Chapter 2
Latent variable models for ordinal data

Silvia Cagnone, Stefania Mignani and Irini Moustaki

2.1 Introduction

Latent variable models with observed ordinal variables are particularly useful for analyzing survey data. Typical ordinal variables express attitudinal statements with response alternatives like “strongly disagree”, “disagree”, “strongly agree” or “very dissatisfied”, “dissatisfied”, “satisfied” and “very satisfied”.

In the literature, there are two main approaches for analyzing ordinal observed variables with latent variables. The most popular one is the *Underlying Variable Approach* (UVA) (Muthén, 1984; Jöreskog, 1990) which assumes that the observed variables are generated by underlying normally distributed continuous variables. This approach is used in structural equation modeling and the relevant methodological developments are available in commercial software such as LISREL (Jöreskog and Sörbom, 1988) and Mplus (Muthén and Muthén, 1998–2007). The other approach is the *Item Response Theory* (IRT) according to which the observed variables are treated as they are. The unit of analysis is the entire response pattern of a subject, so no loss of information occurs. An overview of those type of models can be found in Bartholomew and Knott (1999) and van der Linden and Hambleton (1997). Moustaki and Knott (2000) and Moustaki (2000) discuss a Generalized Linear Latent Variable Model framework (GLLVM) for fitting models with different types of observed variables.

Silvia Cagnone
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy
e-mail: silvia.cagnone@unibo.it

Stefania Mignani
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy
e-mail: stefania.mignani@unibo.it

Irini Moustaki
Department of Statistical Science, London School of Economics, Houghton Street, London, WC2A 2AE, UK, e-mail: i.moustaki@lse.ac.uk

M. Bini et al. (eds.), *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, Contributions to Statistics,
DOI 10.1007/978-3-7908-2385-1_2, © Springer-Verlag Berlin Heidelberg 2009

Several studies (Jöreskog and Moustaki, 2001; Huber et al., 2004; Cagnone et al., 2004) showed that the latter approach is preferable in terms of accuracy of estimates and model fit. This is due to the fact the UVA is based on limited information estimation methods whereas IRT is a full information approach. However, full information methods are much more computationally intensive especially as the number of latent variables increases. Solutions to computational problems for IRT models have been recently proposed by Huber et al. (2004) and Schilling and Bock (2005).

In the following sections we review the latent variable models for ordinal data within the GLLVM framework as introduced by Moustaki (2000). The chapter will focus on the goodness-of-fit issue when sparseness is present (Reiser, 1996; Maydeu-Olivares and Harry, 2005; Cagnone and Mignani, 2007) and the most recent extension to longitudinal data (Cagnone et al., 2009). An application to a subset of the National Longitudinal Survey of Freshmen (NLSF) is also presented.

2.2 The GLLVM for ordinal data

2.2.1 Model specification

Let $\mathbf{y}$ be a vector of K ordinal observed variables each of them with c_k categories and $\boldsymbol{\eta}$ a vector of Q latent variables. The c_k $(k = 1,\ldots,K)$ ordered categories of the variables y_k have associated probabilities $\pi_{1,k}(\boldsymbol{\eta}), \pi_{2,k}(\boldsymbol{\eta}),\ldots,\pi_{c,k}(\boldsymbol{\eta})$ which are functions of the vector of the latent variables $\boldsymbol{\eta}$. Within this framework, the unit of analysis is the response pattern of an individual; for the r-th individual it is defined as $\mathbf{y}_r = (y_1 = s_1, y_2 = s_2,\ldots,y_K = s_K)$. There are $\mathrm{NR} = \prod_{k=1}^{K} c_k$ possible response patterns.

The probability associated to $\mathbf{y}_r$ is given by

$$f(\mathbf{y}_r) = \pi_r = \int_{R_{\boldsymbol{\eta}}} g(\mathbf{y}_r|\boldsymbol{\eta})h(\boldsymbol{\eta})d\boldsymbol{\eta} = \int_{R_{\boldsymbol{\eta}}} \pi(\boldsymbol{\eta})h(\boldsymbol{\eta})d\boldsymbol{\eta} \tag{2.1}$$

where $h(\boldsymbol{\eta})$ is assumed to be a multivariate normal distribution with $\mathbf{0}$ mean and correlation (or covariance) matrix equal to $\boldsymbol{\Phi}$ and $g(\mathbf{y}_r|\boldsymbol{\eta})$ is the conditional probability of the observed variables given the latent variables following a multinomial distribution. Under the assumption of conditional independence:

$$g(\mathbf{y}_r \mid \boldsymbol{\eta}) = \prod_{k=1}^{K} g(y_k \mid \boldsymbol{\eta}) = \prod_{k=1}^{K} \pi_{s,k}{}^{y_{s,k}} = \prod_{k=1}^{K} (\gamma_{s,k} - \gamma_{s-1,k})^{y_{s,k}} \quad s = 2,\cdots,c_k \tag{2.2}$$

where $y_{s,k} = 1$ if a randomly selected individual responds into category s of the kth item and $y_{s,k} = 0$ otherwise. $\gamma_{s,k}$ is the cumulative probability of responding below category s. Unlike the model for binary data, in this case we define the conditional distribution $g(y_k \mid \boldsymbol{\eta})$ in terms of cumulative probabilities $\gamma_{s,k}$ since they take into account the ranking of the categories of the ordinal variables. In more detail $\gamma_{s,k} = \pi_{1,k} + \pi_{2,k} + \ldots + \pi_{s,k}$ is the probability of a response in category s or lower on the variable k. As in the classical generalized linear model, the relation between the observed and the latent variables can be expressed through any monotone differentiable link function. In the case of ordinal variables we can refer to the logit as follows:

$$\ln\left[\frac{\gamma_{s,k}}{(1-\gamma_{s,k})}\right] = \tau_{s,k} - \sum_{q=1}^{Q} \alpha_{kq}\eta_q, \quad s = 1,\ldots,c_k - 1 \tag{2.3}$$

where $\tau_{s,k}$ and α_{kq} can be interpreted as thresholds and factor loadings of the model, respectively. The ordinality is defined properly by the condition $\tau_{1,k} \leq \tau_{2,k} \leq \ldots \leq \tau_{c-1,k}$. We refer to (2.3) as the Proportional Odds Model (POM) (McCullagh and Nelder, 1983).

2.2.2 Model estimation

The parameters of the model are estimated with the E-M algorithm. The E-M has been used for estimating the two-parameter logistic model for binary variables in Bock and Aitkin (1981), and then used for estimating the GLLVM in Bartholomew and Knott (1999). See Moustaki (2000) for the case of ordinal data.
Starting from Eq. (2.1) the joint density of the random variable for the ith individual can be written as

$$f(\mathbf{y}_i, \boldsymbol{\eta}_i) = g(\mathbf{y}_i|\boldsymbol{\eta}_i)h(\boldsymbol{\eta}_i). \tag{2.4}$$

If we consider a sample of size n, the complete log-likelihood is given by:

$$\sum_{i=1}^{n} \log f(\mathbf{y}_i, \boldsymbol{\eta}_i) = \sum_{i=1}^{n} \log[g(\mathbf{y}_i|\boldsymbol{\eta}_i)h(\boldsymbol{\eta}_i)] = \sum_{i=1}^{n} [\log g(\mathbf{y}_i|\boldsymbol{\eta}_i) + \log h(\boldsymbol{\eta}_i)]. \tag{2.5}$$

From the assumption of conditional independence we get:

$$\sum_{i=1}^{n} \log f(\mathbf{y}_i, \boldsymbol{\eta}) = \sum_{i=1}^{n} \left[\sum_{k=1}^{K} \log g(y_{ki}|\boldsymbol{\eta}_i) + \log h(\boldsymbol{\eta}_i)\right]. \tag{2.6}$$

The thresholds and factor loadings are found in the first component of the log-likelihood whereas the parameters related with the covariance matrix of the latent variables are found in the second component.

Estimation of the correlation between latent variables. The E-M algorithm requires first the computation of the expected score function of the correlation terms with respect to the posterior distribution $h(\boldsymbol{\eta}|\mathbf{y})$

$$E_i S(\boldsymbol{\Phi}) = \int_{R_{\eta}} S(\boldsymbol{\Phi}) h(\boldsymbol{\eta}|\mathbf{y}_i) \mathrm{d}\boldsymbol{\eta} \tag{2.7}$$

where

$$S(\boldsymbol{\Phi}) = \frac{\partial \log h(\boldsymbol{\eta}, \boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}} \tag{2.8}$$

that is

$$S(\boldsymbol{\Phi}) = \partial \log\ h(\boldsymbol{\eta}, \boldsymbol{\Phi})/\partial \boldsymbol{\Phi} = -\frac{1}{2}\boldsymbol{\Phi}^{-1} + \frac{1}{2}\boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Phi}^{-1}. \tag{2.9}$$

By substituting (2.9) in Eq. (2.7) we get:

$$E_i S(\boldsymbol{\Phi}) = \int_{R_{\boldsymbol{\eta}}} \left(-\frac{1}{2}\boldsymbol{\Phi}^{-1} + \frac{1}{2}\boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Phi}^{-1}\right) h(\boldsymbol{\eta}|\mathbf{y}_i)\mathrm{d}\boldsymbol{\eta}. \tag{2.10}$$

The integrals can be approximated by using the Gauss-Hermite quadrature points. Since the latent variables are correlated, the approximation is obtained by using the Choleski factorization of the correlation matrix $\boldsymbol{\Phi} = \mathbf{C}\mathbf{C}'$. The Gauss-Hermite approximation will be applied to the integral of the transformed variables as follows

$$f(\mathbf{y}) = (2\pi)^{-n/2} \sum_{w_1,\dots,w_Q} g\left(\mathbf{z} \mid \mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)'\right) h\left(\mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)'\right) \tag{2.11}$$

where $\boldsymbol{\eta} = \mathbf{C}\boldsymbol{\beta}$, $\sum_{w_1,\dots,w_Q} = \sum_{w_1=1}^{v_1} \cdots \sum_{t_n=1}^{v_Q}$ and $v_1, \dots, v_Q$ are the quadrature points. By solving $\sum_{i=1}^{n} E_i S(\boldsymbol{\Phi}) = 0$ using the above approximation, we get explicit solutions for the maximum likelihood estimator of the elements of $\boldsymbol{\Phi}$

$$[\hat{\boldsymbol{\Phi}}]_{lj} = \frac{\sum_{i=1}^{n} \sum_{w_1,\dots,w_Q} \left[\left(\mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)'\right)\left(\mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)'\right)'\right]_{lj} h\left(\mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)|\mathbf{y}_i\right)}{\sum_{i=1}^{n} \sum_{w_1,\dots,w_Q} h\left(\mathbf{C}\left(\beta_{w_1}, \dots, \beta_{w_Q}\right)|\mathbf{y}_i\right)} \tag{2.12}$$

Estimation of the parameters in $g(\mathbf{y}|\boldsymbol{\eta})$. The expected score function of the parameters $\mathbf{a}_k = \left(\tau_{1,k}, \dots, \tau_{c_{k-1},k}, \alpha_{k1}, \dots, \alpha_{kQ}\right), k = 1, \dots, K$ with respect of $h(\boldsymbol{\eta}|\mathbf{y})$ is given by

$$E_i S(\mathbf{a}_k) = \int_{R_{\boldsymbol{\eta}}} S_i(\mathbf{a}_k) h(\boldsymbol{\eta}|\mathbf{y}_i)\mathrm{d}\boldsymbol{\eta}, \tag{2.13}$$

where in this case

$$S_i(\mathbf{a}_k) = \frac{\partial \log g(\mathbf{y}_i|\boldsymbol{\eta})}{\partial \mathbf{a}_k}. \tag{2.14}$$

By solving $E_i S(\mathbf{a}_k) = 0$ we get not-explicit solutions for the parameters $\mathbf{a}_k$. The expressions of the derivatives (2.14) can be found in Moustaki (2000) and Moustaki (2003).

The E-M algorithm works as follows:

- Choose initial estimates for the model parameters.
- E-step: Compute the Expected score functions given in (2.7) and (2.13).

- M-step: Obtain improved estimates for the parameters by solving the non-linear maximum likelihood equations for the parameters of the conditional distribution $g(\mathbf{y}|\boldsymbol{\eta})$ by using a Newton-Raphson iterative scheme and explicit solutions for the correlations between the latent variables.
- Return to step 2 and continue until convergence is achieved.

2.3 The goodness-of-fit of the model

2.3.1 The problem of sparseness

The usual way of testing the goodness-of-fit of latent variable models for ordinal data is to compare the observed and the expected frequencies of all possible response patterns (NR). A test for the model may be based on the usual goodness-of-fit statistics such as the likelihood ratio (LR) and the Pearson chi-square test (GF), defined as follows:

$$\mathrm{LR} = 2n \sum_{r=1}^{\mathrm{NR}} f_r \ln\left(\frac{f_r}{\widehat{\pi}_r}\right), \tag{2.15}$$

$$\mathrm{GF} = n \sum_{r=1}^{\mathrm{NR}} \frac{(f_r - \widehat{\pi}_r)^2}{\widehat{\pi}_r}, \tag{2.16}$$

where f_r is the sample proportion of the r-th response pattern, $\widehat{\pi}_r$ is the corresponding estimated probability $\widehat{\pi}_r = \pi_r(\hat{\mathbf{a}})$ and n is the sample size.

Under regular conditions both statistics are approximately distributed as a χ^2 with degrees of freedom $\mathrm{d}f = \mathrm{NR} - 1 - \#\mathrm{pr}$ where #pr is the number of the estimated parameters. With reference to the contingency table whose cells contain the frequencies of the response patterns, the number of observations in each cell should be large enough to justify the asymptotic approximation of the statistics to the chi-square distribution. Nevertheless, in many cases, contingency tables do not have large numbers of observations and the sparseness problem arise. To solve the sparseness problem a number of theoretical strategies has been proposed. Such strategies have been applied both to the goodness-of-fit statistics and to the residuals calculated from the marginal distributions of the observed variables. For a review of strategies applied to the former see Koheler and Larntz (1980), Agresti and Yang (1987), Read and Cressie (1988), Bartholomew and Tzamourani (1999), and Tollenar and Mooljaart (2003).

An alternative solution to the sparseness problem is to consider the residuals computed from marginal distributions. The residuals express the discrepancies between observed and expected frequencies and can be defined in a number of different ways. Residuals can provide information on how well the model predicts the one and two-way marginal distributions revealing items or pairs of items for which the model does not fit well. In fact, even in the presence of a severe degree of sparseness, almost always the univariate and the bivariate marginal frequencies distributions are quite large so that statistics based on these frequencies are not affected by sparseness. A thorough treatment of the analysis of residuals is given by Reiser (1996) with reference to the two-parameter item response model for binary data. The use of residuals in GLLVM for binary data is discussed in Bartholomew and Tzamourani (1999). They recommend to use them as supplementary analysis to the overall goodness-of-fit testing. In particular, they argue that a good model predicts well all the pairwise associations between observed variables. On the contrary, if some pairs of variables present high bivariate residuals, they indicate that the model does not fit the data. As for POM, Jöreskog and Moustaki (2001) have defined specific measures of fit based on the residuals. For the univariate marginal distributions they have proposed the following measure related to the GF (an equivalent measure is given also for the LR index but it is not reported here because it is outside the scope of this work):

$$\text{GF fit}^{(k)} = n \sum_{s=1}^{c_k} \frac{(f_{s,k} - \hat{\pi}_{s,k})^2}{\hat{\pi}_{s,k}} \quad k = 1, \ldots, K \tag{2.17}$$

where we can define:

$$\hat{\pi}_{s,k} = \sum_{r=1}^{\mathrm{NR}} y_{rs} \hat{\pi}_r, \tag{2.18}$$

and

$$y_{rs} = \begin{cases} 1 \text{ if } & y_k = s \\ 0 \text{ otherwise.} \end{cases} \tag{2.19}$$

The quantities $(f_{s,k} - \hat{\pi}_{s,k})^2 / \hat{\pi}_{s,k}$ $(s = 1, \ldots, c_k)$ are the standardized residuals computed from the univariate marginal distribution of the variable k.

In the same way, for the bivariate marginal distributions of the variables k and l we get:

$$\text{GF fit}^{(kl)} = n \sum_{sk,sl} \frac{(f_{sk,sl} - \hat{\pi}_{sk,sl})^2}{\hat{\pi}_{sk,sl}} \quad k = 1, \ldots, K-1 \quad l = k+1, \ldots, K \tag{2.20}$$

where, as before, we can define:

$$\hat{\pi}_{sk,sl} = \sum_{r=1}^{\mathrm{NR}} y_{rsk} y_{rsl} \hat{\pi}_r, \tag{2.21}$$

and

$$y_{rsk} = \begin{cases} 1 \text{ if } & y_k = s_k \\ 0 \text{ otherwise,} \end{cases} \tag{2.22}$$

$$y_{rsl} = \begin{cases} 1 \text{ if } & y_l = s_l \\ 0 \text{ otherwise.} \end{cases} \tag{2.23}$$

In this case the quantities $(f_{sk,sl} - \hat{\pi}_{sk,sl})^2 / \hat{\pi}_{sk,sl}$ $(s = 1, \ldots, c_k; s = 1, \ldots, c_l)$ are the standardized residuals computed from the bivariate marginal distribution of the variables k and l.

2.3.2 An overall goodness-of-fit test

The residuals based on the marginal distributions can be used for building a overall goodness-of-fit test. To this aim, we need to define the unstandardized residuals for the overall r-th response pattern as:

$$g_r = f_r - \hat{\pi}_r. \tag{2.24}$$

Under regular conditions (Birch, 1964), the NR dimensional vector $\sqrt{n}\mathbf{g}$ converges asymptotically to a gaussian random vector with mean equal to $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega}_\mathbf{g}$ defined as:

$$\boldsymbol{\Omega}_\mathbf{g} = \mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{T}(\mathbf{F}'\mathbf{F})^{-\mathbf{1}}\mathbf{T}', \tag{2.25}$$

where $\mathbf{D}(\boldsymbol{\pi})$ is a diagonal matrix that contains the NR probabilities π_r, the matrix $\mathbf{F}$ is defined as $\mathbf{F} = \mathbf{D}(\boldsymbol{\pi})^{-1/2}\partial\boldsymbol{\pi}/\partial\mathbf{a}$. Finally $\mathbf{T} = \partial\boldsymbol{\pi}/\partial\mathbf{a}$.
The residuals just defined are computed from the overall contingency table of the manifest variables. From these residuals it is possible to obtain the unstandardized residuals associated to the marginal distributions.We refer to the residuals for the bivariate marginal distributions (considering, for simplicity, only the case in which the observed variables have the same number of categories, that is $c_k = c_l = c$). For category a of variable k and category b of variable l they can be defined as:

$$e = (f_{sk,sl} - \hat{\pi}_{sk,sl}). \tag{2.26}$$

$\hat{\pi}_{sk,sl}$ is directly computed by the estimated response probabilities $\hat{\pi}_r$. Passing to the matrix form we can write:

$$\mathbf{e} = \mathbf{M}(\mathbf{f} - \hat{\boldsymbol{\pi}}) = \mathbf{M}\mathbf{g}, \tag{2.27}$$

where $\mathbf{M}$ is a matrix of 0s and 1s. The generic element of $\mathbf{M}$, $m_{sk,sl}$ is given by:

$$m_{sk,sl} = \begin{cases} 1 \text{ if} \quad y_k = s \quad \text{and} \quad y_l = s \\ 0 \text{ otherwise.} \end{cases} \tag{2.28}$$

The elements of $\mathbf{M}$ have been derived in such a way that multiplying $\mathbf{M}$ by the response probabilities $\boldsymbol{\pi}$, we realize the summation across the response patterns obtaining the second-order marginal proportions. From the asymptotic normality of $\mathbf{g}$ and from (2.27) we get:

$$\sqrt{n}\mathbf{e} \to N(\mathbf{0}, \boldsymbol{\Omega}_\mathbf{e}) \tag{2.29}$$

where $\boldsymbol{\Omega}_\mathbf{e} = \mathbf{M}\boldsymbol{\Omega}_\mathbf{g}\mathbf{M}'$.
A consistent estimator for $\boldsymbol{\Omega}_\mathbf{e}$ is given by:

$$\hat{\boldsymbol{\Sigma}}_\mathbf{e} = n^{-1}\mathbf{M}(\mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{T}(\mathbf{F}'\mathbf{F})^{-\mathbf{1}}\mathbf{T}')\mathbf{M}'|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}},\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}. \tag{2.30}$$

The test of fit is developed for assessing the null hypothesis that the theoretical residuals are not significantly different from 0. With this regard we can refer to the statistic:

$$X^2_{\mathbf{e}} = \mathbf{e}'\hat{\boldsymbol{\Sigma}}^{+}_{\mathbf{e}}\mathbf{e} \tag{2.31}$$

that has an asymptotic χ^2 distribution. Since $\hat{\boldsymbol{\Sigma}}_{\mathbf{e}}$ is not a full rank matrix, its inversion can be obtained in different ways. Cagnone and Mignani (2007) propose to use the Moore-Penrose generalized inverse; in the case in which the computational of $\hat{\boldsymbol{\Sigma}}^{+}_{\mathbf{e}}$ is not stable, Maydeu-Olivares and Harry (2005) propose to compute a matrix that has $\hat{\boldsymbol{\Sigma}}^{+}_{\mathbf{e}}$ as generalized inverse. The degrees of freedom of the χ^2 depend on the rank of $\boldsymbol{\Sigma}_{\mathbf{e}}$, that in general results less or equal to the $\min\left(\sum_{k=0}^{2}\binom{p}{k}(c-1)^k, \mathrm{NR}-1-(KQ+K(c-1))\right)$ namely, the minimum between the ranks of $\mathbf{M}$ and $\boldsymbol{\Omega}_{\mathbf{g}}$ (Bishop et al., 1975), respectively.
Reiser (1996) argued that, when sparseness is present, this index can be very useful for the goodness-of-fit of the overall model. In fact, although it is based on partial information, if higher-order interactions are not present (because of the conditional independence assumption) inferences regarding the parameters may be performed without loss of information in smaller marginal table (*collapsibility* of the contingency table). In this case this index produces good results in terms of both Type I error and power of the test (Reiser and Lin, 1999; Cagnone and Mignani, 2007). Nevertheless, when the collapsibility does not hold, this index is not as powerful as the indexes computed from the full contingency table.

2.4 GLLVM for longitudinal ordinal data

When questionnaires are submitted to the same individuals over time, we deal with longitudinal data or repeated measures. Recently many authors focused on latent variable models for longitudinal data with the aim of analyzing traits, attitudes, or any latent constructs over time (Roy and Lin, 2000; Dunson, 2003; Rabe-Hesketh et al., 1996). The latent variable model for ordinal data discussed in the previous sections has been extended to longitudinal data by Cagnone et al. (2009). The key feature of this model is that the inter-relationships among items are explained by time-dependent attitudinal latent variables whereas the associations across time are modelled via item-specific random effects. The time changes in the attitudinal latent variables are measured using a non-stationary autoregressive model. The resulted covariance matrix allows the latent variables to be correlated with unknown variances.

Formally, the model described in Sect. 2.2 is extended to longitudinal data in the following way. Given the vector of the K ordinal observed variables $\mathbf{y}_t$ measured at time t $(t = 1,\dots,T)$, the linear predictor defined in (2.3) becomes

$$\ln\left[\frac{\gamma_{t,k,s}}{(1-\gamma_{t,k,s})}\right] = \tau_{t,k,s} - \alpha_{kt}\eta_t - u_k, \qquad k = 1,\dots,K; s_k = 1,\dots,c_k - 1; t = 1,\dots,T \tag{2.32}$$

where the u_k's are item-specific random effects. The latent variables η_t and their variances allow to explain the associations among the items measured at time t. The associations among the same item measured across time are explained by u_k and the covariances between η_t's. The time dependent latent variables are related through a first order autoregressive structure

$$\eta_t = \phi \eta_{t-1} + \delta_t \tag{2.33}$$

where for identification purposes $\delta_t \sim N(0,1)$ and $\eta_1 \sim N\left(0, \sigma_1^2\right)$. It is also assumed that the random effects u_k are independent of η_t and their common distribution function is $N_K(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \text{diag}_{k=1,\dots,K}\left(\sigma_{uk}^2\right)$. It follows that $\text{Var}(\eta_t) = \phi^{2(t-1)}\sigma_1^2 + I(t \geq 2)\sum_{l=1}^{t-1}\phi^{2(l-1)}$ and $\text{Cov}(\eta_t, \eta_{t'}) = \phi^{t+t'-2}\sigma_1^2 + I(t \geq 2)\sum_{l=0}^{t-2}\phi^{t'-t+2l}$, where $I(.)$ is the indicator function.

As before, model estimation is obtained by using maximum likelihood estimation via the E-M algorithm. The substantial difference with the previous model in terms of estimation procedure is in the matrix $\boldsymbol{\Phi}$ whose elements express the relationships among both latent variables over time and latent variables and random effects. In more detail it is a covariance block matrix given by

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix} \tag{2.34}$$

where $\boldsymbol{\Gamma}$ is the variance covariance matrix of the time dependent latent variables. Its elements depend on the parameters ϕ and σ_1^2 in such a way that

$$\boldsymbol{\Gamma}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} + \phi^2 & -\phi & 0 & \dots & 0 & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \dots & 0 & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \dots & 0 & -\phi & 1 \end{bmatrix}$$

Explicit solutions for the parameters ϕ, σ_1^2 and σ_{uk}^2 $(k = 1, \dots, K)$ are obtained whereas, as before, a Newton Raphson algorithm is used for the thresholds and the factor loadings of the model (Cagnone et al., 2009) .

2.5 Case study: perceptions of prejudice on American campus

In order to illustrate the methodology described above we consider an example extracted from the National Longitudinal Survey of Freshmen (NLSF).[1] The NLSF

[1] This research is based on data from the National Longitudinal Survey of Freshmen, a project designed by Douglas S. Massey and Camille Z. Charles and funded by the Mellon Foundation and the Atlantic Philanthropies.

evaluates the academic and social progress of college students at regular intervals to capture emergent psychological processes, by measuring the degree of social integration and intellectual engagement and to control for pre-existing background differences with respect to social, economic, and demographic characteristics. Data are collected over a period of four waves (1999–2003). The sample was constituted by students of different races and 3,924 completed the survey.
In this analysis we concentrate on the part of questionnaire that investigates the perceptions of prejudice by the undergraduate students. It is composed by 13 ordinal items concerning different aspects of the perceptions of prejudice. After a preliminary exploratory factor analysis, we selected the following most important (in terms of reliability analysis) items:

1. How often, if ever, have students in your college classes ever made you feel uncomfortable or self-conscious because of your race or ethnicity? [StudUnc]
2. Walking around campus, how often, if ever, have you been made to feel uncomfortable or self-conscious because of your race or ethnicity? [CampUnc]
3. How often, if ever, have you felt you were given a bad grade by a professor because of your race or ethnicity [BadProf]
4. How often, if ever, have you felt you were discouraged by a professor from speaking out in class because of your race or ethnicity [DiscProf]

Permitted responses are“Never”,“Rarely”,“Sometimes”, “Often”, “Very often”, “Don’t know”, “Refused”. Since a small proportion of students responded to the last categories, categories from 3 to 5 have been collapsed leaving three categories for each item. Missing data have been treated by means of the listwise deletion. The final sample size is $n = 2,828$. The items are the same only for waves 2000 and 2001, hence in the analysis we consider two time points.
The aim of the analysis is first to fit at each time point a confirmatory factor model and then to perform a longitudinal analysis in order to evaluate if the perceptions of prejudice changes from 2000 to 2001. From a previous exploratory analysis we found that two factors can explain the variability between the items in both time points. Hence a POM model with correlated latent variables has been fitted to the data. In Table 2.1 the results of the estimates are reported.

Table 2.1 Parameter estimates with standard errors in brackets for the POM model, years 2000–2001, NLSF

	2000		2001	
Items	$\hat{\alpha}_{i1}$	$\hat{\alpha}_{12}$	$\hat{\alpha}_{i1}$	$\hat{\alpha}_{12}$
StudUnc	3.55 (0.32)	–	3.80 (0.23)	–
CampUnc	2.71 (0.18)	–	2.86 (0.14)	–
BadProf	–	2.88 (0.16)	–	4.36 (0.18)
DiscProf	–	3.79 (0.27)	–	2.73 (0.20)
ϕ_{12}	0.56(0.02)		0.62(0.01)	

We can observe that the loadings are high and significant for both factors and at both time points. Moreover they are very similar over time, indicating that the measurement invariance assumption is probably satisfied (same loadings over time, (Cagnone et al., 2009)). The correlations are quite high and significant in both observed years.

As for the goodness-of-fit, in year 2000 the LR and GF are equal to 188.52 and 190.58 respectively with $df = 51$ indicating that the two-factor model is rejected. The same result is obtained for year 2001, LR and GF being equal to 200.09 and 170.32 and $df = 51$. However, as discussed above, these tests can be affected by sparse data and therefore limited test statistics are computed instead, $X_{\mathbf{e}}^2$. For 2000, we obtained $X_{\mathbf{e}}^2 = 110.90$ with $df = 21$ and for 2001, we obtained $X_{\mathbf{e}}^2 = 97.51$ with $df = 21$. Both statistics indicate that the two factor model is rejected. If we want to investigate the reason of the poor fit we can look at the GFfits for each pair of items and follow the rule of thumb by which a cell greater than 4 or a total greater than 36 is an indication of poor fit (Jöreskog and Moustaki, 2001). In Table 2.2 the values of the GF fits are reported.

Table 2.2 Bivariate GF fit, years 2000–2001, NLSF

	2000				2001			
StudUnc	–							
CampUnc	53.74	–			56.28	–		
BadProf	17.36	8.76	–		15.71	25.06	–	
DiscProf	19.58	19.94	13.29	–	9.04	26.82	20.07	–

We can observe that the items responsible for bad fit are StudUnc and CampUnc since the value of GF fit is greater than 36.

These results suggest that a longitudinal analysis should be performed only for the second latent variable, that we can interpret as “Professor Prejudice”, since it is well measured by the items Badprof and Discprof. In particular, we want to evaluate if there is a significant change over time of this latent variable.

One fundamental assumption in latent variable models for longitudinal data is the measurement invariance of thresholds and loadings, that is the thresholds and the loadings have to be constrained to be invariant for the same item over time. We first fitted the model described in Sect. 1.4 without imposing any equality measurement constraints (Jöreskog, 2002) but the algorithm did not converge. Then we fitted the model with constrained loadings (ModA) and constrained thresholds and loadings (ModB) and in both cases the algorithm converged. However we found that the latter model has a lower BIC than the former (39495.96 for ModA versus 28183.64 for ModB). The results for ModB are reported in Table 2.3.

We fixed to 1 the loading associated to the same item in the two time points so that the latent variable is identified over time. However the loading estimate associated to DiscProf is very close to 1 and significant, indicating that the two items have the same influence on the latent variable. The variances of the random effects are

Table 2.3 Estimated thresholds and factor loadings with standard errors in brackets for the non-stationary model, NLSF

Items	$\hat{\tau}_{i(1)}$	$\hat{\tau}_{i(2)}$	$\hat{\alpha}_i$	$\hat{\sigma}_{ui}$
Badprof	4.98 (0.08)	7.00 (0.12)	1.00	0.60 (0.13)
Discprof	5.09 (0.08)	6.65 (0.13)	0.92 (0.19)	0.45 (0.06)

significant too, that implies that the random effects explain significantly the variability of the items over time.

The estimated covariance matrix of the latent variable over time is

$$\hat{\boldsymbol{\Gamma}} = \begin{bmatrix} 8.04 & 7.43 \\ 7.43 & 7.86 \end{bmatrix}$$

and the estimated $\hat{\phi} = 0.92(0.03)$ shows a very strong significant correlation between the latent variables in the two time points. Moreover the variability of the latent variable decreases over time.

The results highlight that the two items Badprof and Discprof measure the latent construct with almost the same magnitude. Moreover the perception of prejudice of the students towards the professors does not change substantially over the two observed years.

2.6 Concluding remarks

Latent variable models for ordinal data have been discussed with particular attention to two aspects recently developed, the goodness-of-fit problem and the analysis of longitudinal data. As for the former, a test based on bivariate marginal distributions has been presented. It allows to overcome the sparseness problem, typical of categorical data, that invalidates the classical goodness of fit statistics.

As for the latter, model for ordinal data have been extended to longitudinal data in such a way that different kinds of variability present in the data can be modelled. At this regard the associations among items are explained by means of time dependent latent variables. A non-stationary autoregressive structure allows to evaluate their changes over time. Random effect components capture the variability of the same item over time. The potentiality of this model and the validity of the goodness-of-fit test based on residuals in presence of sparse data have been showed by means of a full application to a subset of the NLSF.