Modelling count data with overdispersion and spatial effects

Susanne Gschlößl Claudia Czado *

February 11, 2005

Abstract

In this paper we consider regression models for count data allowing for overdispersion in a Bayesian framework. Besides the inclusion of covariates, spatial effects are incorporated and modelled using a proper Gaussian conditional autoregressive prior based on Pettitt et al. (2002). Apart from the Poisson regression model, the negative binomial and the generalized Poisson regression model are addressed. Further, zero-inflated models combined with the Poisson and generalized Poisson distribution are discussed. In an application to a data set from a German car insurance company we use the presented models to analyse the expected number of claims. Models are compared according to the deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002). To assess the model fit we use posterior predictive p-values proposed by Gelman et al. (1996). For this data set no significant spatial effects are observed, however the models allowing for overdispersion perform better than a simple Poisson regression model.

*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 calculation of premiums in insurance companies is based on the expected claim frequency and the expected claim size which have to be estimated using statistical models. In this paper we will focus on a model for claim frequency in a Bayesian framework. We consider a regression model which takes both covariate effects and spatial effects into account. This allows the expected claim frequency to depend both on covariates like age and gender and on the geographical location of the policyholders as well. This problem is also addressed in Dimakos and Frigessi (2000). They use a Poisson regression model including spatial effects to model claim frequencies and assume an improper Markov Random Field (MRF) prior for the latent spatial variables. In contrast to this work, we use a proper Gaussian conditional autoregressive prior based on Pettitt et al. (2002) for the spatial effects. Additionally, we consider other models for count data than the Poisson model which allow for overdispersion.

In the Poisson model equality of mean and variance is assumed, however, often the observed variance in the data is higher than expected, that is overdispersion is present. For overdispersed count data the negative binomial distribution or the generalized Poisson distribution introduced by Consul and Jain (1973) and studied in detail in the book by Consul (1989) can be used. Both models allow an independent modelling of mean and variance by the inclusion of an additional parameter.

When dealing with a data set with an excessive number of zeros, zero-inflated models are useful, see for example Winkelmann (2003). Zero inflated models can be used in combination with any model for count data. Additionally to the zero observations arising from the count data model an extra number of zeros is incorporated. Lambert (1992) introduced the zero inflated Poisson regression model, a Bayesian analysis of zero inflated models in general is given in Rodrigues (2003). Zero inflated models in combination with the generalized Poisson distribution have been addressed in Famoye and Singh (2003b) and Famoye and Singh (2003a), a Bayesian analysis without the inclusion of covariates is given in Angers and Biswas (2003). Agarwal et al. (2002) use a zero inflated Poisson regression model for spatial count data.

In this paper, we consider and compare several regression models for count data both allowing for overdispersion and including spatial effects. Besides the Poisson, the negative binomial and the generalized Poisson regression model, we also discuss zero inflated Poisson and zero inflated generalized Poisson regression models with and without the inclusion of spatial effects in a Bayesian context. All these models are applied to a data set from a German car insurance company, the results are compared. Whereas for this data set no significant spatial pattern is observed, the models allowing for overdispersion give slightly better results than the Poisson model. For the models including overdispersion no clear distinction is possible, since they are very close according to parameter estimation and DIC.

This paper is organized as follows. In Section 2 the negative binomial, the generalized Poisson and zero-inflated regression models are presented. The conditional autoregressive prior for the spatial effects is discussed in Section 3, prior assumptions for the regression and model dependent overdispersion parameters are given in Section 4. In order to compare the different models presented and assess the model fit, the deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002) and posterior predictive p-values proposed by Gelman et al. (1996) are used. These criterions are reviewed in Section 5. Finally, in Section 6 an application of the presented models to a data set from a German car insurance company is given, results are summarized in Section 7. Details about the MCMC algorithms can be found in the Appendix.

2 Models for count data including overdispersion

A commonly used model for count data is the Poisson model, where equality of mean and variance is assumed. Since this condition is not satisfied any more if overdispersion is present in the data, we consider models, which allow the variance to be larger than the mean in this section.

2.1 Negative Binomial (NB) distribution and Regression

The density of the negative binomial distribution with parameters $r > 0$ and $\mu > 0$ denoted by NB(r,μ) is defined by

$$P(Y = y|r, \mu) = \frac{\Gamma(y + r)}{\Gamma(r)y!} \cdot \Big(\frac{r}{\mu + r}\Big)^r \cdot \Big(\frac{\mu}{\mu + r}\Big)^y, \qquad y = 0, 1, 2, ... \tag{2.1}$$

with

$$E(Y|r, \mu) = \mu \quad \text{and} \quad Var(Y|r, \mu) = \mu\Big(1 + \frac{\mu}{r}\Big).$$

The variance is multiplied by the positive factor $\varphi := 1 + \frac{\mu}{r}$ and therefore greater than the mean, i.e. overdispersion can be modelled in the negative binomial distribution. We call the factor φ dispersion factor. In the limit $r \to \infty$ the Poisson distribution is obtained, see Johnson et al. (1993). From $\dfrac{P(Y = y + 1)}{P(Y = y)} = \dfrac{y + r}{y + 1}\Big(\dfrac{\mu}{\mu + r}\Big)$, we can derive that $P(Y = y + 1) > P(Y = y)$ if $y < \frac{\mu r - \mu - r}{r} := k$. Therefore, if k is not an integer, the NB distribution is unimodal with mode at $y = \lfloor k \rfloor$, i.e. the integer part of k. If k is an integer there are two modes at $y = k$ and $y = k + 1$. In a regression model with $Y_i \sim \text{NB}(r, \mu_i)$ independent for $i = 1, .., n$, the mean of Y_i is specified in terms of covariates $\mathbf{x}_i$ and unknown regression parameters $\boldsymbol{\beta}$ by

$$E(Y_i|\mathbf{x_i}, \boldsymbol{\beta}) = \mu(\mathbf{x}_i, \boldsymbol{\beta}) := \mu_i > 0.$$

Note, that in the NB regression model the dispersion factor $\varphi_i := 1 + \frac{\mu_i}{r}$ takes observation specific values.

2.2 Generalized Poisson (GP) distribution and Regression

The generalized Poisson distribution has been introduced by Consul and Jain (1973) and is investigated in detail in Consul (1989). A random variable Y is called generalized Poisson distributed with parameters θ and λ, denoted by GP(θ, λ), if

$$P(Y = y|\theta, \lambda) = \begin{cases} \theta(\theta + y\lambda)^{y-1}\dfrac{1}{y!}\exp(-\theta - y\lambda), & y = 0, 1, 2, ... \\ 0 & \text{for}\, y > m \,\text{when}\, \lambda < 0 \end{cases} \tag{2.2}$$

where $\theta > 0$, $\max(-1, -\frac{\theta}{m}) < \lambda \leq 1$ and m (≥ 4) is the largest positive integer for which $\theta + m\lambda > 0$ when λ is negative. Hence, for $\lambda < 0$ the model gets truncated. In this case, the lower limit for λ ensures, that there are at least five classes with positive probability. Further, we have

$$E(Y|\theta,\lambda) = \frac{\theta}{1-\lambda} := \theta \cdot \phi \tag{2.3}$$

where $\phi := \dfrac{1}{1-\lambda}$ and

$$Var(Y|\theta,\lambda) = \frac{\theta}{(1-\lambda)^3} = E(Y|\theta,\lambda) \cdot \frac{1}{(1-\lambda)^2} = E(Y|\theta,\lambda) \cdot \phi^2. \tag{2.4}$$

Hence $\varphi := \frac{1}{(1-\lambda)^2}$ can be interpreted as an dispersion factor for the GP distribution. For $\lambda = 0$, the generalized Poisson distribution reduces to the Poisson distribution with parameter θ, equality of mean and variance are obtained in this case. For $\lambda < 0$ underdispersion is included in the model, whereas for $\lambda > 0$ overdispersion is obtained.

Consul (1989) shows that the GP distribution is unimodal for all values of θ and λ. Covariates can be incorporated in the model as well. For this let $Y_i, i = 1,..,n$ be independent GP(θ_i,λ) distributed response variables and $\boldsymbol{x_i}$ be the corresponding vector of covariates. Further let

$$E(Y_i|\boldsymbol{x_i};\boldsymbol{\beta},\lambda) = \mu(\boldsymbol{x_i};\boldsymbol{\beta}) := \mu_i > 0$$

be the mean of the Y_i. Using (2.3) yields that $\mu_i = \dfrac{\theta_i}{1-\lambda} = \theta_i \cdot \phi$. For this parameterisation the corresponding GP regression model is given by

$$P(Y_i = y_i|\boldsymbol{x_i},\boldsymbol{\beta},\lambda) = \mu_i[\mu_i + (\phi-1)y_i]^{y_i-1}\frac{\phi^{-y_i}}{y_i!}\exp\Big[-\frac{\mu_i + (\phi-1)y_i}{\phi}\Big] \tag{2.5}$$

and denoted by GP(μ_i,λ). In contrast to the NB regression model the dispersion factor $\varphi = \frac{1}{(1-\lambda)^2}$ is the same for any observation here.

2.3 Comparison of NB and GP distribution

In order to compare the behaviour of the NB and the GP distribution, we equate the mean and the variance of a GP(μ,λ) with the mean and the variance of a NB(r,μ) distributed random variable, i.e.

$$\frac{\mu}{(1-\lambda)^2} = \mu(1+\frac{\mu}{r})$$

has to hold and the equation

$$r = \frac{\mu(1-\lambda)^2}{\lambda(2-\lambda)} \tag{2.6}$$

is obtained. In Figure 1 the negative binomial distribution is plotted in comparison to the generalized Poisson distribution with equal mean and variance, i.e. with μ and r chosen according to (2.6). For a better visual comparison the densities of these discrete distributions are presented as line plots. For small values of λ both distributions behave very similarly. With increasing values of λ slight differences between the two distributions can be observed which become greater when

λ tends to 1. In particular, the negative binomial distribution gives more mass to small values of y if a large overdispersion is present.
In the remainder of this paper only the problem of overdispersion will be addressed, i.e. λ will be assumed to take only values in the interval $[0, 1]$.

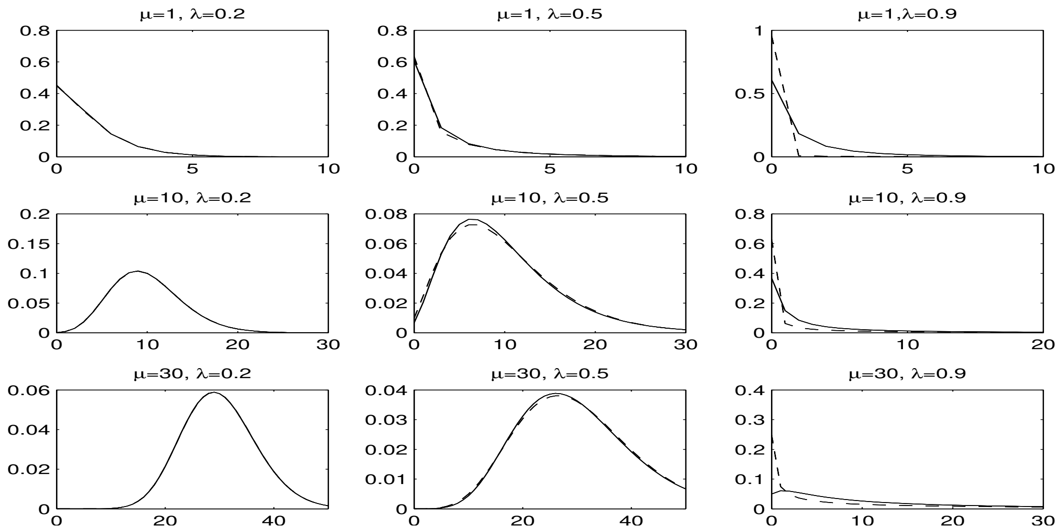

Figure 1: Comparison of the generalized Poisson distribution with $\mu = 1, 10, 30$ and $\lambda = 0.2, 0.5, 0.9$ to the negative binomial distribution with $\mu = 1, 10, 30$ and $r = \frac{\mu(1-\lambda)^2}{\lambda(2-\lambda)}$.

2.4 Zero Inflated (ZI) Models

For count data with an excessive number of zero observations zero inflated (ZI) models can be used. These models allow for a higher number of zeros than can be explained by standard models for count data. Additional to the zero observations arising from the supposed count data distribution, a proportion of extra zeros is assumed. ZI models have been widely used in the literature. A short overview is given in Winkelmann (2003), Lambert (1992) introduced ZI models based on the Poisson distribution including covariates. A Bayesian analysis of ZI models is given in Rodrigues (2003), Agarwal et al. (2002) use ZI models for spatial count data in a Bayesian framework.
Let $\pi(y|\boldsymbol{\theta})$ be a distribution function for count data with unknown parameters $\boldsymbol{\theta}$. Then a zero inflated model with extra proportion $p \in [0, 1]$ of zeros is defined by (see Agarwal et al. (2002))

$$P(Y = y|p, \boldsymbol{\theta}) = \begin{cases} p + (1-p)\pi(y = 0|\boldsymbol{\theta}) & \text{if } y = 0 \\ (1-p)\pi(y|\boldsymbol{\theta}) & \text{if } y > 0 \end{cases} \tag{2.7}$$

Mean and variance are given by

$$E(Y|p, \boldsymbol{\theta}) = (1-p)E_\pi(Y|\boldsymbol{\theta}) \tag{2.8}$$

and

$$Var(Y|p, \boldsymbol{\theta}) = p(1-p)[E_\pi(Y|\boldsymbol{\theta})]^2 + (1-p)Var_\pi(Y|\boldsymbol{\theta}). \tag{2.9}$$

The introduction of latent indicator variables $\mathbf{Z} = (Z_1, .., Z_n)'$ leads to a model which is easier to handle in a Bayesian context and in particular allows a Gibbs step for p. Z_i takes the value $z_i = 0$ for all observations with $y_i > 0$. For all zero observations $y_i = 0$, the latent variable takes the value $z_i = 0$ if observation i arises from the count data distribution $\pi(y|\boldsymbol{\theta})$ and the value $z_i = 1$ if it is an extra zero. The joint distribution of Y_i and Z_i is therefore determined by

$$P(Y_i = 0, Z_i = 1|p_i, \boldsymbol{\theta}) = p_i,\ P(Y_i = 0, Z_i = 0|p_i, \boldsymbol{\theta}) = (1-p_i)\pi(y_i = 0|\theta),$$
$$P(Y_i = y_i, Z_i = 1|p_i, \boldsymbol{\theta}) = 0,\ P(Y_i = y_i, Z_i = 0|p_i, \boldsymbol{\theta}) = (1-p_i)\pi(y_i|\boldsymbol{\theta}),\ y_i > 0$$

which can be written succinctly

$$P(Y_i = y_i, Z_i = z_i|p_i, \boldsymbol{\theta}) = p_i^{z_i}[(1-p_i)\pi(y_i|\boldsymbol{\theta})]^{1-z_i}.$$

Marginally, $Z_i \sim Bernoulli(p_i)$. Using the latent variables $\mathbf{Z}$, the joint likelihood of $\mathbf{Y} = (Y_1, .., Y_n)'$ and $\mathbf{Z}$ is given by

$$\begin{aligned} f(\mathbf{Y}, \mathbf{Z}|p, \boldsymbol{\theta}) &= \prod_{i=1}^{n} p_i^{z_i}[(1-p_i)\pi(y_i|\boldsymbol{\theta})]^{1-z_i} \\ &= \prod_{i:y_i=0} p_i^{z_i}[(1-p_i)\pi(0|\boldsymbol{\theta})]^{1-z_i} \cdot \prod_{i:y_i>0} (1-p_i)\pi(y_i|\boldsymbol{\theta}) \end{aligned}$$

In this paper we will focus on the zero inflated Poisson and the zero inflated generalized Poisson models, which are special cases of the ZI model (2.7).

2.4.1 Zero Inflated Poisson (ZIP) Distribution

Probably the most known ZI model is the ZI Poisson (ZIP) model. The zero inflated Poisson distribution $ZIP(p, \mu)$ is defined by

$$P(Y = y|p, \mu) = \begin{cases} p + (1-p)\exp(-\mu), & \text{if } y = 0 \\ (1-p)\mu^y \exp(-\mu)\frac{1}{y!} & \text{if } y > 0 \end{cases}.$$

Using (2.8) and (2.9), mean and variance of the ZIP distribution are specified by

$$E(Y|p, \mu) = (1-p)\mu$$

and

$$Var(Y|p, \mu) = (1-p)\mu(\mu p + 1) = E(Y|p, \mu)(\mu p + 1).$$

For $p > 0$ the dispersion factor $\varphi := \mu p + 1$ of the ZIP model is positive, i.e. in this case the ZIP model allows for overdispersion.

2.4.2 Zero Inflated Generalized Poisson (ZIGP) Distribution

The ZIGP regression model was already introduced by Famoye and Singh (2003b), in Famoye and Singh (2003a) a generalisation to k-inflated GP regression models is given. A Bayesian analysis of the ZIGP model is presented in Angers and Biswas (2003), however they do not incorporate covariates. The zero inflated generalized Poisson distribution $ZIGP(p, \mu, \lambda)$ is obtained if the density function of the generalized Poisson distribution is chosen for $\pi(y|\theta)$, i.e. $\pi(y|\theta) = gp(y|\mu, \lambda)$, where

$$\begin{aligned} gp(y|\mu,\lambda) &:= P(Y=y|\mu,\lambda) \\ &= \mu[\mu + \frac{\lambda}{1-\lambda}y]^{y-1}(1-\lambda)^y \frac{1}{y!}\exp(-\mu(1-\lambda)-\lambda y),\ y>0 \end{aligned}$$

and

$$gp(0|\mu,\lambda) := P(Y=0|\mu,\lambda) = \exp(-\mu(1-\lambda)),$$

see (2.5). The mean and the variance of the ZIGP distribution are then given by

$$E(Y|p,\mu,\lambda) = (1-p)\mu$$

and

$$\begin{aligned} Var(Y|p,\mu,\lambda) &= p(1-p)\mu^2 + (1-p)\frac{\mu}{(1-\lambda)^2} = (1-p)\mu\Big[p\mu + \frac{1}{(1-\lambda)^2}\Big] \\ &= E(Y|p,\mu,\lambda)\Big[p\mu + \frac{1}{(1-\lambda)^2}\Big]. \end{aligned}$$

The dispersion factor of the ZIGP model is therefore given by $\varphi := p\mu + \frac{1}{(1-\lambda)^2}$. Here, overdispersion can both result from the overdispersion parameter λ of the GP distribution and the extra proportion of zeros p when $p > 0$.

2.4.3 Zero Inflated Regression Models

In a regression model $Y_i \sim ZIP(p_i, \mu_i)$ and $Y_i \sim ZIGP(p_i, \mu_i, \lambda)$, independent for $i = 1,..,n$, respectively, a regression can be performed both for $\mathbf{p} = (p_1,..,p_n)'$ and for $\boldsymbol{\mu} = (\mu_1,..,\mu_n)'$, assuming

$$p_i = \frac{\exp(\tilde{\mathbf{x}}_{\mathbf{i}}'\boldsymbol{\alpha})}{1+\exp(\tilde{\mathbf{x}}_{\mathbf{i}}'\boldsymbol{\alpha})} \quad \text{and} \quad \mu_i = t_i \exp(\mathbf{x_i}'\boldsymbol{\beta}),$$

where $\tilde{\mathbf{x}}_{\mathbf{i}} = (\tilde{x}_{i1},..,\tilde{x}_{im})'$ and $\mathbf{x_i} = (x_{i1},..,x_{ik})', i = 1,..,n$, are vectors of covariates which may but do not have to contain the same covariates and $\boldsymbol{\alpha} = (\alpha_1,..,\alpha_m)'$ and $\boldsymbol{\beta} = (\beta_1,..,\beta_k)'$ are the corresponding vectors of unknown parameters. However, using the same covariates for both the regression on μ and p may cause problems due to overfitting of the model. The focus in this paper will be on ZI regression models assuming a constant p and performing a regression on the mean μ_i only.

3 Spatial effects using a Gaussian conditional autoregressive model

Pettitt et al. (2002) introduce a Gaussian conditional autoregressive model to model spatial dependencies, we consider a special case of their model here. Assume J regions $\{1,...,J\}$ and let

$\boldsymbol{\gamma} = (\gamma_1, .., \gamma_J)'$ denote the vector of spatial effects for each region. Then $\boldsymbol{\gamma}$ is assumed to follow a multivariate normal distribution, in particular

$$\boldsymbol{\gamma} \sim N(0, \sigma^2 Q^{-1}) \tag{3.1}$$

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

$$Q_{ij} = \begin{cases} 1 + |\psi| \cdot N_i & i = j \\ -\psi & i \neq j, i \sim j \\ 0 & \text{otherwise} \end{cases} . \tag{3.2}$$

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 conditional distribution of γ_i, given all the remaining components $\boldsymbol{\gamma}_{-i}, i = 1, .., J$ is then given by

$$\gamma_i | \boldsymbol{\gamma}_{-\boldsymbol{i}} \sim N\Big(\frac{\psi}{1 + |\psi| \cdot N_i} \sum_{j \sim i} \gamma_j, \sigma^2 \frac{1}{1 + |\psi| \cdot N_i}\Big). \tag{3.3}$$

The parameter ψ determines the degree of spatial dependence, for $\psi = 0$ all regions are spatially independent, whereas for $\psi \to \infty$ the degree of dependence increases. Pettitt et al. (2002) show that Q is symmetric and positive definite, therefore (3.1) is a proper distribution. Another convenient feature of this Gaussian conditional autoregressive model is that the determinant of Q which is needed for the update of ψ in a Markov Chain Monte Carlo(MCMC) algorithm can be computed efficiently, see Pettitt et al. (2002) for details.

3.1 Related CAR models

Many other authors have dealt with conditional autoregressive models. An overview about CAR models is given in the book by Banerjee et al. (2004) and in Jin et al. (2004) where also multivariate CAR models are discussed. The most popular model is probably the intrinsic CAR model introduced by Besag and Kooperberg (1995) where the full conditional of γ_i given $\boldsymbol{\gamma}_{-\boldsymbol{i}}$ is given by

$$\gamma_i | \boldsymbol{\gamma}_{-\boldsymbol{i}} \sim N\Big(\sum_{j \sim i} \frac{\gamma_j}{N_i}, \frac{\sigma^2}{N_i}\Big). \tag{3.4}$$

This model can be extended to the weighted version

$$\gamma_i | \boldsymbol{\gamma}_{-\boldsymbol{i}} \sim N\Big(\sum_j \frac{\rho_{ij}}{\sum_j \rho_{ij}} \gamma_j, \frac{\sigma^2}{\sum_j \rho_{ij}}\Big), \tag{3.5}$$

where ρ_{ij} are the elements of a symmetric positive-definite matrix $\boldsymbol{\varrho} = (\varrho_{ij})_{i,j=1,..,J}$ with

$$\varrho_{ij} = \begin{cases} \varrho(d_{ij}), & i \neq j \\ 0, & i = j \end{cases} .$$

d_{ij} denotes the euclidean distance between region i and j. If ρ_{ij} is chosen as

$$\varrho_{ij} = \begin{cases} 1 & i \sim j \\ 0 & \text{otherwise} \end{cases} , \tag{3.6}$$

this model reduces to the unweighted intrinsic CAR model (3.4). The joint density for $\boldsymbol{\gamma}$ in the intrinsic CAR model is improper in contrast to model (3.3) described above which has a proper joint density.
Czado and Prokopenko (2004) consider a modification of model (3.3) given by

$$\gamma_i|\boldsymbol{\gamma_{-i}} \sim N\Big(\frac{\psi}{1+|\psi|\cdot N_i}\sum_{j\sim i}\gamma_j, \sigma^2\frac{1+|\psi|}{1+|\psi|\cdot N_i}\Big) \tag{3.7}$$

where the conditional variance is multiplied by the additional term $1+|\psi|$. This is a proper model as well but in the limit $\psi \to \infty$ it reduces to the intrinsic CAR model.
Another modification of the intrinsic CAR model has been presented by Sun et al. (2000). They introduce the parameter $|\varrho| < 1$ such that

$$\gamma_i|\boldsymbol{\gamma_{-i}} \sim N\Big(\varrho\sum_{j}\frac{\rho_{ij}}{\sum_j \rho_{ij}}\gamma_j, \frac{\sigma^2}{\sum_j \rho_{ij}}\Big) \tag{3.8}$$

to get a proper multivariate normal distribution for $\boldsymbol{\gamma}$. For ρ_{ij} as in (3.6) the simplified model

$$\gamma_i|\boldsymbol{\gamma_{-i}} \sim N\Big(\varrho\sum_{j\sim i}\frac{\gamma_j}{N_i}, \frac{\sigma^2}{N_i}\Big) \tag{3.9}$$

is obtained. Here, the intrinsic CAR model is the limiting case for $\varrho = 1$. A multivariate version of this model has also been used by Gelfand and Vounatsou (2003) to model spatial effects in hierarchical models.

4 Bayesian Inference

Parameter estimation in the regression set up is done using a Bayesian approach. For more information on Bayesian data analysis and MCMC methods see Gilks et al. (1996) and Gelman et al. (2004). Assume the claim frequencies Y_i, $i = 1,..,n$ to be observed at J regions. Besides the well known Poisson regression model $Poi(\mu_i)$ we consider the $NB(r,\mu_i), GP(\mu_i,\lambda)$, $ZIP(p,\mu_i)$ and $ZIGP(p,\mu_i,\lambda)$ models described in the previous sections. In each of these models the mean μ_i, $i = 1,..,n$ is specified by

$$\mu_i = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_{R(i)}), \tag{4.1}$$

where $\mathbf{x}_i = (1, x_{i1},..,x_{ik})$ denotes the vector of covariates for the ith observation and $\boldsymbol{\beta} = (\beta_0,..,\beta_k)$ the vector of unknown regression parameters. Note, that an intercept β_0 is included in the model. To allow for geographical differences in the J regions spatial effects $\boldsymbol{\gamma} = (\gamma_1,..,\gamma_J)$ are introduced, where $R(i) \in \{1,..,J\}$ denotes the region of the i-th observation. t_i gives the exposure time for the ith observation and is treated as an offset. For the zero inflated models we assume a constant p for all observations first. The parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, λ, p and r respectively are taken to be a priori independent and the following prior distributions are chosen:

- $\pi(\boldsymbol{\beta}) \sim N(0, \tau^2 I_{k+1})$, with $\tau^2 = 100$
- $\pi(\boldsymbol{\gamma}|\sigma^2, \psi) \sim N(\mathbf{0}, \sigma^2 Q^{-1})$ with Q specified as in (3.2)

For the hyperparameters σ^2 and ψ the proper priors

- $\pi(\sigma^2) \sim IGamma(a, b)$ with $a = 1$ and $b = 0.005$
- $\pi(\psi) \sim \frac{1}{(1+\psi)^2}$

are assumed. For the model specific parameters the following prior distributions are chosen:

- GP Regression: $\pi(\lambda) \sim U([0, 1])$
- NB Regression : $\pi(r) \sim Gamma(a, b)$, i.e. $\pi(r) = \frac{b^a}{\Gamma(a)} r^{a-1} e^{-rb}$, where $a = 1$ and $\pi(b) \sim Gamma(c, d)$, i.e. $\pi(b) \propto b^{c-1} e^{-bd}$ with $c = 1$ and $d = 0.005$
- ZIP/ZIGP Regression: $\pi(p) \sim U([0, 1])$

Details on the MCMC algorithms can be found in the Appendix.

5 Bayesian Model comparison

5.1 Deviance Information Criterion (DIC)

Spiegelhalter et al. (2002) suggest to use the following criterion for model comparison in Bayesian inference. Assume a probability model $p(\mathbf{y}|\boldsymbol{\theta})$. The Bayesian deviance $D(\boldsymbol{\theta})$, which is used as a measure for goodness of fit, is defined as

$$D(\boldsymbol{\theta}) = -2 \log p(\mathbf{y}|\theta) + 2 \log f(\mathbf{y})$$

where $f(\mathbf{y})$ is some fully specified standardizing term. To measure the model complexity Spiegelhalter et al. (2002) introduce the effective number of parameters p_D defined by

$$\begin{aligned} p_D &:= E[D(\boldsymbol{\theta}|\mathbf{y})] - D(E[\boldsymbol{\theta}|\mathbf{y}]) \\ &= \text{posterior mean of the deviance} - \text{deviance of the posterior means.} \end{aligned}$$

Finally they define the deviance information criterion (DIC) as the sum of the posterior mean of the deviance and the effective number of parameters

$$DIC := E[D(\boldsymbol{\theta}|\mathbf{y})] + p_D.$$

According to this criterion the model with the smallest DIC is to be preferred. p_D and DIC are easily computed using the available MCMC output by taking the posterior mean of the deviance to obtain $E[D(\boldsymbol{\theta}|\mathbf{y})]$ and the plug-in estimate of the deviance $D(E[\boldsymbol{\theta}|\mathbf{y}])$ using the posterior means $E[\boldsymbol{\theta}|\mathbf{y}]$ of the parameter $\boldsymbol{\theta}$.
Bayes factors based on marginal likelihood provide an alternative method for model comparison, see Kass and Raftery (1995). Further, Bayesian Model Averaging (BMA), see for example Hoeting et al. (1999), which is based on Bayes factors, presents a method for model selection taking model uncertainty into account. However, since the computation of Bayes factors requires substantial efforts for complex hierarchical models, see Han and Carlin (2001), we use the DIC for model comparison in this paper.

5.2 Model checking using posterior predictive p-values

Gelman et al. (1996) propose a method for model checking using the posterior predictive distribution for a discrepancy measure. Stern and Cressie (2000) use this method for model checking in disease mapping models. Assume a model with likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta}$ and posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. The posterior predictive distribution for a replication $\mathbf{y}^{rep}$ of the observed data $\mathbf{y}_{obs}$ is defined as

$$p(\mathbf{y}^{rep}|\mathbf{y}_{obs}) = \int p(\mathbf{y}^{rep}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}.$$

Choose a discrepancy measure $D(\mathbf{y}, \boldsymbol{\theta})$ which may depend on $\boldsymbol{\theta}$. Then the posterior predictive p-value for this discrepancy measure is given by

$$p_b(\mathbf{y}_{obs}) = P(D(\mathbf{y}^{rep}, \boldsymbol{\theta}) \geq D(\mathbf{y}_{obs}, \boldsymbol{\theta})|\mathbf{y}_{obs})).$$

This posterior predictive p-value can be computed as follows using a set of MCMC draws $\boldsymbol{\theta}^j, j = 1,.., J$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{obs})$.

- For each draw $\boldsymbol{\theta}^j, j = 1,.., J$ simulate a replicated data set $\mathbf{y}^{repj}$ according to its sampling distribution $p(\mathbf{y}^{rep}|\boldsymbol{\theta}^j)$
- Compute the discrepancy measures $D(\mathbf{y}^{repj}, \boldsymbol{\theta}^j)$ and $D(\mathbf{y}_{obs}, \boldsymbol{\theta}^j)$

If the discrepancy measure $D(\mathbf{y}, \boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, Gelman et al. (1996) propose to produce a scatterplot of the pairs $(D(\mathbf{y}_{obs}, \boldsymbol{\theta}^j), D(\mathbf{y}^{repj}, \boldsymbol{\theta}^j)), j = 1,.., J$. If the model fit is good, about half of the points should fall below and half above the 45 degree line. The p-value can then be estimated by the proportion of pairs for which $D(\mathbf{y}^{repj}, \boldsymbol{\theta}^j) \geq D(\mathbf{y}_{obs}, \boldsymbol{\theta}^j)$, i.e.

$$\hat{p}_b = \frac{1}{J}\sum_{j=1}^{J} I(D(\mathbf{y}^{repj}, \boldsymbol{\theta}^j) \geq D(\mathbf{y}_{obs}, \boldsymbol{\theta}^j)),$$

where $I(\cdot)$ denotes the indicator function. An extreme p-value close to 0 or 1 indicates a lack of fit of the model according to the chosen discrepancy measure. If $D(\mathbf{y}, \boldsymbol{\theta})$ is independent of $\boldsymbol{\theta}$ a histogram of $D(\mathbf{y}^{repj})$ can be displayed and compared to $D(\mathbf{y}_{obs})$.
Possible discrepancy measures include the

- Deviance: $D(\mathbf{y}, \boldsymbol{\theta}) = Dev(\mathbf{y}, \boldsymbol{\theta}) = -2\log p(\mathbf{y}|\boldsymbol{\theta})$
- χ^2-discrepancy: $D(\mathbf{y}, \boldsymbol{\theta}) = \chi^2(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{(y_i - E(Y_i|\boldsymbol{\theta}))^2}{Var(Y_i|\boldsymbol{\theta})}$
- $D(\mathbf{y}) = DIC(\mathbf{y})$.

6 Application

6.1 Data description

In this section an application to a data set of policyholders in Bavaria, with full comprehensive car insurance for one year is presented. There are several covariates given in the data like age and gender of the policyholder, kilometers driven per year, type of car, age of car and the region in which each policyholder is living in. We analyse a subset of these data, in particular we only consider traffic accident data for policyholders with one particular, average type of car. The resulting data set contains about 16300 observations. There is a very large amount of observations with no claim in the dataset and the maximum number of claims is only 4, see Table 1. We first analysed the data set in Splus using a Poisson model without spatial effects to

number of claims	number of observations	percentage
0	15576	0.955
1	692	0.042
2	36	0.002
3	2	$1.3 \cdot 10^{-4}$
4	1	$6.1 \cdot 10^{-5}$

Table 1: Summary of the observed claim frequencies in the data

identify significant covariates and interactions and then used this model as a starting model in our MCMC algorithms. Seven covariates, including age, gender and kilometers driven per year and one interaction turned out to be significant for explaining claim frequency. Since Bavaria is divided into 96 regions, 96 spatial effects are introduced.

6.2 Models

The data is analysed using a simple Poisson regression, a negative binomial and a generalized Poisson regression model. Due to the large amount of zero observations a zero inflated Poisson and a zero inflated generalized Poisson regression model are fitted as well. In the context of car insurance data, zero inflated models allow for two different interpretations for the occurrence of zero observations. On the one hand, a zero observation may simply state that no claim occurred, on the other hand it is also possible that a claim occurred but was not reported in order to preserve the insurers no-claims bonus. The proportion of the not reported claims is modelled by the parameter p in the ZI models.

The MCMC algorithms for these 5 models are run for 6000 iterations using a burnin of $b = 1000$ iterations. The mean and variance structure of the models are summarized here for reference once again with μ_i, $i = 1, .., n$ specified as in (4.1):

$$Y_i \sim \text{Pois}(\mu_i): \quad E(Y_i) = \mu_i \quad Var(Y_i) = E(Y_i) = \mu_i \tag{6.1}$$

$$Y_i \sim NB(\mu_i, r): \quad E(Y_i) = \mu_i, \quad Var(Y_i) = \mu_i(1 + \frac{\mu_i}{r}) \tag{6.2}$$

$$Y_i \sim GP(\mu_i, \lambda): \quad E(Y_i) = \mu_i, \quad Var(Y_i) = \frac{\mu_i}{(1-\lambda)^2} \tag{6.3}$$

$$Y_i \sim ZIP(p, \mu_i): \quad E(Y_i) = (1-p)\mu_i, \quad Var(Y_i) = E(Y_i)(p\mu_i + 1) \tag{6.4}$$

$$Y_i \sim ZIGP(p, \mu_i, \lambda): \quad E(Y_i) = (1-p)\mu_i, \quad Var(Y_i) = E(Y_i)\Big(p\mu_i + \frac{1}{(1-\lambda)^2}\Big) \tag{6.5}$$

$$Y_i \sim ZIP(p_i, \mu_i): \quad E(Y_i) = (1-p_i)\mu_i, \quad Var(Y_i) = E(Y_i)(p_i\mu_i + 1) \tag{6.6}$$

The estimated posterior means together with the 2.5% and 97.5% quantiles for the regression parameters given in Table 2 are very similar for all models. In the ZIP and ZIGP regression models a higher estimated intercept $\hat{\beta}_0$ is observed since about half ($\hat{p} = 0.46, 0.34$, see Table 4) of the zero observations are assumed to be extra zeros. Results for the ZIP regression model (6.6) with additional regression for p are given in Section 6.2.1.

Model	γ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
Poi	yes	-8.39	15.36	-8.90	0.08	-0.15	0.20	-0.27	0.13	0.08
(6.1)		-8.53	7.36	-16.96	0.02	-0.24	0.06	-0.39	0.06	0.02
		-8.26	23.29	-1.14	0.16	-0.07	0.35	-0.15	0.20	0.13
	no	-8.37	15.53	-8.79	0.08	-0.15	0.22	-0.26	0.14	0.08
		-8.47	7.61	-16.87	0.01	-0.24	0.08	-0.38	0.07	0.02
		-8.27	23.22	-1.06	0.15	-0.07	0.36	-0.14	0.20	0.13
NB	yes	-8.37	15.45	-8.94	0.08	-0.15	0.21	-0.27	0.14	0.08
(6.2)		-8.49	6.92	-17.27	0.01	-0.24	0.06	-0.40	0.07	0.02
		-8.26	23.87	-0.93	0.15	-0.07	0.36	-0.14	0.20	0.14
	no	-8.36	15.42	-8.81	0.08	-0.15	0.22	-0.26	0.13	0.08
		-8.47	6.90	-16.81	0.01	-0.24	0.06	-0.39	0.06	0.18
		-8.25	23.70	-0.84	0.15	-0.07	0.37	-0.14	0.21	0.14
GP	yes	-8.37	15.77	-8.62	0.08	-0.15	0.20	-0.26	0.14	0.07
(6.3)		-8.48	7.76	-16.55	0.01	-0.23	0.05	-0.38	0.07	0.01
		-8.26	23.72	-0.67	0.15	-0.06	0.34	-0.13	0.20	0.12
	no	-8.36	15.81	-8.48	0.08	-0.15	0.20	-0.26	0.14	0.07
		-8.47	7.55	-16.71	0.01	-0.23	0.06	-0.39	0.07	0.01
		-8.26	23.75	-0.36	0.15	-0.06	0.35	-0.14	0.20	0.13
ZIP	yes	-7.75	15.17	-8.74	0.08	-0.15	0.22	-0.26	0.13	0.08
(6.4)		-8.04	6.66	-16.82	0.01	-0.24	0.05	-0.40	0.07	0.02
		-7.45	23.09	-0.63	0.15	-0.06	0.38	-0.12	0.20	0.14
	no	-7.77	15.46	-8.78	0.08	-0.15	0.22	-0.26	0.13	0.08
		-8.10	7.36	-16.81	0.01	-0.24	0.05	-0.41	0.06	0.02
		-7.47	23.48	-1.10	0.16	-0.07	0.38	-0.12	0.20	0.14
ZIGP	yes	-7.94	15.53	-8.69	0.08	-0.15	0.21	-0.26	0.14	0.08
(6.5)		-8.35	7.27	-16.81	0.01	-0.24	0.04	-0.40	0.07	0.02
		-7.56	23.85	-0.71	0.15	-0.06	0.37	-0.12	0.21	0.13
	no	-7.91	15.49	-8.78	0.08	-0.15	0.21	-0.26	0.14	0.08
		-8.30	7.76	-16.95	0.01	-0.24	0.04	-0.41	0.07	0.02
		-7.51	23.27	-0.78	0.15	-0.07	0.38	-0.12	0.20	0.13

Table 2: Estimated posterior means (first row) together with 2.5% (second row) and 97.5 % (third row) quantiles for the regression parameters of Models (6.1)-(6.5) with and without spatial effects.

Further note that the range of the estimated posterior mean spatial effects given in Table 3 is much larger in the Poisson model compared to the other models. Apparently, since the Poisson model is the only model not taking overdispersion into account, some extra variation present in the data is explained via the spatial effects here.

For a comparison of the estimated overdispersion in the different models, we consider the

Model	$[\min_j \hat{\gamma}_j \max_j \hat{\gamma}_j]$
Poisson (6.1)	$[-0.26, 0.30]$
NB (6.2)	$[-0.10, 0.15]$
GP (6.3)	$[-0.03, 0.06]$
ZIP (6.4)	$[-0.03, 0.06]$
ZIGP (6.5)	$[-0.04, 0.07]$
ZIP with p-regression(6.6)	$[-0.04, 0.07]$

Table 3: Range of estimated posterior means of the spatial effects of models (6.1)-(6.6)

Parameter	γ	mean	$\hat{\varphi}_i$				
					mean		
		(2.5%,97.5 %)	min	25%	50%	75%	max
r in	yes	1.12			1.048		
NB (6.2)		(0.59, 2.36)	1.000	1.020	1.045	1.070	1.472
	no	1.06			1.050		
		(0.59, 2.03)	1.000	1.021	1.047	1.073	1.536
λ in	yes	0.027			1.055		
GP (6.3)		(0.01, 0.05)					
	no	0.026			1.055		
		(0.01, 0.04)					
p in	yes	0.46			1.041		
ZIP (6.4)		(0.30, 0.59)	1.000	1.017	1.039	1.061	1.423
	no	0.44			1.040		
		(0.26, 0.58)	1.000	1.017	1.038	1.059	1.435
p in	yes	0.34					
ZIGP (6.5)		(0.04, 0.54)			1.051		
λ in	yes	0.012	1.024	1.036	1.050	1.064	1.300
ZIGP (6.5)		$(5 \cdot 10^{-4}, 0.03)$					
p in	no	0.35					
ZIGP (6.5)		(0.06, 0.56)			1.051		
λ in	no	0.011	1.024	1.035	1.050	1.064	1.303
ZIGP (6.5)		$(4 \cdot 10^{-4}, 0.03)$					

Table 4: Estimated posterior means for the model specific parameters in Models (6.2)-(6.5) with and without spatial effects, with the 2.5 % and 97.5 % quantiles given in brackets. Further the mean (upper row), range and quantiles (lower row) of the posterior mean of the dispersion factors $\hat{\varphi}_i$ are given.

estimated dispersion factors φ_i which are defined by $1+\frac{\mu_i}{r}$, $\frac{1}{(1-\lambda)^2}$, $(p\mu_i+1)$ and $p\mu_i+\frac{1}{(1-\lambda)^2}$ for the NB, GP, ZIP and ZIGP regression models. In particular, we compute the mean, minimum, maximum value and quantiles of the posterior means of the dispersion factors $\hat{\varphi}_i := \frac{1}{R-b}\sum_{j=b}^{R}\hat{\varphi}_i^j$

for each model, where $\hat{\varphi}_i^j$ denotes the j-th MCMC iterate for φ_i and R and b give the number of total iterations. The results are reported in Table 4. The average dispersion factors are fairly close in all models. Note, that the dispersion factor in the GP regression model (6.3) is the same for all observations, whereas it depends on the mean $\mu_i, i = 1,..,n$ and therefore is different for each observation in the other models. For a closer comparison the scatterplot of the posterior means of the mean
$\hat{\mu}_i = \frac{1}{R-b}\sum_{j=b}^{R}\hat{\mu}_i^j$, $i = 1,..,n$ with $\hat{\mu}_i^j = \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}}^j + \hat{\gamma}_{R(i)}^j)$ in the GP regression model (6.3) and the NB regression model (6.2) as well as the scatterplot of the posterior mean of the variances $\hat{Var}(y_i) = \frac{1}{R-b}\sum_{j=b}^{R}\frac{\hat{\mu}_i^j}{(1-\hat{\lambda}^j)^2}$ and $\hat{Var}(y_i) = \frac{1}{R-b}\sum_{j=b}^{R}\hat{\mu}_i^j(1+\frac{\hat{\mu}_i^j}{\hat{r}^j})$, $i = 1,..,n$ respectively are given in Figure 2. For increasing values of the mean, the NB model (6.2) tends to give slightly

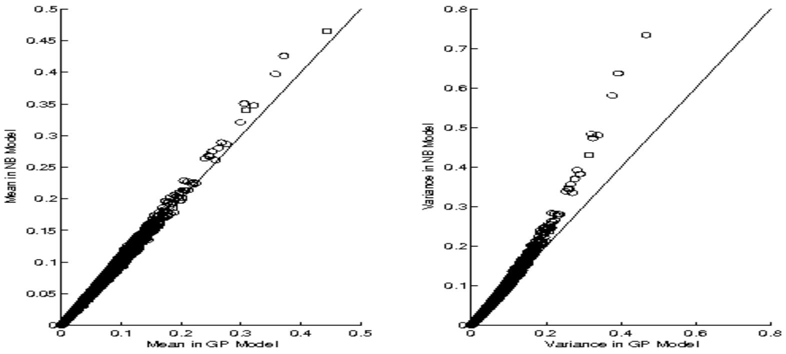

Figure 2: Scatterplots of the estimated posterior means $\hat{\mu}_i = \frac{1}{R-b}\sum_{j=b}^{R}\hat{\mu}_i^j$, $i = 1,..,n$ (left panel) and the posterior mean of the estimated variances $\hat{Var}(y_i) = \frac{1}{R-b}\sum_{j=b}^{R}\frac{\hat{\mu}_i^j}{(1-\hat{\lambda}^j)^2}$ in GP regression model (6.3) and $\hat{Var}(y_i) = \frac{1}{R-b}\sum_{j=b}^{R}\hat{\mu}_i^j(1+\frac{\hat{\mu}_i^j}{\hat{r}^j})$, $i = 1,..,n$ in the NB regression model (6.2) (right panel) with spatial effects, respectively.

higher estimates for the mean and this higher estimation is increased for the variance, since the mean is also involved in the dispersion factor of the negative binomial model. Similar results are obtained for both the ZIP and the ZIGP regression model compared to the GP regression model. Here as well, the posterior means of the variances are estimated slightly higher than in the GP model for increasing values of the means.
To get an idea of the pattern of the spatial effects, map plots are presented in the following. Since the results are very similar for all models only the plots for GP regression model (6.3) are

shown here. On the left hand side of Figure 3 a map of the posterior medians resulting from model (6.3) is given.

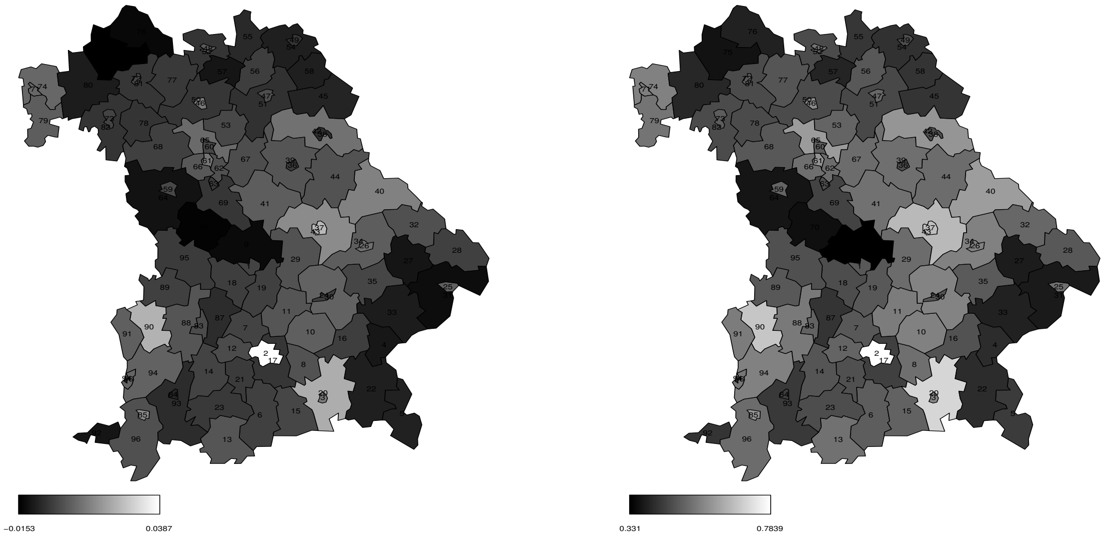

Figure 3: Map of the posterior medians (left) and of proportions of draws greater 0 (right) for the spatial effects for GP regression model (6.3).

On the right hand side in Figure 3, for each region j, j=1,..,J, the proportion of draws with the spatial effects greater than zero is plotted, i.e. $\frac{1}{R-b}\sum_{k=b}^{R} 1(\hat{\gamma}_j^k > 0)$, where $1(\cdot)$ denotes the indicator function. This is the posterior probability that the risk for a claim in region j is greater than average, i.e. $P(\gamma_j > 0|\mathbf{y})$. The spatial patterns in both maps are very similar. However, the range of the spatial effects (see Table 3) is very small and 0 is included in the 95 % credible interval for each spatial effect, suggesting that the spatial effects are not significant. Therefore we also run the MCMC algorithms for the above models without spatial effects, the corresponding estimated posterior means of the regression parameters are reported in Table 2, the estimates for the dispersion parameters can be found in Table 4.

6.2.1 ZIP Model with regression for p and spatial effects

According to the results from the ZIP and ZIGP models in the previous section the extra proportion of zeros p is rather high in the data set. Therefore it seems reasonable to perform a regression on p in order to check whether the occurrence of an extra zero depends on covariates as well. We consider the ZIP model (6.6) with regression on p here. Since in a model with the same covariates $\mathbf{x}_i$ both in the p- and $\boldsymbol{\mu}$-regression convergence problems occurred, each covariate is only incorporated either on the p- or the $\boldsymbol{\mu}$-level. For the p-regression a logistic link is chosen, spatial effects are incorporated in the $\boldsymbol{\mu}$-regression, i.e.

$$p_i = \frac{\exp(\tilde{\mathbf{x}}_i'\boldsymbol{\alpha})}{1+\exp(\tilde{\mathbf{x}}_i'\boldsymbol{\alpha})} \qquad \text{and} \qquad \mu_i = t_i \exp(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_{R(i)})$$

where $\tilde{\mathbf{x}}_i = (1, \tilde{x}_{i1}, .., \tilde{x}_{il})$ and $\mathbf{x}_i = (1, x_{i1}, .., x_{ik})$ are the vectors of non overlapping covariates, both including an intercept. We assume flat normal priors for the p-regression parameters $\boldsymbol{\alpha}$, in particular $\pi(\boldsymbol{\alpha}) \sim N(0, \tau^2 I_{l+1})$, with $\tau^2 = 100$. For the remaining parameters the prior distributions given in Section 4 are used. The final model altogether includes nine covariates, an intercept and three covariates on the p-level and an intercept and four covariates on the μ-level. This model was obtained by first analysing a ZIP model with regression on p only using all available covariates. Here, three covariates and the intercept turned out to have significant effects according to the 95% credible intervals. The remaining non significant covariates on the p-level were then incorporated for regression on the mean μ_i. The probability for an extra zero is higher if the policyholder's car is new and the number of kilometers driven per year is low. For an increasing deductible for the policyholder an extra zero is more likely as well. In contrast factors like age and gender of the policyholder are rather influential on the μ-level, the claim risk is higher for women and depends on age in a polynomial way. The range of the estimated posterior means of the spatial effects again is close to zero, see Table 3, similar to the results given in the previous sections. The average value of the estimated probabilities for an extra zero $\hat{p}_i, i = 1, .., n$ is 0.40, the minimum and maximum values of p_i, $i = 1, .., n$ are $\min_i p_i = 0.02$ and $\max_i p_i = 0.71$.

6.3 Model comparison using DIC

To check if the inclusion of spatial effects gives an overall improvement to our model and to compare the presented models, the DIC is considered. In Table 5, the resulting effective number of parameters and the DIC for each fitted model are given. Only in the Poisson regression case a well defined normalizing constant $f(y)$ (see Section 5.1) exists, while in all other models the likelihood of the saturated model depends on the unknown overdispersion parameters. Therefore we make the choice of setting the normalizing function $f(y)$ to 0. Consequently $E[D(\boldsymbol{\theta}|\mathbf{y})]$ is based only on the unscaled deviance which cannot be directly interpreted as a goodness of fit measure of a specific model. The model fit of a specific model however can be assessed using posterior predictive p-values which are given in the next Section.

The values of the DIC are very close for all models, with slightly higher values for the Poisson regression model, suggesting that the models allowing for overdispersion perform slightly better. However, since the DIC is almost the same for the models with and without spatial effects, the spatial effects do not have a significant effect in our data and therefore can be as well neglected. The effective number of parameters p_D is very close to the true number of parameters which is nine for the Poisson regression model (6.1) and ten for models (6.2)-(6.4) without spatial effects. Only for the ZIGP regression model (6.5) without spatial effects which includes 11 covariates, the effective number of parameter is estimated by ten. For the models including spatial effects p_D takes values between 13.78 and 40.80, while there are 9 regression parameters and 96 spatial effects driven by two hyperparameters in the model. This reflects the fact, that there is no large spatial variation present in the data. Note, that the DIC must be used with care here, since

Model	γ	DIC	$E[D(\boldsymbol{\theta}\|\mathbf{y})]$	p_D
Poisson (6.1)	yes	6200.6	6159.8	40.80
NB (6.2)	yes	6170.5	6149.9	20.61
GP(6.3)	yes	6174.9	6161.0	13.97
ZIP (6.4)	yes	6172.9	6159.1	13.78
ZIGP (6.5)	yes	6173.8	6159.5	14.33
ZIP with p-regression (6.6)	yes	6179.3	6165.2	14.05
Poisson (6.1)	no	6185.7	6177.0	8.74
NB(6.2)	no	6170.5	6160.7	9.81
GP(6.3)	no	6175.9	6166.2	9.71
ZIP (6.4)	no	6173.2	6163.2	9.97
ZIGP (6.5)	no	6174.1	6164.3	9.94
ZIP with p-regression (6.6)	no	6179.2	6170.6	8.67

Table 5: DIC and effective number of parameters p_D for the different models

originally the DIC is only defined for distributions of the exponential family which is not the case for all models considered here. However, the posterior mean of the deviance $E[D(\boldsymbol{\theta}|\mathbf{y})]$ given in Table 5 as well, can be used alternatively to compare the models, since the number of parameters is similar in all models. These values are very close as well for all models, the lowest values are obtained for the NB model in both cases with and without spatial effects. This indicates again that spatial effects are not significant and the models including overdispersion, in particular the NB regression model, might be preferred.
In the ZIP regression model with p-regression (6.6) altogether nine covariates were included. The resulting effective number of parameters p_D is 14.05 and 8.67 in model (6.6) with and without spatial effects, respectively. The values for the DIC are 6179.3 and 6179.2 respectively, indicating again that the spatial effects are not significant.

6.4 Model checking using posterior predictive p-values

In this section we assess the model fit using the method of posterior predictive p-values presented in Section 5.2. We use the deviance $D(\mathbf{y}, \boldsymbol{\theta}) = Dev(\mathbf{y}, \boldsymbol{\theta}) = -2 \log p(\mathbf{y}|\boldsymbol{\theta})$ and the DIC $D(\mathbf{y}) = DIC(\mathbf{y})$ as discrepancy measures for checking the model fit of Models (6.1)-(6.6) with and without spatial effects. The resulting posterior predictive p-values based on 500 replicated data sets are reported in Table 6.

The corresponding scatterplot and histogram for GP regression model (6.3) without spatial effects are given in Figure 4, similar plots are obtained for the remaining models.
Both the plots and the posterior predictive p-values indicate that the models give an adequate fit according to the deviance and the DIC. Only the p-values for the Poisson regression model (6.1) without and including spatial effects are rather small, indicating that the other models allowing for overdispersion might be preferred.

Model	γ	p-value	
		Dev	*DIC*
Poisson (6.1)	yes	0.018	0.076
NB (6.2)	yes	0.206	0.244
GP (6.3)	yes	0.170	0.212
ZIP (6.4)	yes	0.148	0.178
ZIGP (6.5)	yes	0.200	0.232
ZIP with p-regression (6.6)	yes	0.184	0.200
Poisson (6.1)	no	0.038	0.050
NB (6.2)	no	0.216	0.246
GP (6.3)	no	0.170	0.194
ZIP (6.4)	no	0.142	0.168
ZIGP (6.5)	no	0.164	0.194
ZIP with p-regression (6.6)	no	0.166	0.180

Table 6: Posterior predictive pvalues for the discrepancy measures *Dev* and *DIC* for Models (6.1)-(6.6) with and without spatial effects.

7 Conclusions

We have presented several regression models for count data both including spatial effects and allowing for overdispersion. Overdispersion is either modelled by the introduction of an additional parameter as in the negative binomial and generalized Poisson model, by allowing for an extra proportion of zero observations using zero inflated models or by combining zero inflated models with overdispersed distributions. All these models were applied to analyse the expected number of claims in a data set from a German car insurance company. For this data set no significant spatial effects were observed. However, according to the DIC and posterior predicted p-values, the models allowing for overdispersion gave a better fit than an ordinary Poisson regression model. The results for the overdispersed models are very close, a slight preference might be given to the NB regression model. Although the estimated percentage of extra zeros is rather high for the zero inflated models, the overall model fit is not improved by allowing two different sources of zeros for this data set.

A Appendix

In this section the algorithmic schemes for the discussed models are summarized. Most update steps are performed using a single component Metropolis Hastings step. For the proposal distributions either a random walk proposal or an independence proposal is used. In particular, for the independence proposal we take a t-distribution with $v = 20$ degrees of freedom with the same mode and the same inverse curvature at the mode as the target distribution.

A.1 GP regression model

- Sample $\lambda|\mathbf{y}, \boldsymbol{\beta}, \gamma$

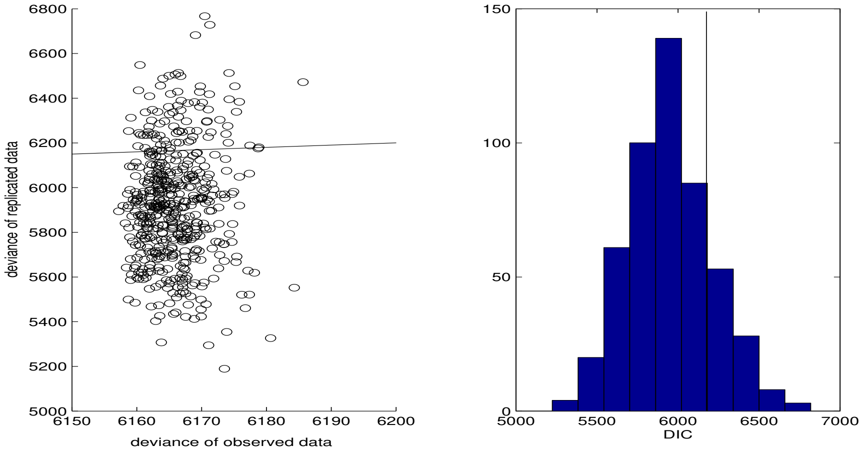

Figure 4: Scatterplot of $(Dev(\mathbf{y}^{repj}, \boldsymbol{\theta}^j), Dev(\mathbf{y}_{obs}, \boldsymbol{\theta}^j))$ where the line gives the 45 degree line and histogram of $DIC(\mathbf{y}^{repj}), j = 1, .., 500$ with $DIC(\mathbf{y}_{obs})$ indicated by the vertical line for GP regression model (6.3) without spatial effects.

- Sample $\beta_j | \mathbf{y}, \lambda, \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}, \qquad j = 0, .., k$
- Update of spatial effects
  - Sample $\frac{1}{\sigma^2} | \boldsymbol{\gamma}, \psi \sim Gamma$
  - Sample $\psi | \boldsymbol{\gamma}, \sigma$
  - Sample $\gamma_j | \mathbf{y}, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \qquad j = 1, .., J$

Since the full conditional of σ^2 is again Inverse Gamma, σ^2 can be sampled directly using a Gibbs step. For the remaining parameters a Metropolis Hastings step is used. In particular, the regression parameters $\boldsymbol{\beta}$ and the spatial effects $\boldsymbol{\gamma}$ are updated component by component using an independence proposal distribution. For the overdispersion parameter λ and the spatial hyper-parameter ψ a random walk proposal is taken. For the Poisson regression model the algorithmic scheme is the same, but with λ set fix to 0.

A.2 NB regression model

- Sample $r | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}$
- Sample $\beta_j | \mathbf{y}, r, \boldsymbol{\beta}_{-j}, \boldsymbol{\gamma}, \qquad j = 0, .., k$
- Update of spatial effects
  - sample spatial hyperparameters $\frac{1}{\sigma^2}$ and ψ as in A.1
  - Sample $\gamma_j | \mathbf{y}, r, \boldsymbol{\beta}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \qquad j = 1, .., J$

In the NB regression model $r, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are updated component by component using a MH step with random walk proposal.

A.3 ZI models

To avoid convergence problems in the ZI models which arose due to the observed correlation between the intercept β_0, p and λ, we use collapsed algorithms, in particular β_0, p and λ are updated with the latent variables $\mathbf{z}$ integrated out, i.e. based on model (2.7). Doing so convergence of the chains was improved a lot.

A.3.1 ZIP model without p-regression

- Updates with $\mathbf{z}$ integrated out
  - Sample $\beta_0|\mathbf{y}, p, \boldsymbol{\beta}, \boldsymbol{\gamma}$
  - Sample $p|\mathbf{y}, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}$
- Sample $z_i|\mathbf{y}, p, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma} \sim Bernoulli\Big(\dfrac{p}{p + (1-p)\exp(-\mu_i)}\Big)$ $\forall i$ with $y_i > 0$
- Sample $\beta_j|\mathbf{y}, \boldsymbol{\beta}_{-j}, \mathbf{z}, \boldsymbol{\gamma}, \qquad j = 1, .., k$
- Update of spatial effects
  - sample spatial hyperparameters $\frac{1}{\sigma^2}$ and ψ as in A.1
  - Sample $\gamma_j|\mathbf{y}, \beta_0, \boldsymbol{\beta}_{-0}, \mathbf{z}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \qquad j = 1, .., J$

The latent variables $\mathbf{z}$ can be updated using a Gibbs step. Since the full conditional of p is log concave, adaptive rejection sampling (ARS) introduced by Gilks and Wild (1992) is used to update p. For the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ a Metropolis Hastings step using an independence proposal distribution is performed.

A.3.2 ZIGP model without p-regression

- Updates with $\mathbf{z}$ integrated out
  - Sample $\beta_0|\mathbf{y}, p, \lambda, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}$
  - Sample $p|\mathbf{y}, \lambda, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}$
  - Sample $\lambda|\mathbf{y}, p, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}$
- Sample $z_i|\mathbf{y}, \lambda, p, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma} \sim Bernoulli\Big(\dfrac{p}{p + (1-p)\exp(-\mu_i)}\Big)$ $\forall i$ with $y_i > 0$
- Sample $\beta_j|\mathbf{y}, \boldsymbol{\beta}_{-j}, \mathbf{z}, \boldsymbol{\gamma}, \qquad j = 1, .., k$
- Update of spatial effects
  - sample spatial hyperparameters $\frac{1}{\sigma^2}$ and ψ as in A.1
  - Sample $\gamma_j|\mathbf{y}, \lambda, \beta_0, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \qquad j = 1, .., J$

For the ZIGP model the same proposal distributions as in the ZIP model are used. For λ an independence proposal is taken.

A.3.3 ZIP model with p-regression

- Updates with $\mathbf{z}$ integrated out
  - Sample $\beta_0|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}$
  - Sample $\alpha_j|\mathbf{y}, \boldsymbol{\alpha}_{-j}, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma}, \qquad j = 1, .., m$
- Sample $z_i|\mathbf{y}, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}_{-0}, \boldsymbol{\gamma} \sim Bernoulli\Big(\dfrac{p_i}{p_i + (1 - p_i)\exp(-\mu_i)}\Big) \quad \forall i$ with $y_i > 0$
- Sample $\beta_j|\mathbf{y}, \boldsymbol{\beta}_{-j}, \mathbf{z}, \boldsymbol{\gamma}, \qquad j = 1, .., k$
- Update of spatial effects
  - sample spatial hyperparameters $\frac{1}{\sigma^2}$ and ψ as in A.1
  - Sample $\gamma_j|\mathbf{y}, \beta_0, \boldsymbol{\beta}, \mathbf{z}, \boldsymbol{\gamma}_{-j}, \psi, \sigma, \qquad j = 1, .., J$

The parameters $\boldsymbol{\alpha}$ are updated in a MH step using independence proposals. The remaining parameters are updated as in section A.3.1.

Acknowledgement

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

Agarwal, D. K., A. E. Gelfand, and S. Citron-Pousty (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics 9*, 341–355.

Angers, J.-F. and A. Biswas (2003). A bayesian analysis of zero-inflated generalized poisson model. *Computational Statistics & Data Analysis 42*, 37–46.

Banerjee, S., B. Carlin, and A. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman & Hall/CRC.

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

Consul, P. (1989). *Generalized Poisson Distributions. Properties and Applications.* New York: Marcel Dekker, Inc.

Consul, P. and G. Jain (1973). A generalization of the poisson distributions. *Technometrics 15*, 791–799.

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

Dimakos, X. and A. Frigessi (2000). Hierarchical bayesian premium rating with latent structure. Statistical research report no.1, Department of mathematics, University of Oslo.

Famoye, F. and K. Singh (2003a). On inflated generalized poisson regression models. *Advances and Applications in Statistics 3 (2)*, 145–158.

Famoye, F. and K. Singh (2003b). Zero inflated generalized poisson regression model. *submitted*.

Gelfand, A. E. and P. Vounatsou (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics 4 (1)*, 11–25.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2004). *Bayesian Data Analysis, Second Edition*. Boca Raton: Chapman & Hall/CRC.

Gelman, A., X. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica 6*, 733–807.

Gilks, W., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman & Hall/CRC.

Gilks, W. and P. Wild (1992). Adaptive rejection sampling for gibbs sampling. *Applied Statistics 41 (2)*, 337–348.

Han, C. and B. Carlin (2001). Markov chain monte carlo methods for computing bayes factors: A comparative review. *Journal of the American Statistical Association 96*, 1122–1132.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14 (4)*, 382–417.

Jin, X., B. Carlin, and S. Banerjee (2004). Generalized hierarchical multivariate car models for areal data. *submitted to Biometrics*.

Johnson, N. L., S. Kotz, and A. W. Kemp (1993). *Univariate Discrete Distributions. Second Edition*. New York: Wiley.

Kass, R. and A. Raftery (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association 90*, 773–795.

Lambert, D. (1992). Zero-inflated poisson regression with and application to defects in manufacturing. *Technometrics 34 (1)*, 1–14.

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.

Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics 32 (2)*, 281–289.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B 64 (4)*, 583–640.

Stern, H. and N. Cressie (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine 19*, 2377–2397.

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.

Winkelmann, R. (2003). *Econometric Analysis of Count Data. 4th Edition.* Berlin: Springer-Verlag.