A Mixed Copula Model for Insurance Claims and Claim Sizes

Claudia Czado, Rainer Kastenmeier, Eike Christian Brechmann[1], Aleksey Min

Center for Mathematical Sciences, Technische Universität München
Boltzmannstr. 3, D-85747 Garching, Germany

Abstract

C. Czado, R. Kastenmeier, E. C. Brechmann, A. Min. A Mixed Copula Model for Insurance Claims and Claim Sizes. Scandinavian Actuarial Journal. A crucial assumption of the classical compound Poisson model of Lundberg (1903) for assessing the total loss incurred in an insurance portfolio is the independence between the occurrence of a claim and its claims size. In this paper we present a mixed copula approach suggested by Song et al. (2009) to allow for dependency between the number of claims and its corresponding average claim size using a Gaussian copula. Marginally we permit for regression effects both on the number of incurred claims as well as its average claim size using generalized linear models. Parameters are estimated using adaptive versions of maximization by parts (Song et al. 2005). The performance of the estimation procedure is validated in an extensive simulation study. Finally the method is applied to a portfolio of car insurance policies, indicating its superiority over the classical compound Poisson model. Key words: GLM, copula, maximization by parts, number of claims, average claim size, total claim size.

[1]Corresponding author. E-mail: brechmann@ma.tum.de.

1 Introduction

Total loss estimation in non-life insurance is an important task of actuaries, e.g., to calculate premiums and price reinsurance contracts. A solid estimation of total loss distributions in an insurance portfolio is therefore essential and can be carried out based on models for average claim size and number of claims. In the classical compound Poisson model going back to Lundberg (1903) average claim size and number of claims are assumed to be independent, where claim sizes follow a Gamma distribution, while the number of claims is modeled by a Poisson distribution. However, this independence assumption may not always hold. Gschlößl and Czado (2007), for instance, analyze a comprehensive car insurance data set using a full Bayesian approach. In their analysis, they allow for some dependency between the average claim size and the number of claims and detect that this dependency turns out to be significant.

Based on an arbitrary set of covariates, we construct a bivariate regression model for average claim size and number of claims allowing for dependency between both variables of interest. Ng et al. (2007) model both numerical and categorical variables by a semi-supervised regression approach. Using Least Squares and K-Modes (a clustering algorithm following the K-Means paradigm) they construct a flexible algorithm that allows to capture dependencies by building clusters. If the variables of interest are count variables, Wang et al. (2003) show how to construct a bivariate zero-inflated Poisson regression model. Their model for injury counts explicitly takes into account a possible high number of observed zeros which is likely to be observed in an insurance portfolio too.

In Song (2000) a large class of multivariate dispersion models is constructed by linking univariate dispersion models (e.g., Poisson, Normal, Gamma) with a Gaussian copula. These models are marginally closed, i.e., their marginals belong to the same distribution class as the multivariate model, and readily yield a flexible class to model error distributions of generalized linear models (GLM's). Based on this work, Song (2007) and Song et al. (2009) develop a multivariate analogue of univariate GLM theory with joint models for continuous, discrete, and mixed outcomes (so-called Vector GLM's). These models have the advantage of being marginally closed and thus allowing for a marginal representation of the regression coefficients.

GLM's are widely used for actuarial problems. For an overview and discussion of several applications see Haberman and Renshaw (1996). The authors, among other things, build a model for premium rating in non-life insurance using models for average claim size and claim frequency. A more detailed analysis on this issue can be found in Renshaw (1994) who considers the influence of covariates on average claim size and claim frequency. Taylor (1989) and Boskov and Verrall (1994) fit adjusted loss ratios with spline functions and a spatial Bayesian model, respectively. However, Boskov and Verrall (1994) conclude that the separate modeling of claim size and claim frequency is preferable. Based on the compound Poisson model, Jørgensen and de Souza (1994) and Smyth and Jørgensen (2002) although use a non-separate approach to model the claim rate. On the other hand, Dimakos and Frigessi (2002) model claim size and claim frequency separately, but rely on the independence assumption of the classical model by Lundberg (1903). Gschlößl and Czado (2007) relax this assumption a bit by allowing the number of claims to enter as a covariate into the model for average claim size. In order to allow for more general dependency we construct a joint regression model by linking a marginal Gamma GLM for the average claim size and a marginal Poisson GLM for the number of claims with a Gaussian copula using the mixed copula approach described above. To estimate the model parameters we develop a new algorithm based on the maximization by parts algorithm, first introduced by Song et al. (2005).

The paper is organized as follows. In Section 2 we construct the mixed copula regression model for average claim size and number of claims. Subsequently an algorithm for parameter estimation in this model is developed in Section 3. We examine this algorithm by means of a simulation study in Section 4. The application of our model to a full comprehensive car insurance data set is presented in Section 5. We compare our results to the classical independent model and finally summarize and discuss our approach.

2 Mixed copula regression model

We are interested in constructing a bivariate model, where the margins follow generalized linear regression models (GLM's). This allows to model dependence between the two components. In particular we are interested in allowing for a Poisson regression and a Gamma regression component. The Poisson regression component represents the number of claims in a group of policy holders of specified characteristics captured in exogenous factors, while the Gamma regression component models the corresponding average claim size of the group. For this we follow the mixed copula approach of Song (2007).

First we specify the marginal distributions. For this let $Y_{i1} \in \mathbb{R}^+$, $i = 1, 2, \ldots, n$, be independent continuous random variables and $Y_{i2} \in \mathbb{N}_0$, $i = 1, 2, \ldots, n$, independent count random variables. Marginally we assume the following two GLM's specified by

$$Y_{i1} \sim Gamma(\mu_{i1}, \nu^2) \text{ with } \ln(\mu_{i1}) = \boldsymbol{x}_i{}'\boldsymbol{\alpha}, \tag{2.1}$$

$$Y_{i2} \sim Poisson(\mu_{i2}) \text{ with } \ln(\mu_{i2}) = \ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}, \tag{2.2}$$

where $\boldsymbol{x}_i \in \mathbb{R}^p$ are covariates for the continuous variable Y_{i1} and $\boldsymbol{z}_i \in \mathbb{R}^q$ are covariates for the count variable Y_{i2}, respectively. In the Poisson GLM we use the offset $\ln(e_i)$, where e_i gives the known time length in which events occur. In our application this corresponds to the total time members of a policy group with specific characteristics were insured. The density of the $Gamma(\mu_{i1}, \nu^2)$ distribution is specified as

$$g_1(y_{i1}|\mu_{i1}, \nu^2) := \frac{1}{\Gamma(\frac{1}{\nu^2})} \left(\frac{1}{\mu_{i1}\nu^2}\right)^{1/\nu^2} y_{i1}^{1/\nu^2-1} e^{-\frac{1}{\mu_{i1}\nu^2} y_{i1}},$$

with $\mu_{i1} := E[Y_{i1}]$ and $Var[Y_{i1}] = \mu_{i1}^2\nu^2$. $G_1(\cdot|\mu_{i1}, \nu)$ denotes the cumulative distribution function (cdf) of Y_{i1}. Further, we assume that the parameter ν is known and does not need to be estimated in the joint regression model. In our example we choose the parameter ν as the dispersion parameter which is estimated in the marginal Gamma GLM. This means that we assume that the signal-to-noise ratio $E[Y_{i1}]/\sqrt{Var[Y_{i1}]} = 1/\nu$ is equal in both models.

The probability mass function of the $Poisson(\mu_{i2})$ distribution is denoted by

$$g_2(y_{i2}|\mu_{i2}) = \begin{cases} 0 & \text{for } y_{i1} < 0; \\ \frac{1}{y_{i2}!}\ \mu_{i2}^{y}\ e^{-\mu_{i2}} & \text{for } y_{i2} = 0, 1, 2, \dots\ , \end{cases}$$

and the corresponding cdf is given by by $G_2(\cdot|\mu_{i2})$.

To construct the joint distribution function of Y_{i1} and Y_{i2} with the two marginal regression models given in (2.1) and (2.2), we adopt a mixed copula approach. For this we choose the bivariate Gaussian copula cdf $C(\cdot, \cdot|\rho)$ as copula function because it is well investigated and directly interpretable in terms of the correlation parameter. Then, by applying Sklar's theorem (Sklar 1959), a joint distribution function for Y_{i1} and Y_{i2} can be constructed as

$$F(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = C(u_{i1}, u_{i2}|\rho) = \Phi_2\{\Phi^{-1}(u_{i1}), \Phi^{-1}(u_{i2})|\Gamma\}, \qquad (2.3)$$

where $u_{i1} := G_1(y_{i1}|\mu_{i1}, \nu^2)$ and $u_{i2} := G_2(y_{i2}|\mu_{i2})$. $\Phi(\cdot)$ denotes the (univariate) standard normal cdf, while $\Phi_2(\cdot, \cdot|\Gamma)$ is the bivariate normal cdf with covariance matrix $\Gamma := \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ and Pearson correlation ρ between the two normal scores $q_{i1} := \Phi^{-1}(u_{i1})$ and $q_{i2} := \Phi^{-1}(u_{i2})$. We like to note that the copulas in Sklar's theorem is no longer unique for discrete margins (see also Genest and Neslehova (2007)). However (2.3) still provides a valid distribution function.

Then, according to equation (6.9) in Song (2007), the joint density function of the continuous margin Y_{i1} and the discrete margin Y_{i2} is given by

$$f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = g_1(y_{i1}|\mu_{i1}, \nu^2)\left[C_1'(u_{i1}, u_{i2}|\rho) - C_1'(u_{i1}, u_{i2}^-|\rho)\right], \qquad (2.4)$$

where $C_1'(u_{i1}, u_{i2}|\rho) := \frac{\partial}{\partial u_1} C(u_1, u_{i2}|\rho)\Big|_{u_1=u_{i1}}$ and $C_1'(u_{i1}, u_{i2}^-|\rho) := \frac{\partial}{\partial u_1} C(u_1, u_{i2}^-|\rho)\Big|_{u_1=u_{i1}}$ with $u_{i2}^- := G_2(y_{i2} - 1|\mu_{i2})$. This equation can be read as

$$f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = g_1(y_{i1}|\mu_{i1}, \nu^2) f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho),$$

where $f_{Y_{i2}|Y_{i1}}(\cdot|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)$ is the conditional density of Y_{i2} given Y_{i1}. We can simplify this (see Lemma 1 in the Appendix) to

$$f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = \begin{cases} g_1(y_{i1}|\mu_{i1}, \nu^2) D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})), & \text{if } y_{i2} = 0; \\ \\ g_1(y_{i1}|\mu_{i1}, \nu^2)[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) & \\ \quad - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))], & \text{if } y_{i2} \geq 1; \end{cases} \qquad (2.5)$$

where

$$D_\rho(u_1, u_{i2}) := \Phi\left(\frac{q_{i2} - \rho q_1}{\sqrt{1 - \rho^2}}\right) = \Phi\left(\frac{\Phi^{-1}(u_{i2}) - \rho\Phi^{-1}(u_1)}{\sqrt{1 - \rho^2}}\right).$$

For determining the total claim size distribution we only consider the groups of policy holders with at least one single claim. Therefore we use the log-likelihood conditional on at least one observed (ascertained) claim as basis for our inference. Let $\boldsymbol{y} := (\boldsymbol{y}_1{}', \cdots, \boldsymbol{y}_n{}')'$ with $\boldsymbol{y}_i = (y_{i1}, y_{i2})'$ be observed pairs of Gamma-Poisson distributed response variables, where y_{i1} is the Gamma distributed margin and y_{i2} denotes the Poisson distributed margin. Further let $\boldsymbol{\theta} := (\boldsymbol{\alpha}', \boldsymbol{\beta}', \gamma)'$ be the unknown parameter vector with $\gamma \in \mathbb{R}$ being Fisher's z-transformation of ρ, i.e., $\gamma = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$. Additionally, we define the design matrices $\boldsymbol{X} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{Z} := (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, where $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ denote covariate vectors associated to y_{i1} and to y_{i2} including intercepts, respectively. Further let $\mathcal{J} := \{i | i = 1, \ldots, n; y_{i2} \geq 1\}$ be the index set of all observations with $y_{i2} \geq 1$ and $\boldsymbol{Z}_{\mathcal{J}}$ and $\boldsymbol{X}_{\mathcal{J}}$ the design matrices restricted to the set $\mathcal{J}$. Therefore the likelihood function conditional on $y_{i2} \geq 1, \forall i \in \mathcal{J}$ is given by

$$L^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}_{\mathcal{J}}, \boldsymbol{Z}_{\mathcal{J}}) = \prod_{i \in \mathcal{J}} \frac{f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho)}{[1 - g_2(0, \mu_{i2})]}, \tag{2.6}$$

and the conditional log-likelihood has the form

$$\begin{aligned} l^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}_{\mathcal{J}}, \boldsymbol{Z}_{\mathcal{J}})) &= \ln(L^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}_{\mathcal{J}}, \boldsymbol{Z}_{\mathcal{J}}) \\ = \; & - \sum_{i \in \mathcal{J}} \ln\{1 - g_2(0, \mu_{i2})\} + \sum_{i \in \mathcal{J}} \ln\{g_1(y_{i1}|\mu_{i1}, \nu^2)\} \qquad (2.7) \\ & + \sum_{i \in \mathcal{J}} \ln\{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))\}, \end{aligned}$$

with $\mu_{i1} = e^{\boldsymbol{x}_i{}'\boldsymbol{\alpha}}$ and $\mu_{i2} = e^{\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}}$. In the following we use $L^c(\boldsymbol{\theta})$ and $l^c(\boldsymbol{\theta})$ as abbreviations of the conditional likelihood (2.6) and the conditional log-likelihood (2.7), respectively.

3 Maximization by Parts Algorithm

Maximization by Parts (MBP) is a fix-point algorithm to solve a score equation for the maximum likelihood estimator (MLE) published in Song et al. (2005). For this method no second order derivatives of the full likelihood are necessary and, while standard maximum likelihood algorithms are well developed for "standard" models, this is not the case for our high-dimensional mixed copula regression model (in our application we have 68 parameters!). The good performance of MBP compared to alternative methods is shown, e.g., in Zhang et al. (2011) for Student t-copula based models and in Liu and Luger (2009) for copula-GARCH models.

The log likelihood function is decomposed into two parts, which are often quite natural, e.g., for copulas and the marginal densities. The first part of the decomposition has to be simple to maximize, i.e., it is straightforward to get the second order derivative. The second part is used to update the solution of the first part to get an efficient estimator and does not require second order derivatives. Here, we employ a variation of the MBP algorithm considered on page 1149 of Song et al. (2005).

We apply the MBP algorithm to maximize our log-likelihood (2.7) and to determine the MLE of $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \gamma)' \in \mathbb{R}^{p+q+1}$. Thus we decompose $\boldsymbol{\theta} = (\boldsymbol{\theta}_1{}', \boldsymbol{\theta}_2{}')'$

with $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}', \boldsymbol{\beta}')' \in \mathbb{R}^{p+q}$ and $\boldsymbol{\theta}_2 = \gamma \in \mathbb{R}$. This leads to a first decomposition

$$l^c(\boldsymbol{\theta}) = \ln(L^c(\boldsymbol{\theta})) = l^c_m(\boldsymbol{\theta}_1) + l^c_d(\boldsymbol{\theta}_1, \gamma) \tag{3.1}$$

for the conditional log-likelihood, where we define

$$l^c_m(\boldsymbol{\theta}_1) := \ln(L^c_m(\boldsymbol{\theta}_1)) := -\sum_{i\in\mathcal{J}} \ln(1 - e^{-\mu_{i2}}) + \sum_{i\in\mathcal{J}} \ln(g_1(y_{i1}|\mu_{i1}, \nu^2))$$

and

$$\begin{aligned} l^c_d(\boldsymbol{\theta}_1, \gamma) := \ln(L^c_d(\boldsymbol{\theta}_1, \gamma)) \;\; := \;\; & \sum_{i\in\mathcal{J}} \ln\{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu), G_2(y_{i2}|\mu_{i2})) \\ & -D_\rho(G_1(y_{i1}|\mu_{i1}, \nu), G_2(y_{i2} - 1|\mu_{i2}))\}. \end{aligned}$$

On the one hand, $l^c_m(\boldsymbol{\theta}_1)$ contains the marginal part of the conditional log-likelihood and is independent of γ, the Fisher transformation of the copula parameter ρ. On the other hand, $l^c_d(\boldsymbol{\theta}_1, \gamma)$ contains the copula part of the conditional log-likelihood and depends on γ, i.e., on the correlation parameter ρ.

For the MBP algorithm we need the score functions of $l^c_m(\boldsymbol{\theta}_1)$ and $l^c_d(\boldsymbol{\theta}_1, \gamma)$. Using the following abbreviations

$$\begin{aligned} G_{i1} &:= G_1(y_{i1}|\mu_{i1}, \nu), \qquad G_{i2} := G_2(y_{i2}|\mu_{i2}), \\ G^-_{i2} &:= G_2(y_{i2} - 1|\mu_{i2}), \qquad d_\rho(u_1, u_2) := \phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right), \end{aligned}$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution, it follows using differentiation that

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\alpha}} l^c_m(\boldsymbol{\theta}_1) &= \sum_{i\in\mathcal{J}} \frac{\partial \ln(g_1(y_{i1}|\mu_{i1}, \nu^2))}{\partial \boldsymbol{\alpha}} = \sum_{i\in\mathcal{J}} \frac{1}{g_1(y_{i1}|\mu_{i1}, \nu^2)} \frac{\partial g_1(y_{i1}|\mu_{i1}, \nu^2)}{\partial \mu_{i1}} \frac{\partial \mu_{i1}}{\partial \boldsymbol{\alpha}} \\ &= \sum_{i\in\mathcal{J}} \frac{1}{g_1(y_{i1}|\mu_{i1}, \nu^2)} \frac{1}{\mu^2_{i1}\nu^2}\, g_1(y_{i1}|\mu_{i1}, \nu^2)\, (y_{i1} - \mu_{i1})\, \mu_{i1}\, \boldsymbol{x}_i, \\ &= \frac{1}{\nu^2} \sum_{i\in\mathcal{J}} \boldsymbol{x}_i\, \mu^{-1}_{i1}(y_{i1} - \mu_{i1}). \end{aligned}$$

Similarly, we get

$$\frac{\partial}{\partial \boldsymbol{\beta}} l^c_m(\boldsymbol{\theta}_1) = -\sum_{i\in\mathcal{J}} \boldsymbol{z}_i \frac{\mu_{i2}}{e^{\mu_{i2}} - 1}.$$

As already mentioned we need second order derivatives for $l^c_m(\boldsymbol{\theta}_1)$ of the MBP algorithm. Straightforward differentiation gives

$$\begin{aligned} \mathcal{I}^c_m(\boldsymbol{\theta}_1) &:= -m^{-1} E\left[\frac{\partial^2 l^c_m(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'}\right] \\ &= m^{-1} \begin{pmatrix} \frac{1}{\nu^2}\sum_{i\in\mathcal{J}} \boldsymbol{x}_i \boldsymbol{x}_i' & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \sum_{i\in\mathcal{J}} \boldsymbol{z}_i \frac{\mu_{i2}(e^{\mu_{i2}} - 1 - \mu_{i2} e^{\mu_{i2}})}{(e^{\mu_{i2}}-1)^2} \boldsymbol{z}_i' \end{pmatrix}, \end{aligned}$$

where m is the number of elements in $\mathcal{J}$. We also need the score function of the dependency part $l_d^c(\boldsymbol{\theta}_1, \gamma)$. Thus we have to compute

$$\frac{\partial\, l_d^c(\boldsymbol{\theta}_1, \gamma)}{\partial \boldsymbol{\theta}} = \left(\tfrac{\partial}{\partial \boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1, \gamma), \tfrac{\partial}{\partial \boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1, \gamma), \tfrac{\partial}{\partial \gamma} l_d^c(\boldsymbol{\theta}_1, \gamma)\right)'.$$

These partial derivatives are given by (see Lemma 4 in the Appendix)

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1, \gamma) &= \sum_{i \in \mathcal{J}} \frac{d_\rho(G_{i1}, G_{i2}) - d_\rho(G_{i1}, G_{i2}^-)}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \frac{G_{i1}^* - G_{i1}}{\phi(\Phi^{-1}(G_{i1}))} \frac{-\rho}{\sqrt{1-\rho^2}} \boldsymbol{x}_i, \\
\frac{\partial}{\partial \boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1, \gamma) &= \sum_{i \in \mathcal{J}} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \left[d_\rho(G_{i1}, Gi2^-) \frac{g_2(y_{i2}-1|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}^-))} \right. \\
&\quad \left. - d_\rho(G_{i1}, Gi2) \frac{g_2(y_{i2}|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}))} \right] \frac{\mu_{i2}}{\sqrt{1-\rho^2}} \boldsymbol{z}_i, \\
\frac{\partial}{\partial \gamma} l_d^c(\boldsymbol{\theta}_1, \gamma) &= \sum_{i \in \mathcal{J}} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \left\{ d_\rho(G_{i1}, G_{i2})[\rho \Phi^{-1}(G_{i2}) - \Phi^{-1}(G_{i1})] \right. \\
&\quad \left. - d_\rho(G_{i1}, G_{i2}^-)[\rho \Phi^{-1}(G_{i2}^-) - \Phi^{-1}(G_{i1})] \right\} \frac{1}{\sqrt{1-\rho^2}}.
\end{aligned}$$

For the convergence of the MBP algorithm Song et al. (2005) common regularity conditions as well as information dominance (see condition (B) on page 1148 of Song et al. (2005)) are needed. Empirical evidence showed that an initial MBP algorithm based on (3.1) does not satisfy information dominance, therefore we modify our initial decomposition. For this we expand our conditional likelihood (2.6) by

$$L_w(\boldsymbol{\theta}_1) = \prod_{i \in \mathcal{J}} |\det(\Sigma_w)|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{\mu})' \Sigma_w^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}) \right\},$$

with $\boldsymbol{\mu} = (\mu_{i1}, \mu_{i2})'$ and $\Sigma_w = \left(\begin{smallmatrix} 1 & \rho_w \\ \rho_w & 1 \end{smallmatrix}\right)$. The correlation ρ_w can be pre-specified at a value estimated from a preliminary analysis of the data, but may be different from the underlying correlation. Additionally, we expand our conditional likelihood (2.6) by the likelihood of the marginal Poisson-GLM (2.2). So we get the expanded likelihood $L^*(\boldsymbol{\theta})$ and its new decomposition as

$$\begin{aligned}
L^*(\boldsymbol{\theta}) &:= L^c(\boldsymbol{\theta}) \frac{L_w(\boldsymbol{\theta}_1) \prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})}{L_w(\boldsymbol{\theta}_1) \prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})} \\
&= \underbrace{L_m^c(\boldsymbol{\theta}_1) L_w(\boldsymbol{\theta}_1) \prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})}_{=: L_m^*(\boldsymbol{\theta}_1)} \underbrace{\frac{L_d^c(\boldsymbol{\theta}_1, \gamma)}{L_w(\boldsymbol{\theta}_1) \prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})}}_{=: L_d^*(\boldsymbol{\theta}_1, \gamma)}.
\end{aligned}$$

The expanded log-likelihood $l^*(\boldsymbol{\theta})$ and its decomposition then have the form

$$l^*(\boldsymbol{\theta}) := \underbrace{\ln(L_m^*(\boldsymbol{\theta}_1))}_{=: l_m^*(\boldsymbol{\theta}1)} + \underbrace{\ln(L_d^*(\boldsymbol{\theta}_1, \gamma))}_{=: l_d^*(\boldsymbol{\theta}_1, \gamma)},$$

with

$$\begin{aligned}
l_m^*(\boldsymbol{\theta}_1) &:= l_m^c(\boldsymbol{\theta}_1) + \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})) + \ln(L_w(\boldsymbol{\theta}_1)) \\
l_d^*(\boldsymbol{\theta}1, \gamma) &:= l_d^c(\boldsymbol{\theta}_1) - \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})) - \ln(L_w(\boldsymbol{\theta}_1))
\end{aligned}$$

It is now easy to determine the corresponding first and second order derivatives of $l^*_m(\boldsymbol{\theta}_1)$ and $l^*_d(\boldsymbol{\theta}1, \gamma)$. In particular the Fisher information corresponding to $l^*_m(\boldsymbol{\theta}_1)$ is given by

$$
\begin{aligned}
\mathcal{I}^*_m(\boldsymbol{\theta}_1) &:= -m^{-1}E\left[\frac{\partial^2 l^*_m(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1{}'}\right] \\
&= \mathcal{I}^c_m(\boldsymbol{\theta}_1) + m^{-1}\begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \sum_{i\in\mathcal{J}}\mu_{i2}\boldsymbol{z}_i\boldsymbol{z}_i{}' \end{pmatrix} \\
&\quad + m^{-1}\sum_{i\in\mathcal{J}}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\Sigma_w^{-1}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}'.
\end{aligned}
$$

Now we hope that the elements of $\mathcal{I}^*_m(\boldsymbol{\theta}_1)$ are large enough to force convergence in the Fisher scoring step of the MBP algorithm. Note that

$$l^c(\boldsymbol{\theta}) = l^c_m(\boldsymbol{\theta}_1) + l^c_d(\boldsymbol{\theta}_1, \gamma) = l^*_m(\boldsymbol{\theta}_1) + l^*_d(\boldsymbol{\theta}_1, \gamma),$$

and hence

$$\frac{\partial l^c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1} = \frac{\partial l^c_m(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} + \frac{\partial l^c_d(\boldsymbol{\theta}_1, \gamma)}{\partial\boldsymbol{\theta}_1} = \frac{\partial l^*_m(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} + \frac{\partial l^*_d(\boldsymbol{\theta}_1, \gamma)}{\partial\boldsymbol{\theta}_1}.$$

Moreover, the expansions are independent of ρ or γ and therefore $\partial l^*_d(\boldsymbol{\theta}_1, \gamma)/\partial\gamma = \partial l^c_d(\boldsymbol{\theta}_1, \gamma)/\partial\gamma$, which we already derived above. The applied MBP algorithm with the expansion of the conditional log-likelihood then proceeds as follows:

Algorithm 1 (MBP algorithm for the Poisson-Gamma regression model)

Step 0 *:*

- **(i)** *The initial value for $\boldsymbol{\theta}_1$ is $\boldsymbol{\theta}_1^0 = [\hat{\boldsymbol{\alpha}}_I{}', \hat{\boldsymbol{\beta}}_I{}']'$, where $\hat{\boldsymbol{\alpha}}_I$ and $\hat{\boldsymbol{\beta}}_I$ are the MLE's of the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of independent GLM's (2.1) and (2.2).*
- **(ii)** *The initial value for γ is γ^0 the result of $\partial l^c_d(\boldsymbol{\theta}_1^0, \gamma)/\partial\gamma = 0$ using bisection.*
- **(iii)** *The pre-specified correlation ρ_w is the empirical correlation between Poisson and Gamma regression residuals determined in Step 0 (i).*

Step k (k $= 1, 2, 3, \ldots$**)** *: First, we update $\boldsymbol{\theta}_1$ by one step of Fisher scoring, i.e.,*

$$\boldsymbol{\theta}_1^k = \boldsymbol{\theta}_1^{k-1} + \{\mathcal{I}^*_m(\boldsymbol{\theta}_1^{k-1})\}^{-1}\left(\left.\frac{\partial l^c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1}\right|_{\substack{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{k-1} \\ \gamma = \gamma^{k-1}}}\right).$$

Then, by solving $\partial l^c_d(\boldsymbol{\theta}^k, \gamma)/\partial\gamma = 0$ using bisection, we obtain the new γ^k.

When the convergence criterion (e.g., $||\boldsymbol{\theta}^k - \boldsymbol{\theta}^{k-1}||_\infty < 10^{-6}$) is met, the algorithm stops and outputs an approximation of the MLE of $\boldsymbol{\theta} = [\boldsymbol{\theta}_1{}', \gamma]'$. Since γ is scalar,

$\partial l_d^c(\boldsymbol{\theta}^k, \gamma)/\partial\gamma = 0$ is a one-dimensional search and the bisection method (see, e.g., Burden and Faires (2004)) works efficiently.

Empirical experience shows that when the fix pre-specified ρ_w is not close enough to the resulting MLE of ρ, the MBP algorithm presented above does not converge. Hence, we modify the MBP algorithm further by updating ρ_w in each step. The changes in the algorithm are as follows:

in Step 0 (iii) Set $\rho_w := \frac{e^{2\gamma^0}-1}{e^{2\gamma^0}+1}$.

in Step k (k $= 1, 2, 3, \ldots$**):** Update ρ_w by setting $\rho_w^k := \frac{e^{2\gamma^k}-1}{e^{2\gamma^k}+1}$.

In the next section we run a simulation study for the MBP algorithm with pre-specified ρ_w and with the adapting ρ_w-update given above. This study shows that both versions of the MBP algorithm provide similar results, but the version with the adapting ρ_w-update has a better convergence behavior in small samples.

We close this section by providing standard error estimates for the MLE of $\boldsymbol{\theta}$. According to Theorem 3 of Song et al. (2005) the MBP algorithm provides an asymptotically normal distribution of the resulting MLE, which we can use to estimate the standard error of the MLE. Let $\hat{\boldsymbol{\theta}}$ be the resulting MLE of the $\boldsymbol{\theta}$ calculated by the MBP algorithm 1. For $k \to \infty$ $\hat{\boldsymbol{\theta}}$ has the asymptotic covariance matrix

$$m^{-1}\mathcal{I}^{-1} = m^{-1}E\left[\left.\frac{\partial^2 l^c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1},$$

where m denotes the number of elements in the index set $\mathcal{J}$. An estimator for the Fisher information matrix $\mathcal{I}$ of the conditional log-likelihood is

$$\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}) := \mathcal{I}_m^c(\hat{\boldsymbol{\theta}}) + \hat{\mathcal{I}}_d^c(\hat{\boldsymbol{\theta}}), \tag{3.2}$$

where

$$\hat{\mathcal{I}}_d(\hat{\boldsymbol{\theta}}) := m^{-1}\sum_{i\in\mathcal{J}} l_d'(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\ l_d'(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)',$$

with $l_d'(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) := \left.\frac{\partial}{\partial\boldsymbol{\theta}} l_d(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

The estimated standard error for $\hat{\boldsymbol{\theta}}$ is then the square root of the diagonal elements of the matrix $m^{-1}\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}})^{-1}$.

4 Simulation study

In this section we study the small sample properties of the MLE's in the Poisson-Gamma regression model determined by the proposed MBP algorithms, one with a fixed choice of ρ_w and one with an adaptive choice of ρ_w. We assume the constant of variation ν in the marginal Gamma regression as known. Several values of ν are studied. Overall 24 scenarios are investigated with a sample size of $N = 1000$ for the Poisson-Gamma pairs. To estimate bias and mean squared error we performed 500 repetitions. For both marginal regression models we specify a single covariate and allow for an intercept. Covariate values are chosen as i.i.d uniform(0,1) realizations and remain fixed for all scenarios and repetitions, i.e., we have

$$\mu_{i1} = \exp(\alpha_1 + x_{i2}\alpha_2) \text{ and } \mu_{i2} = \exp(\beta_1 + z_{i2}\beta_1),$$

with $x_{i2} \in (0,1)$ and $z_{i2} \in (0,1)$ for all i. For the regression parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ of the marginal Gamma GLM we consider the values $(1,1)'$ or $(1,3)'$ so that $\mu_{i1} \in (2.72, 7.39)$ or $\mu_{i1} \in (2.72, 54.60)$. For the regression parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ we choose the values $(-1,3)$ or $(-0.5,3)'$ so that $\mu_{i2} \in (0.37, 7.39)$ or $\mu_{i2} \in (0.61, 12.18)$. For the correlation parameter ρ of the Gaussian copula we consider 0.1 for a small, 0.5 for a medium and 0.9 for a high correlation. The values of the constant coefficients of variation of the Gamma distribution ν are chosen in such a way that the signal-to-noise ratio

$$snr := \frac{E[Y_{i1}]}{\sqrt{Var[Y_{i1}]}} = \frac{\mu_{i1}}{\mu_{i1}\nu} = \frac{1}{\nu}$$

is 1 or 2, i.e., we set $\nu = 0.5$ or $\nu = 1$. The chosen parameter combinations are given in Table 1.

For each scenario we simulate correlated Poisson-Gamma regression responses as follows: To generate a pair (y_{i1}, y_{i2}) of a marginally $Gamma(\mu_{i1}, \nu)$ distributed random variable Y_{i1} and a with ρ correlated marginally $Poisson(\mu_{i2})$ distributed random variable Y_{i2} we use the conditional probability mass function of the Poisson variable Y_{i2} given the Gamma variable Y_{i1}. The joint density function Y_{i1} and Y_{i2} is given in equation (2.5). Therefore the conditional probability mass of Y_{i2} given Y_{i1} is given as

$$f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho) := \frac{f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2})}{g_1(y_{i1}|\mu_{i1}, \nu)} \tag{4.1}$$

$$= \begin{cases} \frac{g_1(y_{i1}|\mu_{i1},\nu) D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}|\mu_{i2}))}{g_1(y_{i1}|\mu_{i1},\nu)}, & \text{if } y_{i2} = 0; \\ \frac{g_1(y_{i1}|\mu_{i1},\nu)[D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}-1|\mu_{i2}))]}{g_1(y_{i1}|\mu_{i1},\nu)}, & \text{if } y_{i2} \geq 1 \end{cases}$$

$$= \begin{cases} D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}|\mu_{i2})), & \text{if } y_{i2} = 0; \\ D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1},\nu), G_2(y_{i2}-1|\mu_{i2})), & \text{if } y_{i2} \geq 1. \end{cases}$$

The algorithm to generate a $Gamma(\mu_{i1}, \nu)$ observation y_{i1} and a $Poisson(\mu_{i2})$ observation y_{i2} with correlation ρ then proceeds as follows:

Algorithm 2 (Generation of Correlated Gamma and Poisson Random Variables)

Step 1: *Sample* y_{i1} *from a* $Gamma(\mu_{i1}, \nu)$ *distribution.*

Step 2: *Calculate* $p_k = f_{Y_{i2}|Y_{i1}}(y_{i2} = k|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)$ *for* $k = 0, 1, \ldots, k^*$, *where* $p_{k^*} \geq \varepsilon$ *and* $p_{k^*+1} < \varepsilon$, $\varepsilon \in (0,1)$.

Step 3: *Sample* y_{i2} *from* $\{0, 1, \ldots, k^*\}$ *with* $P(Y_{i2} = k) = p_k$ *for* $k \in \{0, 1, \ldots, k^*\}$.

Scenario	Parameters α_1	α_2	β_1	β_2	ρ	ν
1	1.0	1.0	−1.0	3.0	0.1	0.5
2	1.0	1.0	−1.0	3.0	0.1	1.0
3	1.0	1.0	−0.5	3.0	0.1	0.5
4	1.0	1.0	−0.5	3.0	0.1	1.0
5	1.0	3.0	−1.0	3.0	0.1	0.5
6	1.0	3.0	−1.0	3.0	0.1	1.0
7	1.0	3.0	−0.5	3.0	0.1	0.5
8	1.0	3.0	−0.5	3.0	0.1	1.0
9	1.0	1.0	−1.0	3.0	0.5	0.5
10	1.0	1.0	−1.0	3.0	0.5	1.0
11	1.0	1.0	−0.5	3.0	0.5	0.5
12	1.0	1.0	−0.5	3.0	0.5	1.0
13	1.0	3.0	−1.0	3.0	0.5	0.5
14	1.0	3.0	−1.0	3.0	0.5	1.0
15	1.0	3.0	−0.5	3.0	0.5	0.5
16	1.0	3.0	−0.5	3.0	0.5	1.0
17	1.0	1.0	−1.0	3.0	0.9	0.5
18	1.0	1.0	−1.0	3.0	0.9	1.0
19	1.0	1.0	−0.5	3.0	0.9	0.5
20	1.0	1.0	−0.5	3.0	0.9	1.0
21	1.0	3.0	−1.0	3.0	0.9	0.5
22	1.0	3.0	−1.0	3.0	0.9	1.0
23	1.0	3.0	−0.5	3.0	0.9	0.5
24	1.0	3.0	−0.5	3.0	0.9	1.0

Table 1: Chosen parameter settings for 24 different scenarios studied.

The value of ε determines where we neglect the tail of the conditional distribution. Pairs which result in a zero Poisson count were removed.

In the MBP algorithms we choose as stopping criterion $||\boldsymbol{\theta}_1^k - \boldsymbol{\theta}_1^{k-1}|| < 10^{-3}$, where $\boldsymbol{\theta}_1^k = (\alpha_1^k, \alpha_2^k, \beta_1^k, \beta_2^k)'$ are the regression parameter after the k-th iteration, and $||\rho^k - \rho^{k-1}|| < 10^{-4}$, where ρ^k denotes the correlation parameter after the k-th iteration. In the version of the MBP algorithm with fixed ρ_w we set ρ_w equal to the empirical correlation between the residuals of the marginal Gamma GLM and the marginal Poisson GLM.

We generated 500 data sets for each scenario and calculated the relative bias for all parameters as well as the maximum relative bias for each scenario. The bias results are summarized in Table 2 assuming a fixed ρ_w and in Table 3 for the adaptive ρ_w-update.

These bias results show a satisfactory small sample behavior of the MBP algorithms. In particular 13 (15) scenarios using the MBP algorithm with pre-specified

Scenario	Relative bias in % α_1	α_2	β_1	β_2	ρ	ν	max
1	1.17	−0.34	0.10	0.14	−6.62	−0.18	6.62
2	−0.19	0.59	−0.20	−0.07	−0.92	−0.20	0.92
3	0.65	−0.12	0.47	0.15	−4.01	−0.30	4.01
4	0.46	−0.95	−0.16	−0.03	−0.37	−0.17	0.95
5	1.28	−0.47	−0.39	−0.13	−4.83	−0.07	4.83
6	−0.20	−0.02	0.18	0.05	−1.68	−0.06	1.68
7	0.68	−0.29	2.05	0.41	−3.11	−0.04	3.11
8	−0.38	−0.04	0.61	0.10	1.28	−0.27	1.28
9	5.85	−1.42	−8.54	−2.46	−1.49	−0.08	8.54
10	0.38	−0.80	0.10	0.04	0.10	−0.05	0.80
11	3.67	−1.37	−6.87	−0.92	−0.40	0.00	6.87
12	−0.26	0.06	0.38	0.02	0.23	−0.03	0.38
13	6.04	−2.12	−5.11	−1.45	−0.89	−0.20	6.04
14	0.06	0.08	−0.13	−0.01	0.09	−0.29	0.29
15	3.34	−1.24	−4.19	−0.58	−0.45	−0.05	4.19
16	0.64	−0.38	−0.54	−0.11	0.04	0.05	0.64
17	5.88	−0.54	−14.08	−3.77	0.26	−0.11	14.08
18	−9.64	1.19	11.43	3.24	−0.44	−0.10	11.43
19	−0.89	−0.04	3.90	0.54	0.03	0.15	3.90
20	−15.54	0.27	33.99	5.36	−0.68	−0.15	33.99
21	5.68	−1.83	−7.32	−1.94	0.11	−0.07	7.32
22	−5.50	1.56	2.01	0.42	−0.27	−0.23	5.50
23	−0.35	0.23	−0.45	−0.12	0.03	0.13	0.45
24	−8.91	3.19	9.39	1.47	−0.23	−0.17	9.39

Table 2: Relative bias and maximal absolute relative bias per scenario of the MLE's determined by the MBP algorithm with fixed ρ_w for the parameters α_1, α_2, β_1, β_2, ρ and ν over 24 scenarios.

	Relative bias in %						
Scenario	α_1	α_2	β_1	β_2	ρ	ν	*max*
1	1.18	−0.17	−0.66	−0.25	−6.66	0.11	6.66
2	−0.71	1.22	−0.22	−0.08	1.72	−0.14	1.72
3	0.84	−0.40	1.09	0.23	−0.71	0.08	1.09
4	0.14	−0.09	0.06	−0.03	−1.17	−0.05	1.17
5	1.30	−0.43	−0.55	−0.22	−3.84	0.11	3.84
6	−0.66	0.45	−0.26	−0.09	1.47	−0.14	1.47
7	0.73	−0.28	1.21	0.25	0.60	0.08	1.21
8	0.22	−0.06	0.03	−0.03	−1.29	−0.05	1.29
9	5.92	−1.43	−8.73	−2.49	−1.56	−0.04	8.73
10	−0.16	0.35	0.44	0.18	−0.14	−0.11	0.44
11	3.91	−1.75	−8.06	−1.22	−0.49	0.13	8.06
12	−0.10	−0.41	0.74	0.15	−0.14	−0.22	0.74
13	6.17	−2.14	−5.23	−1.45	−1.07	−0.04	6.17
14	0.04	−0.00	0.41	0.17	−0.19	−0.11	0.41
15	3.39	−1.24	−4.54	−0.66	−0.38	0.13	4.54
16	−0.02	−0.17	0.69	0.14	−0.20	−0.22	0.69
17	6.16	0.41	−15.74	−4.24	0.30	0.02	15.74
18	−6.44	3.65	6.79	1.95	−0.22	−0.15	6.79
19	−1.04	2.44	0.08	0.05	−0.00	−0.29	2.44
20	−13.27	4.30	25.07	3.87	−0.49	−0.16	25.07
21	6.77	−2.13	−9.35	−2.59	0.07	0.02	9.35
22	1.38	−0.19	−1.34	−0.44	−0.00	−0.15	1.38
23	3.29	−0.47	−9.86	−1.45	0.07	−0.29	9.86
24	−0.09	0.59	−1.33	−0.30	−0.02	−0.16	1.33

Table 3: Relative bias and maximal absolute relative bias per scenario of the MLE's determined by the MBP algorithm with adaptive ρ_w-update for the parameters α_1, α_2, β_1, β_2, ρ and ν over 24 scenarios.

average claim size					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.01	2374.23	4195.27	5755.01	7272.75	49339.06

	number of claims			
	1	2	3	4
#	12472	356	20	2
%	97.06	2.77	0.16	0.02

Table 4: Summary statistics of the responses in the original data set.

ρ_w (with adaptive ρ_w) show a maximum relative bias less than 5 %. A maximum relative bias less than 10% is observed in 21 (22) scenarios for the MBP algorithm with pre-specified ρ_w (with adaptive ρ_w). A relative bias larger than 10% is only observed in scenarios with extreme correlation of $\rho = 0.9$ and small marginal Gamma means.

More detailed results of the simulation study such as mean squared error estimates can be found in Kastenmeier (2008) and show that especially in scenarios with high correlated data, the adaptive ρ_w-update seems to improve the resulting MLE's. For medium to low correlated data the performances of the two versions of the algorithm are almost identical.

5 Application

5.1 Data

The data set contains information on full comprehensive car insurance policies in Germany in the year 2000. Each observation consists of the cumulative and average claim size, the number of claims (which is at least one, as only those policies are recorded) and the exposure time (since not all policies are valid for the whole year) as well as several covariates such as type and age of the car, distance driven per year, age and gender of the policyholder, the claim free years and the deductible. A subset of this data set containing three types of midsized cars with a total of 12'850 policies is analyzed here. We are interested in the joint distribution of the number of claims and the average claim size of a policy allowing for dependency between both in order to estimate the expected total loss.

The average claim size of a policy is given in the currency of DM[2]. It is calculated as the sum of the claim sizes of each policy claim divided by the number of claims. The histogram of the observed average claim sizes in Figure 1 (left panel) shows a right-skewed shape. The mean of the individual claim size (5'755.01 DM) is greater than the median, but smaller than the third quartile (cp. Table 4), so that the right-skewness is not extreme. As the largest observed claim size is only about 0.07% of the sum of all individual claim sizes, the data set does not contain extreme values either. Therefore, there is no need to use a heavy tailed distribution and so a Gamma model is an appropriate choice.

The range of the observed number of claims is very small and about 97 % of the observed policies have one claim only (cp. Table 4). The maximum number

[2]'Deutsche Mark' (DM) is the former German currency, which was replaced by the Euro in 2002 (1 DM $\hat{=}$ 0.51 Euro). Here we also use abbreviation TDM for 1'000 DM.

	number of claims													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
#	5318	1249	462	238	131	63	61	39	32	13	17	8	3	2
%	69.40	16.30	6.03	3.11	1.71	0.82	0.80	0.51	0.42	0.17	0.22	0.10	0.04	0.03
	15	16	17	18	19	20	21	22	26	27	30	31	50	-
#	5	1	6	1	3	2	2	2	1	1	1	1	1	-
%	0.07	0.01	0.08	0.01	0.04	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0.01	-

Table 5: Absolute and relative frequencies of the occurring values of number of claims in the aggregated data set.

of observed claims (four) occurs only twice. This extreme right-skewness results in the fact that the mean is not only greater than the median, but also greater than the third quartile. As we want to model the number of claims by a Poisson GLM, we need a wider range of the observed number of claims and more observations unequal to one in order to get good estimates for the model parameters. This can be achieved by aggregating the policies according to categorical covariates and summing up the number of claims of the policies in the same cell. The covariates we use for the data aggregation are: age of policyholder, regional class, driven distance per year, construction year of the car, deductible and premium rate. The resulting aggregated data set contains 7'663 policy groups. Details regarding the aggregation process and the covariate categories can be found in Kastenmeier (2008)[3]. As we intended, the range of the number of claims is now wider than in the original data set. The absolute and relative frequencies of the observed number of claims are given in Table 5. We like to note that the covariates chosen for aggregation are often used for pricing car insurance policies.

(Figure 1 about here.)

In Figure 1 (right panel) we have the histogram of average claim sizes of the policy groups. The shape of the histogram is quite similar to the one of the histogram of average claim sizes of each policy, given in the left panel of Figure 1. Hence the assumption of a Gamma distribution for the average claim sizes is still appropriate. By summing up the product of the average claim size and the number of claims of the policy groups we obtain the total loss which is 76'071 TDM. Of course, the total loss of the aggregated data set is equal to the one of the individual policies in the original data set.

The plot of the number of claims against the average claim size of the policy groups and the corresponding regression line in Figure 2 shows that the regression line has a small positive slope. This is an indication for a positive correlation between the number of claims and the average claim size. We also checked numerous alternative aggregations to ensure that the positive slope is not due to the specific chosen aggregation.

[3]Note that the data aggregation in this paper is slightly different from the one performed in Kastenmeier (2008). In order to obtain more evenly spread age groups, we merged the youngest and the oldest two of the age groups in the above thesis, respectively.

(Figure 2 about here.)

In the following we use the new (aggregated) data set of the policy groups with the objective to estimate the parameter of the joint distribution function of the number of claims and the average claim size of the policy groups.

5.2 Model Selection

We now apply the mixed copula regression model of Section 2 to the average claim size and the number of claims given in the aggregated data set described above. To calculate the MLEs of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ρ with Algorithm 1 we need initial values $\boldsymbol{\alpha}_I$ and $\boldsymbol{\beta}_I$. As $\boldsymbol{\alpha}_I$ we take the MLE of the regression parameter in the marginal Gamma GLM (2.1) applied to the average claim size. The MLE of the regression parameter in the Poisson GLM (2.2) applied to the number of claims is, however, not a good initial value $\boldsymbol{\beta}_I$ for the MBP Algorithm, because the data set contains only policy groups with at least one claim. Hence we use a zero-truncated Poisson GLM (see Winkelmann (2008)), a GLM using the Poisson distribution conditional on that the count variable is greater or equal to one, to model the number of claims and take the resulting MLE of the regression parameter in this GLM as the initial value $\boldsymbol{\beta}_I$. The probability mass of the Zero-truncated Poisson distribution is given by

$$g_{2|Y>0}(y|\lambda) := P(Y = y|Y > 0) = \frac{P(Y = y)}{P(Y > 0)} = \frac{\lambda^y \; e^{-\lambda}}{y! \; (1 - e^{-\lambda})},$$

with mean $E[Y|Y > 0] = \lambda/(1 - e^{-\lambda})$ and variance $Var[Y|Y > 0] = E[Y|Y > 0](1-e^{-\lambda}-\lambda e^{-\lambda})/(1-e^{-\lambda})$. Let $y_i \in \mathbb{N}$, $i = 1, \dots, n$, with $y_i > 0$ be the observations of Zero-truncated Poisson(λ_i)-distributed random variables Y_i and let $\boldsymbol{z}_i \in \mathbb{R}^p$ be the corresponding covariate vectors. The Zero-truncated Poisson GLM is then specified by

$$Y_i \;\sim\; Zero - truncated\ Poisson(\lambda_i) \text{ with } \ln(\lambda_i) = \ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}, \tag{5.1}$$

with regression parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ and offset $\ln(e_i)$, where e_i is the exposure of Y_i. The mean of Y_i is then calculated as

$$\mu_i := E[Y_i|Y_i > 0] = \frac{\lambda_i}{1 - e^{-\lambda_i}} = \frac{e^{\ln(e_i)+\boldsymbol{z}_i{}'\boldsymbol{\beta}}}{1 - \exp(-e^{\ln(e_i)+\boldsymbol{z}_i{}'\boldsymbol{\beta}})}. \tag{5.2}$$

The corresponding log-likelihood function is

$$l(\beta|\boldsymbol{y}, Z) := \sum_{i=1}^{n} y_i \ln(\lambda_i) - \lambda_i - \ln(1 - e^{-\lambda_i}) + \ln(y_i!),$$

with $\boldsymbol{y} = (y_1, \dots y_n)$ and design matrix $Z = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)'$, where $\lambda_i = \exp(\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta})$. Details about the construction of the Zero-truncated Poisson GLM are given in Kastenmeier (2008).

To construct the joint Gamma-Poisson regression model we dummy code the covariates. The dummy variables of the covariates categories are marked with the prefix d.. The categories of each covariate and the corresponding dummy variable can be found in Kastenmeier (2008). Note that there is one dummy variable less for each covariate than there are covariate categories. In the model selection we consider only those covariates for which there is at least one corresponding dummy variable which is significant at the 5%-level (which is equivalent to a t-value greater than 1.96 or a p-value smaller than 0.05). As stopping criterion for the MBP algorithm we choose $||\boldsymbol{\theta}_1^k - \boldsymbol{\theta}_1^{k-1}|| < 10^{-3}$, where $\boldsymbol{\theta}_1^k = (\boldsymbol{\alpha}^{k\prime}, \boldsymbol{\beta}^{k\prime})'$ are the regression parameters after the k-th iteration, and $||\rho^k - \rho^{k-1}|| < 10^{-4}$, where ρ^k denotes the correlation parameter after the k-th iteration. Additionally, we estimate the standard error for each parameter estimate using the expression for the asymptotic covariance matrix provided in (3.2).

The dummy covariates of the final model with their corresponding estimated regression parameters, standard deviations and p-values can be found in Tables 6 and 7. Table 8 displays the results for the copula correlation parameter which is highly significant. Its value is 0.1366 which shows that there is a small, but significant, dependency between the average claim size and the number of claims of a policy group, as we suspected from Figure 2. A possible reason of this positive correlation is that the deductible is used only for the first claim, while further claims are fully presented. Furthermore, if someone has more than one claim per year, he possibly exhibits a riskier driving behavior and is therefore also more likely to produce higher average claims.

To compare our model with the classical independent model, we also perform the covariate selection for the marginal Gamma GLM (2.1) and the marginal Poisson GLM (5.1). The resulting models with their corresponding estimated regression parameters, standard deviations and p-values can be found in the right parts of Tables 6 and 7 (in the independent Poisson GLM 'age' is only marginally significant but kept in the model for better comparability). The dispersion parameter ν is estimated in the marginal Gamma GLM as 0.5904 and also used in the mixed copula regression model, meaning an equal signal-to-noise ratio in both models.

5.3 Model Evaluation and Total Loss Estimation

In order to evaluate our model and to estimate the expected total loss of the insurance portfolio, we partition our data set into 25 risk groups by the 20%-quantiles of the expected average claim size and the expected number of claims (see Table 9) in the final mixed copula model, respectively, i.e., we classify each observation with respect to its expected claim size and frequency. Then we use Monte-Carlo Estimators (MCE's) to estimate the expected total loss in each risk group. For this we generate $R = 500$ data sets with the size of the original data set using the results of the regression analysis in Section 5.2. For each of the 500 generated data sets we can calculate the total loss in risk group (j, k) with j denoting the level of expected claim frequency and k the level of expected average claim size, i.e.,

$$S^{r(j,k)} = \sum_{\substack{i=1 \\ i \text{ in risk group } (j,k)}}^{7663} Y_{i1}^r \, Y_{i2}^r, \; r = 1, \ldots, R, \; j = 1, \ldots, 5, \; k = 1, \ldots, 5,$$

	Mixed Copula Model			Independent Model		
	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Intercept	1.1357	0.0710	0.0000	1.2244	0.0710	0.0000
d.rcl1	0.0144	0.0440	0.7446	0.0032	0.0440	0.9416
d.rcl2	-0.0132	0.0402	0.7432	-0.0300	0.0402	0.4550
d.rcl3	0.0491	0.0388	0.2048	0.0278	0.0388	0.4729
d.rcl4	0.1520	0.0405	0.0002	0.1343	0.0405	0.0009
d.rcl5	0.1251	0.0431	0.0037	0.1123	0.0431	0.0092
d.rcl6	0.1118	0.0406	0.0059	0.0951	0.0406	0.0193
d.rcl7	0.1646	0.0533	0.0020	0.1677	0.0533	0.0017
d.premrate1	0.1268	0.0232	0.0000	0.1329	0.0232	0.0000
d.premrate2	0.0804	0.0308	0.0089	0.1093	0.0308	0.0004
d.premrate3	0.1656	0.0330	0.0000	0.1980	0.0330	0.0000
d.premrate4	0.2505	0.0342	0.0000	0.2820	0.0342	0.0000
d.premrate5	0.2903	0.0428	0.0000	0.3287	0.0428	0.0000
d.premrate6	0.3961	0.0663	0.0000	0.4402	0.0663	0.0000
d.deductible1	0.2169	0.0574	0.0002	0.2076	0.0574	0.0003
d.deductible2	0.2946	0.0506	0.0000	0.2562	0.0506	0.0000
d.deductible3	0.3560	0.0532	0.0000	0.3410	0.0532	0.0000
d.deductible4	0.3498	0.0859	0.0000	0.3776	0.0859	0.0000
d.drivdist1	0.0292	0.0256	0.2552	0.0301	0.0256	0.2411
d.drivdist2	0.0297	0.0302	0.3245	0.0390	0.0302	0.1958
d.drivdist3	0.0577	0.0253	0.0228	0.0564	0.0253	0.0261
d.drivdist4	0.0509	0.0287	0.0755	0.0545	0.0286	0.0570
d.age1	-0.0518	0.0360	0.1501	-0.0486	0.0360	0.1777
d.age2	-0.0744	0.0352	0.0347	-0.0704	0.0352	0.0457
d.age3	-0.0349	0.0322	0.2781	-0.0383	0.0322	0.2341
d.age4	-0.0135	0.0342	0.6926	-0.0181	0.0342	0.5961
d.age5	-0.0916	0.0367	0.0126	-0.0968	0.0367	0.0084
d.constyear1	0.0768	0.0366	0.0359	0.0775	0.0366	0.0341
d.constyear2	0.1078	0.0339	0.0015	0.1067	0.0339	0.0016
d.constyear3	0.1151	0.0327	0.0004	0.1120	0.0327	0.0006
d.constyear4	0.1272	0.0312	0.0000	0.1174	0.0312	0.0002
d.constyear5	0.1814	0.0314	0.0000	0.1713	0.0314	0.0000
d.constyear6	0.2140	0.0378	0.0000	0.2144	0.0378	0.0000
d.sex	—	—	—	—	—	—

Table 6: Summary for the Gamma regression parameter $\boldsymbol{\alpha}$ of the mixed copula regression model and the independent Gamma GLM.

	Mixed Copula Model			Independent Model		
	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Intercept	-7.2503	0.1161	0.0000	-7.3470	0.1900	0.0000
d.rcl1	0.2193	0.0747	0.0033	0.2338	0.0893	0.0089
d.rcl2	0.3970	0.0667	0.0000	0.4217	0.0805	0.0000
d.rcl3	0.4931	0.0618	0.0000	0.5185	0.0778	0.0000
d.rcl4	0.4230	0.0668	0.0000	0.4420	0.0806	0.0000
d.rcl5	0.3092	0.0739	0.0000	0.3254	0.0856	0.0001
d.rcl6	0.3992	0.0676	0.0000	0.4145	0.0811	0.0000
d.rcl7	-0.1357	0.0865	0.1166	-0.1772	0.1331	0.1832
d.premrate1	-0.1149	0.0440	0.0090	-0.1167	0.0309	0.0002
d.premrate2	-0.6478	0.0522	0.0000	-0.6719	0.0715	0.0000
d.premrate3	-0.6593	0.0551	0.0000	-0.7077	0.0818	0.0000
d.premrate4	-0.5707	0.0565	0.0000	-0.6082	0.0837	0.0000
d.premrate5	-0.7400	0.0664	0.0000	-0.7889	0.1287	0.0000
d.premrate6	-0.8177	0.0977	0.0000	-0.9075	0.2370	0.0001
d.deductible1	0.2503	0.0884	0.0046	0.2653	0.1719	0.1227
d.deductible2	0.9658	0.0773	0.0000	1.0390	0.1537	0.0000
d.deductible3	0.3459	0.0821	0.0000	0.3869	0.1617	0.0167
d.deductible4	-1.1229	0.1223	0.0000	-1.3237	0.5931	0.0256
d.drivdist1	-0.0799	0.0416	0.0550	-0.0828	0.0345	0.0164
d.drivdist2	-0.3224	0.0523	0.0000	-0.3354	0.0506	0.0000
d.drivdist3	-0.0295	0.0427	0.4896	-0.0317	0.0347	0.3610
d.drivdist4	-0.1551	0.0499	0.0019	-0.1692	0.0455	0.0002
d.age1	-0.0878	0.0583	0.1321	-0.0860	0.0832	0.3014
d.age2	-0.1587	0.0589	0.0071	-0.1506	0.0795	0.0582
d.age3	0.0404	0.0524	0.4409	0.0580	0.0739	0.4321
d.age4	-0.0140	0.0572	0.8062	-0.0013	0.0753	0.9857
d.age5	-0.0535	0.0626	0.3928	-0.0405	0.0765	0.5961
d.constyear1	-0.1198	0.0615	0.0513	-0.1330	0.0593	0.0248
d.constyear2	-0.0626	0.0579	0.2795	-0.0733	0.0517	0.1563
d.constyear3	-0.0193	0.0550	0.7255	-0.0296	0.0494	0.5482
d.constyear4	0.0420	0.0516	0.4160	0.0301	0.0439	0.4922
d.constyear5	0.0367	0.0514	0.4755	0.0204	0.0442	0.6440
d.constyear6	0.1354	0.0638	0.0338	0.1061	0.0611	0.0824
d.sex	0.2960	0.0323	0.0000	0.3144	0.0336	0.0000

Table 7: Summary for the Poisson regression parameter $\boldsymbol{\beta}$ of of the mixed copula regression model and the independent Zero-Truncated Poisson GLM.

	Estimate	Std. Error	p-value
ρ	0.1366	0.0094	0.0000

Table 8: Summary for the correlation parameter ρ of the mixed copula regression model.

Variable	Min.	20%	40%	60%	80%	Max.
average claim size	2936.71	4887.85	5349.11	5780.42	6330.51	9764.60
number of claims	1.00	1.11	1.21	1.34	1.68	60.04

Table 9: Quantiles of the expected average claim size and the expected number of claims in the final mixed copula model.

where Y_{i1}^r and Y_{i2}^r are the average claims size and the number of claims of the i-th policy group in the r-th generated data set respectively. The MCE $\hat{S}^{(j,k)}$ of the expected total loss of risk group (j,k) is then calculated by

$$\hat{S}^{(j,k)} = \frac{1}{R}\sum_{r=1}^{R} S^{r(j,k)}.$$

To generate the data sets we use a sampling algorithm similar to Algorithm 2. But, as our insurance portfolio contains only policy groups with at least one claim, we have to sample from the joint Gamma-Poisson distribution with density (2.5) conditional on the Poisson variate being greater than 0. For this, we set $p_k = f_{Y_{i2}|Y_{i1},Y_{i2}\geq 1}(y_{i2} = k|y_{i1},\mu_{i1},\nu,\mu_{i2},\rho)$ for $k = 1,2,\ldots,k^*$, in Algorithm 2, where

$$\begin{aligned} f_{Y_{i2}|Y_{i1},Y_{i2}\geq 1}(y_{i2}|y_{i1},\mu_{i1},\nu,\mu_{i2},\rho) &:= \frac{f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1},\mu_{i1},\nu,\mu_{i2},\rho)}{1 - f_{Y_{i2}|Y_{i1}}(0|y_{i1},\mu_{i1},\nu,\mu_{i2},\rho)} \\ &= \frac{D_\rho(G_1(y_{i1}|\mu_{i1},\nu),G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1},\nu),G_2(y_{i2}-1|\mu_{i2}))}{1 - D_\rho(G_1(y_{i1}|\mu_{i1},\nu),G_2(0|\mu_{i2}))}, \end{aligned}$$

with $y_{i2} \geq 1$ and sample y_{i2} from $\{1,2,\ldots,k^*\}$ with $P(Y_{i2} = k) = p_k$ for $k \in 1,2,\ldots,k^*$. The density $f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1},\mu_{i1},\nu,\mu_{i2},\rho)$ is given in Equation (4.1). The parameter setting for the data simulation is the following:

$$\begin{aligned} \mu_{i1} &= e^{\boldsymbol{x}_i'\hat{\boldsymbol{\alpha}}}, \qquad \nu = \hat{\nu}, \\ \mu_{i2} &= e^{\ln(e_i)+\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}}, \qquad \rho = \hat{\rho}, \end{aligned}$$

where $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$ are the MLE's of the parameters in the mixed copula regression model. For comparison reasons, we perform the same simulation using the results of the independent GLM's with the following parameter setting for the data generation:

$$\begin{aligned} \mu_{i1} &= e^{\boldsymbol{x}_i'\hat{\boldsymbol{\alpha}}_{ind}}, \qquad \nu = \hat{\nu}, \\ \mu_{i2} &= e^{\ln(e_i)+\boldsymbol{z}_i'\hat{\boldsymbol{\beta}}_{ind}}, \qquad \rho = 0, \end{aligned}$$

where $\hat{\boldsymbol{\alpha}}_{ind}$, $\hat{\boldsymbol{\beta}}_{ind}$ are the MLE's of the parameter regression parameter in the independent Gamma GLM and the independent Zero-truncated Poisson GLM (5.1). So we get the total loss $S_{ind}^{r(j,k)}$, $r = 1, 2, \ldots, R$, of the simulated data sets with independent claim frequency and claim size for each risk group (j, k). The corresponding MCE of the expected total loss is then

$$\hat{S}_{ind}^{(j,k)} = \frac{1}{R} \sum_{r=1}^{R} S_{ind}^{r(j,k)}.$$

(Figure 3 about here.)

We can now compare the results of our joint regression model in the following way: first, for the mixed copula model we compute absolute deviations of the simulated total losses from the observed total losses in each risk group weighted by the exposure of the respective group (left panel in Figure 3). As the deviations of the classical independent model are of the same order of magnitude (not displayed here), we compare the deviations of both models directly (right panel in Figure 3: light colors indicate risk groups in which the joint regression model performs better). The plots show that the joint model performs very strongly except for those risk groups with very small expected number of claims and the case when the expected number of claims is very large and the expected average claim size is small. Especially in the latter risk group modeling is unsatisfactory because this risk group makes up 16% of the total exposure, while those risk groups with very small expected number of claims contribute only 5% of the total exposure. This indicates that the choice of the Gaussian copula may not have been the best as it does not allow an asymmetric tail behavior which, apparently, would be appropriate here. However, compared to the independent model the results of the joint regression model are more accurate in 17 of 25 risk groups corresponding to 73% of the total exposure and the average weighted deviations are smaller as well (20.63 vs. 21.42). Standard errors of the mixed copula model lie between 63.97 and 571.70, whereas those of the independent model are in the range from 63.42 to 529.40.

Naturally, the expected total loss of the full comprehensive car insurance portfolio can be estimated as well by summing up the simulated total losses of each risk group:

$$\hat{S} = \sum_{j,k=1}^{5} \hat{S}^{(j,k)},$$

and similary for $\hat{S}_{ind}$. The MCE of the expected total loss using the estimated distribution parameters of the mixed copula regression model provides $\hat{S} = 74'109$ TDM with a standard error of 1143.46 TDM. The estimated expected total loss in the classical independent model (using the estimated regression parameters of the independent Gamma GLM and the independent Zero-truncated Poisson GLM) is $\hat{S}_{ind} = 75'774$ TDM with a standard error of 1041.75 TDM. The total loss of the observed car insurance portfolio has the amount of 76'071 TDM and is about 2.6% higher than $\hat{S}$ and about 0.4% higher than $\hat{S}_{ind}$ (cp. Figure 4).

(Figure 4 about here.)

In the classical independent model, we can also estimate the expected total loss without using a Monte-Carlo estimate. As we assume independency between the number of claims and the average claim size, the theoretical expected total loss is easy to calculate:

$$E[S_{ind}] = E[\sum_{i=1}^{7663} Y_{i1}Y_{i2}] = \sum_{i=1}^{7663} E[Y_{i1}]E[Y_{i2}],$$

with $E[Y_{i1}] = \mu_{i1} = e^{\boldsymbol{x}_i{}'\boldsymbol{\alpha}}$ and $E[Y_{i2}] = \frac{\mu_{i2}}{1-\exp(-\mu_{i2})}$ where $\mu_{i2} = e^{\ln(e_i)+\boldsymbol{z}_i{}'\boldsymbol{\beta}}$ (cp. (5.2)). So an estimator for the expected total loss is given by

$$\widehat{E[S_{ind}]} := \sum_{i=1}^{7663} \hat{\mu}_{i1} \frac{\hat{\mu}_{i2}}{1-\exp(-\hat{\mu}_{i2})},$$

where $\hat{\mu}_{i1} = e^{\boldsymbol{x}_i{}'\hat{\boldsymbol{\alpha}}_{ind}}$ and $\hat{\mu}_{i2} = e^{\ln(e_i)+\boldsymbol{z}_i{}'\hat{\boldsymbol{\beta}}_{ind}}$. The result of $\widehat{E[S_{ind}]}$ is 76'069 TDM. The MCE $\hat{S}_{ind}$ provides a quite similar value as the estimator $\widehat{E[S_{ind}]}$ which shows that the simulation works properly (simulation error of about 0.4%).

We see that the estimated expected total loss using the mixed copula model is about 2% smaller than the estimated expected total loss using the independent regression model. This can be explained by the estimation problems of our model for small claim frequencies and with the positive correlation between the number of claims and the average claim size in combination with the accumulated small number of claims per policy in the observed insurance portfolio. When the number of claims is small, the positive correlation causes a smaller average claim size as in the case of zero correlation, i.e., independency between the claim frequency and the average claim size. On the other hand, in the case of an insurance portfolio with an accumulated high number of claims per policy, we would get a higher total loss with the joint regression model than by using the independent regression models. Whether underestimation of the mixed copula model is systematic cannot be assessed here as we have only one available total loss observation for the insurance portfolio.

6 Summary and conclusion

The paper presents a new approach to modeling and estimating the total loss of an insurance portfolio. We developed a joint regression model with a marginal Gamma GLM for the continuous variable of the average claim size and a marginal Poisson GLM for the discrete variable for the number of claims. The GLM's were linked by the Mixed Copula Approach with a Gaussian Copula which has one parameter to model the dependency structure.

In order to fit the joint Gamma-Poisson regression model to data and to calculate the MLE's of the regression parameters as well as the correlation parameter of the Gaussian copula we constructed an algorithm based on the MBP algorithm and checked its quality by running an extensive simulation study with 24 scenarios which yielded the result that it works quite well in most of the scenarios, especially when the correlation is low or medium.

The application of the model to a full comprehensive car insurance portfolio of a German insurance company showed that there is a significant small positive dependency between the average claim size and the number of claims in this insurance portfolio. As the resulting parameter setting of the real insurance data set falls in the area of the scenario parameter settings for which the algorithm works well, we can act on the assumption that the parameter values for the insurance portfolio are well estimated.

Finally, we used the Monte Carlo method to estimate the expected total loss for the portfolio by using the results of the previous joint regression analysis. In comparison with the classical independent model, it was shown that the expected total loss estimated with the joint regression model is smaller than the one estimated with the classical model. Nevertheless, our joint model performs very well in total, but has problems for extreme values of the variables of interest. This raises the question if the choice of another copula might improve the model, which we will study in the future. Furthermore the marginal GLM for the number of claims might be improved by choosing a Generalized Poisson GLM (Consul and Jain 1973) in order to model over- and underdispersion.

Acknowledgement

C. Czado is supported by the DFG (German Research Foundation) grant CZ 86/1-3. We like to thank Peter Song for contributing valuable ideas and information.

References

Boskov, M. and R. J. Verrall (1994). Premium rating by geographic area using spatial models. *Astin Bull. 24*(1), 131–143.

Burden, R. L. and J. D. Faires (2004). *Numerical Analysis* (8th ed.). Pacific Grove: Brooks Cole Publishing.

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

Dimakos, X. and A. Frigessi (2002). Bayesian premium rating with latent structure. *Scand. Actuar. J. 2002*(3), 162–184.

Genest, C. and J. Neslehova (2007). A primer on copulas for count data. *Astin Bull. 37*, 475–515.

Gschlößl, S. and C. Czado (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scand. Actuar. J. 2007*(3), 202–225.

Haberman, S. and A. E. Renshaw (1996). Generalized linear models and actuarial science. *The Statistician 45*(3), 407–436.

Jørgensen, B. and M. C. P. de Souza (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scand. Actuar. J. 1994*(1), 69–93.

Kastenmeier, R. (2008). Joint regression analysis of insurance claims and claim sizes. Technische Universität München, Mathematical Sciences, Diploma thesis, http://www-m4.ma.tum.de/Diplarb/da_txt.html.

Liu, Y. and R. Luger (2009). Efficient estimation of copula-garch models. *Computational Statistics and Data Analysis. 53*(6), 2284–2297.

Lundberg, F. (1903). *Approximerad framställning afsannollikhetsfunktionen. II. återförsäkring af kollektivrisker.* Uppsala: Almqvist & Wiksells Boktr.

Ng, M. K., E. Y. Chan, M. M. So, and W.-K. Ching (2007). A bivariate zero-inflated Poisson regression model to analyze occupational injuries. *Pattern Recognition 40*(6), 1745–1752.

Renshaw, A. E. (1994). Modeling the claims process in the presence of covariates. *Astin Bull. 24*(2), 265–285.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Smyth, G. K. and B. Jørgensen (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bull. 32*(1), 143–157.

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scand. J. Statist. 2000*(2), 305–320.

Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications* (1st ed.), Volume 365 of *Springer Series in Statistics*. New York: Springer.

Song, P. X.-K., Y. Fan, and J. D. Kalbfleisch (2005). Maximization by parts in likelihood inference. *J. Amer. Statist. Assoc. 100*(472), 1145–1167.

Song, P. X.-K., M. Li, and Y. Yuan (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics 65*(1), 60–68.

Taylor, G. (1989). Use of spline functions for premium rating by geographic area. *Astin Bull. 19*(1), 89–122.

Wang, K., A. H. Lee, K. K. W. Yau, and P. J. W. Carrivick (2003). A semi-supervised regression model for mixed numerical and categorical variables. *Accident Analysis and Prevention 35*(4), 625–629.

Winkelmann, R. (2008). *Econometric analysis of count data* (5th ed.). Berlin: Springer.

Zhang, R., C. Czado, and A. Min (2011). Efficient maximum likelihood estimation of copula based meta t-distributions. *Computational Statistics and Data Analysis. 55*(3), 1196–1214.

Appendix

LEMMA 1 *The joint Gamma-Poisson-density function for Y_{i1} and Y_{i2} is given by*

$$f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = \begin{cases} g_1(y_{i1}|\mu_{i1}, \nu^2) D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})), & \textit{if } y_{i2} = 0; \\ g_1(y_{i1}|\mu_{i1}, \nu^2)[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) \\ \quad - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))], & \textit{if } y_{i2} \geq 1. \end{cases}$$

Proof: First we compute and simplify the derivative $\frac{\partial}{\partial u_1} C(u_1, u_{i2}|\rho)$. For this we use the abbreviation $q_1 := \Phi^{-1}(u_1)$.

$$\begin{aligned}
C_1'(u_1, u_{i2}|\rho) &= \frac{\partial}{\partial u_1} C(u_1, u_{i2}|\rho) \\
&= \frac{\partial}{\partial u_1} \frac{1}{2\pi\sqrt{|\det(\Gamma)|}} \int_{-\infty}^{q_1}\int_{-\infty}^{q_{i2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}'\Gamma^{-1}\boldsymbol{x}\right\} d\boldsymbol{x} \\
&= \frac{1}{2\pi\sqrt{|\det(\Gamma)|}} \int_{-\infty}^{q_{i2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} q_1, & x_2 \end{pmatrix}\Gamma^{-1}\begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\right\} dx_2 \;\times\; \frac{\partial}{\partial u_1}\, q_1 \\
&= \frac{1}{2\pi\sqrt{|\det(\Gamma)|}} \int_{-\infty}^{q_{i2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} q_1, & x_2 \end{pmatrix}\Gamma^{-1}\begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\right\} dx_2 \;\times\; \sqrt{2\pi}\exp\left\{\frac{1}{2}q_1^2\right\} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_{i2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(q_1\rho - x_2)^2\right\} dx_2.
\end{aligned}$$

Using the transformation $x_2 = z\sqrt{1-\rho^2} + \rho q_1$ it follows that

$$C_1'(u_1, u_{i2}|\rho) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_{i2}-\rho q_1}{\sqrt{1-\rho^2}}} \exp\left\{-\frac{1}{2}z^2\right\} dz = \Phi\left(\frac{\Phi^{-1}(u_{i2}) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) =: D_\rho(u_1, u_{i2}).$$

Equivalently we get

$$C_1'(u_1, u_{i2}^-|\rho) = \Phi\left(\frac{\Phi^{-1}(u_{i2}^-) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = D_\rho(u_1, u_{i2}^-).$$

For $y_{i2} = 0$ we obtain $u_{i2}^- = G_2(y_{i2} - 1|\mu_{i2}) = \sum_{k=0}^{-1} g_2(k|\mu_{i2}) = 0$. This implies $\Phi^{-1}(u_{i2}^-) = \Phi^{-1}(0) = -\infty$ and therefore we have

$$D_\rho(u_1, u_{i2}^-) = \Phi\left(\frac{\Phi^{-1}(u_{i2}^-) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = \Phi\left(\frac{-\infty - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = 0. \tag{6.1}$$

Using (2.4) gives the desired result. It is a valid joint density since

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{y_{i2}=0}^{n} [D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))] \\
=\; &\lim_{n\to\infty} [-\underbrace{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(-1|\mu_{i2}))}_{=0 \text{ according to (6.1)}} + D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(n|\mu_{i2}))] \\
=\; &D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), 1) = \Phi\left(\frac{\Phi^{-1}(1) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \\
=\; &\Phi\left(\frac{\infty - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = \Phi(\infty) = 1.
\end{aligned}$$

Lemma 2 *Let $g_1^*(\cdot)$ be the density function and $G_1^*(\cdot)$ the cdf of a Gamma$(a+1, b)$-distribution with $a = \frac{1}{\nu^2}$ and $b = \frac{1}{\mu_{i1}\nu^2}$, then we have*

$$\frac{\partial G_{i1}}{\partial \mu_{i1}} = \frac{1}{\mu_{i1}\nu^2}\,(G_{i1}^* - G_{i1}) \tag{6.2}$$

$$\frac{\partial G_{i2}}{\partial \mu_{i2}} = -g_2(y_{i2}|\mu_{i2}). \tag{6.3}$$

Proof: We have

$$\begin{aligned}
\frac{\partial G_{i1}}{\partial \mu_{i1}} &= \frac{\partial}{\partial \mu_{i1}} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy = \int_0^{y_{i1}} \frac{\partial g_1(y|\mu_{i1}, \nu^2)}{\partial \mu_{i1}}\ dy \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ (y - \mu_{i1})\ dy \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} \frac{1}{\Gamma(\frac{1}{\nu^2})} \left(\frac{1}{\mu_{i1}\nu^2}\right)^{1/\nu^2} y^{1/\nu^2}\ e^{-\frac{1}{\mu_{i1}\nu^2} y}\ dy \right. \\
&\quad \left. -\mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} \frac{1}{\Gamma(\frac{1}{\nu^2}+1)} \left(\frac{1}{\mu_{i1}\nu^2}\right)^{1/\nu^2+1} y^{1/\nu^2}\ e^{-\frac{1}{\mu_{i1}\nu^2} y}\ \mu_{i1}\ dy \right. \\
&\quad \left. -\mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\mu_{i1} \int_0^{y_{i1}} g_1^*(y|\mu_{i1}, \nu^2)\ dy - \mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}\nu^2}\ (G_{i1}^* - G_{i1}).
\end{aligned}$$

For the second part we have

$$\begin{aligned}
\frac{\partial G_{i2}}{\partial \mu_{i2}} &= \frac{\partial}{\partial \mu_{i2}} [\sum_{k=0}^{y_{i2}} \frac{1}{k!}\ \mu_{i2}^k\ e^{-\mu_{i2}}] \\
&= -e^{-\mu_{i2}} + \sum_{k=1}^{y_{i2}} \left(\frac{1}{(k-1)!}\ \mu_{i2}^{k-1}\ e^{-\mu_{i2}} - \frac{1}{k!}\ \mu_{i2}^k\ e^{-\mu_{i2}} \right) \\
&= -e^{-\mu_{i2}} + \sum_{k=0}^{y_{i2}-1} \frac{1}{k!}\ \mu_{i2}^k\ e^{-\mu_{i2}} - \sum_{k=1}^{y_{i2}} \frac{1}{k!}\ \mu_{i2}^k\ e^{-\mu_{i2}} \\
&= -\frac{1}{y_{i2}!}\ \mu_{i2}^{y_{i2}}\ e^{-\mu_{i2}} \\
&= -g_2(y_{i2}|\mu_{i2}).
\end{aligned}$$

Lemma 3

$$\frac{\partial D_\rho(G_{i1}, G_{i2})}{\partial \rho} = \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\ \frac{\rho\Phi^{-1}(G_{i2}) - \Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}.$$

Proof:

$$\begin{aligned}
\frac{\partial D_\rho(G_{i1}, G_{i2})}{\partial \rho} &= \frac{\partial}{\partial \rho}\Phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \frac{\partial}{\partial \rho} \left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&\quad \times \left(\frac{-\Phi^{-1}(G_{i1})\sqrt{1-\rho^2}}{1-\rho^2} + \frac{\rho(1-\rho^2)^{-1/2}(\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1}))}{1-\rho^2} \right)
\end{aligned}$$

$$= \phi\left(\frac{\Phi^{-1}(G_{i2})-\rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\ \frac{\rho\Phi^{-1}(G_{i2})-\Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}.$$

Lemma 4 *The partial derivatives of*

$$\frac{\partial\ l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial\boldsymbol{\theta}} = \left(\tfrac{\partial}{\partial\boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1,\gamma), \tfrac{\partial}{\partial\boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1,\gamma), \tfrac{\partial}{\partial\gamma} l_d^c(\boldsymbol{\theta}_1,\gamma)\right)'$$

are given by

$$\frac{\partial}{\partial\boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1,\gamma) = \sum_{i\in\mathcal{J}} \frac{d_\rho(G_{i1},G_{i2})-d_\rho(G_{i1},G_{i2}^-)}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\ \frac{G_{i1}^*-G_{i1}}{\phi(\Phi^{-1}(G_{i1}))}\ \frac{-\rho}{\sqrt{1-\rho^2}}\ \boldsymbol{x}_i, \tag{6.4}$$

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1,\gamma) &= \sum_{i\in\mathcal{J}} \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\left[d_\rho(G_{i1},Gi2^-)\frac{g_2(y_{i2}-1|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}^-))}\right.\\ &\left. -d_\rho(G_{i1},Gi2)\frac{g_2(y_{i2}|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}))}\right]\frac{\mu_{i2}}{\sqrt{1-\rho^2}}\boldsymbol{z}_i,\end{aligned} \tag{6.5}$$

$$\begin{aligned}\frac{\partial}{\partial\gamma} l_d^c(\boldsymbol{\theta}_1,\gamma) &= \sum_{i\in\mathcal{J}} \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\left\{d_\rho(G_{i1},G_{i2})[\rho\Phi^{-1}(G_{i2})-\Phi^{-1}(G_{i1})]\right.\\ &\left. -d_\rho(G_{i1},G_{i2}^-)[\rho\Phi^{-1}(G_{i2}^-)-\Phi^{-1}(G_{i1})]\right\}\frac{1}{\sqrt{1-\rho^2}}.\end{aligned} \tag{6.6}$$

Proof: We begin with the partial derivative with respect to $\boldsymbol{\alpha}$.

$$\frac{\partial}{\partial\boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1,\gamma) = \sum_{i\in\mathcal{J}} \frac{\partial}{\partial\boldsymbol{\alpha}} ln[D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)]$$

$$= \sum_{i\in\mathcal{J}} \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\left(\frac{\partial D_\rho(G_{i1},G_{i2})}{\partial G_{i1}} - \frac{\partial D_\rho(G_{i1},G_{i2}^-)}{\partial G_{i1}}\right)\frac{\partial G_{i1}}{\partial\boldsymbol{\alpha}},$$

where

$$\begin{aligned}\frac{\partial D_\rho(G_{i1},\cdot)}{\partial G_{i1}} &= \frac{\partial}{\partial G_{i1}}\Phi\left(\frac{\Phi^{-1}(\cdot)-\rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\\ &= \phi\left(\frac{\Phi^{-1}(\cdot)-\rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\frac{-\rho}{\sqrt{1-\rho^2}}\ \frac{1}{\phi(\Phi^{-1}(G_{i1}))},\end{aligned}$$

since $(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}$. Further with (6.2) of Lemma 2

$$\frac{\partial G_{i1}}{\partial\boldsymbol{\alpha}} = \frac{\partial G_{i1}}{\partial\mu_{i1}}\ \frac{\partial\mu_{i1}}{\partial\boldsymbol{\alpha}} = \frac{1}{\mu_{i1}\nu^2}\ (G_{i1}^*-G_{i1})\ \mu_{i1}\ \boldsymbol{x}_i,$$

Combining these parts we get expression (6.4). Next we compute the partial derivative with respect to $\boldsymbol{\beta}$.

$$\frac{\partial}{\partial\boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1,\gamma) = \sum_{i\in\mathcal{J}} \frac{\partial}{\partial\boldsymbol{\beta}}\left\{ln[D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)]\right\}$$

$$= \sum_{i\in\mathcal{J}} \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\left(\frac{\partial D_\rho(G_{i1},G_{i2})}{\partial G_{i2}}\ \frac{\partial G_{i2}}{\partial\boldsymbol{\beta}} - \frac{\partial D_\rho(G_{i1},G_{i2}^-)}{\partial G_{i2}^-}\ \frac{\partial G_{i2}^-}{\partial\boldsymbol{\beta}}\right),$$

where

$$\begin{aligned}\frac{\partial D_\rho(\cdot, G_{i2})}{\partial G_{i2}} &= \frac{\partial}{\partial G_{i2}} \Phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(\cdot)}{\sqrt{1-\rho^2}}\right) \\ &= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(\cdot)}{\sqrt{1-\rho^2}}\right) \frac{1}{\sqrt{1-\rho^2}} \frac{1}{\phi(\Phi^{-1}(G_{i2}))},\end{aligned}$$

and with (6.3) of Lemma 2

$$\frac{\partial G_{i2}}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}}{\partial \mu_{i2}} \frac{\partial \mu_{i2}}{\partial \boldsymbol{\beta}} = -g_2(y_{i2}|\mu_{i2})\ \mu_{i2}\ \boldsymbol{z}_i.$$

Similarly, we derive for G_{i2}^-

$$\frac{\partial G_{i2}^-}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}^-}{\partial \mu_{i2}} \frac{\partial \mu_{i2}}{\partial \boldsymbol{\beta}} = -g_2(y_{i2}-1|\mu_{i2})\ \mu_{i2}\ \boldsymbol{z}_i.$$

All these parts together lead to (6.5). Finally, we consider the derivative with respect to γ:

$$\frac{\partial}{\partial \gamma} l_d^c(\boldsymbol{\theta}_1, \gamma) = \sum_{i \in \mathcal{J}} \frac{\partial}{\partial \gamma} ln[D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)]$$

$$= \sum_{i \in \mathcal{J}} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \left(\frac{\partial D_\rho(G_{i1}, G_{i2})}{\partial \rho} - \frac{\partial D_\rho(G_{i1}, G_{i2}^-)}{\partial \rho}\right) \frac{\partial \rho}{\partial \gamma}.$$

Using Lemma 3 we have

$$\begin{aligned}\frac{\partial D_\rho(G_{i1}, G_{i2})}{\partial \rho} &= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \frac{\rho\Phi^{-1}(G_{i2}) - \Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}, \\ \frac{\partial D_\rho(G_{i1}, G_{i2}^-)}{\partial \rho} &= \phi\left(\frac{\Phi^{-1}(G_{i2}^-) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \frac{\rho\Phi^{-1}(G_{i2}^-) - \Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}.\end{aligned}$$

Further

$$\frac{\partial \rho}{\partial \gamma} = \frac{\partial}{\partial \gamma} \frac{e^{2\gamma}-1}{e^{2\gamma}+1} = 1-\rho^2.$$

Combining all these parts again we finally get expression (6.6).

Corresponding author

Eike Christian Brechmann

Center for Mathematical Sciences
Technische Universität München
Boltzmannstr. 3
D-85747 Garching, Germany

brechmann@ma.tum.de

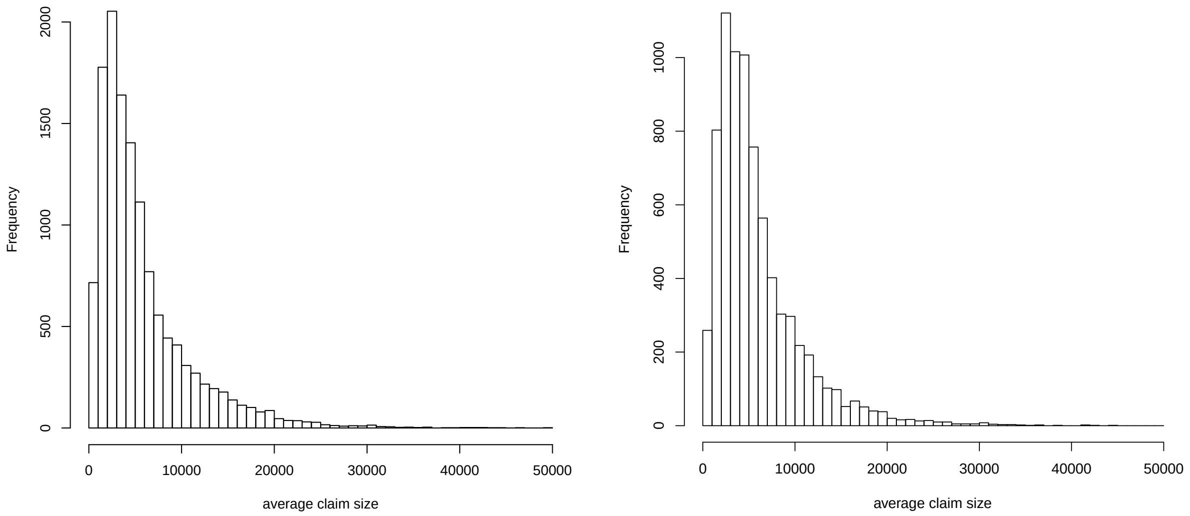

Figure 1: Histogram of the observed average claim size in the original (left panel) and in the aggregated data set (right panel).

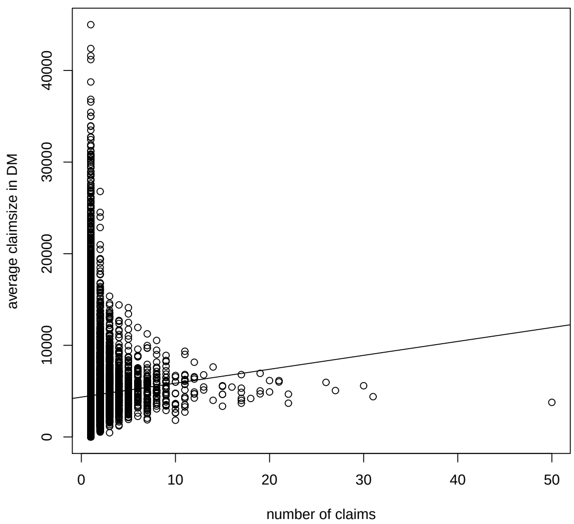

Figure 2: Plot of the number of claims against the average claim size of the groups

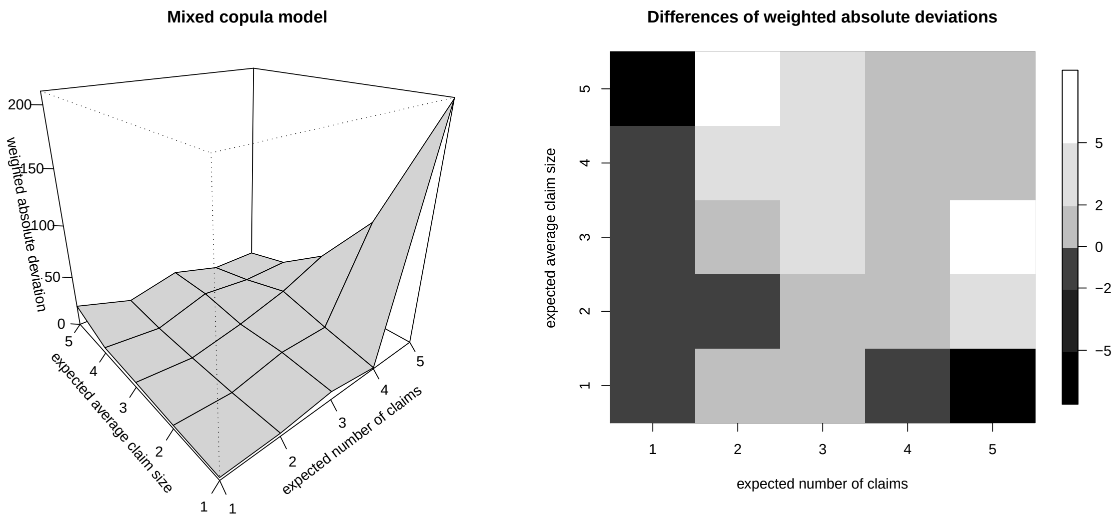

Figure 3: Left panel: plot of absolute deviations from the observed total loss weighted by exposure for each risk group for the mixed copula model. Right panel: comparison of absolute deviations of the joint regression model and classical independent model; light colors indicate risk groups in which the joint regression model performs better. (Classification of risk groups: a small index means a small expected value for the respective value.)

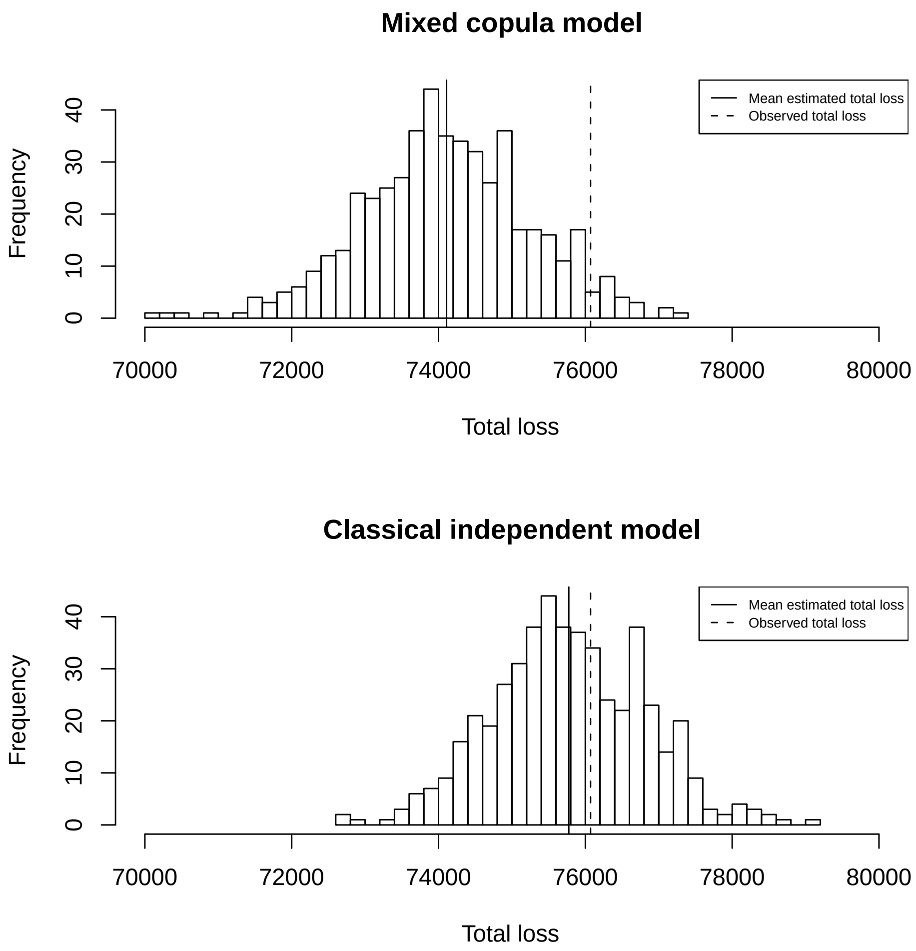

Figure 4: Histogram of the estimated expected total loss using the results of the mixed copula regression model and the independent GLMs, respectively.