

# Hierarchical Bayesian Premium Rating with Latent Structure

X. K. Dimakos\* and A. Frigessi†

## Abstract

We propose a fully Bayesian approach to non-life risk premium rating, based on hierarchical models with latent variables for both claim frequency and claim size. Inference is based on the joint posterior distribution and is performed by Markov Chain Monte Carlo. Rather than plug-in point estimates of all unknown parameters, we take into account all sources of uncertainty simultaneously when the model is used to predict claims and estimate risk premiums. Several models are fitted to both a simulated dataset and a small portfolio regarding theft from cars. We show that interaction among latent variables can improve predictions significantly. We also investigate when interaction is not necessary. We compare our results with those obtained under a standard generalized linear model and show through numerical simulation that geographically located and spatially interacting latent variables can successfully compensate for missing covariates.

*Key words:* Claim frequency, claim size, geopricing, generalized linear models, generalized linear mixed models, hierarchical models, latent variables, Markov chain Monte Carlo, Markov random fields, non-life insurance, posterior credibility intervals, premium rating, risk premium, spatial smoothing, theft from car.

## 1 Introduction

The premium charged to policyholders is usually broken down to a part covering the insured risk (the risk premium) and further components related to administrative costs, marketing strategies and company profit. The risk premium is determined on the basis of the claims that the company expects during the period of exposure. In order to estimate the risk premium a statistical model is built, which aims to describe the stochastic relations between information available on the insured risk and policyholder, and the corresponding claims. In non-life insurance, Generalized Linear Models (GLM) and their variants are common statistical tools, see for instance McCullagh & Nelder (1989, pp. 204-208; 296-300), Brockman & Wright (1992) and Haberman & Renshaw (1996). An appropriate GLM is fitted to the historical data base of the insurance company, which contains covariate information and losses over a certain time period. Once the parameters in the model are estimated, it is standard practice to price the risk premium of a new insurance policy using the corresponding covariates and estimated parameters. This way, the uncertainty associated with these estimates, which is a consequence of the uncertainty regarding the adequacy of the GLM and data variation, is ignored. A remedy is found in the asymptotic theory of maximum likelihood estimation, see Brockman & Wright (1992, Appendix G). However, the asymptotic model may be inadequate and can induce serious bias, especially for claim types characterized by a low claim frequency.

---

\*Department of Mathematics, University of Oslo, P.O.Box 1053 Blindern, N-0316 Oslo, Norway. Email: xeni@math.uio.no

†Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway. Email: arnoldo.frigessi@nr.no

For this reason it is not common practice in the insurance business to associate confidence intervals to estimated risk premiums. But, ignoring the uncertainty of the risk premium reduces the room to maneuver of the company.

The first aim of this paper is to present a statistical framework where uncertainty is propagated to the predicted risk premium. We propose a Bayesian approach and give full account of all sources of randomness in the final estimate of the risk premium of a new policy, allowing the company to determine safety loadings, marketing strategies and financial policies on a more solid basis. Furthermore we compute posterior credibility intervals in a unified inferential framework. This is technically done by Monte Carlo sampling of the posterior distribution.

The second aim of this paper is to extend the toolbox of statistical models useful for risk premium estimation by introducing models with latent variables. The simplest case of a latent variable is an unknown regression intercept common to all policies that can be modeled by GLM. Such an intercept can be interpreted as an unknown covariate, which accounts for the basic common risk factor. If data on the geographical location of the policies are available, it is possible to differentiate the risk geographically by introducing a regionally varying intercept also in the GLM setting. This intercept is interpreted as a regional latent variable and allows for *geopricing*. More generally, latent variables are introduced in order to compensate for non-available information which would be useful in order to quantify the insured risk. Like all other unknown parameters, latent variables are estimated using the historical data base and used when pricing a new policy.

We extend the GLM framework by modeling the latent structure as random effects and we make prior assumptions on their joint distribution. The first step in this direction is Generalized Linear Mixed Models (GLMM), with latent variables that are sampled from a common multivariate normal distribution, see Clayton (1996). Furthermore, complex correlation structures are possible. For instance, when regional random effects are used, it might be appropriate to assume that these tend to vary smoothly from one region to the neighbouring ones. We borrow ideas from disease mapping and spatial statistics to construct Markov Random Field (MRF) models for the latent variables with these features. We build the models hierarchically, following assumptions of conditional independence. Estimation is done using Markov Chain Monte Carlo (MCMC) to sample from the joint posterior distribution of all parameters, including the latent variables.

The latent variables are not only substitutes for unavailable but relevant covariate information. They also favour the sharing of risk among policyholders characterized by the same level of such a latent variable. To enhance the mutual and social benefit of insurance, latent variables collect part of the individual risk of each policyholder in a certain homogeneous group and distribute it evenly over the same group. When there is no information on how to assign this risk to individuals, this is likely to be the best strategy.

In this paper we concentrate on regional latent variables and assume smoothing priors for them. Regionally differentiated pricing, sometimes called *geopricing*, has been addressed by Boskov & Verrall (1994) and Taylor (1989), amongst other. *Geopricing* is easily carried out within the framework we present.

The final aim of this paper is to demonstrate that spatially interacting latent variables are useful for risk premium estimation. This is shown in a simulation example where we investigate the possibility of recovering missing covariates by means of spatially interacting latent variables. The methodology is not only useful to price the risk premium of a policy but also in order to predict the total payouts of a large portfolio, which in turn is useful to take decisions on reserves and reinsurance. Our simulation study shows that latent variables

are useful for recovering missing covariates as prediction of single claims is improved when latent variables are included in the fitted model. However, for predicting the total payouts, there seems to be no significant advantage in introducing complex MRF interaction.

The paper is organized as follows. In Section 2 we describe the basic GLMs for claim frequency and claim size. We use the standard assumptions that the number of claims follows a Poisson distribution and that, conditional on the number of claims, the claim size follows a Gamma distribution. Also in Section 2 we describe the current practice of risk premium estimation. In Section 3 we describe the class of hierarchical GLMs with latent variables and prepare for Bayesian inference by describing prior models for parameters and latent variables. We then derive the posterior distributions on which inference is based. In Section 4 and 5 we describe the MCMC procedure adopted for inference and risk premium estimation. Here we show how credibility intervals for risk premiums can be computed and discuss various point estimates. In Section 6 we explain how we shall compare different models given new validation data. Section 7 contains the simulation experiment. Even if the data are fictitious, we have tried to make this a realistic example. In the simulated data about 12% of the policies have claims over a two year period of exposure. Our results indicate that MRF interaction among latent covariates can give significant benefits. In Section 8 we apply our methods to a small data set of policies on theft from private cars over a one year period. These data, together with a second independent dataset used for validation, are provided by the Norwegian insurance company Gjensidige Forsikring. The data are not easy to analyse as the claim frequency is very small (0.7% of the polices have claims). In this example hierarchical GLMs perform well. The average error per exposure year of the estimated premiums is between 17 and 21 NOK. However, it is not clear if structured MRF interaction among latent variables significantly improves rating performance. Section 9 ends the paper with a short conclusion. The appendix contains details of the MCMC implementation.

## 2 Models and Established Practice

### 2.1 Baseline GLM

We consider a portfolio of  $I$  insurance policies. For each policy  $i = 1, \dots, I$ , the number of claims  $N_i$  and the average claim size  $S_i$  are recorded for an exposure time  $e_i$ . In addition covariates  $\mathbf{x}_i$  characterizing the insured risk and policyholder are available. (Boldface is used to indicate vector-valued variables and all vectors are assumed to be column vectors.) The covariates may be continuous or categorical for describing qualitative factors. Since qualitative factors may be rewritten as zero-one variables we will not distinguish between these two types of covariates in our notation. Interaction terms among covariates (McCullagh & Nelder 1989, pp. 53, 58) can of course also be included.

GLMs extend classical regression by allowing: (i) the response variable to follow any distribution in the one-parameter exponential family; (ii) the additivity of the explanatory variables to occur on a monotone transformation of the mean, rather than the mean itself. Denoting by  $Y$  the response variable with mean  $\mu$ , the density of  $Y$  is on the form  $f(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) - c(y, \phi)\}$ , where  $\theta$  and  $\phi$  are the canonical and dispersion parameters and  $a(\cdot)$ ,  $b(\cdot)$  and  $c(\cdot)$  are known functions. The mean  $\mu$  is such that  $\eta = g(\mu) = \boldsymbol{\beta}'\mathbf{x}$ . The linear predictor  $\eta$  is a linear combination of the explanatory variables  $\mathbf{x}$  with parameters  $\boldsymbol{\beta}$  and is related to the mean by the monotone link function  $g(\cdot)$ .

We follow a traditional approach in the actuarial literature and assume that the number of claims  $N_i$  for policy  $i$  is Poisson distributed with individual intensity  $\lambda_i$  and expectation

$\mu_i = e_i \lambda_i$ , taking account of the exposure time  $e_i$ . The single claim sizes  $S_{ik}, k = 1, \dots, N_i$ , are assumed to be independent and Gamma distributed with expectation  $\xi_i$  and index parameter  $\nu$ . In this paper we use the parameterization of the Gamma density given in McCullagh & Nelder (1989, p. 287). Hence, the average claim size  $S_i = \sum_{k=1}^{N_i} S_{ik}/N_i, k = 1, \dots, N_i$  follows conditionally on  $N_i$  a Gamma distribution with the same expectation  $\xi_i$ . Given the parameter values,  $N_i, i = 1, \dots, I$  are assumed to be conditionally independent. Similarly for  $S_i$  given  $N_i$ . Summarizing we write

$$N_i \sim \text{Poisson}(e_i \lambda_i) \quad \text{and} \quad S_i | N_i \sim \text{Gamma}(\xi_i, \nu N_i) \quad i = 1, \dots, I. \quad (1)$$

Information on the policyholder and region is incorporated through covariates that influence the claim frequency and size through the intensity  $\lambda_i$  and the expectation  $\xi_i$  respectively. The number and type of covariates for the frequency and size component need not be the same. Typically some covariates are shared but most are unique. We make this explicit in notation by letting  $\mathbf{x}_i$  be the collection  $\{\mathbf{x}_i^F, \mathbf{x}_i^S\}$  of the covariates  $\mathbf{x}_i^F$  and  $\mathbf{x}_i^S$  thought to influence the claim frequency and size respectively. The covariates enter via the linear predictors  $\eta_i^F = \boldsymbol{\alpha}' \mathbf{x}_i^F$  and  $\eta_i^S = \boldsymbol{\beta}' \mathbf{x}_i^S$ . For both the claim frequency and size model we choose a log-link as this provides multiplicative models with  $\lambda_i = \exp(\eta_i^F)$  and  $\xi_i = \exp(\eta_i^S)$ . For the Poisson model the exposure time enters as an *offset* since  $\log(E(N_i)) = \log(e_i) + \eta_i$ . Multiplicative models are easy to interpret and use. Alternatively, additive models could be chosen. However, for the Poisson model, an additive model can give rise to negative fitted values of  $\lambda_i$ . Brockman & Wright (1992) argue that multiplicative models produce a more reasonable differentiation between policies.

By introducing a dummy unity covariate for all individuals so that  $\eta_i^F = \alpha_0 + \boldsymbol{\alpha}' \mathbf{x}_i^F$  (similarly for claim size) a risk parameter  $\alpha_0$  shared by all individuals is introduced. If the policies are located on  $J$  geographical regions, regional specific intercepts are modeled by introducing  $J$  dummy zero-one variables for regional belonging so that  $\eta_i^F = \alpha_{R(i)} + \boldsymbol{\alpha}' \mathbf{x}_i^F$ , where  $R(i)$  denotes the region where policy  $i$  is located. More generally,  $R(i)$  might indicate a group to which policy  $i$  belongs.

A few comments on the Poisson/Gamma assumptions are due. There are claim types for which the Poisson/Gamma model might not be well suited. As an example consider fire insurance of major industrial sites. Large fires are very rare but once they occur, losses are huge. A more reasonable model for claim frequency might be the binomial distribution, while claim sizes typically have a heavy tail and should be modeled using an extreme value distribution.

GLMs have been considered for application to insurance by several authors. McCullagh & Nelder (1989, pp. 204-208; 296-300) apply a Poisson model with log link to marine insurance data on the damages to certain cargo carrying vessels from the Lloyd's Register of Shipping and a Gamma claim size distribution with reciprocal link function to the car insurance data of Baxter et al. (1980). Renshaw (1994) presents several GLMs for modeling the claim number and size in the presence of covariates. Particular attention is given to the Poisson and Gamma/Pareto distribution, and questions regarding the choice of link function and use of quasi likelihood are discussed. Haberman & Renshaw (1996) review applications of GLMs to actuarial problems including survival modeling, multistate models for health insurance and non-life premium rating and reserves calculations. The baseline GLM in the present paper is among the models considered for premium rating.

## 2.2 Inference and Premium Estimation

The *risk premium* is defined as  $\rho_i = e_i \lambda_i \xi_i$  for a policy  $i$ . In the simple setting of Section 2.1 it holds that  $\rho_i = \mathbb{E}(\sum_{k=1}^{N_i} S_{ik}) = \mathbb{E}(N_i)\mathbb{E}(S_i|N_i)$  where  $S_{ik}$ ,  $k = 1, \dots, N_i$  are single claim sizes. We call the total payout  $P_i = \sum_{k=1}^{N_i} S_{ik}$ .

In standard GLMs the parameters of the claim frequency and size models are estimated by iterative maximum likelihood. A classical algorithm for this is Fisher scoring, see McCullagh & Nelder (1989, Section 2.5). The standard point estimate of the risk premium for an insured risk with covariates  $\{\mathbf{x}_i^F, \mathbf{x}_i^S\}$  and exposure  $e_i$  is

$$\hat{\rho}_i = e_i \hat{\lambda}_i \hat{\xi}_i = e_i \exp(\hat{\boldsymbol{\alpha}}' \mathbf{x}_i^F + \hat{\boldsymbol{\beta}}' \mathbf{x}_i^S), \quad (2)$$

where  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\boldsymbol{\beta}}$  are the parameter estimates. Notice that an estimate of the parameter  $\nu$  is not needed to compute this point estimate of the premium. However, it is needed to describe the full distribution of the claim size and total payouts.

Classical maximum likelihood theory allows the creation of confidence intervals around (2), as described in Brockman & Wright (1992, Appendix G). Basically, it can be shown that  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\boldsymbol{\beta}}$  are asymptotically uncorrelated as  $I \rightarrow \infty$ , and as they are asymptotically normally distributed, also independent. Since  $\hat{\mu}_i$  and  $\hat{\xi}_i$  are functions of  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\boldsymbol{\beta}}$  respectively, they are also asymptotically independent. Moreover,  $\hat{\mu}_i$ ,  $\hat{\xi}_i$  and  $\hat{\rho}_i$  are asymptotically lognormal. Hence  $\text{Var}(\hat{\rho}_i) = \text{Var}(\hat{\mu}_i \hat{\xi}_i) = \text{Var}(\hat{\mu}_i) \xi_i^2 + \text{Var}(\hat{\xi}_i) \mu_i^2 + \text{Var}(\hat{\mu}_i) \text{Var}(\hat{\xi}_i)$  is estimated by plugging in  $\hat{\mu}_i$  and  $\hat{\xi}_i$  for  $\mu_i$  and  $\xi_i$  respectively and by estimating variances of the lognormal distributions. Confidence intervals are computed from the fitted lognormal distribution. Model bias and plug-ins introduce errors that are difficult to assess.

A traditional actuarial tool in premium rating is tables of rating factors. In our multiplicative models, the estimated rating factor for one level of a covariate, say  $x_r$ , is the exponent of the corresponding term in the estimated linear predictor, i.e.  $\exp(\hat{\alpha}_r x_r)$ . Once the claim frequency and size model are fitted, rating factors for both components of the premium are obtained. As seen in (2) the estimated premium for a policy is the product of the corresponding rating factors in the frequency and size model corrected by exposure time. Later, in Table 8, we derive rating factors for our theft from car data.

## 3 Models with Latent Structure

### 3.1 Hierarchical Generalized Linear Models

Lee & Nelder (1996) define hierarchical GLMs (HGLM) as GLMs where the linear predictor is allowed to have one or more random components in addition to the usual fixed effect  $\boldsymbol{\beta}'\mathbf{x}$ , so that  $\eta = g(\mu) = \boldsymbol{\beta}'\mathbf{x} + \mathbf{u}$ , where  $\mathbf{u}$  is an unobserved random effect that we call latent variable. The Generalized Linear Mixed Model (GLMM) is a HGLM with a normally distributed random effect  $\mathbf{u}$ . Interaction between fixed and random effects is discussed by Clayton (1996, pp. 289–291) who notes that an interaction term which contain factors that are random effects must itself be a random effect.

Let  $\gamma_{R(i)}^F$  be a random effect in the frequency model assigned to each policy  $i$  belonging to group  $R(i)$ . Such latent variables correspond to unmeasured covariates that are either group specific or associated with the policyholder but merely measured through an average effect in the group. In the examples we consider  $R(i)$  will be the region where policy  $i$  is located. More generally it may be any other criteria useful for grouping policies homogeneously. We distinguish latent variables in the frequency model from those in the claim size model, and

denote the latter by  $\gamma_{R(i)}^S$ . Observe that two different grouping criteria could be used for the two models. The linear predictors now take the form

$$\eta_i^F = \boldsymbol{\alpha}' \mathbf{x}_i^F + \gamma_{R(i)}^F \quad \text{and} \quad \eta_i^S = \boldsymbol{\beta}' \mathbf{x}_i^S + \gamma_{R(i)}^S, \quad (3)$$

where  $R(i) \in \{1, \dots, J\}$ . Naturally, the models may be extended to include latent variables for several different grouping criteria.

The inclusion of latent variables in the model is motivated by the following commonly encountered and related problems:

- (i) *Unmeasured covariates*: A common issue is missing information in the form of unreported covariates. A company may fail to collect information on important risk factors because these relate to questions that cannot be asked, or to answers that are unreliable. Furthermore, specific risk factors may be unknown.
- (ii) *Overdispersion*: As a consequence of lack of covariate information, GLMs do not explain variability in full. Including latent variables in the model amounts to introducing extra degrees of freedom and can lead to a better fit.
- (iii) *Risk sharing*: Latent variables enable risk to be distributed among all policies belonging to the same group, thus making prices more uniform.

Random effects have been considered in actuarial sciences before. Nelder & Verrall (1997) show how credibility theory may be encompassed within the theory of HGLMs. Tomberlin (1988) uses random effects to estimate accident frequencies for motor vehicles. Premium rating by geographical area has previously been done by Taylor (1989) and Boskov & Verrall (1994) and applied to the same dataset on Household Contents in Sydney. In Taylor (1989) rating is performed by fitting two-dimensional spline functions, while Boskov & Verrall (1994) use a Bayesian approach and Gibbs sampling to fit a geographically smoothing model. In these latter two papers, data are preprocessed to remove all risk components other than the spatial ones, a difficult and non-optimal approach as all risk factors should be fitted simultaneously. Also, loss-ratios are fitted and there is no separate model for claim frequency and size. Indeed, Boskov & Verrall (1994) state that such analysis would be preferable.

### 3.2 Prior Distributions for Latent Variables

A fundamental part of Bayesian inference is the design of appropriate prior distributions for the unknown parameters. We consider two models for regional latent variables. The first model is a GLMM with conditionally independent regional latent variables specified by

$$\begin{aligned} \gamma_j^r | \mu_r, \sigma_r^2 &\sim N(\mu_r, \sigma_r^2), \quad j = 1, \dots, J, \quad r = F \text{ or } S. \\ \mu_r &\sim N(a_r, b_r^2) \\ 1/\sigma_r^2 &\sim \text{Gamma}(c_r, d_r), \end{aligned} \quad (4)$$

for region  $j = 1, \dots, J$  and where  $N(\cdot, \cdot)$  indicates the normal distribution. The hyperparameters  $a_r, b_r, c_r$  and  $d_r$  are given suitable values, chosen in such a way that if no further information is available, corresponding priors are quite flat. It is also possible to color this level in the hierarchy by assigning distributions to  $a_r, b_r, c_r$  and  $d_r$ .

The second model we consider for the latent variables is a Markov random field (MRF), see for example Besag et al. (1991). For this purpose, we need to introduce a neighbourhood

structure for the regions. For each region  $j$  let  $\delta_j$  denote its neighbouring regions, not including  $j$  itself. For instance, regions are neighbours if they share the same border or if they are within a certain distance of each other. A MRF prior model describing belief in spatially smooth latent variables has a density of the form

$$\gamma^r \sim \pi(\gamma^r) \propto \exp\left\{-\frac{\kappa}{2} \sum_j \sum_{k \in \delta_j} (\gamma_j^r - \gamma_k^r)^2\right\}, \quad r = F \text{ or } S. \quad (5)$$

The interaction or smoothing parameter  $\kappa$  determines the degree of spatial smoothing: the larger  $\kappa$  the more smoothness. With  $\kappa = \infty$  all the latent variables are a posteriori equal and with  $\kappa = 0$  they are a priori independent. These two models are equal to GLMs with one common intercept and  $J$  regional specific intercepts respectively. If indeed the unmeasured covariates exhibit a smooth spatial pattern, a model like (5) can give a better fit. It explains correlated residuals best and often allows for interesting interpretations. Observe that different neighborhood systems and smoothing parameter  $\kappa$  could be used for the frequency and size model. Pairwise interactions, like in (5) are often enough, although more complex MRFs can easily be introduced, see for example Tjelmeland & Besag (1998).

The MRF density (5) is in fact improper, but this is not a problem because the posterior that is used for estimation is proper. An important property of (5) is that, while it induces spatial homogeneity of the latent variables, it leaves the overall level undecided.

Mollié (1996) remarks that the regularity/irregularity of the topology is important when choosing a smoothing spatial prior. For very irregular topologies where some regions have many neighbours and others very few, it may be necessary to assign different smoothing parameters to regions with different numbers of neighbours, for example taking  $\kappa_j = \kappa/n_j$  where  $n_j$  is the number of neighbours of region  $j$ .

We shall not estimate  $\kappa$  in this paper, but perform a sensitivity analysis. To estimate  $\kappa$  is a difficult task, similar to the choice of smoothing parameter in nonparametric statistics. There are several methods, including cross validation, which allows to assess an appropriate value for  $\kappa$ , see Wand & Jones (1995), Fan & Gijbels (1996). A discussion of these methods goes beyond the scope of this paper.

Other possible models for latent variables include the conditional autoregressive (CAR) model, see Carlin & Louis (1996, pp. 263) and modeling sources of risk at unknown geographical locations with latent Gamma fields, see Wolpert & Ickstadt (1998).

### 3.3 Posterior Distribution

The posterior distribution is the tool for inference within the Bayesian framework. In our case we need the posterior distribution of the parameters and latent variables which is conditioned on the observed claim numbers and claim sizes. The posterior distribution fully describes the uncertainty associated with the parameters and latent variables. To obtain posterior estimates for the parameters of interest we shall compute their marginal posterior means (MPM). The joint posterior density for the model specified by (1) and (3) and latent variables with prior  $\pi(\gamma^F)$  and  $\pi(\gamma^S)$  as in (5) is by conditional independence

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma^F, \gamma^S, \nu | \mathbf{N}, \mathbf{S}) \propto \pi(\mathbf{N} | \boldsymbol{\alpha}, \gamma^F) \pi(\mathbf{S} | \mathbf{N}, \boldsymbol{\beta}, \gamma^S, \nu) \pi(\gamma^F) \pi(\gamma^S) \pi(\boldsymbol{\alpha}) \pi(\boldsymbol{\beta}) \pi(\nu) \quad (6)$$

where

$$\pi(\mathbf{N} | \boldsymbol{\alpha}, \gamma^F) \propto \exp\left\{\sum_{i=1}^I [-e_i \exp(\boldsymbol{\alpha}' \mathbf{x}_i^F + \gamma_{R(i)}^F) + N_i(\boldsymbol{\alpha}' \mathbf{x}_i^F + \gamma_{R(i)}^F)]\right\}$$

and

$$\pi(\mathbf{S}|\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\gamma}^S, \nu) \propto \exp\left\{\sum_{i=1}^I [-\log \Gamma(\nu N_i) + (\nu N_i)(\log(\nu N_i) - (\boldsymbol{\beta}' \mathbf{x}_i^S + \gamma_{R(i)}^S)) + (\nu N_i - 1) \log S_i - (\nu N_i S_i) / \exp(\boldsymbol{\beta}' \mathbf{x}_i^S + \gamma_{R(i)}^S)]\right\}.$$

The factors  $\pi(\boldsymbol{\alpha})$ ,  $\pi(\boldsymbol{\beta})$  and  $\pi(\nu)$  are assigned prior densities. We will use flat priors for  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$  and  $\nu$  in the applications of Sections 7 and 8.

Let  $\boldsymbol{\delta} = (\mu_F, \mu_S, \sigma_F, \sigma_S, a_F, a_S, b_F, b_S, c_F, c_S, d_F, d_S)$  be the collection of hyperparameters in (4). Assuming  $a_r, b_r, c_r, d_r$  for  $r = F, S$  to be fixed, we have by conditional independence in (4) that  $\pi(\boldsymbol{\delta}) = \pi_1(\mu_F|a_F, b_F)\pi_2(1/\sigma_F^2|c_F, d_F)\pi_1(\mu_S|a_S, b_S)\pi_2(1/\sigma_S^2|c_S, d_S)$  where  $\pi_1$  and  $\pi_2$  are the normal and Gamma density respectively. Furthermore, we have  $\pi(\boldsymbol{\gamma}^F|\mu_F, \sigma_F^2) = \prod_{j=1}^J \pi_1(\gamma_j^F|\mu_F, \sigma_F^2)$  and similarly for  $\pi(\boldsymbol{\gamma}^S|\mu_S, \sigma_S^2)$ . The posterior for the model specified by (1), (3) and (4) is then given by

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}^F, \boldsymbol{\gamma}^S, \nu, \boldsymbol{\delta}|\mathbf{N}, \mathbf{S}) &\propto \pi(\mathbf{N}|\boldsymbol{\alpha}, \boldsymbol{\gamma}^F)\pi(\mathbf{S}|\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\gamma}^S, \nu)\pi(\boldsymbol{\gamma}^F|\mu_F, \sigma_F^2)\pi(\boldsymbol{\gamma}^S|\mu_S, \sigma_S^2) \\ &\times \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\nu)\pi(\boldsymbol{\delta}). \end{aligned} \quad (7)$$

For, say  $\gamma_j^F$ , the marginal posterior mean (MPM) is given by  $E(\gamma_j^F|\mathbf{N})$  where  $\boldsymbol{\alpha}$  has been integrated out. Given the posterior distribution it is possible to define marginal posterior credibility  $100(1 - \alpha)\%$  intervals for each parameter or latent variable as the interval given by the lower and upper  $\alpha/2$  points of the corresponding marginal posterior. Joint posterior credibility intervals may also be of interest.

## 4 Markov Chain Monte Carlo Inference

### 4.1 The Metropolis-Hastings Algorithm

For a HGLM, estimation can no longer be done by iterative maximum likelihood with Fisher scoring. A complex structure for the latent variables in the model specified by (1) and (3), like the MRF prior (5), requires the use of Markov Chain Monte Carlo (MCMC) algorithms to compute MPM estimates. For the GLMM with prior (4), Breslow & Clayton (1993) propose two maximum likelihood based algorithms to do inference, but MCMC, is usually a good alternative. For all the models we consider, Bayesian inference can be done via MCMC. We shall do so even if Fisher scoring could be used when the latent variables follow prior (5) with  $\kappa = 0$  or  $\kappa = \infty$ .

It is not the purpose of this paper to explain MCMC. For a review of MCMC methods see Gilks et al. (1996). Applied to our setting, the idea of MCMC is to run an ergodic Markov chain which has as stationary distribution the posterior distribution (6) or (7) on the full parameter space (including the latent variables). Starting from arbitrary initial parameter values and latent variables, the chain is run until it is believed to have reached equilibrium. Convergence of the chain can be assessed using a variety of diagnostics, see Brooks & Roberts (1999) for reviews. Thereafter, the chain produces a sequence of dependent samples from the posterior distribution.

We have chosen to use a single-component Metropolis-Hastings algorithm in which one parameter or latent variable is updated at a time. A new value for the selected parameter or latent variable is sampled from a proposal distribution and accepted with a certain probability that guarantees ergodicity. In this paper, one step in the MCMC algorithm includes an update



of all the parameters and latent variables. Since the full posterior (6) is the product of the two posterior components  $\pi(\mathbf{N}|\boldsymbol{\alpha}, \boldsymbol{\gamma}^F)\pi(\boldsymbol{\gamma}^F)\pi(\boldsymbol{\alpha})$  and  $\pi(\mathbf{S}|\mathbf{N}, \boldsymbol{\beta}, \boldsymbol{\gamma}^S, \nu)\pi(\boldsymbol{\gamma}^S)\pi(\boldsymbol{\beta})$  for the claim frequency and size respectively, we may run two separate MCMC chains. Similarly for the posterior (7). This is computationally convenient since the two algorithms may then be run in parallel. A full account of the algorithm and expressions for acceptance probabilities are given in the appendix.

## 4.2 Inference

A crucial advantage of using MCMC is that it allows to approximate by sampling the full posterior distribution. Hence, finding credibility intervals requires no more effort than obtaining point estimates.

Assume that the Metropolis-Hastings algorithm has converged after  $T_0$  steps and that it is then run for an additional  $T$  steps to obtain  $T$  dependent samples from the posterior distribution. Marginal posterior means (MPM) are estimated by empirical sample means along the sampled MCMC trajectories. For instance, let  $\gamma_j^{F(t)}$  be the sampled value for the latent variable  $\gamma_j^F$  in the  $t$ -th MCMC step. By the law of large numbers for Markov chains

$$\hat{\gamma}_j^F = \frac{1}{T} \sum_{t=T_0+1}^{T_0+T} \gamma_j^{F(t)}, \quad j = 1, \dots, J, \quad (8)$$

approximates  $E(\gamma_j^F|\mathbf{N})$  for  $T$  large enough. Similarly for all other parameters. An estimate of the posterior variance  $\text{Var}(\gamma_j^F|\mathbf{N})$  is the empirical variance along the sampled trajectory, i.e.

$$\text{Var}(\hat{\gamma}_j^F|\mathbf{N}) \approx \frac{1}{T} \sum_{t=T_0+1}^{T_0+T} (\gamma_j^{F(t)} - \hat{\gamma}_j^F)^2.$$

Through the sampled trajectory  $\gamma_j^{F(t)}$ ,  $t = T_0 + 1, \dots, T_0 + T$ , we can estimate the marginal posterior density of  $\gamma_j^F$  given  $\mathbf{N}$  by the empirical density. Credibility intervals are created by taking lower and upper  $\alpha/2$  points  $\gamma_j^{F(\alpha/2)}$  and  $\gamma_j^{F(1-\alpha/2)}$  of the MCMC trajectory so that  $(\gamma_j^{F(\alpha/2)}, \gamma_j^{F(1-\alpha/2)})$  describes the posterior variability of  $\gamma_j^F$  at the specified  $1 - \alpha$  level. The parameters and latent variables are not only interesting per se. For instance, the expected posterior claim intensity  $E(\lambda_i|\mathbf{N}) = E(\exp(\eta_i^F)|\mathbf{N})$  for policy  $i$  can be easily approximated by

$$\frac{1}{T} \sum_{t=T_0+1}^{T_0+T} \exp(\eta_i^{F(t)}) = \frac{1}{T} \sum_{t=T_0+1}^{T_0+T} \exp((\boldsymbol{\alpha}^{(t)})' \mathbf{x}_i^F + \gamma_{R(i)}^{F(t)})$$

taken along the MCMC trajectory. The marginal posterior density of the claim intensity  $\lambda_i$  given  $\mathbf{N}$  incorporates all uncertainty, is easy to approximate and can be plotted. The posterior mean  $E(\lambda_i|\mathbf{N})$  should be contrasted to  $\exp\{E(\eta_i^F|\mathbf{N})\}$  for which a natural estimate is the plug-in  $\exp(\hat{\boldsymbol{\alpha}}' \mathbf{x}_i^F + \hat{\gamma}_{R(i)}^F)$ , where  $\hat{\boldsymbol{\alpha}} = \sum_{t=T_0+1}^{T_0+T} \boldsymbol{\alpha}^{(t)}/T$  and  $\hat{\gamma}_{R(i)}^F = \sum_{t=T_0+1}^{T_0+T} \gamma_{R(i)}^{F(t)}/T$ . By Jensen's inequality  $E(\exp(\eta_i^F)|\mathbf{N}) \geq \exp\{E(\eta_i^F|\mathbf{N})\}$ . Table 1 summarizes quantities of interest and the corresponding estimates based on MCMC for both the claim frequency and size model.

	posterior moment	MCMC samples	estimate
F	$E(\alpha_r   \mathbf{N})$ $E(\gamma_j^F   \mathbf{N})$ $e_i E\{\exp(\boldsymbol{\alpha}' \mathbf{x}_i^F + \gamma_{R(i)}^F)   \mathbf{N}\}$ $e_i \exp\{E(\boldsymbol{\alpha}' \mathbf{x}_i^F + \gamma_{R(i)}^F   \mathbf{N})\}$	$\{\alpha_r^{(t)}\}$ $\{\gamma_j^{F(t)}\}$ $\{\mu_i^{(t)}\} = \{e_i \exp((\boldsymbol{\alpha}^{(t)})' \mathbf{x}_i^F + \gamma_{R(i)}^{F(t)})\}$	$\hat{\alpha}_r = \sum_t \alpha_r^{(t)} / T$ $\hat{\gamma}_j^F = \sum_t \gamma_j^{F(t)} / T$ $\hat{\mu}_i^1 = \sum_t \mu_i^{(t)} / T$ $\hat{\mu}_i^2 = e_i \exp(\hat{\boldsymbol{\alpha}}' \mathbf{x}_i^F + \hat{\gamma}_{R(i)}^F)$
S	$E(\beta_r   \mathbf{S}, \mathbf{N})$ $E(\gamma_j^S   \mathbf{S}, \mathbf{N})$ $E\{\exp(\boldsymbol{\beta}' \mathbf{x}_i^S + \gamma_{R(i)}^S)   \mathbf{S}, \mathbf{N}\}$ $\exp\{E(\boldsymbol{\beta}' \mathbf{x}_i^S + \gamma_{R(i)}^S   \mathbf{S}, \mathbf{N})\}$	$\{\beta_r^{(t)}\}$ $\{\gamma_j^{S(t)}\}$ $\{\xi_i^{(t)}\} = \{\exp((\boldsymbol{\beta}^{(t)})' \mathbf{x}_i^S + \gamma_{R(i)}^{S(t)})\}$	$\hat{\beta}_r = \sum_t \beta_r^{(t)} / T$ $\hat{\gamma}_j^S = \sum_t \gamma_j^{S(t)} / T$ $\hat{\xi}_i^1 = \sum_t \xi_i^{(t)} / T$ $\hat{\xi}_i^2 = \exp(\hat{\boldsymbol{\beta}}' \mathbf{x}_i^S + \hat{\gamma}_{R(i)}^S)$
P	$E(\rho_i   \mathbf{N}, \mathbf{S})$ $e_i \exp\{E(\eta_i^F   \mathbf{N}) + E(\eta_i^S   \mathbf{S}, \mathbf{N})\}$	$\{\rho_i^{(t)}\} = \{\mu_i^{(t)} \xi_i^{(t)}\}$ $\{\boldsymbol{\alpha}^{(t)}\}, \{\boldsymbol{\beta}^{(t)}\}, \{\gamma^{F(t)}\}, \{\gamma^{S(t)}\}$	$\hat{\rho}_i^1 = \sum_t \hat{\rho}_i^{(t)} / T$ $\hat{\rho}_i^2 = \hat{\mu}_i^2 \hat{\xi}_i^2$

Table 1: *Summary of the estimation and simulation procedure for the frequency (F), size (S) and risk premium (P) components. The index of  $t$  runs over  $T$  steps in the MCMC sampler following the burn-in  $T_0$ , i.e.  $t = T_0 + 1, \dots, T + 1$ .*

Another level in the hierarchy can be introduced by sampling the posterior number of claims for policy  $i$  from a Poisson distribution with stochastic expectation  $\mu_i$ . This is done by sampling in step  $t$  a number  $N_i^{(t)}$  from a Poisson distribution with expectation  $\mu_i^{(t)}$  (see Table 1). Sampling claim sizes is done conditionally on the sampled claim numbers being non-zero and also involves the index parameter  $\nu$ . Hence,  $S_i^{(t)}$  is sampled from Gamma( $\xi_i^{(t)}, N_i^{(t)} \nu^{(t)}$ ) if  $N_i^{(t)} > 0$ , otherwise the claim size is zero. This introduces a further level of variability. While perhaps important in practice, for the purpose of this paper we shall not pursue this any further.

## 5 Price Predictions with MCMC

The risk premium  $\rho_i = e_i \lambda_i \xi_i = e_i \exp(\eta_i^F + \eta_i^S)$  has a posterior marginal distribution that describes the variability in full. The MCMC algorithm generates samples from the posterior density of  $\rho_i$  given  $\mathbf{N}$  and  $\mathbf{S}$  of the form  $\rho_i^{(t)} = \mu_i^{(t)} \xi_i^{(t)} = e_i \exp((\boldsymbol{\alpha}^{(t)})' \mathbf{x}_i^F + \gamma_{R(i)}^{F(t)}) \exp((\boldsymbol{\beta}^{(t)})' \mathbf{x}_i^S + \gamma_{R(i)}^{S(t)})$ , see Table 1. Estimates of posterior variance and credibility intervals are then constructed as described in Section 4.2. In the right column of Figure 7 we display such posterior densities and credibility intervals for some selected policies. We shall return to Figure 7 in Section 8. Here we only notice that the posterior marginal density visually describes the uncertainty around a risk premium prediction. We synthesize information in the posterior distribution through the posterior expectation  $E(\rho_i | \mathbf{N}, \mathbf{S})$  which is estimated consistently as  $T \rightarrow \infty$  by  $\hat{\rho}_i^1 = \sum_{t=T_0+1}^{T_0+T} \hat{\rho}_i^{(t)} / T$ . A natural alternative quantity of interest is  $e_i \exp\{E(\eta_i^F | \mathbf{N}) + E(\eta_i^S | \mathbf{S}, \mathbf{N})\}$ , estimated consistently as  $T \rightarrow \infty$  by  $\hat{\rho}_i^2 = \hat{\mu}_i^2 \hat{\xi}_i^2$ , where we plug posterior estimates of all parameters into (3) before taking exponents. The reason why we also keep track of these estimates, is that they become equal to (2) since for  $\kappa = 0$  and  $\kappa = \infty$  the hierarchical model reduces to standard GLM as we take flat priors on  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . As  $I \rightarrow \infty$  this is also true if the priors were informative. By Jensen's inequality  $E(\rho_i | \mathbf{N}, \mathbf{S}) \geq e_i \exp\{E(\eta_i^F | \mathbf{N}) + E(\eta_i^S | \mathbf{S}, \mathbf{N})\}$  which implies that  $\hat{\rho}_i^1$  gives a more conservative estimate of the risk premium. Table 1 gives an overview of posterior moments of interest and

estimators based on the MCMC simulations.

Both the suggested premium estimators are easily implemented in an automated pricing routine. To find  $\hat{\rho}_i^1$ , the trajectories  $\{\alpha^{(t)}\}, \{\beta^{(t)}\}, \{\gamma^F(t)\}$  and  $\{\gamma^S(t)\}$ ,  $t = T_0 + 1, \dots, T_0 + T$  are needed. These may be generated by running MCMC for each new pricing or a fixed sample may be stored. For  $\hat{\rho}_i^2$ , only the point estimates  $\hat{\alpha}, \hat{\beta}, \hat{\gamma}^F$  and  $\hat{\gamma}^S$  need to be stored. In both cases risk premium estimates are obtained by straightforward matrix operations.

As indicated at the end of Section 4.2, a further step in the hierarchy involves sampling the actual posterior number of claims and claim sizes. In MCMC step  $t$ , the sampled total claim for a policy  $i$  with one or more claims is  $P_i^{(t)} = N_i^{(t)} S_i^{(t)}$ , otherwise it is  $P_i^{(t)} = 0$ . The sampling induces an extra level of variability, even if MPM estimates of the type  $\sum_t P_i^{(t)} / T$  will be approximately as  $\hat{\rho}_i^1$ .

## 6 Validation on Test Data

Models may be compared and validated on an independent test dataset in terms of their predictive power. In our setting there are three components of interest, the claim frequency, the claim size, and the sum of claims for each policy over the exposure period. Assume that the validation dataset contains  $M$  policies with the same covariates as the dataset used for estimation and information on the number of occurred claims and claim sizes. For policy  $i$  in the validation dataset let  $\theta_i$  be the quantity of interest, that is, either the claim frequency  $N_i$ , the claim size  $S_i$  or the total payout  $P_i$ . For  $\theta_i = N_i$  possible estimates are  $\hat{\mu}_i^1$  and  $\hat{\mu}_i^2$ . Similarly,  $\hat{\xi}_i^1$  and  $\hat{\xi}_i^2$  estimate  $\theta_i = S_i$ . For  $\theta_i = P_i$  possible estimates are  $\hat{\rho}_i^1$  and  $\hat{\rho}_i^2$ . See again Table 1. Denote in general the observed and estimated quantities of interest as  $\theta_i$  and  $\hat{\theta}_i$ ,  $i = 1, \dots, M$  respectively. We define the total error  $\text{TE}^\theta = \sum_{i=1}^M (\theta_i - \hat{\theta}_i)$  and decompose it according to the sign of the error as

$$\text{TE}^\theta = \sum_{i=1}^M (\theta_i - \hat{\theta}_i) I(\theta_i \geq \hat{\theta}_i) + \sum_{i=1}^M (\theta_i - \hat{\theta}_i) I(\theta_i < \hat{\theta}_i) = \text{TE}_u^\theta + \text{TE}_o^\theta. \quad (9)$$

The subscripts  $u$  and  $o$  are used to distinguish the contributions from under- and over-estimation respectively. Because claim size is modeled conditional on the number of claims being at least one, the sum should be over policies with observed claims only when  $\theta_i = S_i$ . We are interested in the total error in claim size for the occurred claims and as the majority of the policies have no claims, the contribution from these would dominate the total error if included in the sum. To estimate the uncertainty of the total error we find

$$\text{TE}^{\theta(t)} = \sum_{i=1}^M (\theta_i - \hat{\theta}_i^{(t)}), \quad t = T_0 + 1, \dots, T_0 + T,$$

with  $\hat{\theta}_i^{(t)}$  denoting the  $t$ -th MCMC iterate of the quantity of interest and use as credibility limits the appropriate empirical posterior percentiles.

The predicted total claim is important to determine company reserves. We shall compare models in terms of their total error. Naturally, a small total error is to be preferred. However, two models might have the same total error, but the absolute values of the under- and over-estimation terms in (9) may differ significantly. Of course, the model with smallest contributing errors allows for improved pricing.

## 7 Simulation Experiment

The aim of this simulation experiment is to explore to what degree models with interacting regional latent variables are able to capture the effects of unmeasured covariates. If so, such models should be preferred to GLMs. Our approach is to generate a simulated portfolio of 5 000 policies with different types of covariates and claims. The models specified by (1) and (3) with priors (4) or (5) are fitted, using the simulated covariates with the exception of some selected ones that are substituted by the latent variables. We compare the various models in terms of their ability to predict the number of claims of a second independent set of simulated policies with the same covariates. Only the frequency component of the premium is considered in this section and hence all superscripts  $F$  or  $S$  are omitted.

### 7.1 Description of the Simulated Data

The simulated polices are located in Norway which has 19 administrative regions. The portfolio was distributed randomly to these 19 counties proportionally to the population census. Each policy has a random uniform exposure between 1 and 24 months. The covariates  $\mathbf{x}_i$  of policy  $i$  located in region  $R(i)$  are as follows:

$$\begin{aligned} \text{URBANIZATION:} & \quad x_{R(i),1} = \log(\text{inhabitants per km}^2 \text{ of region } R(i)) \\ \text{CRIME RATE:} & \quad x_{R(i),2} = \text{number of crimes per 1 000 inhabitants in region } R(i) \\ \text{POLICYHOLDER'S INCOME:} & \quad x_{i3} = I\{\text{high}\} \quad x_{i4} = I\{\text{average}\} \quad x_{i5} = I\{\text{low}\} \\ \text{POLICYHOLDER'S SEX:} & \quad x_{i6} = I\{\text{female}\} \quad x_{i7} = I\{\text{male}\}. \end{aligned}$$

Here  $I(\cdot)$  is the indicator function so that for instance  $x_{i3} = 1$  if the policyholder's income is categorized as "high" and zero otherwise. For each policy located in region  $j = 1, \dots, 19$  the income is sampled from a Gamma distribution with expectation equal to the average individual net income in region  $j$  and standard deviation equal to the difference between the maximum and minimum average income of the 19 regions. This results in a high variability in the sampled incomes. The income is classified as "low" if it is smaller than the 1st empirical quantile of the sampled incomes, "high" if it is larger than the 3rd quantile, and "average" otherwise. The sex of the policyholders is sampled according to the census proportions of women and men in the counties. All data are from Statistics Norway for the years 1997/1998. Because we want our simulation example to mimic smoothly varying regional covariates, the urbanization and crime rate of the capital Oslo (county 19) are adjusted to be the average of the three more rural counties sharing the same border. The linear predictor for each policy is constructed as  $\eta_i = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}_i$  with  $\alpha_0 = -2.5$  as a common intercept for all the policies and  $\boldsymbol{\alpha}' = (0.1, 0.005, -0.7, -0.1, 0, -0.2, 0)$ . The two zero parameter values imply that males with low income are taken as reference level. Finally the number of claims for each policy is generated by sampling  $N_i \sim \text{Poisson}(e_i \exp(\eta_i))$ ,  $i = 1, \dots, 5000$  independently.

The parameter values are chosen to give reasonable differences between the policyholders and covariates groups with the highest and lowest claim frequency, as well as a reasonable percentage of claims in the portfolio. The data are summarized in Table 2. Figure 1 shows the spatial distribution of the urbanization (before taking the logarithm) and the crime rate, as well as the location of the counties.

characteristic	value						
number of policies	5000						
number of policies with claims	601 (12.02%)						
distribution of claims	claims	0	1	2	3	4	5
	count	4399	550	43	6	1	1
	%	87.98	11.0	0.86	0.12	0.02	0.02
male vs. female	1.22						
low income vs. high income	2.01						
low income vs. average income	1.10						
average income vs. high income	1.82						
region with highest vs. lowest frequency	1.83						
policy with highest vs. lowest frequency	4.51						

Table 2: *Summary of the simulated portfolio. The lower part of the table shows ratios between the expected number of claims, according to the model used to create the data, for policies that have all covariates equal except for the indicated ones.*

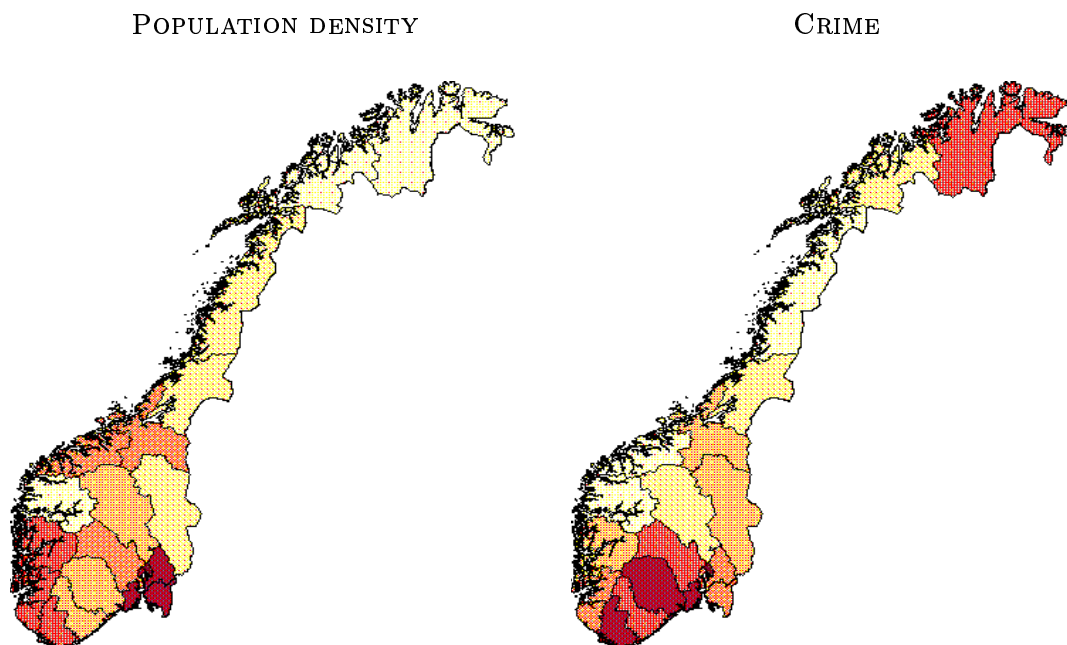


Figure 1: *The regional information used to create the urbanization and crime rate covariates. The left panel shows the population density of the counties of Norway and the right panel shows the number of reported crimes per 1000 inhabitants. Dark color indicates high density/number and the solid lines show the county borders.*

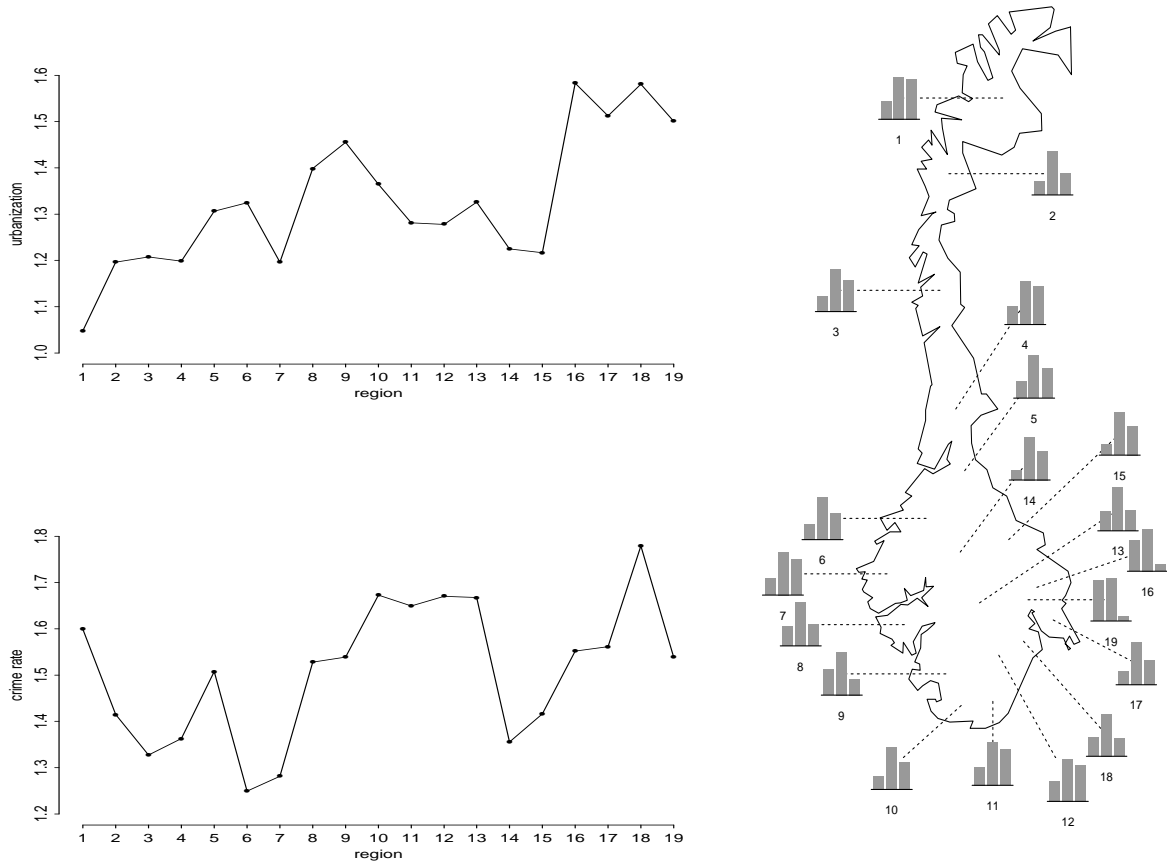


Figure 2: The left panel shows the risk factors  $\exp(0.1x_{R(i),1})$  and  $\exp(0.005x_{R(i),2})$  associated with the urbanization (upper panel) and crime rate (lower panel) covariates for each region. The right panel shows the spatial distribution of the income covariate. The bars in the histograms are (left to right) the number of individuals with income characterized as high, average or low. The number under each histogram is the region number.

Figure 2 shows risk factors due to the urbanization and crime rate covariates,  $\exp(0.1x_{R(i),1})$  and  $\exp(0.005x_{R(i),2})$  respectively, and the spatial distribution of the income covariate. The region numbers are also shown in the figure. Observe that the categorical income covariate seems to be distributed quite homogeneously over the country. In fact, the common intercept  $\alpha_0$  and the terms in the linear predictor  $\eta_i$  involving the income variables, can be substituted by a regional specific parameter. This parameter then measures the average effect of the policyholders' income in a region. A calculation shows these regional specific parameters to be in the range  $-2.59$  to  $-2.53$  with a standard deviation of  $0.04$  when disregarding region 19 (with value  $-2.72$ ).

## 7.2 Description of the Study

We perform two separate experiments. First, the regional covariates urbanization and crime rate are not used when fitting the models so that  $\mathbf{x}_i = (x_{i3}, \dots, x_{i7})$ . Instead, regional latent variables  $\gamma_j$ ,  $j = 1, \dots, 19$  are introduced for each region as in (3). In the second experiment,

the individual income information is discarded instead so that  $\mathbf{x}_i = (x_{R(i),1}, x_{R(i),2}, x_{i6}, x_{i7})$  when fitting the models. Again, seeking to replace this missing information, regional latent variables are introduced. We fit latent variables with priors (4) and (5) for a range of smoothing parameters  $\kappa$  including the limit cases  $\kappa = 0$  and  $\kappa = \infty$ , which are GLMs. Regions are neighbours if they share the same border. The study shows that models with interacting regional latent variables (5) give better results in terms of predictive power only when the missing covariates exhibit a relatively smooth spatial pattern. This shows that the prior distribution for the latent variables should represent the correlation structure among the unmeasured covariates. Of course, it is difficult to obtain such knowledge, but a good guess can help model building.

The predictive abilities of the fitted models applied to validation data are compared in terms of the total error as described in Section 6. We create a validation dataset by generating a new independent portfolio in the same way as was done for the estimation dataset, with all covariates  $\mathbf{x}_i = (x_{R(i),1}, \dots, x_{i7})$ . In the notation of Section 6 we take as quantity of interest  $\theta_i = \mu_i$ , the expected number of claims for policy  $i$ . Since the purpose of this experiment is model comparison, we use  $\mu_i$  instead of  $N_i$  to avoid measuring the effects of the Poisson variation associated with the number of claims which might blur the model comparison. Then  $\hat{\theta}_i$  can be chosen as  $\hat{\mu}_i^1$  or  $\hat{\mu}_i^2$ , see Table 1. The results are similar. Here we report  $\hat{\mu}_i^2$ , showing that even in the more standard setting of using plug-in estimates, interacting latent variables is useful. In order to quantify the uncertainty in the estimated errors due to variability in the simulated data, we repeat the procedure on 1000 independently generated test datasets, always using the same fitted model to predict the claims. Over the 1000 repetitions we obtain the average total error and confidence intervals for the total error.

As starting values for the parameters in the MCMC algorithm, we use the corresponding true values  $\alpha$  and for the latent variables we use the common intercept  $\alpha_0$ . Also, in the GLMM (4)  $\alpha_0$  is used as starting value for  $\mu$  and 10 is used for  $\sigma^2$ . The hyperparameters  $a, b, c$  and  $d$  are fixed to  $\alpha_0, 10, 0.001$  and  $0.001$  respectively. For both simulation experiments, all estimates are based on 100 000 MCMC iterations, following a burn-in of 50 000 iterations. For all our models this is more than sufficient since results indicate that convergence is obtained after 5 000 – 10 000 iterations, depending on the model.

### 7.3 Results

Table 3 shows the average total errors over the 1000 repetitions for the two simulation experiments. In the first experiment, where regional latent variables are used to replace discarded urbanization and crime rate covariates, models with a high (but not infinite) smoothing have the best predictive power. Values of  $\kappa$  in the range 30 to 60 reduces the contributions to the total error from under- and over-estimation. Confidence intervals (not displayed in the table) for  $\text{TE}_u^\mu$  at level 95% for models with  $\kappa = \{30, 40, 50\}$  are (35,37), which is significantly better than all other models, except from models with  $\kappa = \{20, 60\}$ , where significance is not reached at that level. Looking at confidence intervals for  $\text{TE}_o^\mu$ , the best smoothing parameter values are  $\kappa = \{20, 30\}$ , with confidence intervals (23,26). At 95% level these are not significantly different from the models with  $\kappa = \{6, 8, 10, 40, 50, 60\}$ , but significantly better than the rest of the models. Comparing the model with  $\kappa = 30$  to the GLMs (i.e.  $\kappa = \{0, \infty\}$ ), the error in under- and over-estimation, is reduced by more than 50%. This shows that interacting latent variables can be very useful. In this comparison the results from fitting the true model using all covariates, that is the one used to generate data, are used as reference. For the total error  $\text{TE}^\mu$  the differences between models are not significant. Hence, for reserve calculations

	true model	smoothing parameter $\kappa$												GLMM
		0	2	4	6	8	10	20	30	40	50	60	$\infty$	
experiment 1														
$TE_u^\mu$	21	54	48	45	42	41	39	37	36	36	36	36	52	46
$TE_o^\mu$	15	38	33	30	28	27	26	24	24	25	26	26	38	32
$TE^\mu$	6	16	15	14	14	13	13	12	12	11	10	10	14	14
experiment 2														
$TE_u^\mu$	21	93	89	87	86	86	85	83	82	82	82	82	73	90
$TE_o^\mu$	15	76	74	73	72	71	71	69	68	67	67	66	67	73
$TE^\mu$	6	16	15	15	14	14	14	14	14	14	15	16	6	17

Table 3: *Total errors for different fitted models with regional latent variables to replace discarded urbanization and crime rate covariates (experiment 1) and individual income (experiment 2). The first column shows the results obtained by fitting the model from which the data are generated, including all the covariates. Columns 3-14 show the result for varying values of the smoothing parameter  $\kappa$  in (5). The right column shows the results for the GLMM (4). All the total errors are averages over prediction on 1 000 independently generated validation datasets. One would expect a total error of zero for the true model. The resulting value of 6 is due to the average number of claims for the 1 000 validation datasets being 669, whereas the number of claims in the estimation dataset equals 663.*

introducing interacting latent variables might be less important.

In the second experiment, conducted as the first but with latent variables to replace discarded individual income information, the model with  $\kappa = \infty$  is significantly better than all the other models. This is not surprising considering the argument of Section 7.1 where we show that the income covariate is distributed homogeneously over the country. Hence, as we exclude a covariate with no differentiated spatial trend, introducing regional interacting latent variables does not improve predictions with respect to a single common latent intercept. Notice that the GLMM does not compare favourably with the other models.

## 8 Norwegian Theft from Car Data

### 8.1 Description of the Data

We study a dataset of 19 840 policies on theft from cars. The dataset originates from a random sample of 20 000 policies from the 1998 database of the Norwegian insurance company Gjensidige Forsikring. An initial removal of records with missing or invalid data produces the final estimation dataset. An independent validation dataset is created and screened in the same way, leaving 19 844 records.

The two datasets are described in Table 8.1. Theft from cars appear to be a quite low frequent claim type. In the estimation dataset only 0.7% of the policies have claims. The variability in the claim sizes is quite large with the empirical mean and standard deviation being of the same magnitude. Also, as seen in Figure 3, the distribution of the claim size is skewed due to occasional large claims. All claim sizes are given in Norwegian kroner (1 NOK  $\approx$  0.124 USD).



variable	description	estimation data		validation data	
$\sum N_i$	total number of claims	139		159	
$\sum e_i$	total exposure years	14383		14106	
$\sum S_i N_i$	total claim size (1000 NOK)	1145		1407	
		mean	sd	mean	sd
$N_i$	number of claims per policy	0.007	0.086	0.008	0.093
$S_i$	mean claim size (NOK) per claim	8236	8036	8850	7657
EXPO	exposure time in years	0.725	0.338	0.711	0.316
AGE	age of policyholder	45.2	15.3	44.4	14.9
CAR AGE	age of car	10.1	6.0	9.8	5.9
MILEAGE	insured kilometers (in 1000 km)	13.5	6.4	14.0	6.8
SEX	% female	29.4		29.8	
		% records	% claims	% records	% claims
COUNTY	1. Finnmark	0.52	0.0	0.62	0.0
	2. Troms	2.26	0.7	2.15	1.3
	3. Nordland	3.77	1.4	3.75	0.6
	4. Nord-Trøndelag	4.76	5.0	4.88	4.4
	5. Sør-Trøndelag	7.50	7.9	7.21	7.5
	6. Møre og Romsdal	5.40	0.7	4.90	1.3
	7. Sogn og Fjordane	2.48	1.4	2.42	1.3
	8. Hordaland	9.31	7.9	9.45	6.9
	9. Rogaland	7.18	6.5	7.48	6.9
	10. Vest Agder	1.95	0.7	2.12	3.1
	11. Aust Agder	1.99	1.4	2.19	2.5
	12. Telemark	3.33	0.7	3.63	3.8
	13. Buskerud	5.89	5.0	5.75	6.9
	14. Oppland	7.11	2.9	6.95	6.9
	15. Hedemark	7.03	6.5	6.90	1.9
	16. Akershus	11.05	14.4	11.44	15.7
	17. Østfold	6.21	8.6	6.15	6.9
	18. Vestfold	4.15	8.6	4.02	4.4
	19. Oslo	8.11	19.4	7.98	17.6

Table 4: *Summary of covariates and response variables for the estimation and validation dataset. The first three lines show the total number of claims, exposure time and claim size in the dataset, followed by the mean and standard deviation of the number of claims, exposure time and claim size. For the covariate SEX the percentage of women in the dataset is given. The table shows the percentage of policies located in each county and the percentage of claims occurred therein.*

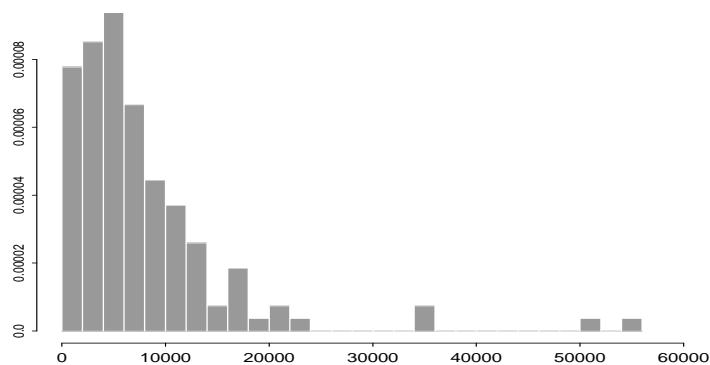


Figure 3: *Histogram of claim size for the estimation dataset. Only policies with claims are included.*

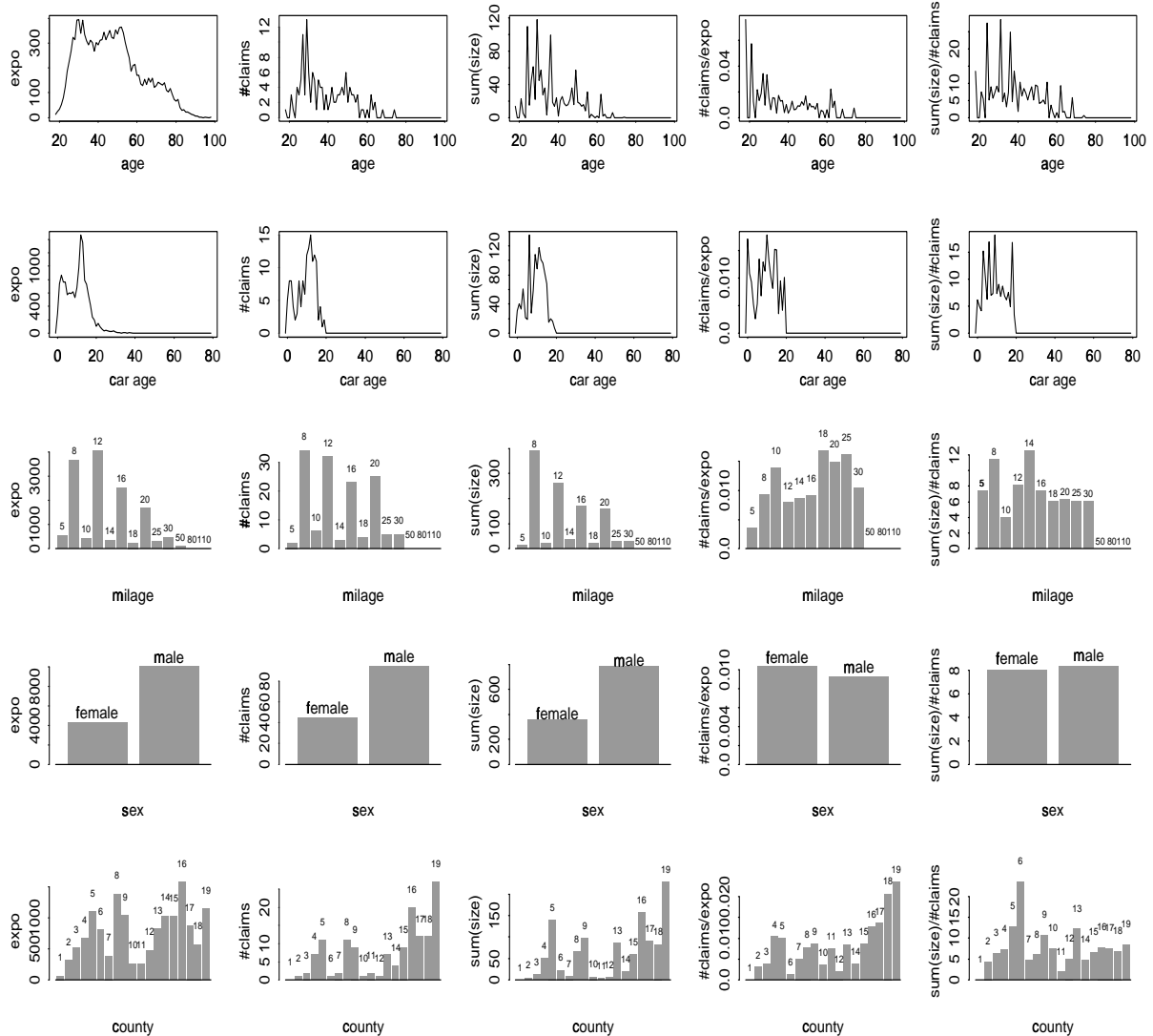


Figure 4: *The plots show (left to right) how the total exposure time, number of claims, total claim size (in 1000 NOK), claims per year and average claim size (in 1000 NOK) vary as functions the covariates AGE, CAR AGE, MILEAGE, SEX and COUNTY for the estimation dataset.*

The available covariates are age of policyholder (AGE), age of the car with -1 indicating a 1999 car model (CAR AGE), maximum allowed mileage measured in 1 000 kilometers (MILEAGE), sex of the policyholder (SEX) and region where the policy is located (COUNTY). Figure 4 shows how several important quantities (total exposure in years, number of claims, sum of claim sizes, number of claims per insured year and average claim size) vary with these different covariates in the estimation dataset. We let AGE, CAR AGE and MILEAGE enter the models as continuous covariates. The covariate SEX is coded as a zero-one variable that takes value 1 for men. Hence, women are used as the reference level in the parameter estimation.

The geographical regions of interest are again the 19 Norwegian administrative counties. Table 8.1 shows that the portfolio is unevenly distributed over the country, with some counties being well represented in the dataset and others not. Also, the distribution of the number of

claim frequency		claim size	
model	AIC	model	AIC
1	1384	1	109
1+AGE	1342	1+AGE	104
COUNTY	1368	COUNTY	124
AGE+COUNTY	1325	AGE+ CAR AGE <sup>2</sup> +MILEAGE	102
AGE+CAR AGE+ MILEAGE+ SEX+ COUNTY	1330	AGE+CAR AGE <sup>2</sup> + MILEAGE+ SEX	103

Table 5: *AIC values for some frequency and size models.*

claims in each county is clearly not proportional to the number of policies therein. There are counties with very few claims, for which estimation of independent regional effects is difficult.

## 8.2 Precise Form of the Linear Predictor

As a first step in our analysis we use Generalized Additive Models with a smoothing spline to determine advantageous transformations of the continuous covariates, see Hastie & Tibshirani (1990). For the claim frequency model this analysis supports the use of linear terms for all these covariates, while for claim size a quadratic term should be used for CAR AGE.

We next consider the explanatory power of each covariate. When choosing among several, possibly nested GLMs, classical methods for model selection like the Akaike information criterion (AIC) may be applied. For models with latent variables that originate from the same prior distribution as in (4) or interact as in (5), this criterion is inapplicable because model dimension is not well defined. As a first step we therefore use the AIC criterion on the GLMs ( $\kappa \in \{0, \infty\}$ ) only. It is reasonable to believe that the effect of the covariates is similar for the GLMM (4) and for models with interacting latent variables (5). For both claim frequency and claim size we fit several models, starting with only an intercept term and adding covariates to see which models give the best fit in terms of smallest AIC. We use the statistical software S-Plus, see Venables & Ripley (1994, p. 177) in this analysis.

In Table 5 some of the fitted models with their AIC values are reported. For the frequency model AGE and COUNTY give the smallest AIC. For claim size, a quadratic term for CAR AGE gives a better fit than a linear term. When including COUNTY the AIC increases considerably for claim size, which is not surprising since there are regions with very few claims (see Table 8.1). Based on our findings we decide to include all the covariates for both the frequency and size model as this is what would be done if a larger dataset was available and is close to optimal.

## 8.3 Inference

We fit the Poisson and Gamma models (1) with linear predictors (3). For the latent variables associated with the 19 regions we either use the normal prior (4) or the MRF (5) with  $\kappa \in \{0, 0.5, 1, 2, 3, 4, 8, 16, 20, \infty\}$ . With  $\kappa = 0$  or  $\kappa = \infty$ , parameters can be estimated without the need for MCMC. The parameter estimates obtained by fitting the claim frequency and size models with one common intercept ( $\kappa = \infty$ ) using Fisher scoring are used as starting values for the Markov chains. The estimated common intercepts are taken as starting values for  $\mu_F$  and  $\mu_S$ , in the GLMM, while 10 is taken as starting value for  $\sigma_F$  and  $\sigma_S$ . Also,  $a_F$  and  $a_S$  are fixed as the estimated common intercepts, while  $b_F = b_S$ ,  $c_F = c_S$  and  $d_F = d_S$  are set to 10, 0.001 and 0.001 respectively to make the priors quite flat. A natural starting value for  $\nu$  is 1 since this corresponds to a model with standard deviation equal to the expectation

for single claims. MCMC convergence time is monitored by visual inspection of the plots of the trajectories of the MPM estimates (8) for all parameters and latent variables and by estimating the integrated autocorrelation time, see Geyer (1992). With  $\kappa = 0$  or  $\kappa = \infty$ , the number of iterations needed to reach equilibrium in the Markov chain can be determined by comparing the Newton-Raphson estimates with the estimated MPM from MCMC. These results can be used as guidelines also when fitting models for other values of  $\kappa$ . Convergence is achieved quickly for low values of  $\kappa$  and convergence time increases in  $\kappa$ . We feel confident that a burn-in of 20 000 iterations for  $\kappa \in \{0, 0.5, 1, 2, 3, 4, \infty\}$  and the GLMM gives reliable estimates. We reparameterize the model with  $\kappa = \infty$  to have only one common intercept (or latent variable) for all the regions, which explains the short convergence time. A burn-in of 100 000 iterations is needed for  $\kappa \in \{8, 16, 20\}$ . For all the models the MPM estimate is based on the 10 000 iterations following the burn-in.

## 8.4 Results

The resulting MPM estimates for a few chosen models are given in Table 6 for claim frequency and in Table 7 for claim size. As there are no claims in county number 1, an independent regional claim size effect as needed when  $\kappa = 0$  cannot be estimated. Also, there are 7 other counties (numbered as 2, 3, 6, 7, 10, 11 and 12) with only one or two claims (see Table 8.1) for which estimation of independent regional latent variables are associated with a larger uncertainty. However, as  $\kappa$  increases these regions share the information of the neighbouring regions and the uncertainty is clearly reduced. Figures 5 and 6 show estimates of the marginal posterior density of the parameters for the frequency model with  $\kappa = 8$  and  $\kappa = \infty$ . The posterior densities of the parameters approach normal distributions as  $\kappa$  increases. For  $\kappa$  zero or small (no figure), the latent variables in the claim frequency model have a slightly skewed density with a tail towards smaller (and negative) values for regions with few claims. For claim size (no figure), the marginal posterior density of the latent variables are skewed towards larger values for regions with large claims. When the normal approximation to the posterior marginal distribution is satisfactory, the posterior standard deviation gives an idea of the dispersion and is given in Tables 6 and 7. For the GLMMs the posterior normal distributions in (4) are found to be  $N(-3.0, 0.78)$  and  $N(10.17, 0.54)$  for the claim frequency and size respectively. The possibility to plot posterior marginal densities is particularly important and useful for the quantities of interest related to the risk premium. Figure 7 shows the posterior densities of  $\mu_i$ ,  $\xi_i$  and of the risk premium  $\rho_i = \mu_i \xi_i$  for four different policies. These policies are selected to show the wide range in the obtained posterior densities. In Figure 7 we use the model with prior (5) with  $\kappa = 8$  for claim frequency and  $\kappa = \infty$  for claim size. The covariates (AGE, CAR AGE, MILEAGE, SEX, COUNTY) are for the first of these policies (43, 0, 8, male, 7) and the exposure time is 0.51 years. We notice the peaked marginal posterior density for the corresponding risk premium. The second policy has 1 year of exposure and covariates (27, 8, 16, female, 8) and has a more dispersed posterior density, with a longer right tail. Both these policies have no claims in the exposed period. The two last policies have one claim each of 4 200 NOK and 16 455 NOK and exposure times of 0.73 and 0.87 years respectively. Their covariates are (29, 13, 8, female, 8) and (24, 5, 14, male, 19). The posterior density of the risk premium of the third policy is centered at approximately 100 NOK and is relatively peaked, while the posterior density of the last policy is centered at approximately 370 and is very dispersed. The point estimates  $\hat{\rho}_i^1$  for the four policies are 56, 161, 103 and 369 respectively, with 95% credibility intervals (30, 95), (92, 269), (57, 174) and (208, 600) respectively.

name	$\kappa = 0$		$\kappa = 8$		$\kappa = \infty$		GLMM	
AGE	-0.04	0.01	-0.04	0.01	-0.04	0.01	-0.04	0.01
CAR AGE	-0.01	0.02	-0.02	0.02	-0.02	0.02	-0.01	0.02
MILEAGE	5e-03	0.01	1e-03	0.02	2e-03	0.01	4e-03	0.01
SEX(male)	-0.07	0.18	-0.04	0.19	-0.08	0.18	-0.01	0.17
county 1	-16.83	9.23	-3.36	0.65	-2.71	0.39	-3.34	0.85
2	-4.54	1.30	-3.32	0.57	-2.71	0.39	-3.51	0.75
3	-3.97	0.88	-3.17	0.51	-2.71	0.39	-3.50	0.68
4	-2.72	0.60	-2.86	0.49	-2.71	0.39	-2.78	0.58
5	-2.76	0.51	-2.78	0.46	-2.71	0.39	-2.80	0.53
6	-5.27	1.27	-2.98	0.47	-2.71	0.39	-3.95	0.73
7	-3.67	0.90	-2.83	0.46	-2.71	0.39	-3.34	0.68
8	-3.03	0.50	-2.80	0.45	-2.71	0.39	-3.02	0.51
9	-2.95	0.52	-2.83	0.46	-2.71	0.39	-2.97	0.53
10	-4.14	1.31	-2.91	0.51	-2.71	0.39	-3.37	0.74
11	-3.25	0.86	-2.85	0.48	-2.71	0.39	-3.12	0.69
12	-4.80	1.28	-2.77	0.46	-2.71	0.39	-3.63	0.73
13	-2.97	0.56	-2.63	0.44	-2.71	0.39	-2.98	0.55
14	-3.76	0.66	-2.74	0.44	-2.71	0.39	-3.51	0.63
15	-2.89	0.54	-2.69	0.47	-2.71	0.39	-2.90	0.54
16	-2.50	0.46	-2.49	0.44	-2.71	0.39	-2.55	0.49
17	-2.39	0.50	-2.37	0.48	-2.71	0.39	-2.49	0.53
18	-1.94	0.50	-2.38	0.48	-2.71	0.39	-2.11	0.55
19	-1.87	0.42	-2.22	0.44	-2.71	0.39	-1.96	0.46

Table 6: *Estimates of the marginal posterior mean (left) and standard deviation (right) for the spatial smoothing prior (5) with  $\kappa \in \{0, 8, \infty\}$  and the GLMM (4) for claim frequency.*

name	$\kappa = 0$		$\kappa = 8$		$\kappa = \infty$		GLMM	
AGE	-0.02	0.01	-0.02	0.01	-0.01	0.01	-0.02	0.01
CAR AGE <sup>2</sup>	-1e-03	1e-03	-2e-03	1e-03	-2e-03	1e-03	-2e-03	1e-03
MILEAGE	-0.02	0.02	-0.03	0.02	-0.03	0.01	-0.04	0.01
SEX(male)	0.03	0.16	0.07	0.17	0.09	0.14	0.15	0.02
county 1	NA	NA	10.09	0.71	10.14	0.33	10.18	0.67
2	9.59	1.00	10.09	0.62	10.14	0.33	10.00	0.59
3	9.97	0.72	10.17	0.54	10.14	0.33	10.09	0.55
4	10.14	0.55	10.30	0.49	10.14	0.33	10.27	0.46
5	10.26	0.39	10.34	0.43	10.14	0.33	10.28	0.36
6	11.26	1.09	10.35	0.47	10.14	0.33	10.50	0.53
7	9.68	0.78	10.24	0.46	10.14	0.33	9.99	0.55
8	9.74	0.44	10.15	0.43	10.14	0.33	9.94	0.41
9	10.25	0.44	10.24	0.43	10.14	0.33	10.37	0.40
10	10.26	1.10	10.17	0.50	10.14	0.33	10.13	0.60
11	8.76	0.76	10.11	0.48	10.14	0.33	9.73	0.59
12	9.78	1.01	10.19	0.45	10.14	0.33	10.05	0.58
13	10.36	0.47	10.29	0.43	10.14	0.33	10.37	0.41
14	9.81	0.63	10.29	0.45	10.14	0.33	10.08	0.51
15	10.10	0.55	10.32	0.47	10.14	0.33	10.27	0.47
16	10.02	0.42	10.30	0.43	10.14	0.33	10.20	0.38
17	10.20	0.52	10.40	0.51	10.14	0.33	10.35	0.45
18	9.94	0.48	10.20	0.47	10.14	0.33	10.13	0.44
19	10.20	0.44	10.38	0.45	10.14	0.33	10.36	0.40

Table 7: *Estimates of the marginal posterior mean (left) and standard deviation (right) for the spatial smoothing prior (5) with  $\kappa \in \{0, 8, \infty\}$  and the GLMM (4) for claim size. In region 1 there are no claims and an independent regional effect cannot be estimated.*

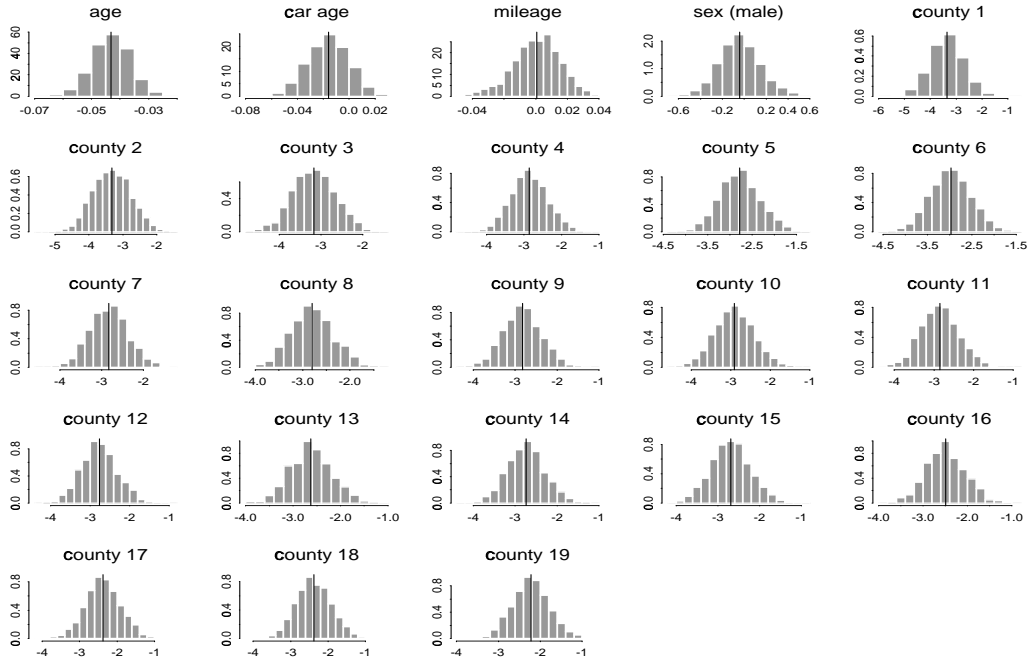


Figure 5: *Estimated posterior density of the parameters of the claim frequency model for  $\kappa = 8$ . The vertical line shows the MPM estimate.*

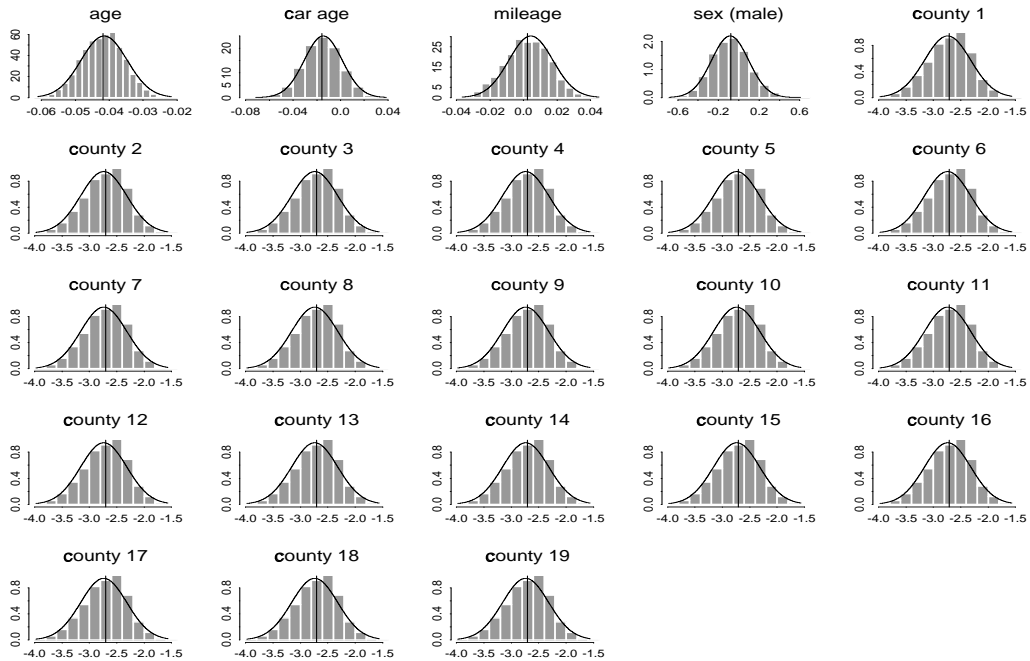


Figure 6: *Estimated posterior density of the parameters of the claim frequency model or  $\kappa = \infty$ . The solid curve shows the asymptotic normal distribution according to maximum likelihood theory. The vertical line shows the MPM estimate.*

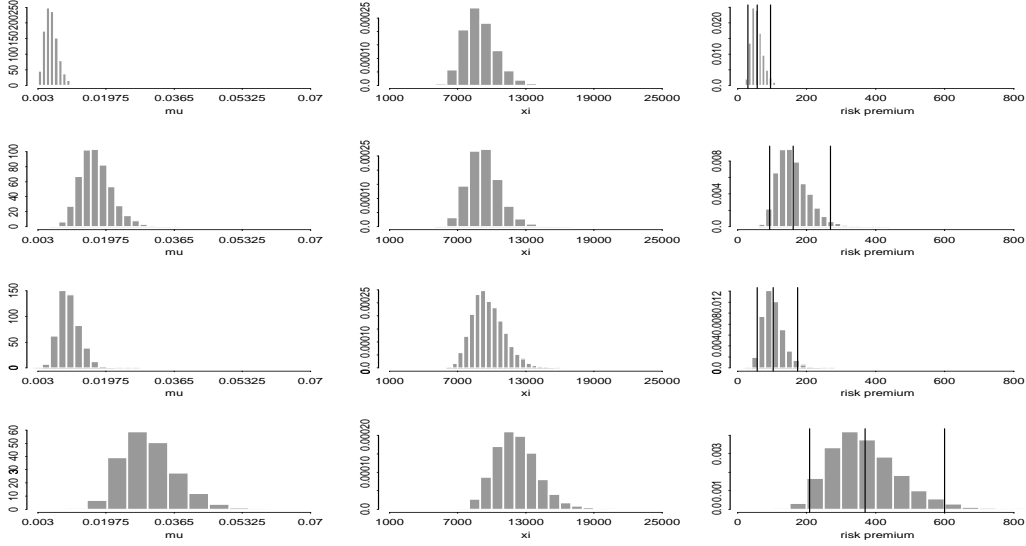


Figure 7: The posterior density of  $\mu_i$  (left),  $\xi_i$  (center) and risk premium  $\rho_i = \mu_i \xi_i$  (right) for 4 selected policies from the validation dataset. The fitted models have prior (5) for the latent variables with  $\kappa = 8$  and  $\kappa = \infty$  for claim frequency and size respectively. The covariates AGE, CAR AGE, MILEAGE, SEX and COUNTY are for the policies (top to bottom) (43, 0, 8, male, 7), (27, 8, 16, female, 8), (29, 13, 8, female, 8) and (24, 5, 14, male, 19). The first two policies have no claims, while the last two have one claim each of 4200 NOK and 16 455 NOK respectively. Exposure times are 0.51, 1, 0.73 and 0.87 years. In the histogram for the risk premium, the two outer vertical lines indicate 95% credibility intervals and the center line indicates the estimate  $\hat{\rho}_i^1$  of Table 1.

Table 8 gives posterior rating factors (see Section 2.2) using the prior (5) with  $\kappa = 0$  for claim frequency and  $\kappa = \infty$  for claim size. For the continuous covariates, factors are given for some chosen values. As an example of premium calculations using these tables, a policy located in region 19 with a male policyholder aged 18 and a 1998 car model (0 years) with annual mileage of 12 000 km has an estimated premium  $\hat{\rho}_i^2 = (0.155 \cdot 0.462 \cdot 0.934 \cdot 1.064) \times (25369 \cdot 0.772 \cdot 1.094 \cdot 0.681) = 1038$  NOK. For comparison, a policy with the same characteristics in county 7 is priced to 174 NOK.

## 8.5 Model Validation and Comparison

Assessing the fit of models with interacting latent variables is difficult. Bayesian model validation is an active research area and we refer to Spiegelhalter et al. (1998) for an interesting idea that could be applied in our context. It goes beyond the scope of this paper to investigate this aspect thoroughly, though we shall perform some standard testing for the standard GLMs in the maximum likelihood setting.

The Pearson statistic (McCullagh & Nelder 1989, pp. 34, 37) is a common goodness of fit measure for GLMs. Our models specified by (1), (3) and (5) with  $\kappa = 0$  or  $\kappa = \infty$  can be evaluated this way in a pure likelihood setting. The Pearson test applied to the estimation data, indicate that the Poisson and Gamma assumptions for these two models are adequate. As the claim frequency is very small we need to group the data in order to check the Poisson assumption. When the Pearson test is applied to the validation dataset, we need to ignore a few large departures from the expected values to obtain a satisfactory fit.

	COUNTY		AGE		CAR AGE		MILEAGE		SEX					
	F	S	F	S	F	S	F	S	F	S				
1	5e-08	25369	18	0.462	0.772	-1	1.013	0.998	8	1.042	0.774	female	1	1
2	0.011	--	20	0.424	0.751	0	1	1	12	1.064	0.681	male	0.934	1.094
3	0.019	--	25	0.342	0.699	1	0.987	0.998	16	1.086	0.599			
4	0.066	--	30	0.276	0.650	2	0.975	0.993	20	1.108	0.527			
5	0.063	--	35	0.223	0.605	3	0.963	0.984	50	1.293	0.201			
6	0.005	--	40	0.180	0.563	4	0.951	0.972	80	1.509	0.077			
7	0.026	--	45	0.145	0.524	5	0.939	0.956						
8	0.048	--	50	0.117	0.488	10	0.882	0.836						
9	0.052	--	55	0.094	0.454	15	0.828	0.669						
10	0.016	--	60	0.076	0.423	20	0.777	0.489						
11	0.039	--	65	0.061	0.394									
12	0.008	--	70	0.050	0.366									
13	0.051	--	75	0.040	0.341									
14	0.023	--	80	0.032	0.317									
15	0.056	--												
16	0.083	--												
17	0.092	--												
18	0.144	--												
19	0.155	--												

Table 8: Rating factors for the claim frequency and size model with prior (5) for the latent variables with  $\kappa = 0$  and  $\kappa = \infty$  respectively.

Test results indicate that there could be some overdispersion in the frequency data, contradicting our Poisson assumption. In the Gamma model (1), the parameter  $\nu$  is taken equal for all policies. Also this assumption is tested and found acceptable.

The residuals of the Poisson model show no outliers or unexpected patterns. For the Gamma model, residuals are markedly large and positive for claim size exceeding approximately 20 000 NOK which is the 95th percentile. This suggests the presence of a heavier tail.

The main issue is to compare the various models in terms of their capacity to predict the number of claims and claim sizes. For this purpose we use the validation dataset. For each policy  $i$  in the dataset we compute the estimates  $\hat{\mu}_i^1$ ,  $\hat{\xi}_i^1$  and  $\hat{\rho}_i^1$  as well as  $\hat{\mu}_i^2$ ,  $\hat{\xi}_i^2$  and  $\hat{\rho}_i^2$  (see Table 1) using its covariates. These estimates can then be compared to the actual number of claims, claim sizes and total payouts which for most policies is zero. The total errors are computed as described in Section 6, but here we also find the average errors. Hence, for claim frequency and total payouts the errors  $\text{TE}^N$  and  $\text{TE}^P$  are divided by the total exposure time  $\sum_i e_i$  in the dataset. For claim size, interest lies with the average error per claim, and  $\text{TE}^S$  is divided by  $\sum_i N_i$ . The fitted models are then compared.

Table 9 shows the best and the worst model in terms of absolute average error using  $\hat{\mu}_i^2$ ,  $\hat{\xi}_i^2$  and  $\hat{\rho}_i^2$  to predict  $N_i$ ,  $S_i$  and  $P_i$  respectively. Compared to  $\hat{\mu}_i^1$ ,  $\hat{\xi}_i^1$  and  $\hat{\rho}_i^1$  these exhibit the largest variability in total error between models. For claim frequency and claim size the models with the smallest average error on the validation dataset is found have prior (5) with  $\kappa = \infty$  and  $\kappa = 16$  respectively. The largest errors are achieved for  $\kappa = 0$  and  $\kappa = 1$ . The best model in terms of average error for the risk premium has  $\kappa = \infty$  in both components, although we did not try all possible smoothing combinations of the two models. The important point is that none of the above differences is significant. In fact, if we approximate the posterior density of the total error  $\text{TE}^N$  by the empirical density of  $\text{TE}^{N(T_0+1)}, \dots, \text{TE}^{N(T_0+T)}$  and compute 95% credibility intervals for each model, these overlap significantly. For example,



		frequency	size	risk premium	
Predict the validation data	best	$\kappa = \infty$ 0.00013 (-0.0007, 0.0027)	$\kappa = 16$ 1064 (-414, 1925)	$\kappa = (\infty, \infty)$ 16.9 (-6.6, 32.1)	model type average error credibility
	worst	$\kappa = 0$ 0.0018 (-0.0008, 0.0027)	$\kappa = 1$ 1185 (-744, 1926)	$\kappa = (0, \infty)$ 21.2 (-7.8, 31.5)	
Predict the estimation data	best	$\kappa = \infty$ 0.0002 (-0.0017, 0.0015)	$\kappa = \infty$ -405 (-1740, 551)	$\kappa = (\infty, \infty)$ 1.2 (-21.0, 15.5)	
	worst	$\kappa = 0$ 0.0007 (-0.0017, 0.0015)	$\kappa = 0$ -752 (-4584, 27.9)	$\kappa = (0, \infty)$ 5.7 (-21.4, 15.3)	

Table 9: Average errors with  $\hat{\mu}_i^2$ ,  $\hat{\xi}_i^2$  and  $\hat{\rho}_i^2$  used to predict the claim frequency, claim size and total payouts for the validation and estimation dataset. The table shows the best and worst model in terms of smallest and largest absolute average error. For each model a 95% credibility interval for the average error is given. For the risk premium the values of  $\kappa$  are for the frequency and size model respectively.

as seen from Table 9, the 95% credibility intervals of the average error for the best and worst model are essentially identical. Similar results are found for claim size and risk premium. We are not able to find any model with predictive properties significantly better than the other ones. The difference between models is minor compared to the uncertainty associated with the predictions for each model. As the smoothing increases, the uncertainty gets smaller since the number of effective parameters is reduced. As the validation dataset features a higher claim frequency and claim size than the estimation dataset (see Table 8.1), all fitted models underestimate claims in the validation dataset. The average error for the risk premium is between 17 and 21 NOK which is a good performance.

For comparison we also repeat the prediction of claims and compute the errors on the original estimation dataset used to fit models. In Table 9 we see that these errors are smaller. The negative average errors for claim size is a consequence of large claims in the tail of the distribution. The average error varies between 1.2 and 5.7 NOK. The small magnitude of these errors illustrates that the models are very well suited for premium estimation.

Considering this dataset it is not surprising that regional latent variables do not significantly improve the ability to predict on the validation dataset compared to standard GLM. First of all, our datasets are small with respect to the low claim frequency. Because claims are rare, a pattern of regional smoothness simply is not present in this data. Hence, latent variables that would stand as proxies for unaccounted smooth regional variation cannot lead to significant improvements. The effect of large claims is similar: the occurrence of a large claim will cause the average claim size of a specific covariate group to be markedly different from that of similar groups. Hence, again smooth latent variables are unnecessary. A larger dataset is needed to verify that interacting latent variables can be helpful in replacing missing covariates, as we believe and as confirmed by our simulation results in Section 7.

## 9 Conclusions

Hierarchical models and Bayesian inference have been applied to the context of risk premium rating and compared to classical GLMs. We have shown that a fully Bayesian analysis is feasible by means of MCMC algorithms. This allows us to account for all sources of uncertainty and errors in the final posterior estimates. These can be given as point estimates but it is important to inspect the posterior marginal density to evaluate dispersion around such estimates and to detect important skewness and heavy tails. Latent variables associated to specific categories of policies or policyholders, have been introduced both in claim frequency and claim size modeling. From our simulation example we conclude that interacting regional latent variables can improve the performance of models in terms of their ability to predict the number of claims, their size and the risk premium. This is particularly true if the portfolio includes unmeasured covariates which display regional patterns, trends or smoothness. Our analysis of a the car theft data shows that the improvement achieved by models with interacting latent variables compared to standard GLMs is not significant if the dataset is too small with respect to its claim frequency.

Hierarchical Bayesian models are very flexible and allow us to incorporate important features like heavy tails, time trends and seasonal effects. There is a wide literature exploring such models in the areas of epidemiology (Glad et al. 2000), disease mapping (Biggeri et al. 1999), image analysis (Rue & Hurn 1999) and environmental monitoring (Wikle et al. 1998). A specific characteristic feature of insurance is that both a frequency and size model are of interest.

## A The Metropolis-Hastings Algorithm

The single-component Metropolis-Hastings algorithm works as follows. Assume that the distribution of interest is  $\pi(\boldsymbol{\psi})$ . Denote by  $\boldsymbol{\psi}^k$  the  $k$ -th iterate of the Markov chain. Assume that at step  $k + 1$  the  $l$ -th component of  $\boldsymbol{\psi}$  is updated. A proposal  $\tilde{\psi}_l$  for the  $l$ -th element of  $\boldsymbol{\psi}$  is sampled from an irreducible proposal density  $q_l(\cdot|\boldsymbol{\psi}^k)$ . Then  $\psi_l^{k+1}$  equals the proposed value with probability

$$p(\tilde{\psi}_l, \boldsymbol{\psi}^k) = \min\left\{1, \frac{\pi(\tilde{\psi}_l, \boldsymbol{\psi}_{-l}^k) q_l(\psi_l^k|\tilde{\psi}_l, \boldsymbol{\psi}_{-l}^k)}{\pi(\boldsymbol{\psi}^k) q_l(\tilde{\psi}_l|\boldsymbol{\psi}^k)}\right\},$$

where  $\boldsymbol{\psi}_{-l}^k$  denotes the current value of all components except the  $l$ -th one. With probability  $1 - p(\tilde{\psi}_l, \boldsymbol{\psi}^k)$  we have  $\psi_l^{k+1} = \psi_l^k$ . In our context the elements of  $\boldsymbol{\psi}$  are updated in the following order. First the latent variables are updated, then the parameters for the fixed effects. For the claim size an additional step for the parameter  $\nu$  of the Gamma distribution is needed. The algorithm for the GLMM has additional steps for updating the parameters  $\mu_r$  and  $\sigma_r$ ,  $r = F, S$  of the normal distribution in (4). During each full update the latent variables are updated in a random permutation order which is drawn independently in each iteration. The same procedure is followed for the other parameters. All proposal densities are uniform distributions over a range centered at the current values. For  $\nu$  in the claim size model and  $\sigma_r$  in the GLMM (4) the densities are truncated at zero to restrict proposals to positive values. The range of the uniform distributions is tuned to give an approximate 50% acceptance probability. Notice that this is a very simple Metropolis-Hastings algorithm. We have made no special effort to achieve fast convergence as should certainly be done in an implementation intended for professional use.

We give the acceptance probabilities for the latent variables and parameters in the claim frequency model. As the Poisson and Gamma distributions belong to the exponential family, the structure of the acceptance probabilities is the same. The acceptance probabilities for the latent variables and the parameters in the claim size model are obtained from the expressions below by replacing  $-e_i$  by  $N_i S_i$ , multiplying the sum over policies by  $-\nu$  and changing the sign of the terms in the linear predictor that go into exponents inside the sum. Hence, only one computer code needs to be implemented. In the following all indexing on iteration number is omitted for notational convenience.

As nearly all computing time is spent calculating the acceptance probabilities, the speed of the algorithm may be increased considerably by precalculating all terms in the acceptance probabilities which involve only the responses and covariates as they need only be calculated once. Also, when updating the latent variables (or parameters), the parameters (or latent variables) are unchanged and terms including these may be precalculated as well.

Denoting by  $n_l$  the number of neighbours of region  $l$ , the acceptance probabilities for the latent variables  $\gamma^F$  in (5) are

$$\begin{aligned} p(\tilde{\gamma}_l^F, \gamma^F) &= \min\left\{1, \frac{\pi(\mathbf{N}|\tilde{\gamma}_l^F, \gamma_{-l}^F, \boldsymbol{\alpha})}{\pi(\mathbf{N}|\gamma^F, \boldsymbol{\alpha})} \frac{\pi(\tilde{\gamma}_l^F, \gamma_{-l}^F)}{\pi(\gamma^F)}\right\} \\ &= \min\left\{1, \exp\left\{\sum_{i:R(i)=l} [(\tilde{\gamma}_l^F - \gamma_l^F)N_i - (e^{\tilde{\gamma}_l^F} - e^{\gamma_l^F})e_i \exp(\boldsymbol{\alpha}'\mathbf{x}_i^F) \right. \right. \\ &\quad \left. \left. - \frac{\kappa}{2}[n_l((\tilde{\gamma}_l^F)^2 - (\gamma_l^F)^2) - 2(\tilde{\gamma}_l^F - \gamma_l^F) \sum_{k \in \delta_l} \gamma_k^F]\right]\right\}. \end{aligned}$$

For the parameters the probabilities are

$$\begin{aligned} p(\tilde{\alpha}_l, \boldsymbol{\alpha}) &= \min\left\{1, \frac{\pi(\mathbf{N}|\gamma^F, \tilde{\alpha}_l, \boldsymbol{\alpha}_{-l})}{\pi(\mathbf{N}|\gamma^F, \boldsymbol{\alpha})} \frac{\pi(\tilde{\alpha}_l, \boldsymbol{\alpha}_{-l})}{\pi(\boldsymbol{\alpha})}\right\} \\ &= \min\left\{1, \exp\left\{\sum_{i=1}^I [-e_i(e^{\tilde{\alpha}_l x_{il}^F} - e^{\alpha_l x_{il}^F}) \exp(\boldsymbol{\alpha}'_{-l}\mathbf{x}_{i,-l}^F + \gamma_{R(i)}^F) + (\tilde{\alpha}_l - \alpha_l)N_i x_{il}^F]\right\}\right\}. \end{aligned}$$

Estimation in the claim size model also involves the index parameter  $\nu$  for which the acceptance probability is

$$\begin{aligned} p(\tilde{\nu}, \nu) &= \min\left\{1, \frac{\pi(\mathbf{S}|\mathbf{N}, \gamma^S, \boldsymbol{\beta}, \tilde{\nu})}{\pi(\mathbf{S}|\mathbf{N}, \gamma^S, \boldsymbol{\beta}, \nu)} \frac{q(\nu|\tilde{\nu})}{q(\tilde{\nu}|\nu)}\right\} \\ &= \min\left\{1, \exp\left\{\sum_{i=1}^I \left[\log \frac{\Gamma(\nu N_i)}{\Gamma(\tilde{\nu} N_i)} + N_i \left(\log \frac{\tilde{\nu}^{\tilde{\nu}}}{\nu^{\nu}}\right) - N_i(\tilde{\nu} - \nu)(-\log(N_i S_i) + \boldsymbol{\beta}'\mathbf{x}_i^S \right. \right. \right. \\ &\quad \left. \left. + \gamma_{R(i)}^S + S_i / \exp(\boldsymbol{\beta}'\mathbf{x}_i^S + \gamma_{R(i)}^S)\right)\right]\right\} \frac{\nu + \sigma_\nu - \max\{0, \nu - \sigma_\nu\}}{\tilde{\nu} + \sigma_\nu - \max\{0, \tilde{\nu} - \sigma_\nu\}}, \end{aligned}$$

where  $\sigma_\nu$  is the semirange of the uniform proposal density for  $\nu$ .

The prior (4) requires that the parameters  $\mu_F$  and  $\sigma_F$  of the normal prior distribution are updated in addition to the latent variables and the parameters. For notational convenience

we introduce  $\tau_F = 1/\sigma_F^2$ . The acceptance probabilities for the latent variables are now

$$\begin{aligned} p(\tilde{\gamma}_l^F, \gamma^F) &= \min\left\{1, \frac{\pi(\mathbf{N}|\tilde{\gamma}_l^F, \gamma_{-l}^F, \boldsymbol{\alpha}) \pi(\tilde{\gamma}_l^F, \gamma_{-l}^F|\mu_F, \sigma_F^2)}{\pi(\mathbf{N}|\gamma^F, \boldsymbol{\alpha}) \pi(\gamma^F|\mu_F, \sigma_F^2)}\right\} \\ &= \min\left\{1, \exp\left\{\sum_{i:R(i)=l} [(\tilde{\gamma}_l^F - \gamma_l^F)N_i - (e^{\tilde{\gamma}_l^F} - e^{\gamma_l^F})e_i \exp(\boldsymbol{\alpha}'\mathbf{x}_i^F) \right. \right. \\ &\quad \left. \left. - \frac{\tau_F}{2}[(\tilde{\gamma}_l^F)^2 - (\gamma_l^F)^2 - 2\mu_F(\tilde{\gamma}_l^F - \gamma_l^F)]\right]\right\}. \end{aligned}$$

The acceptance probability for the claim size model changes similarly. Next we give the acceptance probabilities for  $\mu_F$  and  $\tau_F$ . The proposal densities for  $\mu_F$  and  $\tau_F$  are also uniform, centered at the current values for  $\mu_F$  and  $\tau_F$ . For  $\tau_F$  we truncate at zero. The values of  $a_F, b_F, c_F$  and  $d_F$  are taken to be fixed, for example as in Section 8.3. We obtain

$$\begin{aligned} p(\tilde{\mu}_F, \mu_F) &= \min\left\{1, \frac{\pi(\gamma^F|\tilde{\mu}_F, \sigma_F^2) \pi(\tilde{\mu}_F|a_F, b_F)}{\pi(\gamma^F|\mu_F, \sigma_F^2) \pi(\mu_F|a_F, b_F)}\right\} \\ &= \min\left\{1, \exp\left\{-\frac{\tau_F}{2}[2(\mu_F - \tilde{\mu}_F) \sum_{j=1}^J \gamma_j^F + I(\tilde{\mu}_F^2 - \mu_F^2)] \right. \right. \\ &\quad \left. \left. - \frac{1}{2b_F^2}[\tilde{\mu}_F^2 - \mu_F^2 - 2a_F(\tilde{\mu}_F - \mu_F)]\right]\right\} \end{aligned}$$

and

$$\begin{aligned} p(\tilde{\tau}_F, \tau_F) &= \min\left\{1, \frac{\pi(\gamma^F|\mu_F, \tilde{\sigma}_F^2) \pi(\tilde{\tau}_F|c_F, d_F) q(\tau_F|\tilde{\tau}_F)}{\pi(\gamma^F|\mu_F, \sigma_F^2) \pi(\tau_F|c_F, d_F) q(\tilde{\tau}_F|\tau_F)}\right\} \\ &= \min\left\{1, \left(\frac{\tilde{\tau}_F}{\tau_F}\right)^{J/2+d_F-1} \exp\left\{-(\tilde{\tau}_F - \tau_F)(d_F/c_F + \frac{1}{2} \sum_{j=1}^J (\gamma_j^F - \mu_F)^2)\right\} \right. \\ &\quad \left. \times \frac{\tau_F + \sigma_\tau - \max\{0, \tau_F - \sigma_\tau\}}{\tilde{\tau}_F + \sigma_\tau - \max\{0, \tilde{\tau}_F - \sigma_\tau\}}\right\}, \end{aligned}$$

where  $\sigma_\tau$  is the semirange of the proposal for  $\tau_F$ .

## References

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Wiley, New York.
- Baxter, L. A., Coutts, S. M. & Ross, G. A. F. (1980). Applications of linear models in motor insurance. In *Proc. 21st Int. Congr. Actuaries*, vol. 2, pp. 11–29. Stämpfli, Bern.
- Besag, J. E., York, J. C. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43**, 1–59. With discussion.
- Biggeri, A., Divino, F., Frigessi, A., Lawson, A. B., Böhning, D., Lesaffre, E. & Viel, J. (1999). Introduction to spatial models in ecological analysis. In *Disease mapping and risk assessment for public health* (Eds. A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J. Viel & R. Bertollini), pp. 181–192. Wiley, Chichester.
- Boskov, M. & Verrall, R. J. (1994). Premium rating by geographic area using spatial models. *ASTIN Bull.* **24**(1), 131–143.

- Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9–25.
- Brockman, M. J. & Wright, T. S. (1992). Statistical motor rating: Making effective use of your data. *Journal of the Institute of Actuaries* **119**, 457–543.
- Brooks, S. & Roberts, G. (1999). Assessing convergence of Markov chain Monte Carlo algorithms. *Statist. Comput.* **8**, 319–335.
- Carlin, B. P. & Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall, London.
- Clayton, D. G. (1996). Generalized linear mixed models. In *Markov chain Monte Carlo in practice* (Eds. W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 275–302. Chapman and Hall, London.
- Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.
- Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**(4), 473–511.
- Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In *Markov chain Monte Carlo in practice* (Eds. W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 1–19. Chapman and Hall, London.
- Glad, I., Balducci, M., Frigessi, A., Scalia Tomba, G. & Pezzotti, P. (2000). Bayesian back-calculation with hiv seropositivity notifications. To appear in *Biometrics*.
- Haberman, S. & Renshaw, A. E. (1996). Generalized linear models in actuarial science. *The Statistician* **45**(4), 407–436.
- Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized additive models*, Chapman and Hall, London.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression*. Wiley, London.
- Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *J. R. Statist. Soc. B* **58**(4), 619–678.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Second edition. Chapman and Hall, London.
- Mollié, A. (1996). Bayesian disease mapping. In *Markov chain Monte Carlo in practice* (Eds. W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 359–376. Chapman and Hall, London.
- Nelder, J. A. & Verrall, R. J. (1997). Credibility theory and generalized linear models. *ASTIN Bull.* **27**, 71–82.
- Renshaw, A. E. (1994). Modeling the claims process in the presence of covariates. *ASTIN Bull.* **24**(2), 265–285.
- Rue, H. & Hurn, M. A. (1999). Bayesian object identification. *Biometrika* **86**(3), 649–660.

- Spiegelhalter, D., Carlin, B. & Best, N. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Working paper.
- Taylor, G. C. (1989). Use of spline functions for premium rating by geographic area. *ASTIN Bull.* **19**(1), 89–122.
- Tjelmeland, H. & Besag, J. (1998). Markov random fields with higher-order interactions. *Scand. J. Statist* **25**(3), 415–433.
- Tomberlin, T. J. (1988). Predicting accident frequencies for drivers classified by two factors. *J. Amer. Statist. Assoc.* **83**(402), 309–321.
- Venables, W. N. & Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.
- Wand, M. P. & Jones, M. C. (1995). *Kernel smoothing*. Chapman and Hall, London.
- Wikle, C., Berliner, L. & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and ecological statistics* **5**(2), 117–154.
- Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**(2), 251–267.

## Acknowledgements

This work was supported by The Research Council of Norway (NFR) under grant no. 110673/420 (Numerical Computations in Applied Mathematics) and grant no. 121144/420 (Knowledge, data and decisions). We thank Gjensidige Forsikring for providing the data used in Section 8 and particularly Jørgen Hagen. The first author thanks the Norwegian Computing Center for support. We wish to thank Nicky Best and Clare Marshall for valuable comments. Figure 1 is generated using the software Statistisk Sett.