M2^*$, corresponding to $M1$ and $M2$, i.e. without cluster-level item-specific errors $e_{hj}^{(2)}$. The likelihood ratio test comparing these two models confirms that $M2^*$ is better than $M1^*$ (LR statistic= 51.7, df=3). Next, the same two models on item scores with the addition of cluster-level item-specific errors $e_{hj}^{(2)}$ are fitted: $M1^+$ and $M2^+$. The likelihood ratio test comparing these two models leads to a less clear result (LR statistic= 7.4, df= 3, p -value= 0.06). Therefore the choice of not decomposing the specificities may have unexpected consequences on the selection of the factor structure, suggesting to further investigate this point.

Confirmatory two-level factor analysis with constrained thresholds

At this point, it is interesting to compare model $M2$ with a model having the same factor structure, but constrained thresholds. Specifically, we consider the *equal latent thresholds* structure described in *Model likelihood and identification*. This structure cannot be easily imposed in *Mplus* because the subject-level item-specific standard deviations $\psi_h^{(1)}$ are assumed to be equal across items and cannot

be defined as model parameters. This point can be overcome in two ways.

First, it is possible to add a set of fictitious factors each pointing to one item, except for the reference item. This allows to define the item specific means μ_h and specificities $\psi_h^{(1)}$. However this solution is computationally inefficient, as it increases by $(H-1)$ the number of latent variables and thus the dimension of the integration in the marginal likelihood. Second, it is possible to impose non-linear constraints on the thresholds in order to obtain the desired threshold structure. As previously noted, under the assumption that the thresholds $\gamma_{c,h}$ are constant across items and hence written as γ_c , the *actual* thresholds are $\tau_{c,h} = (\gamma_c - \mu_h)/\psi_h^{(1)}$, for a total of $(C-1) + 2 \cdot (H-1)$ free parameters. In the *Mplus* parametrization μ_h and $\psi_h^{(1)}$ are assumed to be constant across items, so the threshold model for the ordinal variables is characterized by $(C-1) \cdot H$ estimable thresholds: therefore $(C-3) \cdot (H-1)$ constraints must be imposed in order to get the correct number of free parameters. To this end note that the relationship $\tau_{c,h} = (\gamma_c - \mu_h)/\psi_h^{(1)}$ implies the following equalities for any $c = 2, \dots, C-1$, and any pairs of items h and h^* :

$$\frac{\tau_{c,h} - \tau_{1,h}}{\tau_{c,h^*} - \tau_{1,h^*}} = \frac{\psi_{h^*}^{(1)}}{\psi_h^{(1)}}.$$

Hence in the present case ($C = 5$ and $H = 5$) the required $2 \cdot 4 = 8$ constraints on the *actual* thresholds could be:

$$\begin{aligned} \frac{\tau_{3,h} - \tau_{1,h}}{\tau_{3,h^*} - \tau_{1,h^*}} &= \frac{\tau_{2,h} - \tau_{1,h}}{\tau_{2,h^*} - \tau_{1,h^*}} \\ \frac{\tau_{4,h} - \tau_{1,h}}{\tau_{4,h^*} - \tau_{1,h^*}} &= \frac{\tau_{2,h} - \tau_{1,h}}{\tau_{2,h^*} - \tau_{1,h^*}} \end{aligned}$$

for $h^* = 2$ (the reference item) and $h = 1, 3, 4, 5$.

After estimation the item-specific standard deviations and means (with respect to the reference item) can be recovered using appropriate formulae, e.g.

$$\frac{\psi_h^{(1)}}{\psi_{h^*}^{(1)}} = \frac{\tau_{2,h^*} - \tau_{1,h^*}}{\tau_{2,h} - \tau_{1,h}}$$

and

$$\begin{aligned} \mu_h - \mu_{h^*} &= \psi_{h^*}^{(1)} \cdot \tau_{1,h^*} - \psi_h^{(1)} \cdot \tau_{1,h} \\ &= \psi_{h^*}^{(1)} \cdot \left[\tau_{1,h^*} - \frac{\psi_h^{(1)}}{\psi_{h^*}^{(1)}} \cdot \tau_{1,h} \right]. \end{aligned}$$

The program *Mplus* is used to fit a model with the same covariance structure of model *M2*, but with *equal latent thresholds*. This leads to a model with 35 parameters and 8 non-linear constraints on the thresholds. The likelihood ratio test for the *equal latent thresholds* assumption is $LRT = 17.19$ with 8 df giving a p -value=0.028. Therefore, with the data at hand the use of such structure is questionable, although the consequences on the communalities and factor scores are found to be modest.

An alternative software for fitting the models described in this paper is the `gllamm` command of Stata (Rabe-Hesketh *et al.*, 2004b), a highly flexible procedure which allows to fit the two-level factor model for ordinal variables both with unconstrained thresholds and with *equal latent thresholds*. In `gllamm` the *equal latent thresholds* structure is appealing, since it is implemented through a special link function, the *scaled ordered probit* (option `link(soprobit)`), whose scale parameters are the standard deviations of the specificities, $\psi_h^{(1)}$.

The `gllamm` command performs maximum likelihood, using a Newton-Raphson algorithm with adaptive Gaussian quadrature. In our application, the results obtained with `gllamm` are similar to those yielded by *Mplus*, though the computational times are substantially longer.

Concluding remarks

Multilevel factor models for ordinal variables are useful but complex tools, giving rise to problems of specification, identification, estimation and interpretation. At present the major obstacle to a wide use of such models is due to software limitations. To our knowledge the only widespread packages able to yield ML estimates for the models here discussed are *Mplus* and the `gllamm` command of Stata. The application presented in the paper relies on *Mplus*.

In the framework of full information maximum likelihood the maximization techniques can be classified along two dimensions (Rijmen *et al.*, 2003): the method of numerical integration of the intractable integrals used to approximate the marginal likelihood and the type of algorithm used to maximize the approximate marginal likelihood. Numerical integration can be deterministic, such as Gaussian quadrature (adaptive or not), or stochastic, such as Monte Carlo integration. The maximizing algorithm can perform the maximization directly on the marginal likelihood, such as the Newton-Raphson, or indirectly on some variant of the likelihood, such as the EM. In general the

computational time of Gaussian quadrature depends mainly on the number of factors, so models with three or more factors per level may take too much time for being of practical use. In such cases Monte Carlo integration may be more convenient. Anyway, there are promising attempts to improve the efficiency of numerical integration techniques (e.g. spherical quadrature: Rabe-Hesketh *et al.*, 2005; quasi-Monte Carlo: Pan, 2004).

As for the maximizing algorithm, further research is needed to assess the relative merits of Newton-Raphson, EM and their numerous variants, and to assess the interactions with the numerical integration techniques. A promising route is the adaptation of simulation-based methods (Gouriéroux and Monfort, 1996) to the class of multilevel factor models: an interesting example in this respect is the application of Mazzoli (2001) concerning a multilevel SEM with ordinal variables. In the search for approximate but computationally efficient methods, the development of limited information maximum likelihood (Muthén and Satorra, 1996) may be worthwhile.

In the Bayesian paradigm there is a growing research activity aimed at developing efficient MCMC algorithms for models with latent variables: in particular, Ansari and Jedidi (2000) and Goldstein and Browne (2004) treated multilevel factor models with binary responses, while Fox and Glas (2002) considered more general multilevel structural models. Although faster estimation algorithms can be developed, the supplementary computational effort needed to treat the response variables as ordinal, instead of continuous, is inevitably not negligible, so one can legitimately wonder whether the effort is adequately repaid in terms of the quality of statistical inference. A general answer is obviously not possible. The results of Muthén and Kaplan (1985) suggest that in factor models treating the ordinal variables as continuous is not severely harmful on condition that the frequency distributions are unimodal with an internal mode. Anyway, the use of a proper model is always a desirable feature of the analysis and the resulting inferences are generally more reliable.

As a final note, we stress that multilevel factor models should be handled with care. Hence, even if very efficient estimation algorithms were available, it is a good practice, especially in the case of ordinal response variables, to fit a multilevel factor model as the final step of the analysis, after having explored the data with simpler techniques.

References

- Agresti, A. (2002). *Categorical data analysis*. Second edition. New York: Wiley.
- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Third edition. New York: Wiley.
- Ansari, A. & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, 65, 475-496.
- Bartholomew, D.J. & Knott, M., (1999). *Latent variables models and factor analysis*. London: Arnold.
- Fielding, A. (1999). Why use arbitrary point scores?: ordered categories in models of educational progress. *JRSS A*, 162, 303-328.
- Fox, J. P. & Glas, C. A. W. (2002). Modeling measurement error in structural multilevel models. In Marcoulides and Moustaki (Eds.), *Latent Variable and Latent Structure Models*, 243-267. New Jersey: Lawrence Erlbaum.
- Goldstein, H., (2003). *Multilevel statistical models*. Third edition. London: Arnold.
- Goldstein, H. & Browne, W. J. (2002). Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. In Marcoulides and Moustaki (Eds.), *Latent Variable and Latent Structure Models*, 225-242. New Jersey: Lawrence Erlbaum.
- Goldstein, H., & Browne, W. J. (2004). Multilevel factor analysis models for continuous and discrete data. in A. Olivares (Eds.) *Advanced Psychometrics. A Festschrift to Roderick P. McDonald*. New Jersey: Lawrence Erlbaum.
- Goldstein, H. & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.
- Gouriéroux, C. & Monfort, A. (1996). *Simulation Based Econometric Methods*, New York: Oxford University Press.

- Grilli, L. & Rampichini, C. (2003). Alternative specifications of multivariate multilevel probit ordinal response models. *Journal of Educational and Behavioral Statistics*, 28, 31-44.
- Jöreskog, K. G., (2004). Structural Equation Modeling with Ordinal Variables using LISREL, manuscript available at <http://www.ssicentral.com/lisrel/ordinal.htm>
- Jöreskog, K. G. & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.
- Little, R. J. A. and Rubin D. B. (2002). *Statistical Analysis with Missing Data*. Second Edition. New York: Wiley
- Longford, N., & Muthén, B. (1992). Factor analysis for clustered observations. *Psychometrika*, 57, 581-597.
- Mazzolli, B. (2001). A multilevel structural equation model with polytomous and dichotomous data, *mimeo* presented at the Third International Conference on Multilevel Analysis, Amsterdam, April 9-10, 2001.
- Moustaki, I., Jöreskog, K. G. & Mavridis, D. (2004). Factor Models for Ordinal Variables With Covariate Effects on the Manifest and Latent Variables: A Comparison of LISREL and IRT Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 11, 487-513.
- Muthén, B.O. (1994). Multilevel Covariance Structure Analysis. in *Sociological Methods & Research*, 22, 376-398.
- Muthén, B. O. & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557-585.
- Muthén, L. K. & Muthén, B. O. (2004). *Mplus User's Guide (version 3)*. Los Angeles, CA: Muthén & Muthén. *Mplus* web site: <http://www.statmodel.com/index2.html>

- Muthén, B. O. & Satorra, A. (1996). Technical aspects of Muthén's LISCOMP approach to estimation of latent variables relations with a comprehensive measurement model. *Psychometrika*, 60, 489-503.
- Pan, J. (2004). Quasi-Monte Carlo estimation in Generalized linear mixed models. 19th International Workshop on Statistical Modelling, Florence, July 4-8, 2004.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004a). Generalized multilevel structural equation modelling. *Psychometrika*, 69(2), 167-190.
- Rabe-Hesketh S., Skrondal A. & Pickles, A. (2004b). GLLMM Manual, *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160*, <http://www.bepress.com/ucbbiostat/paper160>
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128, 301-323.
- Rasbash, J., Steele, F., Browne, W. & Prosser, B. (2004). *A user's guide to MLwiN version 2.0*. London: Institute of Education.
- Rijmen, F., Tuerlinckx, F., De Boeck, P. & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8, 185-205.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*, Boca Raton, FL: Chapman & Hall/ CRC Press.

Table 1: Univariate distributions of job satisfaction items. 1998's graduates from the University of Florence.

ITEM	LEVEL OF SATISFACTION					TOTAL	
	1	2	3	4	5	%	N
1. Earning	7.8	23.9	38.1	20.5	9.7	100.0	2421
2. Career	11.0	28.2	32.6	18.0	10.2	100.0	2393
3. Consistency	24.5	27.5	24.2	12.5	11.3	100.0	2427
4. Professionalism	26.0	40.3	22.8	7.7	3.2	100.0	2420
5. Interests	21.5	32.7	28.2	10.8	6.8	100.0	2419
TOTAL	18.2	30.5	29.2	13.9	8.2	100.0	12080

Table 2: Univariate ordinal logit variance component models: ICC_h , thresholds and p -value of LRT comparing models with and without random intercept. 1998's graduates from the University of Florence.

ITEM	ICC_h (%)	Thresholds				LRT	
		γ_1	γ_2	γ_3	γ_4	statistic	p -value
1. Earning	5.5	-2.56	-0.81	0.86	2.29	88.40	0.000
2. Career	8.7	-2.35	-0.63	0.84	2.14	145.68	0.000
3. Consistency	6.6	-1.14	0.11	1.25	2.19	98.80	0.000
4. Professionalism	1.8	-1.03	0.71	2.15	3.46	6.18	0.006
5. Interests	2.1	-1.24	0.24	1.63	2.71	17.79	0.000

Table 3: Polychoric correlation matrix of the items. 1998's graduates from the University of Florence.

ITEM	1	2	3	4	5
1. Earning	1.00				
2. Career	0.54	1.00			
3. Consistency	0.11	0.25	1.00		
4. Professionalism	0.28	0.45	0.54	1.00	
5. Interests	0.16	0.33	0.61	0.58	1.00

Table 4: Exploratory factor analysis on the polychoric correlation matrix of the items: varimax rotated factors and communalities. 1998's graduates from the University of Florence.

ITEM	Factor loadings		Communality
	Factor 1	Factor 2	
1. Earning	0.08	0.65	0.43
2. Career	0.26	0.80	0.70
3. Consistency	0.77	0.07	0.60
4. Professionalism	0.68	0.34	0.58
5. Interests	0.78	0.16	0.63

Table 5: Two-level multivariate model on item scores: between percentage of variance (*in italics*) and covariance, and total variance of items. 1998's graduates from the University of Florence.

ITEM	1	2	3	4	5
1. Earning	<i>5.90</i>				
2. Career	12.96	<i>8.63</i>			
3. Consistency	21.02	13.09	<i>7.37</i>		
4. Professionalism	10.55	7.63	6.62	<i>2.30</i>	
5. Interests	9.57	4.75	6.07	2.76	<i>2.36</i>
Total variance	1.15	1.31	1.68	1.04	1.31

Table 6: Two-level multivariate model on item scores: correlation matrix decomposition. 1998's graduates from the University of Florence.

ITEM	1	2	3	4	5
<i>Between</i>					
1. Earning	1.00				
2. Career	0.89	1.00			
3. Consistency	0.36	0.40	1.00		
4. Professionalism	0.72	0.69	0.79	1.00	
5. Interests	0.39	0.32	0.81	0.62	1.00
<i>Within</i>					
1. Earning	1.00				
2. Career	0.46	1.00			
3. Consistency	0.10	0.23	1.00		
4. Professionalism	0.23	0.39	0.48	1.00	
5. Interests	0.14	0.31	0.55	0.52	1.00
<i>Total</i>					
1. Earning	1.00				
2. Career	0.49	1.00			
3. Consistency	0.11	0.24	1.00		
4. Professionalism	0.25	0.40	0.49	1.00	
5. Interests	0.15	0.30	0.55	0.53	1.00

Table 7: Exploratory ML factor analysis on the within and between correlation matrices estimated from item scores: factor loadings and communalities (varimax rotation only for the within part). 1998's graduates from the University of Florence.

ITEM	Within			Between		
	Loadings		Comm.	Loadings		Comm.
	F1	F2		F1	F2	
1. Earning	0.07	0.59	0.35	0.00	1.00	1.00
2. Career	0.25	0.75	0.63	0.08	0.89	0.80
3. Consistency	0.72	0.07	0.53	0.93	0.36	1.00
4. Professionalism	0.64	0.32	0.50	0.57	0.72	0.84
5. Interests	0.74	0.16	0.58	0.71	0.39	0.66

Table 8: Confirmatory two-level factor analysis: model $M2$ parameter estimates. 1998's graduates from the University of Florence.

ITEM	Loadings				Thresholds			
	Within		Between					
	$\lambda_{1h}^{(1)}$	$\lambda_{2h}^{(1)}$	$\lambda_{1h}^{(2)}$	$\lambda_{2h}^{(2)}$	$\gamma_{1,h}$	$\gamma_{2,h}$	$\gamma_{3,h}$	$\gamma_{4,h}$
1. Earning	-	1.09	0.75	-	-3.96	-1.39	1.24	3.29
2. Career	1 ^(*)	1 ^(*)	1 ^(*)	-	-3.64	-0.92	1.46	3.37
3. Consistency	2.30	-	0.32	1 ^(*)	-1.80	0.19	1.98	3.39
4. Professionalism	2.25	0.41	0.31	0.27	-1.79	1.07	3.36	5.27
5. Interests	2.85	-	0.09	0.47	-2.34	0.34	2.83	4.62
Factor variance	0.75	3.09	0.71	0.45				

The symbol ^(*) denotes a fixed value.

Table 9: Confirmatory two-level factor analysis: communalities. 1998's graduates from the University of Florence.

ITEM	% Communality					% Between on Total
	Within		Between		Total	
	FW1	FW2	FB1	FB2		
1. Earning	-	49.9	5.4	-	55.2	9.7
2. Career	9.6	39.4	9.0	-	58.0	15.6
3. Consistency	51.1	-	0.9	5.7	57.7	11.5
4. Professionalism	49.2	6.7	0.9	0.4	57.2	2.3
5. Interests	64.2	-	0.1	1.0	65.3	1.6

Figure 1: Estimated factor scores for the degree programmes

