Fahrmeir, Ludwig; Steinert, Sven

Working Paper

A geoadditive Bayesian latent variable model for Poisson indicators

Discussion Paper, No. 508

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Fahrmeir, Ludwig; Steinert, Sven (2006) : A geoadditive Bayesian latent variable model for Poisson indicators, Discussion Paper, No. 508, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, https://doi.org/10.5282/ubm/epub.1877

This Version is available at:
https://hdl.handle.net/10419/31095

A geoadditive Bayesian Latent Variable Model for Poisson indicators

Ludwig Fahrmeir and Sven Steinert

Department of Statistics, University of Munich

December, 2006

Abstract

We introduce a new latent variable model with count variable indicators, where usual linear parametric effects of covariates, nonparametric effects of continuous covariates and spatial effects on the continuous latent variables are modelled through a geoadditive predictor. Bayesian modelling of nonparametric functions and spatial effects is based on penalized spline and Markov random field priors. Full Bayesian inference is performed via an auxiliary variable Gibbs sampling technique, using a recent suggestion of Frühwirth-Schnatter and Wagner (2006). As an advantage, our Poisson indicator latent variable model can be combined with semiparametric latent variable models for mixed binary, ordinal and continuous indicator variables within an unified and coherent framework for modelling and inference. A simulation study investigates performance, and an application to post war human security in Cambodia illustrates the approach.

Keywords: Latent variable models, Poisson indicators, penalized splines, spatial effects, MCMC.

1 Introduction

In this paper, we introduce a flexible geoadditive latent variable model (LVM), where observed or manifest indicators are count variables. Conditional on common latent variables and possibly, on some covariates, we assume as measurement model that the indicator variables follow a log-linear Poisson model, extending the usual linear predictor constructed from covariates through a linear combination of latent variables and factor loadings. In this way, correlation between the indicators is also automatically accounted for. The effects of further covariates of different type on the latent variables are modelled through a geoadditive predictor, extending the usual linear predictor by adding nonparametric functions for possibly nonlinear effects of continuous covariates, and spatial effects resulting from geographical small area information about the location of units or residence of individuals in the sample. Covariates of this type are present in Section 5, where we illustrate our approach to a latent variable model with count variable indicators for post war human security in Cambodia. This application is motivated by the study in Benini, Owen and Rue (2006) where separate independent geoadditive Poisson regressions are applied to the same indicators. In contrast, our latent variable model automatically accounts for correlation of indicator through a common latent factor.

We develop full Bayesian inference for parameters, functions and spatial effects as well as for the latent variable, using an underlying variable approach (UVA) to facilitate simulation based posterior analysis via Gibbs sampling. Following a recent proposal by Frühwirth-Schnatter and Wagner (2006) in the context of state space models for count variables, we generate auxiliary, unobservable Gaussian variables from the observable indicators. Based on the resulting auxiliary measurement model, Gibbs sampling can be performed along the lines of the sampling scheme proposed in Fahrmeir and Raach (2006) and Raach (2005) for geoadditive latent variable models with mixed binary, ordered categorical and

continuous indicators. Moreover, the LVM for count data presented here can therefore be combined with the latter class of models to flexible semiparametric latent variable models for mixed categorical, continuous and count variable indicators within a unified and coherent framework.

In comparison, LVM presented in the literature so far mostly assume that the effects of covariates on both the observable indicators and the latent variables are modelled in simple linear parametric form, see Skrondal and Rabe-Hesketh (2004) for a recent comprehensive introduction. The origin of the LVM with covariate effects can be traced back to the MIMIC model of Jöreskog and Goldberger (1975). 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. Notable exceptions are nonlinear latent variable models suggested by Arminger and Muthén (1998), Lee and Song (2004), and Song and Lee (2005), but the nonlinear relationship is still of conventional parametric form. The semiparametrically structured geoadditive predictor used in our LVM is described in Fahrmeir, Kneib and Lang (2004), and Brezger and Lang (2006) in the simpler context of semiparametric generalized regression for univariate responses.

This paper is organized as follows: Section 2 presents the measurement model for the observable indicators as well as the corresponding auxiliary Gaussian measurement model, and the geoadditive structural model for the latent variables. Section 3 outlines the Gibbs sampling scheme for Bayesian inference. Section 4 investigates performance in a simulation

study, and Section 5 illustrates the approach by a real data application to a study on post war security in Cambodia.

2 Statistical model

2.1 Measurement model

In our LVM all indicators or manifest variables y_j, $j = 1, \ldots p$, are count data, that means nonnegative integers. Let y_{ij} denote the observed value of indicator y_j, $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{id})'$ a vector of covariates, and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})$, $q < p$, a vector of latent variables for individual i, $i = 1, \ldots, n$. Conditional on covariates and latent variables we assume a log-linear Poisson model

$$y_{ij}|\mu_{ij}, \cdot \sim Po(\mu_{ij})\,,\ \mu_{ij} = \exp\left(\lambda_{j0} + \boldsymbol{\lambda}_j'\boldsymbol{z}_i + \boldsymbol{\alpha}_j'\boldsymbol{\omega}_i\right)\,,\ i = 1, \ldots, n\,, j = 1, \ldots, p. \quad (1)$$

In (1) λ_{j0} is an intercept term, the q-dimensional vector $\boldsymbol{\lambda}_j' = (\lambda_{j1}, \ldots, \lambda_{jq})$ consists of the factor loadings indicating the strength of relationship between latent and manifest variables, and $\boldsymbol{\alpha}_j' = (\alpha_{j1}, \ldots, \alpha_{jd})$ is the vector of direct effects of covariates on y_{ij}. Note that (1) extends the usual linear predictor $\lambda_{j0} + \boldsymbol{\alpha}_j'\boldsymbol{\omega}_i$ of log-linear Poisson models by incorporating the linear effect $\boldsymbol{\lambda}_j'\boldsymbol{z}_i$ of latent variables.

In analogy to latent variable models with binary and ordinal indicators, our concept for Bayesian modelling and inference is based on an underlying variable approach (UVA) for auxiliary Gaussian variables. This facilitates full Bayesian inference via Gibbs sampling, and it allows us to combine geoadditive latent variable models developed in Fahrmeir and Raach (2006) for binary, ordinal and continuous indicators with models for count indicators

considered here. Following a recent suggestion of Frühwirth-Schnatter and Wagner (2006) in the context of state space models for count data, the introduction of two so called data augmentation steps eliminates the nonlinearity of the Poisson model as well as the nonnormality of the error term. The (conditional) distribution of $y_{ij} \,|\, \mu_{ij}, \cdot$ is considered as the distribution of the number of jumps of an unobserved Poisson process in the time interval $[0, 1]$. The first data augmentation step introduces the interarrival times τ_{ijl}, $i = 1, \ldots, n$, $j = 1, \ldots, p$, $l = 1, \ldots, y_{ij} + 1$, of this unobserved Poisson process. Because of the properties of a Poisson process, we know that these interarrival times τ_{ijl} follow an exponential distribution with parameter μ_{ij}, this means $\tau_{ijl} \,|\, \cdot \sim Exp(\mu_{ij}) = Exp(1)/\mu_{ij}$. Taking logarithms we obtain the linear model

$$-\log \tau_{ijl} \,|\, \cdot = \lambda_{j0} + \boldsymbol{\lambda}_j' \boldsymbol{z}_i + \boldsymbol{\alpha}_j' \boldsymbol{\omega}_i + \varepsilon_{ijl}, \quad \varepsilon_{ijl} \sim \log Exp(1).$$

This way the nonlinearity of the Poisson model is eliminated, but the nonnormality of the error term ε_{ijl} still remains. The density $f(\varepsilon_{ijl})$ of the error term is independent of any unknown parameter:

$$f(\varepsilon_{ijl}) = \exp\{\varepsilon_{ijl} - \exp(\varepsilon_{ijl})\}.$$

According to Chib et al. (2002) who approximate the density of a $\log \chi^2$-distribution by a normal mixture, Frühwirth-Schnatter and Wagner (2006) approximate the density of the $\log Exp(1)$-distribution of the error term ε_{ijl} by a mixture of ten normal distributions to obtain a conditionally Gaussian model

$$f(\varepsilon_{ijl}) \approx \sum_{r=1}^{10} w_r \, f_N\left(\varepsilon_{ijl}; m_r, \sigma_r^2\right). \tag{2}$$

To achieve a satisfactory approximation quality the parameters (weights w_r, means m_r and variances σ_r^2) were calculated by minimizing the Kullback-Leibler distance. The cor-

responding values of the weights, means and variances can be found in Table 1.

Table 1: *Parameter values for the normal mixture approximation of the log Exp(1)-distribution*

r	1	2	3	4	5	6	7	8	9	10
ω_r	0.00397	0.0396	0.168	0.147	0.125	0.101	0.104	0.116	0.107	0.088
m_r	-5.09	-3.29	-1.82	-1.24	-0.764	-0.391	-0.0431	0.306	0.673	1.06
σ_r^2	4.5	2.02	1.1	0.422	0.198	0.107	0.0788	0.0766	0.0947	0.146

To demonstrate the quality of the approximation we plot the real as well as the approximated density of this distribution in Figure 1a. The difference of these two densities can be seen in Figure 1b.

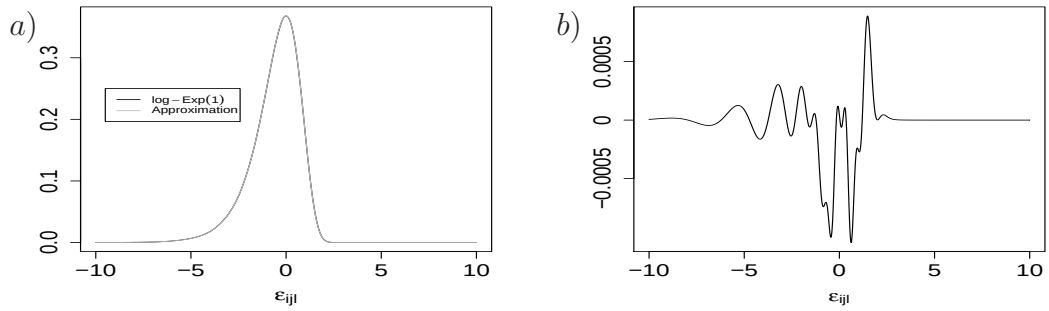

Figure 1: *a) Density of the log Exp(1)-distribution and the density of the mixture distribution, b) Difference between log Exp(1)-distribution and approximation*

The second data augmentation step introduces for every ε_{ijl} the so called component indicator as another unobserved magnitude. Conditional on these component indicators we obtain a Gaussian distribution instead of the Poisson-distribution in (1):

$$-\log \tau_{ijl} \,|\, \cdot, r_{ijl} = \lambda_{j0} + \boldsymbol{\lambda}_j' \boldsymbol{z}_i + \boldsymbol{\alpha}_j' \boldsymbol{\omega}_i - m_{r_{ijl}} + \zeta_{ijl}, \tag{3}$$

where the error term $\zeta_{ijl} \,|\, r_{ijl}$ follows a $N\left(0, \sigma^2_{r_{ijl}}\right)$-distribution. Thus we get:

$$-\log \tau_{ijl} \,|\, \cdot, r_{ijl} \sim N\left(\lambda_{j0} + \boldsymbol{\lambda}_j' \boldsymbol{z}_i + \boldsymbol{\alpha}_j' \boldsymbol{\omega}_i - m_{r_{ijl}},\, \sigma^2_{r_{ijl}}\right). \tag{4}$$

After eliminating the nonnormality of the error term by the second data augmentation

step we define the underlying (or auxiliary) variables y^*_{ijl} as follows:

$$y^*_{ijl} = -\log\left(\tau_{ijl} \,|\, \cdot, r_{ijl}\right) + m_{r_{ijl}}, \quad l = 1, \ldots, y_{ij} + 1.$$

This means, for every y_{ij} we obtain $y_{ij} + 1$ auxiliary Gaussian measurement models

$$y^*_{ijl} \,|\, \cdot, r_{ijl} = \lambda_{j0} + \boldsymbol{\lambda}'_j \boldsymbol{z}_i + \boldsymbol{\alpha}'_j \boldsymbol{\omega}_i + \zeta_{ijl}\,, \tag{5}$$

with $\zeta_{ijl} \,|\, r_{ijl} \sim N\left(0, \sigma^2_{r_{ijl}}\right)$ and $l = 1, \ldots, y_{ij}+1$. Obviously we obtain $y_{ij}+1$ measurement replications $y^*_{ij1}, \ldots, y^*_{ij(v+1)}$ with the same predictor $\lambda_{j0} + \boldsymbol{\lambda}'_j \boldsymbol{z}_i + \boldsymbol{\alpha}'_j \boldsymbol{\omega}_i$ but with different variances $\sigma^2_{r_{ijl}}$ for the error terms $\zeta_{ijl} \,|\, r_{ijl}$, $l = 1, \ldots, y_{ij} + 1$. In the following detailed presentation (6) we shortly write y^*_{ijl} instead of $y^*_{ijl} \,|\, \cdot, r_{ijl}$:

$$\begin{pmatrix} y^*_{ij1} \\ \vdots \\ y^*_{ijl} \\ \vdots \\ y^*_{ij(v+1)} \end{pmatrix} = \begin{pmatrix} \lambda_{j0} + \boldsymbol{\lambda}'_j \boldsymbol{z}_i + \boldsymbol{\alpha}'_j \boldsymbol{\omega}_i \\ \vdots \\ \lambda_{j0} + \boldsymbol{\lambda}'_j \boldsymbol{z}_i + \boldsymbol{\alpha}'_j \boldsymbol{\omega}_i \\ \vdots \\ \lambda_{j0} + \boldsymbol{\lambda}'_j \boldsymbol{z}_i + \boldsymbol{\alpha}'_j \boldsymbol{\omega}_i \end{pmatrix} + \begin{pmatrix} \zeta_{ij1} \\ \vdots \\ \zeta_{ijl} \\ \vdots \\ \zeta_{ij(v+1)} \end{pmatrix}. \tag{6}$$

Defining $\boldsymbol{y}^*_i = \left(y^*_{i11}, \ldots, y^*_{i1(y_{i1}+1)}, \ldots, y^*_{ip1}, \ldots, y_{ip(y_{ip}+1)}\right)'$,
$\boldsymbol{\lambda}^*_0 = (\lambda_{10}, \ldots, \lambda_{10}, \ldots, \lambda_{p0}, \ldots, \lambda_{p0})'$, $\boldsymbol{\Lambda}^* = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p, \ldots, \boldsymbol{\lambda}_p)'$,
$\boldsymbol{A}^* = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p, \ldots, \boldsymbol{\alpha}_p)'$, and $\boldsymbol{\varepsilon} = \left(\zeta^*_{i11}, \ldots, \zeta^*_{i1(y_{i1}+1)}, \ldots, \zeta^*_{ip1}, \ldots, \zeta_{ip(y_{ip}+1)}\right)'$ the underlying Gaussian measurement model is given in matrix notation as

$$\boldsymbol{y}^*_i = \boldsymbol{\lambda}^*_0 + \boldsymbol{\Lambda}^* \boldsymbol{z}_i + \boldsymbol{A}^* \boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i\,, \quad i = 1, \ldots, n. \tag{7}$$

Given the values of the manifest Poisson variables the dimension of the vector $\boldsymbol{y}^*_i$ is

$\dim(\boldsymbol{y}_i^*) = (y_{i1}+1) + \cdots + (y_{ip}+1)$, $i = 1, \ldots, n$. Note that the dimensions of $\boldsymbol{\lambda}_0^*$, $\boldsymbol{\Lambda}^*$ and $\boldsymbol{A}^*$ also depend on i, but we suppress this notationally. We point out that according to (6) all $y_{ij}+1$ underlying variables of the Poisson indicator y_{ij} have the same predictor. This is why we have to repeat the rows λ_{j0}, $\boldsymbol{\lambda}_j'$ and $\boldsymbol{\alpha}_j'$, $j = 1, \ldots, p$, $(y_{ij}+1)$-times in the intercept vector $\boldsymbol{\lambda}_0^*$, in the matrix of the factor loadings $\boldsymbol{\Lambda}^*$ and in the matrix of the regression coefficients $\boldsymbol{A}^*$. The error term $\boldsymbol{\varepsilon}_i$ in (7) is normal,

$$\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i) \quad \text{with} \quad \boldsymbol{\Sigma}_i = \operatorname{diag}\left(\sigma^2_{r_{i11}}, \ldots, \sigma^2_{r_{i1(y_{i1}+1)}}, \ldots, \sigma^2_{r_{ip1}}, \ldots, \sigma^2_{r_{ip(y_{ip}+1)}}\right).$$

The latent factors $\boldsymbol{z}_i$ are assumed to be i. i. d. with $\boldsymbol{z}_i \sim N_q(0, \boldsymbol{I}_q)$. Then the conditional and marginal characterising moments of the measurement model are $\operatorname{Var}\left(y_{ijl}^* \,|\, \boldsymbol{z}\right) = \sigma^2_{r_{ijl}}$, $\operatorname{Cov}\left(y_{ijl}^*, y_{ikm}^* \,|\, \boldsymbol{z}\right) = 0$, $\operatorname{Var}\left(y_{ijl}^*\right) = \sum_{r=1}^{q} \lambda_{jr}^2 + \sigma^2_{r_{ijl}}$, $\operatorname{Cov}\left(y_{ijl}^*, y_{km}^*\right) = \sum_{r=1}^{q} \lambda_{jr}\lambda_{lr}$.

The measurement model faces a well known identification problem because there is an indeterminateness concerning the matrix of factor loadings $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_p)'$ 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 ($\operatorname{Var}(\widetilde{\boldsymbol{z}}_i) = \boldsymbol{V}\boldsymbol{I}_q\boldsymbol{V}' = \boldsymbol{I}_q$). An indefinite number of models exists since all orthogonal rotations of the latent space could occur. The solution lies in the restriction of parameters of $\boldsymbol{\Lambda}$ in a suitable way e. g. Lopes and West (2004). We solve this identification problem by postulating a lower block triangular matrix of factor loadings of full rank with positive diagonal elements. This way to ensure identification is also used by Geweke and Zhou (1996) and Aguilar and West (2000). Since we use only one latent variable in the later presented simulation and application, this restriction is not necessary. The reason for this is that there exists only one orthogonal transformation in a model with only one latent variable. The only possible orthogonal transformation is nothing else than a change of the sign of factor loadings and factor scores. Hence, in a model with one latent

variable the problem of indeterminateness can simply be solved after the computation. Where required, we solely have to multiplicate the loadings and scores by -1, see also Steinert (2006).

We complete this subsection with *prior assumptions on parameters of the measurement model*: Let $\bar{\boldsymbol{\lambda}} = \left(\lambda_{10}, \boldsymbol{\alpha}_1', \boldsymbol{\lambda}_1', \lambda_{20}, \boldsymbol{\alpha}_2', \boldsymbol{\lambda}_2', \ldots, \lambda_{p0}, \boldsymbol{\alpha}_p', \boldsymbol{\lambda}_p', \right)'$ denote the $p \cdot (1 + d + q)$-dimensional vector which contains all intercepts, factor loadings and regression coefficients. For intercepts λ_{j0} and regression coefficients $\boldsymbol{\alpha}_j$ we usually choose noninformative flat priors

$$p\left(\lambda_{j0}\right) \propto const\,, \quad p\left(\boldsymbol{\alpha}_j\right) \propto const\,, \quad j = 1, \ldots, p\,.$$

Choice of (weakly) informative priors is also possible, however, and can easily be integrated in the corresponding Gaussian full conditional of our Gibbs sampling scheme.

However, it may be reasonable to include prior information for the factor loadings in order to prevent the occurrence of Heywood cases in the Bayesian setting. A Heywood case appears when one factor loads up completely on one (sometimes even more) indicator(s), hence the latent variable accounts for the full variability of the respective indicator. Since this result is highly implausible, we choose informative priors with a normal density centered at zero with a certain precision (inverse variance). A recommended 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 only weakly informative and therefore allows to obtain high factor loadings. We suggest three different prior precision settings, weak (0.5), standard (1.0) and strong (4.0). Simulation studies in Raach (2005) show that the probability of Heywood cases decreases as the number of observations n or the number of indicators increases. Moreover, for large sample sizes (some thousands), all three choices for the prior variance lead to practically the same posterior estimates. Note also, that Heywood cases tend to arise of the researcher attempts to extract more latent variables than

the information provided by the data contains. This is not the case here, because we use only one latent variable in simulation studies (Section 4) and in the application (Section 5).

2.2 Structural model

Structural models commonly relate latent variables z_{ir}, $r = 1 \ldots, q$, to a covariate vector $\boldsymbol{u}_i$ through a *linear* model

$$z_{ir} = \boldsymbol{u}_i'\boldsymbol{\gamma} + \delta_{ir}, \qquad i = 1, \ldots, n \tag{8}$$

with i. i. d. errors $\delta_{ir} \sim N(0, 1)$. The variances of the errors are set to 1 for identifiability reasons. The *linear* predictor $\eta_{ir} = \boldsymbol{u}_i'\boldsymbol{\gamma}_r$ assumes linear parametric effects of the covariate vector $\boldsymbol{u}_i$, with coefficient vector $\boldsymbol{\gamma}_r$. Again for identifiability reasons, the linear predictor must not contain an intercept term. We extend the linear predictor to a *geoadditive* predictor

$$\eta_{ir} = f_{r1}(x_{i1}) + \cdots + f_{rg}(x_{ig}) + f_{r,spat}(s_i) + \boldsymbol{\gamma}_r'\boldsymbol{u}_i\,, \tag{9}$$

where $f_{r1}(x_{i1}), \ldots, f_{rg}(x_{ig})$ are nonlinear functions for the effects of additional continuous covariates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_g$ and $f_{r,spat}(s)$ is the spatial effect at location $s \in \{1, \ldots, d\}$, indexing d geographical regions or, more generally, a discrete lattice of spatial locations. If there is no spatial information in form of location variables in the data, the spatial effect $f_{r,spat}(s)$ is deleted in (9), and we obtain an *additive predictor*. The geoadditive predictor has the same form as for geoadditive or structured additive regression models proposed in Fahrmeir, Kneib and Lang (2004), Lang and Brezger (2004), and for semiparametric latent variable models for mixed Gaussian and categorical indicators in Fahrmeir and Raach (2006). In complete analogy we model functions $f_{r1}, \ldots, f_{rg}$ through Bayesian P-splines and spatial

effects through a Markov random field. Then the predictor vector $\boldsymbol{\eta}^{(r)} = (\eta_{1r}, \ldots, \eta_{nr})'$ can always be written in form of a large linear (mixed) model

$$\boldsymbol{\eta}^{(r)} = (\eta_{1r}, \eta_{2r}, \ldots, \eta_{nr})' = \boldsymbol{X}_1\boldsymbol{\beta}_{r1} + \ldots + \boldsymbol{X}_g\boldsymbol{\beta}_{rg} + \boldsymbol{X}_{spat}\boldsymbol{\beta}_{r,spat} + \boldsymbol{U}\boldsymbol{\gamma}_r\,, \qquad (10)$$

where the design matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_g$ contain the function values of B-spline basis functions and $\boldsymbol{X}_{spat}$ is a $(n \times d)$-incidence matrix, where the sth element of row i is 1 if observation i comes from region s and all other elements are 0. The parameter vectors $\boldsymbol{\beta}_{r1}, \ldots, \boldsymbol{\beta}_{r,spat}$ are random with priors defined below. From a frequentist perspective the parameter vector $\boldsymbol{\gamma}_r$ is considered as 'fixed', so that (10) may be interpreted as a linear mixed model predictor. The structural model is completed by prior assumptions on parameters and functions. We assume independent priors for separate functions and parameters as well as for functions and parameters of different predictors, $\boldsymbol{\eta}^{(r)}$, $r = 1, \ldots, q$. To simplify notation, we therefore drop indices. For the parameter $\boldsymbol{\gamma}$ of the linear part of predictors, we routinely assign flat, noninformative priors $p(\boldsymbol{\gamma}) \propto const$, but informative normal priors would also be possible.

Priors for functions of continuous covariates are defined through Bayesian P-splines, based on Lang and Brezger (2004) and Brezger and Lang (2006). 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, and is constructed by a linear combination

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

of $d = D + I$ B-spline basis functions B_c with regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)'$. The characteristics of B-splines are described in the above mentioned literature and in

Dierckx (1993), Eilers and Marx (1996). Smoothness of the function f is achieved by penalizing differences of coefficients of adjacent B-splines. In a Bayesian approach, this penalization is incorporated conveniently by applying a first-order or second-order random walk prior to the B-splines regression coefficients f:

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

with $u_t \sim N(0, \kappa^2)$, respectively. The first-order random walk has a diffuse prior $\beta_1 \propto const$; the second-order random walk additionally has $\beta_2 \propto const$. 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

$$p(\boldsymbol{\beta} \,|\, \kappa^2) = \exp\left(-\frac{1}{2\kappa^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}\right)$$

with appropriately defined penalty matrix $\boldsymbol{K}$. 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 $X_{ic} = B_c(x_i)$. Thus the vector of function evaluations for all observations is given by $\boldsymbol{X}\boldsymbol{\beta}$. In our analysis, we choose B-splines of degree $D = 3$ with $I = 10$ intervals.

The prior for the spatial effect is defined through a Markov random field (MRF). Let us assume that location s_i denotes the region where observation i comes from, and the vector $\boldsymbol{f}_{geo} = (f_{geo}(s_1), \ldots, f_{geo}(s_n))$ of function evaluations $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)$ contains the effect $\beta_s := f_{spat}(s)$, $s = 1, \ldots, d$, of the d different regions. The spatial function evaluations of all observations i can be written as $\boldsymbol{X}\boldsymbol{\beta}$ 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 our context, two regions are considered neighbours when they share a common boundary. We apply the following smoothness prior to the spatial effects β_c, $c = 1, \ldots, d$, for all d regions:

$$\beta_s | \beta_{s'}, s' \neq s, \kappa^2 \sim N\left(\sum_{s' \in \partial_s} \frac{\beta_{s'}}{N_s}, \frac{\kappa^2}{N_s} \right), \tag{11}$$

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. The entire prior distribution follows as $p(\boldsymbol{\beta} \,|\, \kappa^2) \propto \exp(-\boldsymbol{\beta}' \boldsymbol{K} \boldsymbol{\beta} / (2\kappa^2))$ 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 general MRF priors are possible, see Rue and Held (2005).

Priors for smoothing parameters: All priors for nonparametric functions and the spatial effect are defined conditional on the inverse smoothing parameter κ^2. It is automatically estimated in our Bayesian approach. We assign weakly informative but proper inverse Gamma priors

$$\kappa^2 \sim IG\,(a, b)\,,$$

with small values $a = b = \epsilon$, to avoid problems with possibly improper posteriors.

Stacking all regression parameters and smoothing parameters in vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$, the

full prior specification is given by

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\kappa}) &= p(\boldsymbol{\beta} \,|\, \boldsymbol{\kappa}) p(\boldsymbol{\gamma}) p(\boldsymbol{\kappa}) \\ &= \prod_{r=1}^{q} \prod_{h=1}^{spat} p(\boldsymbol{\beta}_{rh}) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r) \\ &\propto \prod_{r=1}^{q} \prod_{h=1}^{spat} \exp(-\tfrac{1}{2\kappa_{rh}^2} \boldsymbol{\beta}_{rh}' \boldsymbol{K}_{rh} \boldsymbol{\beta}_{rh}) p(\kappa_{rh}^2) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r) \,. \end{aligned}$$

3 Bayesian inference

Full Bayesian inference can be carried out via Gibbs sampling in combination with data augmentation, considering underlying variables $\boldsymbol{y}^*$ and latent variables $\boldsymbol{z}$ as additional "parameters". Gathering interarrival times $\boldsymbol{\tau}$ and mixture component indicators $\boldsymbol{r}$ of the underlying variable model, intercepts $\boldsymbol{\lambda}_0$, factor loadings $\boldsymbol{\Lambda}$ and direct effects $\boldsymbol{A}$ of the measurement model as well as parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ of the structural model in the vector $\boldsymbol{\theta}$ of all parameters, the resulting posterior is

$$\begin{aligned} p\left(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z} \,|\, \boldsymbol{y}, \right) &\propto p\left(\boldsymbol{\theta}\right) \cdot p\left(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{z} \,|\, \boldsymbol{\theta}\right) \\ &= p\left(\boldsymbol{\theta}\right) \cdot p\left(\boldsymbol{y}^*, \boldsymbol{z} \,|\, \boldsymbol{\theta}\right) \cdot p\left(\boldsymbol{y} \,|\, \boldsymbol{y}^*, \boldsymbol{\theta}\right) \end{aligned}$$

Gibbs sampling is performed in the following steps:

1. Generate underlying variables $\boldsymbol{y}^*$, including interarrival times $\boldsymbol{\tau}$ and component indicators $\boldsymbol{r}$.
2. Generate latent variables $\boldsymbol{z}$.
3. Draw from the posterior $p\left(\boldsymbol{\gamma} \,|\, \cdot\right)$.
4. Draw from the posteriors $p\left(\boldsymbol{\beta}_h \,|\, \cdot\right)$, $h = 1, \ldots, spat$.
5. Draw from the posteriors $p\left(\kappa_h \,|\, \cdot\right)$ for smoothing parameters.
6. Draw from posteriors $p\left(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A} \,|\, \cdot\right)$.

Steps 2-5 are essentially the same as in Fahrmeir and Raach (2006) and Raach (2005). Therefore we focus here on steps 1 and 6. Further details are given in Steinert (2006).

Step1. Following Frühwirth-Schnatter and Wagner (2006), we proceed as follows.

(i) Generating interarrival times $\boldsymbol{\tau}_j = \{\tau_{ijl}, l = 1, \ldots, (y_{ij}+1), i = 1, \ldots, n\}$:

Set $v = y_{ij}$. If $y_{ij} > 0$ draw $g_{ij1}, \ldots, g_{ijv}$ from the uniform distribution on [0, 1] and construct the resulting order statistic $g_{ij(1)}, \ldots, g_{ij(v)}$. Interarrival times are then given by

$$\tau_{ijl} = g_{ij(l)} - g_{ij(l-1)}\,, \qquad l = 1, \ldots, v\,,$$

where $g_{ij(0)} = 0$, and

$$\tau_{ij(v+1)} = 1 - \sum_{l=1}^{v} \tau_{ijl} + \xi_i\,, \qquad \xi_i \sim Exp(\mu_{ij})\,.$$

(ii) Generating component indicators r_{ijl}, $l = 1, \ldots, (y_{ij}+1)$:

$$P\left(r_{ijl} = k \,|\, \tau_{ijl}, \cdot\right) \propto f\left(\tau_{ijl} \,|\, r_{ijl} = k, \cdot\right) w_k.$$

(iii) Generating underlying variables y^*_{ijl}:

$$y^*_{ijl} \,|\, \cdot \sim N\left(\lambda_{j0} + \boldsymbol{\lambda}_j' \boldsymbol{z}_i + \boldsymbol{\alpha}_j' \boldsymbol{\omega}_i, \sigma^2_{r_{ijl}}\right)$$

Step 6. Generating $\bar{\boldsymbol{\lambda}}^j = \left(\lambda_{0j}, \boldsymbol{\alpha}_j', \boldsymbol{\lambda}_j'\right)'$:

$$\bar{\boldsymbol{\lambda}}^j \,|\, \cdot \sim N\left(\mathrm{E}(\bar{\boldsymbol{\lambda}}^j \,|\, \cdot), \mathrm{Var}(\bar{\boldsymbol{\lambda}}^j \,|\, \cdot)\right).,$$

where

$$\mathrm{E}\left(\bar{\boldsymbol{\lambda}}^j \,|\, \boldsymbol{W}_j \boldsymbol{y}^*_j, \boldsymbol{z}, \boldsymbol{\omega}\right) = \left(\bar{\boldsymbol{\Lambda}}^{*j} + \boldsymbol{L}_j' \boldsymbol{W}_j^{-1} \boldsymbol{L}_j\right)^{-1} \left(\bar{\boldsymbol{\Lambda}}^{*j} \bar{\boldsymbol{\lambda}}^{*j} + \boldsymbol{L}_j' \boldsymbol{W}_j^{-1} \boldsymbol{y}^*_j\right)$$

$$\text{Var}\left(\bar{\boldsymbol{\lambda}}^j \mid \boldsymbol{y}_j^*, \boldsymbol{z}, \boldsymbol{\omega}\right) = \left(\bar{\boldsymbol{\Lambda}}^{*j} + \boldsymbol{L}_j' \boldsymbol{W}_j^{-1} \boldsymbol{L}_j\right)^{-1}.$$

and

$$\boldsymbol{y}_j^* = \left(y_{1j1}^*, \ldots, y_{1j(y_{1j}+1)}^*, \ldots, y_{nj1}^*, \ldots, y_{nj(y_{nj}+1)}^*\right)',$$

$$\boldsymbol{W}_j = diag\left(\sigma^2_{r_{1j1}}, \ldots, \sigma^2_{r_{1j(y_{1j}+1)}}, \ldots, \sigma^2_{r_{nj1}}, \ldots, \sigma^2_{r_{nj(y_{nj}+1)}}\right),$$

$$\boldsymbol{L}_j' = (\boldsymbol{l}_1', \ldots, \boldsymbol{l}_1', \ldots, \boldsymbol{l}_n', \ldots, \boldsymbol{l}_n')'$$

and

$$\boldsymbol{l}_i' = (1, \omega_{i1}, \ldots, \omega_{id}, z_{i1}, \ldots, z_{iq})'.$$

4 Simulation

To investigate performance several simulation studies were conducted in Steinert (2006). We confine presentation to only one LVM with two different numbers of observations, $N_1 = 300$ and $N_2 = 1000$, respectively. This model includes three Poisson distributed indicators and one latent variable. It also includes two indirect covariates, one metric and one spatial covariate. To demonstrate that it is possible to estimate nonparametric effects of a metric covariate in a LVM with Poisson responses we used the function

$$f(x) = sin\left(\frac{2\pi x}{20}\right), \quad x \in [0, 20],$$

which rises and drops with a high curvature. As in Raach (2005) we used the two-dimensional function

$$f_{spat} = sin\left(\frac{2\pi x}{20}\right) \cdot \left(\frac{2\pi y}{20}\right), \quad x = 1, \ldots, 20,\ y = 1, \ldots, 20,$$

to generate a spatial covariate. As mentioned before, here two regions are considered neighbours when they share a common boundary. According to this assumption our regions have four neighbours, apart from the regions at the corners or on the border which clearly have less neighbouring regions as can be seen in Figure 2.

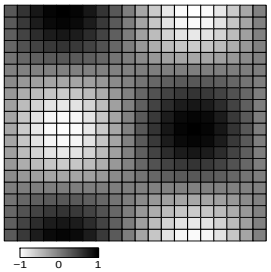

Figure 2: *Map of 400 regions with corresponding true functional values of spatial function* f_{spa}

As mentioned above LVM with mixed responses are possible, too. Results of further simulation studies for models with mixed responses as well as with pure Poisson responses and different covariate combinations and more than one latent factor are given in Steinert (2006).

After performing analysis with different numbers of iterations for the burnin and the sampling phase, we considered 2000 iterations for the burnin phase and 5000 iterations for the sampling phase as satisfactory. In order to judge the estimation quality of the model parameters we generated $S = 50$ different data sets for both numbers of observations. After analysing these data sets as described in the preceding sections, we calculated mean (MEAN), standard deviaton (STD), bias (BIAS) and the mean squared error (MSE) to assess the quality of estimation. Let θ^{true} denote the true value of an arbitrary model parameter, $\hat{\theta}_s$ the estimated value for the sth data set and $\hat{\theta}_s^{std}$ denotes the standard error of the estimation for data set s. Above mentioned characteristic magnitudes can be calculated as follows: MEAN= $\frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s$, STD= $\frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s^{std}$, BIAS= $\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\theta}_s - \theta_{true}\right)$, MSE= $\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\theta}_s - \theta_{true}\right)^2$. By calculating the portion of simulation runs for which the true parameter value θ^{true} is inside the corresponding 95% credible region we obtain the

empirical 95% coverage (COV) probability.

We first depict the estimated parameters of the measurement model. Thereafter we show the results of the structural model.

Mean, standard deviation, bias, mean squared error and coverage of the simulation studies are given in Table 2 for $N_1 = 300$ observations and in Table 3 for $N_2 = 1000$ observations, respectively. As expected STD and MSE decrease for an increasing number of observations. In general we can conclude that the estimates fit the true values very well.

Table 2: *True parameter values and estimated results obtained by simulations of 50 different data sets with* $N_1 = 300$ *observations*

Par.	TRUE	MEAN	STD	BIAS	MSE	COV
λ_{10}	0.3	0.3032	0.0595	0.0032	0.0034	96
λ_{20}	0.5	0.5021	0.0517	0.0021	0.0026	98
λ_{30}	0.8	0.7759	0.0681	-0.0240	0.0051	94
λ_{11}	0.5	0.4823	0.0455	-0.0176	0.0023	98
λ_{21}	0.5	0.4769	0.0470	-0.0230	0.0027	90
λ_{31}	0.8	0.7786	0.0654	-0.0213	0.0046	92

Table 3: *True parameter values and estimated results obtained by simulations of 50 different data sets with* $N_2 = 1000$ *observations*

Par.	TRUE	MEAN	STD	BIAS	MSE	COV
λ_{10}	0.3	0.2967	0.0352	-0.0032	0.0012	96
λ_{20}	0.5	0.5039	0.0517	0.0299	0.0039	98
λ_{30}	0.8	0.8071	0.0375	0.0071	0.0014	96
λ_{11}	0.5	0.4882	0.0277	-0.0117	0.0008	94
λ_{21}	0.5	0.4843	0.0285	-0.0156	0.0010	90
λ_{31}	0.8	0.7765	0.0318	-0.0234	0.0015	82

For the nonparametric effects we used highly diffuse but proper hyperpriors for the smoothing parameters with $a = b = 0.001$, c f. subsection 2.2. Figure 3 shows the results of the estimation of nonparametric effects of a metric covariate and the corresponding 10%- and 90%-quantiles. Since the nonparametric estimates fit the true values very well differences between estimated and true function are hardly distinguishable for both numbers of observations. Furthermore it is observed that with increasing numbers of observations the bias decreases and the 80% credible region narrows.

The mean of the estimated spatial effects and the corresponding bias are also plotted in Figure 3 for both numbers of observations. The plots show that the quality of estimation

increases for $N_2 = 1000$ observations. This is justified by the fact that on average we have only 0.75 observations for one region for $N_1 = 300$ observations. However, in both cases the estimates have the right tendency. Another common property is that high function values are estimated too low and low function values are estimated too high, as can be seen in the bias graphs. This is due to our choice of a MRF prior, but might not be representative for real data sets with smoother spatial effects than our function f_{spat} whose function values change quite suddenly between high and low values.

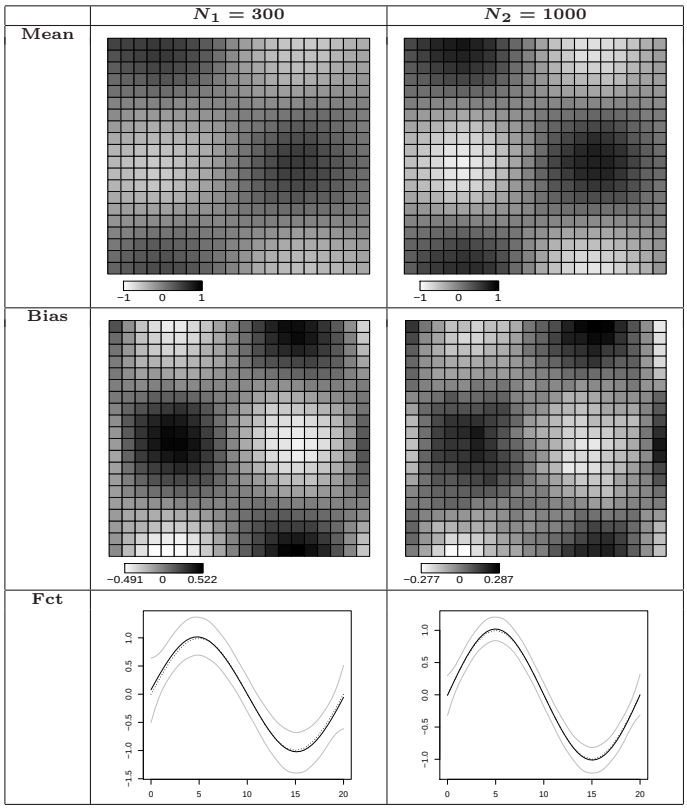

Figure 3: *Mean (top) and bias (middle) of the estimated spatial effects. The function estimates (below) show the mean (black line), the 10%- and the 90%-quantiles (grey lines). The dotted line represents the true function f.*

The estimates of the smoothing parameters κ^2_{spa} and κ^2_{fct} are given in table 4. The estimated values as well as the standard deviations decrease with an increasing number of observations.

Table 4: *Estimation of smoothing parameters* κ^2_{spa} *and* κ^2_{fct}

	$N_1 = 300$		$N_2 = 1000$	
Par.	**Mean**	**STD**	**Mean**	**STD**
κ^2_{spa}	0.5945	0.2794	0.4240	0.0937
κ^2_{fct}	0.4847	0.1585	0.3390	0.0705

We can conclude that besides a very good estimation of model parameters corresponding to the measurement model in a LVM with Poisson responses, nonparametric effects of an indirect covariate as well as spatial effects of an indirect spatial covariate can be recovered quite reasonably. Furthermore, as expected, the quality of estimation becomes better with an increasing number of observations.

5 Application

We illustrate our approach with an application to data from a study on post war human security in Cambodia. The conflict and violence data in this study has been collected by the monitoring arm of the Government of Cambodia's decentralisation program SEILA, the Khmer word for foundation stone. We use data collected for the year 2002 and obtained from headmen and leaders of over 13000 villages and urban neighbourhoods. More details on the data as well as sociological and political background is given in Benini, Owen and Rue (2006). They used separate geoadditive count data models to analyse the impact of the legacy of war, poverty and resource competition, urbanity, and governance quality on the three dependent variables

- number of serious crime committed
- number of land conflicts
- number of households known to have domestic violence problems.

We apply a Poisson indicator LVM to these three indicators, focusing on the latent variable "disposition for violence". Instead of the total numbers of counts per year, we use the

monthly averages y_1, y_2 and y_3 of the three count variables as target variables. Because the yearly numbers are only estimates provided by local leaders, the effect of averaging can be neglected, and it helps to make data analysis computationally feasible.

Based on the study of Benini, Owen and Rue (2006), we formulate the following Poisson indicator LVM:

$$y_{ij} = \exp(\mu_{ij}) \qquad \mu_{ij} = \lambda_{j0} + \lambda_j z_i + \alpha_j \cdot nrfam_i$$

$$z_i = f_1(usbomb_i) + f_2(contam_i) + f_{spat}(community_i) + \delta_i$$

For $i = 1, \ldots, n = 1619$,

- y_{i1} is the monthly average of serious crime ($j = 1$) in community i,
- y_{i2} is the monthly average of land conflicts ($j = 2$) in community i,
- y_{i3} is the monthly average of domestic violence ($j = 3$) in community i,
- $nrfam_i$ is the logarithm of the numbers of families living in community i,
- $usbomb_i$ is $\log_{10}(pound_i + 1)$, where $pound$ is the US bombing load in community i given in pounds,
- $contam_i$ is $\log_{10}(sq\,m + 1)$, where $sq\,m$ is the contaminated area in square meters in community i,
- $community_i$ is the spatial location of community i, compare the map in Figure 5.

The $\log_{10}$-transformations were chosen also in Benini, Owen and Rue (2006), motivated by the belief that the destruction in the heavily bombed and mined regions went hand in hand with the magnitude of the bombing or mining, rather than the absolute bombing density in terms of load per surface. We use $nrfam_i$ as a direct covariate to adjust for the effect of size of the communities on the rates μ_{ij}.

The functions f_1 and f_2 are modelled as cubic P-splines with ten knots, and the spatial effects as a Markov random field. Because data were missing for nine communities, esti-

mation based on the remaining communities only. As for the simulations before, we used 2000 iterations for the burnin and 5000 iterations for the sampling phase, and additionally a thinning parameter of five. Again we used highly diffuse hyperpriors for the smoothing parameters with $a = b = 0.001$. Furthermore in Figure 4 we plot the autocorrelation of the sampling paths of some parameters, i. e. the parameters λ_{30} (Figure 4a) and λ_{31} (Figure 4b).

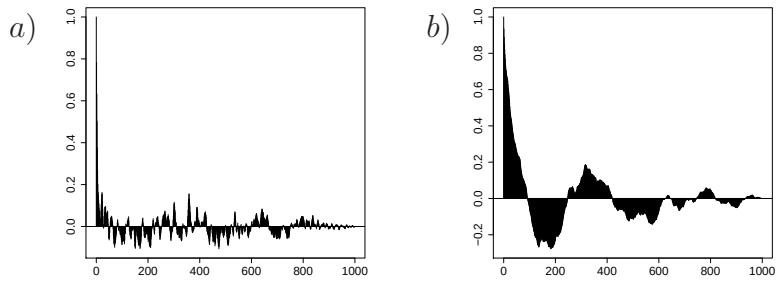

Figure 4: *a) Autocorrelation for λ_{30}, b) autocorrelation for λ_{31}; The x-axis denotes the lag.*

Table 5 shows the parameter estimates for the measurement model. The factor loading $\lambda_{11} = 0.4378$ for serious crime is the highest, but still of comparable magnitude with the two other factor loadings, so that no single factor loading is strongly dominant.

Table 5: *Results*

Parameter	MW	STD	2.5%-Quantil	Median	97.5%-Quantil
λ_{10}	-9.7060	0.4951	-10.6835	-9.6862	-8.7985
λ_{20}	-6.1013	0.3585	-6.8249	-6.0949	-5.4142
λ_{30}	-4.7471	0.2826	-5.3056	-4.7472	-4.1829
λ_{11}	0.4378	0.0385	0.3709	0.4363	0.5165
λ_{21}	0.3568	0.0276	0.3069	0.3567	0.4141
λ_{31}	0.3202	0.0243	0.2758	0.3216	0.3629
α_{11}	1.2536	0.0653	1.1295	1.2510	1.3842
α_{21}	0.8619	0.0489	0.7681	0.8610	0.9581
α_{31}	0.7681	0.0383	0.6937	0.7692	0.8408
κ^2_{gontam}	0.0476	0.1729	0.0007	0.0127	0.2866
κ^2_{usbomb}	0.0177	0.0494	0.0005	0.0052	0.1041
$\kappa^2_{community}$	3.8455	1.0804	2.1231	3.7536	6.0977

The map of estimated spatial effects in Figure 5 provides clear evidence of significant spatial variability. For interpretation, it is important to note that our measurement model already adjust for population density in the communities through the effect of the number of families (*nrfam*) living in a community. Therefore, the map in Figure 5 shows spatial

effects after adjusting for population density. For comparison, Figure 6 shows observed rates of land conflicts for the provinces of Cambodia, rated to the population number in the provinces. A map of domestic violence rated to the population has a similar pattern.

The map of community-specific spatial effects on the latent variable "disposition of violence", after adjustment for the number of families, roughly has a similar pattern: Disposition for violence seems to be significantly below average in the north-east at the border to Laos and Vietnam, in the south-east, in particular the Mekong delta, and in parts of the west. On the other side, there is a significant increase in the east at the border to Vietnam, and in the north-east at the border to Thailand. This evidence motivates to search for underlying determinants, to support politics and governance.

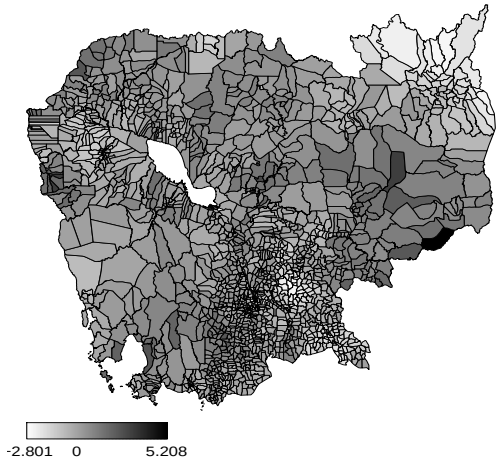

Figure 5: *Map of Cambodia with the estimated spatial effects for all 1628 communities*

Figure 7 shows the estimated effects f_1 (*usbomb*) and f_2 (*contam*). (Note that both variables have already been transformed logarithmically). Function f_2 indicates a monotonically increasing effect of the amount of contaminated area, while the effect of the intensity of US bombing seems to be linear and small. The latter was confirmed in a second analysis, where the effect of *usbomb* was assumed to be linear.

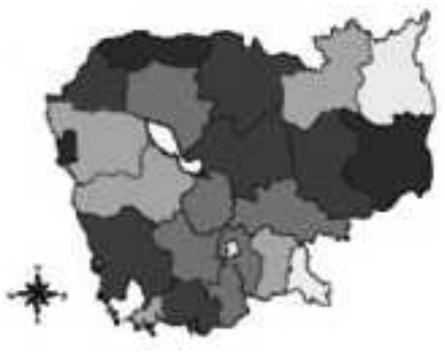

Figure 6: *Province rates for land conflicts, rated to population (light: below average, dark: above average)*

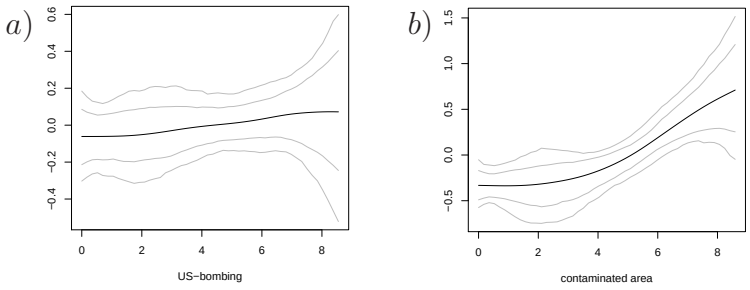

Figure 7: *a) Estimated effects $f_1\,(usbomb)$, b) estimated effects $f_2\,(contam)$*

6 Conclusion

Modern Bayesian inference based on MCMC technology is particular useful in developing flexible models incorporating latent variables. Our Poisson indicator variable model can be combined with corresponding LVM for continuous, ordinal and binary indicators to a broad class of latent variable models useful in many applications, from social sciences to medicine and biology.

Acknowledgment

We thank Aldo Benini for making the Cambodia data available and for useful advice on substantive problems, Malis Min, formerly with Legal Aid Cambodia, for discussion on land conflicts, and Alexander Raach for providing his software as the basis for the methods implemented in this work. Financial support from the SFB 386 "Statistical Analysis of Discrete Structures" is gratefully acknowledged.

References

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics* **18**, 338–357.

Arminger, G. and 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.

Benini, A., Owen, T. and Rue, H. (2006). A semi-parametric spatial regression approach to post-war human security: Cambodia 2002-2004, *Technical Report*.

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

Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics*, **1**08, 281-316.

Dierckx, P. (1993). *Curve and surface fitting with splines*. Oxford: Clarendon Press.

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

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

Fahrmeir, L. and Raach, A. (2006). A Bayesian semiparametric latent variable model for mixed responses. *Discussion paper 471*, revised for *Psychometrika*.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary Mixture Sampling for parameter-driven Models of time Series of Counts with Application to State Space Modelling, to appear in *Journal of Computational and Graphical Statistics*.

Geweke, J. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* **9**, 557–587.

Jöreskog, K. G. and 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.

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

Lee, S.-Y. and Song, X.-Y. (2004). Bayesian model comparison of nonlinear sructural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology* **57**, 131–150.

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

Moustaki, I., Jöreskog, K. G. and 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. and 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.

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, Department of Statistics, University of Munich.

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

Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society B* **59**, 667–678.

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

Song, X.-Y. and Lee, S.-Y. (2005). Statistical analysis of a two-level nonlinear structural equation model with fixed covariates. *Journal of Educational and Behavioral Statistics* **30**, 1–26.

Steinert, S. (2006). *Semiparametrische Latente-Variablen-Modelle.* Diploma Thesis, Department of Statistics, University of Munich.

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