Insurance: Mathematics and Economics 62 (2015) 194–201

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Bayesian total loss estimation using shared random effects

Carolin Baumgartner, Lutz F. Gruber *, Claudia Czado

Lehrstuhl für Mathematische Statistik, Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany

ARTICLE INFO

Article history:
Received April 2014
Received in revised form
January 2015
Accepted 25 February 2015
Available online 31 March 2015

Keywords:
Total loss
Claim size
Claim count
Shared parameter model
Dependence
Generalized linear mixed model
Bayesian inference
Markov Chain Monte Carlo

ABSTRACT

The pricing of insurance policies requires estimates of the total loss. The traditional compound model imposes an independence assumption on the number of claims and their individual sizes. Bivariate models, which model both variables jointly, eliminate this assumption. A regression approach allows policy holder characteristics and product features to be included in the model. This article presents a bivariate model that uses joint random effects across both response variables to induce dependence effects. Bayesian posterior estimation is done using Markov Chain Monte Carlo (MCMC) methods. A real data example demonstrates that our proposed model exhibits better fitting and forecasting capabilities than existing models.

1. Introduction

Estimating the total loss of an insurance portfolio is a prime objective in actuarial science. Traditionally, a compounding approach is used. A first such model has been given by Lundberg (1903). Here individual claims are aggregated to model the total loss under the assumption that individual claim sizes and the claim frequency are mutually independent variables. This assumption has been fundamental in actuarial risk theory (see for example Willmot and Lin, 2001, Klugman et al., 2004 and Kaas et al., 2009). Recent extensions include Bayesian (e.g. Ausin et al., 2011) and nonparametric estimation (e.g. Vilar et al., 2009) approaches. However this approach does not easily allow the inclusion of policy holder characteristics and thus a different approach will be followed.

In particular the total loss can be written as the product of the average claim size and the number of claims. We call this approach a product approach. This approach allows characteristics of the policy holder and the insurance product to be included in the model as regression variables. The inclusion of such characteristics is important for the pricing of an insurance product. While separate analyses using a compound approach are also feasible, this would result in information loss compared to a joint analysis considered here.

The standard product model uses a Poisson generalized linear model (GLM) for the claim counts and a Gamma GLM for the average claim sizes. In the first applications the independence of the two components was assumed (e.g. Dimakos and Frigessi, 2002). This Poisson–Gamma model for total loss can therefore be extended in two ways: first, by the use of different models for claim counts and claim sizes; second, by eliminating the independence assumption of claim counts and average claim sizes. There are many examples of different marginal models discussed in existing literature (Haberman and Renshaw, 1996; deJong and Heller, 2008; Ohlsson and Johansson, 2010). We pursue the second approach and show that the standard Poisson–Gamma model can be improved dramatically by adding a dependence model for claim counts and claim sizes. It is shown in Gschlößl and Czado (2007) and Krämer et al. (2013) that real data does not back the independence assumption. Our work is inspired by Krämer et al. (2013), who used a copula to model the dependence characteristics between claim counts and average claim sizes. We will allow for more structured dependence patterns by using shared random effects.

Common multivariate distributions such as the multivariate normal distribution require that all margins follow the same distribution. This is not suitable for our situation, because we want to model the dependence characteristics between a metric variable, average claim size, and a count variable, claim count. Copulas can combine arbitrary marginal distributions to form a multivariate distribution (Joe, 1997; Nelsen, 2006). The copula approach has been applied to actuarial problems such as bivariate loss distributions (DeLeon and Wu, 2011) and describing dependencies among

* Corresponding author.
E-mail addresses: lutz.gruber@ma.tum.de (L.F. Gruber), cczado@ma.tum.de (C. Czado).

http://dx.doi.org/10.1016/j.insmatheco.2015.02.008
0167-6687/

loss triangles (De Jong, 2012; Shi and Frees, 2011). For example, Krämer et al. (2013) modeled total loss distributions using a combination of regression models and a copula.

We propose to include shared fixed effects and random effects in the regression models for average claim size and claim count. This allows the joint model's dependence effects to vary with different covariate values. In contrast, the copula-based model considered in Krämer et al. (2013) uses a dependence model that does not change with the covariates. While linear mixed models and generalized linear mixed models are a common occurrence in the literature (Fahrmeir et al., 2013; de Leon and Chough, 2013), multivariate mixed models are less common. Shared parameter models have been used in survival analysis, but we did not find evidence that they were previously used in insurance statistics. Our main contribution is a Bayesian formulation of a shared parameter model for total loss estimation and an MCMC-based implementation for parameter inference. A real-data example shows that our model is superior to existing ones, which suggests that existing models inadequately reflect varying dependence effects found in real data.

The paper is organized as follows. We discuss the general setup of a mixture model (Section 2) and propose a suitable joint model for the claim size and claim count (Section 3). Section 4 presents theoretical considerations about Bayesian inference as well as a Monte Carlo Markov Chain implementation of our model; Section 5 does the same for Krämer et al. (2013)'s copula model. A simulation study investigates the small sample performance of the shared random intercept model (Section 6). A case study examines the descriptive capabilities of our model and the reference model of Krämer et al. (2013) using the same German car insurance data set as in Krämer et al. (2013) (Section 7). We conclude with a brief discussion and an outlook on future research in Section 8.

2. General shared parameter and mixture models

2.1. General set-up

Suppose $\mathbf{V} = (V_1, \ldots, V_P)^\top$ is a P-dimensional random vector with components $V_p,\ p = 1, \ldots, P$. The general shared parameter model induces dependence in this random vector through a common random parameter γ, which has a density function g. The joint density of $\mathbf{V}$ follows as

$$f_{\mathbf{V}}(\mathbf{v}) = \int_{-\infty}^{\infty} f(v_1, \ldots, v_P|\gamma)g(\gamma)d\gamma.$$

Assuming conditional independence between the random variables V_p given γ for all $p = 1, \ldots, P$, this simplifies to

$$f_{\mathbf{V}}(\mathbf{v}) = \int_{-\infty}^{\infty} \prod_{p=1}^{P} f_p(v_p|\gamma)g(\gamma)d\gamma. \tag{1}$$

In model (1), the shared parameter γ acts as a frailty parameter, which can also be interpreted as a random effect with density g. Shared parameter models are particularly popular in survival analysis (see Hogan and Laird, 1997 for repeated measures and event times and Vonesh et al., 2006 for longitudinal data and event times). Model (1) is easily extended to allow for component-specific random effects γ_p for $p = 1, \ldots, P$ with joint distribution $g(\gamma_1, \ldots, \gamma_P)$; the density function of the corresponding model is

$$f_{\mathbf{V}}(\mathbf{v}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{p=1}^{p} f_p(v_p|\gamma_p)g(\gamma_1, \ldots, \gamma_P)d\gamma_1 \cdots d\gamma_P. \tag{2}$$

Dependent random effects occur when the joint distribution $g(\gamma_1, \ldots, \gamma_P)$ cannot be written as a product of densities, but the use of independent random effects is more common. The construction underlying (2) can be used to construct a variety of multivariate discrete distributions (Sarabia and Gómez-Déniz, 2008).

2.2. Interpretation as GLMMs

Generalized linear models (GLM) are regression models where the response variable can take a non normal distribution, such as a Poisson or binomial distribution McCullagh and Nelder (1989); Dobson and Barnett (2008). The books deJong and Heller (2008); Ohlsson and Johansson (2010) present applications within an insurance context. A generalized linear mixed model (GLMM) is an extension of the generalized linear model that has an independent random effect in addition to the fixed effects (West et al., 2007; McCulloch et al., 2008; Demidenko, 2013). Recently they also have been used within actuarial science (Antonio and Beirlant, 2007 and Chapter 16 of Frees et al., 2014). The response Y_p takes the role of V_p from (2). The random effect γ_p and fixed effects $\mathbf{w}_p$ connect to the distribution of Y_p through the link function h,

$$h(\mu_p^Y) = \mathbf{w}_p^\top \boldsymbol{\beta} + \gamma_p,$$

where μ_p^Y is the conditional expectation of Y_p given γ_p and $\boldsymbol{\beta}$ are the regression coefficients of the fixed effects $\mathbf{w}_p$. Typically, the conditional distribution of Y_p given γ_p is assumed to follow an exponential family distribution with density f_Y.

A common assumption is that the random effects $\gamma_p,\ p = 1, \ldots, P$, are i.i.d. from the same univariate distribution $g(\cdot; \tau)$, where τ is a hyper parameter. If $f_Y(\cdot; \boldsymbol{\beta}, \gamma_p, \mathbf{w}_p)$ denotes the density of Y_p with covariates $\mathbf{w}_p$, regression coefficients $\boldsymbol{\beta}$ and random effects γ_p, the joint density of $\mathbf{Y} = (Y_1, \ldots, Y_P)^\top$ is

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\beta}, \tau) = \prod_{p=1}^{P} \left[\int_{-\infty}^{\infty} f_Y(y_p; \boldsymbol{\beta}, \gamma_p, \mathbf{w}_p)g(\gamma_p; \tau)d\gamma_p \right]. \tag{3}$$

Again, this result holds under the assumption that the Y_p are conditionally independent given γ_p for all $p = 1, \ldots, P$. To accommodate groups of response variables $Y_{ij},\ i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, we allow for independent, group-specific random effects $\gamma_j,\ j = 1, \ldots, J$, with joint density $g(\cdot, \tau)$. In this case the index p is replaced by the double index ij and the conditional mean μ_{ij}^Y given γ_j is

$$h_Y(\mu_{ij}^Y) = \mathbf{w}_{ij}^\top \boldsymbol{\beta} + \gamma_j, \tag{4}$$

for some link function h_Y and $\mathbf{w}_{ij} := (w_{ij1}, \ldots, w_{ijK})^\top$. Assuming conditional independence of $Y_{i,j}$ given γ_j, the joint density can be expressed as

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\beta}, \tau) = \prod_{j=1}^{J} \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n_j} f_Y(y_{ij}; \boldsymbol{\beta}, \gamma_j, \mathbf{w}_{ij})g(\gamma_j; \tau)d\gamma_j \right]. \tag{5}$$

2.3. Multivariate GLMMs

In a final step we extend the GLMM (4) and (5) to a multivariate setting. In this article, we restrict ourselves to the bivariate case, because this is the relevant case for our application to total loss estimation. The first response variable X_{ij} is the average claim size with mean μ_{ij}^X; the second response variable Y_{ij} is the claim count with mean μ_{ij}^Y. The covariates for claim size X_{ij} are denoted by $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijL})^\top$ and the corresponding regression coefficients are $\boldsymbol{\alpha}$. The covariates for the claim counts Y_{ij} are denoted by $\mathbf{w}_{ij}$ and their regression coefficients are $\boldsymbol{\beta}$. The dependence between the claim size X_{ij} and claim counts Y_{ij} is modeled by a shared random effects γ_p of the form

$$h_X(\mu_{ij}^X) = \mathbf{w}_{ij}^\top \boldsymbol{\beta} + \gamma_j$$

$$h_Y(\mu_{ij}^Y) = \mathbf{z}_{ij}^\top \boldsymbol{\alpha} + s\gamma_j, \tag{6}$$

where h_Y and h_X are appropriate known link functions and s is a scale parameter. The scale parameter s allows the conditional means of X_{ij} and Y_{ij} to be affected differently by the shared random effect γ_p. Assuming that X_{ij} and Y_{ij} are independent given γ_j, the joint density of $(\mathbf{Y}, \mathbf{X})$ is

$$f_{\mathbf{Y},\mathbf{X}}(\mathbf{y}, \mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\alpha}, s, \tau) = \prod_{j=1}^{J} \left[\int_{-\infty}^{\infty} \prod_{i=1}^{n_j} f_X(x_{ij}; \boldsymbol{\beta}, \gamma_j, \mathbf{w}_{ij}) \times f_Y(y_{ij}; \boldsymbol{\alpha}, s, \gamma_j, \mathbf{z}_{ij}) g(\gamma_j; \tau) d\gamma_j \right]. \quad (7)$$

2.4. Inference in GLMMs

Methods for inference in GLMMs can be found, for example, in Fahrmeir et al. (2013), de Leon and Chough (2013) and Demidenko (2013) and Berridge and Crouchley (2011) give special emphasis to multivariate GLMMs. While there are frequentist inference methods for linear regression models with normally distributed random effects, such as restricted maximum likelihood (REML), there are no frequentist estimation techniques for GLMMs. The lack of an asymptotic frequentist theory in these models is a major reason to use a Bayesian approach instead (Fahrmeir et al., 2013). While learning Bayesian models is computationally more expensive than just solving closed-form equations, computing the integrals that appear in the densities (3), (7) requires the use of numerical procedures.

3. Shared random effects model for claim size and claim count

We model the total loss as the product of average claim size and claim counts; this is the same approach proposed by Krämer et al. (2013). To induce dependence between claim size and claim counts, we add categorical random effects to both models whose coefficients are linearly dependent (see Section 2.3).

For this we assume that the portfolio can be divided into J groups $j = 1, \ldots, J$, each of which has n_j group members $i = 1, \ldots, n_j$ and $n := n_1 + \cdots + n_J$. The assumption that the insurance portfolio can be divided into groups is a natural one, given that policy holders that have common traits are typically expected to be more likely to have a similar risk profile than others.

3.1. Model for average claim size

The average claim size X_{ij} of policy holder i in group j is a positive, continuous random variable. The Gamma distribution is right-skewed, non-negative and continuous, which makes it a suitable choice to model insurance claims (deJong and Heller, 2008). We use the mean parametrization of the Gamma distribution, which has location parameter μ, dispersion parameter $\delta > 0$, to model X_{ij}

$$X_{ij} \sim G(\mu_{ij}, \delta),$$

$$f_X(x; \mu, \delta) = \frac{1}{x\Gamma\left(\frac{1}{\delta}\right)} \left(\frac{x}{\mu\delta}\right)^{\frac{1}{\delta}} \exp\left(-\frac{x}{\mu\delta}\right) \quad \text{for } x > 0. \quad (8)$$

3.2. Model for claim counts

For the total loss only claim counts greater than zero are relevant, there it is enough to specify a model for $Y_{i,j} \geq 1$. The number of claims $Y_{ij} \geq 1$ of policy holder i in group j will be modeled by a discrete distribution that has as support the set of positive natural numbers $\mathbb{N}$. Following existing literature (Denuit et al., 2007; Winkelmann, 2008), we choose the zero-truncated Poisson (ZTP) distribution to model the claim counts $Y_{ij} \geq 1$,

$$Y_{ij} \sim ZTP(\lambda_{ij}),$$

$$f_Y(y; \lambda) = \frac{\lambda^y \exp(-\lambda)}{y!(1 - \exp(-\lambda))} \quad \text{for } y = 1, 2 \ldots. \quad (9)$$

3.3. Shared random effects model formulation

A shared random effect accounts for the dependence between average claim size and claim counts in our model. Furthermore, there are K fixed effect covariates $w_{1ij}, \ldots, w_{Kij}$ in the model for claim size X_{ij} and L fixed effect covariates $z_{1ij}, \ldots, z_{Lij}$ in the model for claim counts Y_{ij}. The conditional means μ_{ij} and λ_{ij}, given the shared random effect γ_j, are

$$\ln(\mu_{ij}(\boldsymbol{\beta}, \gamma_j)) = \beta_0 + \beta_1 \cdot w_{1ij} + \cdots + \beta_K \cdot w_{Kij} + \gamma_j$$
$$\ln(\lambda_{ij}(\boldsymbol{\alpha}, \gamma, s)) = \alpha_0 + \alpha_1 \cdot z_{1ij} + \cdots + \alpha_L \cdot z_{Lij} + s * \gamma_j + \ln(e_{ij}). \quad (10)$$

Here e_{ij} is the known exposure time of policy ij, $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_K)^\top$ and $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_L)^\top$. The scaling parameter s adjusts the sign and effect size of the shared parameters on the conditional means.

In the case of a single member per group ($J = n$), the model formulation (10) includes the case of shared policy-specific random intercepts, that is, we have $i = 1, \ldots, n$ policies

$$\ln(\mu_i(\boldsymbol{\beta}, \gamma_i)) = \beta_0 + \beta_1 \cdot w_{1i} + \cdots + \beta_K \cdot w_{Ki} + \gamma_i$$
$$\ln(\lambda_i(\boldsymbol{\alpha}, \gamma_i, s)) = \alpha_0 + \alpha_1 \cdot z_{1i} + \cdots + \alpha_L \cdot z_{Li} + s * \gamma_i + \ln(e_i). \quad (11)$$

The number of random effects are different in both models. Model (10) has J shared random effects, while model (11) has n shared random intercepts.

3.4. Likelihood functions

We derive the likelihood functions of the shared random effects models under the conditional independence assumptions from Section 2. The regression setup assumes that X_i and Y_i are independent given the random effects $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_J)$ across individuals $i = 1, \ldots, n$. The dependence model assumes that X_i and Y_i are independent given the shared random effect $\boldsymbol{\gamma}$.

The conditional joint likelihood of model (10) is therefore given by

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, s | \mathbf{x}, \mathbf{y}, \boldsymbol{\gamma}) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} f_X(x_{ij}; \mu_{ij}(\boldsymbol{\beta}, \gamma_j)) \cdot f_Y(y_{ij}; \lambda_{ij}(\boldsymbol{\alpha}, \gamma_j, s)). \quad (12)$$

Here $\mathbf{x}$ denotes the observations of the average claim sizes x_{ij} and $\mathbf{y}$ denotes the observations of the claim counts. The density f_X and the probability mass function f_Y are specified in (8) and (9). Similarly the conditional joint likelihood of model (11) is derived as

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, s | \mathbf{x}, \mathbf{y}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} f_X(x_i; \mu_i(\boldsymbol{\beta}, \gamma_i)) \cdot f_Y(y_i; \lambda_i(\boldsymbol{\alpha}, \gamma_i, s)). \quad (13)$$

Recall that the unconditional likelihood of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, s)$ requires integration over $\gamma_j, j = 1, \ldots, P$, of (12) and $\gamma_i, \; i = 1, \ldots, n$, of (13), respectively.

4. Parameter inference and posterior simulation of total loss

Bayesian parameter inference estimates the posterior distribution, whose density is proportional to the product of the likelihood and the prior density of all parameters. Since there is no closed-form solution of the normalizing constant of the posterior density, we use Markov Chain Monte Carlo (MCMC) methods (Brooks et al., 2011; Roberts and Rosenthal, 2004) to generate approximate samples from the posterior distribution.

4.1. Priors

We have to choose priors for the random effects γ_j in (10), γ_i in (11), scale parameter s, regression parameters $\beta_1, \ldots, \beta_K$ and $\alpha_1, \ldots, \alpha_L$, and the dispersion parameter δ of claim size.

We choose normal priors for the shared random effect parameters of (10) and (11),

$$\gamma_j \sim N(0, \sigma_\gamma^2) \quad \text{i.i.d. for } j = 1, \ldots, J \text{ and}$$

$$\gamma_i \sim N(0, \sigma_\gamma^2) \quad \text{i.i.d. for } i = 1, \ldots, n, \text{ respectively.}$$

The hyper parameter σ_γ^2 follows an inverse Gamma prior with shape $\alpha_\gamma = 8\frac{1}{4}$ and rate $\beta_\gamma = 1\frac{13}{16}$ (see Table 2) and the scale parameter s of the shared random effects is assumed to be standard normal,

$$\sigma_\gamma^2 \sim IG(\alpha_\gamma, \beta_\gamma) \quad \text{and}$$

$$s \sim N(0, 1).$$

We assume that the covariates are either standardized or on a similar scale. Therefore, we use normal priors with a common fixed variance for the regression coefficients,

$$\beta_k \sim N(0, \sigma_\beta^2) \quad \text{for } k = 1, \ldots, K \text{ and}$$

$$\alpha_l \sim N(0, \sigma_\alpha^2) \quad \text{for } l = 1, \ldots, L.$$

In the real-data example, we will specify the prior only vaguely by setting the prior variances $\sigma_\beta^2 = 100^2$ and $\sigma_\alpha^2 = 100^2$. We use a Gamma prior with shape $\alpha_\delta = 11\frac{1}{9}$ and rate $\beta_\delta = 111\frac{1}{9}$ (see Table 2) for the dispersion parameter δ of the claim size,

$$\delta \sim G(\alpha_\delta, \beta_\delta).$$

We form the joint prior of all parameters $\boldsymbol{\theta} := (\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, \sigma_\gamma^2, s)$ and random effects $\boldsymbol{\gamma}$ as the product distribution of all individual priors. More specifically, let $\pi_{IG}(\cdot, \alpha, \beta)$ and $\pi_G(\cdot, \alpha, \beta)$ denote the density of an inverse Gamma and Gamma random variable with shape α the rate β, respectively, and let ϕ denote the standard normal density. Then the joint prior is

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\gamma}) &= \pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, \sigma_\gamma^2, s, \boldsymbol{\gamma}) \\ &= \left[\prod_{j=1}^{J} \phi\left(\frac{\gamma_j}{\sigma_\gamma}\right)\right]\left[\prod_{k=1}^{K} \phi\left(\frac{\beta_k}{\sigma_\beta}\right)\right]\left[\prod_{l=1}^{L} \phi\left(\frac{\alpha_l}{\sigma_\alpha}\right)\right] \\ &\quad \times \phi(s)\pi_{IG}(\sigma_\gamma^2 | \alpha_\gamma, \beta_\gamma)\pi_G(\sigma_\delta | \alpha_\delta, \beta_\delta). \end{aligned} \tag{14}$$

4.2. Posterior estimation

The posterior distribution of the parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{\gamma}$ is

$$p(\boldsymbol{\theta}, \boldsymbol{\gamma} | \mathbf{x}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, \sigma_\gamma^2, s, \boldsymbol{\gamma} | \mathbf{x}, \mathbf{y}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, \sigma_\gamma^2, s, \boldsymbol{\gamma}). \tag{15}$$

We generate R samples from the posterior distribution p using the Metropolis Hastings algorithm (MH; Metropolis et al., 1953; Hastings, 1970), which is generally described in Roberts and Rosenthal (2004). The MCMC iterates are indexed by $r = 1, \ldots, R$. The components of the multivariate parameter vector $\boldsymbol{\theta}$ are updated one-by-one with normally distributed random walk proposals. Random walk proposals are centered at the previous state of the MCMC sampling chain. The proposal variance for each parameter component is tuned in pilot runs to achieve acceptance rates between 20% and 80%.

4.3. Posterior simulation of total loss

For each MCMC iteration $r = 1, \ldots, R$, we draw $N := \sum_{j=1}^{N} n_j$ or n average claim sizes and claim counts from model specification (10) or (11), respectively, to generate a forecast data set $(\mathbf{x}^r, \mathbf{y}^r)$ using the current posterior sample as the model's parameters. The corresponding total loss iterate $\hat{L}^r$ from this forecast data set $(\mathbf{x}^r, \mathbf{y}^r)$ is computed as

$$\hat{L}^r := \sum_{j=1}^{J} \sum_{i=1}^{n_j} x_{ij}^r y_{ij}^r. \tag{16}$$

The total loss forecasts $\{\hat{L}^r,\ r = 1, \ldots, R\}$ are a Monte Carlo sample from the posterior distribution of the total loss. They can be used to estimate the posterior density of the total loss, to construct point estimates, or to describe uncertainty through credible intervals.

5. Posterior inference for a copula-based model for total loss

Krämer et al. (2013) modeled the dependence between claim size and claim counts using a bivariate copula. We use (Baumgartner, 2013)'s Bayesian extension of Krämer et al. (2013)'s model as a benchmark in our real-data example.

5.1. Model specification

We model the ith average claim size X_i by a Gamma regression model with density $f_X(\cdot; \mu_i := \exp(\boldsymbol{\beta}^\top \mathbf{w}_i), \delta)$ and distribution function $F_X(\cdot; \mu_i, \delta)$, were $\mathbf{w}_i$ are the covariates including an intercept for X_i. The corresponding number of claims Y_i are modeled by a zero-truncated Poisson regression model with density $f_Y(\cdot; \lambda_i) := \exp(\boldsymbol{\alpha}^\top \mathbf{z}_i + \ln(e_i))$ and distribution function $F_Y(\cdot; \lambda_i)$. Here $\mathbf{z}_i$ are the covariates including an intercept for Y_i.

The joint density of (X_i, Y_i) in the copula based model is given by

$$\begin{aligned} f_{X,Y}(x_i, y_i) &= \frac{\partial}{\partial x_i} P(X_i \le x_i, Y_i = y_i) = f_X(x_i; \mu_i, \delta) \\ &= f_X(x_i; \mu_i, \delta)\,[D_1(F_X(x_i; \mu_i, \delta), F_Y(y_i; \lambda_i); \theta) \\ &\quad - D_1(F_X(x_i; \mu_i, \delta), F_Y(y_i - 1; \lambda_i); \theta)], \end{aligned} \tag{17}$$

where $D_1(u, v; \theta) := \frac{\partial}{\partial u} C(u, v; \theta)$ and $C(\cdot, \cdot; \theta)$ is a copula with parameter θ (see Krämer et al., 2013).

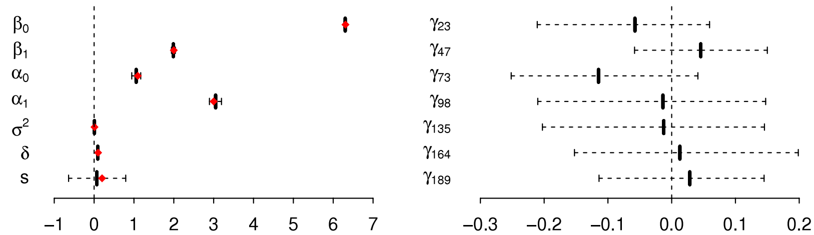
5.2. Candidate copula families

We allow for the bivariate Gauss, Clayton, Gumbel and Frank copula families. The copulas are parametrized by Kendall's $\tau \in [-1, 1]$ instead of the natural copula parameters. Furthermore, the classical Poisson–Gamma model, which assumes independence between claim counts and claim sizes, is included to put the results in perspective.

5.3. Priors

We utilize similar priors as for the shared random effect models (see Table 1, compare Section 4.1) and a normal distribution truncated to the interval $[-1, 1]$ for the Kendall's τ parameters,

$$\tau \sim N_{[-1,1]}(\mu_\tau, \sigma_\tau^2).$$



(a) Parameters β_0, β_1, α_0, α_1, σ_γ^2, δ and s; the diamonds indicate the true values.

(b) Shared random effect parameters γ_{23}, γ_{47}, γ_{73}, γ_{98}, γ_{135}, γ_{164}, and γ_{189}.

Fig. 1. Estimated 95% credible intervals and posterior modes (thick vertical bars).

Table 1
Prior distributions for the copula based total loss regression model.

Parameter	Prior	Parameter	Prior
δ	$G(\nu_\delta, \beta_\delta)$	β_k	$N(0, \sigma_\beta^2)$
τ	$N_{[-1,1]}(\mu_\tau, \sigma_\tau^2)$	α_l	$N(0, \sigma_\alpha^2)$

5.4. *Posterior estimation*

We apply a basic MH sampler for posterior estimation, which was developed by Baumgartner (2013, Chapter 4). Again we update the components of the multivariate parameter vector $\boldsymbol{\theta} := (\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta, \tau)^\top$ one-by-one with normally distributed random walk proposals. As before, the proposal variance for each parameter component is tuned in pilot runs to achieve acceptance rates between 20% and 80%.

6. Simulation study: shared random intercept model

To illustrate the performance of the MH algorithm we present a simulation study using the shared random intercept model (11). The results presented here are based on Chapter 5 of Baumgartner (2013).

6.1. *Study set-up*

We choose $n = 200$, which allows for 200 shared random effects and include a single covariate for each component that takes uniform values in $[-1, 1]$. We choose the regression parameters such that the expected average claim size and number of claims are 1000 and 10, respectively. All exposure times e_i are set to 1. This model has regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^\top$, dispersion δ, shared random effects γ_j for $j = 1, \ldots, 200$, and scale s. The parameters' true values and prior specifications are summarized in Table 2.

6.2. *Results*

The results are based on 10,000 MH iterations; the sampling chain did not exhibit burn-in behavior, given that suitable starting values were used. Furthermore, the effective sample size (Plummer et al., 2005) of all parameters is always large enough for posterior inference (Table 2).

Fig. 1(a) shows the posterior mode estimates and 95% credible intervals of the parameters β_0, β_1, α_0, α_1, σ_γ^2, δ and s. All posterior mode estimates are extremely close to the true value of the parameters and the 95% credible intervals of all parameters except for s are very tight (Table 2). Fig. 1(b) shows the posterior modes and 95% credible intervals of seven shared random effect parameters $\gamma_i, i = 23, 47, 73, 98, 135, 164, 189$. Dispersed around zero, the estimates are in agreement with their $N(0, \sigma_\gamma^2)$ priors. These results validate our MCMC posterior sampler.

7. Real-data example: German car insurance data

We provide a novel analysis of a German car insurance data set that was previously discussed in Czado et al. (2012) and Krämer et al. (2013) using our shared random effects model. The data was collected in 2000 and contains 7663 policy groups and includes the seven categorical covariates sex (`sex`), regional class (`rcl`), bonus (`bonus`), deductible (`ded`), distance driven (`dist`), age group (`age`) and construction year (`const`) as well as information on the exposure time.

7.1. *Description of the models*

Copula model. The full copula model (17) uses all covariates as fixed effects. We estimate that model and then exclude any covariates of which the 95% credible interval of the parameter estimate contains zero. Table 3 shows the selected covariates, which will be used in the remainder. There are 13 parameters specifying the average claim size model; this number includes the dispersion parameter δ and the intercept β_0. The claim counts model has 11 regression coefficients including the intercept α_0. In addition, the random effects variance σ_γ^2 and scale parameter s specify the dependence between claim size and claim counts. Note that our copula model is different from Krämer et al. (2013)'s in that we include `rcl` in the model for claim sizes, but do not include `const`. Furthermore, Krämer et al. (2013) include `dist` as a fixed effect for claim counts and we do not.

Shared random effects model. The shared random effects model uses the same covariates as the copula model, however `rcl` is used as a random effect and not as a fixed effect. Using `rcl` as a random effect will introduce spatial clustering of the claim size–claim counts dependence structure.

7.2. *Results: parameter inference*

This section only discusses the parameter estimates of our shared random effects model. Section 7.3 will compare the estimated total loss distributions of the copula-based model with our model.

MCMC output. The MCMC chain produced by our MH sampler did not exhibit burn-in behavior, given that suitable starting values were used. The posterior estimates discussed in this section are based on 10,000 states of the sampling chain. Our choice of the number of MCMC updates is based on the results of the simulation study (Section 6) and an in-depth analysis of the sampling chains and trace plots. Furthermore, we found that no burn-in period is necessary, given our good starting values. Our computer took four hours to generate 10,000 MCMC-based posterior draws for this data set.

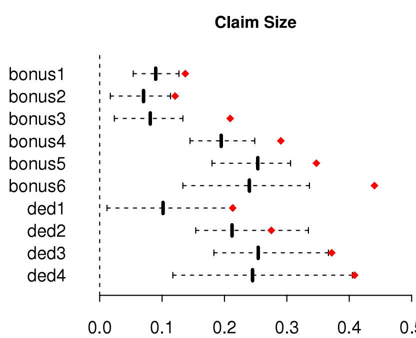
Fixed effects. Figs. 2(a) and (b) show the posterior mode estimates as well as 95% credible intervals of the fixed effect regression coefficients of the models for average claim size and claim counts,

Table 2
True parameter values, priors, posterior summary and effective sample size (ESS) based on 10,000 MH iterations.

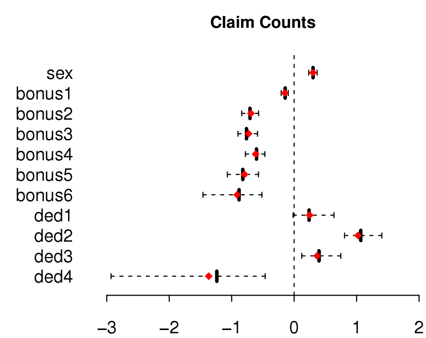
Parameter	True value	Prior			Sample posterior		ESS
		Distribution	Mode	Std. dev.	Mode	95% Cred. int.	
β_0	6.31	$N(0, 100^2)$	0	100	6.30	(6.27, 6.33)	5561
β_1	2.00	$N(0, 100^2)$	0	100	1.99	(1.94, 2.04)	460
α_0	1.10	$N(0, 100^2)$	0	100	1.06	(0.94, 1.17)	183
α_1	3.00	$N(0, 100^2)$	0	100	3.05	(2.89, 3.20)	207
σ_γ^2	0.01	$\Gamma^{-1}(8\frac{1}{4}, 1\frac{13}{16})$	0.20	0.1	0.01	(0.01, 0.01)	212
$\gamma_{j=1:200}$		$N(0, \sigma_\gamma^2)$	0	σ_γ			(1740, 2442)
δ	0.10	$\Gamma(11\frac{1}{9}, 111\frac{1}{9})$	0.09	0.03	0.09	(0.07, 0.12)	395
s	0.20	$N(0, 1)$	0	1	0.07	(−0.64, 0.79)	566

Table 3
Fixed effects (x) and random effects (r) used to model average claim size and claim counts.

Name	Description	# Categories	Claim size		Claim counts	
			Us	Krämer et al. (2013)	Us	Krämer et al. (2013)
`sex`	Driver's sex	2	–	–	x	x
`rcl`	Regional class	8	x/r	–	x/r	x
`bonus`	No-claims bonus	7	x	x	x	x
`ded`	Type of deductible	5	x	x	x	x
`dist`	Distance driven	5	–	–	–	x
`age`	Driver's age	6	–	–	–	–
`const`	Car construction year	7	–	x	–	–



(a) Gamma model for claim size.



(b) ZTP model for claim counts.

Fig. 2. Posterior mode estimates (vertical lines) and 95% credible intervals. The diamonds indicate Krämer et al. (2013)'s point estimates.

respectively. None of the coefficients' credible intervals contain the zero, which suggests that our models cannot be reduced any further. The use of different covariates in our models and Krämer et al. (2013)'s models explains the discrepancies between our estimates and theirs. Asymmetric posterior distributions are reflected in asymmetric posterior credible intervals, see, for example, `ded2` in Fig. 2(a) and `bonus6` as well as `ded4` in Fig. 2(b). A Bayesian approach, which can capture such asymmetries, will yield superior uncertainty estimates and risk forecasts than frequentist methods.

Shared random effects. Fig. 3(a) shows the posterior estimates of the shared random effects $\gamma_{\texttt{rcl1}}, \ldots, \gamma_{\texttt{rcl8}}$. The 95% credible interval of $\gamma_{\texttt{rcl8}}$ is the only one not to contain zero. The scale factor s is estimated at $\hat{s} = -3.98$ (95% credible interval: $(-5.13, -2.81)$; Fig. 3(b)). The negative sign of s shows that claim size and claim counts are negatively associated, which means that in a given year, there are typically many small claims and only few big claims. This is a more favorable situation than independence or positive association, which would lead to higher loss estimates.

Joint model. The parameters σ^2 and δ of the joint model have very tight 95% credible intervals (Fig. 3(b)). The posterior estimate of the dispersion parameter δ (95% credible interval (0.30, 0.32)) indicates little variation of the dispersion parameter for average claim size distribution. The posterior estimate of the variance parameter σ^2 (95% credible interval (0.01, 0.01) not including zero) indicates the presence of a small within group association.

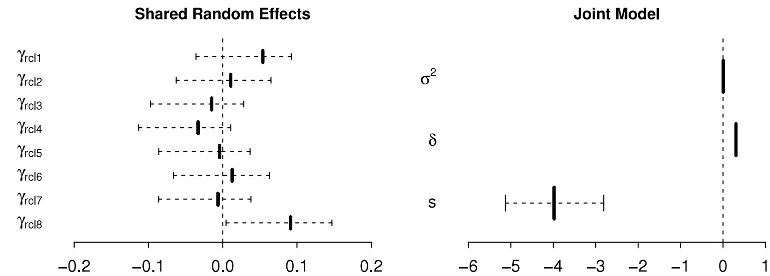
7.3. *Results: total loss estimation*

We use the last 200 MH iterations, $r = 9801 : 10,000$, to sample from the total loss distribution as described in Section 4. Fig. 4 shows the loss estimates of our shared random effects model in comparison to several copula-driven models from Section 5. Our shared random effects model performs superior to each of the copula-based models. Our shared random effects model and the independence model are the only models whose 95% prediction intervals, $I_{\text{shared random effects}} = (74\,818, 80\,281)$ and $I_{\text{independence}} = (76\,033, 81\,921)$, respectively, contain the observed loss $L = 76\,071$.

8. Discussion and outlook

This paper presents a total loss model based on a joint bivariate model of the claim size and claim count. Our novel approach introduces a bivariate GLMM with shared parameters between the two response variables to capture their dependence properties.

The major innovation of our model over Lundberg (1903)'s compound model is that we do not assume independence of claim



(a) Shared random effects $\gamma_{\mathtt{rcl1}}, \ldots, \gamma_{\mathtt{rcl8}}$. (b) Parameters σ^2, δ, s of the joint model.

Fig. 3. Posterior mode estimates (vertical lines) and 95% credible intervals.

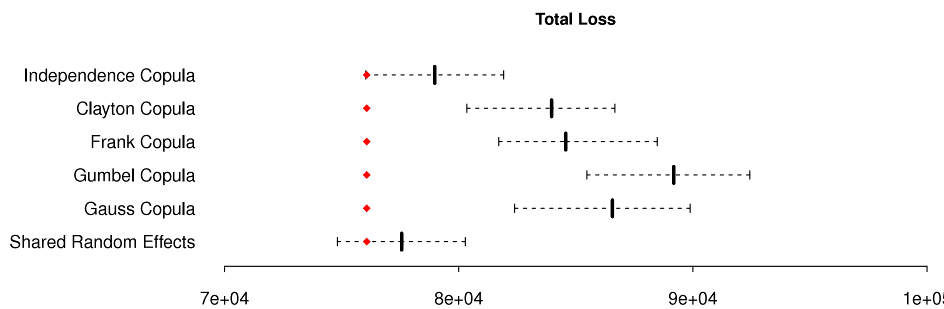


Fig. 4. Estimated 95% prediction intervals of the total loss distribution of our shared random effects model and some copula-based models. The diamonds indicate the observed total loss.

counts and claim size. This allows us to use different factors that influence the distribution parameters of the average claim size and the claim frequency.

The shared random effects allow different dependence characteristics for each grouping category. An obvious extension of this model can include more than one grouping factor and could even include metrical shared mixed effects. While Krämer et al. (2013)'s copula-based approach used the same basic setup, they did not include group-specific dependence characteristics in their model.

The analysis of a German car insurance portfolio shows that the dependence structure varies across different geographical regions. Our shared random effects model can reflect such varying dependence characteristics, while Krämer et al. (2013)'s copula-based approach cannot.

The improved uncertainty estimates and loss forecasts of our shared random effects model provide crucial information to estimate and mitigate risk. This is highly relevant to insurance companies, whose long-term viability relies on policies being priced according to their risk profile.

Future research will investigate an extension of Krämer et al. (2013)'s copula-based model to include group-specific dependencies.

Acknowledgments

The numerical computations were performed on a Linux cluster supported by DFG grant INST 95/919-1 FUGG.

We thank the Editor and an unnamed referee for constructive comments on our original manuscript.

References

Antonio, K., Beirlant, J., 2007. Actuarial statistics with generalized linear mixed models. Insurance Math. Econom. 40 (1), 58–76.
Ausin, M., Vilar, J., Cao, R., Gonzalez-Fragueiro, C., 2011. Bayesian analysis of aggregate loss models. Math. Finance 21, 257–279.
Baumgartner, C., 2013. Comparison of multivariate mixed and copula models for the estimation of losses (Master's thesis), Technische Universität München.
Berridge, D., Crouchley, R., 2011. Multivariate Generalized Linear Mixed Models Using R. A Chapman & Hall book. Taylor & Francis.
Brooks, S., Gelman, A., Jones, G.L., Meng, X.-L., 2011. Handbook of Markov Chain Monte Carlo. Chapman and Hall.
Czado, C., Kastenmeier, R., Brechmann, E.C., Min, A., 2012. A mixed copula model for insurance claims and claim sizes. Scand. Actuar. J. 2012 (4), 278–305.
De Jong, P., 2012. Modeling dependence between loss triangles. N. Am. Actuar. J. 16 (1), 74–86.
deJong, P., Heller, G.Z., 2008. Generalized Linear Models for Insurance Data. In: International Series on Actuarial Science, Cambridge University Press.
de Leon, A., Chough, K., 2013. Analysis of Mixed Data: Methods & Applications. Taylor & Francis.
DeLeon, A., Wu, B., 2011. Copula-based regression models for a bivariate mixed discrete and continuous outcome. Stat. Med. 30, 175–185.
Demidenko, E., 2013. Mixed Models: Theory and Applications with R. In: Wiley Series in Probability and Statistics, Wiley.
Denuit, M., Marechal, X., Pitrebois, S., Walhin, J., 2007. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. Wiley.
Dimakos, X., Frigessi, A., 2002. Bayesian premium rating with latent structure. Scand. Actuar. J. 3, 162–184.
Dobson, A., Barnett, A., 2008. An Introduction to Generalized Linear Models, third ed. Chapman and Hall.
Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2013. Regression: Models, Methods and Applications. Springer.
Frees, E., Derrig, R., Meyers, G., 2014. Predictive Modeling Applications in Actuarial Science. In: International Series on Actuarial Science, Bd. 1. Cambridge University Press.
Gschlößl, S., Czado, C., 2007. Spatial modelling of claim frequency and claim size in non-life insurance. Scand. Actuar. J. 202–225.
Haberman, S., Renshaw, A., 1996. Generalized linear models and actuarial science. Statistician 45, 407–436.
Hastings, W., 1970. Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57 (1), 97–109.
Hogan, J., Laird, N., 1997. Mixture models for the joint distribution of repeated measures and event times. Stat. Med. 16, 239–257.
Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall.
Kaas, R., Goovaerts, M., Dhaene, J., 2009. Modern Actuarial Risk Theory. Springer-Verlag, Berlin, Heidelberg.
Klugman, S., Panjer, H., Willmot, G., 2004. Loss Models: From Data to Decisions. In: Wiley Series in Probability and Statistics, Wiley.
Krämer, N., Brechmann, E.C., Silvestrini, D., Czado, C., 2013. Total loss estimation using copula-based regression models. Insurance Math. Econom. 53, 829–839.
Lundberg, F., 1903. Approximerad framställning afsannollikhetsfunktionen: II. återförsäkring af kollektivrisker... (Ph.D. thesis), Almqvist & Wiksells Boktr.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models. Chapman and Hall.
McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. Generalized, Linear, and Mixed Models. John Wiley and Sons, Inc.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculation by fast computing machines. J. Chem. Phys. 21, 1087–1092.
Nelsen, R.B., 2006. An Introduction to Copulas. Springer.
Ohlsson, E., Johansson, B., 2010. Non-Life Insurance Pricing with Generalized Linear Models. In: EAA Series, Springer.
Plummer, M., Best, N., Cowles, K., Vines, K., 2005. Output analysis and diagnostics for MCMC. R package version 0.10-3, URL http://cran.rproject.org.

Roberts, G.O., Rosenthal, J.S., 2004. General state space Markov chains and MCMC algorithms. Probab. Surv. 1, 20–71.
Sarabia, J.M., Gómez-Déniz, E., 2008. Construction of multivariate distributions: a review of some recent results. SORT 32 (1), 3–36.
Shi, P., Frees, E.W., 2011. Dependent loss reserving using copulas. Astin Bull. 41 (02), 449–486.
Vilar, J., Cao, R., Ausin, M., Gonzalez-Fraguerio, C., 2009. Nonparametric analysis of aggregate loss models. J. Appl. Stat. 36 (2), 149–166.
Vonesh, E., Greene, T., Schluchter, M., 2006. Shared parameter models for the joint analysis of longitudinal data and event times. Stat. Med. 25, 143–163.
West, B.T., Welch, K.B., Galecki, A.T., 2007. Linear Mixed Models. Chapman and Hall.
Willmot, G., Lin, X., 2001. Compound distributions. In: Lundberg Approximations for Compound Distributions with Insurance Applications. In: Lecture Notes in Statistics, vol. 156. Springer, New York, pp. 51–80.
Winkelmann, R., 2008. Econometric Analysis of Count Data. Springer, Berlin, Heidelberg.