Fahrmeir, Raach:

A Bayesian semiparametric latent variable model for mixed responses

Sonderforschungsbereich 386, Paper 471 (2006)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

A Bayesian semiparametric latent variable model for mixed responses

Ludwig Fahrmeir and Alexander Raach
Department of Statistics, University of Munich

January, 2006

Abstract

In this article we introduce a latent variable model (LVM) for mixed ordinal and continuous responses, where covariate effects on the continuous latent variables are modelled through a flexible semiparametric predictor. We extend existing LVM with simple linear covariate effects by including nonparametric components for nonlinear effects of continuous covariates and interactions with other covariates as well as spatial effects. Full Bayesian modelling is based on penalized spline and Markov random field priors and is performed by computationally efficient Markov chain Monte Carlo (MCMC) methods. We apply our approach to a large German social science survey which motivated our methodological development.

Keywords: Latent variable models, mixed responses, penalized splines, spatial effects, MCMC.

1 Introduction

In many scientific fields, latent variable models (LVM) are used successfully to explain the interrelationships between the components of multivariate observable responses, to measure underlying unobservable constructs, and to assess the influence of covariates on observable and latent variables. A recent comprehensive introduction to LVM is provided by Skrondal and Rabe-Hesketh (2004). LVM presented in the literature so far mostly assume that the effects of covariates on both the observable responses or indicators and the latent variables are modelled in simple linear parametric form. In many research settings, as in the application which motivated our methodological development, this assumption is often too restrictive for revealing the true functional relationships between the covariates and the latent variables.

In this paper, we introduce a flexible semiparametric LVM with mixed binary, ordinal and continuous indicators. The effects of covariates of different type are modelled through a semiparametrically structured additive predictor including the usual linear parametric component, nonparametric functions for possibly nonlinear effects of continuous covariates as well as nonparametric interactions with other covariates, and – as a particular feature – spatial effects resulting from geographical information about the location or residence of units or individuals in the sample. Covariate effects of this type are present in our application in Section 4,

where we investigate a substantive research question emerging from the social science internet survey "Prospect Germany". Based on binary and ordinal indicators we analyze two latent variables: the attitude of German citizens if social coverage should be taken care primarily by the citizens or by the state, and the ambition of the respondent to achieve something in job and in society. Apart from effects of categorical covariates, we explore nonlinear effects of age and its interaction with sex as well as the spatial effect of the district in Germany where the respondent lives.

The origin of the LVM with covariate effects can be traced back to Jöreskog and Goldberger (1975) who named and discussed a multiple indicators and multiple causes model (MIMIC). Muthén (1989) extended the MIMIC model to include binary and ordinal manifest variables. Sammel, Ryan and Legler (1997) discussed a LVM with covariates for mixed outcomes in the Item Response Theory (IRT) context. A comparison of different approaches for ordinal indicators including covariate effects is provided by Moustaki, Jöreskog and Mavridis (2004). Zhu, Eickhoff and Yan (2005) firstly discussed the influence of spatial covariates on the latent variables using a ML approach. A latent variable model for mixed categorical and survival data has been recently suggested by Moustaki and Steele (2005). In all this work the effects of covariates are modelled through a simple linear predictor. A notable exception are nonlinear latent variables suggested by Arminger and Muthén (1998), but the nonlinear relationship is still of conventional parametric form. The semiparametrically structured additive predictor used in our LVM is described in Fahrmeir, Kneib and Lang (2004), and Brezger and Lang (2006) in the context of semiparametric generalized regression for univariate responses.

Traditionally most LVM rely on frequentist estimation methods such as maximum likelihood or weighted least squares. In this paper a fully Bayesian approach is employed where all unknown population parameters are considered to be random variables. This includes the specification of prior distributions for all parameters that have to be estimated. The posterior distribution of those parameters is obtained by using Markov chain Monte Carlo (MCMC) methods. A small primer of Bayesian models within the context of social sciences is provided by Rupp, Dey and Zumbo (2004).

This paper is structured in the following way: Section 2 presents the statistical model of the LVM which consists of two parts, the measurement model and the structural equation including the prior distributions of all parameters; Section 3 discusses the Bayesian model including the posterior distribution (all individual MCMC steps are summarized in Appendix 5); Section 4 demonstrates the application of our semiparametric LVM with the survey "Prospect Germany".

2 Statistical model

The LVM with covariate effects consists of two components: the measurement model for continuous, binary and ordinal response with covariates influencing the indicators directly (direct effects); and the structural model explaining the modification of the latent variables by covariates (indirect effects).

2.1 Measurement model

In all LVM, p different indicators or manifest variables are observed for n observations. Each indicator j ($1 \leq j \leq p$) can be of continuous, binary or ordinal type. The response value of indicator j of individuum i ($1 \leq i \leq n$) is denoted by y_{ij}. All indicators of a single individuum are contained in the $p \times 1$-dimensional vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})'$. For notational convenience we sort the manifest variables in such a way that the first p_1 indicators are binary or ordinal valued, and the remaining $p_2 = p - p_1$ indicators are continuous.

For ordinal indicators, an underlying unobserved variable y_{ij}^* is postulated. Let's assume that ordinal indicator j has K_j categories and its cutpoints are denoted by τ_{jk} ($0 \leq k \leq K_j$). The discrete value of an ordinal indicator y_{ij} is generated by the underlying variable y_{ij}^* through the threshold mechanism

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

for $1 \leq j \leq p_1$. Since ordinal categories are ordered, we have to impose an order restriction on the cutpoints as stated by $-\infty =: \tau_{j0} < \tau_{j1} < \tau_{j2} < \ldots < \tau_{jK_j} := \infty$. The distribution of the underlying variable is governed by the equation $y_{ij}^* = \mu_{ij} + \varepsilon_{ij}$ where μ_{ij} denotes the mean value and ε_{ij} is a random error variable drawn from the standard normal distribution. Let Φ denote the respective cumulative distribution function. Using (1) the probability p_{ijk} that category k for individual i and indicator j is observed, leads to

$$p_{ijk} = P(y_{ij} = k|\mu_{ij}) = P(\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}|\mu_{ij}) = \Phi(\tau_{jk} - \mu_{ij}) - \Phi(\tau_{j,k-1} - \mu_{ij})\,.$$

The logistic distribution function could also be used instead of the standard normal distribution function which leads to the logit model commonly used in the IRT approach; we use the standard normal distribution function because parameter estimates for both function lead to very similar results in prediction (Moustaki, 2003) and the Gibbs sampler can be employed. Due to identification restrictions, the cutpoints τ_{j1} of all indicators j are fixed to zero. For more information on ordinal modelling we refer to Fahrmeir and Tutz (2001), or Johnson and Albert (1999). Continuous variables are observed directly, i. e. $y_{ij}^* = y_{ij}$ for $p_1 < j \leq p$.

The relationship between the $\boldsymbol{y}_i^*$ variables and the q latent variables $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})'$ is given by the measurement model according to

$$\boldsymbol{y}_i^* = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{z}_i + \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{\varepsilon}_i\,, \tag{2}$$

with $\boldsymbol{\varepsilon}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. The q-dimensional vector $\boldsymbol{z}$ contains q latent variables which explain the relationships between the indicators $\boldsymbol{y}_i^*$. The p indicators $\boldsymbol{y}_i^*$ result from a linear combination of q latent variables plus individual error terms for each indicator. The $p \times q$ matrix $\boldsymbol{\Lambda}$ is composed of the factor loadings indicating the strength of relationship between latent factors and indicators; $\boldsymbol{\lambda}_0$ is a p-dimensional intercept vector. The direct covariates are summarized in the d-dimensional vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{id})'$ and the $p \times d$-dimensional matrix $\boldsymbol{A}$ contains the respective regression coefficients. As in factor analysis, the latent variables are i. i. d. with $\boldsymbol{z}_i \sim N_q(\boldsymbol{0}, I_q)$. Then the properties of the measurement model are conveniently described by its characterising moments (i) $\mathrm{Var}(y_{ij}|\boldsymbol{z}, \boldsymbol{w}) = \sigma_j^2$, (ii) $\mathrm{Var}(y_{ij}) = \sum_{r=1}^q \lambda_{jr}^2 + \sigma_j^2$, (iii) $\mathrm{Cov}(y_{ij}, y_{il}|\boldsymbol{z}, \boldsymbol{w}) = 0$, and (iv) $\mathrm{Cov}(y_{ij}, y_{il}) = \sum_{r=1}^q \lambda_{jr}\lambda_{lr}$.

Although direct effects are typically not in the focus of an analysis – most of the times the researcher is more interested in the factor loadings, the latent scores, and the influence of covariates on the latent variables – they still provide additional information about the data structure by including associations between indicators $\boldsymbol{y}_i$ and covariates $\boldsymbol{w}_i$, and not only among the $\boldsymbol{y}_i$ themselves (see Muthén, 1989).

The model still holds two identification problems; the first problem is associated with the modelling of ordinal variables, and the second problem is related to the uniqueness of the factor loadings matrix $\boldsymbol{\Lambda}$ and factor scores. Firstly, we recognize that the probability $P(y_{ij} \leq k) = P(y_{ij}^* \leq \tau_{jk}) = \boldsymbol{\Phi}\left((\tau_{jk} - \lambda_{j0})/\sqrt{\Phi_{jj}}\right)$ is not altered when we multiply the nominator and the denominator by a constant c_2 implying that τ_{jk}, λ_{j0} and ϕ_{jj} are only identified up to a multiplicative constant. In our model, we dispose of this identification problem by setting the error variances $V(\varepsilon_{ij}) = \sigma_j^2 = 1$. In the frequentist literature instead, it is common to fix the total variance of ordinal indicators to one. For that reason the parameter estimates of the Bayesian and the frequentistic models can not be directly compared, but the parameters can be converted easily. The reason why we proceed in a different way in the Bayesian context lies in the fact that full conditional distributions used in MCMC are of an easier form for the constraint $V(\varepsilon_{ij}) = \sigma_j^2 = 1$. Secondly, there is an indeterminacy with respect to the factor loadings matrix and factor scores. The model is invariant under transformations with any orthogonal $q \times q$ matrix $\boldsymbol{V}$ of the form $\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\boldsymbol{V}'$ and $\widetilde{\boldsymbol{z}}_i = \boldsymbol{V}\boldsymbol{z}_i$ because this transformation keeps the variance of the latent scores unchanged ($V(\boldsymbol{z}_i) = \boldsymbol{V}\boldsymbol{I}_k\boldsymbol{V}' = \boldsymbol{\Psi}$). An indefinite number of models exists again since all orthogonal rotations of the latent space could occur. The solution lies in the restriction of parameters of the factor loadings matrix $\boldsymbol{\Lambda}$ in a suitable way (e. g. Seber, 1984).

2.2 Structural model

Now we allow covariates or indirect effects to modify the latent variables by introducing the structural equation part of the model, i. e.

$$\boldsymbol{z}_i = \boldsymbol{\eta}_i + \boldsymbol{\xi}_i\,,$$

with $\boldsymbol{\xi}_i \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$ and a structured additive predictor vector $\boldsymbol{\eta}_i$. In the literature (e. g. Moustaki, Jöreskog and Mavridis, 2004), the predictor $\boldsymbol{\eta}_i$ is always of the linear form $\boldsymbol{\eta}_i = \boldsymbol{\gamma}\boldsymbol{u}_i$ where $\boldsymbol{\gamma}$ is a $q \times m$ matrix of regression coefficients, and $\boldsymbol{u}_i$ is a $m \times 1$ vector of fixed covariates of observation i which are summarized in the $n \times m$ matrix $\boldsymbol{U} = \{u_{il}\}$, $1 \leq l \leq m$. These parametric effects imply that the means of the latent variables are linearly dependent on the covariates $\boldsymbol{u}_i$ which is a severe restriction in many real-life research settings: firstly, for continuous covariates the assumption of a strictly linear effect on the predictor may not be appropriate – additionally, effects of continuous covariates may vary for different subgroups of the population; secondly, the latent variables might be spatially correlated. To incorporate those issues, we employ a more versatile predictor

$$\eta_{ir} = f_{r1}(x_{i1}) + \ldots + f_{rg}(x_{ig}) + f_{r,spat}(s_i) + \boldsymbol{\gamma}_r'\boldsymbol{u}_i\,, \tag{3}$$

where g denotes the number of nonparametric functions $f_{rh'}$ of continuous covariates $x_{ih'}$ $(1 \leq h' \leq g)$, $f_{r,spat}$ is the spatial effect of the location s_i and $\boldsymbol{\gamma}_r$ is the vector of values in the r-th row of the $q \times m$ dimensional matrix $\boldsymbol{\gamma}$ of regression coefficients. The index h comprises

all g functions of continuous covariates plus the spatial effect, i. e. $h \in 1, \ldots, g, spat$. We recognize that a separate function per covariate has to be estimated for each of the latent variables, leading to

$$\boldsymbol{\eta}_i = \boldsymbol{f}_1(x_{i1}) + \ldots + \boldsymbol{f}_g(x_{ig}) + \boldsymbol{f}_{spat}(s_i) + \boldsymbol{\gamma}\boldsymbol{u}_i, \tag{4}$$

where $\boldsymbol{f}_h$ are now q-dimensional vector valued functions. If (4) does not contain a spatial effect $\boldsymbol{f}_{spat}$ and all covariates $x_{ih'}$ are continuous, an additive LVM is obtained. The structure of the predictor stems from the class of generalized additive models (GAM) as described by Hastie and Tibshirani (1990). If (4) additionally contains a spatial effect $\boldsymbol{f}_{spat}$, a geoadditive LVM is obtained whose predictor is structured as in geoadditive models (see Kammann and Wand, 2003). Finally, interactions of continuous and categorical covariates can be included when the functions are of the form $f_{rh'}(x_{ih'}) = f_{rh'}(\tilde{x}_{ih'}, v_{ih'}) = g_{rh'}(\tilde{x}_{ih'})v_{ih'}$. This leads to the class of varying coefficient LVM derived from varying coefficient models (VCM) according to Hastie and Tibshirani (1993). Hence we obtain a structured additive predictor (STAR) which embodies a wide range of models (see Fahrmeir, Kneib and Lang, 2004).

We assume the structure of the linear predictor to be equal for all latent variables, although in general a different structure of the predictor might be used for each latent variable. In component notation, the predictor of equation (4) yields

$$\boldsymbol{\eta}_i = \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \vdots \\ \eta_{iq} \end{pmatrix} = \begin{pmatrix} f_{11}(x_{i1}) \\ f_{21}(x_{i1}) \\ \vdots \\ f_{q1}(x_{i1}) \end{pmatrix} + \ldots + \begin{pmatrix} f_{1g}(x_{ig}) \\ f_{2g}(x_{ig}) \\ \vdots \\ f_{qg}(x_{ig}) \end{pmatrix} + \begin{pmatrix} f_{1,spat}(s_i) \\ f_{2,spat}(s_i) \\ \vdots \\ f_{q,spat}(s_i) \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \ldots & \gamma_{1m} \\ \gamma_{21} & \ldots & \gamma_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{q1} & \ldots & \gamma_{qm} \end{pmatrix} \cdot \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{im} \end{pmatrix}.$$

For each function f_{rh}, a comparably large number of parameters d_h have to be estimated. Let $\boldsymbol{\beta}_{rh} = (\beta_{rh,1}, \beta_{rh,2}, \ldots, \beta_{rh,d_h})'$ denote the coefficient vector of function f_{rh}. The vectors of function evaluations $\boldsymbol{\beta}_{rh}$ allow a feasible notation of the vector $\boldsymbol{\eta}_r$ which contains the predictor values of all observations i in the following way:

$$\boldsymbol{\eta}^{(r)} = (\eta_{1r}, \eta_{2r}, \ldots, \eta_{nr})' = \boldsymbol{X}_1\boldsymbol{\beta}_{r1} + \ldots + \boldsymbol{X}_g\boldsymbol{\beta}_{rg} + \boldsymbol{X}_{spat}\boldsymbol{\beta}_{r,spat} + \boldsymbol{U}\boldsymbol{\gamma}_r\,, \tag{5}$$

with suitably defined $n \times d_h$ dimensional design matrices $\boldsymbol{X}_h$ whose entries depend on the type of function and modelling approach. The most compact form of the predictor $\boldsymbol{\eta}$ is then obtained by

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \ldots & \eta_{1q} \\ \vdots & \ldots & \vdots \\ \eta_{n1} & \ldots & \eta_{nq} \end{pmatrix} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \ldots + \boldsymbol{X}_g\boldsymbol{\beta}_g + \boldsymbol{X}_{spat}\boldsymbol{\beta}_{spat} + \boldsymbol{U}\boldsymbol{\gamma} = \sum_{h=1}^{spat} \boldsymbol{X}_h\boldsymbol{\beta}_{rh},$$

where the $d_h \times q$ dimensional matrix $\boldsymbol{\beta}_h = (\boldsymbol{\beta}_{1h}, \boldsymbol{\beta}_{2h}, \ldots, \boldsymbol{\beta}_{qh})$ contains all vectors parameterizing functions. Additionally one smoothing parameter κ_{rh} has to be estimated for each function f_{rh} which is further explained in Section 2.3.2. In total, the structural model contains $q \cdot g$ functions plus $q \cdot m$ regression coefficients plus $q \cdot g$ smoothing parameters, adding up to $q \cdot (\sum_{h=1}^{g} d_h + m + g)$ parameters.

It has to be noted that there is no constant intercept allowed in the predictor due to identification restrictions. This can be readily seen by assuming a simple model with one latent

factor z_i, no direct effects, an intercept γ_0 in the predictor of the structural equations and no further indirect effects. This leads to $\mathrm{E}(y_{ij}^*) = \lambda_{j0} + \lambda_{j1}\gamma_0$. Accordingly, the intercept γ_0 in the structural equation can not be estimated independently from the intercepts λ_{j0} in the measurement model because adding a constant c_3 to γ_0 can be offset by reducing λ_{j0} with $c_3\lambda_{j1}$. Another consequence of this restriction is the necessity to center all functions f_{rh} around zero.

2.3 Prior distributions

This section presents the prior specifications of all parameters. Since the prior distributions of the underlying variables $\boldsymbol{y}^*$ and the latent variables $\boldsymbol{z}$ are implicitly determined by the prior distributions of all other parameters and the distributional assumptions about $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\xi}_i$, only the prior distributions for the parameter vector $\boldsymbol{\theta} = \mathrm{vec}\{\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}$ have to be specified. From now on, let's assume that the indivual parts of the model are stochastically independent, thus the prior distribution yields

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}) \cdot p(\boldsymbol{\Sigma}) \cdot p(\boldsymbol{\tau}) \cdot p(\boldsymbol{\beta}, \boldsymbol{\gamma})\,. \tag{6}$$

2.3.1 Prior distribution of parameters excluding functions and spatial effects

Regarding the intercepts, factor loadings and direct effects, we define a $p \cdot (1 + q + d)$ dimensional vector $\bar{\boldsymbol{\lambda}}$ which contains all parameters of $\boldsymbol{\lambda}_0$, $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$ arranged as follows $\bar{\boldsymbol{\lambda}} := (\lambda_{10}, \lambda_{11}, \ldots, \lambda_{1q}, a_{11}, \ldots, a_{1d}, \ldots\ldots, \lambda_{p0}, \lambda_{p1}, \ldots, \lambda_{pq}, a_{p1}, \ldots, a_{pd})$. The prior distribution selected for $\bar{\boldsymbol{\lambda}}$ is a $p\cdot(1+q+d)$ dimensional multivariate normal density with the mean $\bar{\boldsymbol{\lambda}}^*$ and the precision matrix $\bar{\boldsymbol{\Lambda}}^*$, i. e. $\bar{\boldsymbol{\lambda}} \sim N(\bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{\Lambda}}^{*-1})$. If the researcher has no prior information about the parameters, noninformative priors should be used. In order to compare Bayesian parameter estimates with frequentist estimates, it is recommended to use noninformative priors because then the posterior solely depends on the likelihood part, and both parameter estimates coincide. For our purposes, we choose noninformative priors for the intercepts $\boldsymbol{\lambda}_0$ and the regression coefficients $\boldsymbol{A}$ of the direct effects. However we are forced to include prior information for the factor loadings for ordinal indicators in order to prevent the occurrence of Heywood cases in the Bayesian setting (Heywood, 1931; Bartholomew, 1987). A Heywood case appears when one latent factor loads up completely on one indicator, hence the latent variable accounts for the full variability of the respective indicator, and the corresponding communality equals 1. Since this result is highly implausible, informative priors are chosen with a normal density centered at zero with a certain variance. A standard choice in applications (Lopes and West, 2004; Quinn, 2004) is a prior variance of one because this prevents the occurrence of Heywood cases, is highly diffuse and therefore does not influence the estimation of the factor loadings. In this work two different MCMC algorithms are used – the second algorithm sometimes requires a stronger normal prior on the factor loadings. Hence our prior precision matrix $\bar{\boldsymbol{\Lambda}}^*$ equals zero for the off-diagonal elements, and the diagonal elements also equal zero except for the factor loadings parameters λ_{jr} which are set to one of the values 1.0 (standard prior) or 4.0 (strong prior).

For continuous indicators, error variances have to be estimated. Since the error variance matrix $\boldsymbol{\Sigma}$ is diagonal, its prior distribution can be given by specifying the individual prior

distributions of σ_j^2. The standard conjugate prior choice for error variances in a linear model with normally distributed errors is the inverse Gamma distribution, hence $\sigma_j^2 \sim IG\left(\nu/2, \nu s^2/2\right)$ for $j = p_1 + 1, \ldots, p$ with two hyperparameters: the degrees of freedom $\nu > 0$, and the scale parameter $s > 0$ (see Gelman et al., 2004). The prior density is $p(\sigma_j^2) \propto (\sigma_j^2)^{-(\nu/2+1)} \exp(-\nu s^2/(2\sigma_j^2))$ for $\sigma_j^2 > 0$. A noninformative prior distribution is obtained for $\nu \longrightarrow 0$ and $\nu s^2 \longrightarrow 0$, and results in the improper prior distribution $p(\sigma_j^2) \propto 1/\sigma_j^2$ with $\sigma_j^2 \in [0; \infty[$. We eschew the use of improper priors for the error variances σ_j^2 to prevent Heywood cases and improper posteriors. Hyperparameters ν and s should be chosen in such a way to correctly include prior information if available. If noninformative priors are to be used, the prior distributions should have a vanishing probability for $\sigma_j^2 \longrightarrow 0$ to prevent Heywood cases (Lopes and West, 2004).

For the cutpoints $\boldsymbol{\tau}$, we choose noninformative diffuse prior distributions so that the order condition $0 < \tau_{j2} < \tau_{j3} < \ldots < \tau_{j,K_j-1} < \infty$ for $j = 1, \ldots, p_1$ is fulfilled.

The conjugate prior distribution of the vector of regression coefficients $\boldsymbol{\gamma}_r$ is a m-dimensional multivariate normal density with the mean $\boldsymbol{\gamma}_r^*$ and the precision matrix $\boldsymbol{\Gamma}_r^*$, i. e. $\boldsymbol{\gamma}_r \sim N(\boldsymbol{\gamma}_r^*, \boldsymbol{\Gamma}_r^{*-1})$. In our analyses, we always choose noninformative priors for all regression parameters $\boldsymbol{\gamma}_r$, hence all values of $\boldsymbol{\Gamma}_r^*$ are set to zero.

2.3.2 Prior distributions for functions of continuous covariates

In this section, priors for the nonparametric function parameters $\boldsymbol{\beta}$ and for the parametric effects $\boldsymbol{\gamma}$ are specified. We assume the independence of prior specifications between separate functions and parametric effects, and between functions and parametric effects of different latent variables, thus $p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{r=1}^{q} \prod_{h=1}^{spat} p(\boldsymbol{\beta}_{rh}) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r)$. Priors for nonparametric continuous covariates are based on Gaussian smoothness priors (see Fahrmeir and Tutz, 2001), and priors for spatial covariates are based on Markov random fields (see Besag and Kooperberg, 1995). Conveniently, in the Bayesian approach both types of covariates can be treated in a unifying framework involving the use of a penalty matrix $\boldsymbol{K}$.

The nonparametric effects for continuous covariates are modelled as P-splines. Dropping indices to simplify the notation, f denotes a nonparametric function, $\boldsymbol{\beta}$ is the vector of function parameters, x represents the continuous covariate, and d denotes the dimension of the vector of function parameters. Our Bayesian approach is based on the work of Lang and Brezger (2004), and Brezger and Lang (2006) who give a detailed account on Bayesian P-splines in various settings. The unknown function f of a continuous covariate x is approximated by a polynomial spline of degree D defined on a set of equally spaced knots $x^{min} = \varrho_0 < \varrho_1 < \ldots < \varrho_{I-1} < \varrho_I = x^{max}$ with I intervals. This polynomial spline is constructed by a linear combination of $d = D + I$ B-spline basis functions B_c in the following way:

$$f(x) = \sum_{c=1}^{d} \beta_c B_c(x) \,.$$

The vector of function parameters now contains the regression coefficients or weights of the individual B-spline basis functions, short $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{D+I})'$. The characteristics of B-splines are described in the above mentioned literature. The smoothness of the function

f is achieved by penalizing too high differences of coefficients of adjacent B-splines. In a Bayesian approach, this penalization is incorporated conveniantly by applying a first-order or second-order random walk prior to the B-splines regression coefficients f. The first-order and second-order random walks are defined as

$$\beta_t = \beta_{t-1} + u_t \quad \text{and} \quad \beta_t = 2\beta_{t-1} - \beta_{t-2} + u_t$$

with $u_t \sim N(0, \kappa^2)$, respectively. First-order random walk has a diffuse prior $\beta_1 \propto constant$; second-order random walk additionally has $\beta_2 \propto constant$. The variance κ^2 determines the smoothness of the resulting function f, and acts as an inverse smoothing parameter. The entire prior distribution of a function f can equivalently be rewritten in form of a global smoothness prior

$$p(\boldsymbol{\beta}) = \prod_{t=2}^{d} p(\beta_t|\beta_{t-1}, \kappa^2) \propto \exp(-\frac{1}{2\kappa^2}\sum_{t=2}^{d}(\beta_t - \beta_{t-1})^2) = \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta})\,;$$

in the case of the first-order random walk the penalty matrix leads to

$$\boldsymbol{K} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

Classic first-order and second-order random walks are obtained when P-Splines of degree 0 are chosen and the knots are equal to the unique observations of x. The design matrix $\boldsymbol{X}$ is constructed in the following way: each row i of $\boldsymbol{X}$ contains the values of the B-spline basis functions evaluated at x_i, hence $X_{ic} = B_c(x_i)$. $\boldsymbol{X}$ consists of $D + I$ columns, and each row in X has $D + 1$ non-zero values. Thus the vector of function evaluations for all observations i is given by $\boldsymbol{X}\boldsymbol{\beta}$. In our analyses, we choose B-splines of degree $D = 3$ with $I = 10$ intervals.

Considering nonparametric interactions in a VCM, the function is of the form $f(x_i) = f(\tilde{x}_i, v_i) = g(\tilde{x}_i)v_i$ where the effect modifiers $\tilde{x}_i$ are continuous covariates, and the interacting variables v_i are continuous or categorical. We restrict our model to cope with categorical interacting variables. Since the differences between two categories of an ordinal or categorical variable are not interpretable, we apply a dummy coding for v (see Fahrmeir and Tutz, 2001). Let's assume that v has K categories, then we define

$$v_i^{(k)} = \begin{cases} 1 & \text{, if sample } i \text{ observes category } k \\ 0 & \text{, else} \end{cases} \qquad , k = 1, \ldots, K\,.$$

The dummy coding implies the estimation of K different functions $f^{(k)}$ with function parameter values $\boldsymbol{\beta}^{(k)}$, so that the predictor for the function f results in

$$f = f^{(1)} + \ldots + f^{(K)} = X^*\boldsymbol{\beta}^{(1)} + diag(v_1^{(2)}, \ldots, v_n^{(2)})X^*\boldsymbol{\beta}^{(2)} + \ldots + diag(v_1^{(K)}, \ldots, v_n^{(K)})X^*\boldsymbol{\beta}^{(K)}.$$

Here the reference category was set to category 1, but arbitrary reference categories are possible. The design matrix X^* is the usual design matrix associated with the continuous function $g(\tilde{x})$.

2.4 Prior distribution for the spatial effect

Let us assume the covariate s_i denotes the region of observation i, and the vector of function evaluations $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)$ contains the estimates of the d different regions. The spatial function evaluations of all observations i can be written as $\boldsymbol{X\beta}$ with the $n \times d$ dimensional design matrix $\boldsymbol{X}$, where $X_{is} = 1$ if observation i is associated to region s; all other values of row i equal zero. The basic assumption is that adjacent regions should have a similar impact on the latent scores whereas two regions far apart from each other do not exhibit such a similarity. In order to make a prior specification, the full neighborhood structure for each region has to be known. In our context, two regions are considered neighbours when they share a common boundary. We apply the following spatial smoothness prior to the function evaluations β_c $(c = 1, \ldots, d)$ for all d regions:

$$\beta_s | \beta_{s'}, s' \neq s, \kappa^2 \sim N\left(\sum_{s' \in \partial_s} \frac{\beta_{s'}}{N_s}, \frac{\kappa^2}{N_s} \right) , \tag{7}$$

where N_s indicates the number of adjacent sites of region s, and $s' \in \partial_s$ denotes all regions s' being neighbours of region s. Hence the conditional mean of β_s is an unweighted average of the function values of all adjacent regions. Since spatial data, e. g. regions, does not inhibit a natural ordering, a symmetric conditioning is applied. Different definitions of neighbourhood are given in Besag, York and Mollie (1991) and can be modelled by a more general prior distribution including equation (7) as a special case as described in Fahrmeir and Lang (2001b). The entire prior distribution follows as $p(\boldsymbol{\beta}) \propto \exp(-\boldsymbol{\beta}' K \boldsymbol{\beta}/(2\kappa^2))$ with the d-dimensional penalty matrix $\boldsymbol{K}$ whose entries are

$$k_{ss} = N_s \qquad \text{and} \qquad k_{ss'} = \begin{cases} -1 & , s' \in \partial_s \, , \\ 0 & , \text{otherwise} \, . \end{cases}$$

More information about Markov random fields is given in Rue (2005).

2.4.1 Prior distributions of smoothing parameters

All priors for nonparametric functions and the spatial effect are defined conditional on the inverse smoothing parameter κ^2, i. e. $p(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\kappa^2)p(\kappa^2)$. It is automatically estimated in our Bayesian approach. To complete the prior specification for nonparametric effects, we define the prior of the hyperparameter κ^2 to be $p(\kappa^2) \propto \frac{1}{(\kappa^2)^{a+1}} \exp(-b/\kappa^2)$ where $a \in \mathbb{R}$ and $b > 0$. If $a > 0$, this expression corresponds to an inverse Gamma distribution IG(a,b). The parameters a and b have to be chosen appropriately. Common choices include $a = b = 0.001$ leading to an almost noninformative prior for κ^2; or $a = 1$ and b equal a very small value, e. g. $b = 0.005$ as proposed by Besag et al. (1995). The choice of such highly vague but proper priors prevents problems associated with noninformative priors such as the nonconvergence of the Gibbs sampler (see Hobert and Casella, 1996) due to improper posterior distributions. However, further studies (see Raach, 2005) show that certain noninformative priors (Sun, Tsutakawa and He, 2001) lead to proper posteriors and correct function estimates.

To conclude, we want to emphasize that priors of all nonparametric effects (continuous, spatial, and interaction) can be modelled in a unifying framework with $p(\boldsymbol{\beta}) \propto \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}' K \boldsymbol{\beta})$

and a suitably defined penalty matrix $\boldsymbol{K}$. Finally, the full prior specification is given by

$$p(\boldsymbol{\beta},\boldsymbol{\gamma}) = \prod_{r=1}^{q}\prod_{h=1}^{spat} p(\boldsymbol{\beta}_{rh}) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r) \propto \prod_{r=1}^{q}\prod_{h=1}^{spat} \exp(-\frac{1}{2\kappa_{rh}^2}\boldsymbol{\beta}_{rh}'\boldsymbol{K}_{rh}\boldsymbol{\beta}_{rh}) p(\kappa_{rh}^2) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r)\,.$$

3 Bayesian inference

The posterior distribution is obtained via Bayes' formula according to $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the parameter vector that has to be estimated and $\boldsymbol{y}$ represents the given data. The parameter vector that has to be estimated is $\boldsymbol{\theta} = \text{vec}\{\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}$. The data $\boldsymbol{y}$ includes indicators $\boldsymbol{y}_i$, direct effects $\boldsymbol{w}_i$, indirect nonparametric effects $\boldsymbol{x}_i$, and indirect parametric effects $\boldsymbol{u}_i$ for $i = 1, \ldots, n$. It turns out that this Bayesian setup does not lead to a posterior that can be estimated by a convenient MCMC algorithm due to high-dimensional integrals which have to be solved. This problem can be resolved by extending the parameter space with nonobservable data, i. e. the underlying variables $\boldsymbol{y}^*$ for ordinal indicators, and the latent variables $\boldsymbol{z}$. This is a common approach in Bayesian methodology called data augmentation which was introduced by Tanner and Wong (1987). Albert and Chib (1993) implemented this approach for the estimation of regression parameters and cutpoints for ordinal response. The resulting Bayesian model yields

$$p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})\,. \tag{8}$$

The augmented posterior distribution is not just a mere technicality to enable efficient and easy sampling, it also empowers the interpretation of latent variables because the values of all latent variables z_{ir} are automatically estimated. For example, this enables to rank observations according to their respective latent variable value, and statements can be made about the probability that observation i_1 has a higher latent variable value than observation i_2. This property is an important advantage of the Bayesian approach compared to the frequentist approach. The underlying variables y^* are also automatically estimated but usually are of minor importance to the researcher. The following subsection deals with the prior specifications $p(\boldsymbol{\theta})$.

The posterior distribution is obtained by multiplying the prior distributions with the likelihood, leading to

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) &\propto p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \\ &= p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \cdot p(\boldsymbol{y}|\boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \\ &= p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \cdot p(\boldsymbol{y}|\boldsymbol{y}^*, \boldsymbol{\tau}). \end{aligned}$$

The simplification from the second to the third row is induced by equation (1) which states that the ordinal response $\boldsymbol{y}_{ij}$ is solely determined by the corresponding underlying variable $\boldsymbol{y}_{ij}^*$ and the cutpoints $\boldsymbol{\tau}_j$. Accordingly, the likelihood splits into two separate parts, the joint distribution of the underlying and latent variables given the parameters and covariates, and the distribution of the actual response given the underlying variables and the cutpoints. The joint distribution of $\boldsymbol{y}_i^*$ and $\boldsymbol{z}_i$ results from a combination of the measurement model and the structural equation, and is easily obtained by standard statistical calculus for multivariate

densities, hence $\boldsymbol{y}_i^*, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i$ results in

$$\begin{bmatrix} \boldsymbol{y}_i^* \\ \boldsymbol{z}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\lambda_0} + \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{\Lambda}\boldsymbol{\eta}_i \\ \boldsymbol{\eta}_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}' & \boldsymbol{I}_{m \times m} \end{bmatrix} \right), \tag{9}$$

where the predictor $\boldsymbol{\eta}_i$ is a function of $\boldsymbol{x}_i$ and $\boldsymbol{u}_i$ according to equation (4). The second part of the likelihood is trivial for continuous indicators. The probability density is constant because no underlying variable is necessary, and $\boldsymbol{y}_i^*$ equals the actual response $\boldsymbol{y}_i$. For ordinal indicators however, the response y_{ij} is unambiguously determined by the underlying variable y_{ij}^* and the cutpoints $\boldsymbol{\tau}_j$ through equation (1). Assuming independently and identically distributed observations, we obtain

$$p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[p(\boldsymbol{y}_i^*, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \left(\prod_{j=1}^{p_1} \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \right) \right],$$

with $p(\boldsymbol{y}_i^*, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i)$ as given in Equation (9). Obviously the analytical calculation of this high-dimensional density is impossible, and direct sampling is a difficult task to do and highly inefficient. For that reason we introduce two different MCMC algorithms which sample the parameters in a sequential fashion as outlined in the Appendix 5. In particular, we will make extensive use of the Gibbs sampler that is well suited to sample from normal densities.

The first MCMC procedure for the estimation of cutpoints for ordinal indicators was presented by Albert and Chib (1993) based on Gibbs sampling steps; unfortunately this algorithm exhibits poor convergence properties. Therefore our two samplers are based on two improved algorithms which both improve convergence considerably: firstly, Cowles (1996) introduced a Metropolis-Hastings step for the sampling of the cutpoints based on the idea of collapsed sampling (this algorithm is therefore called MH sampler or MHS); secondly, Liu and Sabatti (2000) use a transformation of some parameters at the end of each MCMC iteration.

In the MCMC context, the theorem of Liu and Sabatti (2000) can be used by applying a transformation to a group of parameters. Since the transformation maintains the distribution of those parameters, the stationary posterior distribution is not altered by this procedure. A Gibbs sampler with such a transformation is called a Generalized Gibbs sampler (GGS) and the transformation in each iteration is called a Generalized Gibbs move. Whether a GGS shows a better convergence than its parent MCMC algorithm depends crucially on the choice of the transformation group, and the form of the posterior distribution. Clearly the slowly converging parameters should be transformed by the Generalized Gibbs move in order to improve convergence and sampling the transformation parameters should be straightforward and fast. Note that in practice it is often difficult to find a suitable transformation group which both improves convergence and allows efficient sampling of the transformation members. However this method has proved to enhance the convergence of cutpoint parameters in ordinal probit models with an underlying latent variable (e. g. see Liu and Sabatti, 2000). The full conditionals of the MHS and the GGS are presented in Appendix A in detail.

Performance of MCMC inference is investigated in Raach (2005) through several simulation studies. Results can be summarized as follows. For both the MHS and GGS, convergence properties and quality of parameter estimates are very satisfactory. The MHS should be used for low number of observations because the probability of obtaining a Heywood case is lower

than for the GGS; for a high number of observations, the GGS should be preferred due to its slightly better convergence properties and its far superior computational speed. For both samplers, parameter estimates of the semiparametric predictor are excellent apart from a bias which occurs for estimated functions belonging to non-reference categories for nonparametric interactions when a low number of samples are observed.

4 Application

In this section the practicability of our LVM is demonstrated with the real data set of an internet survey from 2001 called "Perspektive Deutschland 1" or "Prospect Germany 1", abbreviated by PD1. This internet survey was initiated by the companies McKinsey&Comp., stern.de and T-Online, and approximately 170.000 people participated in the survey. The general goal of the survey was to receive answers from the population in which areas of life people are willing to bear responsibility, and in which areas they consider the state to fulfill the duty; another focus of the study is to measure the happiness of the population with the living conditions and the state's infrastructure offerings at their place of living. As common for social surveys, almost all of the variables are of binary and ordinal type. The only continuous variable is given by age, and one spatial variable is provided by the administrative district of the participant.

Our analysis includes two latent variables and ten indicators. We have generated a subsample which contains valid answers for all employed indicators and covariates, resulting in $N = 6804$ observations. The first latent construct used in our analysis is supposed to reflect the participant's attitude if social coverage should be ensured by the citizens on one's own responsibility, or if the state has to take care of social coverage; for example, a person with a high value of the latent construct would rather like citizens to take on more responsibility, to lay aside a certain amount of money in case of unemployment, to save privately for their retirement and to possess a private health insurance according to their needs. The second latent construct shall reflect the ambition of the person to achieve something in job and in society. Since both latent variables can not be observed directly, we try to measure them by using ten manifest variables which are described in Table 1.

Four indirect covariates are used in our analyses. There are two categorical covariates (*Sex*, *Inc* denoting income), one continuous covariate (*Age*), and one spatial covariate (*Reg*), all summarized in Table 2. The covariate *Inc* is based on Germany's old currency, the "Deutsche Mark" (DM), and one EURO equals 1.95583 DM. The spatial covariate *Reg* is determined by the 402 administrative districts (excluding the island of Rügen).

The factor loading λ_{12} is fixed to zero due to identification restrictions. The measurement

Explanation of the ten indicators			
No.	**Indicator (type)**	**Question/statement**	**Response categories**
1	*Health* (ordinal)	To what extent should the state make provisions for the citizens' health care ?	1. Completely. 2, 3, 4, 5, ... 6. Not at all.
2	*System* (binary)	Which type of social system would you prefer ?	1. The state guarantees each citizen a sufficient social coverage. The associated costs are payed by all citizens in the form of taxes and contributions according to their income. 2. Citizens can decide by themselves if and to which extent they want to cover themselves and their families in the case of illness, unemployment, retirement and nursing. Everybody who is not insured in order to save contributions will have to bear the risks.
3	*Initiative* (ordinal)	I think it's correct that in future every individual must increasingly take care about his/her provision for old age than it's the case today.	1. Absolutely true. 2, 3, 4, 5, ... 6. Absolutely wrong.
4	*Retirement* (ordinal)	To what extent should the state take care of the provision of old age ?	1. Completely. 2, 3, 4, 5, ... 6. Not at all.
5	*Emergency* (ordinal)	To what extent should the state take care of the citizens' protection in difficult life circumstances and emergencies ?	1. Completely. 2, 3, 4, 5, ... 6. Not at all.
6	*Perform* (ordinal)	I consider it important to perform better than other people.	1. Absolutely true, 2, 3, 4, 5, ... 6. Absolutely wrong.
7	*Society* (ordinal)	I want to achieve something in the society.	1. Absolutely true. 2, 3, 4, 5, ... 6. Absolutely wrong.
8	*Reputation* (ordinal)	How important is the following regarding your job: To gain respect and a good reputation in the public.	1. Absolutely important. 2, 3, 4, 5, ... 6. Not important at all.
9	*Salary* (ordinal)	How important is the following regarding your job: a high salary.	1. Absolutely important. 2, 3, 4, 5, ... 6. Not important at all.
10	*Career* (ordinal)	How important is the following regarding your job: To make a career.	1. Absolutely important. 2, 3, 4, 5, ... 6. Not important at all.

Table 1: *Variable names, variable types, questions/statements, and response categories of the ten indicators.*

Covariates used in the analyses		
Covariate name	**Covariate type**	**Response categories/range**
Sex	Categorical	1. Male 2. Female
Inc	Categorical	1. Less than 2500 DM net household income per month. 2. Between 2500 DM and 4500 DM net household income per month. 3. Between 4500 DM and 7500 DM net household income per month. 4. More than 7500 DM net household income per month.
Age	Metric	20, 21, ..., 70 years of age.
Reg	Spatial	1, 2, ..., 402 regions of Germany.

Table 2: *Variable names, variable types, and response categories of the four indirect covariates used in the analyses.*

model results in

$$\begin{pmatrix} y_{i1}^* \\ y_{i2}^* \\ y_{i3}^* \\ y_{i4}^* \\ y_{i5}^* \\ y_{i6}^* \\ y_{i7}^* \\ y_{i8}^* \\ y_{i9}^* \\ y_{i,10}^* \end{pmatrix} = \begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \lambda_{30} \\ \lambda_{40} \\ \lambda_{50} \\ \lambda_{60} \\ \lambda_{70} \\ \lambda_{80} \\ \lambda_{90} \\ \lambda_{10,0} \end{pmatrix} + \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \\ \lambda_{91} & \lambda_{92} \\ \lambda_{10,1} & \lambda_{10,2} \end{pmatrix} \cdot \begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \\ \varepsilon_{i6} \\ \varepsilon_{i7} \\ \varepsilon_{i8} \\ \varepsilon_{i9} \\ \varepsilon_{i,10} \end{pmatrix} \quad \text{, with} \quad \varepsilon_{ij} \sim N(0,1)\,.$$

The predictor of the structural equation is

$$\boldsymbol{\eta}_i = \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} f_{11}(Age_i) \\ f_{21}(Age_i) \end{pmatrix} + \begin{pmatrix} f_{12}(Age_i) \cdot Sex2_i \\ f_{22}(Age_i) \cdot Sex2_i \end{pmatrix} + \begin{pmatrix} f_{13}(Reg_i) \\ f_{23}(Reg_i) \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{pmatrix} \cdot \begin{pmatrix} Inc2_i \\ Inc3_i \\ Inc4_i \end{pmatrix}.$$

In this context indicator variables are used; for example $Sex2_i$ is set to one if observation i is female, otherwise it is set to zero. The estimates of the factor loadings and parametric regression coefficients are summarized in Table 3, and the smooth functions for the basic age effect and the age effect of females (all modelled by P-splines priors of degree 3 and 10 intervals) are drawn in Figure 1; the spatial functions of covariate *Reg* are shown in Figure 2.

Looking at the estimated mean factor loadings, it follows that the variation of all ten indicators is based significantly on the two latent constructs. The factor loadings indicate that the first latent variable loads onto the first five indicators, and indicators 6 to 10 measure the second latent variable. Concerning the first latent variable, indicator 4 (*Retirement*) stands out by having the highest factor loading of 1.3 because the corresponding question regarding the old-age provision aims very closely at the idea of the latent construct. For the second latent variable, indicator 10 (*Career*) shows the highest factor loading. In general, LVM including covariate effects have slightly lower factor loadings than models without covariate effects; the reason for this behaviour lies in the fact that the covariates explain some of the variation of the latent variables and thus also of the manifest variables.

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings of first latent variable						
1. *Health*	λ_{11}	0.906	0.024	0.874	0.905	0.937
2. *System*	λ_{21}	0.854	0.030	0.817	0.853	0.893
3. *Initiative*	λ_{31}	-0.926	0.026	-0.959	-0.925	-0.891
4. *Retirement*	λ_{41}	1.280	0.035	1.235	1.280	1.326
5. *Emergency*	λ_{51}	0.692	0.019	0.667	0.691	0.717
6. *Perform*	λ_{61}	-0.178	0.020	-0.203	-0.178	-0.153
7. *Society*	λ_{71}	0.070	0.022	0.042	0.070	0.099
8. *Reputation*	λ_{81}	0.145	0.020	0.119	0.145	0.170
9. *Salary*	λ_{91}	0.069	0.020	0.044	0.069	0.094
10. *Career*	$\lambda_{10,1}$	0.050	0.037	0.004	0.051	0.097
Factor loadings of second latent variable						
1. *Health*	λ_{12}	0.000	0.000	0.000	0.000	0.000
2. *System*	λ_{22}	0.264	0.026	0.231	0.264	0.297
3. *Initiative*	λ_{32}	-0.266	0.023	-0.295	-0.266	-0.236
4. *Retirement*	λ_{42}	-0.024	0.027	-0.058	-0.025	0.010
5. *Emergency*	λ_{52}	-0.027	0.019	-0.051	-0.027	-0.004
6. *Perform*	λ_{62}	-0.630	0.019	-0.654	-0.630	-0.606
7. *Society*	λ_{72}	-0.790	0.021	-0.817	-0.790	-0.763
8. *Reputation*	λ_{82}	-0.680	0.019	-0.704	-0.679	-0.656
9. *Salary*	λ_{92}	-0.644	0.019	-0.668	-0.644	-0.621
10. *Career*	$\lambda_{10,2}$	-1.570	0.052	-1.637	-1.569	-1.503
Parametric indirect effects of first latent variable						
*Inc*2	γ_{11}	0.150	0.042	0.097	0.150	0.202
*Inc*3	γ_{12}	0.477	0.044	0.421	0.477	0.533
*Inc*4	γ_{13}	0.941	0.051	0.876	0.942	1.007
Parametric indirect effects of second latent variable						
*Inc*2	γ_{21}	0.152	0.042	0.100	0.152	0.205
*Inc*3	γ_{22}	0.297	0.044	0.240	0.297	0.354
*Inc*4	γ_{23}	0.587	0.054	0.517	0.588	0.656

Table 3: *Estimates of factor loadings and parametric indirect effects.*

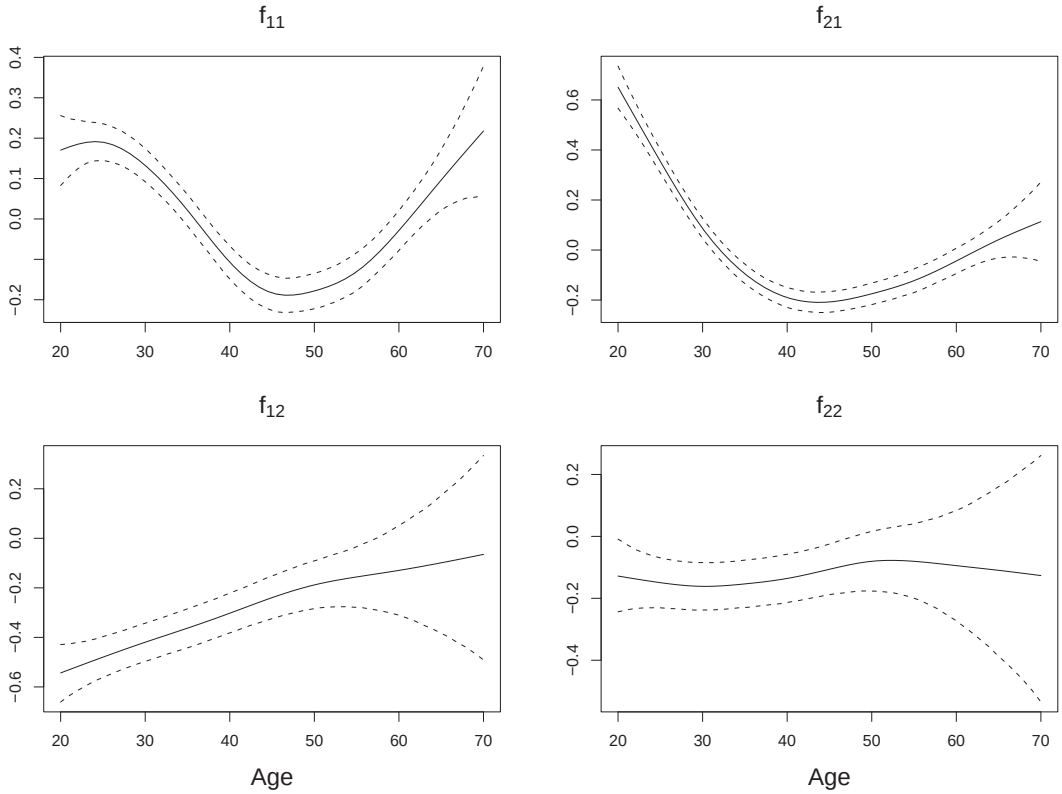

Figure 1: *Estimates of the smooth functions $f_1(Age)$ and $f_2(Age)$ of model 2 modelled by P-splines priors. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

Continuing the discussion with parametric effects, the covariate *Inc* exerts a very strong influence on the first latent construct concerning the attitude towards social coverage. With increasing income, the mean of the latent variable increases considerably, for example about 0.941 for a person in income category 4 compared to the reference category 1. This effect might be explained by the fact that people with high incomes generally show a higher inititiative of one's own and a higher readiness to take risks than people with lower incomes. In addition, big earners make high monetary contributions to the social system without getting an adequate service in return. The covariate *Inc* also has a similar effect on the second latent variable, although with lower absolute values. People with a higher ambition to have success will naturally have higher incomes.

The smooth function f_{11} of covariate *Age* shows a distinctive sinusoidal shape for the first latent variable. Given the same sex and income, young people prefer to take care of themselves, while medium aged people shortly before retirement seem to prefer a strong state. Moving to even older ages, respondents seem to become more progressive again. This result might be explained by the fact that young people suffer from the contributions to the social systems and do not get an appropriate service in return when they are old due to the ageing population. Additionally, young people are prepared to take bigger risks than older people. Medium-aged people around the age between 40 and 60 have to take care of their children, maybe pay mortgages on their property, save for their retirements, and hence would suffer tremendously

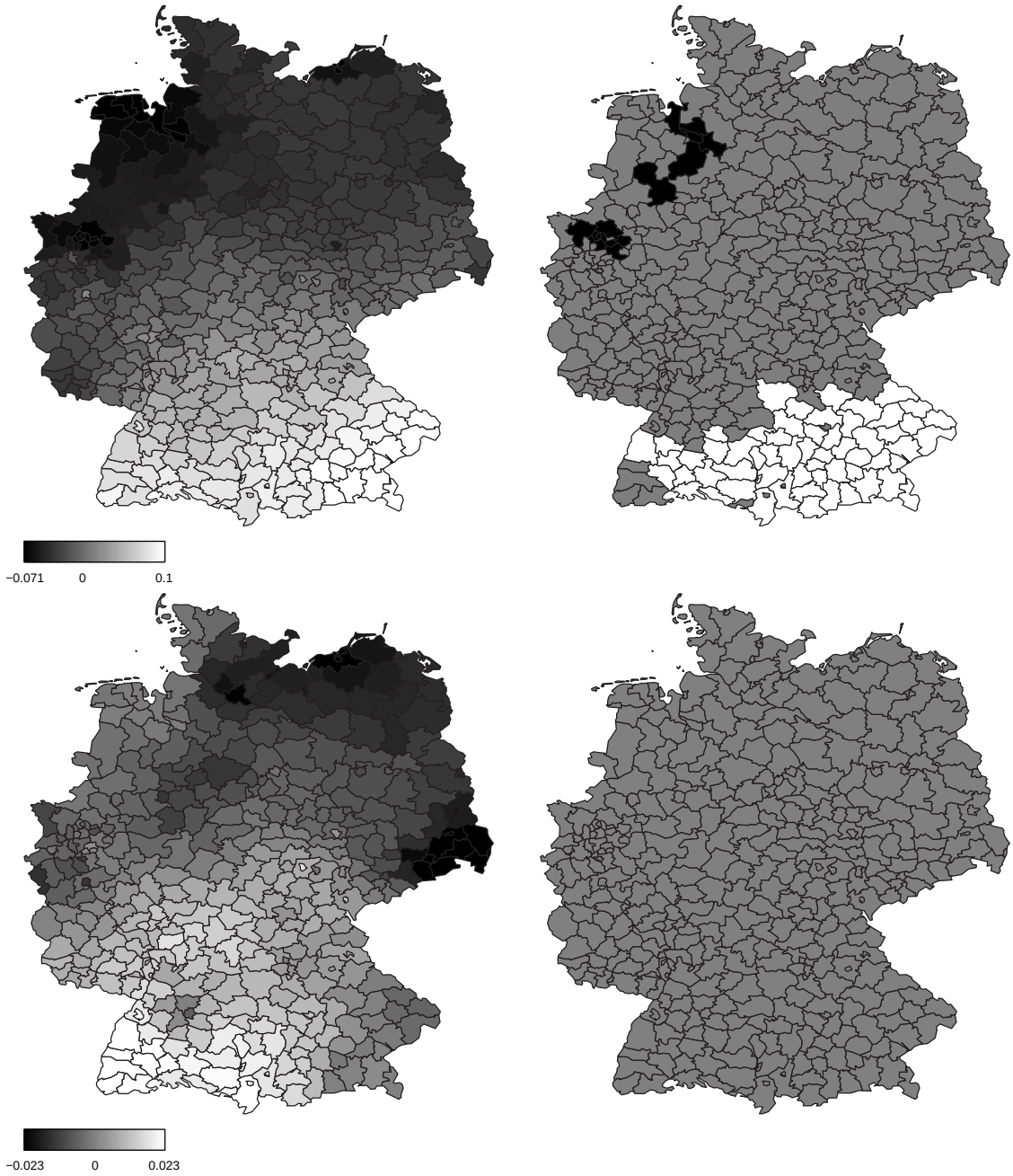

Figure 2: ***Left:*** *Function estimates $f_{1,spat}$ (top) and $f_{2,spat}$ (bottom) of the spatial covariate Reg. Minimum and maximum values are set to 2.5% and 97.5% quantiles of the observed function estimates, respectively.* ***Right:*** *Regions with a significant negative effect (black), a significant posivite effect (white), or a non-significant effect (grey) are plotted for $f_{1,spat}$ (top) and $f_{2,spat}$ (bottom).*

from a loss of job. Therefore those people would rather avoid risks and prefer a stronger state. This sinusoidal smooth function of the age effect also demonstrates that a big error is made if the covariate *Age* was only treated by a linear, parametric effect; we have carried out MCMC simulations where *Age* is included as a parametric linear effect which only recovers the almost linear drop between the age of 30 and 50 which is the range of age where most of the observations are located. The smooth function f_{21} of *Age* of the second latent variable shows an even more pronounced effect; young people seem to have a much higher ambition to be successful in job and society than older people which seems quite natural. This positive effect drops fast and approaches its minimum at the age of 45 and rises very slowly again for increasing age; obviously people at 45 are more concerned with other values, e. g. their family, than with success.

Having a short look at the smooth additional functional effect of females, we can see that the first latent variable seems to be more negative for females than for males (f_{12}). This negative effect is strongest for the youngest female participants, and rises slowly for increasing age. Women seem to prefer a social system where risks are hedged and the state transfers money from richer people to the less fortunate. For the second latent variable (f_{22}), however, there is only a very weak, constant and almost non-significant difference between males and females. It seems that the ambition to succeed in job and society is universal for both males and females. We see that the estimation of a VCM for the first latent variable provided additional insight, whereas a parametric effect for *Sex* would suffice for the second latent variable.

For the first latent variable, the spatial function estimate shows that the south has a small preference for a system based on the inititiative of one's own, and the inhabitants from north-west and eastern Germany would rather have a strong state caring for them. Please remark that this effect is significant for the north-west part, but not significant for eastern Germany. This coincides with the political landscape where the north-western area which was traditionally governed by the social-democratic party of Germany, and the south by the conservative party. The second latent variable only shows a very weak and non-significant dependancy on the geographical location.

This analysis of the real data set of "PD1" confirms that a nonparametric predictor influencing the latent variable can be estimated and that the results can deliver additional insight about the latent variable and its interrelationships with covariates. Furthermore, our model allows to make predictions based on indirect covariates for the expected latent scores even if the indicator response is not available for the new observation.

5 Conclusions

In this thesis we introduce a latent variabe model (LVM) for mixed continuous, binary, and ordinal response including covariates that can influence both the manifest and the latent variables. This thesis expands the existing models in the literature by using a semiparametric instead of a pure parametric predictor in the structural equation – this enables various types of covariates such as parametric effects, smooth functions of continuous and spatial covariates, and interactions of continuous and categorical variables (VCM) to influence the latent variables. This model allows the applied statistician to conduct much more detailed analyses

about the influence of covariates on the latent variables than currently available models.

Apart from applications in other disciplines, the following methodological research is of interest: model choice is an important issue for answering questions on variable selection, linear versus nonlinear effects, etc. We have experimented with the deviance information criterion (DIC), but a better understanding of DIC and other model selection methods in LVM requires additional research. Extensions to other types of observable responses such as count variables, non-negative continuous variables and survival times will be of relevance in various fields of application.

A MCMC algorithms

Two different MCMC algorithms are presented in this section; the first algorithm uses the work of Cowles (1996) for the sampling of cutpoint parameters, the second one is based on the paper of Liu and Sabatti (2000). We start the discussion with the sampling steps A–F that are identical for both algorithms – sampling step G differs for the two sampler types. Simulation is done with the statistical software R and the code uses functions provided by MCMCpack (2005).

A. Full conditionals of the underlying variables. For continuous indicators this step can be ommitted because there is no underlying variable y_{ij}^*, hence $y_{ij}^* = y_{ij}$. For ordinal indicators, the problem of sampling $\boldsymbol{y}^*$ simplifies to the sampling of the individual $\boldsymbol{y}_i^*$ because the observations are independently and identically distributed. Furthermore, the multivariate vector $\boldsymbol{y}_i^*$ can be obtained by sampling the individual components y_{ij}^* because the conditional covariance matrix $V(\boldsymbol{y}_i^*|\boldsymbol{\theta}, \boldsymbol{z}_i, \boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = \boldsymbol{\Sigma}$ is diagonal:

$$\begin{aligned} y_{ij}^*|\boldsymbol{z}_i, \boldsymbol{\theta} \setminus \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}, y_{ij}, \boldsymbol{w}_i \quad \sim \quad & N\left(\lambda_{j0} + \sum_{c=1}^{d} a_{jc} w_{ic} + \sum_{r=1}^{q} \lambda_{jr} z_{ir}, 1\right) \times \\ & \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \,. \end{aligned} \tag{10}$$

B. Full conditionals of the latent variables. The full conditional $p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) = p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$ springs from the joint distribution of $\boldsymbol{y}_i^*$ and $\boldsymbol{z}_i$ given $\boldsymbol{\theta}$, $\boldsymbol{x}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{w}_i$ in (9), and is a multivariate normal distribution with the expectation vector

$$E(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{y}_i^*, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = \boldsymbol{\eta}_i + \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{y}_i^* - \boldsymbol{\lambda}_0 - \boldsymbol{A}\boldsymbol{w}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\,, \tag{11}$$

and covariance matrix

$$V(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{y}_i^*, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = I_{q\times q} - \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1}\boldsymbol{B}\,, \tag{12}$$

where $\boldsymbol{\eta}_i$ denotes the predictor of the structural part of the model as defined in (4). A conditioning on $\boldsymbol{y}_i$ is not necessary since $\boldsymbol{y}_i$ is implicitly known through $\boldsymbol{y}_i^*$. Random samples from this multivariate density are generated by sampling from a multivariate standard normal density with dimension q, multiplying the result with the Cholesky matrix of the covariance matrix (12), and adding the expectation vector (11). Since the number of indicators p is typically much higher than the number of latent variables q, it is computationally more efficient to calculate the inversion of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$ by applying the matrix inversion lemma, i. e.

$$(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_{q\times q} + \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\,.$$

Now, only the $q \times q$ matrix $\boldsymbol{I}_{q\times q} + \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}$ has to be inverted instead of the $p \times p$ matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$, and the diagonal matrix $\boldsymbol{\Sigma}$ is inverted easily.

C. Full conditionals of the nonparametric indirect effects. The structural part of the model in equation (5) forms the basis of the full conditional for all indirect effects. Since the error variance matrix of the latent variables $\boldsymbol{z}_i$ is diagonal and priori information about the

nonparametric indirect effects is defined per latent variable and function, we can draw the parameter vectors $\boldsymbol{\beta}_{rh}$ of functions f_{rh} sequentially. The conditional expectation vector yields

$$E(\boldsymbol{\beta}_{rh}|\boldsymbol{\beta}_r \setminus \{\boldsymbol{\beta}_{rh}\}, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = \left(\boldsymbol{X}_h'\boldsymbol{X}_h + \frac{1}{\kappa_h}\boldsymbol{K}_h\right)^{-1} \boldsymbol{X}_h'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\,, \tag{13}$$

and the covariance matrix is given by

$$V(\boldsymbol{\beta}_{rh}|\boldsymbol{\beta}_r \setminus \{\boldsymbol{\beta}_{rh}\}, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = M_h^{-1} = \left(\boldsymbol{X}_h'\boldsymbol{X}_h + \frac{1}{\kappa_h}\boldsymbol{K}_h\right)^{-1}. \tag{14}$$

The term $\tilde{\boldsymbol{\eta}}_r$ contains the sum of all remaining parts of the predictor, i. e.

$$\tilde{\boldsymbol{\eta}}_r = X_1\boldsymbol{\beta}_{r1} + \ldots + X_{h-1}\boldsymbol{\beta}_{r,h-1} + X_{h+1}\boldsymbol{\beta}_{r,h+1} + \ldots + X_g\boldsymbol{\beta}_{rg} + X_{spat}\boldsymbol{\beta}_{r,spat} + \boldsymbol{U}\boldsymbol{\gamma}_r\,.$$

Although the parameter vector $\boldsymbol{\beta}_{rh}$ is drawn from a standard multivariate normal distribution, efficient sampling is not straightforward because linear equation systems with high-dimensional precision matrices M_h have to be solved in every iteration of the MCMC algorithm. We implemented an approach presented by Rue (2001) as described in the manual of BayesX (2005). Since the parameter vector is high-dimensional and often contains several hundred parameters, the solving of linear equations has be done in a very efficient way to limit the computational task. Therefore all penalty matrices $\boldsymbol{K}$ should be transformed into a band matrix like structure (e. g. using the reverse Cuthill-McKee algorithm) and all computational operations concerned with solving linear equations should be optimized for band matrices in an appropriate way.

After having sampled the parameter vector $\boldsymbol{\beta}_{rh}$, it is necessary to center the sampled parameter vector appropriately around zero because there is no intercept allowed in the structural part of the model due to identification restrictions. If there is no proper centering, different offsets will appear in all nonparametric functions in the predictor leading to highly fluctuating or even non-converging parameter estimates.

The smoothing parameter is a priori inverse Gamma distributed, short $\kappa_{rh} \sim IG(a_{rh}, b_{rh})$. Accordingly the full conditional distributions are also inverse Gamma distributed, i. e. $\kappa_{rh} \sim IG(a'_{rh}, b'_{rh})$ with $a'_{rh} = a_{rh} + \text{rank}(\boldsymbol{K}_h)/2$ and $b'_{rh} = b_{rh} + \boldsymbol{\beta}'_{rh}\boldsymbol{K}_{rh}\boldsymbol{\beta}_{rh}/2$.

D. Full conditionals of the parametric indirect effects. Since the priori information of the parametric indirect effects is also defined for each latent variable r, $\boldsymbol{\gamma}_r \sim N(\boldsymbol{\gamma}_r^*, \boldsymbol{\Gamma}_r^{*-1})$, we can again estimate the regression parameters $\boldsymbol{\gamma}_r$ sequentially due to the diagonal covariance matrix of $\boldsymbol{\xi}_i$. Ergo for each latent variable r, we obtain a linear model with Gaussian response and the full conditional distribution for the parametric indirect effects yields the expectation vector

$$E(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{\Gamma}_r^* + \boldsymbol{U}'\boldsymbol{U})^{-1}\left(\boldsymbol{\Gamma}_r^*\boldsymbol{\gamma}_r^* + \boldsymbol{U}'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\right)\,,$$

and the covariance matrix

$$V(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{\Gamma}_r^* + \boldsymbol{U}'\boldsymbol{U})^{-1}\,,$$

with the $n \times m$-dimensional design matrix $\boldsymbol{U}$ defined in the usual way, containing the covariates u_{il}. Again the term $\tilde{\boldsymbol{\eta}}_r$ contains the sum of all values of the remaining parts of the

predictor. We always use diffuse priori distributions for parametric indirect effects leading to the following simplified expressions

$$\begin{aligned} E(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) =& (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\,, \qquad \text{and} \\ V(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) =& (\boldsymbol{U}'\boldsymbol{U})^{-1}\,. \end{aligned}$$

E. Full conditionals of the intercepts, direct effects and factor loadings. The vector $\bar{\boldsymbol{\lambda}}$ contains all parameters $\boldsymbol{\lambda}_0$, $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$. For an informative prior distribution $N(\bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{\Lambda}}^{*-1})$, the full conditional $p(\bar{\boldsymbol{\lambda}}|\boldsymbol{\theta} \setminus \{\bar{\boldsymbol{\lambda}}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) = p(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w})$ is a $p \cdot (1+d+q)$ dimensional multivariate normal distribution with expectation vector

$$E(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w}) = \left(\boldsymbol{\Delta}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_i\right)^{-1} \left(\bar{\boldsymbol{\Lambda}}^*\bar{\boldsymbol{\lambda}}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_i^*\right),$$

and covariance matrix

$$V(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w}) = \left(\boldsymbol{\Delta}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_i\right)^{-1}.$$

The $p \times p(1 + d + q)$ dimensional matrix $\boldsymbol{L}_i$ is defined as $\boldsymbol{L}_i = \boldsymbol{I}_{p\times p} \otimes \boldsymbol{l}_i$ with $\boldsymbol{l}_i = (1, w_{i1}, \ldots, w_{id}, z_{i1}, \ldots, z_{iq})$.

For diffuse priors, priors with a diagonal prior precision matrix $\bar{\boldsymbol{\Lambda}}^*$, and priors with a precision matrix $\bar{\boldsymbol{\Lambda}}^*$ with off-diagonal zero-entries for parameters for different indicators, the full conditional can be calculated sequentially for each indicator because $V(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$ is diagonal. Let the vector $\bar{\boldsymbol{\lambda}}^j$ contain the parameters of row j, short $\bar{\boldsymbol{\lambda}}^j = (\lambda_{j0}, a_{j1}, \ldots, a_{jd}, \lambda_{j1}, \ldots, \lambda_{jm})'$. Ergo the full conditional distribution of $\bar{\boldsymbol{\lambda}}^j$ is a $(1 + m + q)$-dimensional multivariate normal distribution with expectation vector

$$E(\bar{\boldsymbol{\lambda}}^j|\sigma_j^2, \boldsymbol{y}_j^*, \boldsymbol{z}, \boldsymbol{w}) = (\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\boldsymbol{y}_j^*\,,$$

and covariance matrix

$$V(\bar{\boldsymbol{\lambda}}^j|\sigma_j^2, \boldsymbol{y}_j^*, \boldsymbol{z}, \boldsymbol{w}) = \sigma_j^2(\boldsymbol{L}'\boldsymbol{L})^{-1}\,.$$

The $n \times (1 + d + q)$ dimensional matrix $\boldsymbol{L}$ is defined by $\boldsymbol{L} = (\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n)'$ with the rows $\boldsymbol{l}_i = (1, w_{i1}, \ldots, w_{id}, z_{i1}, \ldots, z_{iq})$. For our choice of priors, we can always use the second way of sampling the parameters which provides faster computation.

Here it becomes clear why we fixed the error variances for ordinal indicators to 1. In order to obtain a standardized solution, we had to sample the factor loadings under the restriction $\sum_{r=1}^{q} \lambda_{jr}^2 + \theta^2 = 1$. Since this can not be achieved by a standard full conditional, we fix the error variances of ordinal manifest variables to 1 instead. A standardization is easily possible after the simulation run.

F. Full conditionals of the error variances. For ordinal variables, the sampling of error variances is ommitted because error variances are set to 1 due to identification restrictions of the model. Since error variances are distributed normally with a diagonal error variance

matrix $\boldsymbol{\Sigma}$, the full conditional $p(\sigma_j^2|\boldsymbol{\theta}\backslash\{\boldsymbol{\sigma}^2\}, \boldsymbol{y}_i^*, \boldsymbol{z}, \boldsymbol{w})$ results to an inverse gamma distribution with $n+\nu$ degrees of freedom and scale parameter s^2, i. e.

$$\sigma_j^2|\boldsymbol{\theta} \setminus \{\boldsymbol{\sigma}^2\}, \boldsymbol{y}_i^*, \boldsymbol{z}, \boldsymbol{w} \quad \sim \quad IG\left(\frac{n+\nu}{2}, \frac{(n+\nu)S^2}{2}\right), \tag{15}$$

with

$$S^2 = \frac{1}{n+\nu}\left(\sum_{i=1}^{n}\left(y_{ij}^* - \lambda_{j0} - \sum_{c=1}^{d} a_{jc}w_{ic} - \sum_{r=1}^{q}\lambda_{jr}z_{ir}\right)^2 + \nu s^2\right), \tag{16}$$

where σ_j^2 is a priori $IG(\nu/2, \nu s^2/2)$ distributed. In the case of a noninformative prior distribution for σ_j^2, the full conditional distribution is obtained by setting the values of ν and s in (15) and (16) to zero.

So far, all sampling steps are identical for the MHS and the GGS – the final sampling step differs for both sampler types.

G1. MHS – full conditionals of the cutpoints. Cowles (1996) proposed the following Metropolis-Hastings step (also described in Johnson and Albert, 1999) for the cutpoints in order to improve convergence compared to the method of Albert and Chib (1993):

Repeat for $j = 1, \ldots, p_1$

- *Draw a set of proposal cutpoints* $\tilde{\boldsymbol{\tau}}$
  For $k = 2, \ldots, K_j - 1$, sample $\tilde{\tau}_k \sim N(\tau_{jk}^{(t-1)}, \sigma_{MH})$ truncated to the interval $[\tilde{\tau}_{k-1}, \tau_{j,k+1}^{(t-1)}]$.

- *Compute the acceptance ratio* R

$$R = \prod_{i=1}^{n} \frac{\boldsymbol{\Phi}(\tilde{\tau}_{y_i} - \mu_{ij}^{(t-1)}) - \boldsymbol{\Phi}(\tilde{\tau}_{y_i-1} - \mu_{ij}^{(t-1)})}{\boldsymbol{\Phi}(\tau_{y_i}^{(t-1)} - \mu_{ij}^{(t-1)}) - \boldsymbol{\Phi}(\tau_{y_i-1}^{(t-1)} - \mu_{ij}^{(t-1)})} \times \prod_{k=2}^{K_j-1} \frac{\boldsymbol{\Phi}((\tau_{k+1}^{(t-1)} - \tau_k^{(t-1)})/\sigma_{MH}) - \boldsymbol{\Phi}((\tilde{\tau}_{k-1} - \tau_k^{(t-1)})/\sigma_{MH})}{\boldsymbol{\Phi}((\tilde{\tau}_{k+1} - \tilde{\tau}_k)/\sigma_{MH}) - \boldsymbol{\Phi}((\tau_{k-1}^{(t-1)} - \tilde{\tau}_k)/\sigma_{MH})}$$

  $\tilde{\tau}_{y_i}$ denotes the cutpoint proposal corresponding to the ordinal value of observation y_i, similarly $\tau_{y_i}^{(t-1)}$ is the actual value of the cutpoint corresponding to the observed ordinal category at iteration $t-1$. The term $\mu_i^{(t-1)}$ denotes the value of the linear predictor in the measurement model at iteration $t-1$ for observation i and indicator j, short $\mu_i^{(t-1)} = \lambda_{j0}^{(t-1)} + \sum_{c=1}^{d} a_{jc}^{(t-1)} w_{ic} + \sum_{r=1}^{q} \lambda_{jr}^{(t-1)} z_{ir}^{(t-1)}$.

- *Accept or reject proposal value* $\tilde{\boldsymbol{\tau}}$
  Set $\boldsymbol{\tau}_j^{(t)} = \tilde{\boldsymbol{\tau}}$ with probability R, otherwise set $\boldsymbol{\tau}_j^{(t)} = \boldsymbol{\tau}_j^{(t-1)}$.

The value σ_{MH} can be considered a tuning parameter and has to be set by the researcher before starting the simulation. A rule of thumb recommends setting $\sigma_{MH} = 0.05/K_j$ which should lead to acceptance ratios of 25-50%. If necessary, a different σ_{MH}^j for each indicator could be employed to achieve proper acceptance ratios for all indicators j, for example if

the number of ordinal categories vary across indicators. This popular standard algorithm of sampling the cutpoints still inhibits some drawbacks. Convergence is still not optimal, a MH step has to be employed with the serious drawback of setting and adjusting the tuning parameter σ_{MH}, and the calculation of the acceptance ratio is computationally demanding, especially in an analysis involving many ordinal indicators.

G2. GGS – full conditionals of the cutpoints and parameter transformation. For ordinal indicators, the full conditionals of τ_{jk} $(2 \leq k \leq K_j - 1)$ given $\boldsymbol{y}_j$, $\boldsymbol{y}_j^*$ and $\boldsymbol{\theta} \setminus \{\tau_{jk}\}$ are uniform distributions in the interval $[l_{\tau k}, r_{\tau k}]$, with the interval borders (see Albert and Chib, 1993)

$$\begin{aligned} l_{\tau,k} :=& \max\left\{\tau_{j,k-1}, \max_{i=1,\ldots,n}\{y_{ij}^* | y_{ij} = k\}\right\}, \\ r_{\tau,k} :=& \min\left\{\tau_{j,k+1}, \min_{i=1,\ldots,n}\{y_{ij}^* | y_{ij} = k+1\}\right\}, \end{aligned}$$

where $\max(\emptyset) = -\infty$ and $\min(\emptyset) = \infty$. Now it becomes very clear why this sampling algorithm exhibits bad convergence properties. The cutpoints τ_{jk} have almost no room to move in the small interval $[l_{\tau k}, r_{\tau k}]$, especially for a moderate or high number of observations. This also leads to a poor convergence of some other parameters of the model, especially the intercepts $\boldsymbol{\lambda}_0$. In order to receive even better convergence properties than the MHS and compensate for its drawbacks, a Generalized Gibbs move is carried out for some parameters based on the work of Liu and Sabatti (2000). The difficulty is to find a suitable transformation group Γ so that the resulting distribution allows a fast sampling of the transformation members γ (not related to the regression coefficients of the structural model). Since a Generalized Gibbs move for the whole posterior can not be deducted, we develop a Generalized Gibbs move for each of the p_1 linear submodels in the measurement model for all ordinal indicators. This is possible due to the diagonal form of the error variance matrix $\boldsymbol{\Sigma}$. We identified the partial scale group on $\boldsymbol{S}$ to be a suitable transformation group:

$$\Gamma_v := \{\gamma > 0 : \gamma(\boldsymbol{\theta}) = (\gamma\theta_1, \ldots, \gamma\theta_v, \theta_{v+1}, \ldots, \theta_{dim})\}.$$

Here only v components are transformed, the others remain fixed. The left-Haar measure for this group is $\gamma^{-1}d\gamma$ as for the total scale group. The determinant of the Jacobian is $\det(\partial\gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) = \gamma^v$. This yields $\pi(\gamma(\boldsymbol{\theta}))\,|J_\gamma(\boldsymbol{\theta})|\,L(d\gamma) = \gamma^{v-1}\pi(\gamma(\boldsymbol{\theta}))\,d\gamma$. Now we specify suitable subsets $\boldsymbol{\theta}_j$ of the total parameter vector $\boldsymbol{\theta}$, so that we can transform these subsets' parameters for each indicator j. Accordingly we define the p_1 parameter vectors $\boldsymbol{\theta}_j$ per ordinal indicator j as $\boldsymbol{\theta}_j = (y_{j1}^*, \ldots, y_{jn}^*, \lambda_{j0}, a_{j1}, \ldots, a_{jd}, \lambda_{j1}, \ldots, \lambda_{jm}, \tau_{j2}, \ldots, \tau_{jK_{j-1}})$ each of which contains $v = n + d + m + K_j - 1$ parameters. Thus we get p_1 different Generalized Gibbs moves

$$\gamma_j(\boldsymbol{\theta}_j) = (\gamma y_{j1}^*, \ldots, \gamma y_{jn}^*, \gamma\lambda_{j0}, \gamma a_{j1}, \ldots, \gamma a_{jd}, \gamma\lambda_{j1}, \ldots, \gamma\lambda_{jm}, \gamma\tau_{j2}, \ldots, \gamma\tau_{jK_{j-1}}),$$

that transform the corresponding parameter set $\boldsymbol{\theta}_j$. All other parameters of the full parameter vector remain constant and are not transformed. The individual parameter sets $\boldsymbol{\theta}_j$ can be

derived by arranging the posterior distribution in the following way:

$$
\begin{aligned}
& p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[p(\boldsymbol{y}_i^*, \boldsymbol{z}_i | \boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \left(\prod_{j=1}^{p_1} \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \right) \right] \\
=& p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[p(\boldsymbol{z}_i | \boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \cdot p(\boldsymbol{y}_i^* | \boldsymbol{\theta}, \boldsymbol{z}_i, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \left(\prod_{j=1}^{p_1} \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \right) \right] \\
=& p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[p(\boldsymbol{z}_i | \boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \prod_{j=p_1+1}^{p} p(y_{ij}^* | \boldsymbol{z}_i, \boldsymbol{\theta}_j \setminus \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}, \boldsymbol{w}_i) \right. \\
& \qquad \left. \times \prod_{j=1}^{p_1} \left(p(y_{ij}^* | \boldsymbol{z}_i, \boldsymbol{\theta}_j \setminus \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}, \boldsymbol{w}_i) \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \right) \right] \\
=& p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(\boldsymbol{z}_i | \boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \prod_{i=1}^{n} \prod_{j=p_1+1}^{p} p(y_{ij}^* | \boldsymbol{z}_i, \boldsymbol{\theta}_j \backslash \{\Gamma\}, \boldsymbol{w}_i) \\
& \qquad \times \prod_{j=1}^{p_1} \underbrace{\left[\prod_{i=1}^{n} \Big(p(y_{ij}^* | \boldsymbol{z}_i, \boldsymbol{\theta}_j \backslash \{\Gamma\}, \boldsymbol{w}_i) \Big) \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y_{ij}^* \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \right]}_{} .
\end{aligned}
$$

Thus we deploy p_1 distinct Generalized Gibbs moves for the p_1 components of the posterior distribution which are underbraced in the above formula. Based on the underbraced part of the posterior, we can formulate the densities $\gamma_j^{v-1} \pi(\gamma_j \boldsymbol{\theta}_j | \cdot)$ to be proportional to

$$
\begin{aligned}
& \gamma_j^{v-1} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} \left(\gamma_j y_{ij}^* - \gamma_j \lambda_{j0} - \sum_{c=1}^{d} \gamma_j a_{jc} w_{ic} - \sum_{r=1}^{q} \gamma_j \lambda_{jr} z_{ir} \right)^2 \right\} \\
&= \left(\gamma_j^2\right)^{\frac{v-1}{2}} \exp \left\{ -\frac{1}{2} \gamma_j^2 \sum_{i=1}^{n} \left(y_{ij}^* - \lambda_{j0} - \sum_{c=1}^{d} a_{jc} w_{ic} - \sum_{r=1}^{q} \lambda_{jr} z_{ir} \right)^2 \right\} .
\end{aligned}
\tag{17}
$$

We dropped the right hand side of the underbraced formula because the right term remains constant under the transformation according to

$$
\sum_{k=1}^{K_j} \mathbb{1}_{\gamma_j \tau_{j,k-1} < \gamma_j \bar{y}_{ij} \leq \gamma_j \tau_{jk}} \mathbb{1}_{y_{ij}=k} \quad \Longleftrightarrow \quad \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < \bar{y}_{ij} \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k} \, .
$$

It follows from equation (17) that γ_j^2 follows a Gamma distribution $\Gamma(a_j, b_j)$ with parameters

$$
\begin{aligned}
a_j &= \frac{v+1}{2} = \frac{n+d+m+K_j}{2} , \\
b_j &= \frac{\sum_{i=1}^{N} \left(y_{ij}^* - \lambda_{j0} - \sum_{c=1}^{d} a_{jc} w_{ic} - \sum_{r=1}^{q} \lambda_{jr} z_{ir} \right)}{2} ,
\end{aligned}
\tag{18}
$$

and the density of $\Gamma(a, b)$ is given by $f(x|a,b) = b^a x^{a-1} e^{-bx} / \Gamma(a)$ for $x \geq 0$. After having sampled the cutpoints from the uniform distributions as described above, p_1 Generalized Gibbs moves ($j = 1, \ldots, p_1$) are carried out in the following way:

- Draw γ_j^2 from $\Gamma(a_j, b_j)$ with a_j and b_j as defined in equations (18), and update all parameters of the subset $\boldsymbol{\theta}_j$ in the following way:

$$
\begin{aligned}
y_{\cdot j}^{* \, new} &\longleftarrow \gamma_j \, y_{\cdot j}^{* \, current} \\
\lambda_{j\cdot}^{new} &\longleftarrow \gamma_j \, \lambda_{j\cdot}^{current} \\
a_{j\cdot}^{new} &\longleftarrow \gamma_j \, a_{j\cdot}^{current} \\
\tau_{j\cdot}^{new} &\longleftarrow \gamma_j \, \tau_{j\cdot}^{current} \, .
\end{aligned}
$$

For both samplers, starting values are as follows: intercepts $\boldsymbol{\lambda}_0$ and regression parameters of direct effects $\boldsymbol{A}$ are set to zero; the first free factor loading parameter of each latent variable is set to 0.7 to promote a positive solution for those parameters, all other factor loadings are also set to zero; variances θ_j^2 for continuous indicators start at the value 1; the nonparametric function parameters $\boldsymbol{\beta}$ and standard regression coefficients $\boldsymbol{\gamma}$ start at zero; cutpoints τ_{jk} are set to be $1, \ldots, K_j - 2$ for $k = 2, \ldots, K_j - 1$. This choice of starting values implies zero values for the latent variables $\boldsymbol{z}$, while specifying starting values for the underlying variables y^* is not necessary because they are sampled in the first iteration. We performed a sensitivity analysis and tested variations of different starting values for all involved parameters. The results show that starting values do not affect the resulting parameter estimates, as long as we refrain from using implausible and far-fetched starting values (e. g. factor loadings higher than 5).

References

Albert, J. H., Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Arminger, G., Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271–300.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. Charles Griffin & Company Ltd., London, Oxford University Press, New York. 2nd edition: Bartholomew, D. J., Knott, M. (1999). Arnold, London.

BayesX (2005). *Software for Bayesian inference* by Brezger, A., Kneib, T., Lang, S. URL: http://www.stat.uni-muenchen.de/~bayesx/bayesx.html

Besag, J. E., York, J., Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Besag, J. E., Green, P., Higdon, D., Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

Besag, J. E., Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

Brezger, A., Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* **50**, 967–991.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**, 101–111.

Eilers, P. H. C., Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with comments). *Statistical Science* **11**, 89–121.

Fahrmeir, L., Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer, New York.

Fahrmeir, L., Lang, S. (2001b). Bayesian semiparametric regression analysis of multicategorical time-space data. *Annals of the Institute of Statistical Mathematics* **53**, 10–30.

Fahrmeir, L., Kneib, T., Lang, S. (2004). Penalized structured additive regression of space-time data: a Bayesian perspective. *Statistica Sinica* **14**, 731–761.

Gamerman, D. (1997). Efficient sampling from the posterior distribution in generalized linear models. *Statistics and Computing* **7**, 57–68.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). *Bayesian data analysis* (2nd edition). Chapman and Hall.

Hastie, T., Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall.

Hastie, T., Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B* **55**, 757–796.

Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society A* **134**, 486–501.

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.

Johnson, V. E., Albert, J. H. (1999). *Ordinal data modeling*. Springer, New York.

Jöreskog, K. G., Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* **70**, 631–639.

Kammann, E. E., Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society C* **52**, 1–18.

Lang, S., Fronk, E.-M., Fahrmeir, L. (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics* **17**, 479–500.

Lang, S., Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183–212.

Lawley, D. N., Maxwell, A. E. (1963). *Factor analysis as a statistical method*. 2nd edition (1971). Butterworths, London.

Liu, J. S., Sabatti, C. (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika* **87**, 353–369.

Lopes, H. F., West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67.

MCMCPack (2005). *R-package for Markov chain Monte Carlo* by Martin, A. D., Quinn, K. M. URL: http://mcmcpack.wustl.edu/

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* **56**, 337–357.

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* **11**, 487–513.

Moustaki, I., Steele, F. (2005). Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modeling* **5**, 327–342.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54**, 557–585.

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* **12**, 338–353.

Raach, A. W. (2005). *A Bayesian semiparametric latent variable model for binary, ordinal and continuous response.* Dissertation, to appear.

Rao, M. M. (1987). *Measure theory and integration.* John Wiley, New York.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society B* **63**, 325–338.

Rue, H., Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications.* Chapman and Hall.

Rupp, A. A., Dey, D. K., Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: applications of bayesian methodology to modeling. *Structural Equation Modeling* **11**, 424–451.

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.

Seber, G. A. F. (1984). *Multivariate observations.* John Wiley, New York.

Skrondal, A., Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* Chapman and Hall.

Sun, D., Tsutakawa, R. K., He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica* **11**, 77–95.

Tanner, M. A., Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Zhu, J., Eickhoff, J. C., Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**, 674–683.