Gschlößl, Czado:

Introducing and evaluating a Gibbs sampler for spatial Poisson regression models

Projektpartner

MAX-PLANCK-GESELLSCHAFT

Introducing and evaluating a Gibbs sampler for spatial Poisson regression models

Susanne Gschlößl   Claudia Czado *

March 30, 2005

Abstract

In this paper we present a Gibbs sampler for a Poisson model including spatial effects. Frühwirth-Schnatter and Wagner (2004b) show that by data augmentation via the introduction of two sequences of latent variables a Poisson regression model can be transformed into a 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 are 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 data set. In general, both large and small simulated spatial effects are estimated accurately by the Gibbs samplers and reasonable low autocorrelations are obtained when the data variability is rather large. However, for data with very low heterogeneity, the autocorrelations resulting from the Gibbs samplers are very high, withdrawing the computational advantage over a Metropolis Hastings independence sampler which exhibits very low autocorrelations in all settings.

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

---

*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

The Poisson distribution is a common model for count data. Often, a regression is performed on the mean of the Poisson model taking covariates into account. If the data are geographically distributed, spatial effects may be additionally introduced to allow for spatial dependencies. Consider for example data of policyholders from a car insurance company distributed on J geographical regions. Since the risk for claims may depend on the region the policyholder is living in, models for the expected number of claims for each policyholder might be improved by allowing for spatial effects. In a Bayesian approach, a conditional autoregressive model (CAR), see for example Pettitt et al. (2002), is a popular spatial prior allowing for a certain spatial dependency structure.

Since the full conditional distributions of a spatial Poisson regression model do not follow any standard distribution, usually single component Metropolis Hasting (MH) steps are performed in a MCMC setting, see for example 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. 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.

In this paper we describe a straightforward Gibbs sampler for spatial Poisson regression models based on the approach by Frühwirth-Schnatter and Wagner (2004a) who 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. The basic idea is to identify Poisson observations as the number of jumps of an unobserved Poisson process. The first step of the data augmentation consists in the introduction of unobserved inter-arrival times of the Poisson process as latent variables. Thus, linearity of the observation equation is obtained, however the resulting model is still non-normal. In a second step the non-normal error term is approximated by a mixture of normal distributions and the component indicators of these mixtures are introduced as the second sequence of latent variables. Conditioning on both the inter-arrival times and the component indicators a linear normal model is obtained and a Gibbs sampler can be applied.

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

a similar methodology, a Gibbs sampler for logistic models is developed in Frühwirth-Schnatter and Waldl (2004). In this paper we show that this Gibbs sampler for a Poisson regression model can be extended to a spatial Poisson regression model in a straightforward manner.

It is well known, that mixing and convergence of the Gibbs sampler depends crucially on several implementation issues, see for example Dellaportas and Roberts (2003) for an overview and 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 parameterisation for hierarchical normal linear models, Papaspiliopoulos et al. (2003) address parameterisation issues for several classes of hierarchical models and introduce partially non-centered parameterisations. An overview on model parameterisations and efficiency is given in Frühwirth-Schnatter (2004) with special emphasis on state space models.

We consider both centered and non-centered parameterisations for our model and discuss various algorithmic schemes. In a simulation study we examine algorithms updating the parameters in several blocks as well as collapsed algorithms. In general, the lowest autocorrelations are obtained when a collapsed algorithm is used. The estimation of both small and large simulated spatial effects is very accurate, the mixing of the Gibbs samplers is reasonable well when the variability of the data is rather large. However, if the heterogeneity of the data is very low, high autocorrelations are obtained for all Gibbs sampler schemes, diminishing the computational advantage in comparison to a single component MH sampler with independence proposal.

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 for the MCMC update of all parameters and latent variables are given, in particular several algorithmic schemes are considered for the update of the regression parameters and the spatial effects using both centered and non-centered model parameterisations. In Section 3 the developed Gibbs sampler schemes are examined and compared to a single component MH sampler with independence proposals in two simulation studys. In the first study, the influence of the size of the spatial effects on the resulting autocorrelations and parameter estimation is addressed, while the second study considers the influence of the data heterogeneity on the performance of the Gibbs samplers. We also apply the Gibbs samplers to model the expected number of claims in a real data set from a German car insurance company. A comparison of the computational costs of the considered samplers is given in Section 4, 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. Further $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ denotes the vector of regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, .., \beta_p)$ and spatial 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 and 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

As indicated in the introduction, the idea 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. For each observation y_i we introduce $y_i + 1$ hidden inter-arrival times $\tau_{ij}, j = 1, .., y_i + 1$. From the properties of a Poisson process, see for example Mikosch (2004), it is well known that the inter-arrival times

are independent and follow an exponential distribution with parameter μ_i, i.e.

$$\tau_{ij}|\boldsymbol{\alpha} \sim Exponential(\mu_i) = \frac{Exponential(1)}{\mu_i}.$$

Taking the logarithm we obtain

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

Denote by $\boldsymbol{\tau} = \{\tau_{ij}, i = 1, .., n, j = 1, .., y_i + 1\}$ the collection of all inter-arrival times. Then the posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\tau}$

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \mathbf{y}, \boldsymbol{\tau}) = p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\tau})$$

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). The error term in (2.5) can be approximated by a mixture of R normal distributions with mean m_r, variance s_r^2 and weight $w_r, r = 1, .., R$, i.e.

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

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) show that $R = 5$ is sufficient to obtain a close approximation to the normal distribution. They also give the corresponding values for m_r, s_r^2 and w_r. The second step of the data augmentation then consists in the introduction of the component indicators $r_{ij} \in \{1, .., 5\}$ as latent variables. We denote the set of all component indicators by $\mathbf{R} = \{r_{ij}, i = 1, .., n, j = 1, .., y_i + 1\}$. Conditional on $\mathbf{R}$ we have

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

i.e.

$$(\log \tau_{ij} + \log t_i - m_{r_{ij}})|\boldsymbol{\alpha}, r_{ij} \sim N(-\mathbf{z}_i'\boldsymbol{\alpha}, s_{r_{ij}}^2). \tag{2.6}$$

Therefore we are dealing with a normal model which is linear in $\boldsymbol{\alpha}$ now. The posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\tau}$ and $\mathbf{R}$ is given by

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) \propto \pi(\boldsymbol{\alpha}|\boldsymbol{\theta}) \prod_{i=1}^{n} \prod_{j=1}^{y_i+1} \frac{1}{s_{r_{ij}}} \exp\left[-\frac{1}{2s_{r_{ij}}^2}(\log \tau_{ij} + \log t_i - m_{r_{ij}} + \mathbf{z}_i'\boldsymbol{\alpha})^2\right].$$

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 $\boldsymbol{\theta}$ 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) later in more detail. The spatial hyperparameter ψ is updated using a Metropolis Hastings 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 following sections.

2.4 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

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

It is well known that, 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

7

$\frac{1}{\mu_i} = \frac{1}{t_i \exp(\mathbf{z}_i' \boldsymbol{\alpha})}$ truncated at $1 - \sum_{j=1}^{y_i} \tau_{ij}$. 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 $\zeta_i \sim Exponential(\mu_i)$

2.5 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

$$
\begin{aligned}
p(r_{ij} = k|\tau_{ij}, \boldsymbol{\alpha}) &= \frac{p(r_{ij} = k, \tau_{ij}, \boldsymbol{\alpha})}{p(\tau_{ij}, \boldsymbol{\alpha})} = \frac{p(\tau_{ij}|r_{ij} = k, \boldsymbol{\alpha})p(r_{ij} = k)}{p(\tau_{ij}, \boldsymbol{\alpha})} \\
&\propto p(\tau_{ij}|r_{ij} = k, \boldsymbol{\alpha})w_k
\end{aligned}
\tag{2.7}
$$

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

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

r_{ij} can therefore be sampled from the discrete distribution (2.7) with $R = 5$ categories.

2.6 Starting values

Starting values for the component indicators r_{ij} are obtained by drawing random numbers from 1 to R. For τ_{ij} starting values are generated according to the sampling procedure described in Section 2.4. For zero observations we sample $\zeta_i \sim Exponential(0.1)$, for observations greater zero $\zeta_i \sim Exponential(y_i)$, as suggested in Frühwirth-Schnatter and Wagner (2004b).

2.7 Sampling $\boldsymbol{\alpha}$

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

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

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 (2.6) is written as

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

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.

2.7.1 Joint 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

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

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,

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

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

The calculation of the posterior covariance matrix Σ_α^{-1} in Section 2.7.1 can be computationally intensive if the number of regression parameters and spatial effects is large. Therefore it might

be more efficient to update $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in two separate blocks. Recall that $\tilde{\mathbf{z}}' = (\tilde{\mathbf{x}}', \tilde{\mathbf{v}}')$. The full conditional of $\boldsymbol{\beta}$ is given by

$$
\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) \;\; &\propto \;\; \pi(\boldsymbol{\beta}) \prod_{i=1}^{N} \exp\Big(-\frac{1}{2\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \gamma_{R(i)})^2\Big) \\
&\propto \;\; \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}'V_0^{-1}\boldsymbol{\beta} + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \gamma_{R(i)})^2\Big]\Big\} \\
&\propto \;\; \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}'\Sigma_\beta\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\mu}_\beta\Big]\Big\},
\end{aligned}
$$

where $\Sigma_\beta := V_0^{-1} + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i'$ and $\boldsymbol{\mu}_\beta := -\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i(\tilde{y}_i + \gamma_{R(i)})$.
Hence,

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

Similarly, we can show that

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

where $\Sigma_\gamma := \frac{1}{\sigma^2}Q + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'$ and $\boldsymbol{\mu}_\gamma := -\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})$.

2.7.3 Joint update of the intercept β_0 and $\boldsymbol{\gamma}$

Due to the high correlation 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 we perform a joint block update of β_0 and $\boldsymbol{\gamma}$ 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

$$
\Sigma_{\beta_{-0}} := V_{0\beta_{-0}}^{-1} + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_{\beta_{-0}i}\tilde{\mathbf{x}}_{\beta_{-0}i}', \quad \boldsymbol{\mu}_{\beta_{-0}} := -\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_{\beta_{-0}i}(\tilde{y}_i + \gamma_{R(i)} + \beta_0)
$$

and

$$
\Sigma_{\gamma\beta_0} := \begin{pmatrix} \tau^{-2} & 0 \\ 0 & \frac{1}{\sigma^2}Q \end{pmatrix} + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(1, \tilde{\mathbf{v}}_i)(1, \tilde{\mathbf{v}}_i)', \quad \boldsymbol{\mu}_{\gamma\beta_0} := -\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(1, \tilde{\mathbf{v}}_i)(\tilde{y}_i + \tilde{\mathbf{x}}_{\beta_{-0}i}'\boldsymbol{\beta}_{-0}).
$$

Here $V_{0\beta_{-0}} = \tau^2 I_p$ and $\tilde{\mathbf{x}}_{\beta_{-0}i} = (\tilde{x}_{i1}, .., \tilde{x}_{ip})$.

2.7.4 Collapsed algorithm for a model parameterisation with a non-centered mean

Another possibility is to use a collapsed algorithm, that is to integrate out particular components and perform an update based on the marginal distribution. 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 that

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

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

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

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

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 2.7.2

2.7.5 Collapsed algorithm for a model parameterisation with a non-centered mean and scale

Up to now, we only considered models with the non-centered mean parameterisation specified by (2.2) and 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

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

Following the lines of Section 2.7.4 we obtain for $\boldsymbol{\beta}$ the marginalized posterior distribution

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

where

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

and

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

Here A^* is given by $A^* := \sigma^2\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + Q$. The full conditional distribution for $\boldsymbol{\gamma}^*$ is given by

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

with

$$\Sigma_\gamma^* := \sigma^2\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + Q \quad \text{and} \quad \boldsymbol{\mu}_\gamma^* := -\sigma\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta}).$$

For the spatial hyperparameter ψ the full conditional distribution is given by

$$p(\psi|\boldsymbol{\gamma}^*) \propto |Q|^{\frac{1}{2}}\exp(-\frac{1}{2}\boldsymbol{\gamma}^{*\prime}Q\boldsymbol{\gamma}^*)\pi(\psi).$$

For this parametrisation we choose a normal prior for σ, in particular $\sigma \sim N(0,\tau_\sigma^2)$. The full conditional distribution of σ is then given by

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

where

$$\Sigma_\sigma^* := \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(\gamma_{R(i)}^*)^2 + \tau_\sigma^{-2} \quad \text{and} \quad \mu_\sigma^* := -\sum_{i=1}^{N}\gamma_{R(i)}^*\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta}).$$

2.7.6   Collapsed algorithm for a model parameterisation with a non-centered mean and variance

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

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

here and obtain the observation equation

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

where L is a lower triangular matrix resulting from the Cholesky decompostion $Q^{-1} = LL'$. 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

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

where

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

and

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

Here A^{**} is given by $A^{**} := \sigma^2\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{L'}\mathbf{v}_i\tilde{\mathbf{v}}_i'L + I$.
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

$$\Sigma_\gamma^{**} := \sigma^2\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'L + I \quad\text{and}\quad \boldsymbol{\mu}_\gamma^{**} := -\sigma\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta}).$$

The full conditional distribution of ψ is given by

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

For σ again the normal prior $\sigma \sim N(0, \tau_\sigma^2)$ is assumed and we obtain the following full conditional distribution for σ:

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

where

$$\Sigma_\sigma^{**} := \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(\boldsymbol{\gamma}^{**\prime}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}L'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'L)\boldsymbol{\gamma}^{**} + \tau_\sigma^{-2} \quad\text{and}\quad \mu_\sigma^{**} := -\sum_{i=1}^{N}\mathbf{v}_i'L\boldsymbol{\gamma}^{**}\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta}).$$

2.7.7 Centered CAR-Model

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 2.7.3 but with $\boldsymbol{\mu}_{\beta_{-0}}$ replaced by $-\sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2} \tilde{\mathbf{x}}_{\beta_{-0}i}(\tilde{y}_i + \gamma^c_{R(i)})$.

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 given as in Section 2.7.2 and

$$\boldsymbol{\mu}_\gamma^{cent} := \frac{\beta_0}{\sigma^2} Q\mathbf{1} - \sum_{i=1}^{N} \frac{1}{\tilde{s}_i^2} \tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}'_{\beta_{-0}i}\boldsymbol{\beta}_{-0})$$

and $\mathbf{1}$ represents a $J \times 1$ vector of ones.

β_0 is updated in an extra Gibbs step, in particular

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

where $\Sigma_{\beta_0} := \frac{1}{\sigma^2} \sum_{i,j=1}^{J} Q_{ij} + \frac{1}{\tau^2}$ and $\mu_{\beta_0} := \frac{1}{\sigma^2} \mathbf{1}'Q\boldsymbol{\gamma}^c$.

3 Simulation studies and application

We want to use the developed Gibbs sampler to analyse the expected number of claims in a data set from a German car insurance company. The data consider 16307 policyholders in Bavaria with full comprehensive car insurance for 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 we will first examine the developed Gibbs sampler schemes according to their mixing and estimation properties on simulated data. Additionally we will also use a Metropolis Hastings algorithm 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. We will consider two studies. In the first study the influence of the size of the spatial effects on mixing behaviour and parameter estimation is examined, while in the second study the impact of data heterogeneity is the focus.

14

3.1 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

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

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 $[\min(\boldsymbol{\gamma}) \max(\boldsymbol{\gamma})] = [-0.08, 0.08]$. For these two data sets we run the Gibbs sampler with the following three update schemes:

- parameterisation with non-centered mean: $\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1})$

  - block update of $\boldsymbol{\beta}_{-0}|\beta_0, \boldsymbol{\gamma}$ and $(\beta_0, \boldsymbol{\gamma})|\boldsymbol{\beta}_{-0}$ given in Section 2.7.3 (*block*)
  - collapsed algorithm given in Section 2.7.4 (*collapsed*)

- centered parameterization given in Section 2.7.7: $\boldsymbol{\gamma}^c \sim N(\beta_0, \sigma^2 Q^{-1})$ (*centered*)

Additionally, a single component Metropolis Hastings sampler with *independence* proposals is used. The algorithms are run for 5000 iterations, a burnin of 1000 iterations is taken. For data set $\mathbf{y}_1$ with large spatial effects the estimated empirical autocorrelations for the regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and the first nine spatial effects $(\gamma_1, .., \gamma_9)$ are plotted in Figure 1 for the different update strategies, respectively. Except for the intercept β_0 the lowest autocorrelations are obtained for the independence sampler. The autocorrelations resulting from the three Gibbs samplers in general are only slightly higher. Further, the differences between the autocorrelations of the Gibbs update schemes are very small, suggesting that the mixing properties of the three algorithms are very similar. However, the autocorrelations resulting form the collapsed algorithm tend to be slightly lower, especially for β_1.

The corresponding plots for data set $\mathbf{y}_2$ with small spatial effects are given in Figure 2. Here, clearly the worst mixing is obtained if the centered parameterization for the spatial effects 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 Gibbs sampler using the block and the collapsed update and the MH independence sampler we get similar results as for data set $\mathbf{y}_1$.
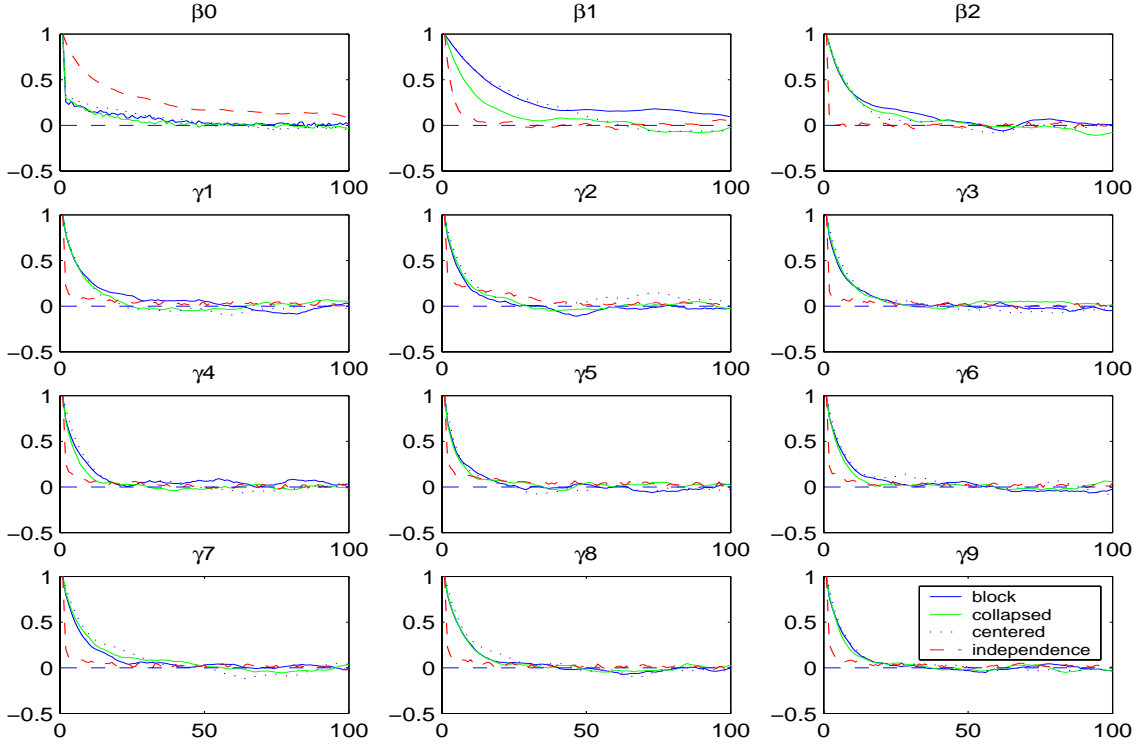
Figure 1: Estimated empirical autocorrelations for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and $(\gamma_1, .., \gamma_9)$ for data set $\mathbf{y}_1$ using the *block, collapsed, centered* and *independence* update strategies

Again, the independence sampler produces the lowest autocorrelations, but, particularly for the spatial effects, the autocorrelations resulting from the block and collapsed Gibbs sampler are reasonable small as well. According to the empirical autocorrelations for $\boldsymbol{\beta}$ again the collapsed Gibbs algorithm is to be preferred to the Gibbs sampler using a block update of β_0 and $\boldsymbol{\gamma}$.

In order to compare the four algorithms according to their estimation of the spatial effects we consider the following sum of squared differences taken over the MCMC iterates $j = burnin, .., R$ where $burnin = 1000$ and $R = 5000$

$$\sum_{i=1}^{J} \frac{1}{R - burnin} \sum_{j=burnin}^{R} [(\hat{\beta}_0^j + \hat{\gamma}_i^j) - (\beta_0^{true} + \gamma_i^{true})]^2. \tag{3.1}$$

Here $\hat{\beta}_0^j$ and $\hat{\gamma}_i^j$ denote the MCMC iterates for β_0 and $\gamma_i, i = 1, .., J$, respectively. For the centered parameterisation (3.1) changes to

$$\sum_{i=1}^{J} \frac{1}{R - burnin} \sum_{j=burnin}^{R} [\hat{\gamma}_i^{cj} - (\beta_0^{true} + \gamma_i^{true})]^2 \tag{3.2}$$

since the intercept β_0 is the spatial prior mean here and therefore already included in $\hat{\gamma}_i^{cj}$. The resulting values for data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ are given in Table 2. The magnitude of this quantity is
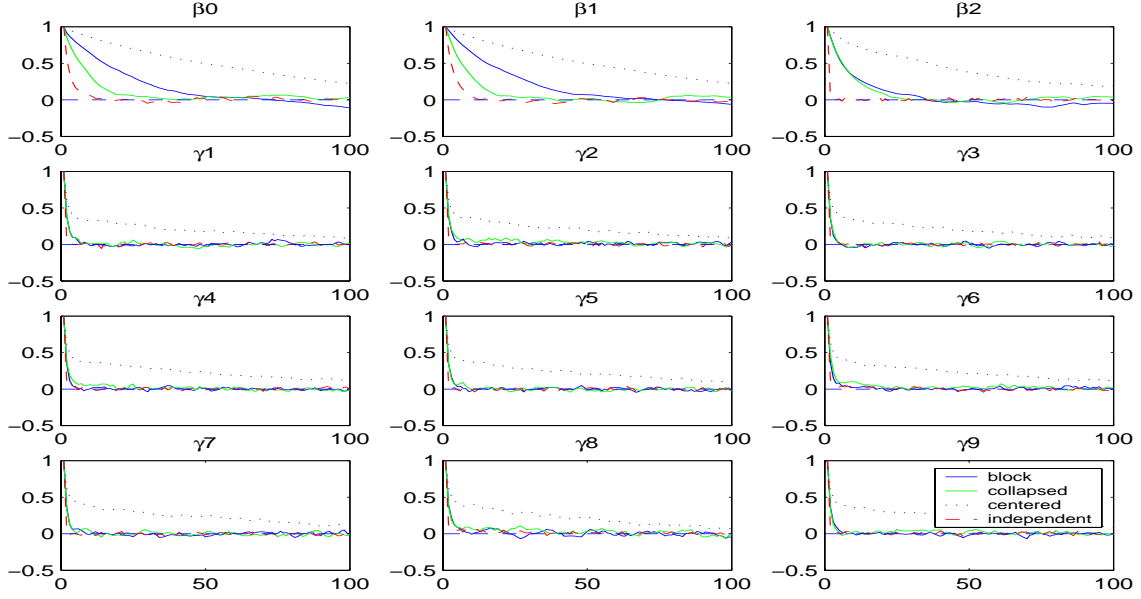
16

Figure 2: Estimated empirical autocorrelations for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and $(\gamma_1, .., \gamma_9)$ for data set $\mathbf{y}_2$ using the *block*, *collapsed*, *centered* and *independence* update strategies

Algorithm:	$\sigma^2 = 1$	$\sigma^2 = 0.01$
Gibbs block	6.95	0.40
Gibbs collapsed	6.54	0.36
Gibbs centered	6.50	0.41
Independence	7.51	1.44

Table 2: Distance (3.1) for *block*, *collapsed* Gibbs sampler and *independece* MH sampler and (3.2) for *centered* Gibbs sampler for data sets $\mathbf{y}_1$ ($\sigma^2 = 1$) and $\mathbf{y}_2$ ($\sigma^2 = 0.01$).

about the same for all algorithms for data set $\mathbf{y}_1$ with large spatial effects, indicating that the estimation of the true spatial effects is very similar for all of the four algorithms. For data set $\mathbf{y}_2$ with small spatial effects however, (3.1) is much higher for the independence sampler, i.e. estimation of the spatial effects is worse here. To illustrate these results we also present a boxplot of the MCMC iterates $\hat{\beta}_0^j + \hat{\gamma}_i^j, j = burnin, .., R$ (for the centered parameterization a boxplot of $\hat{\gamma}_i^{cj}, j = burnin, .., R$ is given) for one particular spatial effect γ_i in Figure 3. These plots show that the variability of the MCMC estimates for $\beta_0 + \gamma_i$ is very similar for all algorithms for the data set with large spatial effects, for the data set with small spatial effects however the MCMC iterates resulting from the independence sampler are much more variable.

The variance of the two simulated data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ takes the values $var(\mathbf{y}_1) = 0.51$ and $var(\mathbf{y}_2) = 0.49$. 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 will conduct a second simulation
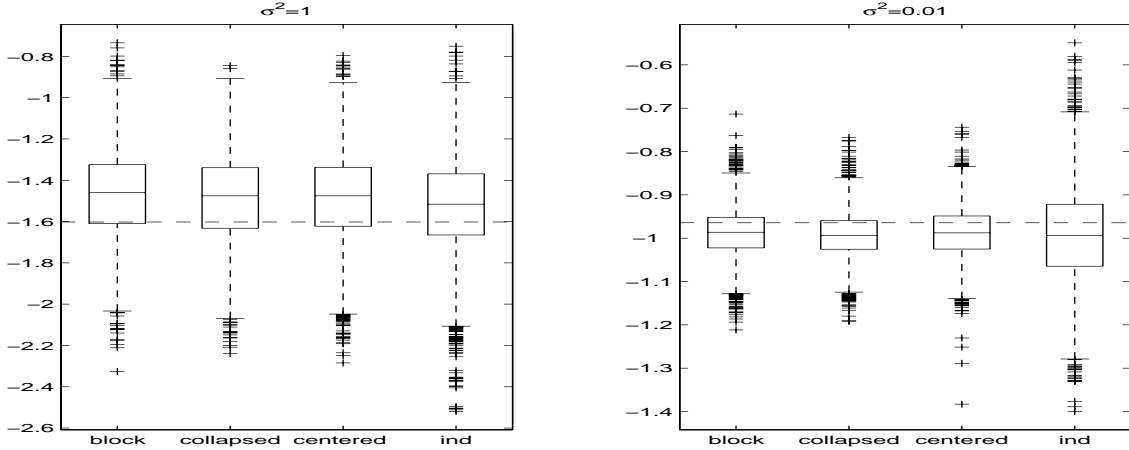
17

Figure 3: Boxplots of $\hat{\beta}_0^j + \hat{\gamma}_i^j, j = burnin, .., R$, for the Gibbs sampler using *block* and *collapsed* update and the MH *independence* sampler and boxplot of $\hat{\gamma}_i^{cj}, j = burnin, .., R$ for the Gibbs sampler using the *centered* parameterisation for one particular spatial effect $\gamma_i, i \in \{1, .., J\}$. The dotted horizontal line gives the true value $\beta_0 + \gamma_i$.

study where we examine whether the heterogeneity of the data influences mixing as well.

3.2 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. 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 $\sigma^2 = 0.01$, that is 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 $Var(\mathbf{y}_4) = 0.05$, that is data set $\mathbf{y}_4$ has very low heterogeneity and is very 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 set with high heterogeneity. We run the block and the collapsed Gibbs sampler based on the model parameterisation with non-centered mean and the Gibbs sampler based on the centered parameterisation for 5000 iterations, taking a burnin of 1000 iterations. For comparison again the MH independence sampler is applied. The estimated

empirical autocorrelations for $\boldsymbol{\beta}$ and for nine of the spatial effects $\boldsymbol{\gamma}$ are given in Figures 4 and 5 for the two data sets. For the high heterogeneity data set $\mathbf{y}_3$ similar results as in study 1 are obtained. This is not surprising, since the heterogeneity in data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ is about the same as in data set $\mathbf{y}_3$.
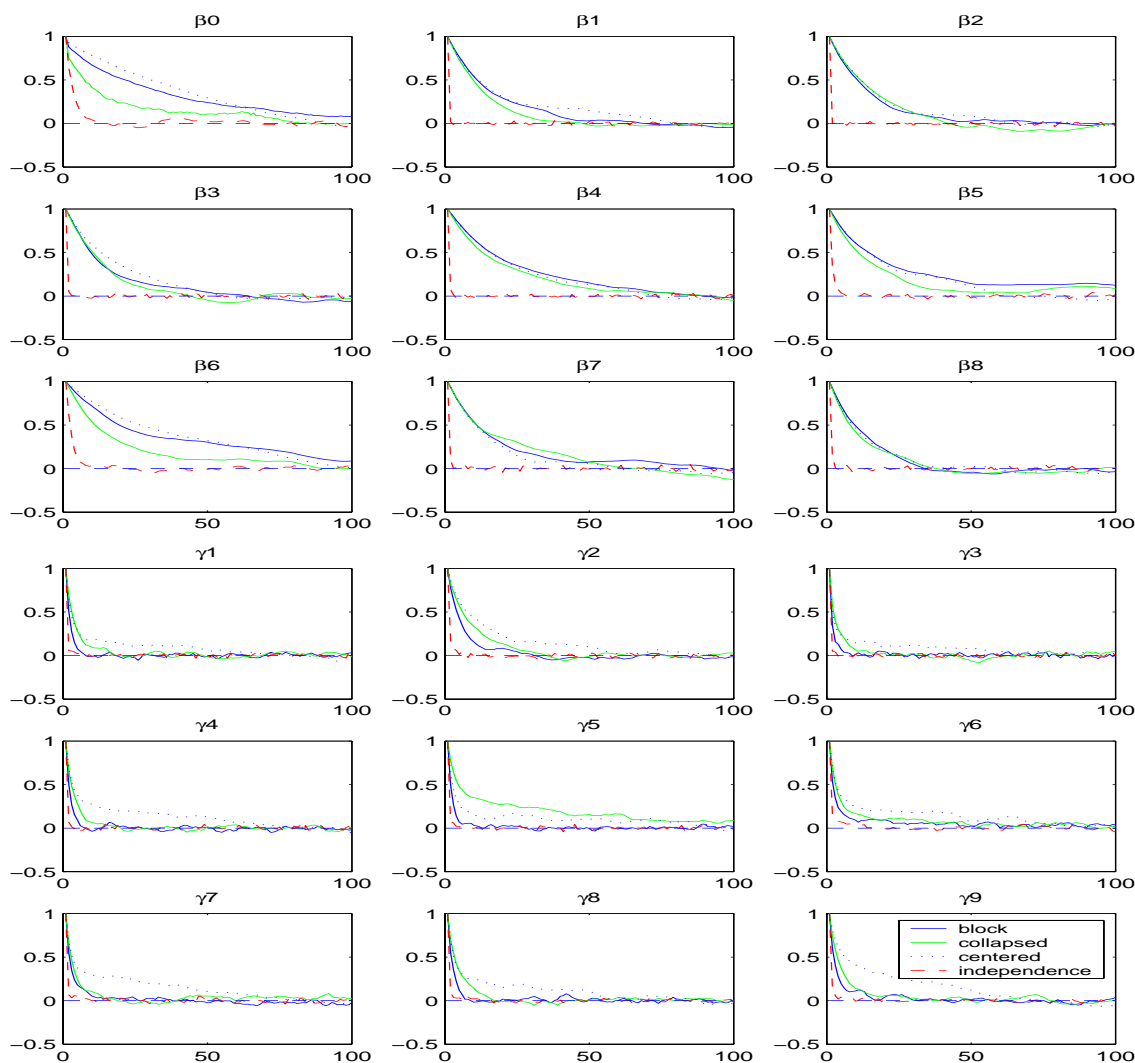

Figure 4: Estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $(\gamma_1, .., \gamma_9)$ for data set $\mathbf{y}_3$ with 'high' heterogeneity using the *block*, *collapsed*, *centered* and *independence* update strategies

The estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in data set $\mathbf{y}_4$ with low heterogeneity however are very high for the three Gibbs sampler schemes, especially for the regression parameters. For this data set the Gibbs sampler is clearly inferior to the MH independence sampler which has very low autocorrelations again.
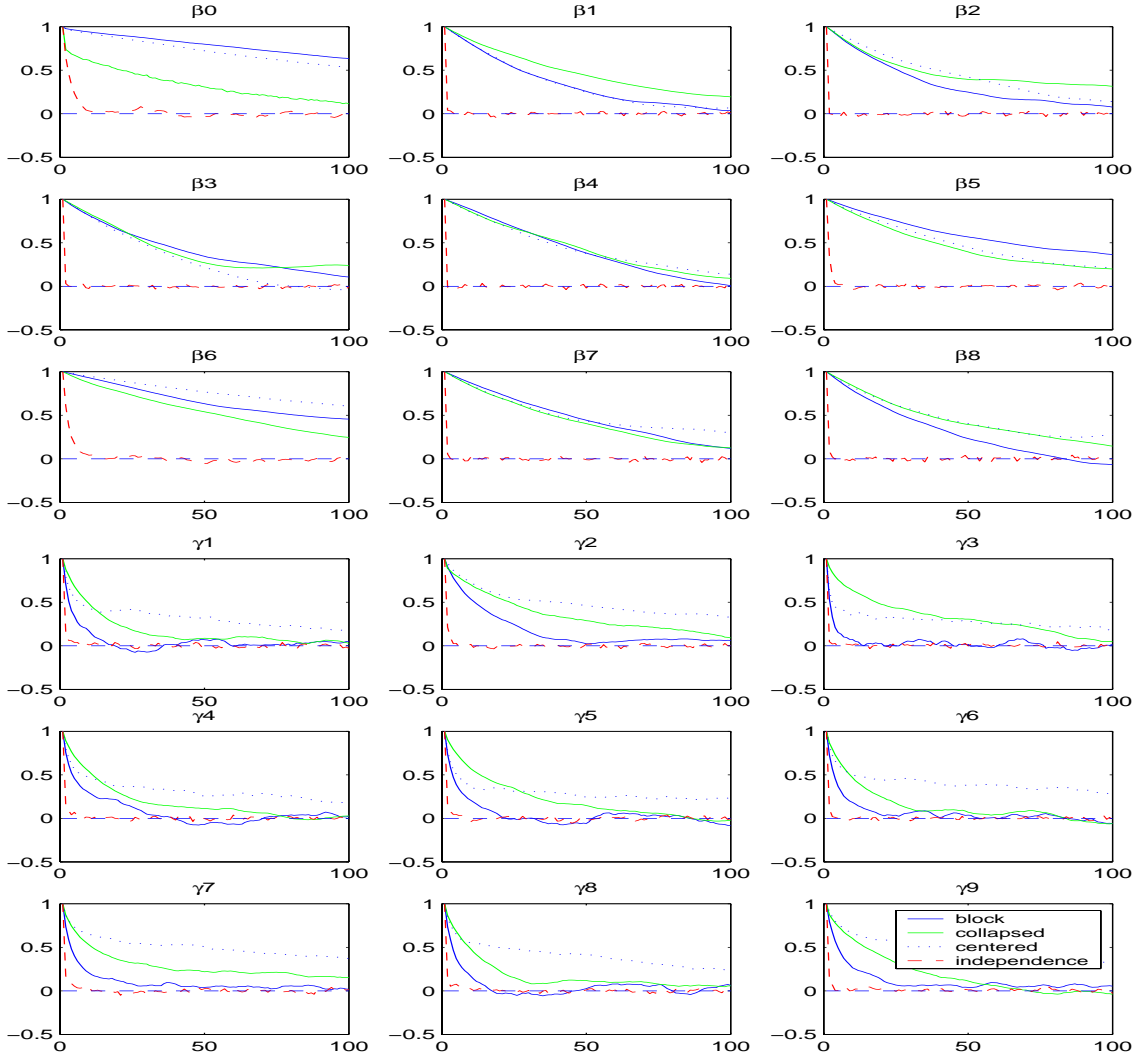
Figure 5: Estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $(\gamma_1, .., \gamma_9)$ for data set $\mathbf{y}_4$ with low heterogeneity using the *block, collapsed, centered* and *independence* update strategies

3.3 Application to car insurance data

Finally we apply the *block, collapsed* and *centered* Gibbs sampler as well as the *independence* MH sampler on the car insurance data set described at the beginning of this section. The empirical autocorrelations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are plotted in Figure 6. Similar results to data set $\mathbf{y}_4$ which is very close to our real data, are observed. In particular for the regression parameters, the autocorrelations resulting from the Gibbs sampler schemes are very high compared to the good mixing of the MH independence sampler.
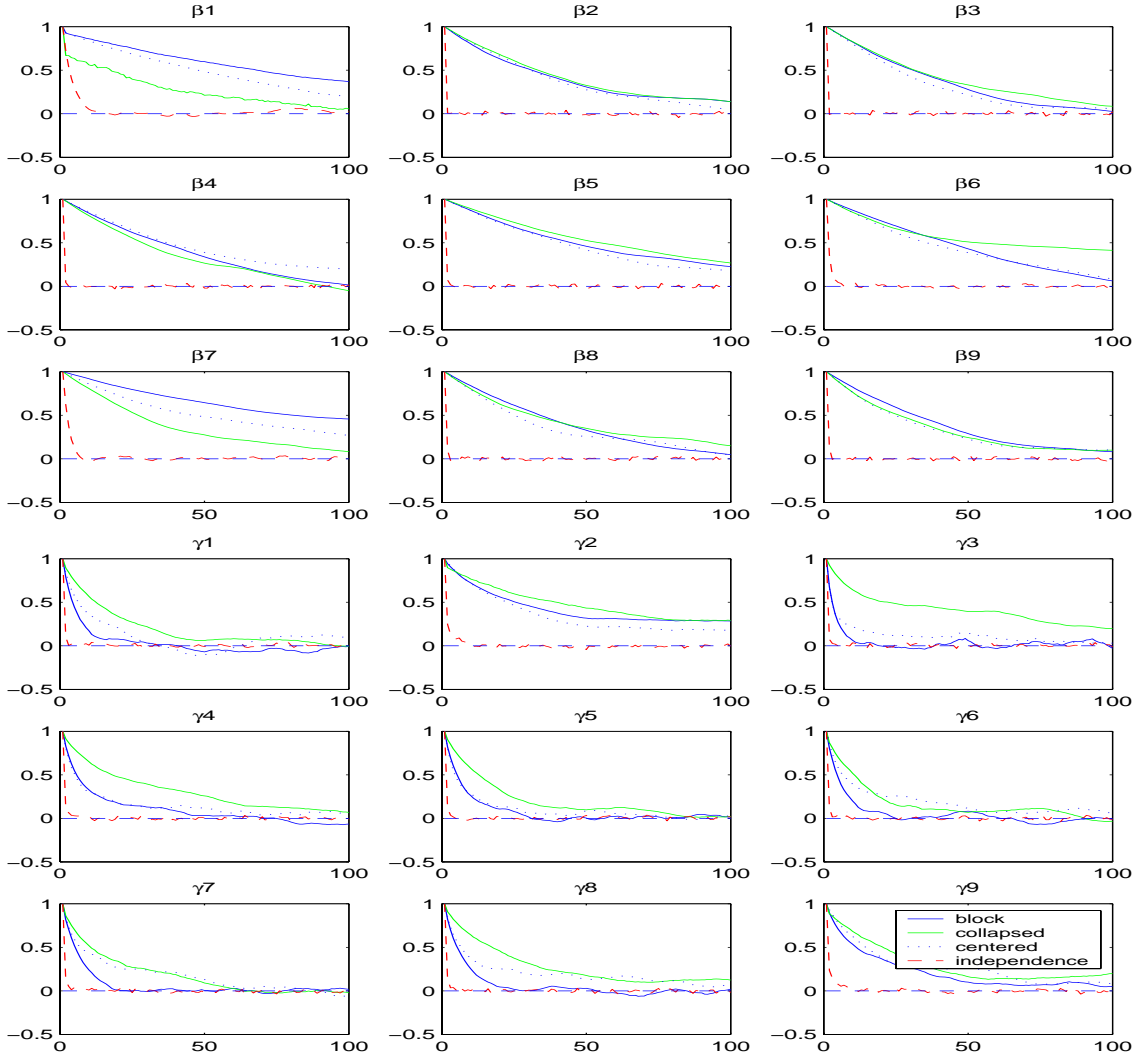
Figure 6: Estimated empirical autocorrelations for $\boldsymbol{\beta}(\gamma_1, .., \gamma_9)$ and for the real data set using the *block*, *collapsed*, *centered* and *independence* update strategies

3.3.1 Parameterisations with non-centered mean, scale and variance

Since neither the centered nor the non-centered mean parameterisations led to satisfying results for the real data set, we will check, if mixing can be improved by using collapsed algorithms with parameterisations which are non-centered in the mean, scale and variance discussed in Sections 2.7.5 and 2.7.6. First, we apply collapsed algorithms for these non-centered parameterisation for the simulated data sets $\mathbf{y}_1$ and $\mathbf{y}_2$, the resulting estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $\gamma_1, .., \gamma_9$ are given in Figures 7 and 8. Here *coll1* denotes the algorithm given in Section 2.7.4, whereas *coll2* and *coll3* denote the algorithms for the parameterisations discussed in Sections 2.7.5 and 2.7.6 respectively. For both data sets the three algorithms perform very similar in general, the mixing is not improved using the parameterisations additionally non-centered in
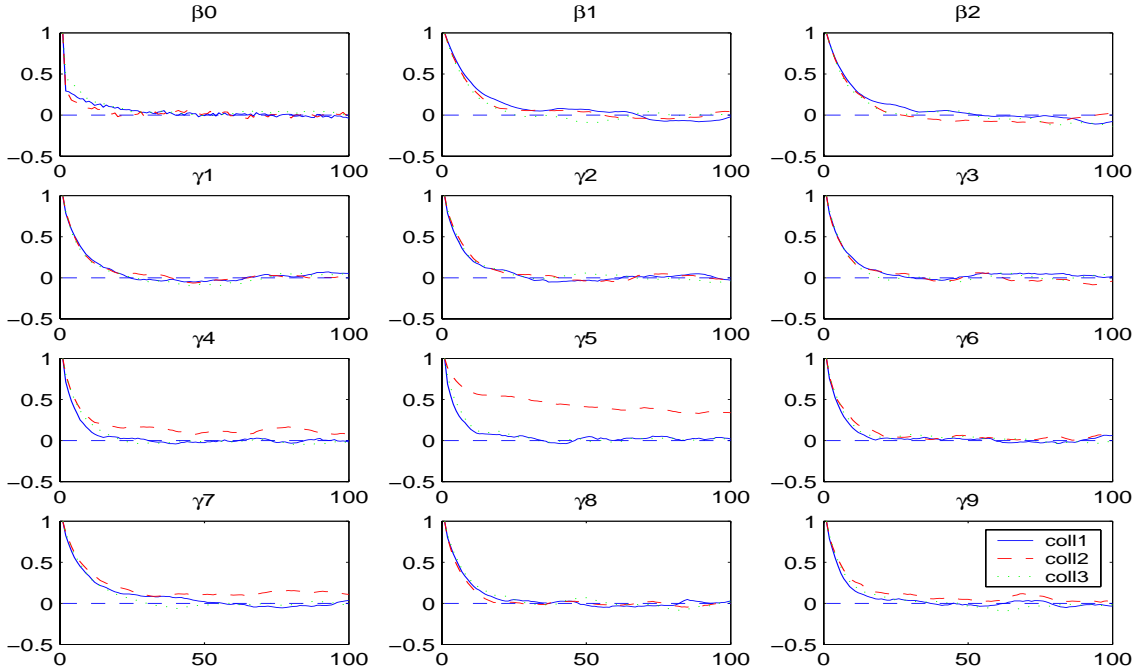
scale and variance respectively.

Figure 7: Estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $\gamma_1, .., \gamma_9$ for data set $\mathbf{y}_1$ using a collapsed algorithm for model parameterisations *coll1*, *coll2* and *coll3*.

The resulting empricial autocorrelations for $\boldsymbol{\beta}$ and $\gamma_1, .., \gamma_9$ in the real car insurance data set are plotted in Figure 9. Here, at least for the spatial effects slightly lower autocorrelations are obtained when using the non-centered scale and variance parameterisations. For the regression parameters however, mixing is not improved.

4 Computational costs

Additional to mixing and parameter estimation we also want to compare the computational costs of the Gibbs sampler schemes and the MH independence sampler. 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, that is 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 the Gibbs sampler is always faster than the MH independence sampler. However, the computational advantage of the Gibbs sampler
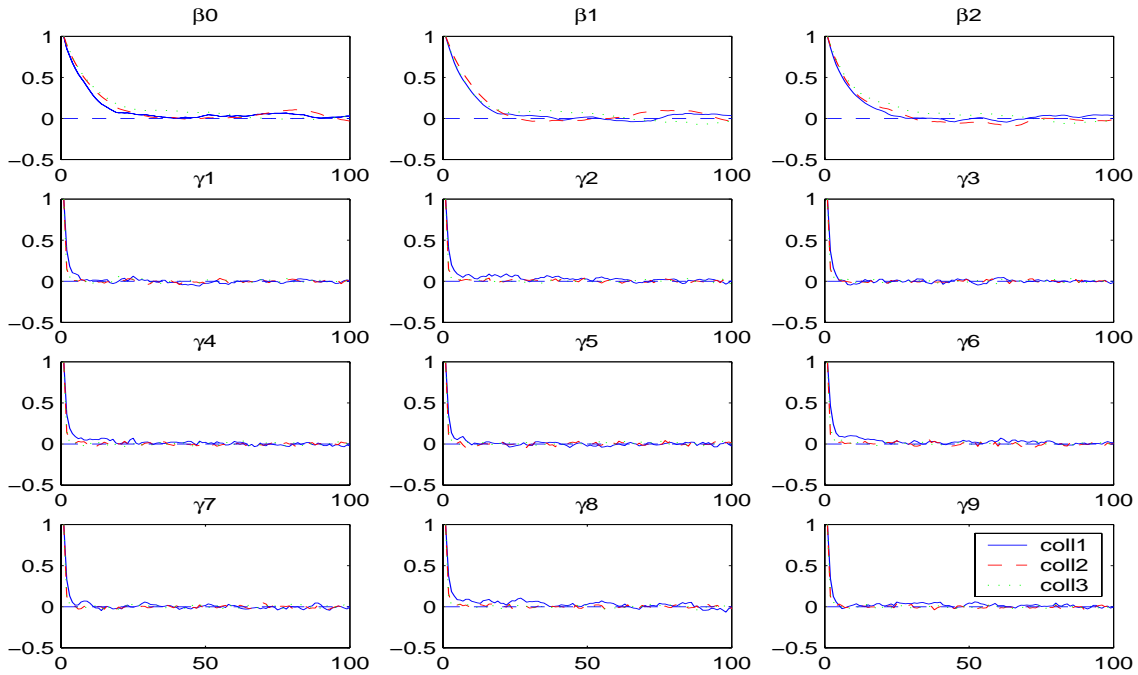
Figure 8: Estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $\gamma_1, .., \gamma_9$ for data set $\mathbf{y}_2$ using a collapsed algorithm for model parameterisations *coll1*, *coll2* and *coll3*.

depends on the complexity of the model. For the setting in study 1 with 5000 observations, an intercept and two covariates for example, the MH independence sampler takes 1.5 times as long as the Gibbs sampler using the block update. For the setting in study 2 the MH independence sampler even takes 3.6 times as long. For the different Gibbs sampler schemes the computational effort differs only slightly.

5   Summary and conclusions

We have described a Gibbs sampler for a spatial Poisson regression model using data augmentation and discussed several update schemes for the regression parameters and spatial effects. Centered and non-centered model parameterisations have been considered.

For data which are not too homogeneous, the Gibbs samplers give reasonable low autocorrelations and a good estimation of the spatial effects. In general, the best mixing is achieved using a collapsed Gibbs sampler in a model parameterisation with a non-centered mean. Additionally non-centering the scale or the variance of the spatial prior distribution did not lead to significant improvement.

However, if the variance of the spatial effects is very small, the centered parameterisation is not very efficient any more, the Gibbs samplers based on the model parameterisations with non-centered mean perform much better. For data with low heterogeneity none of the Gibbs sampler
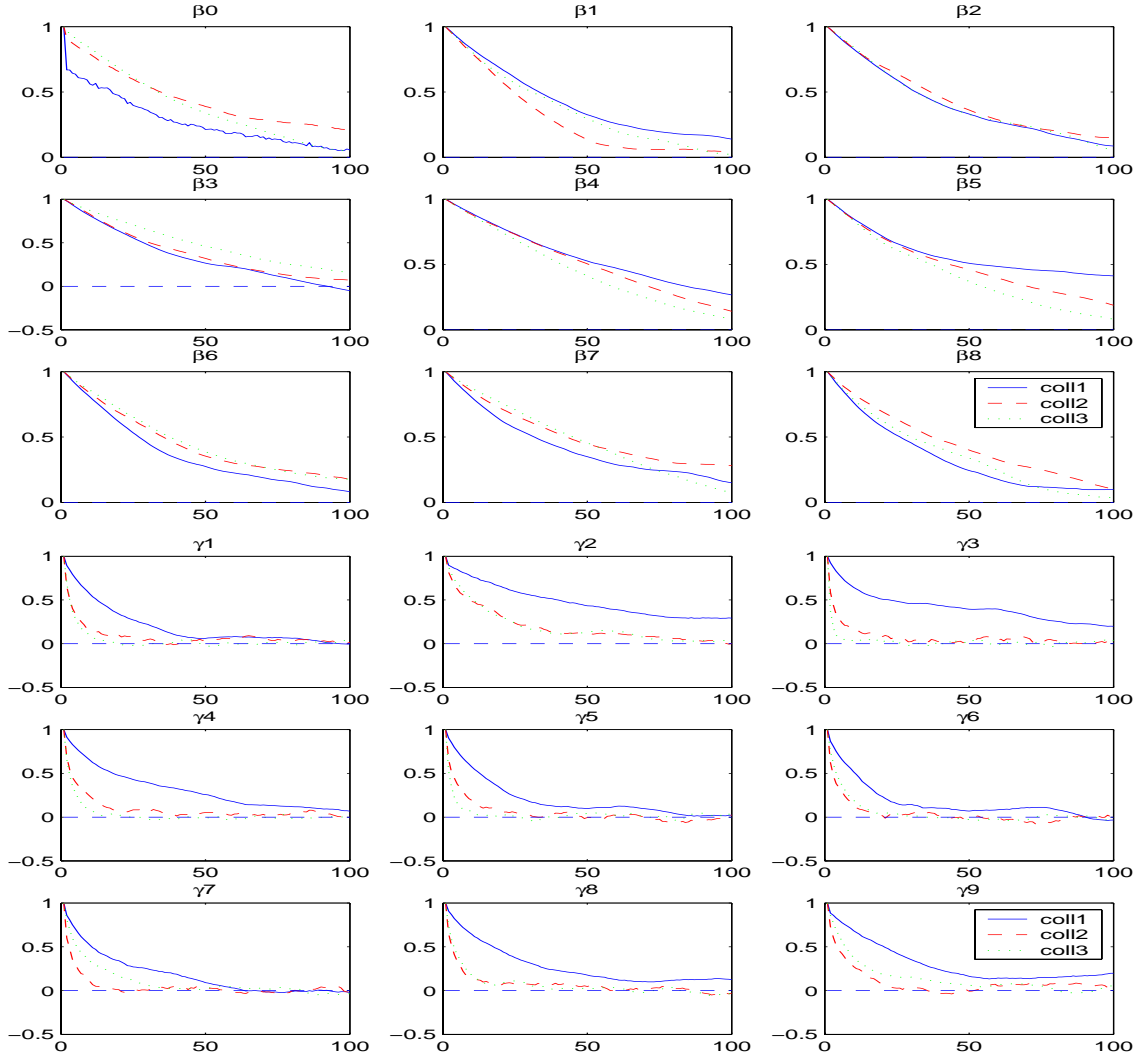
Figure 9: Estimated empirical autocorrelations for $\boldsymbol{\beta}$ and $(\gamma_1, .., \gamma_9)$ for the real data set using a collapsed algorithm for model parameterisations *coll1*, *coll2* and *coll3*.

schemes gives satisfying results, especially for the regression parameters high autocorrelations are obtained, compensating the computational advantage of the Gibbs samplers over the MH independence sampler. The MH independence sampler in contrast, leads to very low autocorrelations for both large and small spatial effects and also works well for low heterogeneity data. The price for this to pay is the higher computational effort. But, taking the high autocorrelations of the Gibbs sampler for low heterogeneity data into account, the MH independence sampler outperforms the Gibbs samplers despite the additional effort. However, the estimation of small spatial effects is worse for the MH independence sampler and the variability of the MCMC estimates of the spatial effects is higher than for the Gibbs samplers.

6 Appendix

For the collapsed algorithm in Section 2.7.4 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

$$
\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{R}) \quad \propto \quad & \exp\left\{-\frac{1}{2}\Big[\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta} + \tilde{\mathbf{v}}_i'\boldsymbol{\gamma})^2 + \boldsymbol{\gamma}'\sigma^{-2}Q\boldsymbol{\gamma} + \boldsymbol{\beta}'\tau^{-2}I\boldsymbol{\beta}\Big]\right\} \\
= \quad & \exp\left\{-\frac{1}{2}\Big[\boldsymbol{\beta}'\tau^{-2}I\boldsymbol{\beta} + \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})^2\Big]\right\} \\
\times \quad & \exp\left\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'\Big(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i' + \sigma^{-2}Q\Big)\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i(\tilde{y}_i + \tilde{\mathbf{x}}_i'\boldsymbol{\beta})\Big]\right\} \\
:= \quad & c(\boldsymbol{\beta}) \times \exp\left\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'a\Big]\right\} \quad\quad\quad\quad (6.3)
\end{aligned}
$$

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

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

and therefore

$$
\begin{aligned}
\int \exp\left\{-\frac{1}{2}\Big[\boldsymbol{\gamma}'A\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'a\Big]\right\}d\boldsymbol{\gamma} \quad \propto \quad & (2\pi)^{\frac{J}{2}}|A|^{-\frac{1}{2}}\exp\left\{\frac{1}{2}(A^{-1}a)'A(A^{-1}a)\right\} \\
\propto \quad & \exp\left\{\frac{1}{2}(A^{-1}a)'A(A^{-1}a)\right\} \quad\quad\quad (6.4)
\end{aligned}
$$

From (6.3) and (6.4) it then follows that

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

Finally, with

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

it follows that

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

$$- \quad 2\boldsymbol{\beta}'\Big((\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{\mathbf{x}}_i')'A^{-1}(\sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{v}}_i\tilde{y}_i) - \sum_{i=1}^{N}\frac{1}{\tilde{s}_i^2}\tilde{\mathbf{x}}_i\tilde{y}_i\Big)\Big]\Big\},$$

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

References

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

Czado, C. and S. Prokopenko (2004). Modeling transport mode decisions using hierarchical binary spatial regression models with cluster effects. *submitted*.

Dellaportas, P. and G. O. Roberts (2003). An introduction to MCMC. In *Spatial statistics and computational methods (Aalborg, 2001)*, Volume 173 of *Lecture Notes in Statist.*, pp. 1–41. New York: Springer.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation. In *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.

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.

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.

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.