Nikele, Fahrmeir:

A latent variable probit model for multivariate ordered categorical responses

Sonderforschungsbereich 386, Paper 164 (1999)

A latent variable probit model for multivariate ordered categorical responses

Marion Nikele

Max-Planck-Institut für Psychologische Forschung, München

Ludwig Fahrmeir

Institut für Statistik, Universität München

Abstract

This paper presents a fully Bayesian approach via Gibbs sampling for MIMIC models with ordered categorical outcomes. The method is of particular interest for moderate or medium sample size data situations as in the study to be presented. Compared to frequentist methods that are based on large sample theory, estimates and standard errors of parameters are more reliable. Experience from simulations and the application to the particular study on changes of styles of marital conflict resolution suggest that the approach provides a useful supplementary tool in combination with traditional methods.

We wish to thank Prof. F. E. Weinert, director of the Max-Planck-Institut for Psychological Research in Munich, for offering the scientific context for this study, the Max-Planck society for the financial support and last not least G. Nunner-Winkler for motivating this work. We also thank G. Arminger for helpful comments.

1 Introduction

The methodological innovation presented in this paper has been developed in the context of a study on changes in moral understanding conducted at the Max-Planck-Institute for Psychological Research in Munich (see, e.g. Nunner-Winkler, 1999). One aspect concerns early socialization experiences in the family system. A specific aim in this study was to investigate the impact of educational levels and age cohorts on the attitude of parents in conflict situations and their behaviour in solving them. These styles of conflict resolution of father and mother are considered as latent variables or constructs which cannot be directly measured. Instead, subjects are asked a variety of questions on indicator variables or response items which are supposed to characterize these latent constructs. Since the focus is a description of behaviour from an observer's perspective, not a positively biased self-presentation, the items were not presented to the parents themselves. Instead, a sample of 245 subjects, coming from three different age cohorts (young, medium and old generation), answered to questions about possible parental conflict resolution strategies. Answers are typically given in five ordered categories like "This behaviour is/was very typical for my mother (or father)" to "It is/was very untypical." We will make use of 15 most important items selected from a larger list of items developed by Döbert and Nunner-Winkler (1983). The substantive question - effect of observables on latent constructs - and the data situation - a large number of ordered categorical outcomes and a medium sample size - are not uncommon for many studies in psychological or social science studies. Generally, structural equation models with latent variables (see, e.g. Bollen, 1989) are adequate approaches for confirmatory data analysis. For the case study at hand, a specific MIMIC (multiple indicators, multiple causes) model will be applied, which relates a large number of ordered multicategorical responses to a small number of latent variables by a factor analytic model, and models the effect of covariates on the latent variables by a multivariate linear model.

LISCOMP (Muthèn, 1988) and MECOSA3 (Arminger, Wittenberg and Schepers, 1996) are well known programs for fitting such models. Parameter estimation is based on the likelihood principle, but due to the complicated structure of the full likelihood approximate or pseudo ML estimation, respectively weighted or unweighted least square methods, are used in the estimation steps, see, e.g., Browne and Arminger (1995). A possible drawback is that inference relies on large sample theory, so that interpretation of results for small or medium sample size in combination with many variables and parameters - as in our study - may become questionable.

For MIMIC models with mixed binary and continuous outcomes, Sammel, Ryan and Legler (1997) recently proposed an empirical Bayes approach involving the EM algorithm. However, the computational burden can become quite heavy due to numerical integrations necessary for E-steps.

Here, we propose a fully Bayesian approach using the Gibbs sampler to simulate from posterior distributions of parameters and latent variables. Bayesian inference via the Gibbs sampler or more general Markov chain Monte Carlo (MCMC) techniques seems to be particularly useful for latent variable models. They do not rely on large sample theory and are therefore well suited even for models with many parameters compared to sample size. Also, as a by-product of the estimation procedure, realizations of the latent variables itself can be estimated along with structural parameters, which may be of substantive interest in its own.

Scheines, Hoijtink and Boomsma (1999) use the Gibbs sampler for a Bayesian estimation approach of structural equation models given covariance data. For nonlinear latent variable models with Gaussian outcomes, a Bayesian approach relying on the Gibbs sampler and the Metropolis-Hastings algorithm has been recently developed by Arminger and Muthèn (1998). Bayesian inference for models with mixed binary and continuous responses is also suggested in an unpublished paper by Muthèn and Arminger (1995).

In this paper we focus on latent variable probit models for multivariate responses with ordered categories, suitable for analysis of the case study we are considering. Extensions to more general settings, like mixed multicategorical and continuous responses are mentioned, with details given in Nikele (1999). The approach is based on a threshold model linking multicategorical responses to a latent linear Gaussian factor analytic model, together with diffuse or informative Gaussian priors for parameters, see Section 2. Posterior estimation via the Gibbs sampler is described in Section 3, and Section 4 contains the application to the empirical study sketched at the beginning.

2 A latent variable probit model for multivariate ordered categorical responses

Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{ip})'$ be a vector of responses, observed for each individual i of a cross section of size n, together with a vector $\boldsymbol{z}_i = (\boldsymbol{x}_i', \boldsymbol{w}_i')'$ of covariates. We focus on the case of ordered categorical responses $Y_{ij} \in \{1, 2, \cdots, K_j\}, j = 1, \cdots, p$. Extensions to more general settings, such as mixed continuous and categorical outcomes, are outlined further below.

Responses are related to observed covariates and unobservable latent variables or parameters in several stages through a Bayesian hierarchical model.

The first stage of this hierarchy links ordered categorical responses y_{ij} to latent continuous variables y_{ij}^* via the threshold mechanism

$$y_{ij} = k \leftrightarrow \tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}, \tag{1}$$

$j = 1, \cdots, p,\ k = 1, \ldots, K_j$, with $\tau_{j0} = -\infty, \tau_{jK_j} = \infty$ and a vector $\boldsymbol{\tau}_j = (\tau_{j1}, \cdots, \tau_{j,K_j-1})'$ of thresholds. The second stage assumes a linear factor analytic model for the vector $\boldsymbol{y}_i^* = (y_{i1}^*, \cdots, y_{ip}^*)'$, conditionally upon a $m \times 1$ $(m < p)$

vector $\boldsymbol{\eta}_i$ of latent factors:

$$\boldsymbol{y}_i^* = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\theta}). \tag{2}$$

Here $\boldsymbol{\lambda}_0$ is a $p \times 1$ intercept vector, $\boldsymbol{\Lambda}$ a $p \times m$ matrix of factor loadings and $\boldsymbol{A}$ a matrix of fixed effects of observable covariates $\boldsymbol{w}_i$. The introduction of these fixed effects is a slight extension of usual factor analytic models, proposed by Sammel and Ryan (1996) for the case of observable $\boldsymbol{y}_i^*$'s. Latent factors $\boldsymbol{\eta}_i$ and errors $\boldsymbol{\epsilon}_i$ are independent. Conditional upon the latent factor $\boldsymbol{\eta}_i$, (1) and (2) define a multivariate probit model.

In the third stage, latent factors $\boldsymbol{\eta}_i$ are connected to observed covariates $\boldsymbol{x}_i$ through a linear regression model:

$$\boldsymbol{\eta}_i = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}\boldsymbol{x}_i + \boldsymbol{\xi}_i, \quad \boldsymbol{\xi}_i \sim N(\boldsymbol{0}, \boldsymbol{\psi}). \tag{3}$$

Without further restrictions on the unknown parameters, the model (1)-(3) is not identifiable. First, $\boldsymbol{\tau}_j$ and the corresponding components λ_{0j} and γ_{0j} in $\boldsymbol{\lambda}_0$ and $\boldsymbol{\gamma}_0$ are identifiable only up to additive constants. One option to circumvent this, is to set

$$\boldsymbol{\lambda}_0 = \boldsymbol{\gamma}_0 = \boldsymbol{0}. \tag{4}$$

For cross-sectional data as in our application it is natural to assume that $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\theta}$ is diagonal. Conditional upon $\boldsymbol{\eta}_i$, (1) and (2) define probit models for $y_{ij}, j = 1, \cdots, p$, but $\boldsymbol{\tau}_j$ and the components of row $\boldsymbol{\lambda}_j$ of $\boldsymbol{\Lambda}$ are identifiable only up to a constant factor in these conditional models. As common in standard probit models, we therefore make the assumption

$$\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\theta} = \boldsymbol{I}. \tag{5}$$

Following the classical model of factor analysis we also assume

$$\mathrm{Cov}(\boldsymbol{\xi}_i) = \boldsymbol{\psi} = \boldsymbol{I}, \tag{6}$$

implying that latent factors are independent and normalized.

For simplicity, we also omit fixed effects $\boldsymbol{A}\boldsymbol{w}_i$ in (2), and thus focus on the following *basic model*: Conditional upon the latent variables y_{ij}^*, responses y_{ij} are multinomially distributed:

$$\begin{aligned} y_{ij}|y_{ij}^* &\sim M(1, \boldsymbol{\pi}_{ij} = (\pi_{ijk}, k = 1, \cdots, K_j)), \\ \pi_{ijk} &= \Pr(\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}) \end{aligned} \tag{7}$$

Conditional upon the vector $\boldsymbol{\eta}_i$ of latent factors, $\boldsymbol{y}_i^* = (y_{i1}^*, \cdots, y_{ip}^*)'$ is multivariate normal:

$$\boldsymbol{y}_i^*|\boldsymbol{\eta}_i \sim N(\boldsymbol{\Lambda}\boldsymbol{\eta}_i, \boldsymbol{I}). \tag{8}$$

Given observed covariates $\boldsymbol{x}_i$, latent factors are independent and normalized:

$$\boldsymbol{\eta}_i|\boldsymbol{x}_i \sim N(\boldsymbol{\Gamma}\boldsymbol{x}_i, \boldsymbol{I}). \tag{9}$$

From (8) and (9) we obtain the marginal model

$$\boldsymbol{y}_i^*|\boldsymbol{x}_i \sim N(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{x}_i, \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{I}) \tag{10}$$

for $\boldsymbol{y}_i^*$, given the covariates and parameters. Thus, in the basic model (7)-(9), correlation between responses is induced by common latent factors alone.

For a fully Bayesian analysis, the models have to be supplemented with priors for unknown parameters in an additional stage of the hierarchy. For the basic model (7)-(9) we choose the following priors:

$$p(\boldsymbol{\tau}_j) \quad \text{diffuse}, \quad j = 1, \cdots, p \tag{11}$$

with thresholds obeying order restrictions, normal priors for factor loadings $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda})$,

$$\boldsymbol{\lambda} \sim N(\boldsymbol{l}, \boldsymbol{L}), \tag{12}$$

and normal or diffuse $\boldsymbol{G}^{-1} \to \boldsymbol{0}$ priors for $\boldsymbol{\gamma} = \text{vec}(\boldsymbol{\Gamma})$,

$$\boldsymbol{\gamma} \sim N(\boldsymbol{g}, \boldsymbol{G}). \tag{13}$$

Hyperparameters $\boldsymbol{l}, \boldsymbol{L}$ and $\boldsymbol{g}, \boldsymbol{G}$ can be obtained, for example, from preliminary analysis via MECOSA or LISCOMP. Whereas a diffuse prior for $\boldsymbol{\gamma}$ is generally uncritical, diffuse priors for $\boldsymbol{\lambda}$ can cause problems. This can be explained as follows: A diffuse prior for $\boldsymbol{\lambda}$ can correspond to a diffuse prior for the "random effect" $\boldsymbol{\Lambda\eta}_i$. As Hobert and Casella (1996) have shown for linear mixed models, this can lead to improper posteriors and a breakdown of the Gibbs sampler.

For the general model (1)-(3), diffuse or normal priors are a natural choice for parameters $\boldsymbol{\lambda}_0, \boldsymbol{\gamma}_0$ and $\boldsymbol{A}$, see Nikele (1999) for details. Priors for non-diagonal covariance matrices $\boldsymbol{\psi}$ or $\boldsymbol{\theta}$ will involve inverse Wishart distributions as, for example, in Arminger and Muthèn (1998), or correlation matrix priors as in Chib and Greenberg (1998). However, here we do not pursue this issue further. The whole approach can be extended to the case of mixed continuous and ordered categorical responses. Then a part of the components of $\boldsymbol{y}_i^*$ is directly observable, and the threshold mechanism (1) is simply dropped for these components. As a consequence, corresponding elements in the diagonal of $\boldsymbol{\theta}$ are not set to 1, but have to be estimated along with remaining parameters. This can be accomplished by imposing inverse Gamma priors on them, see Nikele (1999) for details. Other types of nonnormal responses could be considered as well, for example censored dependent variables that map latent variables to observed responses, see Browne und Arminger (1995).

Another extension would be the following: In the factor analytic model (2) and the regression model (3), the effect of covariates $\boldsymbol{w}_i$ and $\boldsymbol{x}_i$ are assumed to be linear. This may be appropriate for appropriately coded categorical covariates, but can be doubtful for metrical covariates with possibly nonlinear effects $\boldsymbol{A}(\boldsymbol{w}_i)$ and $\boldsymbol{\Gamma}(\boldsymbol{x}_i)$. Nonlinear parametric forms for $\boldsymbol{A}$ and $\boldsymbol{\Gamma}$ can be included, see Arminger and Muthèn (1998) for the case of continuous responses. Another, more flexible possibility, would be a nonparametric Bayesian approach for modeling covariate effects as in Fahrmeir and Lang (1999).

3 Posterior Analysis

Let $p(\boldsymbol{\vartheta})$ summarize our prior information about $\boldsymbol{\vartheta} = \text{vec}\{\boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}\}$ with $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_j)'$. Then the posterior density of $\boldsymbol{\vartheta}$ is given by

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}, \boldsymbol{x}) \propto p(\boldsymbol{\vartheta}) \cdot p(\boldsymbol{y}|\boldsymbol{\vartheta}, \boldsymbol{x}).$$

This form of the posterior density is not particularly useful for Bayesian estimation because the evaluation of the likelihood function is computationally intensive or even intractable. For example, in the case, where all response variables are binary the probability of a certain observed vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})'$ is given by

$$p(\boldsymbol{y}_i|\boldsymbol{\vartheta}, \boldsymbol{x}_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{\overline{c}_{i1}} \cdots \int_{\overline{c}_{ip}} \phi_p(\boldsymbol{y}_i^*|\boldsymbol{\eta}_i, \boldsymbol{\vartheta}) \phi_m(\boldsymbol{\eta}_i|\boldsymbol{\vartheta}, \boldsymbol{x}_i) dy_{ip}^* \ldots dy_{i1}^* d\eta_{im} \ldots d\eta_{i1}. \tag{14}$$

In (14) $\phi_p(\boldsymbol{y}_i^*|\boldsymbol{\eta}_i, \boldsymbol{\vartheta})$ is the $N(\boldsymbol{\Lambda}\boldsymbol{\eta}_i, \boldsymbol{I})$ pdf, $\phi_m(\boldsymbol{\eta}_i|\boldsymbol{\vartheta}, \boldsymbol{x}_i)$ is the $N(\boldsymbol{\Gamma}\boldsymbol{x}_i, \boldsymbol{I})$ pdf and $\overline{c}_{ij}$ denotes the integration domain for y_{ij}^* if the jth y variable takes on the value y_{ij}. For example, in the binary case the integration domain is either $(-\infty, \tau_{j1}]$ or $(\tau_{j1}, +\infty)$.

In contrast, our approach is based on the work by Albert (1992). The main idea is to focus on the joint posterior distribution of the parameter vector $\boldsymbol{\vartheta}$ and the unobservables $\boldsymbol{y}^* = (\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_n^*)'$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)'$. From Bayes' theorem the posterior density can then be found as follows:

$$p(\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}|\boldsymbol{y}, \boldsymbol{x}) \propto p(\boldsymbol{\vartheta}) \cdot p(\boldsymbol{y}^*, \boldsymbol{\eta}|\boldsymbol{\vartheta}, \boldsymbol{x}) \cdot p(\boldsymbol{y}|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}, \boldsymbol{x}) \tag{15}$$

where $p(\boldsymbol{y}^*, \boldsymbol{\eta}|\boldsymbol{\vartheta}, \boldsymbol{x})$ denotes the joint distribution of $\boldsymbol{y}^*$ and $\boldsymbol{\eta}$ given $\boldsymbol{\vartheta}$ and $\boldsymbol{x}$. This distribution is implied by the model; for an observation i it takes the following form:

$$\begin{bmatrix} \boldsymbol{y}_i^* \\ \boldsymbol{\eta}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{x}_i \\ \boldsymbol{\Gamma}\boldsymbol{x}_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{I}_{p\times p} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}' & \boldsymbol{I}_{m\times m} \end{bmatrix} \right). \tag{16}$$

Considering the threshold mechanism (1) the third term in (15) is

$$p(\boldsymbol{y}_i|\boldsymbol{y}_i^*, \boldsymbol{\eta}_i, \boldsymbol{\vartheta}, \boldsymbol{x}_i) = p(\boldsymbol{y}_i|\boldsymbol{y}_i^*, \boldsymbol{\tau}) = \prod_{j=1}^{p} \sum_{k=1}^{K_j} I(\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}) I(y_{ij} = k)$$

where $I(M)$ is the indicator function for an event M. Finally, since the sample is i.i.d., the posterior density function is

$$p(\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}|\boldsymbol{y}, \boldsymbol{x}) \propto$$
$$p(\boldsymbol{\vartheta}) \prod_{i=1}^{n} \left[p(\boldsymbol{y}_i^*, \boldsymbol{\eta}_i|\boldsymbol{\vartheta}, \boldsymbol{x}_i) \times \left(\prod_{j=1}^{p} \sum_{k=1}^{K_j} I(\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}) I(y_{ij} = k) \right) \right].$$

Obviously, it is not easy to draw samples from this distribution. Therefore we use the Gibbs sampler and draw samples from the following five conditional distributions that are described in detail in the subsequent sections:

1. $p(\boldsymbol{y}^*|\boldsymbol{\eta}, \boldsymbol{\vartheta}, \boldsymbol{y}, \boldsymbol{x})$
2. $p(\boldsymbol{\eta}|\boldsymbol{y}^*, \boldsymbol{\vartheta}, \boldsymbol{y}, \boldsymbol{x})$
3. $p(\boldsymbol{\Gamma}|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}\backslash\{\boldsymbol{\Gamma}\}, \boldsymbol{y}, \boldsymbol{x})$
4. $p(\boldsymbol{\Lambda}|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}\backslash\{\boldsymbol{\Lambda}\}, \boldsymbol{y}, \boldsymbol{x})$
5. $p(\boldsymbol{\tau}|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}\backslash\{\boldsymbol{\tau}\}, \boldsymbol{y}, \boldsymbol{x})$

3.1 The fully conditional distribution of $\boldsymbol{y}^*$

First of all we note that it is sufficient to consider the fully conditional distribution of $\boldsymbol{y}_i^*$ for each observation i because we assume an i.i.d. sample. Furthermore, since $V(\boldsymbol{\epsilon}_i) = \boldsymbol{I}$ is diagonal, the multivariate problem reduces to a univariate for each component y_{ij}^*. Given y_{ij}, $\boldsymbol{x}_i$, $\boldsymbol{\vartheta}$ and $\boldsymbol{\eta}_i$ the distribution of y_{ij}^* is restricted to an intervall defined by the observed value y_{ij}. This distribution can easily be derived from equations (1) and (8) that contain the entire information about y_{ij}^*. The fully conditional distribution of y_{ij}^* is therefore a truncated univariate normal distribution (for sampling from a truncated normal see Geweke, 1991):

$$y_{ij}^*|\boldsymbol{\eta}_i, \boldsymbol{\vartheta}\backslash\{\boldsymbol{\Gamma}\}, y_{ij} \sim N\left(\sum_{l=1}^{m}\lambda_{jl}\eta_{il}, 1\right)\sum_{k=1}^{K_j} I(\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}) I(y_{ij} = k).$$

3.2 The fully conditional distribution of $\boldsymbol{\eta}_i$

Due to independence of observations we can focus on the fully conditional distribution for $\boldsymbol{\eta}_i$. This distribution results from the joint distribution of $\boldsymbol{y}_i^*$ and $\boldsymbol{\eta}_i$ given $\boldsymbol{x}_i$ and $\boldsymbol{\vartheta}$ in (16). The conditional density $p(\boldsymbol{\eta}_i|\boldsymbol{y}_i^*, \boldsymbol{\vartheta}, \boldsymbol{y}_i, \boldsymbol{x}_i) = p(\boldsymbol{\eta}_i|\boldsymbol{y}_i^*, \boldsymbol{\vartheta}, \boldsymbol{x}_i)$ is the density of a multivariate normal distribution with mean

$$\mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{y}_i^*, \boldsymbol{\vartheta}, \boldsymbol{x}_i) = \boldsymbol{\Gamma}\boldsymbol{x}_i + \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{I})^{-1}(\boldsymbol{y}_i^* - \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{x}_i)$$

and covariance matrix

$$\mathrm{V}(\boldsymbol{\eta}_i|\boldsymbol{y}_i^*, \boldsymbol{\vartheta}, \boldsymbol{x}_i) = \boldsymbol{I}_{m\times m} - \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{I})^{-1}\boldsymbol{\Lambda}.$$

It is noteworthy that conditioning on $\boldsymbol{y}_i$ is vacuous because $\boldsymbol{y}_i$ cannot provide any further information; if $\boldsymbol{y}_i^*$ is known, $\boldsymbol{y}_i$ is also known by (1) .

3.3 The fully conditional distribution of regression parameters

Ascertaining the fully conditional distribution of the regression parameters $\boldsymbol{\gamma}$ we note that conditioning on all quantities except $\boldsymbol{x}$ and $\boldsymbol{\eta}$ is vacuous:

$$p(\gamma|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}\backslash\{\gamma\}, \boldsymbol{y}, \boldsymbol{x}) = p(\gamma|\boldsymbol{\eta}, \boldsymbol{x}).$$

Therefore it is sufficient to consider (9) which is a usual linear model. Since $\boldsymbol{\xi}_i$ is multivariate normal assuming an informative prior distribution $N(\boldsymbol{g}, \boldsymbol{G})$, γ is (e. g. Arminger und Muthén, 1998) multivariate normal with mean

$$\mathrm{E}(\gamma|\boldsymbol{\eta}, \boldsymbol{x}) = \left(\boldsymbol{G}^{-1} + \sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\left(\boldsymbol{G}^{-1}\boldsymbol{g} + \sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{\eta}_i\right)$$

and covariance matrix

$$\mathrm{V}(\boldsymbol{\gamma}|\boldsymbol{\eta}, \boldsymbol{x}) = \left(\boldsymbol{G}^{-1} + \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1}.$$

For each observation i the $(m \times mq)$ regression matrix $\boldsymbol{X}_i$ is defined as $\boldsymbol{X}_i = \boldsymbol{I}_{m\times m} \otimes \boldsymbol{x}_i'$.

3.4 The fully conditional distribution of the factor analytic model parameters

The fully conditional distribution of the parameters in the factor analytic model is derived the same way as has been shown for the regression parameters. It is sufficient to consider equation (8). Assuming an informative prior distribution $N(\boldsymbol{l}, \boldsymbol{L})$, the fully conditional distribution $p(\boldsymbol{\lambda}|\boldsymbol{y}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}\backslash\{\boldsymbol{\lambda}\}, \boldsymbol{y}, \boldsymbol{x}) = p(\boldsymbol{\lambda}|\boldsymbol{y}^*, \boldsymbol{\eta})$ has the density of a $p \cdot m$-dimensional normal distribution with mean

$$\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y}^*, \boldsymbol{\eta}) = \left(\boldsymbol{L}^{-1} + \sum_{i=1}^{n} \boldsymbol{U}_i' \boldsymbol{U}_i\right)^{-1} \left(\boldsymbol{L}^{-1}\boldsymbol{l} + \sum_{i=1}^{n} \boldsymbol{U}_i'^{-1} \boldsymbol{y}_i^*\right)$$

and covariance matrix

$$\mathrm{V}(\boldsymbol{\lambda}|\boldsymbol{y}^*, \boldsymbol{\eta}) = \left(\boldsymbol{L}^{-1} + \sum_{i=1}^{n} \boldsymbol{U}_i' \boldsymbol{U}_i\right)^{-1}.$$

The $(p \times pm)$ matrix $\boldsymbol{U}_i$ is defined as $\boldsymbol{U}_i = \boldsymbol{I}_{p\times p} \otimes \boldsymbol{\eta}_i'$.

3.5 The fully conditional distribution for the threshold parameters

The fully conditional distribution of τ_{jk} $(1 \le j \le p, 1 \le k \le K_j - 1)$ given $\boldsymbol{y}$, $\boldsymbol{y}^*$ and $\boldsymbol{\vartheta}\backslash\{\tau_{jk}\}$ is given up to a proportionality constant by

$$\prod_{i=1}^{n} [I(\tau_{j,k-1} < y_{ij}^* \le \tau_{jk}) I(y_{ij} = k) + I(\tau_{jk} < y_{ij}^* \le \tau_{j,k+1}) I(y_{ij} = k+1)].$$

Therefore the fully conditional distribution of τ_{jk} is a uniform distribution on the intervall

$$[\max\{\max\{y^*_{ij} : y_{ij} = k\}, \tau_{j,k-1}\}, \min\{\min\{y^*_{ij} : y_{ij} = k+1\}, \tau_{j,k+1}\}]$$

mit $\max(\emptyset) = -\infty$ und $\min(\emptyset) = \infty$. On the one hand, this intervall results from the restriction $\tau_{j,k-1} < \tau_{jk} < \tau_{j,k+1}$. On the other hand, for all y^*_{ij} with $y_{ij} = k$ τ_{jk} has to follow the condition $\tau_{jk} \leq y^*_{ij}$ and for all y^*_{ij} with $y_{ij} = k+1$ τ_{jk} has to follow the condition $\tau_{jk} < y^*_{ij}$ (e.g. Albert & Chib, 1993; Knorr-Held, 1995).

3.6 Performance of the Gibbs sampler

Implementing the Gibbs sampler is straightforward and can be done by many programming languages. In this connection, important questions concern the burn in period and the number of samples needed for a given target accuracy. Strategies for handling these questions are debated in the literature (see e.g. Cowles and Carlin, 1996, for an overview).

In simulation studies the performance of the Gibbs sampler in the MIMIC model was explored. With regard to the priors for the parameters diffuse priors for $\boldsymbol{\lambda}$ are found to be critical. Especially for high loadings the mixing of the time series can be very slow and bad. This is due to the fact that the marginal variance of an unobservable latent variable y^*_{ij} (given $\boldsymbol{\vartheta}$ and $\boldsymbol{x}_i$) depends on the loading via (10) and has no upper bound (for details see Nikele, 1999). A similar problem is observed in linear mixed models (Hobert & Casella, 1996): Diffuse priors for variance components lead to improper posteriors, implying possible nonconvergence of the Gibbs sampler. In our situation, $\boldsymbol{\lambda}$ determines variance components through the covariance matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{I}$ of the marginal Gaussian distribution of $\boldsymbol{y}^*_i|\boldsymbol{x}_i$ in (10).

This problem of slow and bad mixing can be avoided by using informative priors for high loadings. The prior does not have to be very informative. For the simple situation, when there are no exogenous variables and only one latent factor, simulation studies showed that it is sufficient if the prior information amounts to about 1% in relation to the data information. Some further discussion on this issue is also given in the application section 4.

In additional simulation studies we compared the Bayesian analysis with the frequentist approach that is widely used in computer programs such as LISCOMP or MECOSA. Estimation strategies involve several stages. At last, the estimator is found as the one which minimizes a quadratic form containing the differences of first and second order population and sample moments. There are two versions of the estimator we focus on. In the first case, the differences of population and sample moments are weighted by the estimated asymptotic covariance matrix of the moments. The corresponding estimator is the WLS (weighted least squares) estimator. This estimator is consistent and asymptotically normal, but its asymptotic behaviour is questionable for small and medium sample sizes (see, e.g. Muthén & Satorra, 1995). Therefore, an alternative estimator is available where the weight matrix is set to $\boldsymbol{I}$ (ULS = unweighted least squares). This estimator is consistent, but not asymptotic efficient. Therefore, in programs like LISCOMP no asymptotic standard errors are available, for example. In small simulation studies we compared the ULS and WLS estimation strategies with our Bayesian approach using diffuse or slightly informative priors. The following results were found:

1. For small sample sizes (in relation to the number of variables) the estimates are sometimes biased. However, as a rule, the bias is less large for the ULS method and the Bayesian approach than for the WLS method.
2. For small and medium sample sizes the estimated WLS standard errors are often too low whereas the Bayesian standard deviations are estimated

correctly. For the ULS method no standard deviations were available.

3. The Bayes estimator is superior to the WLS estimator as far as the efficiency is concerned. The Bayesian root mean squared error is always much smaller than the WLS one.

4. For large sample sizes the three point estimators approach each other.

Besides estimating simulated data sets we also analyzed the empirical data set of the study sketched at the beginning.

4 Application

To examine the question whether the parental conflict resolution styles differ across the three age cohorts and the three educational levels we formulated a MIMIC model. Explorative analyses of the factor structure on the basis of the polychoric correlations showed that three latent factors were sufficient to summarize the common information of the 15 items (see table 1).

Given this data structure, for our confirmatory analysis the following factor analytic model was derived that includes these three latent factors characterizing different parental styles in handling conflicts:

$$\boldsymbol{y}_i^* = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i. \tag{17}$$

For identification some parameters in $\boldsymbol{\Lambda}$ are fixed to 0. The 15×1 vector $\boldsymbol{y}_i^*$ contains the answers to the 15 items presented. These items contain possible reactions of parents in conflicts. Each item is to be rated on a scale with five possible categories ranging from "The behaviour is (was) very untypical for my mother/father." (category 1) to "The behaviour is (was) very typical for my mother/father." (category 5). For example, the following items were presented:

y_j	factor loadings $\hat{\lambda}_{jr}$ 1st factor ($r = 1$)	2nd factor ($r = 2$)	3rd factor ($r = 3$)
mother			
y_1: soothes father	.020	.038	.753
y_2: tries to talk it over calmly	-.070	.210	.587
y_3: gives in and comes around	-.013	-.024	.851
y_4: keeps harping on it	.425	.326	-.467
y_5: does not let father get a word in edgewise	.544	.390	-.553
y_6: tries to get the children to her side	.613	.105	-.222
y_7: makes sarcastic remarks	.648	.166	-.432
father			
y_8: soothes mother	-.031	.689	.056
y_9: tries to talk it over calmly	-.131	.720	.175
y_{10}: gives in and comes around	-.098	.807	-.094
y_{11}: keeps harping on it	.600	-.406	.208
y_{12}: does not let mother get a word in edgewise	.679	-.399	.070
y_{13}: takes out his anger on the children	.713	-.350	.170
y_{14}: makes sarcastic remarks	.646	-.192	.015
mother and father			
y_{15}: mother and father yell at each other	.762	.049	-.271

Table 1: Results of an explorative ULS analysis of the factor analytic model with LISCOMP using the polychoric correlations.

- My mother/father waits till both have calmed down and then tries to talk it over calmly.
- My mother/father gives in and comes around.
- My mother/father cannot yield and keeps harping on it.

Since the answers come from an ordinal scale for every item we formulated a threshold mechanism with four thresholds to estimate. The 3×1 vector $\boldsymbol{\eta}_i$ contains the three latent factors and the 15×1 vector $\boldsymbol{\epsilon}_i$ contains the uncorrelated and standardized error terms with $V(\boldsymbol{\epsilon}_i) = \boldsymbol{I}$.

The three latent factors were supposed to be influenced by two exogenous variables: age cohort and educational level. Each exogenous variable can take on one of three values and is coded as a dummy variable:

$$x_{ia}^{(A)} = \begin{cases} 1, & \text{if person } i \text{ belongs to age cohort } a \ (a = 1 \text{ (young)}, 2 \text{ (middle)}) \\ 0 & \text{else} \end{cases}$$

$$x_{ib}^{(B)} = \begin{cases} 1, & \text{if person } i \text{ has educational level } b \ (b = 1 \text{ (low)}, 2 \text{ (middle)}) \\ 0 & \text{else.} \end{cases}$$

Therefore a person who belongs to the oldest age cohort and has a high educational level belongs to the reference category.

Combining these considerations the structural model has the following form:

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \end{pmatrix} \begin{pmatrix} x_{i1}^{(A)} \\ x_{i2}^{(A)} \\ x_{i1}^{(B)} \\ x_{i2}^{(B)} \end{pmatrix} + \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \xi_{i3} \end{pmatrix}. \tag{18}$$

The MIMIC model constituted by equations (17) and (18) together with the threshold mechanism is analyzed from a Bayesian point of view. Besides from estimating the model we wanted to compare the estimation results using different priors. First of all we used diffuse priors containing no prior information about the parameters. We ran the Gibbs sampler for 15000 iterations and then for each single marginal component we considered the resulting time series of the samples. For all parameters the time series showed very high autocorrelations for a lag of 1 ranging from 0.43 to 0.95. Especially those elements in $\mathbf{\Lambda}$ with high estimates in the explorative analysis show strong interdependencies in the samples. Parameters λ_{33} and $\lambda_{10,2}$ with the highest absolute loadings of more than 0.80 in the explorative analysis show even autocorrelations higher than 0.40 for a lag of 100.

To explore the Gibbs sampler's behaviour in more detail we performed a very long Gibbs sampler run with 506000 iterations with only every 100th iteration results stored. For nearly all parameters the autocorrelations disappeared already

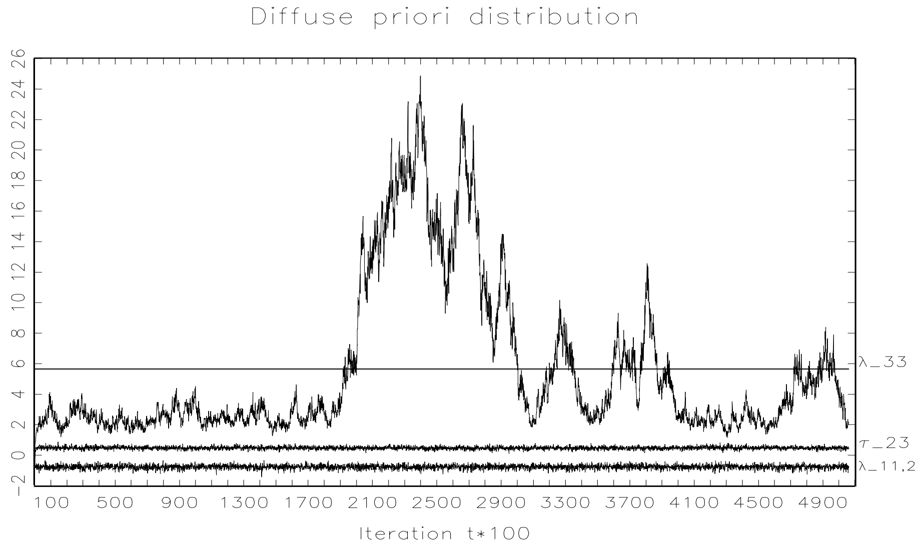

Figure 1: Time series for λ_{33}, $\lambda_{11,2}$ and τ_{23} of a Gibbs sampler run with 506000 iterations using diffuse priori distributions where only every 100th iteration results were stored.

for a lag of 1 or 2. For some parameters, however, considerable correlations were found. Especially for the high loading λ_{33} the autocorrelations are very high and the mixing of the chain is very poor, indeed (see figure 1). The parameter takes on values that are higher than 20.

As already discussed in section 3.6, this serious problem can be avoided by using a slightly informative normal prior $N(\boldsymbol{l}, \boldsymbol{L})$. For the critical loading λ_{33} the corresponding mean in $\boldsymbol{l}$ was fixed to the ULS solution and the variance in $\boldsymbol{L}$ was set to the value 1/2.45. This means that for λ_{33} the prior information is about 1% relative to the data information (Nikele, 1999). For all other parameters that caused no problems with a diffuse prior the corresponding mean was set to 0 and the variance to 10000. Additionally, the parameters were assumed to be a priori

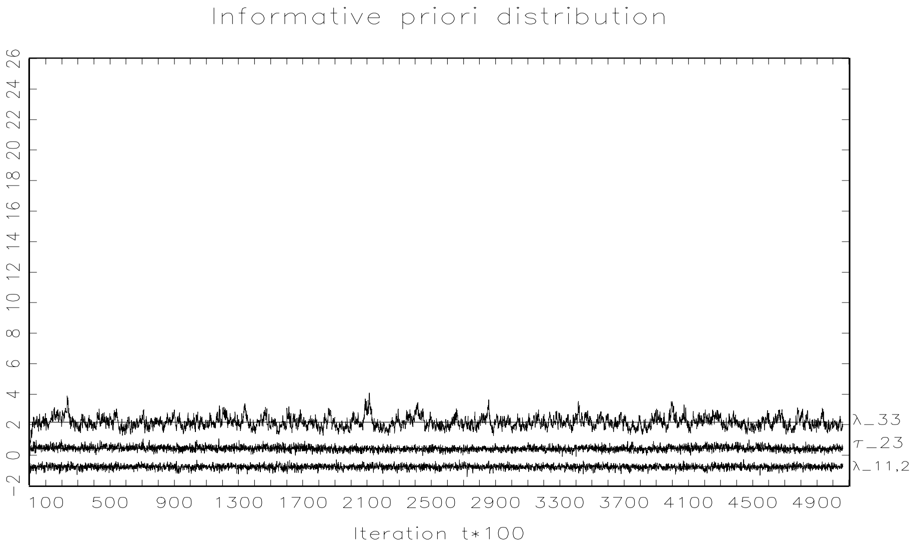

Figure 2: Time series for λ_{33}, $\lambda_{11,2}$ and τ_{23} of a Gibbs sampler run with 506000 iterations using a slightly informative priori distributions for λ_{33} where only every 100th iteration results were stored.

independent. The Gibbs sampler was run for 506000 iterations and the chain was thinned by storing only every 100th iteration. The result was remarkable: Some time series did contain high auto correlations, especially λ_{33}, but the mixing improved considerably (see figure 2).

Besides comparing diffuse and informative priors we also wanted to contrast our Bayesian analysis with a frequentist one. Tables 2 and 3 contain the Bayesian results for the parameters in $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ in the standardized solution where the variance of each y^*_{ij} is standardized to 1 to make the Bayesian and frequentist results comparable (the 60 threshold estimates are not listed). We computed the posterior means and standard deviations using the Gibbs sampler results of the long run for a model with a slightly informative normal prior. We discarded the first

6000 iterations. Also, table 2 and table 3 contain the ULS and WLS estimations together with the estimated WLS standard errors. On average the WLS estimates deviate most strongly from the other two approaches, while the Bayesian and the ULS estimates are very similar. Especially the WLS $\mathbf{\Gamma}$ parameters show great deviations from the corresponding Bayesian and ULS estimates. For example, γ_{34} is estimated for the Bayesian approach at 0.524 (0.183) (respectively at 0.466 for the ULS method) whereas for the WLS method the corresponding estimate is nearly twice as high with 1.072 (0.128).

Furthermore, the WLS standard errors are in all cases lower than the Bayesian standard deviations. Since the simulation studies also showed that the WLS standard errors are underestimated and the Bayesian standard deviations are estimated correctly, the latter are obviously more reliable.

Although from a quantitative point of view there are differences with respect to the estimation results, these deviations are not that large that for the three methods the interpretation of the results differs widely. For example, in a qualitative sense the latent factor structure is the same for the three different approaches. Factor 1 is characterized by variables that show an irrational conflict resolution pattern where both parents are scoffing, do not give each other a hearing, draw the children into their quarreling and shout at each other (items y_4 bis y_7, y_{11} bis y_{15}). The second factor has high loadings on variables that concern the behaviour of the father who is prepared to compromise and looks for more tranquile discourse (items y_8 bis y_{10}). The third factor is characterized by the mother's behaviour who shows affective self-control, gives in and comes around (items y_1 bis y_3).

The results of the structural part of the MIMIC model give indications that the parental styles of conflict resolution changed across generations. Estimates γ_{11} and γ_{21} show an influence of the age cohort on the first factor. That means that in the young and middle age cohort parents act in more irrational ways and are less prepared to make compromises than those in the elder age cohort. Parents of the younger cohorts show less affective self-control. An explanation of this finding may be the growing instability of marriage and family, which is caused by changed economic conditions and by the changes in the understanding of partnership.

A second effect concerns the style in handling conflicts by the mother (γ_{34}). On average, persons who have a medium educational level represent their mothers as more rational and more prepared to make compromises than subjects with a low or high educational level, who report that their mothers have less affective self-control and are more uncompromising.

y_j	1st factor ($r = 1$) Bayes	WLS	ULS	2nd factor ($r = 2$) Bayes	WLS	ULS	3rd factor ($r = 3$) Bayes	WLS	ULS
Mother									
y_1	.067 (.082)	.029 (.045)	-.012	.000	.000	.000	.778 (.039)	.780 (.026)	.719
y_2	.000	.000	.000	.199 (.068)	.314 (.039)	.159	.537 (.056)	.531 (.033)	.565
y_3	.038 (.086)	.051 (.045)	.000	.000	.000	.000	.895 (.030)	.847 (.023)	.926
y_4	.478 (.069)	.429 (.034)	.377	.312 (.074)	.317 (.038)	.315	-.442 (.068)	-.447 (.034)	-.465
y_5	.608 (.071)	.605 (.044)	.486	.368 (.079)	.397 (.047)	.360	-.474 (.077)	-.606 (.038)	-.544
y_6	.630 (.058)	.609 (.033)	.647	.050 (.086)	.164 (.052)	.058	-.200 (.082)	-.270 (.044)	-.236
y_7	.680 (.059)	.645 (.028)	.625	.102 (.084)	.198 (.048)	.150	-.441 (.076)	-.450 (.038)	-.475
Father									
y_8	.082 (.066)	.114 (.036)	.053	.689 (.049)	.619 (.027)	.645	.000	.000	.000
y_9	-.021 (.066)	-.144 (.031)	-.021	.718 (.050)	.695 (.029)	.707	.182 (.062)	.202 (.033)	.180
y_{10}	.000	.000	.000	.834 (.038)	.811 (.028)	.863	.000	.000	.000
y_{11}	.538 (.063)	.541 (.040)	.529	-.479 (.070)	-.492 (.037)	-.474	.153 (.075)	.202 (.043)	.167
y_{12}	.650 (.060)	.647 (.033)	.525	-.444 (.078)	-.549 (.036)	-.418	.066 (.082)	.034 (.037)	.061
y_{13}	.692 (.066)	.713 (.041)	.616	-.398 (.087)	-.515 (.040)	-.420	.121 (.087)	.031 (.040)	.095
y_{14}	.626 (.057)	.612 (.030)	.814	-.246 (.083)	-.362 (.041)	-.304	.000	.000	.000
Mother and Father									
y_{15}	.771 (.048)	.772 (.029)	.670	-.022 (.088)	-.110 (.048)	.006	-.242 (.084)	-.273 (.043)	-.256

Table 2: Standardized point estimations of the loadings λ_{jr}. The estimated Bayesian standard deviations respectively WLS standard errors are given in brackets.

	1. factor, $r = 1$			2. factor, $r = 2$			3. factor, $r = 3$		
	Bayes	MDE	ULS	Bayes	MDE	ULS	Bayes	MDE	ULS
γ_{r1}	.804 (.200)	1.148 (.151)	.993	-.348 (.189)	.033 (.140)	-.322	-.329 (.195)	-.453 (.157)	-.237
γ_{r2}	.707 (.199)	.846 (.156)	.943	-.268 (.191)	.138 (.140)	-.232	-.115 (.193)	-.235 (.153)	-.037
γ_{r3}	-.038 (.175)	.086 (.131)	.004	-.217 (.167)	-.018 (.112)	-.200	.238 (.165)	.082 (.120)	.189
γ_{r4}	.096 (.183)	.032 (.140)	.114	.092 (.180)	-.147 (.129)	.104	.524 (.183)	1.072 (.128)	.466

Table 3: Point estimations of the regressions coefficients in $\mathbf{\Gamma}$. The estimated Bayesian standard deviations respectively WLS standard errors are given in brackets.

5 Conclusion

Simulation results and the real data application in Section 4 suggest that fully Bayesian methods via Gibbs sampling or other MCMC techniques provide a useful supplementary approach for inference in MIMIC or more general latent variable models. Since they do not rely on large sample theory, they provide more reliable point estimates and standard errors of parameters also in medium sample size data situations compared to more traditional frequentist methods. However, preliminary analysis with the latter methods helps to formulate informative priors for high factor loadings, thus avoiding slow mixing of the Gibbs sampler when only diffuse or rather vague priors are imposed.

Due to the modular structure, another advantage of the hierarchical Bayesian modeling approach is its flexibility concerning modifications or generalizations. Interesting extensions for future work are: Incorporation of nondiagonal covariance matrices $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ in (1) and (2), based on reparametrization suggestions in Chen and Dey (1999), or inclusion of nonlinear or nonparametric effects of covariates or latent factors, following ideas in Arminger and Muthén (1998) and Fahrmeir and Lang (1999), respectively.

References

[1] Albert, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, **17**, 251-269.

[2] Albert, J. H. & Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669-679.

[3] Arminger, G. & Muthén, B. O. (1998). A Bayesian Approach to Non-Linear Latent Variable Models Using the Gibbs-Sampler and the Metropolis-Hastings Algorithm. *Psychometrika*, **63**, 271-300.

[4] Arminger, G., Wittenberg, J. & Schepers, A. (1996). *MECOSA 3 USER GUIDE*, Friedrichsdorf/Ts.: ADDITIVE GmbH.

[5] Bollen, K. A. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

[6] Browne, M. W. & Arminger, G. (1995). Specification and Estimation of Mean- and Covariance-Structure-Models. In G. Arminger, C. C. Clogg & M. E. Sobel (Hrsg.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (S. 185-249). New York: Plenum Press.

[7] Chen, M.-H. & Dey, D. K. (1999). Bayesian Analysis for Correlated Ordinal Data Models. In D. K. Dey, S. K. Ghosh & B. K. Mallicle (eds.), *Generalized linear models: A Bayesian perspective.* New York: Marcel-Dekker, to appear.

[8] Chib, S. & Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, **85**, 347-361.

[9] Cowles, M. K. & Carlin, B. P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statis-*

tical Association, **91**, 883-904.

[10] Döbert, R. & Nunner-Winkler, G. (1983). Moralisches Urteilsniveau und Verläßlichkeit. Die Familie als Lernumwelt für kognitive und motivationale Aspekte des moralischen Bewußtseins in der Adoleszenz. In G. Lind, H. A. Hartmann & R. Wakenhut (Hrsg.), *Moralisches Urteilen und Soziale Umwelt. Theoretische, methodologische und empirische Untersuchungen* (S. 95-122). Weinheim/Basel: Beltz.

[11] Fahrmeir, L. & Lang, S. (1999). Bayesian Inference for Generalized Additive Regression based on Dynamic Models. Discussion paper Nr. 134, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München. Available from the World Wide Web: http://www.stat-uni-muenchen.de/sfb386/publikation.html.

[12] Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface* (S. 571-578). Alexandria, VA: American Statistical Association.

[13] Hobert, J. P. & Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, **91**, 1461-1473.

[14] Knorr-Held, L. (1995). Dynamic Cumulative Probit Models for Ordinal Panel-Data; a Bayesian Analysis by Gibbs Sampling. Discussion paper Nr. 2, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München. Available from the World Wide Web: http://www.stat-uni-muenchen.de/sfb386/publikation.html.

[15] Muthén, B. O. (1988). *LISCOMP: Analysis of Linear Structural Equations with a Comprehensive Measurement Model* (2. Auflage). Mooresville: Scientific Software, Inc.

[16] Muthén, B. O. & Arminger, G. (1995). Bayesian Latent Variable Regression for Binary and Continuous Response Variables Using the Gibbs Sampler. Unpublished manuscript, Bergische Universität Wuppertal.

[17] Muthén, B. O. & Satorra, A. (1995). Technical Aspects of Muthén's LISCOMP Approach to Estimation of Latent Variable Relations with a Comprehensive Measurement Model. *Psychometrika*, **60**, 489-503.

[18] Nikele, M. (1999). *Ein Modell mit latenten Variablen für stetige und ordinale Response-Variablen: Bayesianische und frequentistische Schätzstrategien mit einem Anwendungsbeispiel aus der Soziologie.* PhD thesis. Berlin: Logos Verlag.

[19] Nunner-Winkler, G. (1999). Sozialisationsbedingungen moralischer Motivation. In H.-R. Leu & L. Krappmann (Eds.), *Zwischen Autonomie und Verbundenheit - Bedingungen und Formen der Behauptung von Subjektivität* (p. 299-329). Frankfurt a. M.: Suhrkamp Verlag.

[20] Sammel, M. D. & Ryan, L. M. (1996). Latent Variable Models with Fixed Effects. *Biometrics*, **52**, 650-663.

[21] Sammel, M. D., Ryan, L. M. & Legler, J. M. (1997). Latent Variable Models for Mixed Discrete and Continuous Outcomes. *Journal of the Royal Statistical Society*, B **59**, 667-678.

[22] Scheines, R., Hoijtink, H. & Boomsma, A. (1999). Bayesian Estimation and Testing of Structural Equation Models. *Psychometrika*, **64**, 37-52.