Gschlößl, Czado:

Does a Gibbs sampler approach to spatial Poisson regression models outperform a single site MH sampler?

Sonderforschungsbereich 386, Paper 460 (2005)

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

Does a Gibbs sampler approach to spatial Poisson regression models outperform a single site MH sampler?

Susanne Gschlößl Claudia Czado *

December 6, 2005

Abstract

In this paper we present and evaluate a Gibbs sampler for a Poisson regression model including spatial effects. The approach is based on Frühwirth-Schnatter and Wagner (2004b) who show that by data augmentation using the introduction of two sequences of latent variables a Poisson regression model can be transformed into an approximate normal linear model. We show how this methodology can be extended to spatial Poisson regression models and give details of the resulting Gibbs sampler. In particular, the influence of model parameterisation and different update strategies on the mixing of the MCMC chains is discussed. The developed Gibbs samplers are analysed in two simulation studies and applied to model the expected number of claims for policyholders of a German car insurance company. The mixing of the Gibbs samplers depends crucially on the model parameterisation and the update schemes. The best mixing is achieved when collapsed algorithms are used, reasonable low autocorrelations for the spatial effects are obtained in this case. For the regression effects however, autocorrelations are rather high, especially for data with very low heterogeneity. For comparison a single component Metropolis Hastings algorithms is applied which displays very good mixing for all components. Although the Metropolis Hastings sampler requires a higher computational effort, it outperforms the Gibbs samplers which would have to be run considerably longer in order to obtain the same precision of the parameters.

Key words: block updates, collapsing, data augmentation, Gibbs sampler, model parameterisation, spatial Poisson count data

*Both at Center of Mathematical Sciences, Munich University of Technology, Boltzmannstr.3, D-85747 Garching, Germany,email: susanne@ma.tum.de, cczado@ma.tum.de, http://www.ma.tum.de/m4/

1 Introduction

In this paper we present a straightforward Gibbs sampler for spatial Poisson regression models using data augmentation techniques. In particular, we aim to investigate whether this Gibbs sampler is found to be superior to a conventional single site Metropolis Hastings (MH) sampler. The issue of model parameterisation and several update schemes for the parameters in the Gibbs sampler is thoroughly addressed. The performance of the developed Gibbs sampler schemes and the MH sampler is investigated in two simulation studies as well as on real data from a German car insurance company. Performance of the samplers is measured in the computational costs required to obtain the same precision of the posterior means of the parameters.

Since the full conditional distributions of a spatial Poisson regression model do not follow any standard distribution, often single site MH steps are performed in a MCMC setting, see for example Diggle et al. (1998), Dimakos and Frigessi (2002) or Gschlößl and Czado (2005). However, this requires the choice of appropriate proposal distributions in order to achieve reasonable acceptance rates and a good mixing of the MCMC chains. Advanced independence proposals, like for example a normal proposal with the same mode and inverse curvature at the mode as the target distribution, can lead to high acceptance rates and low autocorrelations but involve considerable computational efforts.

Frühwirth-Schnatter and Wagner (2004a) developed a Gibbs sampler for Poisson regression models for small counts. They show that by data augmentation via the introduction of two sequences of latent variables a linear normal model is obtained. In Frühwirth-Schnatter and Wagner (2004b) an application of this Gibbs sampler to state space models is given, in Frühwirth-Schnatter and Wagner (2004a) the same methodology is applied for standard Poisson regression models and Poisson regression models with overdispersion. Using similar techniques, a Gibbs sampler for logistic models is developed in Frühwirth-Schnatter and Waldl (2004).

The aim of this paper is to show that this methodology can be extended to spatial Poisson regression models in a straightforward manner allowing for a Gibbs update of both regression parameters and spatial effects. Although we only consider spatial Poisson data distributed on regions in this paper, the presented methodology could also be applied on geostatistical Poisson models, see Diggle et al. (1998).

It is well known, that mixing and convergence of the Gibbs sampler depends crucially on several implementation issues, see for example Roberts and Sahu (1997) for a detailed discussion. High autocorrelations can be reduced by updating several parameters in one block or using collapsed algorithms, another important issue is model parameterisation. Gelfand et al. (1995) discuss the efficiency of centered and non-centered parameterisations for hierarchical normal linear models, Papaspiliopoulos et al. (2003) address parameterisation issues for several classes of hierarchical models and introduce partially non-centered parameterisations. Christensen et al. (2005) propose the standardization and orthogonalization of all model components leading to efficient and robust MCMC algorithms.

In this paper both centered and non-centered model parameterisations are considered, various

algorithmic schemes, in particular a joint block update of the intercept and the spatial effects as well as collapsed algorithms, see Liu et al. (1994), are discussed. The performance of the samplers is examined and compared to a single site MH sampler with independence proposals in two simulation studies. In the first study, the samplers are applied on data with both large and small spatial effects, while the second study considers the influence of the data heterogeneity on the performance of the samplers. The performance of the samplers is measured in the computational costs required in order to obtain a certain precision of the posterior means of the regression parameters and spatial effects. This is done by taking both the Monte Carlo error of the posterior means of the parameters and the computational time required for one iteration into account. A very similar approach for comparing the performance of MCMC samplers is conducted by Christensen and Waagepetersen (2002). Among the Gibbs samplers collapsed algorithms perform best. In particular for data with small spatial effects, the Monte Carlo errors of the spatial effects are considerably reduced when collapsed samplers and model parameterisations with non-centered scale or variance are used. The Monte Carlo errors of the regression parameters however are rather high, especially for data with low heterogeneity. The MH independence sampler in contrast, exhibits very low Monte Carlo errors and good mixing for both regression and spatial effects in all settings. Although the MH sampler requires a higher computational effort, this drawback is compensated by the high precision of the posterior means of the parameters. In order to obtain the same precision the Gibbs samplers would have to be run considerably longer, diminishing the computational advantage in comparison to the MH sampler. Therefore we have to conclude that the proposed Gibbs sampler for spatial Poisson regression models can not outperform a single site MH sampler using independence proposals.
This paper is organized as follows. In Section 2 the spatial Poisson regression model is specified and the two steps of the data augmentation scheme are described for this specific model. Details on several algorithmic schemes for updating the regression and spatial effects are given in Section 3. In Section 4 the developed Gibbs sampler schemes are examined and compared to a single component MH sampler with independence proposals in two simulation studys. We also apply the Gibbs samplers to model the expected number of claims in a real data set from a German car insurance company. Section 5 gives a summary and draws conclusions.

2 Data augmentation and Gibbs sampler for spatial Poisson regression models

We assume that observations Y_i, $i = 1, .., n$ observed at J regions follow a Poisson model

$$y_i \sim Poisson(\mu_i). \tag{2.1}$$

The mean μ_i is specified by

$$\mu_i = t_i \exp(\mathbf{z}_i'\boldsymbol{\alpha}) := t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{v}_i'\boldsymbol{\gamma}) = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_{R(i)}) \tag{2.2}$$

where $\mathbf{z}_i' = (\mathbf{x}_i', \mathbf{v}_i')$ denotes the covariate vector $\mathbf{x}_i = (1, x_{i1}, .., x_{ip})'$ and the incidence vector $\mathbf{v}_i = (v_{i1}, .., v_{iJ})'$ for the regions, i.e. $v_{ij} = \begin{cases} 1, & \text{if} \quad R(i) = j \\ 0 & \text{otherwise} \end{cases}$, with $R(i) \in \{1, .., J\}$ denoting the region of the i-th observation. Note, that we do not only observe one single but several observations in each region. Further $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ denotes the vector of regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, .., \beta_p)$ and spatial random effects $\boldsymbol{\gamma} = (\gamma_1, ..\gamma_J)$. By the inclusion of spatial effects we allow for geographical differences in the J regions. The quantity t_i gives the exposure time for the i-th observation.

We assume a normal prior distribution centered around zero with a large standard deviation for the regression parameters $\boldsymbol{\beta}$, in particular

$$\boldsymbol{\beta} \sim N_{p+1}(0, V_0)$$

where $V_0 = \tau^2 I_{p+1}$ with $\tau^2 = 100$. Here $N_p(\mu, \Sigma)$ denotes the p-variate Normal distribution with mean μ and covariance matrix Σ. For the spatial effects a conditional autoregressive (CAR) prior based on Pettitt et al. (2002) is used. In particular, we assume

$$\boldsymbol{\gamma}|\psi, \sigma^2 \sim N_J(0, \sigma^2 Q^{-1})$$

where the elements of the precision matrix $Q = (Q_{ij})$, $i, j = 1, .., J$ are given by

$$Q_{ij} = \begin{cases} 1 + |\psi| \cdot N_i & i = j \\ -\psi & i \neq j, i \sim j \\ 0 & \text{otherwise} \end{cases} . \tag{2.3}$$

We write $i \sim j$ for regions i and j which are contiguous and assume regions to be neighbours if they share a common border. N_i denotes the number of neighbours of region i. The spatial hyperparameter ψ determines the degree of spatial dependence, for $\psi = 0$ independence of the spatial effects is obtained whereas for $\psi \to \infty$ the degree of spatial dependency increases. Note, that this prior is a proper distribution in contrast to the well known intrinsic CAR model introduced by Besag and Kooperberg (1995). Other proper spatial prior distributions have been considered, see for example Czado and Prokopenko (2004) who use a modification of Model (2.3) and Sun et al. (2000).

Therefore, we have a multivariate normal prior distribution for the regression and spatial parameters $\boldsymbol{\alpha}$ which is given by

$$\alpha|\boldsymbol{\theta} \sim N_{p+1+J}(0, \Sigma) \tag{2.4}$$

with $\Sigma = \begin{pmatrix} V_0 & 0 \\ 0 & \sigma^2 Q^{-1} \end{pmatrix}$. For the spatial hyperparameters $\boldsymbol{\theta} = (\psi, \sigma^2)$ the proper prior distributions

$$\psi \sim \frac{1}{(1+\psi)^2} \quad \text{and} \quad \sigma^2 \sim IGamma(1, 0.005)$$

are assumed. The parameterisation of this model described by Observation Equation (2.2) and Prior Specification (2.4) is called non-centered in the mean, since the intercept β_0 appears in the observation equation, but not in the spatial prior formulation. Other possible model parameterisations include parameterisations additionally non-centered in the scale or variance of the spatial prior as well as a centered parameterisation, where the intercept β_0 only appears as the mean of the spatial prior. These parameterisations are summarized in Table 1. For a summary on existing parameterisation techniques see for example Frühwirth-Schnatter (2004). Initially, our investigations are based on the non-centered mean parameterisation given by (2.2)

parameterisation	spatial prior	observation equation
centered	$\boldsymbol{\gamma}^c \sim N(\beta_0, \sigma^2 Q^{-1})$	$\mu_i = t_i \exp(\mathbf{x}'_{i-0}\boldsymbol{\beta}_{-0} + \mathbf{v}'_i\boldsymbol{\gamma}^c)$
non-centered mean	$\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1})$	$\mu_i = t_i \exp(\beta_0 + \mathbf{x}'_{i-0}\boldsymbol{\beta}_{-0} + \mathbf{v}'_i\boldsymbol{\gamma})$
non-centered mean and scale	$\boldsymbol{\gamma}^* \sim N(0, Q^{-1})$	$\mu_i = t_i \exp(\beta_0 + \mathbf{x}'_{i-0}\boldsymbol{\beta}_{-0} + \sigma\mathbf{v}'_i\boldsymbol{\gamma}^*)$
non-centered mean and variance	$\boldsymbol{\gamma}^{**} \sim N(0, I)$	$\mu_i = t_i \exp(\beta_0 + \mathbf{x}'_{i-0}\boldsymbol{\beta}_{-0} + \sigma\mathbf{v}'_i L\boldsymbol{\gamma}^{**})$ where $LL' = Q^{-1}$

Table 1: Spatial prior and observation equation for different model parameterisations, where $\mathbf{x}_{i-0} := (x_{i1}, .., x_{ip})'$ and $\boldsymbol{\beta}_{-0} := (\beta_1, .., \beta_p)$

and (2.4). Necessary changes when other parameterisations are used will be indicated specifically.

2.1 Step 1: Introduction of hidden inter-arrival times

The basic idea of the data augmentation scheme developed by Frühwirth-Schnatter and Wagner (2004b) is to regard the Poisson observations $y_i, i = 1, .., n$, as the number of jumps of an unobserved Poisson process with intensity μ_i within the unit interval. The first step of the data augmentation consists in the introduction of y_i+1 hidden inter-arrival times $\tau_{ij}, j = 1, .., y_i+1$ for each observation y_i. Using that the inter-arrival times are independent and follow an exponential distribution with parameter μ_i, see for example Mikosch (2004), i.e. $\tau_{ij}|\boldsymbol{\alpha} \sim Exponential(\mu_i) = \frac{Exponential(1)}{\mu_i}$, we obtain

$$\log \tau_{ij}|\boldsymbol{\alpha} \quad = \quad -\log t_i - \mathbf{z}'_i\boldsymbol{\alpha} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim \log(Exponential(1)). \tag{2.5}$$

Denote by $\boldsymbol{\tau} = \{\tau_{ij}, i = 1, .., n, j = 1, .., y_i + 1\}$ the collection of all inter-arrival times. Since the posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\tau}$ is independent of $\mathbf{y}$, conditional on $\boldsymbol{\tau}$ we are now dealing with model (2.5) which is linear in the parameters $\boldsymbol{\alpha}$, but still has a non-normal error term.

2.2 Step 2: Mixture approximation for error term

The second step of the data augmentation scheme eliminates the non-normality of model (2.5). As shown by Frühwirth-Schnatter and Wagner (2004b), the error term in (2.5) can be approx-

imated sufficiently close to a normal distribution by a mixture of five normal distributions, i.e.

$$p(\epsilon_{ij}) = \exp(\epsilon_{ij} - \exp(\epsilon_{ij})) \approx \sum_{r=1}^{5} w_r f_N(\epsilon_{ij}; m_r, s_r^2),$$

where $f_N(\cdot; m_r, s_r^2)$ denotes the density of the normal distribution with mean m_r and variance s_r^2. Frühwirth-Schnatter and Wagner (2004b) also give the corresponding values for m_r, s_r^2 and the weights w_r.
In the second step of the data augmentation the component indicators $r_{ij} \in \{1, .., 5\}$ are introduced as latent variables. Denoting the set of all component indicators by $\mathbf{R} = \{r_{ij}, i = 1, .., n, j = 1, .., y_i + 1\}$, we have conditional on $\mathbf{R}$

$$\log \tau_{ij} | \boldsymbol{\alpha}, r_{ij} = -\log t_i - \mathbf{z}_i' \boldsymbol{\alpha} + m_{r_{ij}} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, s_{r_{ij}}^2),$$

i.e. we are dealing with a normal model which is linear in $\boldsymbol{\alpha}$ now. Since the prior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{\theta})$ is normal as well, the resulting posterior distribution is multivariate normal and a Gibbs sampler can be applied. Note, that by performing this data augmentation we are no longer dealing with n but with $\sum_{i=1}^{n}(y_i + 1)$ observations. Therefore this Gibbs Sampler is mainly useful for count data with small counts only, otherwise the data set might get very large.

2.3 Algorithmic scheme

The algorithmic scheme for the above Gibbs Sampler is the following:
Choose appropriate starting values for the component indicators $\mathbf{R}$ and the inter-arrival times $\boldsymbol{\tau}$.

(1) sample regression and spatial parameters $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ given $\boldsymbol{\tau}, \mathbf{R}, \boldsymbol{\theta}$

(2) sample spatial hyperparameters $\boldsymbol{\theta} = (\psi, \sigma^2)$ given $\boldsymbol{\alpha}$

(3) sample the inter-arrival times τ_{ij} given $\boldsymbol{\alpha}, \mathbf{y}$

(4) sample the component indicators r_{ij} given $\boldsymbol{\tau}, \boldsymbol{\alpha}$

Step (1) consists of sampling from a multivariate normal distribution. This can be done in one block, however it might be computationally more efficient to perform an update in several smaller blocks. We will consider several update strategies for step (1) in Section 3 in more detail. The spatial hyperparameter ψ is updated using a MH step, whereas σ^2 can be updated using a Gibbs step. Steps (3) and (4), elaborated in Frühwirth-Schnatter and Wagner (2004b), are described in the Appendix A and B, for the choice of starting values see Appendix C.

3 Updating schemes for the regression and spatial parameters in the Gibbs Sampler

For $\boldsymbol{\alpha}$ several update schemes are possible and will be discussed in this section. For notational convenience we define with $N := \sum_{i=1}^{n}(y_i + 1)$

$$\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_1, .., \tilde{\tau}_N) := (\tau_{11}, .., \tau_{1,y_1+1}, \tau_{21}, .., \tau_{2,y_2+1}, .., \tau_{n1}, .., \tau_{n,y_n+1}),$$

$$\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, .., \tilde{\epsilon}_N) := (\epsilon_{11}, .., \epsilon_{1,y_1+1}, \epsilon_{21}, .., \epsilon_{2,y_2+1}, .., \epsilon_{n1}, .., \epsilon_{n,y_n+1}),$$

$$\tilde{\boldsymbol{m}} = (\tilde{m}_1, .., \tilde{m}_N) := (m_{r_{11}}, .., m_{r_{1,y_1+1}}, m_{r_{21}}, .., m_{r_{2,y_2+1}}, .., m_{r_{n1}}, .., m_{r_{n,y_n+1}})$$

and

$$\tilde{\boldsymbol{s}}^{\mathbf{2}} = (\tilde{s}_1^2, .., \tilde{s}_N^2) := (s_{r_{11}}^2, .., s_{r_{1,y_1+1}}^2, s_{r_{21}}^2, .., s_{r_{2,y_2+1}}^2, .., s_{r_{n1}}^2, .., s_{r_{n,y_n+1}}^2).$$

Let $\tilde{\boldsymbol{t}} = (\tilde{t}_1, .., \tilde{t}_N)$ denote the vector where t_i is repeated $y_i + 1$ times. Further define

$$\tilde{\mathbf{y}} = (\tilde{y}_1, .., \tilde{y}_N) := (\log \tilde{\tau}_1 - \tilde{m}_1 + \log \tilde{t}_1, .., \log \tilde{\tau}_N - \tilde{m}_N + \log \tilde{t}_N).$$

Using this notation we have according to (2.5)

$$\tilde{y}_i | \boldsymbol{\alpha}, \mathbf{R} \sim N(-\tilde{\mathbf{z}}_i' \boldsymbol{\alpha}, \tilde{s}_i^2)$$

where $\tilde{\mathbf{z}} = \begin{pmatrix} \tilde{\mathbf{z}}_1' \\ \vdots \\ \tilde{\mathbf{z}}_N' \end{pmatrix}$ is a $N \times (p + 1 + J)$-matrix where $\mathbf{z}_i$ is repeated $y_i + 1$ times.

3.1 Block update of $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$

For a joint update of the regression parameters $\boldsymbol{\beta}$ and the spatial effects $\boldsymbol{\gamma}$ in one block we have to consider the full conditional of $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ which is given by

$$\begin{aligned} p(\boldsymbol{\alpha} | \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) &\propto \pi(\boldsymbol{\alpha} | \boldsymbol{\theta}) \prod_{i=1}^{N} \exp\Big(-\frac{1}{2\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{z}}_i' \boldsymbol{\alpha})^2\Big) \\ &\propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\alpha}' \Sigma^{-1} \boldsymbol{\alpha} + \sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{z}}_i' \boldsymbol{\alpha})^2\Big]\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\alpha}' \Sigma_\alpha \boldsymbol{\alpha} - 2\boldsymbol{\alpha}' \boldsymbol{\mu}_\alpha\Big]\Big\}, \end{aligned}$$

where $\Sigma_\alpha := \Sigma^{-1} + \sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2} \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i'$ and $\boldsymbol{\mu}_\alpha := -\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2} \tilde{\mathbf{z}}_i \tilde{y}_i$.
Hence,

$$\boldsymbol{\alpha} | \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R} \sim N_{p+1+J}(\Sigma_\alpha^{-1} \boldsymbol{\mu}_\alpha, \Sigma_\alpha^{-1}).$$

3.2 Separate update of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$

The calculation of the posterior covariance matrix Σ_α^{-1} in Section 3.1 can be computationally expensive if the number of regression parameters and spatial effects is large as is the case in most spatial applications. Therefore it might be more efficient to update $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in two separate blocks. The full conditional distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are given by

$$\boldsymbol{\beta}|\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R} \sim N_{p+1}(\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta, \Sigma_\beta^{-1}) \quad \text{and} \quad \boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R} \sim N_J(\Sigma_\gamma^{-1}\boldsymbol{\mu}_\gamma, \Sigma_\gamma^{-1}).$$

The explicit formulas for Σ_β, $\boldsymbol{\mu}_\beta$, Σ_γ and $\boldsymbol{\mu}_\gamma$ are given in Table 2.

3.3 Block update of the intercept β_0 and $\boldsymbol{\gamma}$ (*block*)

Due to identifiability problems between the intercept β_0 and the spatial effects $\boldsymbol{\gamma}$ mixing and convergence is not very good when $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are updated in two separate blocks. Better results are achieved if a joint block update of β_0 and $\boldsymbol{\gamma}$ is performed, whereas the remaining parameters $\boldsymbol{\beta}_{-0} = (\beta_1, .., \beta_p)$ are still updated in one separate block. With this setting the posterior distributions are given by

$$\boldsymbol{\beta}_{-0}|\beta_0,\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R} \sim N_p(\Sigma_{\beta_{-0}}^{-1}\boldsymbol{\mu}_{\beta_{-0}}, \Sigma_{\beta_{-0}}^{-1})$$

and

$$\boldsymbol{\gamma},\beta_0|\boldsymbol{\beta}_{-0},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R} \sim N_{J+1}(\Sigma_{\gamma\beta_0}^{-1}\boldsymbol{\mu}_{\gamma\beta_0}, \Sigma_{\gamma\beta_0}^{-1})$$

with $\Sigma_{\beta_{-0}}$, $\boldsymbol{\mu}_{\beta_{-0}}$, $\Sigma_{\gamma\beta_0}$ and $\boldsymbol{\mu}_{\gamma\beta_0}$ as given in Table 2.

3.4 Collapsed algorithm for a model parameterisation with a non-centered mean (*coll1*)

Another possibility is to use a collapsed algorithm. This means, that particular components of the posterior are integrated out and an update based on the marginal distribution is performed. In our context the joint posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be written as

$$p(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R}) \propto p(\boldsymbol{\beta}|\boldsymbol{\tau},\mathbf{R})p(\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R})$$

where $p(\boldsymbol{\beta}|\boldsymbol{\tau},\mathbf{R}) = \int p(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R})d\boldsymbol{\gamma}$ is the marginalised posterior density of $\boldsymbol{\beta}$ with $\boldsymbol{\gamma}$ integrated out. It is shown in the Appendix D that

$$\boldsymbol{\beta}|\boldsymbol{\tau},\mathbf{R} \sim N_{p+1}(\Sigma_{col}^{-1}\boldsymbol{\mu}_{col}, \Sigma_{col}^{-1})$$

with Σ_{col} and $\boldsymbol{\mu}_{col}$ as given in Table 2.

Step (1) in the algorithmic scheme presented in Section 2.3 is then the following for the collapsed algorithm:

- sample $\boldsymbol{\beta}$ from $N_{p+1}(\Sigma_{col}^{-1}\boldsymbol{\mu}_{col}, \Sigma_{col}^{-1})$
- sample $\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R}$ as in Section 3.2

Section	
3.2	$\Sigma_\beta := V_0^{-1} + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i'$ $\boldsymbol{\mu}_\beta := -\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i(\tilde{y}_i + \gamma_{R(i)})$ $\Sigma_\gamma := \frac{1}{\sigma^2}Q + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'$ $\boldsymbol{\mu}_\gamma := -\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$
3.3 (*block*)	$\Sigma_{\beta_{-0}} := V_{0\beta_{-0}}^{-1} + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_{\beta_{-0}i}\tilde{\mathbf{x}}_{\beta_{-0}i}'$ $\boldsymbol{\mu}_{\beta_{-0}} := -\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_{\beta_{-0}i}(\tilde{y}_i + \gamma_{R(i)} + \beta_0)$ $\Sigma_{\gamma\beta_0} := \begin{pmatrix} \tau^{-2} & 0 \\ 0 & \frac{1}{\sigma^2}Q \end{pmatrix} + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}(1, \tilde{\mathbf{v}}_i)(1, \tilde{\mathbf{v}}_i)'$ $\boldsymbol{\mu}_{\gamma\beta_0} := -\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}(1, \tilde{\mathbf{v}}_i)(\tilde{y}_i + \tilde{\mathbf{x}}_{\beta_{-0}i}'\boldsymbol{\beta}_{-0})$ $V_{0\beta_{-0}} = \tau^2 I_p$ $\tilde{\mathbf{x}}_{\beta_{-0}i} = (\tilde{x}_{i1}, .., \tilde{x}_{ip})$
3.4 (*coll1*)	$\Sigma_{col} := \tau^{-2}I + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' - (\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')' A^{-1} (\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')$ $\boldsymbol{\mu}_{col} := (\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')' A^{-1} (\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i) - \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i$ $A := \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + \sigma^{-2}Q$
3.5 (*coll2*)	$\Sigma_{col}^* := \tau^{-2}I + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' - (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')' (A^*)^{-1} (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')$ $\boldsymbol{\mu}_{col}^* := (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')(A^*)^{-1} (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i') - \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i$ $A^* := \sigma^2 \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + Q$ $\Sigma_\gamma^* := \sigma^2 \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + Q$ $\boldsymbol{\mu}_\gamma^* := -\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$ $\Sigma_\sigma^* := \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}(\gamma_{R(i)}^*)^2 + \tau_\sigma^{-2}$ $\mu_\sigma^* := -\sum_{i=1}^N \gamma_{R(i)}^* \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$
3.6 (*coll3*)	$\Sigma_{col}^{**} := \tau^{-2}I + \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' - (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')' (A^{**})^{-1} (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')$ $\boldsymbol{\mu}_{col}^{**} := (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')(A^{**})^{-1} (\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{y}_i') - \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i$ $A^{**} := \sigma^2 \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'L + I$ $\Sigma_\gamma^{**} := \sigma^2 \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'L + I$ $\boldsymbol{\mu}_\gamma^{**} := -\sigma \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$ $\Sigma_\sigma^{**} := \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}(\boldsymbol{\gamma}^{**\prime}(\sum_{i=1}^N \frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'L)\boldsymbol{\gamma}^{**} + \tau_\sigma^{-2}$ $\mu_\sigma^{**} := -\sum_{i=1}^N \mathbf{v}_i'L\boldsymbol{\gamma}^{**} \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$
3.7 (*centered*)	$\boldsymbol{\mu}_\gamma^{cent} := \frac{\beta_0}{\sigma^2}Q\mathbf{1} - \sum_{i=1}^N \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_{\beta_{-0}i}'\boldsymbol{\beta}_{-0})$ $\Sigma_{\beta_0} := \frac{1}{\sigma^2}\sum_{i,j=1}^J Q_{ij} + \frac{1}{\tau^2}$ $\mu_{\beta_0} := \frac{1}{\sigma^2}\mathbf{1}'Q\boldsymbol{\gamma}^c$

Table 2: Covariance and mean specifications for the update strategies in Sections 3.2-3.7.

3.5 Collapsed algorithm for a model parameterisation with a non-centered mean and scale (*coll2*)

Up to now, we only considered models with the non-centered mean parameterisation specified by (2.2) and the spatial prior $\boldsymbol{\gamma}|\psi, \sigma \sim N_J(0, \sigma^2 Q^{-1})$. In this section we consider a model where the prior of the spatial effects is not only non-centered in the mean, but in the scale as well, i.e. the third model parameterisation given in Table 1. By assuming $\boldsymbol{\gamma}^*|\psi \sim N_J(0, Q^{-1})$, σ appears as an unknown parameter in the observation equation, in particular we have

$$\mu_i = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \sigma\gamma^*_{R(i)}).$$

For this parameterisation and $\pi(\cdot)$ denoting the prior distributions, the joint posterior of $\boldsymbol{\beta}, \boldsymbol{\gamma}^*, \psi$ and σ is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}^*, \psi, \sigma|\tilde{y}, \boldsymbol{\tau}, \mathbf{R}) \propto \exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \sigma\tilde{\mathbf{v}}_i'\boldsymbol{\gamma}^*)^2\Big\}\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}^*|\psi)\pi(\psi)\pi(\sigma).$$

Following the lines of Section 3.4 we obtain for $\boldsymbol{\beta}$ the marginalized posterior distribution

$$\boldsymbol{\beta}|\sigma, \boldsymbol{\tau}, \mathbf{R} \sim N_{p+1}((\Sigma^*_{col})^{-1}\boldsymbol{\mu}^*_{col}, (\Sigma^*_{col})^{-1}).$$

The full conditional distribution for $\boldsymbol{\gamma}^*$ is given by

$$\boldsymbol{\gamma}^*|\boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{R}, \sigma, \psi \quad \sim \quad N_J((\Sigma^*_{\gamma})^{-1}\boldsymbol{\mu}^*_{\gamma}, (\Sigma^*_{\gamma})^{-1}).$$

The definitions of Σ^*_{col}, $\boldsymbol{\mu}^*_{col}$, Σ^*_{γ} and $\boldsymbol{\mu}^*_{\gamma}$ can be found in Table 2. The spatial hyperparameter ψ is again updated using a MH step since the full conditional distribution can not be sampled from directly. For this model parameterisation we choose a normal prior for σ, in particular $\sigma \sim N(0, \tau_\sigma^2)$. Note, that σ is not restricted to take positive values, leading to nonidentifiability, since the same likelihood results for $(\sigma, \boldsymbol{\gamma}^*)$ and $(-\sigma, -\boldsymbol{\gamma}^*)$. However, as pointed out by Frühwirth-Schnatter (2004), this leads to an improved mixing for models with small scales σ^2 since boundary problems for σ are avoided. The full conditional distribution of σ is then again normal, in particular

$$\sigma|\boldsymbol{\beta}, \boldsymbol{\gamma}^*, \boldsymbol{\tau}, \mathbf{R} \quad \sim \quad N((\Sigma^*_{\sigma})^{-1}\mu^*_{\sigma}, (\Sigma^*_{\sigma})^{-1}),$$

see Table 2 for details on Σ^*_{σ} and μ^*_{σ}.

3.6 Collapsed algorithm for a model parameterisation with a non-centered mean and variance (*coll3*)

In this section we consider the model parameterisation non-centered in both mean and variance, also given in Table 1. In contrast to the non-centered parameterisation in scale only considered in the previous section, we now assume the prior

$$\boldsymbol{\gamma}^{**} \sim N_J(0, I).$$

The spatial structure incorporated in the precision matrix Q is now moved to the observation equation given by

$$\mu_i = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \sigma \mathbf{v}_i' L \boldsymbol{\gamma}^{**}),$$

where L is a lower triangular matrix resulting from the Cholesky decomposition $Q^{-1} = LL'$. The resulting joint posterior distribution of $\boldsymbol{\beta}, \boldsymbol{\gamma}^{**}, \psi$ and σ is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}^{**}, \psi, \sigma | \tilde{y}, \boldsymbol{\tau}, \mathbf{R}) \propto \exp\Big\{-\frac{1}{2}\sum_{i=1}^{n} \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \sigma \tilde{\mathbf{v}}_i' L \boldsymbol{\gamma}^{*})^2\Big\} \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}^{**})\pi(\psi)\pi(\sigma).$$

The marginalized posterior distribution of $\boldsymbol{\beta}$ changes to

$$\boldsymbol{\beta} | \sigma, \boldsymbol{\tau}, \mathbf{R} \sim N_{p+1}((\Sigma_{col}^{**})^{-1}\boldsymbol{\mu}_{col}^{**}, (\Sigma_{col}^{**})^{-1}),$$

the full conditional distribution of $\boldsymbol{\gamma}^{**}$ is given by

$$\boldsymbol{\gamma}^{**} | \boldsymbol{\beta}, \boldsymbol{\tau}, \mathbf{R}, \sigma, \psi \quad \sim \quad N_J((\Sigma_{\gamma}^{**})^{-1}\boldsymbol{\mu}_{\gamma}^{**}, (\Sigma_{\gamma}^{**})^{-1}),$$

with Σ_{col}^{**}, $\boldsymbol{\mu}_{col}^{**}$, Σ_{γ}^{**} and $\boldsymbol{\mu}_{\gamma}^{**}$ as given in Table 2. While ψ is again updated using a MH step, the full conditional distribution of σ is given by

$$\sigma | \boldsymbol{\beta}, \boldsymbol{\gamma}^{**}, \psi, \boldsymbol{\tau}, \mathbf{R} \quad \sim \quad N((\Sigma_{\sigma}^{**})^{-1}\mu_{\sigma}^{**}, (\Sigma_{\sigma}^{**})^{-1}),$$

see Table 2 for details on Σ_{σ}^{**} and μ_{σ}^{**}. Here again the normal prior $\sigma \sim N(0, \tau_{\sigma}^2)$ is assumed.

3.7 Centered CAR-Model (*centered*)

Alternatively, the centered spatial prior $\boldsymbol{\gamma}^c | \beta_0 \sim N(\beta_0, \sigma^2 Q^{-1})$ with $\beta_0 \sim N(0, \tau^2)$ and $\boldsymbol{\beta}_{-0} \sim N(0, \tau^2 I_p)$ can be used. For this model the posterior distribution for $\boldsymbol{\beta}_{-0}$ is the same as in Section 3.3 but with $\boldsymbol{\mu}_{\beta_{-0}}$ replaced by $-\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_{\beta_{-0}i}(\tilde{y}_i + \gamma_{R(i)}^c)$.
The posterior distribution for $\boldsymbol{\gamma}^c$ is given by

$$\boldsymbol{\gamma}^c | \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}, \mathbf{y} \sim N_J(\Sigma_{\gamma}^{-1}\boldsymbol{\mu}_{\gamma}^{cent}, \Sigma_{\gamma}^{-1})$$

where Σ_{γ} is the same as in Section 3.2 and $\boldsymbol{\mu}_{\gamma}^{cent}$ is given in Table 2.
β_0 is updated in an extra Gibbs step, in particular

$$\beta_0 | \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}, \mathbf{y} \sim N(\Sigma_{\beta_0}^{-1}\mu_{\beta_0}, \Sigma_{\beta_0}^{-1})$$

with Σ_{β_0} and μ_{β_0} defined as in Table 2.

4 Simulation studies and application

We aim to apply the developed Gibbs samplers to analyse the expected number of claims in a data set from a German car insurance company. The data include 16307 policyholders in Bavaria with

full comprehensive car insurance within one year and contain information on several covariates like age and gender of the policyholders, kilometers driven per year and the geographical region each policyholder is living in. Bavaria is divided into 96 regions. The variability of these data is very small, 95% of the observations are zero observations, the highest number of claims observed is only four. The data have been already analysed by Gschlößl and Czado (2005) who considered both a spatial Poisson regression model as well as spatial models taking overdispersion into account. They show that the spatial effects are very small for these data and have no significant contribution to explaining the expected claim number.

In this section, the performance of the Gibbs sampler schemes developed in Sections 2 and 3 will be examined on simulated data first. For comparison, we additionally use a single site Metropolis Hastings algorithm for spatial Poisson regression models with an independence proposal where both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are updated component by component. In particular, we use a t-distribution with 20 degrees of freedom as proposal which has the same mode and inverse curvature at the mode as the target distribution.

The performance of the samplers is measured in terms of the computation time required in order to obtain a certain precision of the estimated posterior means of the parameters. The posterior mean of a variable θ is given by $\bar{\theta} := \sum_{j=1}^{R} \hat{\theta}^j$ with $\hat{\theta}^j, j = 1, .., R$ denoting the MCMC iterates of θ after burnin. The precision of $\bar{\theta}$ is given by the Monte Carlo standard error of $\bar{\theta}$ which is defined as $\sigma_{MC}(\bar{\theta}) := \frac{\sigma_{asy}(\bar{\theta})}{\sqrt{R}}$ where

$$\sigma^2_{asy}(\bar{\theta}) := Var(\theta)\Big(1 + 2\sum_{k=1}^{\infty} \rho_k(\theta)\Big)$$

denotes the asymptotic variance of $\bar{\theta}$, $Var(\theta)$ the sample variance and $\rho_k(\theta)$ the autocorrelation of the MCMC iterates $\hat{\theta}^1, .., \hat{\theta}^R$ at lag k. The asymptotic variance will be estimated using the initial monotone sequence estimator (see Geyer (1992)), defined by

$$\hat{\sigma}^2_{asy}(\bar{\theta}) := \hat{Var}(\theta)(1 + 2\sum_{j=1}^{2m+1} \hat{\rho}_k(\theta)),$$

where m is chosen to be the largest integer such that the sequence $\Gamma_m = \hat{\rho}_{2m}(\theta) + \hat{\rho}_{2m+1}(\theta)$ is positive and monotone. Here $\hat{Var}(\theta) := \hat{\gamma}_0$, $\hat{\rho}_k(\theta) := \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$, $\hat{\gamma}_k := \frac{1}{R}\sum_{j=1}^{R-k}(\hat{\theta}^j - \bar{\theta})(\hat{\theta}^{j+k} - \bar{\theta})$. We additionally require the estimated empirical autocorrelations $\hat{\rho}_{2m+1}(\theta)$ to fall below 0.1.

In order to obtain a certain precision k, $R = \frac{\hat{\sigma}^2_{asy}}{k^2}$ samples are needed. Hence, the computation time required to obtain a precision k for an algorithm with computational costs m per iteration, is given by $R \cdot m$. For a direct comparison of the Gibbs sampler schemes to the MH independence sampler we consider the computational costs relative to the costs of the MH sampler required to obtain the same precision of the posterior means of the parameters. This is given by $R_{rel} \cdot m_{rel} := \frac{\hat{\sigma}^2_{asy}}{\hat{\sigma}^2_{asy,ind}} \cdot \frac{m}{m_{ind}}$, where $\hat{\sigma}^2_{asy,ind}$ and m_{ind} denote the estimated asymptotic variance and the computational costs for one iteration of the MH independence sampler.

We consider two studies. In the first study the influence of the size of the spatial effects on mixing behaviour is examined, while in the second study the impact of data heterogeneity is

investigated. In both studies the Gibbs samplers described in Sections 3.3-3.7, i.e. the following model parameterisations and update schemes are assumed:

- non-centered mean:
  - block update of $\boldsymbol{\beta}_{-0}|\beta_0, \boldsymbol{\gamma}$ and $(\beta_0, \boldsymbol{\gamma})|\boldsymbol{\beta}_{-0}$ given in Section 3.3 (*block*)
  - collapsed algorithm given in Section 3.4 (*coll1*)
- non-centered mean and scale: collapsed algorithm given in Section 3.5 (*coll2*)
- non-centered mean and variance: collapsed algorithm given in Section 3.6 (*coll3*)
- centered parameterisation: algorithm given in Section 3.7 (*centered*)

In the following we will refer to these samplers as *block*, *coll1*, *coll2*, *coll3* and *centered*.

4.1 Computational costs

Recall, that by using the data augmentation scheme described above, we are no longer dealing with n observations, but with $N = \sum_{i=1}^{n}(y_i + 1)$ latent inter-arrival times τ_{ij} and mixture component indicators r_{ij}. Both $\boldsymbol{\tau}$ and $\mathbf{R}$ have to be updated, therefore the number of variables to sample from in each iteration is $2N+J+p+1$(+2 hyperparameters) in comparison to $J+p+1$(+2 hyperparameters) variables in the MH independence sampler. The MH independence sampler in contrast requires the calculation of the posterior mode and the inverse curvature at the posterior mode for each of the $J + p + 1$ components in every iteration. The posterior mode may be obtained using the bisection method for example. In our simulation studies, except the sampler *coll3*, the Gibbs samplers are always faster than the MH independence sampler. However, the computational advantage of the Gibbs samplers depends on the complexity of the model. The computational costs m_{rel} relative to the costs of the MH sampler for one iteration are reported in Table 3. For the setting in Study 1 with 5000 observations, an intercept and two covariates for example, the *centered* Gibbs sampler only takes 0.86 times as long as the MH independence sampler. For the setting in Study 2 with a larger data set the *centered* Gibbs sampler even takes only 0.26 times as long. Among the Gibbs samplers the *centered* Gibbs sampler is the fastest, followed closely by the Gibbs sampler using a block update. The collapsed Gibbs samplers non-centered in the mean (*coll1*) and non-centered in mean and scale (*coll2*) require slightly more time than the *centered* Gibbs sampler. The computational effort for the Gibbs sampler in the model parameterisation non-centered in the mean and the variance (*coll3*) however is more than twice as large. In this algorithm a Cholesky decomposition of the precision matrix Q^{-1} has to be performed in every iteration.

4.2 Study 1: Influence of the size of the spatial effects

We consider two simulated data sets of size 5000 with $y_i \sim Poisson(\mu_i), i = 1, .., 5000$. For both data sets the mean μ_i is specified by

$$\mu_i = \exp(\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \gamma_{R(i)})$$

sampler	Study 1	Study 2
independence	1	1
block	0.87	0.27
centered	0.86	0.26
coll1	0.96	0.30
coll2	0.99	0.31
coll3	2.18	0.57

Table 3: Computation times m_{rel} for the different samplers relative to the MH independence sampler for the settings in Study 1 and Study 2.

where $\mathbf{x}_1$ is an indicator variable and $\mathbf{x}_2$ a continuous standardized variable. The exposure is assumed to be $t_i = 1$ for all observations. We assume a simple spatial structure, namely 100 regions on a 10×10 grid. The spatial effects $\boldsymbol{\gamma}$ are generated according to the CAR prior $\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1})$ with spatial dependence parameter $\psi = 3$. For the first simulated data set $\mathbf{y}_1$ we assume $\sigma^2 = 1$ resulting in a range of $[\min(\boldsymbol{\gamma}) \max(\boldsymbol{\gamma})] = [-0.86, 0.85]$ for the spatial effects, whereas for the second data set $\mathbf{y}_2$ we take $\sigma^2 = 0.01$ resulting in a range of $[\min(\boldsymbol{\gamma}) \max(\boldsymbol{\gamma})] = [-0.08, 0.08]$. The Gibbs samplers *block*, *coll1*, *coll2*, *coll3* and *centered* as well as the independence MH sampler are run for 5000 iterations, a burnin of 1000 iterations is taken. As described above, the performance of the samplers is measured in terms of the Monte Carlo standard error of the posterior means of the parameters and the required computation times. Since estimation of the Monte Carlo error is based on the estimated empirical autocorrelations, this quantity also depends on the mixing of the samplers. For a fair comparison of the Monte Carlo error of the spatial effects the model parameterisation of each sampler has to be taken into account. Therefore we compute the Monte Carlo error for $\beta_0 + \boldsymbol{\gamma}$ for the MH independence sampler and the samplers *block* and *collapsed*, while for the *centered* sampler the standard error of $\boldsymbol{\gamma}$ is considered since here the intercept is the spatial prior mean and therefore already included in $\boldsymbol{\gamma}$. For the *coll2* and *coll3* samplers the Monte Carlo errors for $\beta_0 + \sigma\boldsymbol{\gamma}$ and $\beta_0 + \sigma L\boldsymbol{\gamma}$, respectively, are computed.
In the left panel of Table 4, for each sampler the Monte Carlo standard errors and the performance relative to the MH independence sampler $R_{rel} \cdot m_{rel}$ are reported for the regression parameters β_1, β_2 and the spatial effects in data set $\mathbf{y}_1$. For the spatial effects the average error, taken over all J components, is given. Additionally plots of the empirical estimated autocorrelations are presented in Figure 1. In the left panel the autocorrelations for 25 of the spatial effects, in the right panel autocorrelations for the regression effects are plotted. Mixing for all Gibbs samplers is reasonable well, in average autocorrelations of the spatial effects are below 0.1 at a lag of about 16 to 18. The average Monte Carlo error for the spatial effects is around 0.01 for all Gibbs samplers. The Monte Carlo error of the regression parameters however is lower for the collapsed Gibbs samplers, for the *block* and the *centered* Gibbs sampler especially the autocorrelations of β_1 decrease rather slowly.
The independence MH sampler in contrast, displays the smallest Monte Carlo error for both

spatial effects and regression parameters. In average the autocorrelations of $\beta_0 + \gamma_j$ are below 0.1 at a lag 3 already, the autocorrelations for the regression parameters decrease rapidly as well. Considering the computational effort relative to the MH independence sampler, given by $R_{rel} \cdot m_{rel}$, the MH independence sampler outperforms the Gibbs samplers considerably. The computational effort required to obtain the same precision of the posterior means of the spatial effects is more than 5 times as large for the Gibbs samplers compared to the independence sampler.

sampler	Data set $\mathbf{y}_1$			Data set $\mathbf{y}_2$		
	spatial effects	β_1	β_2	spatial effects	β_1	β_2
independence	0.0041	0.0015	0.0032	0.0021	0.0013	0.0030
	1	1	1	1	1	1
block	0.0100	0.0039	0.0130	0.0031	0.0042	0.0108
	5.18	5.88	14.36	1.90	9.08	11.28
centered	0.0102	0.0045	0.0115	0.0061	0.0078	0.0279
	5.32	7.74	11.11	7.26	30.96	74.38
coll1	0.0101	0.0031	0.0117	0.0022	0.0024	0.0097
	5.83	4.10	12.83	1.05	3.27	10.04
coll2	0.0099	0.0027	0.0105	0.0025	0.0029	0.0114
	5.77	3.21	10.66	1.40	4.93	14.30
coll3	0.0101	0.0024	0.0102	0.0023	0.0026	0.0133
	13.23	5.58	22.15	2.62	8.72	42.85

Table 4: Estimated $\hat{\sigma}_{MC}$ (upper row) for the regression parameters β_1, β_2 and average estimated $\hat{\sigma}_{MC}$ for the spatial effects $\boldsymbol{\gamma} + \beta_0$ in the *independence*, *block*, *coll1* sampler, $\boldsymbol{\gamma}$ in the *centered*, $\beta_0 + \sigma\boldsymbol{\gamma}$ in the *coll2* and $\beta_0 + \sigma L\boldsymbol{\gamma}$ in the *coll3* sampler, as well as $R_{rel} \cdot m_{rel}$ (lower row) for all parameters for data set $\mathbf{y}_1$ and $\mathbf{y}_2$ using different update strategies in Study 1.

The corresponding results for data set $\mathbf{y}_2$ with small spatial effects are reported in the right panel in Table 4, plots of the estimated empirical autocorrelations are given in Figure 1. Here, clearly the lowest precision and worst mixing is obtained if the Gibbs sampler based on the centered model parameterisation is used. This confirms the results given in Gelfand et al. (1995). They show that for a hierarchical normal linear model with random effects the centered parameterisation is efficient if the variance of the random effects dominates the variance in the data. However, if the variance of the random effects is very small in contrast to the variability of the data (as it is the case in data set $\mathbf{y}_2$), high posterior correlations result. For the *block* and particularly the collapsed Gibbs samplers a considerably lower Monte Carlo error is obtained. The average Monte Carlo error of the spatial effects in the collapsed sampler *coll1* is almost as small as in the MH independence sampler. For the regression effects however, the MH independence sampler

exhibits lower Monte Carlo standard errors. The computational costs $R_{rel} \cdot m_{rel}$ relative to the MH sampler, which are required to obtain the same precision of the posterior means of the parameters are greater 1 for all Gibbs samplers for both spatial effects and regression parameters. Hence, the independence sampler gives the best performance for data set $\mathbf{y}_2$ as well.
The variance of the two simulated data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ takes the values $\text{var}(\mathbf{y}_1) = 0.51$ and $\text{var}(\mathbf{y}_2) = 0.49$. However, the variability of our real data from a car insurance company is very small, the variance of these data is only 0.05. Therefore we conduct a second simulation study where we examine whether the heterogeneity of the data influences the performance of the samplers.

4.3 Study 2: Influence of data heterogeneity

We simulate two data sets based on the design of the real data where, according to Gschlößl and Czado (2005), eight covariates significant for explaining the expected claim number y_i were observed, i.e. $y_i \sim Poisson(\mu_i), i = 1, .., 16307$ with

$$\mu_i = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_{R(i)}).$$

Here $\mathbf{x}_i = (1, x_{i1}, .., x_{i8})$ and $x_{ik}, k = 1, .., 8$ are standardized categorical and metrical covariates, the observation specific exposure t_i takes values up to one year. In this setting we have 96 irregular regions in Bavaria. The spatial effects $\boldsymbol{\gamma}$ again are generated according to the CAR prior $\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1})$ with $\psi = 8$ and $\sigma^2 = 0.01$. This results in small spatial effects with a range of $[-0.06 \quad 0.08]$, i.e. spatial effects similar to the ones observed in our real data set. For the first data set $\mathbf{y}_3$ the intercept β_0 is taken to be -1, whereas for the second data set $\mathbf{y}_4$ we take $\beta_0 = -2.5$. For the remaining regression parameters the same values are assumed for both data sets. The resulting variances of $\mathbf{y}_3$ and $\mathbf{y}_4$ are $Var(\mathbf{y}_3) = 0.46$ and $Var(\mathbf{y}_4) = 0.05$, i.e. data set $\mathbf{y}_4$ has very low heterogeneity and is close to our real data. The variance of data set $\mathbf{y}_3$ is not particularly high either, but in comparison to data set $\mathbf{y}_4$ we will refer to this data set as data with high heterogeneity.
The *block*, *centered*, *coll1*, *coll2* and *coll3* Gibbs samplers are run for 5000 iterations, the first 1000 iterations are discarded for burnin. For comparison again the MH independence sampler is applied. The Monte Carlo errors for the posterior means of the regression parameters $\beta_1, .., \beta_8$, the spatial effects $\boldsymbol{\gamma}$ in the centered, $\beta_0 + \boldsymbol{\gamma}$ in the non-centered mean, $\beta_0 + \sigma\boldsymbol{\gamma}$ in the non-centered mean and scale and $\beta_0 + \sigma L\boldsymbol{\gamma}$ in the non-centered mean and variance model parameterisation and the quantities $R_{rel} \cdot m_{rel}$ are reported in Table 5. For the high heterogeneity data set $\mathbf{y}_3$ the collapsed Gibbs samplers *coll2* and *coll3* exhibit the lowest Monte Carlo errors for the spatial effects among the Gibbs samplers. The sampler *coll2* even only requires 38 % of the computational effort of the MH sampler in order to obtain the same precision for the spatial effects. The precision and autocorrelations of the regression effects however are considerably smaller in the independence sampler compared to all Gibbs samplers. The precision and autocorrelations (see Figure 2) of the regression effects however are considerably smaller in the independence sampler compared to all Gibbs samplers. In order to achieve a high precision like in the MH sampler

for all parameters, for each Gibbs sampler the maximum relative effort $R_{rel} \cdot m_{rel}$, occurring for spatial and regression parameters, is required. Since the maximum values $R_{rel} \cdot m_{rel}$ are considerably greater than 1 for each Gibbs sampler, the MH sampler is clearly superior to the Gibbs samplers.

The average Monte Carlo error for the spatial effects in data set $\mathbf{y}_4$ with low heterogeneity is rather high for the three Gibbs sampler schemes *block*, *centered* and *coll1* for both spatial effects and regression parameters, the estimated empirical autocorrelations plotted in Figure 2 decrease very slowly. While for the high heterogeneity data $\mathbf{y}_3$ the computational costs in order to obtain the same precision for the spatial effects of the *block* Gibbs sampler are only 0.65 times as large as of the MH sampler, for the data $\mathbf{y}_4$ the performance of the Gibbs sampler is clearly worse with $R_{rel} \cdot m_{rel} = 5.37$. Results are improved for the collapsed algorithms based on the model parameterisations non-centered in the scale (*coll2*) and in the variance (*coll3*). The sampler *coll2* performs even better than the MH sampler ($R_{rel} \cdot m_{rel} = 0.45$). As indicated in Section 3.5, the model parameterisation with non-centered scale is supposed to improve mixing particularly for models with small scale σ^2 which is the case for data sets $\mathbf{y}_3$ and $\mathbf{y}_4$. However, the Monte Carlo errors for the regression parameters are rather high for all Gibbs samplers and in particular considerably higher than for the high heterogeneity data $\mathbf{y}_3$. The MH independence sampler in contrast exhibits a high precision for all parameters again. Compared to data set $\mathbf{y}_3$, the standard errors for all parameters resulting from the MH sampler are about twice as large for data set $\mathbf{y}_4$, this loss of precision however is much smaller than for the Gibbs samplers. According to the performance measure $R_{rel} \cdot m_{rel}$ for the regression parameters, the MH sampler outperforms the Gibbs samplers considerably. For example, although the Gibbs sampler *coll2* sampler only requires 31 % of the computation time of the MH sampler for one iteration (see Table 3), 30.33 ($R_{rel} \cdot m_{rel}$ for β_2) times the effort of the MH sampler for data set $\mathbf{y}_4$ would be needed in order to obtain for all parameters a precison comparable to the MH sampler.
Note that, compared to the collapsed algorithm *coll2*, the collapsed algorithm *coll3* does not display significantly lower standard errors, neither in Study 1 nor in Study 2. The additional computational effort required for *coll3* which is more than twice as large as for *coll2*, see Table 3, does not pay off.

4.4 Application to car insurance data

Finally we apply the discussed Gibbs samplers as well as the independence MH sampler on the car insurance data set described at the beginning of this section. The Monte Carlo errors for the posterior means of the regression and the spatial effects as well as the corresponding values of $R_{rel} \cdot m_{rel}$ are reported in Table 6. Similar results as for data set $\mathbf{y}_4$ which is very close to our real data, are observed. In particular for the regression parameters, the performance of all Gibbs samplers is considerably worse than the performance of the MH independence sampler. When using the non-centered scale and variance parameterisations at least for the spatial effects reasonable low errors are obtained, however, according to the relative effort $R_{rel} \cdot m_{rel}$ the MH sampler is still superior.

data	sampler	spatial effects	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
$\mathbf{y}_3$	ind	0.0020	0.0346	0.0327	0.0003	0.0003	0.0008	0.0010	0.0003	0.0002
		1	1	1	1	1	1	1	1	1
	block	0.0031	0.1771	0.1627	0.0013	0.0022	0.0042	0.0039	0.0015	0.0009
		0.65	7.07	6.68	5.07	14.54	7.44	4.11	6.75	5.47
	centered	0.0036	0.1955	0.1768	0.0015	0.0020	0.0038	0.0045	0.0013	0.0010
		0.84	8.30	7.60	6.50	11.56	5.87	5.27	4.88	6.50
	coll1	0.0040	0.1487	0.1635	0.0011	0.0018	0.0032	0.0032	0.0015	0.0009
		1.20	5.54	7.50	4.03	10.80	4.80	3.07	7.50	6.08
	coll2	0.0022	0.1561	0.1736	0.0014	0.0021	0.0030	0.0031	0.0012	0.0010
		0.38	6.31	8.74	6.75	15.19	4.36	2.98	4.96	7.75
	coll3	0.0024	0.1899	0.1505	0.0014	0.0022	0.0031	0.0028	0.0016	0.0011
		0.82	17.17	12.07	12.41	30.65	8.56	4.47	16.21	17.24
$\mathbf{y}_4$	ind	0.0048	0.0673	0.0611	0.0006	0.0006	0.0017	0.0021	0.0006	0.0005
		1	1	1	1	1	1	1	1	1
	block	0.0214	0.5199	0.3323	0.0038	0.0041	0.0217	0.0211	0.0076	0.0030
		5.37	16.11	7.99	10.83	12.61	43.99	27.26	43.32	9.72
	centered	0.0114	0.5906	0.4910	0.0040	0.0052	0.0150	0.0209	0.0060	0.0055
		1.47	20.02	16.79	11.56	19.53	20.24	25.75	26.00	31.46
	coll1	0.0189	0.6749	0.6181	0.0052	0.0057	0.0129	0.0133	0.0049	0.0049
		4.65	30.17	30.70	22.53	27.08	17.27	12.03	20.01	28.81
	coll2	0.0058	0.5505	0.6044	0.0038	0.0048	0.0076	0.0070	0.0050	0.0041
		0.45	20.74	30.33	12.43	19.84	6.20	3.44	21.53	20.84
	coll3	0.0091	0.5301	0.4789	0.0052	0.0056	0.0097	0.0096	0.0048	0.0044
		2.05	35.36	35.02	42.81	49.65	18.56	11.91	36.48	44.14

Table 5: Estimated $\hat{\sigma}^2_{MC}$ (upper row) for the regression parameters $\beta_1, .., \beta_8$ and estimated average $\hat{\sigma}^2_{MC}$ for the spatial effects $\boldsymbol{\gamma} + \beta_0$ in the *independence*, *block*, *coll1* sampler, $\boldsymbol{\gamma}$ in the *centered*, $\beta_0 + \sigma\boldsymbol{\gamma}$ in the *coll2* and $\beta_0 + \sigma L\boldsymbol{\gamma}$ in the *coll3* sampler, as well as $R_{rel} \cdot m_{rel}$ (lower row) for data set $\mathbf{y}_3$ and $\mathbf{y}_4$ using different update strategies in Study 2.

5 Summary and conclusions

We have presented a new MCMC methodology for spatial Poisson regression models, extending the approach by Frühwirth-Schnatter and Wagner (2004b). Using data augmentation we have shown that a straightforward Gibbs sampler for spatial Poisson models is available. Several update schemes like a joint block update of the intercept and the spatial effects as well as collapsed algorithms have been discussed. Further we have addressed the issue of model parameterisation, centered as well as non-centered model parameterisations in the mean, the scale and the variance have been considered. The performance of the Gibbs sampler based on different model parameterisations and update schemes has been compared to a single site MH independence sampler on simulated and real data. Performance is measured in terms of the computational

sampler	spatial effects	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
independence	0.0046	0.0673	0.0628	0.0006	0.0007	0.0017	0.0020	0.0006	0.0005
	1	1	1	1	1	1	1	1	1
block	0.0192	0.5201	0.3823	0.0051	0.0083	0.0116	0.0203	0.0037	0.0047
	4.70	16.13	10.01	19.51	37.96	12.57	27.82	10.27	23.86
centered	0.0138	0.5465	0.5847	0.0045	0.0068	0.0126	0.0145	0.0048	0.0028
	2.34	17.14	22.54	14.63	24.54	14.28	13.67	16.64	8.15
coll1	0.0207	0.5967	0.5753	0.0040	0.0073	0.0155	0.0082	0.0043	0.0032
	6.08	23.58	25.28	13.33	32.63	24.94	5.04	15.41	12.29
coll2	0.0116	0.4359	0.6172	0.0046	0.0063	0.0122	0.0096	0.0057	0.0044
	1.97	13.00	29.94	18.22	25.11	15.97	7.14	27.98	24.01
coll3	0.0100	0.5167	0.5945	0.0056	0.0060	0.0115	0.0110	0.0054	0.0036
	2.69	33.60	51.08	49.65	41.88	26.08	17.24	46.17	29.55

Table 6: Estimated $\hat{\sigma}_{MC}$ (upper row) for the regression parameters $\beta_1, .., \beta_8$ and average estimated $\hat{\sigma}_{MC}$ for the spatial effects $\boldsymbol{\gamma} + \beta_0$ in the *independence*, *block*, *coll1* sampler, $\boldsymbol{\gamma}$ in the *centered*, $\beta_0 + \sigma\boldsymbol{\gamma}$ in the *coll2* and $\beta_0 + \sigma L\boldsymbol{\gamma}$ in the *coll3* sampler, as well as $R_{rel} \cdot m_{rel}$ (lower row) for the car insurance data using different update strategies.

costs required in order to obtain the same precision of the posterior means of the parameters.
For data which are not too homogeneous, the Gibbs samplers display good mixing and reasonable small Monte Carlo errors. In particular for data with small spatial random effects, the performance is improved when collapsed Gibbs samplers are used, while the centered parameterisation is not very efficient any more in this case. The MH independence sampler however exhibits the smallest Monte Carlo errors for all parameters for data with both small and large spatial effects. Taking additionally the required computation times of the samplers into account, the MH sampler gives the best performance.
For data with low heterogeneity the Monte Carlo errors increase significantly for all Gibbs samplers, mixing of the samplers is much worse. The MH sampler in contrast also mixes well for low heterogeneity data, the precision of the posterior means of the parameters is considerably higher than for the Gibbs samplers. Considering the computation times of the samplers and the required MCMC iterations in order to obtain the same precision for all parameters, the MH sampler clearly outperforms the Gibbs samplers for low heterogeneity data. Similar results are observed for the real data which also display low heterogeneity.
In the literature various approaches for MCMC estimation in spatial Poisson models are provided. Knorr-Held and Rue (2002) discuss efficient block sampling MH algorithms for Markov random field models in disease mapping, based on the methodology developed in Rue (2001). Haran et al. (2003) study MH algorithms with proposal distributions based on Structured MCMC, introduced by Sargent et al. (2000), for spatial Poisson models, while Christensen et al. (2005) discuss Langevin-Hastings updates in spatial GLMM's. Rue et al. (2004) present non-Gaussian

approximations to hidden Markov random fields and give applications in disease mapping and geostatistical models. These methods have been found to be superior to a conventional MH sampler only performing individual updates of the parameters. Therefore, since a single site MH sampler clearly outperformed the Gibbs samplers developed in this paper, a comparison of the Gibbs samplers to these methods seems to be unnecessary. However, the performance of the Gibbs samplers might be improved by applying the reparameterisation techniques presented in Christensen et al. (2005), which is subject of current research.

A Sampling the inter-arrival times

Given $\mathbf{y}$ and $\boldsymbol{\alpha}$, the inter-arrival times for different observations $i = 1, .., n$ are independent. For fixed i however, $\tau_{i1}, .., \tau_{i,y_i+1}$ are stochastically dependent, but independent of the component indicators $\mathbf{R}$. The inter-arrival times $\tau_{i1}, .., \tau_{iy_i}$ are independent of $\boldsymbol{\alpha}$ and only depend on the number of jumps, whereas τ_{i,y_i+1} depends on the model parameters. Using this we have

$$p(\tau|\mathbf{y}, \boldsymbol{\alpha}, \mathbf{R}) = \prod_{i=1}^{n} p(\tau_{i,y_i+1}|y_i, \boldsymbol{\alpha}, \tau_{i1}, .., \tau_{iy_i}) p(\tau_{i1}, .., \tau_{iy_i}|y_i)$$

Given $y_i = n$, the n arrival times of a Poisson process are distributed as the order statistics of n $U([0, 1])$ distributed random variables, see for example Mikosch (2004). The last inter-arrival time τ_{i,y_i+1}, given $y_i, \tau_{i1}, .., \tau_{iy_i}$, is exponentially distributed with mean $\frac{1}{\mu_i} = \frac{1}{t_i \exp(\mathbf{z}_i'\boldsymbol{\alpha})}$ conditionally on being greater than $1 - \sum_{j=1}^{y_i} \tau_{ij}$. Using the lack of memory property of the exponential distribution this corresponds to sampling τ_{i,y_i+1} from an exponential distribution with mean $\frac{1}{\mu_i}$ plus an "offset" $1 - \sum_{j=1}^{y_i} \tau_{ij}$. Therefore the inter-arrival times can be sampled as follows:

- If $y_i > 0$
  - sample y_i random numbers $u_{i1}, .., u_{iy_i} \sim U([0, 1])$
  - sort these random numbers: $u_{i,(1)}, .., u_{i,(y_i)}$
  - define τ_{ij} as the increments $\tau_{ij} = u_{i,(j)} - u_{i,(j-1)}, \quad j = 1, .., y_i$ where $u_{j,(0)} := 0$
  - sample $\tau_{i,y_i+1} = 1 - \sum_{j=1}^{y_i} \tau_{ij} + \zeta_i$, where $\zeta_i \sim Exponential(\mu_i)$
- If $y_i = 0$ sample $\tau_{i1} = 1 + \zeta_i$, where $\zeta_i \sim Exponential(\mu_i)$

B Sampling the component indicators

The component indicators $\mathbf{R}$ are mutually independent given $\boldsymbol{\tau}, \boldsymbol{\alpha}$, therefore $p(\mathbf{R}|\boldsymbol{\tau}, \boldsymbol{\alpha}) = \prod_{i=1}^{n} \prod_{j=1}^{y_i+1} p(r_{ij}|\tau_{ij}, \boldsymbol{\alpha})$. Further

$$\begin{aligned} p(r_{ij} = k|\tau_{ij}, \boldsymbol{\alpha}) &= \frac{p(r_{ij} = k, \tau_{ij}, \boldsymbol{\alpha})}{p(\tau_{ij}, \boldsymbol{\alpha})} = \frac{p(\tau_{ij}|r_{ij} = k, \boldsymbol{\alpha}) p(r_{ij} = k)}{p(\tau_{ij}|\boldsymbol{\alpha})} \\ &\propto p(\tau_{ij}|r_{ij} = k, \boldsymbol{\alpha}) w_k \end{aligned} \tag{2.1}$$

since $w_k = p(r_{ij} = k)$. Since $\log \tau_{ij}|\boldsymbol{\alpha}, r_{ij} \sim N(-\log \mu_i + m_{r_{ij}}, s^2_{r_{ij}})$, τ_{ij} is log normal distributed, i.e.

$$p(\tau_{ij}|r_{ij} = k, \boldsymbol{\alpha}) \propto \frac{1}{s_k \tau_{ij}} \exp\Big[-\frac{1}{2}\Big(\frac{\log(\tau_{ij}) + \log \mu_i - m_k}{s_k}\Big)^2\Big].$$

r_{ij} can therefore be sampled from the discrete distribution (2.1) with five categories.

C Starting values

Starting values for the component indicators r_{ij} are obtained by drawing random numbers from 1 to 5. For τ_{ij} starting values are generated according to the sampling procedure described in Appendix A. For observations equal to zero we sample $\zeta_i \sim Exponential(0.1)$, for observations greater than zero $\zeta_i \sim Exponential(y_i)$, as suggested in Frühwirth-Schnatter and Wagner (2004b).

D Details on algorithm in Section 3.4

For the collapsed algorithm in Section 3.4 we consider $p(\boldsymbol{\beta}|\boldsymbol{\tau}, \mathbf{R}) = \int p(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) d\boldsymbol{\gamma}$. We have

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) &\propto \exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \tilde{\mathbf{v}}_i'\boldsymbol{\gamma})^2 + \boldsymbol{\gamma}'\sigma^{-2}Q\boldsymbol{\gamma} + \boldsymbol{\beta}'\tau^{-2}I\boldsymbol{\beta}\Big]\Big\} \\
&= \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}'\tau^{-2}I\boldsymbol{\beta} + \sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2\Big]\Big\} \\
&\times \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'\Big(\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + \sigma^{-2}Q\Big)\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})\Big]\Big\} \\
&:= c(\boldsymbol{\beta}) \times \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'a\Big]\Big\} \qquad (4.2)
\end{aligned}$$

where $A := \sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + \sigma^{-2}Q$. Further

$$\begin{aligned}
&\exp \Big\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'a\Big]\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'A(A^{-1}a) + (A^{-1}a)'A(A^{-1}a) - (A^{-1}a)'A(A^{-1}a)\Big]\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}\Big[(\boldsymbol{\gamma} + A^{-1}a)'A(\boldsymbol{\gamma} + A^{-1}a) - (A^{-1}a)'A(A^{-1}a)\Big]\Big\}
\end{aligned}$$

and therefore

$$\begin{aligned}
\int \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'a\Big]\Big\} d\boldsymbol{\gamma} &\propto (2\pi)^{\frac{J}{2}}|A|^{-\frac{1}{2}} \exp\Big\{\frac{1}{2}(A^{-1}a)'A(A^{-1}a)\Big\} \\
&\propto \exp\Big\{\frac{1}{2}(A^{-1}a)'A(A^{-1}a)\Big\} \qquad (4.3)
\end{aligned}$$

From (4.2) and (4.3) it then follows that

$$\begin{aligned}
&\int \quad p(\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{\tau},\mathbf{R})d\boldsymbol{\gamma} \\
&\propto \quad c(\boldsymbol{\beta})\exp\Big\{\frac{1}{2}(A^{-1}a)'A(A^{-1}a)\Big\} \\
&\propto \quad \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}'(\tau^{-2}I+\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i')\boldsymbol{\beta}+2\boldsymbol{\beta}'\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{x}_i\tilde{y}_i-a'A^{-1}a\Big]\Big\}
\end{aligned}$$

Finally, with

$$\begin{aligned}
a'A^{-1}a &= \Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i+\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i'\boldsymbol{\beta}\Big)'A^{-1}\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i+\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i'\boldsymbol{\beta}\Big) \\
&\propto \boldsymbol{\beta}'\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i'\Big)'A^{-1}\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i'\Big)\boldsymbol{\beta}+2\boldsymbol{\beta}'\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i'\Big)'A^{-1}\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i\Big)
\end{aligned}$$

it follows that

$$\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{\tau},\mathbf{R}) \quad\propto\quad & \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}'\Big(\tau^{-2}I+\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i'-(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')'A^{-1}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')\Big)\boldsymbol{\beta} \\
&- \quad 2\boldsymbol{\beta}'\Big((\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')'A^{-1}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i)-\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i\Big)\Big]\Big\},
\end{aligned}$$

i.e.

$$\boldsymbol{\beta}|\boldsymbol{\tau},\mathbf{R}\sim N(\Sigma_{col}^{-1}\boldsymbol{\mu}_{col},\Sigma_{col}^{-1})$$

with

$$\Sigma_{col}:=\tau^{-2}I+\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i'-(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')'A^{-1}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')$$

and

$$\boldsymbol{\mu}_{col}:=(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')'A^{-1}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i)-\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i.$$

Acknowledgement

We would like to thank Sylvia Frühwirth-Schnatter for fruitful discussions and helpful comments and suggestions. The first author is supported by a doctoral fellowship within the Graduiertenkolleg *Angewandte Algorithmische Mathematik*, while the second author is supported by Sonderforschungsbereich 386 *Statistische Analyse Diskreter Strukturen*, both sponsored by the *Deutsche Forschungsgemeinschaft*.

References

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

Christensen, O., G. Roberts, and M. Sköld (2005). Robust MCMC methods for spatial GLMM's. *to appear in Journal of Computational and Graphical Statistics*.

Christensen, O. and R. Waagepetersen (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics 58*, 280–286.

Czado, C. and S. Prokopenko (2004). Modeling transport mode decisions using hierarchical binary spatial regression models with cluster effects. *Discussion paper 406, SFB 386 Statistische Analyse diskreter Strukturen*. http://www.stat.uni-muenchen.de/sfb386/.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed (1998). Model-based geostatistics. *J. Roy. Statist. Soc. Ser. C 47*(3), 299–350. With discussion and a reply by the authors.

Dimakos, X. and A. Frigessi (2002). Bayesian premium rating with latent structure. *Scandinavian Actuarial Journal* (3), 162–184.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation. In S. K. A.C. Harvey and N. Shephard (Eds.), *State space and unobserved component models*, pp. 123–151. Cambridge: Cambridge Univ. Press.

Frühwirth-Schnatter, S. and H. Wagner (2004a). Data augmentation and Gibbs sampling for regression models of small counts. *IFAS Research Paper Series 2004-04*.

Frühwirth-Schnatter, S. and H. Wagner (2004b). Gibbs sampling for parameter-driven models of time series of small counts with applications to state space modelling. *IFAS Research Paper Series 2004-01*.

Frühwirth-Schnatter, S. and H. Waldl (2004). Data augmentation and Gibbs sampling for logistic models. *IFAS Research Paper Series 2004-06*.

Gelfand, A., S. Sahu, and B.P.Carlin (1995). Efficient parametrisations for normal linear mixed models. *Biometrika 82 (3)*, 479–488.

Geyer, C. (1992). Practical Markov Chain Monte Carlo. *Statistical Science 7*(4), 473–511.

Gschlößl, S. and C. Czado (2005). Modelling count data with overdispersion and spatial effects. *Discussion paper 412, SFB 386 Statistische Analyse diskreter Strukturen*. http://www.stat.uni-muenchen.de/sfb386/.

Haran, M., J. Hodges, and B. Carlin (2003). Accelerating computation in Markov random field models for spatial data via structured MCMC. *Journal of Computational and Graphical Statistics 12*(2), 249–264.

Knorr-Held, L. and H. Rue (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics. Theory and Applications 29*, 325–338.

Liu, J., W. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika 81 (1)*, 27–40.

Mikosch, T. (2004). *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes*. New York: Springer.

Papaspiliopoulos, O., G. O. Roberts, and M. Sköld (2003). Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian statistics, 7 (Tenerife, 2002)*, pp. 307–326. New York: Oxford Univ. Press. With a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors.

Pettitt, A., I. Weir, and A. Hart (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing 12 (4)*, 353–367.

Roberts, G. and S. Sahu (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society, B 59 (2)*, 291–317.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society. Series B. Statistical Methodology 63*(4), 597–614.

Rue, H., I. Steinsland, and S. Erland (2004). Approximating hidden Gaussian Markov random fields. *Journal of the Royal Statistical Society. Series B. Statistical Methodology 66*(4), 877–892.

Sargent, D., J. Hodges, and B. Carlin (2000). Structured Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics 9*, 217–234.

Sun, D., R. K. Tsutakawa, H. Kim, and Z. He (2000). Bayesian analysis of mortality rates with disease maps. *Statistics in Medicine 19*, 2015–2035.

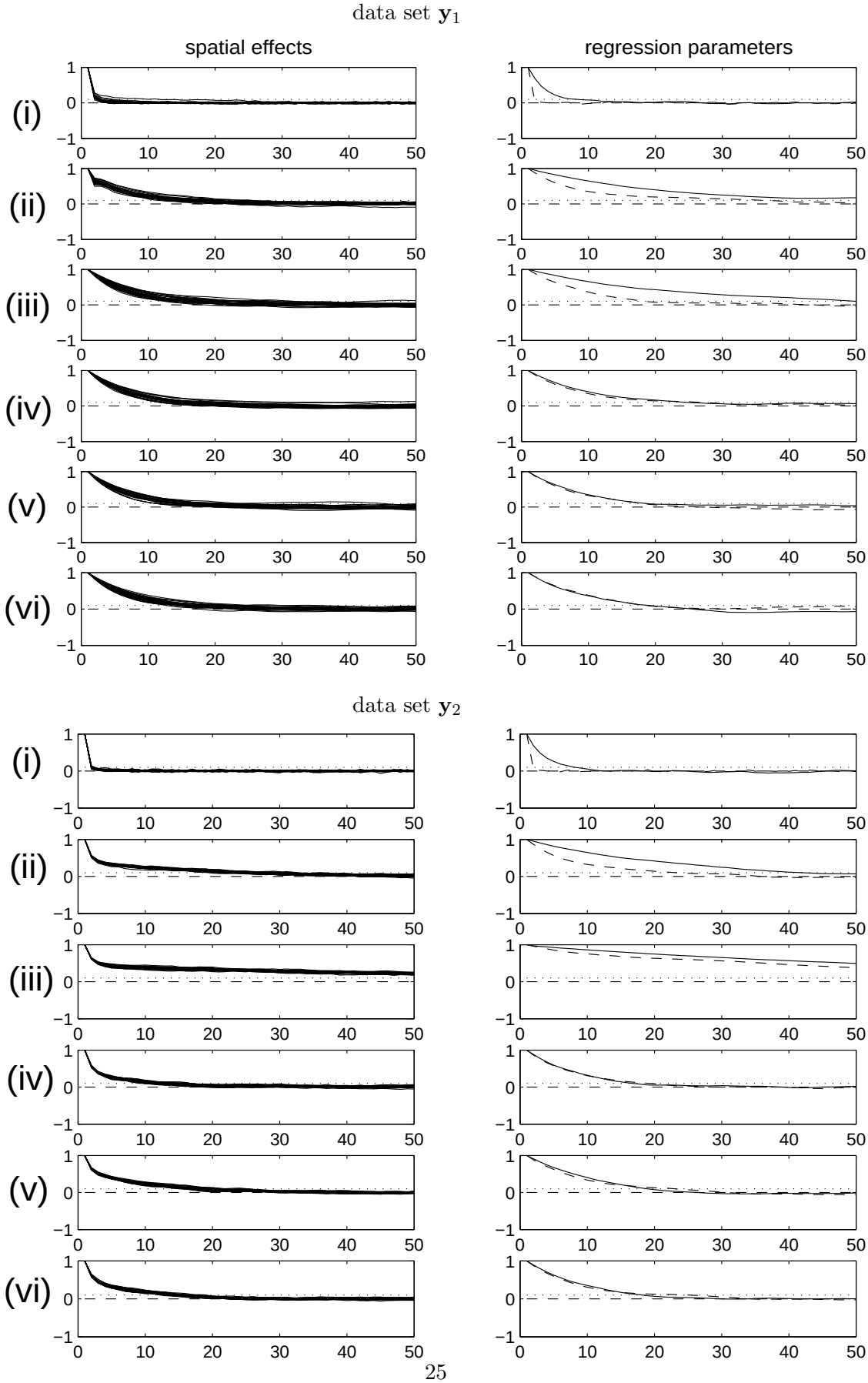

Figure 1: Estimated empirical autocorrelations for the spatial effects (left panel) and the regression parameters β_1 (solid),β_2 (dashed) (right panel) for the *independence* MH sampler (i), the *block* (ii), *centered* (iii), *coll1* (iv), *coll2* (v) and *coll3* (vi) Gibbs samplers for data sets $\mathbf{y}_1$ and $\mathbf{y}_2$.

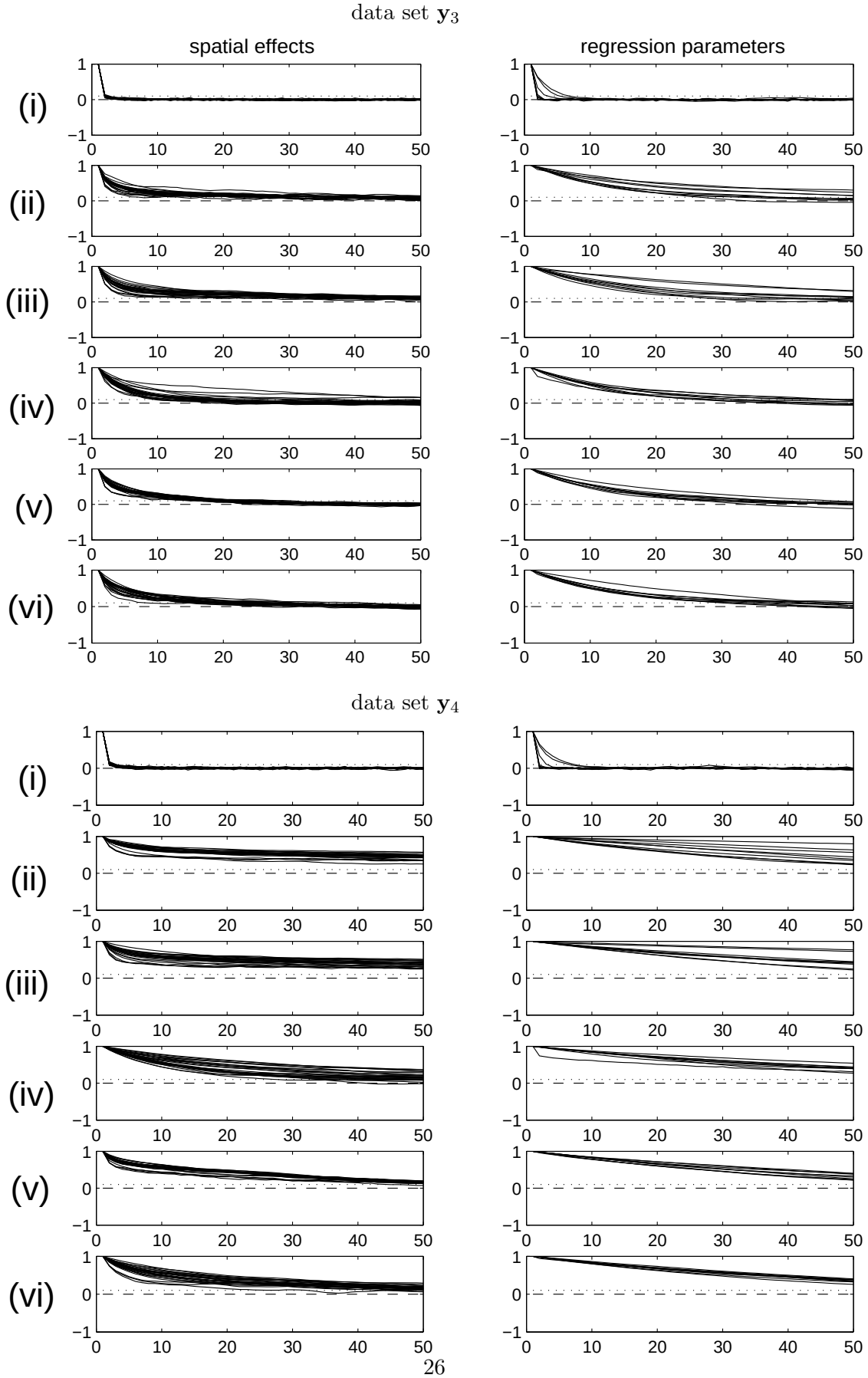

Figure 2: Estimated empirical autocorrelations for the spatial effects (left panel) and the regression parameters $\beta_1, .., \beta_8$ (right panel) for the *independence* MH sampler (i), the *block* (ii), *centered* (iii), *coll1* (iv), *coll2* (v) and *coll3* (vi) Gibbs samplers for data sets $\mathbf{y}_3$ and $\mathbf{y}_4$.