Technische Universität München

Zentrum Mathematik

Joint Regression Analysis of Insurance Claims and Claim Sizes

Diplomarbeit

von

Rainer Kastenmeier

Themenstellerin: Prof. Dr. Claudia Czado
Betreuer: Dr. Aleksey Min
Abgabetermin: 23. April 2008

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Wiesbaden, den 22. April 2008

Acknowledgments

First, I would like to express my gratitude to Ms Prof. Dr. Claudia Czado for making this dioloma thesis possible. I am very thankful for the frequent meetings with the many helpful discussions and hints which enriched the progress of my work a lot.
I also want to address many thanks to Mr Dr. Aleksey Min for the fruitful discussions and the support with the numerous simulations and calculations running for me on the server during the last months.

Contents

List of Tables

List of Figures

1 Introduction

The estimation of total loss distributions plays an important role in the daily work of actuaries, like in the areas of pricing of non-life reinsurance contracts, reserve estimation and calculation or risk management. Therefore, a solid estimation of total loss distributions of an insurance portfolio is essential and thus constitutes a responsible but also extensive and complex research field.

The total loss of an insurance portfolio is calculated by the sum-up of the product of the average claim size and the number of claims of every insurance policy being part of this insurance portfolio. In the classical model, going back on the theory of Lundberg (1903), the average claim size and the number of claims are assumed to be independent (see also Mack (1997), Chapter 1). Thereby, the average claim size follows a Gamma distribution, while the average claim size is modeled with a Poisson distribution. However, certain cases exist for which the independency assumption does not hold.[1] Gschlößl and Czado (2007), for instance, analyze a comprehensive car insurance data set using a full Bayesian approach in their work. In this analysis, they allow for dependency between the average claim size and the number of claims and detect that the dependency between these variables turns out to be significant.

The main objective of this present thesis now is to construct a joint regression model for the average claim size and the number of claims taking into account the possible dependency between both variables. Therefore, we assume a Gamma generalized linear model (GLM) for the average claim size, whereas the number of claims are modeled with a Poisson GLM. We connect these marginal models with a Gaussian copula by using the Mixed Copula Approach given in Song (2007). In order to apply the model on a full comprehensive car insurance data set, we also develop a new algorithm for the estimation of the model parameters. This algorithm is based on the Maximization by Parts algorithm, first introduced by Song et al. (2005).

The content of the thesis is structured as follows: In Chapter 2 we give a brief introduction to the main mathematical theories and methods used in the paper. The joint distribution and the regression model are constructed in Chapter 3 by using a Mixed Copula Approach with a marginal Gamma GLM and a marginal Poisson GLM. We also calculate the density function for this Gamma-Poisson distribution here. In Chapter 4 we introduce the Maximization by Parts algorithm. We apply the Maximization by Parts algorithm to the Gamma-Poisson regression model in order to receive maximum likelihood estimators for the regression parameters and the copula parameter. We then run a simulation study of the applied algorithm in Chapter 5 to check its quality. For the simulation study, we also construct a sampling algorithm to generate observations from the joint Gamma-Poisson distribution. The model finally is applied to a real insurance data set in Chapter 6. After calculating the maximum likelihood estimator for the parameters of the joint regression model with the Maximization by Parts algorithm, we use these results to calculate the expected total loss for the insurance portfolio. The last part, Chapter 7, summarizes the main aspects and results of the paper and concludes with a short outlook.

[1] cp. Mack (1997), pp. 107

2 Background

Before starting with the main part of the paper, this chapter gives an outline of the main mathematical terms and methods used in this thesis.

2.1 Distributions

In this section we give the definitions of the densities and the cumulative distribution function of the distributions we use in the paper.

2.1.1 Gamma Distribution

The first distribution is the Gamma distribution, that is a two-parametrical continuous probability distribution. In the paper we use two different parameterizations of this distribution: the shape-rate parameterization and the mean parameterization. Let $G_1(\cdot|a,b)$ (or $g_1(\cdot|a,b)$) denote the cumulative distribution function (cdf) (or density function) of a $Gamma(a,b)$ distribution with shape parameter $a > 0$ and rate parameter $b > 0$. The density is then given as

$$g_1(y|a,b) = \frac{1}{\Gamma(a)}\; y^{a-1}\; b^a\; e^{-bx}. \tag{2.1}$$

A $Gamma(a,b)$-distributed random variable Y has then the mean $E[y] = \frac{a}{b}$ and the variance $Var[y] = \frac{a}{b^2}$. Here, we mainly use the mean parameterization of the Gamma distribution, so that the density has the form

$$g_1(y|\mu,\nu^2) = \frac{1}{\Gamma(\frac{1}{\nu^2})}\; \left(\frac{1}{\mu\nu^2}\right)^{1/\nu^2}\; y^{1/\nu^2-1}\; e^{-\frac{1}{\mu\nu^2}y}, \tag{2.2}$$

with $\mu := E[Y] = \frac{a}{b} > 0$ and $\nu^2 := \frac{1}{a} > 0$, where $Var[y] = \mu^2\nu^2$. The parameter ν is called the constant coefficient of variation and ν^2 is equal to the dispersion parameter φ of the Gamma distribution. $G_1(\cdot|\mu,\nu)$ (or $g_1(\cdot|\mu,\nu)$) denotes the cumulative distribution function (cdf) (or density function) of the Gamma distribution with mean parameterization.

2.1.2 Poisson and Zero-truncated Poisson Distribution

Further, we need the Poisson distribution in this paper. The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed time period when these events occur with a known average frequency rate and independently of the time since the last event. Let $G_2(\cdot|\lambda)$ (or $g_2(\cdot|\lambda)$) denote the cdf (or probability mass) of the $Poisson(\lambda)$-distribution. The probability mass has the form

$$g_2(y|\lambda) = P(Y=y) = \begin{cases} 0 & ,\quad \text{for } y < 0; \\ \frac{1}{y!}\;\lambda^y\; e^{-\lambda} & ,\quad \text{for } y = 0,1,2,\dots\;, \end{cases} \tag{2.3}$$

with the frequency rate parameter $\lambda > 0$. The cdf is then given as

$$G_2(y|\lambda) = P(Y < y) = \sum_{k=1}^{y} g_2(k|\lambda).$$

When Y is a $Poisson(\lambda)$-distributed random variable it holds that

$$E[Y] = \lambda = Var[Y].$$

We also use a conditional Poisson distribution in the paper. Let $Y \in \mathbb{N}$ with $Y > 0$ be $Poisson(\lambda)$-distributed conditional on $Y > 0$. The probability mass of the conditional Poisson distribution is then given by

$$g_{2|Y>0}(y|\lambda) := P(Y = y|Y > 0) = \frac{P(Y = y)}{P(Y > 0)} = \frac{\lambda^y\ e^{-\lambda}}{y!\ (1 - e^{-\lambda})}. \tag{2.4}$$

We call the discrete distribution with the probability mass (2.4) Zero-truncated Poisson distribution. The mean and the variance of this distribution are calculated as

$$\begin{aligned}
E[Y|Y > 0] &= \sum_{y=1}^{\infty} y\ g_{2|Y>0}(y|\lambda) \\
&= \sum_{y=1}^{\infty} y\ \frac{\lambda^y\ e^{-\lambda}}{y!\ (1 - e^{-\lambda})} \\
&= \frac{1}{1 - e^{-\lambda}} \sum_{y=1}^{\infty} \underbrace{y\ \frac{\lambda^y\ e^{-\lambda}}{y!}}_{\widehat{=}\ E[Y]} \\
&= \frac{\lambda}{1 - e^{-\lambda}}; \qquad (2.5)
\end{aligned}$$

$$\begin{aligned}
E[Y^2|Y > 0] &= \sum_{y=1}^{\infty} y^2\ g_{2|Y>0}(y|\lambda) \\
&= \sum_{y=1}^{\infty} y^2\ \frac{\lambda^y\ e^{-\lambda}}{y!\ (1 - e^{-\lambda})} \\
&= \frac{1}{1 - e^{-\lambda}} \sum_{y=1}^{\infty} \underbrace{y^2\ \frac{\lambda^y\ e^{-\lambda}}{y!}}_{\widehat{=}\ E[Y^2]=Var[Y]+(E[Y])^2} \\
&= \frac{\lambda + \lambda^2}{1 - e^{-\lambda}};
\end{aligned}$$

$$\begin{aligned}
Var[Y|Y > 0] &= E[Y|Y > 0] - (E[Y|Y > 0])^2 \\
&= \frac{\lambda + \lambda^2}{1 - e^{-\lambda}} - \frac{\lambda^2}{(1 - e^{-\lambda})^2} \\
&= \lambda \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{(1 - e^{-\lambda})^2} \\
&= E[Y|Y > 0] \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{(1 - e^{-\lambda})}, \qquad (2.6)
\end{aligned}$$

where λ is equal to the frequency rate parameter of the underlying common $Poisson(\lambda)$ distribution.

2.1.3 Normal Distribution

Other distributions we frequently use in this paper are the normal distribution and the 2-dimensional normal distribution. Let $\Phi(\cdot)$ denote the distribution function of the normal distribution $N(\mu, \sigma)$ with zero mean and unit variance, i.e. $\mu = 0$ and $\sigma = 1$, and $\phi(\cdot)$ the corresponding density function. $\Phi_2(\cdot|\Sigma)$ denotes the distribution function of the 2-dimensional normal distribution $N_2(\boldsymbol{\mu}, \Sigma)$ with zero mean, i.e. $\boldsymbol{\mu} = (0, 0)'$, and covariance matrix Σ. The density function has the form

$$\phi_2(\boldsymbol{x}|\Sigma) = \frac{1}{2\pi\sqrt{|\det(\Sigma)|}} \exp\left\{-\frac{1}{2}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}\right\},$$

with $\boldsymbol{x} \in \mathbb{R}^2$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$, where ρ denotes the correlation between the two margins of the 2-dimensional normal distribution. Further, we use the property of the multivariate normal distribution, that the conditional distribution of one margin given the other is a normal distribution again.

2.2 Maximum Likelihood Estimation

Assuming that we have a random variable X with observation $x_1, \dots, x_n$ and we know that X follows the cdf $F_{\boldsymbol{\theta}}(\cdot)$ with the unknown parameter $\boldsymbol{\theta} \in \Theta$, where Θ denotes the parameter space. To get the full specified distribution of X we can estimate the parameter $\boldsymbol{\theta}$ by using the observed values of X. One method to do this is to calculate the so called Maximum Likelihood Estimators (MLE) of the unknown parameter $\boldsymbol{\theta}$. The MLE of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, maximize the likelihood

$$L(\boldsymbol{\theta}|x_1, \dots, x_n) = \prod_{i=1}^{n} f_{\boldsymbol{\theta}}(x_i)$$

or log-likelihood function

$$l(\boldsymbol{\theta}|x_1, \dots, x_n) = \log(L(\boldsymbol{\theta}|x_1, \dots, x_n)) = \sum_{i=1}^{n} f_{\boldsymbol{\theta}}(x_i),$$

where $f_{\boldsymbol{\theta}}(\cdot)$ is the density of the cdf $F_{\boldsymbol{\theta}}(\cdot)$. In order to receive the MLEs, we maximize the log-likelihood function by calculating the partial derivatives of the log-likelihood for our parameters, the so called score functions. By setting them equal to 0 (then we have the score equations) and solving them under regularity conditions, we get the MLE for each parameter. The regularity conditions are as follows:

(RC1) the parameter space $\Theta \subset \mathbb{R}^n$ is an open interval.

(RC2) the first and second derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$ exist for all $\boldsymbol{\theta} \in \Theta$.

(RC3) the Fisher information $\mathcal{I}(\boldsymbol{\theta}^*) := -E\left[\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]$ exists and is not equal to 0 for all $\boldsymbol{\theta}^* \in \Theta$.

Usually, the score equations are non linear and can only be solved with a numerical algorithm.

2.3 Generalized Linear Model (GLM)

2.3.1 Short Introduction to GLMs

When constructing the margins of our model we use the so called generalized linear model (GLM), which is a generalization of the common least squares regression method. According to Nelder, J. A. and Wedderburn, R. W. M. (1972), in a GLM, the independent observations $y_1, \dots, y_n$ have to be from a random variable Y whose distribution is one of the exponential family. The density of the exponential family is given by

$$f(y, \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right], \tag{2.7}$$

with parameters θ and ϕ, where the functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known. The parameter ϕ is called dispersion parameter. The mean and the variance of a random variable Y whose distribution is one of the exponential family is given by

$$E[Y] = b'(\theta) \tag{2.8}$$

and

$$Var[Y] = b''(\theta)a(\phi), \tag{2.9}$$

where $b'(\cdot)$ and $b''(\theta)$ denote the first and the second derivative of the function $b(\cdot)$.
Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be the covariate vectors of the observations. Typically, we want to relate the parameter θ_i of the distribution to the predictors $\eta_i := \boldsymbol{x}_i{}'\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the parameter vector of the predictor. This is done by modeling

$$E[Y] = b'(\theta_i),$$

with

$$\theta_i = g^{-1}(\boldsymbol{x}_i{}'\boldsymbol{\beta}),$$

where $g(\cdot)$ is a known function, called the link function. Assuming that the observations are a sample of a normal distribution, which falls in the category of the exponential distribution family, the parameter θ is identical with the mean parameter μ of the normal distribution. By taking the identity as link function, we then get a common least squares regression model.
In the paper we construct a joint regression model using a $Gamma(\mu, \nu^2)$-distributed and a $Poisson(\mu)$-distributed margin. The Gamma as well as the Poisson distribution are two other representatives of the exponential distribution family. Hence, we use a GLM for the Gamma distributed sample and a GLM for a Poisson distributed sample. In both cases we do the regression on the positive mean parameters of the distributions and so we need a link function mapping $\mathbb{R}^+$ on $\mathbb{R}$. This property is provided by the natural logarithm $\ln(\cdot)$ as link function.
To calculate an estimator for the GLM parameter $\boldsymbol{\beta}$ we take the solution of the maximum likelihood equation, the MLE $\hat{\boldsymbol{\beta}}$. The fitting of the parameter and the matching to the observed data can be measured quantitatively by the so called deviance. This goodness-of-fit measure is defined as

$$D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}) = -2(l(\hat{\boldsymbol{\theta}}|\boldsymbol{y}, X) - l(\tilde{\boldsymbol{y}}|\boldsymbol{y}, X)), \tag{2.10}$$

where $\hat{\boldsymbol{\theta}} = (g^{-1}(\boldsymbol{x}_1{}'\hat{\boldsymbol{\beta}}, \dots, g^{-1}(\boldsymbol{x}_n{}'\hat{\boldsymbol{\beta}}))'$ and $\tilde{\boldsymbol{y}} = (g^{-1}(y_1), \dots, g^{-1}(y_n))'$. The deviance is asymptotic χ^2-distributed and so we can construct an asymptotic goodness-of-fit test which is called the Partial Deviance Test.[1] The test proceeds with a significant level of $1 - \alpha$ as follows:

[1]Czado (2004), Lecture 2: Introduction to GLM's, slide 18

$$\begin{array}{lll} \eta = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 & \text{Model F with deviance } D_F & \boldsymbol{\beta}_1 \in \mathbb{R}^{p_1},\ \boldsymbol{\beta}_2 \in \mathbb{R}^{p_2} \\ \eta = X_1\boldsymbol{\beta}_1 & \text{Model R with deviance } D_R & p_1 + p_2 = p \end{array}$$

$H_0: \ \boldsymbol{\beta}_2 = 0 \quad H_1: \boldsymbol{\beta}_2 \neq 0$
Reject $H_0 \Leftrightarrow D_R - D_F > \chi^2_{p-p_2=p_1,1-\alpha}$

Later on, we use the Partial Deviance Test to select reasonable regression models.

2.3.2 Zero-truncated Poisson GLM

We want to apply our joint regression model on a data set containing a count response with all observations greater one. Therefore, we have to construct a regression model with the Zero-truncated Poisson distribution. The probability mass of the Zero-truncated Poisson distribution is given in 2.4. It is easy to see that the Zero-truncated Poisson distribution is a distribution of the exponential family. By setting

$$\begin{aligned} \theta &= \ln(\lambda), \\ b(\theta) &= \exp^{\theta} + \ln(1 - \exp(-e^{\theta})), \\ \phi &= 1, \\ a(\phi) &= \phi, \\ c(y, \phi) &= \phi \ln(y!) \end{aligned}$$

we can transfer the probability mass of the Zero-truncated Poisson distribution in the exponential family density form. So it is possible to construct a GLM with the Zero-truncated Poisson distribution.
Let $y_i \in \mathbb{N}$, $i = 1, \ldots, n$, 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. We are interested in estimating the frequency rate parameter λ_i for the Zero-truncated Poisson distributions of the random variables Y_i. Therefore, we use a regression model specified by

$$Y_i \ \sim \ Zero-truncated\ Poisson(\lambda) \text{ with } \ln(\lambda_i) = \ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}, \tag{2.11}$$

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

$$\begin{aligned} \mu_i := E[Y_i|Y_i > 0] &= \frac{\lambda_i}{1 - e^{-\lambda_i}} \\ &= \frac{e^{\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}}}{1 - \exp(-e^{\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}})} \\ &= g^{-1}(\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}), \end{aligned} \tag{2.12}$$

where $g^{-1}(\cdot)$ denotes the inverse of the link function $g(\cdot)$. Equation (2.12) can be estimated by

$$g(y_i) = ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}. \tag{2.13}$$

As we can not calculate $g(\cdot)$ analytically, we linearize the inverse link function $g^{-1}(\cdot)$ using the first order Taylor approximation:

$$\begin{aligned} \mu_i &= g^{-1}(\lambda_i) = \frac{\lambda_i}{1 - e^{-\lambda_i}} \\ &\approx \frac{a_i}{1 - e^{-a_i}} + \frac{1 - e^{-a_i} - a_i e^{-a_i}}{(1 - e^{-a_i})^2}(\lambda_i - a_i), \end{aligned} \tag{2.14}$$

with $a_i \in \mathbb{R}^+$. An approximation of the link function $g(\cdot)$ is then given by

$$\begin{aligned}\tilde{g}(\mu_i) &:= a\left(1 - \frac{1-e^{-a_i}}{1-e^{-a_i}-a_i e^{-a_i}}\right) + \frac{(1-e^{-a_i})^2}{1-e^{-a_i}-a_i e^{-a_i}}\,\mu_i \\ &\approx \ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}. \end{aligned} \tag{2.15}$$

Thereby, a_i denotes the development point of the Taylor approximation and should be as close as possible to the value of λ_i in order to get a good approximation of $g(\cdot)$.
To calculate the MLE of the regression parameter $\boldsymbol{\beta}$ we need the log-likelihood of the Zero-truncated Poisson GLM, which is

$$l(\beta|\boldsymbol{y}, Z) := \sum_{i=1}^{n} y_i \ln(\lambda_i) - \lambda_i - \ln(1-e^{-\lambda_i}) + \ln(y_i!), \tag{2.16}$$

with $\boldsymbol{y} = (y_1, \dots x_n)$ and the design matrix $Z = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)'$, where $\lambda_i = \exp(\ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta})$.
The score function of the log-likelihood is given by

$$\frac{\partial l(\beta|y_1,\dots,y_n;\boldsymbol{z}_1,\dots,\boldsymbol{z}_n)}{\partial\beta} = \sum_{i=1}^{n} y_i\boldsymbol{z}_i - \lambda_i\boldsymbol{z}_i - \frac{\lambda_i e^{-\lambda_i}}{1-e^{-\lambda_i}}\boldsymbol{z}_i. \tag{2.17}$$

The R-function `mle.tuncpois` calculates the MLE of the regression parameter vector $\boldsymbol{\beta}$ using the score function. The code of the R-function can be found in Appendix C.1.
Now let $\hat{\boldsymbol{\beta}}$ be the MLE of the regression parameter vector $\boldsymbol{\beta}$. Then, the deviance $D(\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ of the Zero-truncated Poisson GLM is given by

$$D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}) = -2(l(\hat{\boldsymbol{\lambda}}|\boldsymbol{y}, Z) - l(\tilde{\boldsymbol{y}}|\boldsymbol{y}, Z)), \tag{2.18}$$

where $\hat{\boldsymbol{\lambda}} = (\hat{\lambda_1}, \dots, \hat{\lambda_m})'$ with $\hat{\lambda}_i = \exp(\ln(e_i) + \boldsymbol{z}_i{}'\hat{\boldsymbol{\beta}})$ are the estimated parameters and $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$ with

$$\tilde{y}_i = \tilde{g}^{-1}(y_i) = a\left(1 - \frac{1-e^{-a_i}}{1-e^{-a_i}-a_i e^{-a_i}}\right) + \frac{(1-e^{-a_i})^2}{1-e^{-a_i}-a_i e^{-a_i}}\,y_i$$

are the transformed Zero-truncated Poisson observations. Note that the deviance given in (2.18) corresponds only approximately with the real deviance of the Zero-truncated Poisson GLM as $\tilde{\boldsymbol{y}}$ are not exact transformations. Despite that we can not exactly calculate the deviance of the Zero-truncated Poisson GLM, the Partial Deviance Test for two nested Zero-truncated Poisson GLMs works properly. The reason is that in the test statistic of this test the parts with $\tilde{\boldsymbol{y}}$ of the two deviances compensate themselves and so the test statistic is exact.

2.4 Copula

2.4.1 Short Introduction to Copulas

We use a copula in order to connect the two margins in a joint distribution function. A two-dimensional copula is a function C from $[0,1] \times [0,1]$ to $[0,1]$ with the following properties:

(C1) For every $u, v \in [0,1]$

$$C(u,0) = 0 = C(0,v),$$

and

$$C(u,1) = u \text{ and } C(1,v) = v.$$

(C2) For every $u_1, u_2, v_1, v_2 \in [0,1]$ with $u_1 \leq u_2$ and $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Because of these properties a copula can also be interpreted as a bivariate cdf with $[0,1]$-uniform distributed margins. With Sklar's Theorem[2] we have the tool to "couple" two margins with a copula in one distribution. The theorem says that when C is a copula and F_1 and F_2 are distribution functions, then the function

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)),$$

with $x_1 \in F_1^{-1}([0,1])$ and $x_2 \in F_2^{-1}([0,1])$, is a joint distribution function with margins F_1 and F_2.[3]
There is a wide range of functions fulfilling the properties of a copula. Here, we use a one-parametrical copula, the Gaussian copula. The Gaussian copula is defined with help of the standard normal distribution and has the form

$$C(u_1, u_2|\rho) = \Phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\middle|\,\Gamma\right), \tag{2.19}$$

with $u_1, u_2 \in [0,1]$ and $\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The parameter ρ measures the association between the marginal random variables $X1$ with distribution F_1 and X_2 with distribution F_2. As $u_1 = F_1^{-1}(x_1)$ and $u_2 = F_2^{-1}(x_2)$, the parameter ρ is the Pearson correlation between the so called Normal-Scores $q_1 = \Phi^{-1}(u_1)$ and $q_2\Phi^{-1}(u_2)$ and is defined by

$$\rho = Cor(q_1, q_2) = \frac{Cov(q_1, q_2)}{\sqrt{Var[q_1]Var[q_2]}} = \frac{E[(q_1 - E[q_1])(q_2 - E[q_2])]}{\sqrt{Var[q_1]Var[q_2]}}. \tag{2.20}$$

Note that the normal scoring can also be used to measure the association between two discrete random variables. In this case, however, the above interpretation is not applicable.

2.4.2 Mixed Copula Approach

Now we expand the above copula approach to get a joint distribution function for a continuous random variable and a discrete random variable. For this let Y_1 be a continuous variable with distribution function F_1 and Y_2 be a discrete variable with distribution function F_2. Using the copula approach by ignoring that both random variables have to be continuous, we get this joint distribution function:

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = F(y_1, y_2) = C(F_1(y_1), F_2(y_2)). \tag{2.21}$$

To receive a general formulation of the joint density $f(y_1, y_2)$ we first calculate

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 = y_2) &= P(Y_1 \leq y_1, Y_2 \leq y_2) - P(Y_1 \leq y_1, Y_2 \leq y_2 - 1) \\ &= C(F_1(y_1), F_2(y_2)) - C(F_1(y_1), F_2(y_2 - 1)). \end{aligned} \tag{2.22}$$

With this result it follows that the joint density is computed by

$$\begin{aligned} f(y_1, y_2) &= \frac{\partial}{\partial y_1} P(Y_1 \leq y_1, Y_2 = y_2) \\ &= f_1(y_1)\left[C_1'(F_1(y_1), F_2(y_2)) - C_1'(F_1(y_1), F_2(y_2 - 1))\right], \end{aligned} \tag{2.23}$$

where $C_1'(u_1, u_2) = \frac{\partial}{\partial u_1} C(u_1, u_2)$. In the case of using a Gaussian copula the result is given in Song (2007) equation (6.9).

[2] cp. Sklar (1959)
[3] cp. Nelsen (1999), p.18

2.5 Numerical Methods

The algorithm maximizing the log-likelihood used here applies two different numerical methods. The first one we implement is the Fisher scoring method, which is an iterative method to find the root of the first derivative of the log-likelihood based on the Newton-Raphson method. The difference between the Newton-Raphson method and the Fisher scoring method is that we replace the inverse of the second derivative matrix of the Newton-Raphson method by the inverse of Fisher's information matrix. So one step of the method has the following form:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} + [\mathcal{I}(\boldsymbol{\theta}^{k})]^{-1} \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{k}}. \tag{2.24}$$

As we see, we need the first and the second derivative of the log-likelihood for this method.
On the other hand, we utilize the one-dimensional bisection method, which we also use to find the root of a first derivative. Let us assume that in the interval $[a, b]$, $a, b \in \mathbb{R}$ with $a < b$, one root of the function $f : \mathbb{R} \leftarrow \mathbb{R}$ exists. We choose a and b in such way that $f'(a) * f'(b) \leq 0$. Further, let $\epsilon \in \mathbb{R}$ denote the accuracy parameter. Then the one-dimensional bisection method proceeds as follows:

Algorithm 2.1 (One-dimensional Bisection Method)**.**

Set $m = \frac{a+b}{2}$.
While $|f(m)| \leq \epsilon$ *do*

  If $f(a) * f(m) > 0$
    Set $a_{new} = m$ *and* $m = \frac{a_{new}+b}{2}$
  Else
    Set $b_{new} = m$ *and* $m = \frac{a+b_{new}}{2}$.

When the bisection stops, m is an approximation for the root of $f(\cdot)$. In contrast to the Fisher scoring method, the bisection method does not require a second derivative. But, when we apply the bisection method to a high-dimensional search, it does not work efficiently anymore.

2.6 Estimator Quality

In order to test the quality of the estimator provided by a calculation algorithm, we need to measure the goodness of the estimators. The following definitions and formulas are based on the chapter 8.1 of Stekeler (2004). We expect that a good parameter estimator provides a result corresponding approximately with the true parameter value. The estimator should not systematically under- or overestimate the true value. The under- and overestimation of an estimator is measured by the bias.

Definition 2.2 (Bias and unbiasedness of an estimator)**.**
Let $\hat{\theta}$ be the estimator of the parameter with its true value θ. The bias of the estimator $\hat{\theta}$ is then given as

$$b(\theta, \hat{\theta}) := E[\hat{\theta}] - \theta. \tag{2.25}$$

When the expected value of the estimator is equal to the parameter, i.e.

$$b(\theta, \hat{\theta}) = 0 \Leftrightarrow E[\hat{\theta}] = \theta,$$

the estimator $\hat{\theta}$ is called unbiased.

An accurate estimator should not only have a small bias, it should also have a small variance. Otherwise, the values of the estimator scatter in a wide range around $E[\hat{\theta}]$. Hence, we expect of a good estimator that twice the standard deviation of the estimator is smaller than its absolute expectation value, i.e.

$$\frac{|E[\hat{\theta}]|}{\sqrt{Var[\hat{\theta}]}} > 2. \tag{2.26}$$

The unbiasedness of an estimator is a preferable property, but it is only relevant when the average expected deviation between the estimator and the true value of the parameter is rather small. Another useful measure of goodness for an estimator is the mean squared error:

Definition 2.3 (Mean squared error)**.**
Let $\hat{\theta}$ be the estimator of the parameter with its true value θ. The mean squared error (mse) of the estimator is then given as

$$mse(\theta, \hat{\theta}) := E[(\hat{\theta} - \theta)^2]. \tag{2.27}$$

The mean squared error gives the the expected squared deviation between the estimator $\hat{\theta}$ and the true value θ. When we transform the expression of equation (2.27), we see, that we can rewrite the mean squared error as the sum of the variance of the estimator $\hat{\theta}$ and its squared bias:

$$\begin{aligned} mse(\theta, \hat{\theta}) &= E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)] + E[(E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2E[\hat{\theta}E[\hat{\theta}] - E[\hat{\theta}]^2 - \hat{\theta}\theta + E[\hat{\theta}]\theta] + E[(E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2(E[\hat{\theta}]E[\hat{\theta}] - E[\hat{\theta}]^2 - E[\hat{\theta}]\theta + E[\hat{\theta}]\theta) + E[(E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2 \\ &= Var[\hat{\theta}] + b(\theta, \hat{\theta})^2. \end{aligned} \tag{2.28}$$

With expression (2.28) of the mean squared error we see that the mean squared error of an estimator depends on the bias and on the variance of the estimator. Therefore, in order to get a minimal mean squared error, the bias and the variance must be minimal. We also recognize that the variance of an estimator is only a good measure of goodness when the estimator is unbiased. We want to use measures (2.26) and (2.27) to check the goodness of the MLEs for the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the marginal GLMs, for the correlation parameter ρ of the Gaussian copula and of the constant of variation ν of the marginal Gamma distribution. As we do not know the distribution of the estimators, we can not calculate their expected values and their variances. Hence, we have to estimate the expected value and the variance of the estimators. For this purpose, we produce R independent identically distributed (iid) estimators for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ρ and ν and derive the asymptotical behavior of the estimator from the R observed estimators of each estimator.

An unbiased and consistent estimator of the expected value is the sample mean (see e.g. Georgii (2002), p.194 and p.202). Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_R$ be iid realizations of an estimator for the parameter θ. The empirical mean $\overline{\theta}$ is then given as

$$\overline{\theta} = \frac{1}{R}\sum_{i=1}^{R}\hat{\theta}_i. \tag{2.29}$$

The estimator of the bias is therefore calculated as

$$\widehat{b}(\theta, \hat{\theta}) = \overline{\theta} - \theta = \frac{1}{R}\sum_{i=1}^{R}\hat{\theta}_i - \theta. \tag{2.30}$$

The variance of the estimated bias $\hat{b}$ is given as

$$\begin{aligned} Var[\hat{b}(\theta, \hat{\theta})] &= Var[\overline{\theta} - \theta] = Var[\overline{\theta}] \\ &= Var\left[\frac{1}{R}\sum_{i=1}^{R}\hat{\theta}_i\right] = \frac{1}{R^2}\sum_{i=1}^{R}Var[\hat{\theta}_i] \\ &= \frac{1}{R}Var[\hat{\theta}], \end{aligned} \tag{2.31}$$

where the last step results from the iid property of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_R$. An unbiased and consistent estimator for $Var[\hat{\theta}]$ is the sample variance s^2 (see e.g. Georgii (2002), p.194 and p.202) , which is specified by

$$s^2 := \frac{1}{R-1}\sum_{i=1}^{R}(\hat{\theta}_i - \overline{\theta})^2. \tag{2.32}$$

With (2.32) we have the following estimator for the variance of the estimated bias:

$$\widehat{Var}[\hat{b}(\theta, \hat{\theta}] = \frac{1}{R}s^2 = \frac{1}{R}\left(\frac{1}{R-1}\sum_{i=1}^{R}(\hat{\theta}_i - \overline{\theta})^2\right). \tag{2.33}$$

As we already have calculated the sample variance and the estimator of the bias, we use version (2.28) of the mean squared error to get an estimator for the mean squared error. By using the sample variance and the estimator of the bias we receive the estimator of the mean squared error $\widehat{mse}$ as follows:

$$\widehat{mse}(\theta, \hat{\theta}) = s^2 + \widehat{b}(\theta, \hat{\theta})^2. \tag{2.34}$$

We also want to compute the variance of the mean squared error. Therefore, we use version (2.27) of the mean squared error with the estimator

$$\widehat{mse}^*(\theta, \hat{\theta}) = \frac{1}{R}\sum_{i=1}^{R}(\hat{\theta}_i - \theta)^2. \tag{2.35}$$

We then can calculate the variance of the estimator for the mean squared error $\widehat{mse}^*$:

$$
\begin{aligned}
Var[\widehat{mse}^*(\theta,\hat{\theta})] &= Var\left[\frac{1}{R}\sum_{i=1}^{R}(\hat{\theta}_i-\theta)^2\right]\\
&= \frac{1}{R^2}Var\left[\sum_{i=1}^{R}(\hat{\theta}_i-\theta)^2\right]\\
&= \frac{R}{R^2}Var[(\hat{\theta}-\theta)^2]\\
&= \frac{1}{R}Var[\hat{\theta}^2-2\hat{\theta}\theta+\theta^2]\\
&= \frac{1}{R}Var[\hat{\theta}^2-2\hat{\theta}\theta]\\
&= \frac{1}{R}\left(E[(\hat{\theta}^2-2\hat{\theta}\theta)^2]-\left(E[\hat{\theta}^2-2\hat{\theta}\theta]\right)^2\right)\\
&= \frac{1}{R}\left(E[\hat{\theta}^4-4\hat{\theta}^3\theta+4\hat{\theta}^2\theta^2]-\left(E[\hat{\theta}^2]-2E[\hat{\theta}]\theta\right)^2\right)\\
&= \frac{1}{R}\left(E[\hat{\theta}^4]-4E[\hat{\theta}^3]\theta+4E[\hat{\theta}^2]\theta^2-\left(E[\hat{\theta}^2]\right)^2\right.\\
&\quad \left.-4E[\hat{\theta}^2]E[\hat{\theta}]\theta+4E[\hat{\theta}]^2\theta^2\right). \qquad (2.36)
\end{aligned}
$$

In order to get an estimator for the variance of the mean squared error, we substitute the moments $E[\hat{\theta}^k]$, $k=1,2,3,4$, by their estimators, the empirical moments $m_k := \frac{1}{R}\sum_{i=1}^{R}\theta_i^k$. This results in the following estimator for the variance of the mean squared error:

$$
\widehat{Var}(mse(\theta,\hat{\theta})) = \frac{1}{R}\left(m_4-4\theta m_3+4\theta^2 m_2-m_2^2-4\theta m_2 m_1+4\theta^2 m_1^2\right). \qquad (2.37)
$$

3 Gamma-Poisson Regression Model for Average Claim Size and Number of Claims

In this chapter we construct a joint regression model for the continuous varible of the average claim size and the discrete count variable of the number of claims of an insurance policy. As it is into doing, we assume a Gamma distribution for the continuous average claim size and a Poisson distribution for the number of claims. The detailed assumptions for the joint regression model, the Gamma-Poisson regression model, are specified in the next section.

3.1 Marginal Assumptions

Let $Y_{i1} \in \mathbb{R}^+$, $i = 1, 2, \ldots, n$, be independent continuous random variables and let be $Y_{i2} \in \mathbb{N}_0$, $i = 1, 2, \ldots, n$, independent discrete random variables. We suppose for the random variables the two following GLMs specified by

$$Y_{i1} \sim \quad Gamma(\mu_{i1}, \nu^2), \quad \text{with } \ln(\mu_{i1}) = \boldsymbol{x_i}'\boldsymbol{\alpha}, \tag{3.1}$$

$$Y_{i2} \sim \quad Poisson(\mu_{i2}), \quad \text{with } \ln(\mu_{i2}) = \ln(e_i) + \boldsymbol{z_i}'\boldsymbol{\beta}, \tag{3.2}$$

where $\boldsymbol{x}_i \in \mathbb{R}^p$ are the covariates of the continuous variable Y_{i1} and $\boldsymbol{z}_i \in \mathbb{R}^q$ are the covariates of the discrete variable Y_{i2}. In the Poisson GLM we use the offset $\ln(e_i)$, where e_i gives the known time length in which events occur. Recall that the density of the $Gamma(\mu_{i1}, \nu^2)$ distribution is given by (2.2) and the probability mass function of the $Poisson(\mu_{i2})$ distribution by (2.3). Further, we assume that the parameter ν is known and has not to be estimated in the joint regression model. Note that as ν^2 is identical with dispersion parameter of the Gamma distribution it can be specified in the marginal Gamma GLM. Additionally, we suppose that the two margins are not independent, i.e. the pair Y_{i1} and Y_{i2} is not independent for all i. In the next section we construct a joint distribution function for Y_{i1} and Y_{i2} taking into account the possible dependence and calculate the corresponding density function.

3.2 Joint Density Function

To construct the joint distribution function of Y_{i1} and Y_{i2} with the two marginal regression models given in (3.1) and (3.2), we adopt the Mixed Copula approach. Thereby, we use the Gaussian copula as copula function. Then, by applying Sklar's theorem, the joint distribution function of Y_{i1} and Y_{i2} has the form

$$\begin{aligned} F(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) &= C(u_{i1}, u_{i2}|\rho) &(3.3)\\ &= \Phi_2\{\Phi^{-1}(u_{i1}), \Phi^{-1}(u_{i2})|\Gamma\}, &(3.4)\end{aligned}$$

where $u_{i1} := G_1(y_{i1}|\mu_{i1}, \nu^2)$, $u_{i2} := G_2(y_{i2}|\mu_{i2})$. $\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ is the covariance matrix of the 2-dimensional normal distribution with the Pearson correlation ρ between the two normal scores $q_{i1} := \Phi^{-1}(u_{i1})$ and $q_{i2} := \Phi^{-1}(u_{i2})$.

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

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

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 $u_{i2}^- := G_2(y_{i2}-1|\mu_{i2})$.

To get an explicit form of the joint density of Y_{i1} and Y_{i2} we now compute and simplify the derivative $\frac{\partial}{\partial u_1}C(u_1, u_{i2}|\rho)$. Thereby, we use the abbreviation $q_1 := \Phi^{-1}(u_1)$.

$$\begin{aligned}
C_1'(u_1, u_{i2}|\rho) &= \frac{\partial}{\partial u_1}C(u_1, u_{i2}|\rho) \\
&= \frac{\partial}{\partial u_1}\frac{1}{2\pi\sqrt{|\det(\Gamma)|}}\int_{-\infty}^{q_1}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2}\boldsymbol{x}'\Gamma^{-1}\boldsymbol{x}\right\}d\boldsymbol{x} \\
&= \frac{1}{2\pi\sqrt{|\det(\Gamma)|}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}q_1, & x_2\end{pmatrix}\Gamma^{-1}\begin{pmatrix}q_1\\ x_2\end{pmatrix}\right\}dx_2 \\
&\quad\times\frac{\partial}{\partial u_1}\,q_1 \\
&= \frac{1}{2\pi\sqrt{|\det(\Gamma)|}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}q_1, & x_2\end{pmatrix}\Gamma^{-1}\begin{pmatrix}q_1\\ x_2\end{pmatrix}\right\}dx_2 \\
&\quad\times\sqrt{2\pi}\exp\left\{\frac{1}{2}q_1^2\right\} \\
&= \frac{1}{\sqrt{2\pi|\det(\Gamma)|}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}q_1, & x_2\end{pmatrix}\Gamma^{-1}\begin{pmatrix}q_1\\ x_2\end{pmatrix}+\frac{1}{2}q_1^2\right\}dx_2 \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\begin{pmatrix}q_1, & x_2\end{pmatrix}\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix}\begin{pmatrix}q_1\\ x_2\end{pmatrix}\right. \\
&\quad\left.+\frac{1}{2}q_1^2\right\}dx_2 \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(q_1^2-2q_1x_2\rho+x_2^2)+\frac{1}{2}q_1^2\right\}dx_2 \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[q_1^2-2q_1x_2\rho+x_2^2-(1-\rho^2)q_1^2\right]\right\}dx_2 \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{q_{i2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(q_1\rho-x_2)^2\right\}dx_2.
\end{aligned}$$

Using the transformation

$$\begin{aligned}
x_2 &= z\sqrt{1-\rho^2}+\rho q_1 \iff z = \frac{x_2-\rho q_1}{\sqrt{1-\rho^2}} \\
dx_2 &= \sqrt{1-\rho^2}dz
\end{aligned}$$

it follows that

$$
\begin{aligned}
C_1'(u_1, u_{i2}|\rho) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_{i2}-\rho q_1}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(\rho q_1 - z\sqrt{1-\rho^2} - \rho q_1)^2\right\} \\
&\quad \times \sqrt{1-\rho^2} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_{i2}-\rho q_1}{\sqrt{1-\rho^2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}(-z\sqrt{1-\rho^2})^2\right\} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_{i2}-\rho q_1}{\sqrt{1-\rho^2}}} \exp\left\{-\frac{1}{2}z^2\right\} dz \\
&= \Phi\left(\frac{q_{i2}-\rho q_1}{\sqrt{1-\rho^2}}\right) = \Phi\left(\frac{\Phi^{-1}(u_{i2}) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \\
&:= D_\rho(u_1, u_{i2}).
\end{aligned} \tag{3.5}
$$

Equivalently we get

$$
\begin{aligned}
C_1'(u_1, u_{i2}^-|\rho) &= \Phi\left(\frac{\Phi^{-1}(u_{i2}^-) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \\
&= D_\rho(u_1, u_{i2}^-).
\end{aligned} \tag{3.6}
$$

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

$$
D_\rho(u_1, u_{i2}^-) = \Phi\left(\frac{\Phi^{-1}(u_{i2}^-) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = \Phi\left(\frac{-\infty - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) = \Phi(-\infty) = 0. \tag{3.7}
$$

So the joint Gamma-Poisson-density function for Y_{I1} and Y_{I2} is given by

$$
f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho) = \begin{cases} g_1(y_{i1}|\mu_{i1}, \nu^2) D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})), & \text{if } y_{i2} = 0; \\ \\ g_1(y_{i1}|\mu_{i1}, \nu^2)[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) \\ \quad -D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))], & \text{if } y_{i2} \geq 1. \end{cases} \tag{3.8}
$$

This is a valid joint density, since $f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho)$ is normalized to 1. To show this, we first compute the following infinite sum.

$$
\begin{aligned}
& \lim_{n\to\infty} \sum_{y_{i2}=0}^{n} [D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))] \\
= \ & \lim_{n\to\infty} [-\underbrace{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(-1|\mu_{i2}))}_{=0 \text{ according to (3.7)}} + D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(n|\mu_{i2}))] \\
= \ & D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), 1) \\
= \ & \Phi\left(\frac{\Phi^{-1}(1) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \Phi\left(\frac{\infty - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \\
&= \Phi(\infty) = 1.
\end{aligned}
$$

With this result it is simple to prove that $f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2}, \rho)$ is normalized to 1.

$$
\begin{aligned}
& \int_0^\infty \sum_{y_{i2}=0}^{\infty} f(y_{i1}, y_{i2}|\mu_{i1}, \mu_{i2}, \rho) dy_{i1} \\
&= \int_0^\infty \sum_{y_{i2}=0}^{\infty} g_1(y_{i1}|\mu_{i1}, \nu^2) \left[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2}))\right. \\
& \left. -D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}-1|\mu_{i2}))\right] \, dy_{i1} \\
&= \int_0^\infty g_1(y_{i1}|\mu_{i1}, \nu^2) \sum_{y_{i2}=0}^{\infty} \left[D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2}))\right. \\
& \left. -D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}-1|\mu_{i2}))\right] dy_{i1} \\
&= \int_0^\infty g_1(y_{i1}|\mu_{i1}, \nu^2) \, dy_{i1} \\
&= 1.
\end{aligned}
$$

3.3 Log-Likelihood with Ascertainments

Further on in the paper, we want to calculate the MLEs for the regression parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the marginal GLMs (3.1) and (3.2) as well as for the correlation parameter ρ of the joint density (3.8) for a data set of a car insurance. As we can only observe an average claim size for a policy when the policy holder caused at minimum one claim, the data set contains only insurance policies reporting at minimum one claim during the observation period. That is why we use the log-likelihood conditional on at least one observed (ascertained) claim as basis for our inference. To calculate a MLE we need a twice differentiable likelihood function of the joint density function. As for $Y_{i2} = y_{i2} = 0$ some partial derivatives are infinite it is also necessary that we use the likelihood conditional on $Y_{i2} = y_{i2} \geq 1$.

Let $\boldsymbol{y} := \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}$ with $\boldsymbol{y}_i = \begin{pmatrix} y_{i1}, & y_{i2} \end{pmatrix}$ be observed pairs of Gamma-Poisson distributed random variables, where y_{i1} is the Gamma distributed margin and y_{i2} notes the Poisson distributed margin. Further let $\boldsymbol{\theta} := \begin{pmatrix} \boldsymbol{\alpha}', & \boldsymbol{\beta}', & \gamma \end{pmatrix}'$ be the unknown parameter vector, where γ is Fisher's z-transformation of ρ, i.e.

$$
\gamma = \frac{1}{2} \ln \frac{1+\rho}{1-\rho},
$$

with the reverse transformation given by

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

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 the to y_{i1} associated and the to y_{i2} associated covariate vectors including the intercept. Further let $\mathcal{J} := \{i|i = 1, \ldots, n; y_{i2} \geq 1\}$ be the index set of all observations with $y_{i2} \geq 1$. So the likelihood function conditional on $y_{i2} \geq 1$ is given by

$$L^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \forall i\ y_{i2} \geq 1) = \prod_{i \in \mathcal{J}} \frac{f(y_{i1}, y_{i2}|\mu_{i1}, \mu_{i2}, \rho)}{[1 - P(y_{i2} = 0)]}, \tag{3.9}$$

and therefore the conditional log-likelihood has the form

$$\begin{aligned} l^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \forall i\ y_{i2} \geq 1) &= \ln(L^c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \forall i\ y_{i2} \geq 1)) \\ &= -\sum_{i \in \mathcal{J}} \ln\{1 - g_2(0, \mu_{i2})\} + \sum_{i \in \mathcal{J}} \ln\{g_1(y_{i1}|\mu_{i1}, \nu^2)\} \\ &\quad + \sum_{i \in \mathcal{J}} \ln\{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2}|\mu_{i2})) \\ &\quad - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu^2), G_2(y_{i2} - 1|\mu_{i2}))\}, \end{aligned} \tag{3.10}$$

with $\mu_{i1} = e^{\boldsymbol{x}_i{}'\boldsymbol{\alpha}}$ and $\mu_{i2} = e^{\boldsymbol{z}_i{}'\boldsymbol{\beta}}$. Below we use $L^c(\boldsymbol{\theta})$ and $l^c(\boldsymbol{\theta})$ as abbreviations for the expression of the conditional likelihood (3.9) and the conditional log-likelihood (3.9).

4 Maximization by Parts Algorithm

Maximization by Parts (MBP) is a new fix-point algorithm to solve a score equation for the maximum likelihood estimator published in Song et al. (2005). For this method no second order derivatives of the full likelihood are necessary. The likelihood function is decomposed into two parts. The first part 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. Here, we use a variation of the MBP algorithm, which is also introduced in Song et al. (2005).[1]

4.1 Framework and General Formulation

Let $y_1, \ldots, y_n$ be independent observations of a parametric distribution with density $p_i(y_i|\boldsymbol{\theta})$, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta \subset \mathbb{R}^p$ with $p = p_1 + p_2$ and $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1},\ \boldsymbol{\theta}_2 \in \mathbb{R}^{p_2}$. The corresponding log-likelihood for $\boldsymbol{\theta}$ is then given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(p_i(y_i|\boldsymbol{\theta})) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta}).$$

Further, we consider the following decomposition of the log-likelihood

$$l(\boldsymbol{\theta}) = l_m(\boldsymbol{\theta}_1) + l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

The corresponding score equation then has the form

$$l'(\boldsymbol{\theta}) = \begin{pmatrix} l'_{m(1)}(\boldsymbol{\theta}_1) + l'_{d(1)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ l'_{d(2)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{pmatrix} = \mathbf{0},$$

with $l'_{m(1)}(\boldsymbol{\theta}_1) = \frac{\partial}{\partial \boldsymbol{\theta}_1} l_m(\boldsymbol{\theta}_1) \in \mathbb{R}^{p_1}$, $l'_{d(1)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{\partial}{\partial \boldsymbol{\theta}_1} l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{p_1}$ and $l'_{d(2)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{\partial}{\partial \boldsymbol{\theta}_2} l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{p_2}$.
We assume that the score equation $l'(\boldsymbol{\theta}) = \mathbf{0}$ has a unique solution, the MLE $\hat{\boldsymbol{\theta}}$. In addition we assume that $l'(\boldsymbol{\theta}) = \mathbf{0}$ is too complex to solve, since the second order derivative of the log-likelihood is difficult to calculate. Hence we want to avoid the determination of the second order derivative of the log-likelihood and choose $l_m(\boldsymbol{\theta}_1)$ and $l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in such a way that $l'_m(\boldsymbol{\theta}_1) = \mathbf{0}$ and $l'_{d(2)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$ for a fixed $\boldsymbol{\theta}_1$ are comparatively easy to solve.
We use the following variation of the MBP algorithm:

Algorithm 4.1 (Variation of the MBP)**.**

Step 1: *Obtain initial* $\boldsymbol{\theta}_{1,n}^1$ *as the solution of* $l'_{m(1)}(\boldsymbol{\theta}_1) = \mathbf{0}$*;*
obtain initial $\boldsymbol{\theta}_{2,n}^1$ *as the solution of* $l'_{d(2)}(\boldsymbol{\theta}_{1,n}^1, \boldsymbol{\theta}_2) = \mathbf{0}$.

[1] cp. Song et al. (2005), pp. 1149

Step k: *Get* $\boldsymbol{\theta}^k_{1,n}$ *as the solution of* $l'_{m(1)}(\boldsymbol{\theta}_1) = -l'_{d(1)}(\boldsymbol{\theta}^{k-1}_{1,n}, \boldsymbol{\theta}^{k-1}_{2,n})$*;*
get $\boldsymbol{\theta}^k_{2,n}$ *as the solution of* $l'_{d(2)}(\boldsymbol{\theta}^{k-1}_{1,n}, \boldsymbol{\theta}_2) = \mathbf{0}$*,* $k = 2, 3, \ldots$ *.*

We stop the algorithm, when $||\boldsymbol{\theta}^k_n - \boldsymbol{\theta}^{k-1}_n|| < \epsilon$ with $\boldsymbol{\theta}^k_n = ({\boldsymbol{\theta}^k_{1,n}}', {\boldsymbol{\theta}^k_{2,n}}')'$, where $||\cdot||$ denotes the Euclidean norm and $\epsilon \in \mathbb{R}$ is a small value, e.g. 10^{-6}. For the asymptotic properties of $\boldsymbol{\theta}^k_n = (\boldsymbol{\theta}^k_{1,n}, \boldsymbol{\theta}^k_{2,n})'$ we need to assume the following regularity conditions. Let $\mathcal{U}_0 = \{\boldsymbol{\theta} \in \Theta : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta\}$ be the neighborhood of the true parameter $\boldsymbol{\theta}_0 = ({\boldsymbol{\theta}_{1,0}}', {\boldsymbol{\theta}_{2,0}}')'$.

(MBP1) $\Theta \subset \mathbb{R}^p$ is an open interval.

(MBP2) For $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] \in \mathcal{U}_0$ the first and second derivatives of $l(\boldsymbol{\theta})$, $l_m(\boldsymbol{\theta}_1)$ and $l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ exist.

(MBP3) The Fisher information $\mathcal{I}(\boldsymbol{\theta}') := -n^{-1} E\left[\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\right]$ exists and is not equal to 0 for all $\boldsymbol{\theta}' \in \mathcal{U}_0$.

(MBP4) $l'_m(\cdot)$, $l'_{d(1)}(\cdot)$ and $l'_{d(2)}(\cdot)$ are unbiased estimation functions, i.e.
$E[l'_m(\boldsymbol{\theta}_1)] = 0$, $E[l'_{d(1)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)] = 0$ and $E[l'_{d(2)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)] = 0$ for all $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] \in \mathcal{U}_0$.

Then, according to Song et al. (2005)[2], the estimator $\boldsymbol{\theta}^k_n$ is consistent for $\boldsymbol{\theta}$ for each k. Moreover, conditional on the data, $\boldsymbol{\theta}^k_n$ will converge to the MLE $\hat{\boldsymbol{\theta}}$ as $k \to \infty$. Additionally, for each fixed k, $\boldsymbol{\theta}^k_n$ is asymptotically normal distributed. To specify the asymptotic distribution of $\boldsymbol{\theta}^k_n$ we introduce the following notations:

$$
\begin{aligned}
\mathcal{I}_{m(11)}(\boldsymbol{\theta}_0) &:= -n^{-1} E\left[\left.\frac{\partial^2 l_m(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial{\boldsymbol{\theta}_1}'}\right|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_{01}}\right] \in \mathbb{R}^{p1\times p1}, \\
\mathcal{I}_{d(11)}(\boldsymbol{\theta}_0) &:= -n^{-1} E\left[\left.\frac{\partial^2 l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_1\partial{\boldsymbol{\theta}_1}'}\right|_{[\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]=[\boldsymbol{\theta}_{01},\boldsymbol{\theta}_{02}]}\right] \in \mathbb{R}^{p1\times p1}, \\
\mathcal{I}_{d(12)}(\boldsymbol{\theta}_0) &:= -n^{-1} E\left[\left.\frac{\partial^2 l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_1\partial{\boldsymbol{\theta}_2}'}\right|_{[\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]=[\boldsymbol{\theta}_{01},\boldsymbol{\theta}_{02}]}\right] \in \mathbb{R}^{p1\times p2}, \\
\mathcal{I}_{d(21)}(\boldsymbol{\theta}_0) &:= -n^{-1} E\left[\left.\frac{\partial^2 l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2\partial{\boldsymbol{\theta}_1}'}\right|_{[\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]=[\boldsymbol{\theta}_{01},\boldsymbol{\theta}_{02}]}\right] \in \mathbb{R}^{p2\times p1}, \\
\mathcal{I}_{d(22)}(\boldsymbol{\theta}_0) &:= -n^{-1} E\left[\left.\frac{\partial^2 l_d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial\boldsymbol{\theta}_2\partial{\boldsymbol{\theta}_2}'}\right|_{[\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]=[\boldsymbol{\theta}_{01},\boldsymbol{\theta}_{02}]}\right] \in \mathbb{R}^{p2\times p2}.
\end{aligned}
$$

Further, let be

$$
\begin{aligned}
D(\boldsymbol{\theta}_0) &:= \begin{pmatrix} \mathcal{I}_{m(11)}(\boldsymbol{\theta}_0) & \mathbf{0}_{p1\times p2} \\ \mathbf{0}_{p2\times p1} & \mathcal{I}_{d(22)}(\boldsymbol{\theta}_0) \end{pmatrix} \in \mathbb{R}^{p\times p}, \\
T(\boldsymbol{\theta}_0) &:= -\begin{pmatrix} \mathcal{I}_{d(11)}(\boldsymbol{\theta}_0) & \mathcal{I}_{d(12)}(\boldsymbol{\theta}_0) \\ \mathcal{I}_{d(21)}(\boldsymbol{\theta}_0) & \mathbf{0}_{p2\times p2} \end{pmatrix} \in \mathbb{R}^{p\times p}, \\
L(\boldsymbol{\theta}_0) &:= -\begin{pmatrix} \mathbf{0}_{p1\times p1} & \mathbf{0}_{p1\times p2} \\ \{\mathcal{I}_{d(22)}(\boldsymbol{\theta}_0)\}^{-1}\mathcal{I}_{d(21)}(\boldsymbol{\theta}_0)\{\mathcal{I}_{m(11)}(\boldsymbol{\theta}_0)\}^{-1} & \mathbf{0}_{p2\times p2} \end{pmatrix} \in \mathbb{R}^{p\times p},
\end{aligned}
$$

[2] cp. Song et al. (2005), p. 1150

and $\Gamma(\boldsymbol{\theta}_0) := \{D(\boldsymbol{\theta}_0)\}^{-1}T(\boldsymbol{\theta}_0)$, where $\mathbf{0}_{p_1\times p_2}$ is a zero matrix of dimension $p_1 \times p_2$. I_p denotes the identity matrix of dimension $p \times p$ and

$$\Omega = \lim_{n\to\infty} n^{-1} \begin{pmatrix} E[l'_{m(1)}(\theta_{10})l'_{m(1)}(\theta_{10})'] & E[l'_{m(1)}(\theta_{10})l'_{d(2)}(\theta_0)'] & E[l'_{m(1)}(\theta_{10})l'_{d(1)}(\theta_0)'] & \mathbf{0}_{p_1\times p_2} \\ E[l'_{d(2)}(\theta_0)l'_{m(1)}(\theta_{10})'] & E[l'_{d(2)}(\theta_0)l'_{d(2)}(\theta_0)'] & E[l'_{d(2)}(\theta_{10})l'_{d(1)}(\theta_0)'] & \mathbf{0}_{p_2\times p_2} \\ E[l'_{d(1)}(\theta_0)l'_{m(1)}(\theta_{10})'] & E[l'_{d(1)}(\theta_0)l'_{d(2)}(\theta_0)'] & E[l'_{d(1)}(\theta_0)l'_{d(1)}(\theta_0)'] & \mathbf{0}_{p_1\times p_2} \\ \mathbf{0}_{p_2\times p_2} & \mathbf{0}_{p_2\times p_2} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_2} \end{pmatrix} \in \mathbb{R}^{2p\times 2p}$$

is the asymptotic variance matrix of the estimating function

$$n^{-1/2}[l'_{m(1)}(\boldsymbol{\theta}_{01}), l'_{d(2)}(\boldsymbol{\theta}_0), l'_{d(1)}(\boldsymbol{\theta}_0), \mathbf{0}_{p_2}]',$$

where $\mathbf{0}_{p_2}$ is included to match dimensions.
Under the regularity conditions $\boldsymbol{\theta}_n^k$ is asymptotically normal distributed with mean $\boldsymbol{\theta}_0$ and variance $n^{-1}\Sigma_k$, with $\Sigma_k = B_k\Omega B_k'$, $B_k = [B_{k1}, B_{k2}] \in \mathbb{R}^{2p\times p}$, where

$$B_{k1} = [I_p - \{\Gamma(\boldsymbol{\theta}_0)\}^k]\{\mathcal{I}(\boldsymbol{\theta}_0)\}^{-1} + \{\Gamma(\boldsymbol{\theta}_0)\}^{k-1}L(\boldsymbol{\theta}_0) \;\in \mathbb{R}^{p\times p},$$

and

$$B_{k2} = \{I_p - \{\Gamma(\boldsymbol{\theta}_0)\}^{k-1}\}\{\mathcal{I}(\boldsymbol{\theta}_0)\}^{-1} \;\in \mathbb{R}^{p\times p}.$$

Moreover, if $\{\Gamma(\boldsymbol{\theta}_0)\}^k \to \mathbf{0}$ as $k \to \infty$, then $\Sigma_k \to \{\mathcal{I}(\boldsymbol{\theta}_0)\}^{-1}$.

4.2 Application to Gamma-Poisson Regression Model

Now we apply the MBP algorithm to maximize our likelihood (3.9) and get a MLE for $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \gamma)' \in \mathbb{R}^{p+q+1}$. Here $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}', \boldsymbol{\beta}')' \in \mathbb{R}^{p+q}$ and $\boldsymbol{\theta}_2 = \gamma \in \mathbb{R}$. To decompose the conditional log-likelihood (3.10) as we need it for using the MBP algorithm, we rewrite the conditional likelihood (3.9) as

$$L^c(\boldsymbol{\theta}) = L^c_m(\boldsymbol{\theta}_1)\; L^c_d(\boldsymbol{\theta}_1, \gamma). \tag{4.1}$$

This leads to the decomposition

$$l^c(\boldsymbol{\theta}) = \ln(L^c(\boldsymbol{\theta})) = l^c_m(\boldsymbol{\theta}_1) + l^c_d(\boldsymbol{\theta}_1, \gamma) \tag{4.2}$$

for the conditional log-likelihood with $\boldsymbol{\theta}_1 = [\boldsymbol{\alpha}', \boldsymbol{\beta}']'$, 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

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

The part $l^c_m(\boldsymbol{\theta}_1)$ 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, the part $l^c_d(\boldsymbol{\theta}_1, \gamma)$ contains the copula part of the conditional log-likelihood and depends on the correlation parameter ρ, respectively γ.
Further, we need the score functions of the parts $l^c_m(\boldsymbol{\theta}_1)$ and $l^c_d(\boldsymbol{\theta}_1, \gamma)$ in order to use the MBP algorithm for the MLE calculation.

4.2.1 Score Functions

In this section we calculate the partial derivatives of the log-likelihood (3.9) for our parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and γ. Using the decomposition (4.2) we obviously can write the first derivative of the log-likelihood as

$$\frac{\partial}{\partial \boldsymbol{\theta}}\ l(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\ l_m^c(\boldsymbol{\theta}_1) + \frac{\partial}{\partial \boldsymbol{\theta}}\ l_d^c(\boldsymbol{\theta}_1, \gamma).$$

In the following we compute the partial derivatives of $l_m^c(\boldsymbol{\theta}_1)$ and $l_d^c(\boldsymbol{\theta}_1, \gamma)$. For a better handling we use the following abbreviations:

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

Score Function of $l_m^c(\boldsymbol{\theta}_1)$

First, we calculate the score function of the marginal part

$$\frac{\partial}{\partial \boldsymbol{\theta}_1} l_m^c(\boldsymbol{\theta}_1) = \left(\tfrac{\partial}{\partial \boldsymbol{\alpha}} l_m^c(\boldsymbol{\theta}_1), \tfrac{\partial}{\partial \boldsymbol{\beta}} l_m^c(\boldsymbol{\theta}_1), 0\right)'. \tag{4.3}$$

We now have to figure out the partial derivatives. The derivative with respect to the regression parameter $\boldsymbol{\alpha}$ of the marginal gamma model is given by

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\alpha}} l_m^c(\boldsymbol{\theta}_1) &= \sum_{i \in \mathcal{J}} \frac{\partial \ln(g_1(y_{i1}|\mu_{i1}, \nu^2))}{\partial \boldsymbol{\alpha}} \\
&= \sum_{i \in \mathcal{J}} \frac{1}{g_1(y_{i1}|\mu_{i1}, \nu^2)}\ \frac{\partial g_1(y_{i1}|\mu_{i1}, \nu^2)}{\partial \mu_{i1}} \frac{\partial \mu_{i1}}{\partial \boldsymbol{\alpha}} \\
&= \sum_{i \in \mathcal{J}} \frac{1}{g_1(y_{i1}|\mu_{i1}, \nu^2)}\ \frac{\partial g_1(y_{i1}|\mu_{i1}, \nu^2)}{\partial \mu_{i1}}\ \mu_{i1}\ \boldsymbol{x}_i,
\end{aligned} \tag{4.4}$$

with

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

$$
\begin{aligned}
&= \frac{1}{\Gamma(1/\nu^2)}\,\frac{1}{y_{i1}}\left(\frac{y_{i1}}{\mu_{i1}\nu^2}\right)^{1/\nu^2} e^{-\frac{y_{i1}}{\mu_{i1}\nu^2}}\left(\frac{y_{i1}}{\mu_{i1}^2\nu^2}-\frac{1}{\mu_{i1}\nu^2}\right)\\
&= \frac{1}{\mu_{i1}^2\nu^2}\; g_1(y_{i1}|\mu_{i1},\nu^2)\;(y_{i1}-\mu_{i1}). \qquad (4.5)
\end{aligned}
$$

This result leads us to

$$
\frac{\partial}{\partial \boldsymbol{\alpha}} l_m^c(\boldsymbol{\theta}_1) = \frac{1}{\nu^2}\sum_{i\in\mathcal{J}} x_i\;\mu_{i1}^{-1}(y_{i1}-\mu_{i1}). \qquad (4.6)
$$

The other derivative with respect to the regression parameter $\boldsymbol{\beta}$ of the marginal Poisson model is given by

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} l_m^c(\boldsymbol{\theta}_1) &= -\sum_{i\in\mathcal{J}} \frac{\partial \ln(1-g_2(0|\mu_{i2}))}{\partial \boldsymbol{\beta}}\\
&= -\sum_{i\in\mathcal{J}} \frac{1}{1-e^{-\mu_{i2}}}\frac{\partial(1-e^{-\mu_{i2}})}{\partial \mu_{i2}}\;\frac{\partial \mu_{i2}}{\partial \boldsymbol{\beta}}\\
&= -\sum_{i\in\mathcal{J}} \frac{e^{-\mu_{i2}}}{1-e^{-\mu_{i2}}}\;\mu_{i2}\boldsymbol{z}_i\\
&= -\sum_{i\in\mathcal{J}} \boldsymbol{z}_i\frac{\mu_{i2}}{e^{\mu_{i2}}-1}. \qquad (4.7)
\end{aligned}
$$

Score Function of $l_d^c(\boldsymbol{\theta}_1,\gamma)$

We also need the score function of the dependency part $l_d^c(\boldsymbol{\theta}_1,\gamma)$. Thus we have to compute

$$
\frac{\partial\; l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial \boldsymbol{\theta}} = \left(\tfrac{\partial}{\partial \boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1,\gamma),\, \tfrac{\partial}{\partial \boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1,\gamma),\, \tfrac{\partial}{\partial \gamma} l_d^c(\boldsymbol{\theta}_1,\gamma)\right)'. \qquad (4.8)
$$

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

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1,\gamma) &= \frac{\partial}{\partial \boldsymbol{\alpha}}\left\{ln[D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)]\right\}\\
&= \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)}\left(\frac{\partial D_\rho(G_{i1},G_{i2})}{\partial G_{i1}}-\right.\\
&\qquad \left.\frac{\partial D_\rho(G_{i1},G_{i2}^-)}{\partial G_{i1}}\right)\frac{\partial G_{i1}}{\partial \boldsymbol{\alpha}}, \qquad (4.9)
\end{aligned}
$$

where the derivative of the copula with respect to its first argument is calculated as

$$
\begin{aligned}
\frac{\partial D_\rho(G_{i1},\cdot)}{\partial G_{i1}} &= \frac{\partial}{\partial G_{i1}}\Phi\left(\frac{\Phi^{-1}(\cdot)-\rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\\
&= \phi\left(\frac{\Phi^{-1}(\cdot)-\rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\frac{-\rho}{\sqrt{1-\rho^2}}\;\frac{1}{\phi(\Phi^{-1}(G_{i1}))}, \qquad (4.10)
\end{aligned}
$$

since $(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}$. The derivative of the gamma cdf with respect to the regression parameter $\boldsymbol{\alpha}$ is

$$
\frac{\partial G_{i1}}{\partial \boldsymbol{\alpha}} = \frac{\partial G_{i1}}{\partial \mu_{i1}}\;\frac{\partial \mu_{i1}}{\partial \boldsymbol{\alpha}} = \frac{\partial G_{i1}}{\partial \mu_{i1}}\;\mu_{i1}\;\boldsymbol{x}_i, \qquad (4.11)
$$

with

$$
\begin{aligned}
\frac{\partial G_{i1}}{\partial \mu_{i1}} &= \frac{\partial}{\partial \mu_{i1}} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \\
&= \int_0^{y_{i1}} \frac{\partial g_1(y|\mu_{i1}, \nu^2)}{\partial \mu_{i1}}\ dy \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ (y - \mu_{i1})\ dy \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} y\ g_1(y|\mu_{i1}, \nu^2)\ dy - \mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} \frac{1}{\Gamma(\frac{1}{\nu^2})} \left(\frac{1}{\mu_{i1}\nu^2} \right)^{1/\nu^2} y^{1/\nu^2}\ e^{-\frac{1}{\mu_{i1}\nu^2} y}\ dy \right. \\
&\quad \left. -\mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} \frac{1}{\frac{1}{\nu^2}\Gamma(\frac{1}{\nu^2})} \left(\frac{1}{\mu_{i1}\nu^2} \right)^{1/\nu^2+1} y^{1/\nu^2}\ e^{-\frac{1}{\mu_{i1}\nu^2} y}\ \frac{\mu_{i1}\nu^2}{\nu^2}\ dy \right. \\
&\quad \left. -\mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\int_0^{y_{i1}} \frac{1}{\Gamma(\frac{1}{\nu^2}+1)} \left(\frac{1}{\mu_{i1}\nu^2} \right)^{1/\nu^2+1} y^{1/\nu^2}\ e^{-\frac{1}{\mu_{i1}\nu^2} y}\ \mu_{i1}\ dy \right. \\
&\quad \left. -\mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} \left(\mu_{i1} \int_0^{y_{i1}} g_1^*(y|\mu_{i1}, \nu^2)\ dy - \mu_{i1} \int_0^{y_{i1}} g_1(y|\mu_{i1}, \nu^2)\ dy \right) \\
&= \frac{1}{\mu_{i1}^2 \nu^2} (\mu_{i1}\ G_{i1}^* - \mu_{i1}\ G_{i1}) \\
&= \frac{1}{\mu_{i1} \nu^2} (G_{i1}^* - G_{i1}). \qquad (4.12)
\end{aligned}
$$

Here $g_1^*(\cdot)$ is the density function and $G_1^*(\cdot)$ is the cdf of a $Gamma(a+1, b)$-distribution with $a = \frac{1}{\nu^2}$ and $b = \frac{1}{\mu_{i1}\nu^2}$. Composing all these parts we receive

$$
\frac{\partial}{\partial \boldsymbol{\alpha}} l_d^c(\boldsymbol{\theta}_1, \gamma) = \sum_{i \in \mathcal{J}} \frac{d_\rho(G_{i1}, G_{i2}) - d_\rho(G_{i1}, G_{i2}^-)}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)}\ \frac{G_{i1}^* - G_{i1}}{\phi(\Phi^{-1}(G_{i1}))}\ \frac{-\rho}{\sqrt{1-\rho^2}}\ \boldsymbol{x}_i. \qquad (4.13)
$$

Now we have a closer look at the partial derivative with respect to $\boldsymbol{\beta}$:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1, \gamma) &= \sum_{i \in \mathcal{J}} \frac{\partial}{\partial \boldsymbol{\beta}} \left\{ ln[D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)] \right\} \\
&= \sum_{i \in \mathcal{J}} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \left(\frac{\partial D_\rho(G_{i1}, G_{i2})}{\partial G_{i2}}\ \frac{\partial G_{i2}}{\partial \boldsymbol{\beta}} \right. \\
&\quad \left. - \frac{\partial D_\rho(G_{i1}, G_{i2}^-)}{\partial G_{i2}^-}\ \frac{\partial G_{i2}^-}{\partial \boldsymbol{\beta}} \right). \qquad (4.14)
\end{aligned}
$$

In detail we have to compute all the derivatives, namely

$$\begin{aligned}\frac{\partial D_\rho(\cdot, G_{i2})}{\partial G_{i2}} &= \frac{\partial}{\partial G_{i2}} \Phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(\cdot)}{\sqrt{1-\rho^2}}\right) \\ &= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(\cdot)}{\sqrt{1-\rho^2}}\right) \frac{1}{\sqrt{1-\rho^2}} \; \frac{1}{\phi(\Phi^{-1}(G_{i2}))}, \end{aligned} \tag{4.15}$$

$$\frac{\partial G_{i2}}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}}{\partial \mu_{i2}} \frac{\partial \mu_{i2}}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}}{\partial \mu_{i2}} \; \mu_{i2} \; \boldsymbol{z}_i, \tag{4.16}$$

with

$$\begin{aligned}\frac{\partial G_{i2}}{\partial \mu_{i2}} &= \frac{\partial}{\partial \mu_{i2}} [\sum_{k=0}^{y_{i2}} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}}] \\ &= \frac{\partial}{\partial \mu_{i2}} \; [e^{-\mu_{i2}}] \; + \; \sum_{k=1}^{y_{i2}} \frac{\partial}{\partial \mu_{i2}} \; [\frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}}] \\ &= -e^{-\mu_{i2}} \; + \; \sum_{k=1}^{y_{i2}} \left(\frac{1}{(k-1)!} \; \mu_{i2}^{k-1} \; e^{-\mu_{i2}} - \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} \right) \\ &= -e^{-\mu_{i2}} \; + \; \sum_{k=0}^{y_{i2}-1} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} - \sum_{k=1}^{y_{i2}} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} \\ &= -e^{-\mu_{i2}} + e^{-\mu_{i2}} \; + \; \sum_{k=1}^{y_{i2}-1} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} - \sum_{k=1}^{y_{i2}} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} \\ &= -\frac{1}{y_{i2}!} \; \mu_{i2}^{y_{i2}} \; e^{-\mu_{i2}} \\ &= -g_2(y_{i2}|\mu_{i2}). \end{aligned} \tag{4.17}$$

Similarly, we derive for G_{i2}^-

$$\frac{\partial G_{i2}^-}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}^-}{\partial \mu_{i2}} \frac{\partial \mu_{i2}}{\partial \boldsymbol{\beta}} = \frac{\partial G_{i2}^-}{\partial \mu_{i2}} \; \mu_{i2} \; \boldsymbol{z}_i, \tag{4.18}$$

with

$$\begin{aligned}\frac{\partial G_{i2}^-}{\partial \mu_{i2}} &= \frac{\partial}{\partial \mu_{i2}} \sum_{k=0}^{y_{i2}-1} \frac{1}{k!} \; \mu_{i2}^k \; e^{-\mu_{i2}} \\ &= -g_2(y_{i2}-1|\mu_{i2}). \end{aligned} \tag{4.19}$$

All these parts together lead to

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}} l_d^c(\boldsymbol{\theta}_1, \gamma) &= \sum_{i \in \mathcal{J}} \frac{1}{D_\rho(G_{i1}, G_{i2}) - D_\rho(G_{i1}, G_{i2}^-)} \left[d_\rho(G_{i1}, Gi2^-) \frac{g_2(y_{i2}-1|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}^-))} \right. \\ & \left. -d_\rho(G_{i1}, Gi2) \frac{g_2(y_{i2}|\mu_{i2})}{\phi(\Phi^{-1}(G_{i2}))} \right] \frac{\mu_{i2}}{\sqrt{1-\rho^2}} \boldsymbol{z}_i. \end{aligned} \tag{4.20}$$

Finally, we figure out the derivative with respect to γ:

$$\begin{aligned}
\frac{\partial}{\partial\gamma} l_d^c(\boldsymbol{\theta}_1,\gamma) &= \frac{\partial}{\partial\gamma}\left\{ln[D_\rho(G_{i1},G_{i2}) - D_\rho(G_{i1},G_{i2}^-)]\right\} \\
&= \frac{1}{D_\rho(G_{i1},G_{i2}) - D_\rho(G_{i1},G_{i2}^-)}\left(\frac{\partial D_\rho(G_{i1},G_{i2})}{\partial\rho}\right. \\
&\qquad \left. -\frac{\partial D_\rho(G_{i1},G_{i2}^-)}{\partial\rho}\right)\frac{\partial\rho}{\partial\gamma}. \qquad (4.21)
\end{aligned}$$

We now compute each of the partial derivatives. The first is

$$\begin{aligned}
\frac{\partial D_\rho(G_{i1},G_{i2})}{\partial\rho} &= \frac{\partial}{\partial\rho}\Phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\frac{\partial}{\partial\rho}\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&\quad \times\left(\frac{-\Phi^{-1}(G_{i1})\sqrt{1-\rho^2}}{1-\rho^2}\right. \\
&\quad \left. +\frac{\rho(1-\rho^2)^{-1/2}(\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1}))}{1-\rho^2}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right) \\
&\quad \times\left(\frac{\rho\Phi^{-1}(G_{i2})(1-\rho^2)^{-1/2}}{1-\rho^2}\right. \\
&\quad \left. -\frac{\Phi^{-1}(G_{i1})(1-\rho^2)^{-1/2}\left((1-\rho^2)+\rho^2\right)}{1-\rho^2}\right) \\
&= \phi\left(\frac{\Phi^{-1}(G_{i2}) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\frac{\rho\Phi^{-1}(G_{i2}) - \Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}. \qquad (4.22)
\end{aligned}$$

Similarly, we get

$$\frac{\partial D_\rho(G_{i1},G_{i2}^-)}{\partial\rho} = \phi\left(\frac{\Phi^{-1}(G_{i2}^-) - \rho\Phi^{-1}(G_{i1})}{\sqrt{1-\rho^2}}\right)\frac{\rho\Phi^{-1}(G_{i2}^-) - \Phi^{-1}(G_{i1})}{(1-\rho^2)^{3/2}}. \qquad (4.23)$$

The last one is given by

$$\begin{aligned}
\frac{\partial\rho}{\partial\gamma} &= \frac{\partial}{\partial\gamma}\frac{e^{2\gamma}-1}{e^{2\gamma}+1} \\
&= \frac{2e^{2\gamma}(e^{2\gamma}+1) - 2e^{2\gamma}(e^{2\gamma}-1)}{(e^{2\gamma}+1)^2} \\
&= \frac{2e^{4\gamma}+2e^{2\gamma}-2e^{4\gamma}+2e^{2\gamma}\ -e^{4\gamma}+e^{4\gamma}+1-1}{(e^{2\gamma}+1)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(e^{2\gamma}+1)^2-(e^{2\gamma}-1)^2}{(e^{2\gamma}+1)^2} \\
&= 1-\rho^2. \qquad (4.24)
\end{aligned}
$$

Combining all these parts again we finally get

$$
\begin{aligned}
\frac{\partial}{\partial\gamma} l_d^c(\boldsymbol{\theta}_1,\gamma) &= \sum_{i\in\mathcal{J}} \frac{1}{D_\rho(G_{i1},G_{i2})-D_\rho(G_{i1},G_{i2}^-)} \left\{ d_\rho(G_{i1},G_{i2})[\rho\Phi^{-1}(G_{i2})-\Phi^{-1}(G_{i1})] \right. \\
&\quad \left. -d_\rho(G_{i1},G_{i2}^-)[\rho\Phi^{-1}(G_{i2}^-)-\Phi^{-1}(G_{i1})] \right\} \frac{1}{\sqrt{1-\rho^2}}. \qquad (4.25)
\end{aligned}
$$

4.2.2 Applied MBP Algorithm

Now we have all necessary components for using the variation of the MBP algorithm in order to calculate the MLE of $\boldsymbol{\theta}$. Our MBP algorithm to maximize our log-likelihood (3.9) proceeds as follows:

Algorithm 4.2 (Applied Variation of the MBP)**.**

Step 0: *Calculate the initial values:*

- **Step 0.1:** *Calculate initial value* $\boldsymbol{\theta}_1^0 = [\widehat{\boldsymbol{\alpha}_I}', \widehat{\boldsymbol{\beta}_I}']'$ *for* $\boldsymbol{\theta}_1$.
- **Step 0.2:** *Calculate the initial value* γ^0 *for* γ *by solving* $\frac{\partial l_d^c(\boldsymbol{\theta}_1^0,\gamma)}{\partial\gamma} = 0$ *with the bisection method.*

Step k (k $= 1,2,3,\ldots$**):** *First we update* $\boldsymbol{\theta}_1$ *by one step of the Fisher scoring method (2.24), i.e.*

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

Then, by solving $\frac{\partial l_d^c(\boldsymbol{\theta}^k,\gamma)}{\partial\gamma} = 0$ *with the bisection method (2.1), we receive the new* γ^k.

Thereby,

$$
\begin{aligned}
\mathcal{I}_m^c(\boldsymbol{\theta}_1^{k-1}) &:= -n^{-1} E\left[\left. \frac{\partial^2 l_m^c[\boldsymbol{\theta}_1]}{\partial\boldsymbol{\theta}_1 \partial\boldsymbol{\theta}_1'} \right|_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_1^{k-1}} \right] \\
&= n^{-1} \begin{pmatrix} \frac{1}{\nu^2}\sum_{i\in\mathcal{J}} \boldsymbol{x}_i \frac{y_{i1}}{\mu_{i1}} \boldsymbol{x}_i' & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \sum_{i\in\mathcal{J}} \boldsymbol{z}_i \frac{\mu_{i2}(e^{\mu_{i2}}-1-\mu_{i2}e^{\mu_{i2}})}{(e^{\mu_{i2}}-1)^2} \boldsymbol{z}_i' \end{pmatrix} \qquad (4.26)
\end{aligned}
$$

is the Fisher information of the marginal part l_m^c of the log likelihood at point $\boldsymbol{\theta}_1^{k-1}$. By assuming independency between the marginal Gamma GLM (3.1) and the marginal Poisson GLM (3.2) the initial value vector $\widehat{\boldsymbol{\alpha}_I}$ is the MLE of the regression parameter from the marginal Gamma GLM and $\widehat{\boldsymbol{\beta}_I}$ is the MLE of the regression parameter form the marginal Poisson GLM. When the convergence criterion (e.g. $||\boldsymbol{\theta}^k - \boldsymbol{\theta}^{k-1}||_\infty < 10^{-6}$) is met, the algorithm stops and puts out

an approximation of the MLE for parameter $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \gamma]'$. Since γ is a scalar, $\frac{\partial l_d^c(\boldsymbol{\theta}^k, \gamma)}{\partial \gamma} = 0$ is a one-dimensional search and the bisection method (2.1) works efficiently.
A problem in the application of the MBP algorithm is to ensure convergence. The crux here is the Fisher scoring update of $\boldsymbol{\theta}_1$. Often, there is not enough information in the marginal part of the conditional log-likelihood l_m^c and so the elements of $\mathcal{I}_m^c(\boldsymbol{\theta}_1^k)$ are not large enough to force convergence in the Fisher scoring step. In order to master this problem we try to pack as much information as possible in the first part $l_m^c(\boldsymbol{\theta}_1)$ of the decomposition. According to Song et al. (2005)[3] we follow the idea of Liang and Zeger (1986)[4] and expand our likelihood (3.9) 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 $\Sigma_w = \begin{pmatrix} 1 & \rho_w \\ \rho_w & 1 \end{pmatrix}$ to get the information deficit under control. 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 likelihood (3.9) by the likelihood of the marginal Poisson-GLM (3.2). 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 has the form

$$l^*(\boldsymbol{\theta}) := \underbrace{\ln(L_m^*(\boldsymbol{\theta}_1))}_{=:l_m^*(\boldsymbol{\theta}_1)} + \underbrace{\ln(L_d^*(\boldsymbol{\theta}_1, \gamma))}_{=:l_d^*(\boldsymbol{\theta}_1, \gamma)}, \tag{4.27}$$

with

$$\begin{aligned} l_m^*(\boldsymbol{\theta}_1) &:= l_m^c(\boldsymbol{\theta}_1) + \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})) + \ln(L_w(\boldsymbol{\theta}_1)) \\ &= l_m^c(\boldsymbol{\theta}_1) + \sum_{i \in \mathcal{J}} \ln(g_2(y_{i2}|\mu_{i2})) \\ &\quad -\frac{1}{2} \sum_{i \in \mathcal{J}} \left[\ln(1 - \rho_w^2) + (\boldsymbol{y}_i - \boldsymbol{\mu})'\Sigma_w^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right] \end{aligned} \tag{4.28}$$

and

$$\begin{aligned} l_d^*(\boldsymbol{\theta}1, \gamma) &:= l_d^c(\boldsymbol{\theta}_1) - \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})) - \ln(L_w(\boldsymbol{\theta}_1)) \\ &= l_d^c(\boldsymbol{\theta}_1) - \sum_{i \in \mathcal{J}} \ln(g_2(y_{i2}|\mu_{i2})) \\ &\quad +\frac{1}{2} \sum_{i \in \mathcal{J}} \left[\ln(1 - \rho_w^2) + (\boldsymbol{y}_i - \boldsymbol{\mu})'\Sigma_w^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})\right]. \end{aligned} \tag{4.29}$$

[3]cp. Song et al. (2005), p.1156
[4]cp. Liang and Zeger (1986), p.15

Now we want to see how the Fisher information of the $l_m^*(\boldsymbol{\theta}_1)$ looks like. Therefore, we begin with the calculation of the first derivative of $l_m^*(\boldsymbol{\theta}1)$. According to equation (4.28) it follows

$$\frac{\partial}{\partial \boldsymbol{\theta}} l_m^*(\boldsymbol{\theta}_1) = \frac{\partial}{\partial \boldsymbol{\theta}_1} l_m^c(\boldsymbol{\theta}_1) + \frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2})) + \frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(L_w(\boldsymbol{\theta}_1)).$$

We already derived $\frac{\partial}{\partial \boldsymbol{\theta}_1} l_m^c(\boldsymbol{\theta}_1)$ in (4.3), so we just have to figure out $\frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2}))$ and $\frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(L_w(\boldsymbol{\theta}_1))$. The derivative of the log-likelihood of the Poisson-GLM according to θ_1 is calculated as

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(\prod_{i \in \mathcal{J}} g_2(y_{i2}|\mu_{i2}) &= \sum_{i \in \mathcal{J}} \frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(g_2(y_{i2}|\mu_{i2}) \\
&= \left(0, \sum_{i \in \mathcal{J}} \frac{\partial}{\partial \boldsymbol{\beta}} \ln(g_2(y_{i2}|\mu_{i2}))\right)' \\
&= \left(0, \sum_{i \in \mathcal{J}} \frac{1}{g_2(y_{i2}|\mu_{i2})} \frac{\partial g_2(y_{i2}|\mu_{i2})}{\partial \mu_{i2}} \mu_{i2} z_i\right)', \qquad (4.30)
\end{aligned}$$

with

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

Yet we still need the derivative of $\ln(L_w(\boldsymbol{\theta}_1))$ being calculated by

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}_1} \ln(L_w(\boldsymbol{\theta}_1)) &= -\frac{1}{2} \sum_{i \in \mathcal{J}} \frac{\partial}{\partial \boldsymbol{\theta}_1} \left[\ln(1-\rho_w^2) + (\boldsymbol{y}_i - \boldsymbol{\mu})' \Sigma_w^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu})\right] \\
&= \sum_{i \in \mathcal{J}} \begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix} \begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix} \Sigma_w^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}). \qquad (4.31)
\end{aligned}$$

The results (4.6), (4.7), (4.30) and (4.31) lead us to the following first derivative of $l_m^* \boldsymbol{\theta}_1)$ with respect to $\boldsymbol{\theta}_1$:

$$\begin{aligned}
\frac{\partial l_m^*(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} &= \begin{pmatrix} \frac{1}{\nu^2} \sum_{i \in \mathcal{J}} \boldsymbol{x}_i\, \mu_{i1}^{-1} (y_{i1} - \mu_{i1}) \\ -\sum_{i \in \mathcal{J}} \boldsymbol{z}_i \frac{\mu_{i2}}{e^{\mu_{i2}} - 1} \end{pmatrix} + \begin{pmatrix} 0 \\ \sum_{i \in \mathcal{J}} \boldsymbol{z}_i (\boldsymbol{y}_{i2} - \mu_{i2}) \end{pmatrix} \\
&\quad + \sum_{i \in \mathcal{J}} \begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix} \begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix} \Sigma_w^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}). \qquad (4.32)
\end{aligned}$$

To get the Fisher information of $l_m^*(\cdot)$ the second derivative is necessary, which is given by

$$\begin{aligned}\frac{\partial^2 l_m^*(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1{}'} &= \begin{pmatrix} -\frac{1}{\nu^2}\sum_{i\in\mathcal{J}}\boldsymbol{x}_i\,\frac{y_{i1}}{\mu_{i1}}\boldsymbol{x}_i{}' & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & -\sum_{i\in\mathcal{J}}\boldsymbol{z}_i\frac{\mu_{i2}(e^{\mu_{i2}}-1-\mu_{i2}e^{\mu_{i2}})}{(e^{\mu_{i2}}-1)^2}\boldsymbol{z}_i{}' \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & -\sum_{i\in\mathcal{J}}\boldsymbol{z}_i\mu_{i2}\boldsymbol{z}_i{}' \end{pmatrix} \\ &- \sum_{i\in\mathcal{J}}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\Sigma_w^{-1}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}'.\end{aligned} \tag{4.33}$$

Finally, we get the Fisher information $\mathcal{I}_m^*(\boldsymbol{\theta}_1)$ of $l_m^*(\boldsymbol{\theta}_1)$ with respect to $\boldsymbol{\theta}_1$ with the above result (4.33):

$$\begin{aligned}\mathcal{I}_m^*(\boldsymbol{\theta}_1) &= -n^{-1}E\left[\frac{\partial^2 l_m^*(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1{}'}\right] \\ &= \underbrace{n^{-1}\begin{pmatrix} \frac{1}{\nu^2}\sum_{i\in\mathcal{J}}\boldsymbol{x}_i\boldsymbol{x}_i{}' & 0 \\ 0 & \sum_{i\in\mathcal{J}}\boldsymbol{z}_i\frac{\mu_{i2}(e^{\mu_{i2}}-1-\mu_{i2}e^{\mu_{i2}})}{(e^{\mu_{i2}}-1)^2}\boldsymbol{z}_i{}' \end{pmatrix}}_{=\mathcal{I}_m^c(\boldsymbol{\theta}_1)} \\ &+ n^{-1}\begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \sum_{i\in\mathcal{J}}\mu_{i2}\boldsymbol{z}_i\boldsymbol{z}_i{}' \end{pmatrix} \\ &+ n^{-1}\sum_{i\in\mathcal{J}}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\Sigma_w^{-1}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}'.\end{aligned} \tag{4.34}$$

The Fisher information $\mathcal{I}_d^*(\boldsymbol{\theta}_1,\gamma)$ of $l_d^*(\boldsymbol{\theta}_1,\gamma)$ with respect to $\boldsymbol{\theta}_1$ has the analog form as $\mathcal{I}_m^*(\boldsymbol{\theta}_1)$ with reverse algebraic signs of the expansions, i.e.

$$\begin{aligned}\mathcal{I}_d^*(\boldsymbol{\theta}_1,\gamma) &= -n^{-1}E\left[\frac{\partial^2 l_d^*(\boldsymbol{\theta}_1,\gamma)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1{}'}\right] \\ &= \mathcal{I}_d^c(\boldsymbol{\theta}_1,\gamma) \\ &\quad -n^{-1}\begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & \sum_{i\in\mathcal{J}}\mu_{i2}\boldsymbol{z}_i\boldsymbol{z}_i{}' \end{pmatrix} \\ &\quad -n^{-1}\sum_{i\in\mathcal{J}}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\Sigma_w^{-1}\begin{pmatrix} \mu_{i1} & 0 \\ 0 & \mu_{i2} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_i & \mathbf{0}_p \\ \mathbf{0}_q & \boldsymbol{z}_i \end{pmatrix}'.\end{aligned} \tag{4.35}$$

Now we hope that the elements of $\mathcal{I}_m^*(\theta_1)$ are large enough to force convergence in the Fisher scoring step. Note that

$$l^c(\boldsymbol{\theta}) = l_m^c(\boldsymbol{\theta}_1) + l_d^c(\boldsymbol{\theta}_1,\gamma) = l_m^*(\boldsymbol{\theta}_1) + l_d^*(\boldsymbol{\theta}_1,\gamma),$$

and hence

$$\frac{\partial l^c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1} = \frac{\partial l_m^c(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} + \frac{\partial l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial\boldsymbol{\theta}_1} = \frac{\partial l_m^*(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} + \frac{\partial l_d^*(\boldsymbol{\theta}_1,\gamma)}{\partial\boldsymbol{\theta}_1},$$

and

$$\mathcal{I}^c(\boldsymbol{\theta}) = \mathcal{I}_m^c(\boldsymbol{\theta}_1) + \mathcal{I}_d^c(\boldsymbol{\theta}_1,\gamma) = \mathcal{I}_m^*(\boldsymbol{\theta}_1) + \mathcal{I}_d^*(\boldsymbol{\theta}_1,\gamma).$$

Moreover, the expansions are independent of ρ or γ and therefore $\frac{\partial l_d^*(\boldsymbol{\theta}_1,\gamma)}{\partial\gamma} = \frac{\partial l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial\gamma}$, which we already derived in (4.25).

The applied MBP algorithm with the expansion of the conditional log-likelihood proceeds then as follows:

Algorithm 4.3 (Applied Variation of the MBP)**.**

Step 0: *Assume that the Gamma regression (3.1) and the Poisson regression (3.2) are independent. The initial values then are taken as follows:*

Step 0.1: *The initial value for* $\boldsymbol{\theta}_1$ *is* $\boldsymbol{\theta}_1^0 = [\widehat{\boldsymbol{\alpha}_I}', \widehat{\boldsymbol{\beta}_I}']'$*, where* $\widehat{\boldsymbol{\alpha}_I}$ *and* $\widehat{\boldsymbol{\beta}_I}$ *are the MLEs of the regression coefficients* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *of the GLMs (3.1) and (3.2).*

Step 0.2: *The initial value for* γ *is* γ^0 *the result of* $\frac{\partial l_d^c(\boldsymbol{\theta}_1^0, \gamma)}{\partial \gamma} = 0$ *solved with the bisection method.*

Step 0.3: *The pre-specified correlation* ρ_w *is the empirical correlation of both regression models' residuals.*

Step k (k $= 1, 2, 3, \ldots$**):** *First, we update* $\boldsymbol{\theta}_1$ *by one step of the Fisher scoring method (2.24), i.e.*

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

Then, by solving $\frac{\partial l_d^c(\boldsymbol{\theta}^k, \gamma)}{\partial \gamma} = 0$ *with the bisection method (2.1), we receive 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 puts out an approximation of the MLE for parameter $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \gamma]'$. Since γ is a scalar, $\frac{\partial l_d^c(\boldsymbol{\theta}^k, \gamma)}{\partial \gamma} = 0$ is a one-dimensional search and the bisection method (2.1) works efficiently.
It shows that when the fix pre-specified ρ_w is not close enough to the resulting MLE for ρ the MBP algorithm given in 4.2 does not converge. Hence, we modify the applied variation of the MBP by updating ρ_w. The changes in the algorithm are the following:

in Step 0.3: 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 chapter we run a simulation study for the MBP algorithm with a fix ρ_w and with a ρ_w-update. This study shows that both versions of the MBP algorithm provide similar results, but the version with the ρ_w-update has a better convergence behavior.

4.3 Standard Error Estimation for the MLE

To get an idea of the precision of the MLE $\hat{\boldsymbol{\theta}}$ for the parameter vector $\boldsymbol{\theta}$ we are interested in the standard error of the estimator. According to Theorem 3 of Song et al. (2005) (see also Section 4.1) the MBP algorithm provides an asymptotically normal distribution of the resulting MLE, which we can use to estimate the standard error for the MLE. Let $\hat{\boldsymbol{\theta}}$ be the resulting MLE of the $\boldsymbol{\theta}$ calculated by the MBP algorithm 4.3. For $k \to \infty$ $\hat{\boldsymbol{\theta}}$ has the asymptotic covariance matrix

$$m^{-1}\mathcal{I}^{-1} = m^{-2} E \left[\left. \frac{\partial^2 l^c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \right], \tag{4.36}$$

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

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

where

$$\hat{\mathcal{I}}_d(\hat{\boldsymbol{\theta}}) = \sum_{i \in \mathcal{J}} l'_d(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\; l'_d(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)',$$

with $l'_d(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) := \frac{\partial}{\partial \boldsymbol{\theta}} l_d(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.
The estimated standard error for $\hat{\boldsymbol{\theta}}$ is then the square root of the diagonal elements of the matrix $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}})$. Note that we assume some asymptotic conditions for the conditional log-likelihood which we can not prove as it would need the second derivative of the conditional log-likelihood. But calculating the second derivative of the conditional log-likelihood contradicts the intention of using the MBP algorithm which avoids the determination of the very complex second derivative.

5 Simulation Study of the MLEs provided by the MBP Algorithm

In this chapter we want to study the goodness of the MLE for the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as well as for the correlation parameter ρ provided by the MBP algorithm given in Algorithm 4.2. We also compare the two versions of the algorithm: the one without ρ_w-update and the one with ρ_w-update. Note that we assume the constant of variation ν as known and so it is not calculated with the MBP algorithm. But we prespecify the MLE for ν in the marginal Gamma GLM (3.1) and for completeness the goodness of its MLE is also analyzed here.
Therefore, we simulate R data sets with N observations using the same values of the parameter $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ρ and ν. For each of the N observations we also generate covariates. Afterwards, we prespecify ν and then calculate the MLEs of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ρ with the MBP algorithm given in Algorithm 4.2. So we get R independent identically distributed realizations of the corresponding MLE for each parameter. With these R realizations we are able to estimate goodness measures like the bias or the mean squared error of the MLEs (see 2.6) in order to check their quality.

5.1 Generation of a Correlated Gamma-Poisson Observation

Before we start with the simulation study we need an algorithm that produce random observation of the Gamma-Poisson distribution. To generate an observation set (y_{i1}, y_{i2}) of a $Gamma(\mu_{i1}, \nu)$ distributed random variable Y_{i1} and a with ρ correlated $Poisson(\mu_{i2})$ distributed random variable Y_{i2} we use the conditional probability mass of the Poisson variable Y_{i2} given the Gamma variable Y_{i1}. The joint density function f of Y_{i1} and Y_{i2} is given in equation (3.8). With this knowledge the conditional probability mass of Y_{i2} given Y_{i1} is given as

$$f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho) := \frac{f(y_{i1}, y_{i2}|\mu_{i1}, \nu, \mu_{i2})}{g_1(y_{i1}|\mu_{i1}, \nu)} \tag{5.1}$$

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

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

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

Algorithm 5.1 (Generation of Correlated Gamma and Poisson Random Variables with Correlation ρ).

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

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

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

The ϵ defines when we assume that $P(Y_{i2} = k^* + 1|Y_{i1} = y_{i1}) = 0$. In addition, we need a pair of covariate vectors for each observed value pair for our marginal GLMs. Let $\boldsymbol{x}_i \in \mathbb{R}^{p_1}$ be the covariates of y_{i1} and $\boldsymbol{z}_i \in \mathbb{R}^{p_2}$ be the covariates of y_{i2}. Thereby, we set $x_{i1} = 1$ and $z_{i1} = 1$ as we assume an intercept in our regression models. Further, we choose the coefficient vectors $\boldsymbol{\alpha} \in \mathbb{R}^{p_1}$ and $\boldsymbol{\beta} \in \mathbb{R}^{p_2}$ fix for all i and define the mean of the Gamma distribution of Y_{i1} as

$$\mu_{i1} := \exp(\boldsymbol{x}_i{}'\boldsymbol{\alpha}) \ = \exp(\alpha_1 + \sum_{j=2}^{p1} x_{ij}\alpha_j)$$

and the mean of the Poisson distribution of Y_{i2} as

$$\mu_{i1} := \exp(\boldsymbol{z}_i{}'\boldsymbol{\beta}) \ = \exp(\beta_1 + \sum_{j=2}^{p2} z_{ij}\beta_j).$$

5.2 Simulation Study

In the simulation study we test the quality of the MBP algorithm given by Algorithm 4.2 and the expanded version of the algorithm with the ρ_w-update in 24 scenarios. For each scenario we simulate $R = 500$ data sets and calculate the MLEs for the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ρ and ν for each data. As already mentioned, the MLE for ν is not calculated in the joint regression model, but it is prespecified in the marginal Gamma GLM. Each of the repetition is based on a data set with $N = 1000$ observed Gamma-Poisson random variable pairs. The observations of the Gamma-Poisson random variable pairs are generated by using the sample algorithm given in Algorithm 5.1 using a single uniform-$(0, 1)$ distributed covariate for the marginal Gamma GLM as well as for the marginal Poisson GLM, i.e.

$$\mu_{i1} = \exp(\alpha_1 + x_{i2}\alpha_2)$$

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. The covariates are fixed for all repetitions and scenarios. For the regression parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ of the marginal Gamma GLM we take 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 take 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 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 $\nu = 1$. In Table 5.1 we find an overview of the parameter settings used in the 24 scenarios. In the MBP algorithm we take the stop criterion $\boldsymbol{\theta}_1^k - \boldsymbol{\theta}_1^{k-1} < 10^{-3}$ where

	Parameters					
Scenario	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 5.1: Parameter settings used for the 24 different scenarios

$\boldsymbol{\theta}_1^k = (\alpha_1^k, \alpha_2^k, \beta_1^k, \beta_2^k)'$ are the regression parameter after the k-th iteration and $||\rho^k - rho^k|| < 10^{-4}$, where ρ^k denotes the correlation parameter after the k-th iteration. In the version of the MBP algorithm without ρ_w-update we set ρ_w fix equal to the empirical correlation between the residuals of the marginal Gamma GLM and the residuals of the marginal Poisson GLM.

The MBP algorithm turns out to work quite reasonable for the most of the 24 scenarios. Table 5.2 and Table 5.3 give a summary of the relative bias of each scenario for both algorithm versions. Each table contains the relative bias in percent of all parameters for the 24 scenarios as well as the maximal bias of the parameters in each scenario given in column *max*. The relative bias is calculated by dividing the bias $\hat{\theta}$ by the real value of the parameter. The formula of the estimated bias can be found in equation (2.26). More detailed information about the performance of both versions of the MBP algorithm can be found in Appendix B. When we look at able 5.2 and Table 5.3 we see that, besides three or two exceptions, the maximal bias in a scenario is small. But it is remarkable that for an increasing ρ the results of the algorithm lose their precision, as the maximal bias of the parameter has an increasing tendency. It is also shown that the MBP algorithm works in the scenarios with $\mu_{i1} \in (2.72, 54.60)$, i.e. $\boldsymbol{\beta} = (1.0, 3.0)'$, better than in the corresponding scenarios with $\mu_{i1} \in (2.72, 7.39)$, i.e. $\boldsymbol{\beta} = (1.0, 1.0)'$. On the other hand, the choice of the range for the mean of the marginal Poisson GLM has seemingly no effect on the

	Relative bias in %						
Scenario	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 5.2: Relative bias of the MLEs calculated with the MBP algorithm using no ρ_w-update for the parameters α_1, α_2, β_1, β_2, ρ and ν in the 24 different scenarios with max as greatest relative bias of each scenario.

performance of the algorithm. Additionally, the tables uncover that when the signal-to-noise ratio of the marginal Gamma observation is high, the relative bias of the MLE provided by the MBP algorithm is higher than when the signal-to-noise ratio is small. Further, we can say that the most problematical parameters seem to be ρ, but only when it is small, and the intercept parameter β_1 of the marginal Poisson GLM. The intercept parameter is the parameter that is responsible for the highest relative bias with -14% or -16% in scenario 17, 11% in scenario 18 in the version without ρ_w-update and 34% or 25% in scenario 20. The reason for the suboptimal performance of the MBP algorithm for β_1 could be the fact that we do not take account of all Gamma-Poisson observations generated by the sampling algorithm as we use the likelihood conditional on $Y_{i2} > 0$.

Comparing the MBP algorithm versions, the one with and the one without ρ_w-update, we see that the maximal relative bias for the scenario seems to be slightly smaller in the version with ρ_w-update. Especially in the scenarios with the high correlated data sets, the ρ_w-update seems to improve the resulting MLEs. For medium to low correlated data sets the performances of the two versions are almost identical according to Table 5.2 and Table 5.3.

Scenario	Relative bias in % 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 5.3: Relative bias of the MLEs calculated with the MBP algorithm using ρ_w-update for the parameters α_1, α_2, β_1, β_2, ρ and ν in the 24 different scenarios with max as greatest relative bias of each scenario.

6 Data Analysis and Marginal Regression Models

Our aim in this paper is to estimate the average total loss for a full comprehensive car insurance portfolio. The data set we want to analyze is a sub set of a data set from a German insurance company of the year 2000. As the total loss is calculated by aggregating the total loss of the single policies or policy groups, which is the product of the average claim size and the number of claims, we need the joint distribution function of the average claim size and the number of claims. Thereby, we assume no independency between the average claim size and the number of claims and hence we model the joint distribution by applying the Gamma-Poisson regression model of Chapter 3 to the car insurance portfolio. To calculate the parameter MLEs of the joint distribution we use the MBP algorithm constructed in Section 4.2. The selection of the covariates and the calculation of the initial values for the algorithm is done in the marginal GLMs of the joint regression model. Finally, we use the Monte Carlo method to estimate the expected total loss for the German full comprehensive car insurance portfolio.

6.1 Description and Aggregation of the Data Set

The data set contains information on full comprehensive car insurance policies in Germany within the year 2000. There are only policies recorded which have at minimum one claim within the regarded period and not all policies are valid for the whole year. The exposure time of each policy is known, however. An observation consists of the cumulative and average claim size, the number of claims and several covariates like type and age of the car, distance driven per year, age and gender of the policyholder, the claim free years and the deductible. We only analyze a subset of this data set containing three types of midsized cars. The resulting number of observations is 12'850. We are interested in the joint distribution of the number of claims and the average claim size of a policy respecting possible dependency between both to estimate the expected total loss. In Table 6.1 we can see that most of the observations have one claim within the period and the maximum number of observed claims (four) appears 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 (cp. Table 6.1). As the number of claims variable is a count variable we use a Poisson GLM for modeling.

	number of claims			
values	1	2	3	4
#	12472	356	20	2
%	97.058%	2.770%	0.155%	0.015%

Table 6.1: Absolute and relative frequencies of the occurring values of number of claims

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
number of claims	1.00	1.00	1.00	1.03	1.00	4
average claim size	0.01	2374.23	4195.27	5755.01	7272.75	49339.06

Table 6.2: Quartiles, means, minimum and maximum of the responses

The average claim size of the policies is given in the currency of DM[1]. It is calculated by the sum-up of the claim size of each policy claim divided by the number of claims. The histogram of the observed average claim sizes in Figure 6.1 shows a right-skew 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 6.1). This shows that the right-skewness is not as extreme as the one of the number of claims. 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. Therefore, there is no need to use a heavy tailed distribution and so a Gamma model is a good 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 6.1). In order to get good estimates for the model parameters, we need a wider range of the observed number of claims and more observations unequal to one. This can be achieved by the aggregation of the policies according 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 claim free years. The covariates of regional class and deductible are already categorical, but the covariates driven distance per year, age of policyholder, construction year of the car and claim free years are metric or continuous and have to be categorized. In Tables A.1 - A.6 of Appendix A we can find the categories with their absolute and relative frequencies of all the covariates we use in the aggregation. Since the claim free years affect the premium rate, we transform the covariate claim free years into the covariate premium rate, which is a categorical covariate. The result can be seen in Table A.3.
We aggregate the policies using the R-function `aggregate()`. Using these covariates with their categories we receive $2 \cdot 8 \cdot 7 \cdot 5 \cdot 5 \cdot 8 \cdot 7 = 156'800$ possible groups, but only in 7'955 groups one or more policies are allocated. For each resulting group of policies we calculate the number of claims occurred in the group and the cumulative claim size as well as the average claim size. The average claim size per policy group is calculated by dividing the cumulative claim size of the group by its number of claims. The exposure of a group is the sum of the exposures of the policies falling in the group. Now we have a new data set with 7'955 observations containing the number of claims and the average claim size of policy groups, the categorical or categorized covariates car type, regional class, premium rate, deductible, driven kilometers per year, age of the policyholder and construction year of the car as well as the cumulative exposure time. As we intended, the range of the number of claims is wider than in the original data set. The absolute and relative frequencies of the observed number of claims are given in Table 6.1. In Figure 6.2 we have the histogram of the average claims sizes of the policy groups. The shape of the histogram is quite the same as the one of the histogram of the average claim sizes of each policy, given in Figure 6.1. By summing up the product of the average claim size and the number of claims of the policy groups we receive 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.

[1] 'Deutsche Mark' (DM) is the former German currency, which was replaced by the Euro in the European currency union in 2002 (1 DM $\hat{=}$ 0.51 Euro). Here we also use abbreviation TDM for 1000 DM

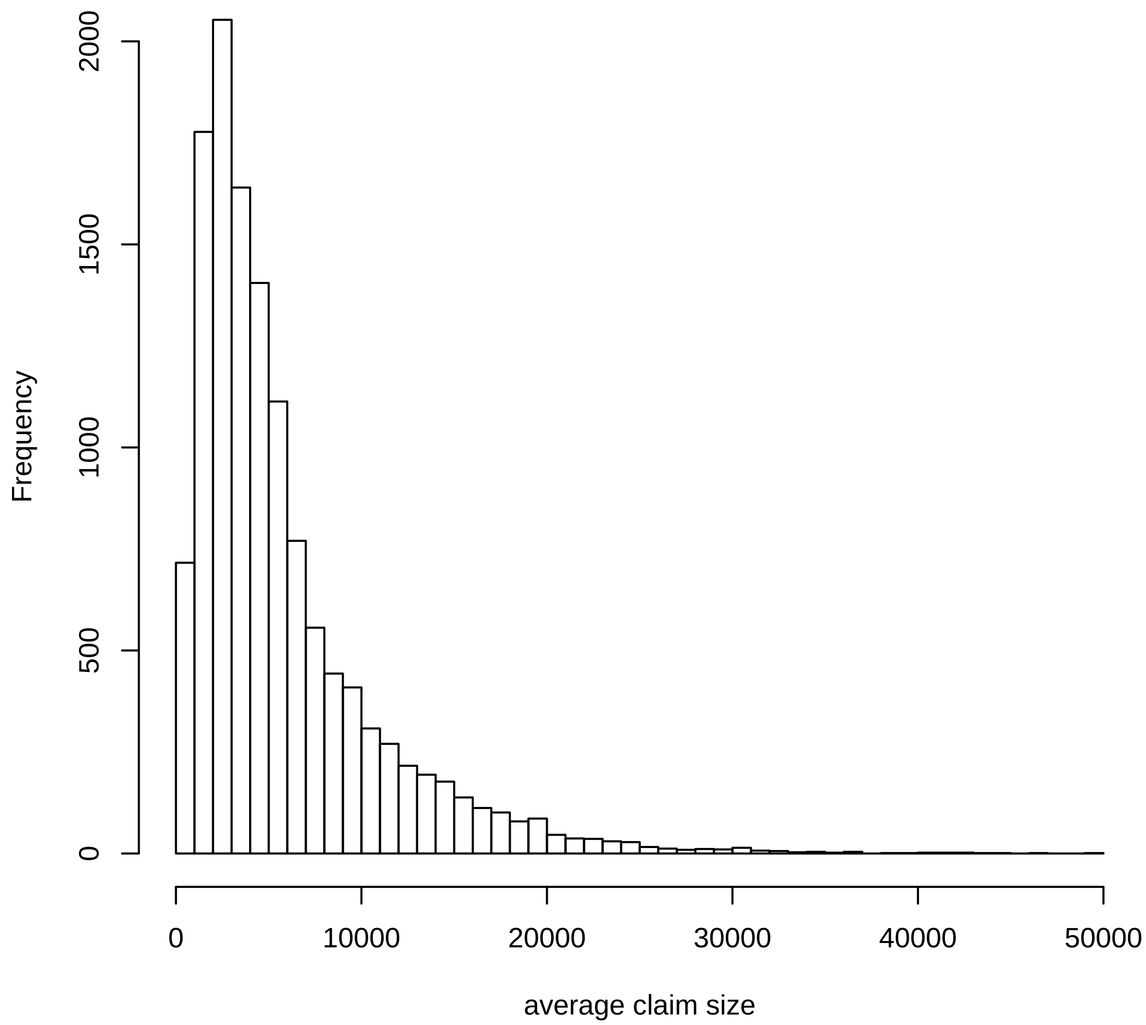

Figure 6.1: Histogram of the observed average claim size

	number of claims										
values	1	2	3	4	5	6	7	8	9	10	11
#	5613	1278	453	236	127	63	55	40	30	14	17
%	70.56	16.07	5.69	2.97	1.60	0.79	0.69	0.50	0.38	0.18	0.21
values	12	13	14	15	16	17	19	20	21	22	37
#	3	5	5	5	2	4	1	1	1	1	1
%	0.04	0.06	0.06	0.06	0.03	0.05	0.01	0.01	0.01	0.01	0.01

Table 6.3: Absolute and relative frequencies of the occurring values of number of claims in the aggregated data set

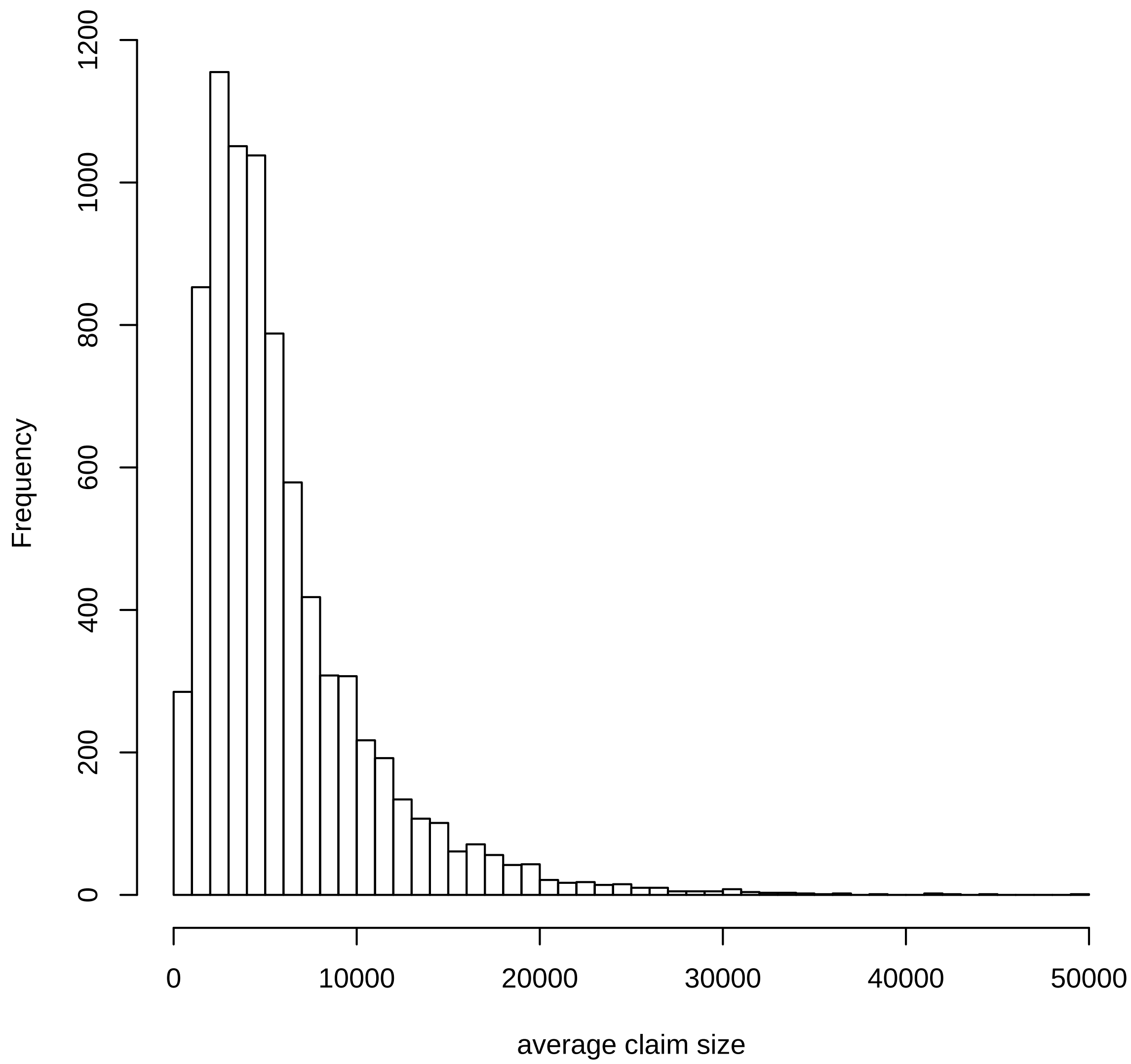

Figure 6.2: Histogram of the observed average claim size in the aggregated data set

When we look at Figure 6.3, which shows a plot of the number of claims against the average claim size of the policy groups and the corresponding regression line, we see that the regression line has a small positive gradient. This could be a hint for a positive correlation between the number of claims and the average claim size.
In the following we use the new 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.

6.2 Marginal Regression Models

In this section we construct a Gamma GLM for the average claim size and a Zero-truncated Poisson GLM (see Section 2.3.2) for the number of claims and independently estimate the regression parameter for each of the two GLMs. We take the resulting regression parameter values as initial values for the MBP algorithm to calculate the parameter of the joint regression model of the number of claims and the average claim size in the next section. For the number of claims we chose a Zero-truncated Poisson model as the data set of the car insurance portfolio only contains policies with at minimum one observed claim. When the data set would also contain policies with zero claim, we could easily take a Poisson GLM to select the covariates and get a good initial value for the MBP algorithm. But in this case we have to take the Zero-truncated Poisson GLM.
Our procedure to construct the GLMs is the following: First we make an explorative data analysis for each GLM to get a general idea of how the covariates influence the response. Thereby, we explore the categorical covariates with plots of the empirical log means for each class of the covariates. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ be the observed response vector Y of the GLM and let $I_1, I_2, \ldots, I_k$ be the classes of a covariate X with the observed values $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$. Moreover, let

$$n_l := |\{x_i : x_i \in I_l\}|,\ l = 1, \ldots, k$$

be the number of observations of $\boldsymbol{x}$ in class l,

$$m_l := \frac{1}{n_l} \sum_{h \in \{i : x_i \in I_l\}} y_h,\ l = 1, \ldots, k$$

be the mean of the response values of each class I_l and

$$\mathit{log\ mean}_l := \ln(m_l),\ l = 1, \ldots, k$$

be the empirical log mean of class I_l. We plot the empirical log mean in one figure for each class of the covariate $\boldsymbol{x}_j$. To do this, we use the R-functions `gamma.main()` and `poisson.main()`.[2] The dashed lines in the plot mark the upper and lower borders of the 95% confidence interval of the empirical log means. Below the figure the range between the highest and the lowest log mean of the classes is specified in parenthesis. For the Gamma GLM we also exploratively check the interactions to get a main idea about the interactions between the covariates. For two covariates X_1 and X_2 with the classes $I_1^1, \ldots, I_{k_1}^1$ and $I_1^1, \ldots, I_{k_1}^1$ let

$$n_{lr} := |y_i : x_{i1} \in I_l^1 \wedge x_{i2} \in I_r^2|,\ l = 1, \ldots, k_1,\ r = 1, \ldots, k_2,$$

[2] The code for the R-functions can be found in Appendix C.

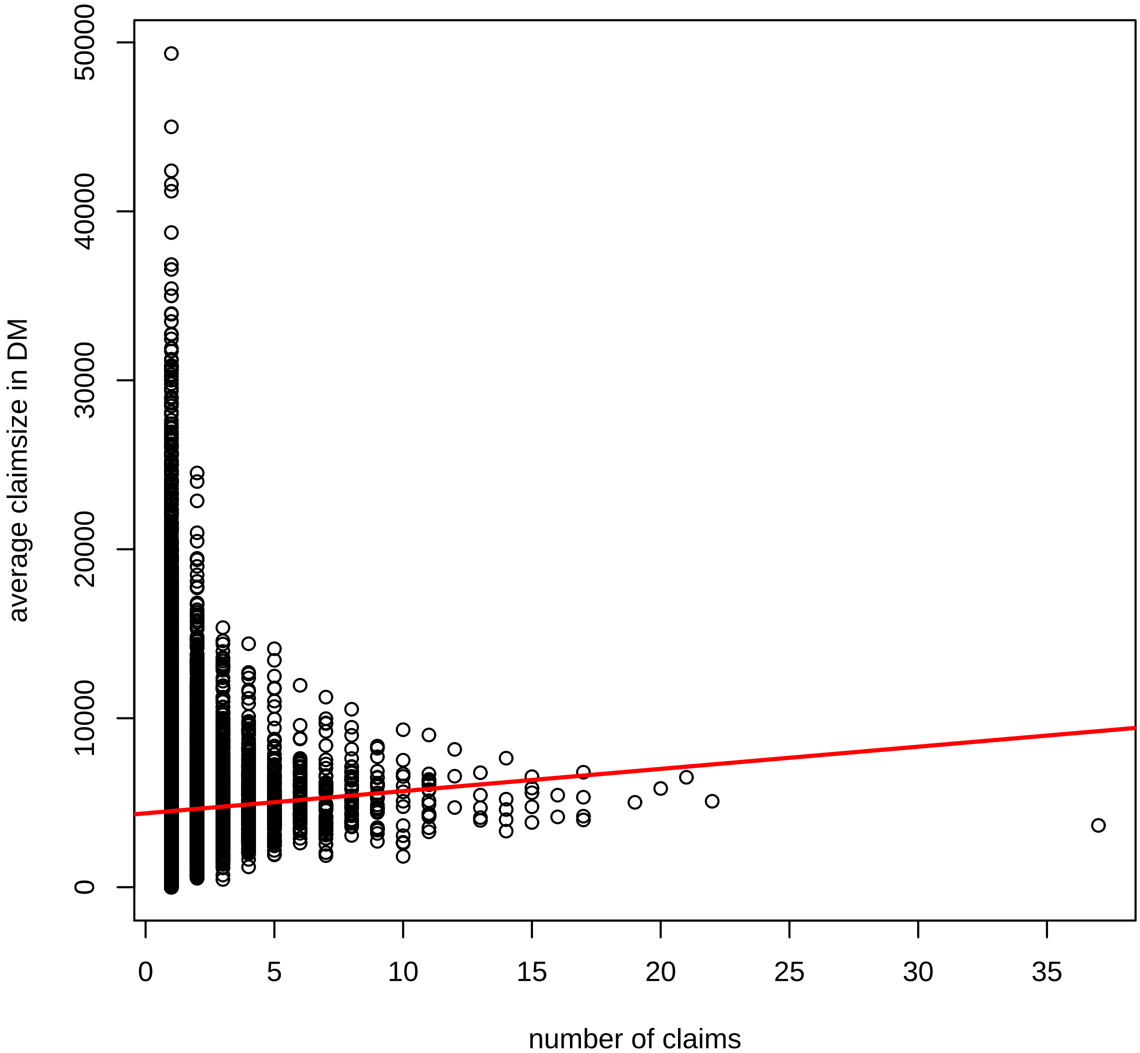

Figure 6.3: Plot of the number of claims against the average claim size of the groups

be the number of observation of Y falling in the cell (l, r),

$$m_{lr} := \frac{1}{n_{lr}} \sum_{h \in \{i: x_{i1} \in I_l^1 and x_{i2} \in I_r^2} y_h, \; l = 1, \ldots, k_1, \; r = 1, \ldots, k_2,$$

be the mean of the observed response values in cell (l, r) and

$$log\ mean_{lr} := \ln(m_{lr}), \; l = 1, \ldots, k_1, \; r = 1, \ldots, k_2,$$

be the empirical log mean of the cell (l, r). The R-function `gamma.inter()` calculates all these variables and plots then the empirical log means for the different cells. Interaction effects can be identified when the frequency polygons are not approximately parallel.
To construct the GLMs we dummy code the covariates. The names of the dummy variables for each covariate can be found in the tables of Appendix A. For the variable selection we use the Partial Deviance Test. Our variable selection procedure is as follows: We start with a GLM containing all covariates (i.e. all dummy variables of each covariate) and test whether there is any possible interaction between two of these covariates which adding is not rejected by the Partial Deviance Test with a significant level of 0.95. When there are any interactions, we add the interaction term with the smallest p-value to our regression model and look if we can add another interaction to the new GLM (the one containing already the not rejected interaction term). When all other addings of interaction terms are rejected we use the Partial Deviance Test to find the covariate with the greatest p-value and drop it. But we only drop covariates which are not used in an added interaction term. When the p-values of all covariates are smaller than 0.5 or all covariates with a p-value greater than 0.5 are used in an interaction term and there is no more interaction term which can be added we found the final GLM. This GLM we then use as the marginal GLM in the joint regression model.
In the next subsections the details of the contraction of the Gamma GLM for the average claim size and the Zero-truncated Poisson GLM for the number of claims are described.

6.2.1 Marginal Gamma GLM

In this subsection we construct the marginal model for the expected average claim size of the policies. We assume that the average claim size Y_{i1} of the policy groups, $i = 1, 2, \ldots, 7'993$, are Gamma distributed with mean parameter μ_{i1} and constant coefficient of variation ν. According to the Gamma GLM (3.1) it holds

$$E[Y_{i1}] = \mu_{i1} = \exp(\boldsymbol{x}_i{}'\boldsymbol{\alpha}),$$

where $\boldsymbol{x}_i$ is the vector of the observed covariates including the 1 for the intercept of the i-th observation and $\boldsymbol{\alpha}$ is the vector of the corresponding parameter vector of the predictor $\boldsymbol{x}_i{}'\boldsymbol{\alpha}$. We can estimate this equation by

$$\ln(Y_{i1}) = \boldsymbol{x}_i{}'\boldsymbol{\alpha}.$$

It follows that $\ln(Y_{i1})$ should have a linear behavior in $\boldsymbol{x}_i$.

Before we start to construct the Gamma GLM, we make an explorative data analysis to get a general idea of how the covariates influence the response (the average claim size of the policy groups). First, we have a look at the plots of the empirical average claim size log means of the classes of all covariates, generated by the R-function `gamma.main()`. In the calculation of the

log means we set $y_i = y_{i1}$, where y_{i1} stands for the i-th observed average claim size. In Figures 6.4 (1)-(7) we see the results.
When we look at Figure 6.4 (1), we discover that the empirical log means of the covariate 'sex' are almost the same (the range between the log empirical means of the two categories is only 0.1, which is the smallest of all covariates), what leads to the presumption that the gender of a policyholder plays a subordinate role in determining the height of the average claim size. Hence, we do not use this covariate in the GLM. The greatest influence on the response seem to have the covariates premium rate and deductible, as they have the widest range (0.44 and 0.66) between the highest and the lowest log mean of their classes.
To explore the interaction between the covariates we look at Figures 6.5 and 6.6. The plots are produced with the R-function `gamma.inter()`. We uncover interaction between covariates when the frequency polygons of the different classes are not roughly parallel. This non parallel behavior can be seen for a lot of covariate combinations. It seems that the strongest interactions are the interactions between the covariates 'regional class' and 'deductible', 'regional class' and 'driven distance per year', 'age and 'premium rate', 'age' and 'construction year of the car', 'age and 'deductible' as well as 'age and 'driven distance per year'.
As our covariates are all categorical we use dummy coding for the covariates to construct the Gamma GLM. The available dummy variables and their definitions are listed in Appendix A. We construct the Gamma GLM by using the Partial Deviance Test. The procedure of the covariate selection is described above. The resulting design matrix X for our Gamma GLM is

```
X <- cbind( d.deduct1, d.deduct2, d.deduct3, d.deduct4,
            d.age1, d.age2, d.age3, d.age4, d.age5, d.age6, d.age7,
            d.constyear1, d.constyear2, d.constyear3, d.constyear4,
            d.constyear5, d.constyear6,
            d.age1*d.constyear1, d.age1*d.constyear2, d.age1*d.constyear3,
            d.age1*d.constyear4, d.age1*d.constyear5, d.age1*d.constyear6,
            d.age2*d.constyear1, d.age2*d.constyear2, d.age2*d.constyear3,
            d.age2*d.constyear4, d.age2*d.constyear5, d.age2*d.constyear6,
            d.age3*d.constyear1, d.age3*d.constyear2, d.age3*d.constyear3,
            d.age3*d.constyear4, d.age3*d.constyear5, d.age3*d.constyear6,
            d.age4*d.constyear1, d.age4*d.constyear2, d.age4*d.constyear3,
            d.age4*d.constyear4, d.age4*d.constyear5, d.age4*d.constyear6,
            d.age5*d.constyear1, d.age5*d.constyear2, d.age5*d.constyear3,
            d.age5*d.constyear4, d.age5*d.constyear5, d.age5*d.constyear6,
            d.age6*d.constyear1, d.age6*d.constyear2, d.age6*d.constyear3,
            d.age6*d.constyear4, d.age6*d.constyear5, d.age6*d.constyear6,
            d.age7*d.constyear1, d.age7*d.constyear2, d.age7*d.constyear3,
            d.age7*d.constyear4, d.age7*d.constyear5, d.age7*d.constyear6 ).
```

To calculate the MLEs of the regression parameter we use the R-function `glm()`. The resulting MLEs are given in the summary in Table 6.4. There, we find the value, the standard error and the t-value with the corresponding p-value of the estimated regression parameters. The estimate of the dispersion parameter is 0.626, which shows an underdispersion in the Gamma GLM. The residual deviance is 4780.0 on 7895 degrees of freedom. The p-value of the residual deviance test is 0.98, which shows that this model is reasonable and states a good fit.

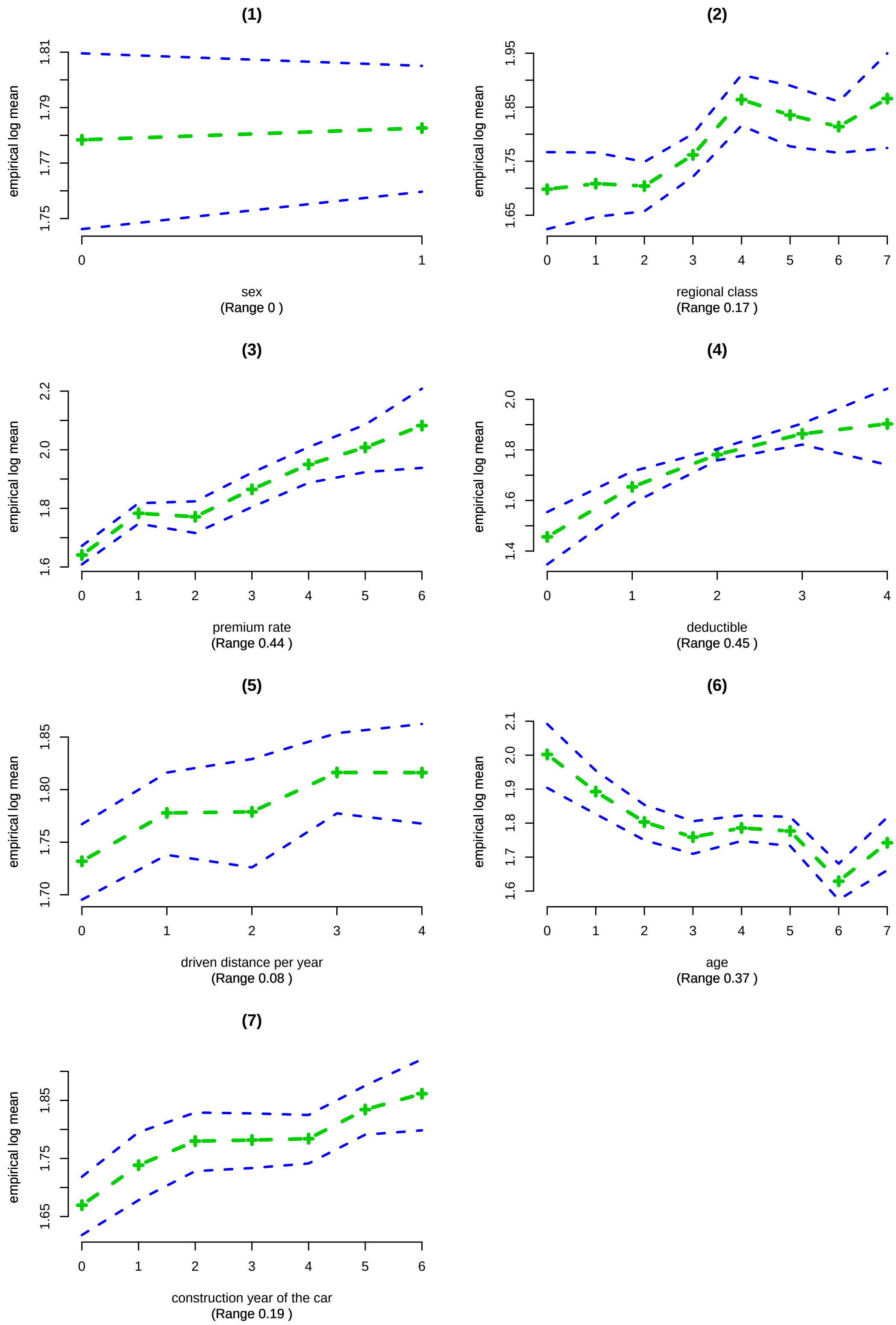

Figure 6.4: Empirical log mean of average claim size for each class of the categorical covariates; (1) - log means for sex, (2) - log means for regional class, (3) - log means for premium rate, (4) - log means for deductible, (5) - log means for driven distance per year, (6) - log means for age, (7) - log means for construction year

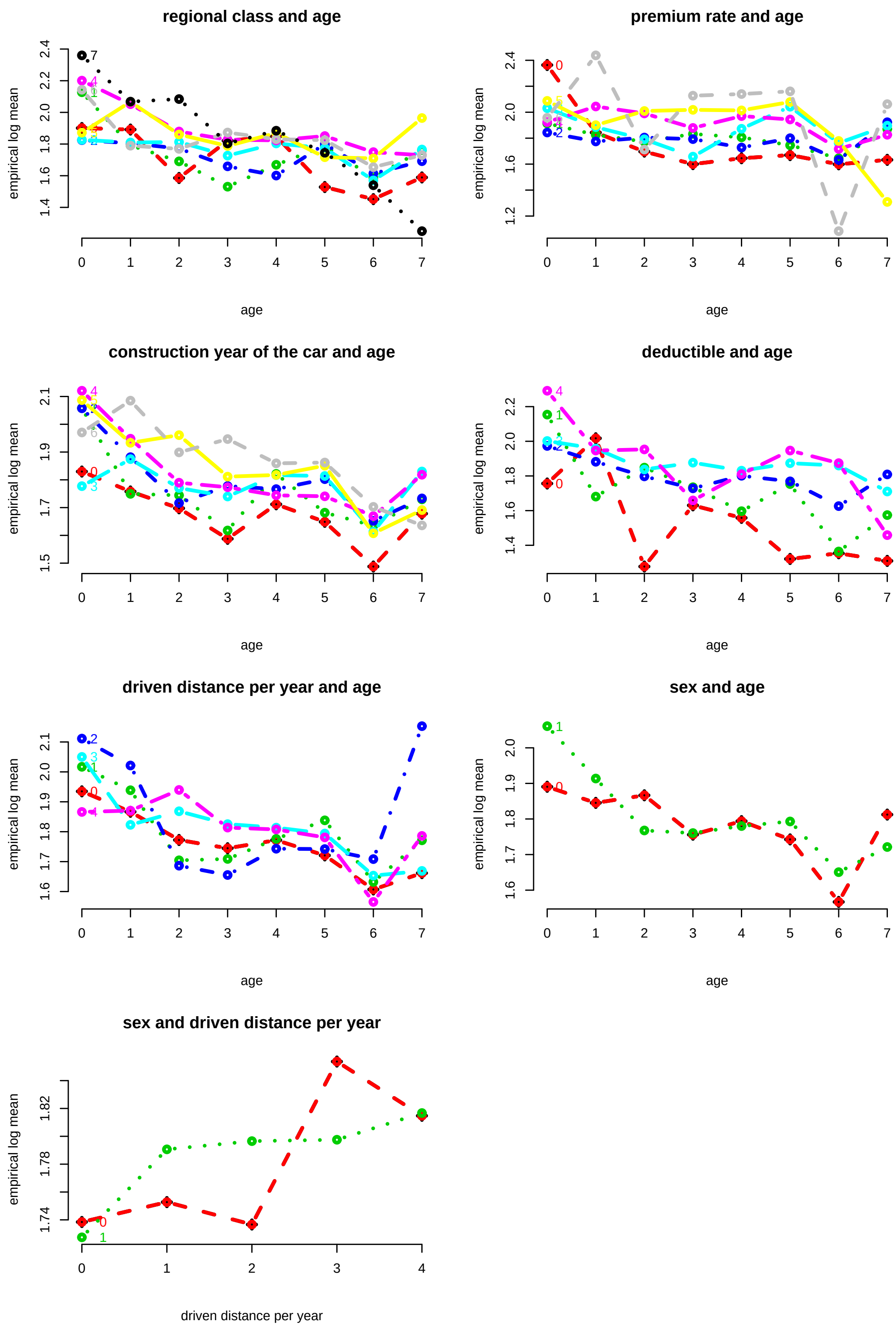

Figure 6.5: Empirical log means of average claim size to explore the 2-dimensional Interactions

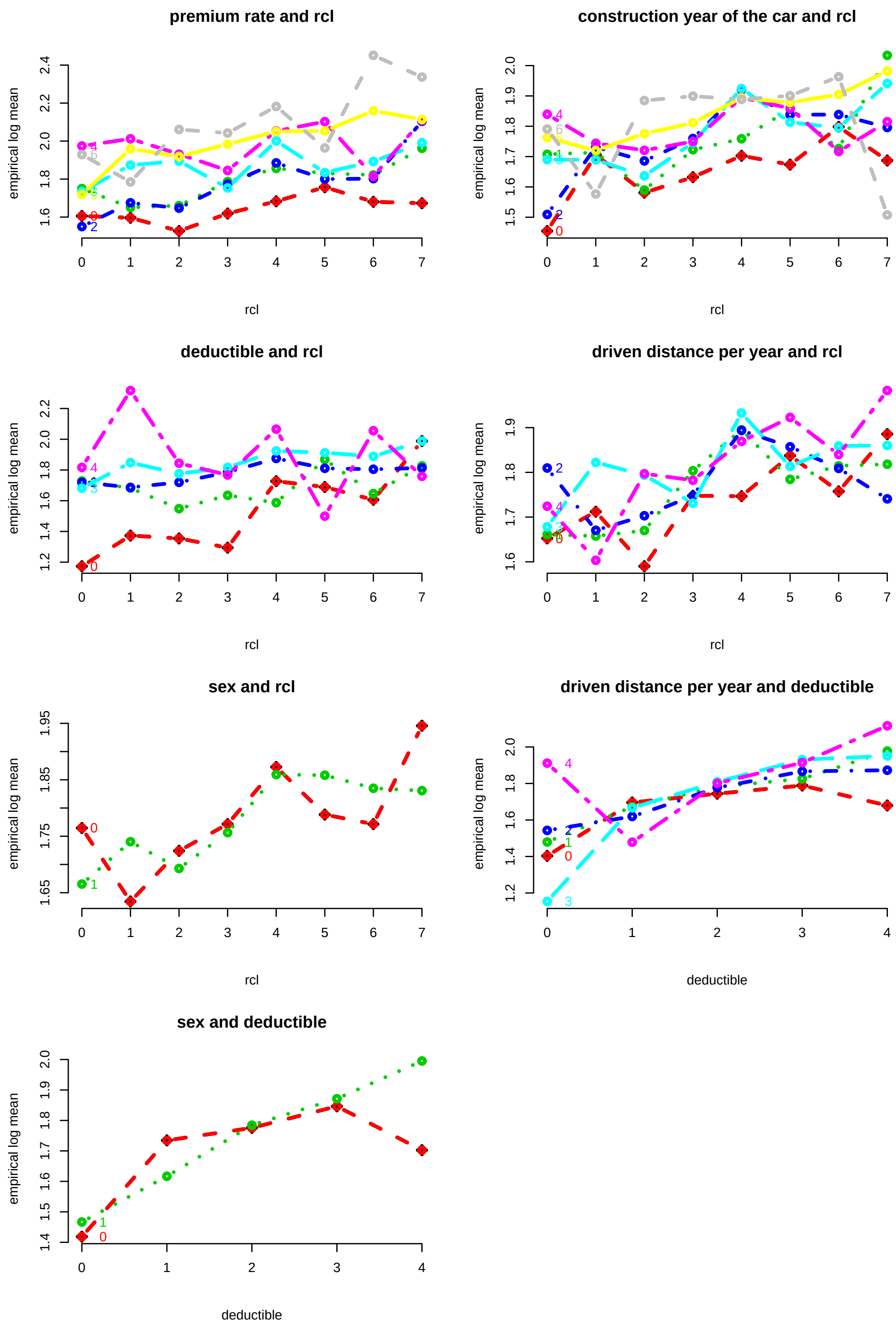

Figure 6.6: Empirical log means of average claim size to explore the 2-dimensional Interactions

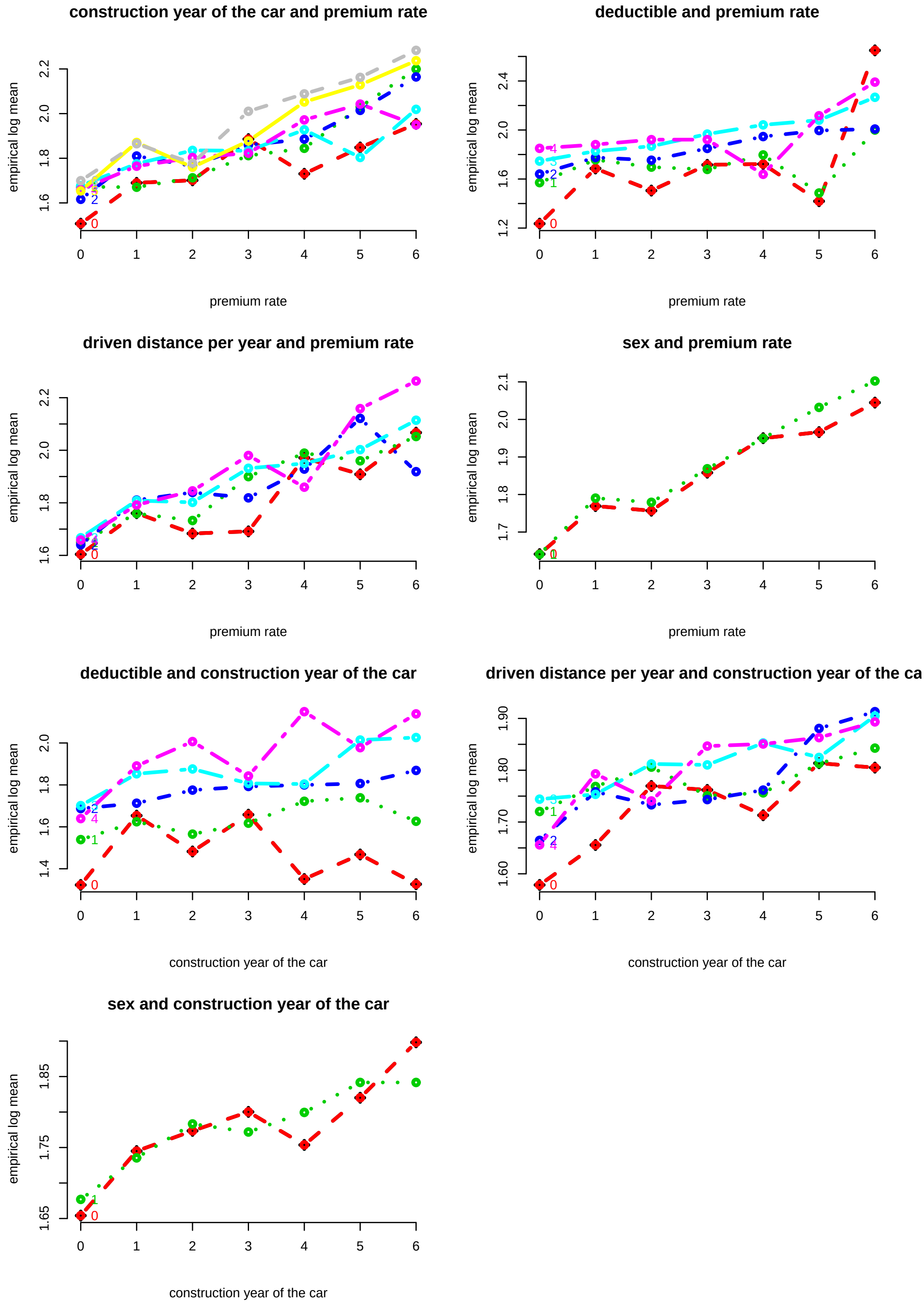

Figure 6.7: Empirical log means of average claim size to explore the 2-dimensional Interactions

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.5544	0.1254	12.40	0.0000
d.deduct1	0.1795	0.0585	3.07	0.0022
d.deduct2	0.2916	0.0513	5.68	0.0000
d.deduct3	0.3803	0.0540	7.04	0.0000
d.deduct4	0.4186	0.0882	4.75	0.0000
d.age1	−0.1040	0.1430	−0.73	0.4672
d.age2	−0.1531	0.1345	−1.14	0.2548
d.age3	−0.2691	0.1332	−2.02	0.0433
d.age4	−0.1469	0.1244	−1.18	0.2376
d.age5	−0.2010	0.1274	−1.58	0.1145
d.age6	−0.3691	0.1389	−2.66	0.0079
d.age7	−0.1900	0.1586	−1.20	0.2311
d.constyear1	0.1807	0.1869	0.97	0.3336
d.constyear2	0.1791	0.1642	1.09	0.2753
d.constyear3	−0.0952	0.1682	−0.57	0.5712
d.constyear4	0.2584	0.1559	1.66	0.0976
d.constyear5	0.2170	0.1542	1.41	0.1593
d.constyear6	0.1052	0.1911	0.55	0.5819
d.age1:d.constyear1	−0.1725	0.2277	−0.76	0.4487
d.age1:d.constyear2	−0.0405	0.2022	−0.20	0.8411
d.age1:d.constyear3	0.2184	0.2033	1.07	0.2829
d.age1:d.constyear4	−0.0579	0.1909	−0.30	0.7618
d.age1:d.constyear5	−0.0445	0.1876	−0.24	0.8127
d.age1:d.constyear6	0.2077	0.2399	0.87	0.3866
d.age2:d.constyear1	−0.1399	0.2145	−0.65	0.5143
d.age2:d.constyear2	−0.1641	0.1906	−0.86	0.3892
d.age2:d.constyear3	0.1610	0.1909	0.84	0.3989
d.age2:d.constyear4	−0.1720	0.1796	−0.96	0.3382
d.age2:d.constyear5	0.0494	0.1785	0.28	0.7820
d.age2:d.constyear6	0.0936	0.2247	0.42	0.6769
d.age3:d.constyear1	−0.1328	0.2123	−0.63	0.5316
d.age3:d.constyear2	−0.0002	0.1868	−0.00	0.9991
d.age3:d.constyear3	0.2460	0.1896	1.30	0.1946
d.age3:d.constyear4	−0.0682	0.1771	−0.38	0.7004
d.age3:d.constyear5	0.0147	0.1765	0.08	0.9338
d.age3:d.constyear6	0.2581	0.2163	1.19	0.2329
d.age4:d.constyear1	−0.0743	0.2002	−0.37	0.7107
d.age4:d.constyear2	−0.1207	0.1776	−0.68	0.4968
d.age4:d.constyear3	0.2179	0.1804	1.21	0.2271
d.age4:d.constyear4	−0.2105	0.1680	−1.25	0.2102
d.age4:d.constyear5	−0.1049	0.1667	−0.63	0.5292
d.age4:d.constyear6	0.0504	0.2060	0.24	0.8066
d.age5:d.constyear1	−0.1415	0.2058	−0.69	0.4916
d.age5:d.constyear2	−0.0192	0.1821	−0.11	0.9160
d.age5:d.constyear3	0.2585	0.1848	1.40	0.1618
d.age5:d.constyear4	−0.1587	0.1716	−0.92	0.3550
d.age5:d.constyear5	−0.0074	0.1698	−0.04	0.9653
d.age5:d.constyear6	0.1184	0.2088	0.57	0.5707
d.age6:d.constyear1	−0.0158	0.2174	−0.07	0.9422
d.age6:d.constyear2	−0.0103	0.1952	−0.05	0.9578
d.age6:d.constyear3	0.2227	0.1973	1.13	0.2590
d.age6:d.constyear4	−0.0450	0.1831	−0.25	0.8060
d.age6:d.constyear5	−0.0725	0.1816	−0.40	0.6897
d.age6:d.constyear6	0.1461	0.2191	0.67	0.5051
d.age7:d.constyear1	−0.0972	0.2477	−0.39	0.6947
d.age7:d.constyear2	−0.0924	0.2283	−0.40	0.6858
d.age7:d.constyear3	0.2714	0.2238	1.21	0.2252
d.age7:d.constyear4	−0.0805	0.2058	−0.39	0.6958
d.age7:d.constyear5	−0.1514	0.2055	−0.74	0.4612
d.age7:d.constyear6	−0.1087	0.2484	−0.44	0.6616

Table 6.4: Summary of the Gamma GLM

In the following $\hat{\boldsymbol{\alpha}}_m$ denotes the estimated values of the regression parameter $\boldsymbol{\alpha}$ given in Table 6.4 and $\hat{\nu} = \sqrt{\hat{\varphi}} = \sqrt{0.626} \approx 0.791$ is the corresponding estimated constant coefficient of variation, where $\hat{\varphi}$ is the estimated dispersion parameter of the above Gamma GLM.

6.2.2 Marginal Zero-truncated Poisson GLM

In this subsection we construct the model for the expected number of claims per group. As the number of claims variable is a count variable with values grater than 0, we use the Zero-truncated Poisson GLM introduced in section 2.3.2. The number of claims per policy obviously depends on the duration of the policy. Hence, we use the cumulative policy duration e_i as exposure in the Zero-truncated Poisson regression. As we want to use the Zero-truncated Poisson regression to model the expectation of the number of claims, we must assume that the number of claims y_{i2} of the policy groups are observations of Zero-truncated Poisson distributions. According to the Zero-truncated Poisson regression model 2.12 it then holds

$$E[Y_{i2}|Y_{i2} > 0] = \frac{\mu_{i2}}{1 - e^{-\mu_{i2}}} \text{ with } \mu_{i2} = e_i \exp(\boldsymbol{z}_i{}'\boldsymbol{\beta}), \tag{6.1}$$

where $\boldsymbol{z}_i$ is the observed covariate vector including the 1 for the intercept of the i-th observation and $\boldsymbol{\beta}$ is the corresponding regression parameter vector of the predictor $\boldsymbol{z}_i{}'\boldsymbol{\beta}$. μ_{i2} is the parameter of the underlying common Poisson distribution. Equation 6.1 can not be solved analytically and therefore, as shown in subsection 2.3.2, we linearize the right part of equation 6.1 with the Taylor formula. Now we estimate equation 6.1 by

$$\ln(\tilde{Y}_{i2}) = \ln(e_i) + \boldsymbol{z}_i{}'\boldsymbol{\beta}, \tag{6.2}$$

where

$$\tilde{Y}_{i2} = Y_{i2} \frac{(1 - e^{-a_i})^2}{1 - e^{-a_i} - a_i e^{-a_i}} - a_i \frac{1 - e^{-a_i}}{1 - e^{-a_i} - a e^{-a_i}} + a_i \tag{6.3}$$

are the approximate transformed observations. The development point a_i of the Taylor approximation has to be as to be in the neighborhood of the real value of μ_{i2} to have a good approximation. Here, we take $a_i = Y_{i2} - 1 + c$ with $c = 0.0001$. We have to add c as a_i has to be greater than 0, as otherwise equation 6.3 is not defined. It follows that $\ln(\tilde{Y}_{i2} - \ln(e_i))$ should be linear in $\boldsymbol{x}_i$.

Again, we first make an explorative data analysis in order to get a general idea of how the covariates influence the response variable (the number of claims of the policy groups). Therefore, we have a look at the plots of the empirical log means of the transformed number of claims $\tilde{Y}_{i2}$ for each class of all covariates, generated by the R-function `poisson.main()`. To calculate the log mean we set $y_i = \frac{\tilde{Y}_{i2}}{e_i}$. The plots can be seen in Figures 6.8 (1)-(7).

When we look at the wide of range between the highest and the lowest log mean of the covariate categories, given under each plot, we find that the covariate 'deductible' seems to have the highest influence on the response as the range is the widest with a value of 2.91 (cp. 6.8 (4)). The lowest influence on the response variable probably have the covariates 'sex' and 'construction year of the car' since the ranges between the highest and the lowest log mean are smallest with value 0.25 and 0.26 respectively (cp. 6.8 (7)).

As our covatiates are again all categorical we use dummy coding to construct the Zero-truncated Poisson GLM. The available dummy variables and their definitions are listed in Appendix A. We construct the Zero-truncated Poisson GLM by using the Partial Residual Test. The procedure of how the covariates are selected is described above. The resulting design matrix Z for our Zero-truncated Poisson GLM then is

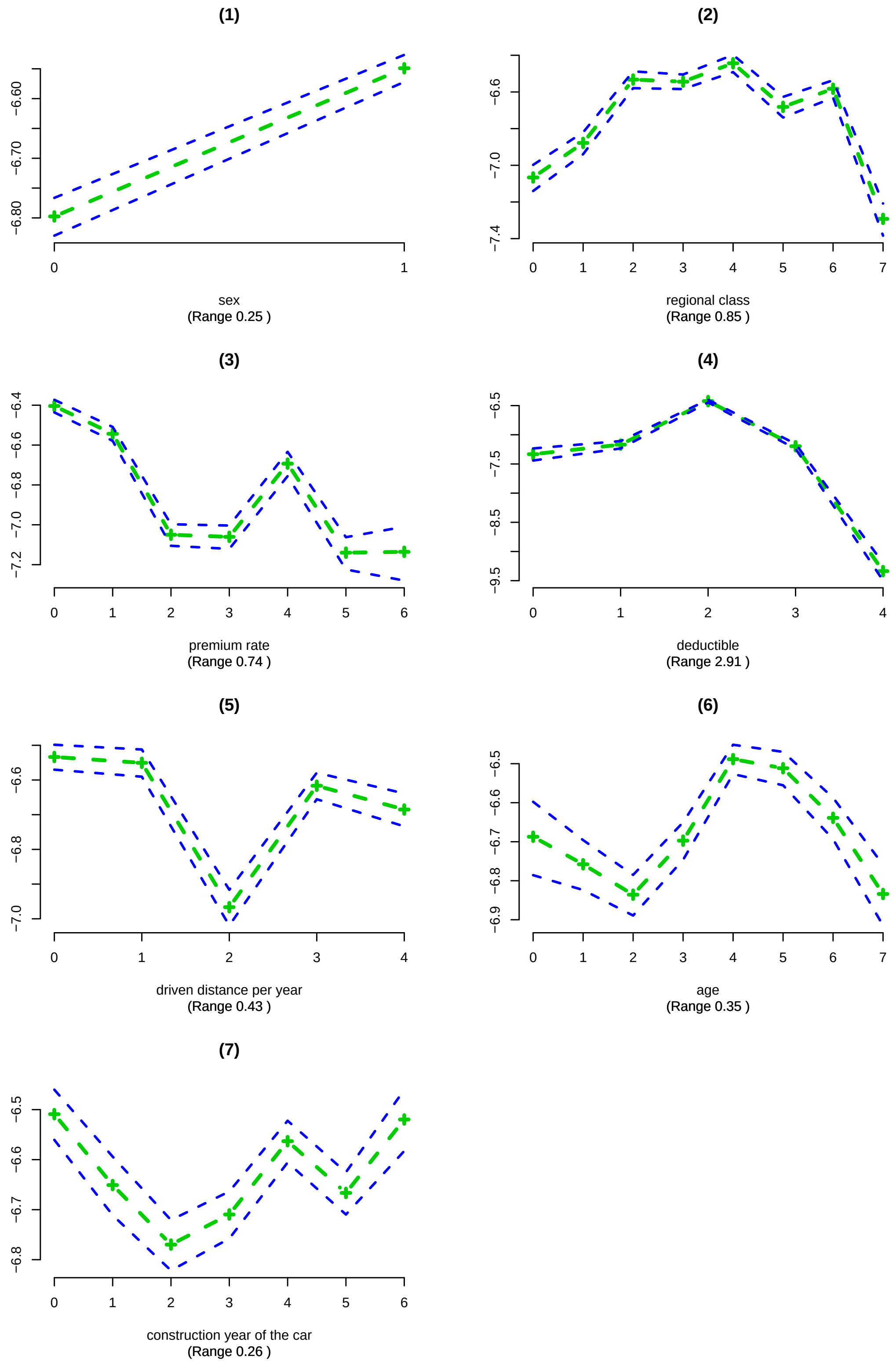

Figure 6.8: Empirical log means of transformed number of claims for each class of the categorical covariates; (1) - log means for sex, (2) - log means for regional class, (3) - log means for premium rate, (4) - log means for deductible, (5) - log means for driven distance per year, (6) - log means for age, (7) - log means for construction year

```
Z <- cbind( d.sex1,
            d.rcl1, d.rcl2, d.rcl3, d.rcl4, d.rcl5, d.rcl6,
            d.prem1, d.prem2, d.prem3, d.prem4, d.prem5, d.prem6,
            d.deduct1, d.deduct2, d.deduct3,  d.deduct4,
            d.drivdist1, d.drivdist2, d.drivdist3, d.drivdist4,
            d.age1, d.age2, d.age3, d.age4, d.age5, d.age6, d.age7 )
```

We use the R-function `mle.truncpois()` given in Appendix C.1 to calculate the MLE of the regression parameters. In the summary table (cp. Table 6.5) we find the value, the standard error and the z-value of the MLE for the regression parameters. In the following $\hat{\boldsymbol{\beta}}_m$ denotes the resulting MLE values of the marginal Zero-Truncated Poisson GLM.

6.3 Parameter Estimation for the Joint Gamma-Poisson Distribution of the Average Claim Size and the Number of Claims

In this section we estimate the joint distribution of the average claim size and the number of claims for the policies groups of the German full comprehensive car insurance. Thereby, we use the aggregated data set as described in Section 6.1. In Section 6.2 we made the covariate selection for the marginal Gamma GLM of the average claim size and for the marginal Poisson GLM using a Zero-truncated Poisson GLM. Further, we calculated the MLE of the marginal regression parameters $\hat{\boldsymbol{\alpha}}_m$ and $\hat{\boldsymbol{\beta}}_m$ of both GLMs and the constant coefficient of variation ν (the squared MLE $\hat{\nu}^2$ is equal to the MLE of the dispersion parameter $\hat{\varphi}$) of the marginal Gamma GLM. So we now have the necessary information to start with the simultaneous calculation of the parameter MLEs of the joint Gamma-Poisson regression model (see Chapter 3) for the average claim size and the number of claims. For this purpose, we use the MBP algorithm given in Algorithm 4.3 with the ρ_w-update. As already mentioned, we take the results $\hat{\boldsymbol{\alpha}}_m$ and $\hat{\boldsymbol{\beta}}_m$ provided by the marginal parameter estimation of the Gamma GLM (see Subsection 6.2.1) and of the Zero-truncated Poisson GLM as the initial values in the algorithm. Note that the constant coefficient of variation is not changed in the MBP algorithm and so the marginal calculated $\hat{\nu}$ is already the ultimate value. As stop criterion for the maximization algorithm we take $\boldsymbol{\theta}_1^k - \boldsymbol{\theta}_1^{k-1} < 10^{-3}$, where $\boldsymbol{\theta}_1^k = (\boldsymbol{\alpha}^{k\prime}, \boldsymbol{\beta}^{k\prime})'$ are the regression parameters after the k-th iteration and $||\rho^k - rho^k|| < 10^{-4}$, where ρ^k denotes the correlation parameter after the k-th iteration. Additionally, we estimate the standard error for each parameter using the inference given in Section 4.3.

Table 6.6, Table 6.7 and Table 6.8 show the resulting MLE values for the insurance portfolio calculated by applying the MBP algorithm 4.3 on the joint Gamma-Poisson regression model. In Table 6.6 we find the result $\hat{\boldsymbol{\alpha}}$ for the regression parameter $\boldsymbol{\alpha}$ of the marginal Gamma GLM, in Table 6.7 the result $\hat{\boldsymbol{\beta}}$ for the regression parameter $\boldsymbol{\alpha}$ of the marginal Poisson GLM and in Table 6.8 the resulting MLE for the correlation parameter of the Gaussian copula ρ. Table 6.6 and Table 6.7 show the name of the covariates for which the regression parameter is estimated in the left column, the second column named "Estimate" gives the estimated value of the regression parameter and in the one denoted with "Std.Error" we can see the estimated standard error for the given estimation value. The last column shows the t-value which is the quotient of the "Estimate" and the "Std. Error".

The estimated value for the correlation parameter ρ of the Gaussian copula is 0.1750 with an estimated standard error of 0.0088. This matches with our guess of the explorative data analysis

	Estimate	Std. Error	t value
(Intercept)	−7.5949	0.8736	−8.69
d.sex1	0.3190	0.7947	0.40
d.rcl1	0.2429	0.2182	1.11
d.rcl2	0.4427	0.3578	1.24
d.rcl3	0.5509	0.4941	1.11
d.rcl4	0.4568	0.3563	1.28
d.rcl5	0.3494	0.2602	1.34
d.rcl6	0.4403	0.3438	1.28
d.rcl7	−0.1785	0.0988	−1.81
d.prem1	−0.1277	0.4564	−0.28
d.prem2	−0.6594	0.1638	−4.03
d.prem3	−0.7978	0.1333	−5.98
d.prem4	−0.6682	0.1328	−5.03
d.prem5	−0.8860	0.0824	−10.75
d.prem6	−0.9871	0.0443	−22.27
d.deduct1	0.2711	0.1321	2.05
d.deduct2	1.1402	0.8349	1.37
d.deduct3	0.4440	0.2094	2.12
d.deduct4	−1.1795	0.0196	−60.11
d.drivdist1	−0.0950	0.4185	−0.23
d.drivdist2	−0.3551	0.2430	−1.46
d.drivdist3	−0.0492	0.4412	−0.11
d.drivdist4	−0.1775	0.2911	−0.61
d.age1	−0.0779	0.1338	−0.58
d.age2	0.0605	0.2336	0.26
d.age3	−0.0160	0.2855	−0.06
d.age4	0.1926	0.4733	0.41
d.age5	0.1373	0.4447	0.31
d.age6	0.0280	0.3742	0.07
d.age7	−0.1215	0.2054	−0.59

Table 6.5: Summary of Zero-truncated Poisson GLM

in which it seems to exist a small positive correlation between the average claim size and the number of claims of the policy groups.
Comparing the regression parameter values resulting of the joint regression inference with the regression parameter values estimated in the marginal GLMs we see that the parameter values of the joint regression inference still has the same algebraic signs as the one of the marginal GLMs, but the influence intensity, i.e. the absolute values, have changed, even when it is not extreme and not with one general direction.
The range of the fitted parameters $\mu_{i1} = \boldsymbol{x}_i{}'\boldsymbol{\alpha}$ is $[2.98, 7.72]$ and the range of the fitted parameters $\mu_{i2} = \boldsymbol{z}_i{}'\boldsymbol{\beta}$ is $[0.00, 43.54]$, but there are only 33 ($\hat{=}$ 0.4%) fitted μ_{i2} greater than 12.18. More over, $\hat{\rho} = 0.1750$ and $\nu = 0.79$ so that we can say that the calculation for the car insurance data set is falling in the parameter framework of scenario 3 or 4 of the simulation study (see Section 5.2). In this two scenarios the MBP algorithm with ρ_w-update works quite well. Hence we can assume that the results for the insurance data set are good, too.

6.4 Monte-Carlo Estimation of the Expected Total Loss

In order to estimate the expected total loss of the insurance portfolio, we use a Monte-Carlo Estimator (MCE). Therefore, we generate $R = 500$ data sets with the size of the original aggregated data set using the results of the joint regression analysis of Section 6.3. For each of the 500 generated data sets we can calculate the total loss

$$S^r = \sum_{i=1}^{n=7955} Y_{i1}^r \, Y_{i2}^r, \; r = 1, 2, \ldots, R,$$

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. The MCE $\hat{S}$ for the expected total loss of the insurance portfolio is then calculated by

$$\hat{S} = \frac{1}{R} \sum_{r=1}^{R} S^r.$$

We use a sampling algorithm similar to Algorithm 5.1. But, as our insurance portfolio contains only number of claims, we have to sample from the joint Gamma-Poisson distribution with density (3.8) conditional on the Poisson variate is greater 0. For this, we set $p_k = f_{Y_{i2}|Y_{i1}, Y_{i2} \geq 1}(y_{i2} = k | y_{i1} inAlgorithm 5.1, \mu_{i1}, \nu, \mu_{i2}, \rho)$ for $k = 1, 2, \ldots, k^*$, where

$$\begin{aligned} & f_{Y_{i2}|Y_{i1}, Y_{i2} \geq 1}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho) := \\ & \frac{f_{Y_{i2}|Y_{i1}}(y_{i2}|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)}{1 - f_{Y_{i2}|Y_{i1}}(0|y_{i1}, \mu_{i1}, \nu, \mu_{i2}, \rho)} = \\ & \frac{D_\rho(G_1(y_{i1}|\mu_{i1}, \nu), G_2(y_{i2}|\mu_{i2})) - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu), G_2(y_{i2} - 1|\mu_{i2}))}{1 - D_\rho(G_1(y_{i1}|\mu_{i1}, \nu), G_2(0|\mu_{i2}))}, \end{aligned} \tag{6.4}$$

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 (5.1). The parameter setting for the data simulation is the following:

$$\begin{aligned} \mu_{i1} &= \boldsymbol{x}_i{}'\hat{\boldsymbol{\alpha}} \\ \nu &= \hat{\nu} \\ \mu_{i2} &= \boldsymbol{z}_i{}'\hat{\boldsymbol{\beta}} \\ \rho &= \hat{\rho}. \end{aligned}$$

Gamma GLM Regression Parameter			
	Estimate	Std. Error	t value
(Intercept)	1.3978	0.1235	11.32
d.deduct1	0.1898	0.0583	3.26
d.deduct2	0.3347	0.0511	6.55
d.deduct3	0.3912	0.0538	7.27
d.deduct4	0.3739	0.0878	4.26
d.age1	−0.0806	0.1409	−0.57
d.age2	−0.1217	0.1325	−0.92
d.age3	−0.2311	0.1312	−1.76
d.age4	−0.1112	0.1224	−0.91
d.age5	−0.1478	0.1255	−1.18
d.age6	−0.3067	0.1371	−2.24
d.age7	−0.1240	0.1570	−0.79
d.constyear1	0.1756	0.1834	0.96
d.constyear2	0.1849	0.1613	1.15
d.constyear3	−0.0840	0.1655	−0.51
d.constyear4	0.2545	0.1530	1.66
d.constyear5	0.2391	0.1517	1.58
d.constyear6	0.1208	0.1882	0.64
d.age1:d.constyear1	−0.1706	0.2238	−0.76
d.age1:d.constyear2	−0.0522	0.1990	−0.26
d.age1:d.constyear3	0.2014	0.2004	1.01
d.age1:d.constyear4	−0.0565	0.1878	−0.30
d.age1:d.constyear5	−0.0736	0.1848	−0.40
d.age1:d.constyear6	0.1809	0.2364	0.77
d.age2:d.constyear1	−0.1279	0.2109	−0.61
d.age2:d.constyear2	−0.1588	0.1877	−0.85
d.age2:d.constyear3	0.1501	0.1881	0.80
d.age2:d.constyear4	−0.1597	0.1766	−0.90
d.age2:d.constyear5	0.0345	0.1759	0.20
d.age2:d.constyear6	0.0754	0.2216	0.34
d.age3:d.constyear1	−0.1151	0.2087	−0.55
d.age3:d.constyear2	0.0007	0.1839	0.00
d.age3:d.constyear3	0.2378	0.1869	1.27
d.age3:d.constyear4	−0.0596	0.1742	−0.34
d.age3:d.constyear5	−0.0014	0.1739	−0.01
d.age3:d.constyear6	0.2269	0.2132	1.06
d.age4:d.constyear1	−0.0568	0.1966	−0.29
d.age4:d.constyear2	−0.1114	0.1747	−0.64
d.age4:d.constyear3	0.2197	0.1777	1.24
d.age4:d.constyear4	−0.1896	0.1651	−1.15
d.age4:d.constyear5	−0.1091	0.1642	−0.66
d.age4:d.constyear6	0.0343	0.2030	0.17
d.age5:d.constyear1	−0.1283	0.2023	−0.63
d.age5:d.constyear2	−0.0259	0.1793	−0.14
d.age5:d.constyear3	0.2428	0.1821	1.33
d.age5:d.constyear4	−0.1469	0.1687	−0.87
d.age5:d.constyear5	−0.0261	0.1673	−0.16
d.age5:d.constyear6	0.0892	0.2058	0.43
d.age6:d.constyear1	−0.0110	0.2140	−0.05
d.age6:d.constyear2	−0.0225	0.1925	−0.12
d.age6:d.constyear3	0.2118	0.1947	1.09
d.age6:d.constyear4	−0.0403	0.1804	−0.22
d.age6:d.constyear5	−0.0928	0.1792	−0.52
d.age6:d.constyear6	0.1029	0.2162	0.48
d.age7:d.constyear1	−0.1060	0.2445	−0.43
d.age7:d.constyear2	−0.1142	0.2257	−0.51
d.age7:d.constyear3	0.2511	0.2213	1.13
d.age7:d.constyear4	−0.0895	0.2032	−0.44
d.age7:d.constyear5	−0.1837	0.2032	−0.90
d.age7:d.constyear6	−0.1557	0.2455	−0.63

Table 6.6: Summary of the Marginal Gamma GLM Parameter calculated by MBP with ρ_p-update

Poisson GLM Regression Parameter			
	Estimate	Std. Error	t value
(Intercept)	−7.2345	0.0978	−73.99
d.sex1	0.2878	0.0210	13.70
d.rcl1	0.2015	0.0516	3.91
d.rcl2	0.3817	0.0460	8.30
d.rcl3	0.4695	0.0437	10.73
d.rcl4	0.3645	0.0463	7.87
d.rcl5	0.2657	0.0500	5.32
d.rcl6	0.3617	0.0466	7.76
d.rcl7	−0.2022	0.0655	−3.09
d.prem1	−0.1504	0.0232	−6.48
d.prem2	−0.6265	0.0363	−17.25
d.prem3	−0.7665	0.0389	−19.68
d.prem4	−0.7182	0.0403	−17.80
d.prem5	−0.9741	0.0519	−18.76
d.prem6	−1.1289	0.0824	−13.70
d.deduct1	0.2194	0.0721	3.04
d.deduct2	1.0060	0.0628	16.01
d.deduct3	0.3660	0.0670	5.46
d.deduct4	−0.9518	0.1122	−8.48
d.drivdist1	−0.0970	0.0251	−3.87
d.drivdist2	−0.3391	0.0332	−10.23
d.drivdist3	−0.0591	0.0251	−2.35
d.drivdist4	−0.1678	0.0307	−5.47
d.age1	−0.1372	0.0694	−1.98
d.age2	−0.0365	0.0663	−0.55
d.age3	−0.1254	0.0656	−1.91
d.age4	0.0657	0.0629	1.04
d.age5	0.0112	0.0641	0.18
d.age6	−0.1012	0.0665	−1.52
d.age7	−0.2361	0.0735	−3.21

Table 6.7: Summary of the Marginal Poisson GLM Parameter calculated by MBP with ρ_p-update

Correlation Parameter of Gaussian Copula			
	Estimate	Std. Error	t value
ρ	0.1750	0.0088	19.79

Table 6.8: Summary of the Correlation Parameter ρ calculated by MBP with ρ_p-update

For comparison reasons, we do the same using the results $\hat{\boldsymbol{\alpha}}_m$ and $\hat{\boldsymbol{\beta}}_m$ of the independent marginal regression analysis of the Gamma GLM (see Section 6.2.1) and the Zero-truncated Poisson GLM (see Section 6.2.2) under the assumption of independency between the number of claims and the average claim size. So the parameter setting for the data generation is as follows:

$$\begin{aligned}\mu_{i1} &= \boldsymbol{x}_i{}'\hat{\boldsymbol{\alpha}}_m \\ \nu &= \hat{\nu} \\ \mu_{i2} &= \boldsymbol{z}_i{}'\hat{\boldsymbol{\beta}}_m \\ \rho &= 0.\end{aligned}$$

So we get the total loss S^r_{ind}, $r = 1, 2, \ldots, R$, of the simulated data sets with independent claim frequency and claim size. The MCE of the expected total loss is then

$$\hat{S}_{ind} = \frac{1}{R}\sum_{r=1}^{R} S^r_{ind}.$$

The MCE for the expected total loss of the full comprehensive car insurance portfolio using the resulting distribution parameter of the joint regression analysis provides 74'482 TDM with a standard error of 1'119 TDM. When we assume independency between the number of claims and the average claim size, the resulting MCE for the total loss of the insurance portfolio is 77'715 TDM with a standard error of 1'110 TDM. The total loss of the observed car insurance portfolio lies between both estimators and has the amount of 76'071 TDM.
Figure 6.4(1) shows the histogram of the 500 simulated total losses for the joint regression model not assuming independency between the number of claims and the average claim size, while Figure 6.4 represents the histogram of the the independent marginal GLMs. The red vertical line marks the total loss (76'071 TDM) of the observed car insurance portfolio. The green vertical line tags the value of the MCE $\hat{S}$ in Figure 6.4(1) and the value of the MCE $\hat{S}_{ind}$ in Figure 6.4(2). For the classical approach, 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:

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

with $\mu_{i1} = \boldsymbol{x}_i{}'\boldsymbol{\alpha}$ and $\mu_{i2} = \boldsymbol{z}_i{}'\boldsymbol{\beta}$. So an estimator for the expected total loss is given by

$$\widehat{E[S_{ind}]} := \sum_{i=1}^{n} \widehat{\mu_{i1}}\widehat{\mu_{i2}},$$

where $\widehat{\mu_{i1}} = \boldsymbol{x}_i{}'\hat{\boldsymbol{\alpha}}_m$ are the fits of the the marginal Gamma GLM of Section 6.2.1 and $\mu_{i2} = \boldsymbol{z}_i{}'\hat{\boldsymbol{\beta}}_m$ are the fits of the marginal Zero-truncated Poisson GLM of Section 6.2.2. The result for $\widehat{E[S_{ind}]}$ is 77'899 TDM. In Figure 6.4(2), $\widehat{E[S_{ind}]}$ is marked with the blue vertical line. The MCE $\hat{S}_{ind}$ provides a quite similar value as the estimator $\widehat{E[S_{ind}]}$, which shows that the simulation works properly.
We see that the estimated expected total loss using the joint regression model is about 5% smaller than the estimated expected total loss using the independent regression models. This can be explained 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.
Another point to mention here is the possible over- or underestimation in the different approaches. When we regard the joint regression model, we see that the MCE for the total loss underestimates the observed total loss of the insurance portfolio, while the estimator using the independent regression models overestimates. But what we can not say here is whether the over- or underestimation can be considered to be systematic as we only have one available total loss observation for the insurance portfolio.

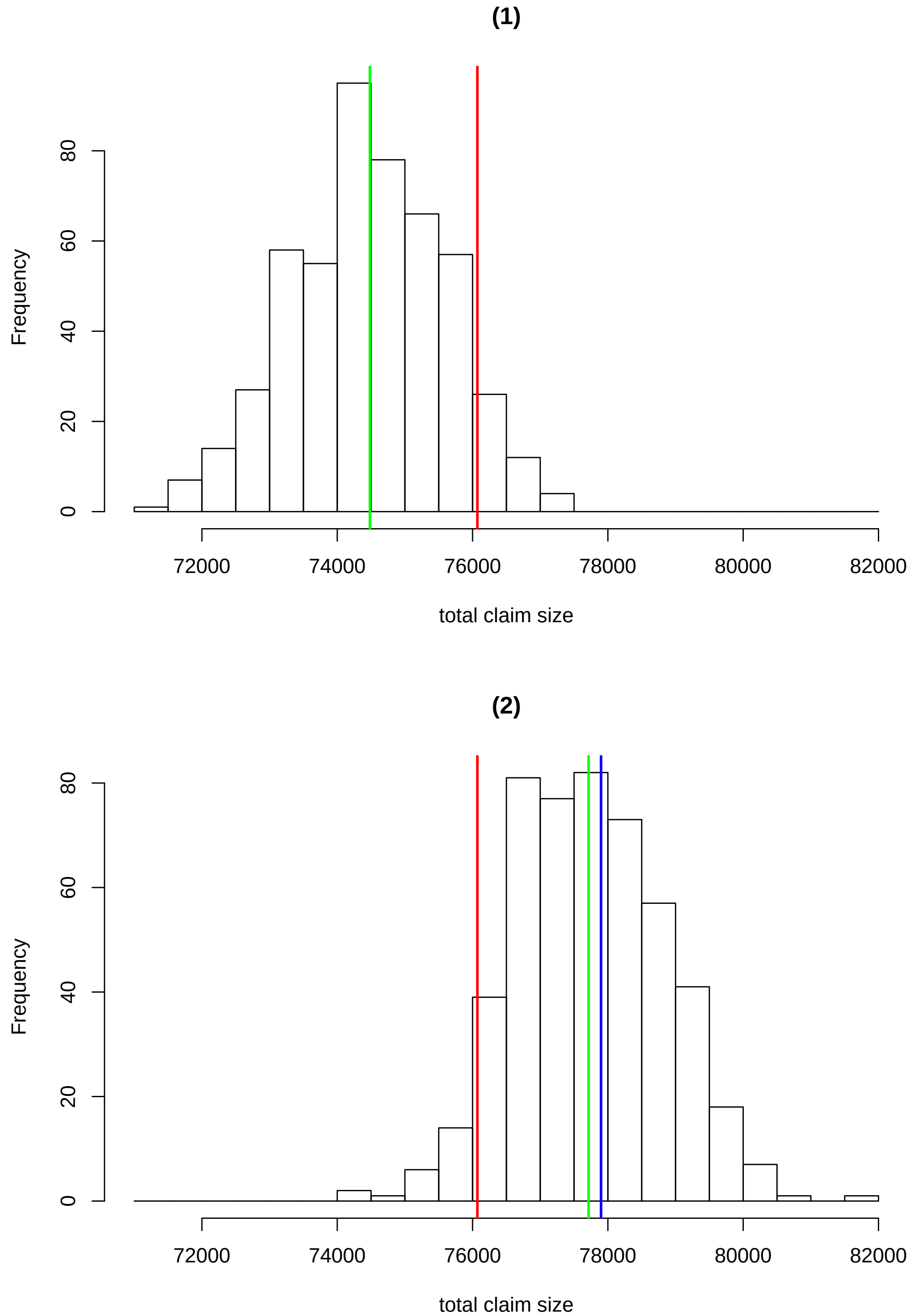

Figure 6.9: Histogram of the estimated expected total loss using the results of the joint regression model (Figure (1)) and of the independent marginal GLMs (Figure (2))

7 Conclusion

The intention of the paper was to construct a model for the average claim size and the number of claims of insurance policies which involves a possible dependency between both variables. The target was to improve the estimation of the total loss of an insurance portfolio. Therefore, 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. We connected both GLMs with the Mixed Copula Approach and got the Gamma-Poisson regression model. As copula function we took the Gaussian Copula which has one parameter to model the dependency structure between the average claim size and the number of claims. Here, it would also have been possible to use another copula function, for example a two parametric copula function which might map the dependency structure in a better way. Another possible expansion of the used joint regression model would have been to take a Generalized Poisson GLM - instead of the Poisson GLM - in order to model over- and underdispersion in the count variable "Number of Claims".
In order to fit the joint Gamma-Poisson regression model to a data set and to calculate the MLEs for the regression parameters as well as the correlation parameter of the Gaussian copula we used the MBP algorithm first developed by Song, Fan, and Kalbfleisch (2005). Based on this approach we constructed a MBP algorithm for our joint Gamma-Poisson regression model. We checked the quality of this algorithm by running a simulation study with 24 scenarios. For this propose we developed a sampling algorithm to generate observations from the Gamma-Poisson model. The result of the simulation study is that the constructed MBP algorithm for the Gamma-Poisson regression model works quite well in most of the scenarios, especially when a low or medium value was chosen for the correlation parameter. But this should also be checked theoretically for a better understanding of the algorithm's behavior, what was not discussed in this thesis.
We then applied the model on a full comprehensive car insurance portfolio of a German insurance company. It turned out that there is a 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, too.
Finally, we used the Monte Carlo method to estimate the expected total loss for the portfolio by using the results of the joint regression analysis of the real data set with the Gamma-Poisson regression model. In comparison with the classical model, where the average claim size and the number of claims are modeled independently, it was shown that the expected total loss estimated with the joint regression model is smaller than the one estimated with the classical model. In order to check whether this behavior is systematic more data needs to be analyzed.
As there are many topics which we have not examined yet in this paper, there still is a lot of future research to do.

Bibliography

Czado, C. (2004). Lecture slides on glm. http://www-m4.ma.tum.de/courses/GLM/slides.html.

Georgii, H.-O. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Volume 139. Berlin; New York: de Gruyter.

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

Liang, K. Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

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

Mack, T. (1997). *Schadenversicherungsmathematik*. Karlsruhe: Verlag Versicherungswirtschaft.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General) 135*(3), 370–384.

Nelsen, R. B. (1999). *An introduction to copoulas*, Volume 139 of *Lecture notes in statistics*. New York: Springer.

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

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.

Stekeler, D. (2004, October). Verallgemeinerte poissonregression und daraus abgeleitete zero-inflated und zero-hurdle regressionsmodelle. Diploma thesis, Technische Unversität München.

A Policy Groups Covariates

sex

category	0	1
dummy variable	*	d.sex1
values	female	male
#	2697	5258
%	33.9	66.1

Note: * marks the reference category

Table A.1: Categories with the absolute and relative frequencies of the covariate 'sex'

regional class

category	0			1	2	3	4	5	6	7	
dummy variable	*			d.rcl1	d.rcl2	d.rcl3	d.rcl4	d.rcl5	d.rcl6	d.rcl7	
values	0	1	2	3	4	5	6	7	8	9	10
#	535			764.0	1295	1718	1233	856	1200	354	
%	6.73			9.60	16.28	21.60	15.50	10.76	15.08	4.45	

Note: * marks the reference category

Table A.2: Categories with the absolute and relative frequencies of the covariate 'regional class'

premium rate

category	0		1		2		3		4		5		6	
dummy variable	*		d.premrate1		d.premrate2		d.premrate3		d.premrate4		d.premrate5		d.premrate6	
values	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	90%	100%	115%	190%
#	2723		2220		929		789		729		415		150	
%	34.23		27.91		11.68		9.92		9.16		5.22		1.89	

Note: * marks the reference category

Table A.3: Categories with the absolute and relative frequencies of the covariate 'premium rate' deductible

category	0	1	2	3	4	
dummy variable	*	d.deduct1	d.deduct2	d.deduct3	d.deduct4	
values	0	300	650	1000	1500	2000
#	254	674	5339	1567	121	
%	3.19	8.47	67.12	19.70	1.52	

Note: * marks the reference category

Table A.4: Categories with the absolute and relative frequencies of the covariate 'deductible' in DM

driven distance per year

category	0	1	2	3	4
dummy variable	*	d.drivdist1	d.drivdist2	d.drivdist3	d.drivdist4
values	[1,9]	(9,12]	(12,15]	(15,20]	(20,300]
#	2100	1766	1021	1861	1207
%	26.40	22.2	12.83	23.39	15.17

Note: * marks the reference category

Table A.5: Categories with the absolute and relative frequencies of the covariate 'driven distance per year' in 1000 km

age

category	0	1	2	3	4	5	6	7
dummy variable	*	d.age1	d.age2	d.age3	d.age4	d.age5	d.age6	d.age7
values	[18,25]	(25,30]	(30,35]	(35,40]	(40,50]	(50,60]	(60,70]	(70,100]
#	308	664	1000	1176	1915	1480	963	449
%	3.87	8.35	12.57	14.78	24.07	18.60	12.11	5.64

Note: * marks the reference category

Table A.6: Categories with the absolute and relative frequencies of the covariate 'age' in years construction year of the car

category	0	1	2	3	4	5	6
dummy variable	*	d.constyear1	d.constyear2	d.constyear3	d.constyear4	d.constyear5	d.constyear6
values	[1966, 1994]	(1994, 1995]	(1995, 1996]	(1996, 1997]	(1997, 1998]	(1998, 1999]	(1999, 2000]
#	1076	792	1066	1223	1555	1518	725
%	13.53	9.96	13.40	15.37	19.55	19.08	9.11

Note: * marks the reference category

Table A.7: Categories with the absolute and relative frequencies of the covariate "construction year of the car"

B Simulation Study Results

In Appendix B we find the detailed results of the simulation study. In the simulation study we test both versions of the MBP algorithm 4.3, the one without ρ_w-update and the one with ρ_w-update. In Table B.1 - Table B.6 the results of the version without ρ_w-update are given and in Table B.7 - Table B.12 we find the results by using the version with ρ_w-update. Each table contains four different scenarios. In the tables the first column gives the parameter names and the second column, labeled with $\boldsymbol{\theta}$, contains the true values of the parameter. In column $\overline{\boldsymbol{\theta}}$ we find the mean of the 500 independent identically distributed realizations of the MLEs for each parameter, calculated according to equation (2.29). The standard deviation $s_{\overline{\theta}}$ of the 500 realizations of the MLEs scaled with 10^2 is given in the next column. It is the square root of the sample variance given in equation (2.32). The following two columns contain the with 10^2 scaled estimated bias $\widehat{b}$ and its with 10^3 scaled estimated standard deviation $s_{\widehat{b}}$ which is the square root of the variance of the estimated bias calculated according to equation (2.31). The formula of the estimated bias can be found in equation (2.26). The column $\widehat{mse}$ shows the by 10^3 scaled mean squared error of the 500 realized MLEs provided by the MBP algorithm, calculated with the formula of equation (2.34). Its estimated standard deviation $s_{\widehat{mse}} = \sqrt{Var[\widehat{mse}]}$ scaled with 10^3 is given in the next column. The calculation formula of $Var[\widehat{mse}]$ can be found in equation (2.37). In the last column we can see the ratio of the sample mean and the sample standard deviation $\overline{\theta}/s_{\overline{\theta}}$. This ratio has to be greater than 2 in order to have a significant estimator of the corresponding parameter.

B.1 Results using the MBP Algorithm without ρ_w-update

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
1	α_1	1.00	1.01	3.55	1.17	1.59	1.39	0.09	28.54
	α_2	1.00	1.00	5.99	−0.34	2.68	3.60	0.22	16.64
	β_1	−1.00	−1.00	9.48	−0.10	4.24	8.98	0.53	10.56
	β_2	3.00	3.00	11.97	0.43	5.35	14.35	0.85	25.09
	ρ	0.10	0.09	3.78	−0.66	1.69	1.48	0.09	2.47
	ν	0.50	0.50	1.29	−0.09	0.58	0.17	0.01	38.80
2	α_1	1.00	1.00	6.82	−0.19	3.05	4.65	0.30	14.64
	α_2	1.00	1.01	11.67	0.59	5.22	13.65	0.86	8.62
	β_1	−1.00	−1.00	9.72	0.20	4.35	9.45	0.57	10.27
	β_2	3.00	3.00	12.21	−0.21	5.46	14.91	0.91	24.55
	ρ	0.10	0.10	4.00	−0.09	1.79	1.60	0.09	2.48
	ν	1.00	1.00	2.86	−0.20	1.28	0.82	0.05	34.93
3	α_1	1.00	1.01	3.36	0.65	1.50	1.17	0.08	29.99
	α_2	1.00	1.00	6.09	−0.12	2.72	3.71	0.29	16.39
	β_1	−0.50	−0.50	6.48	−0.23	2.90	4.21	0.27	7.75
	β_2	3.00	3.00	8.44	0.46	3.77	7.14	0.44	35.61
	ρ	0.10	0.10	3.82	−0.40	1.71	1.47	0.09	2.52
	ν	0.50	0.50	1.26	−0.15	0.56	0.16	0.01	39.64
4	α_1	1.00	1.00	6.32	0.46	2.83	4.02	0.27	15.88
	α_2	1.00	0.99	10.87	−0.95	4.86	11.90	0.81	9.11
	β_1	−0.50	−0.50	6.65	0.08	2.97	4.43	0.29	7.50
	β_2	3.00	3.00	8.57	−0.10	3.83	7.35	0.49	34.98
	ρ	0.10	0.10	3.78	−0.04	1.69	1.43	0.09	2.64
	ν	1.00	1.00	3.22	−0.17	1.44	1.04	0.05	30.99

Table B.1: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.1$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
5	α_1	1.00	1.01	3.67	1.28	1.64	1.51	0.09	27.63
	α_2	3.00	2.99	6.12	−1.41	2.74	3.95	0.24	48.76
	β_1	−1.00	−1.00	9.62	0.39	4.30	9.26	0.64	10.36
	β_2	3.00	3.00	11.90	−0.38	5.32	14.17	0.99	25.18
	ρ	0.10	0.10	3.72	−0.48	1.67	1.41	0.09	2.56
	ν	0.50	0.50	1.31	−0.04	0.58	0.17	0.01	38.27
6	α_1	1.00	1.00	6.42	−0.20	2.87	4.12	0.27	15.56
	α_2	3.00	3.00	11.10	−0.07	4.96	12.31	0.80	27.03
	β_1	−1.00	−1.00	9.21	−0.18	4.12	8.49	0.50	10.87
	β_2	3.00	3.00	11.68	0.16	5.22	13.63	0.83	25.71
	ρ	0.10	0.10	4.03	−0.17	1.80	1.63	0.11	2.44
	ν	1.00	1.00	3.13	−0.06	1.40	0.98	0.06	31.90
7	α_1	1.00	1.01	3.34	0.68	1.49	1.16	0.08	30.17
	α_2	3.00	2.99	5.71	−0.88	2.55	3.34	0.20	52.38
	β_1	−0.50	−0.51	6.71	−1.02	3.00	4.61	0.26	7.60
	β_2	3.00	3.01	8.62	1.22	3.85	7.57	0.43	34.96
	ρ	0.10	0.10	3.76	−0.31	1.68	1.42	0.09	2.58
	ν	0.50	0.50	1.25	−0.02	0.56	0.16	0.01	40.02
8	α_1	1.00	1.00	6.07	−0.38	2.71	3.70	0.24	16.41
	α_2	3.00	3.00	10.61	−0.11	4.74	11.25	0.71	28.27
	β_1	−0.50	−0.50	6.48	−0.31	2.90	4.21	0.30	7.76
	β_2	3.00	3.00	8.48	0.31	3.79	7.19	0.50	35.43
	ρ	0.10	0.10	3.85	0.13	1.72	1.49	0.09	2.63
	ν	1.00	1.00	3.08	−0.27	1.38	0.96	0.06	32.36

Table B.2: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.1$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
9	α_1	1.00	1.06	3.39	5.85	1.51	4.56	0.19	31.26
	α_2	1.00	0.99	5.93	−1.42	2.65	3.72	0.23	16.61
	β_1	−1.00	−0.91	8.05	8.54	3.60	13.78	0.67	11.37
	β_2	3.00	2.93	9.73	−7.38	4.35	14.91	0.80	30.08
	ρ	0.50	0.49	3.16	−0.74	1.41	1.06	0.07	15.57
	ν	0.50	0.50	1.23	−0.04	0.55	0.15	0.01	40.75
10	α_1	1.00	1.00	6.08	0.38	2.72	3.71	0.25	16.52
	α_2	1.00	0.99	10.07	−0.80	4.50	10.21	0.73	9.85
	β_1	−1.00	−1.00	8.50	−0.10	3.80	7.22	0.41	11.78
	β_2	3.00	3.00	10.43	0.11	4.67	10.89	0.64	28.76
	ρ	0.50	0.50	3.01	0.05	1.35	0.90	0.05	16.64
	ν	1.00	1.00	3.23	−0.05	1.44	1.04	0.07	30.97
11	α_1	1.00	1.04	2.99	3.67	1.34	2.24	0.11	34.64
	α_2	1.00	0.99	5.36	−1.37	2.40	3.06	0.19	18.41
	β_1	−0.50	−0.47	5.68	3.44	2.54	4.41	0.28	8.20
	β_2	3.00	2.97	7.07	−2.75	3.16	5.75	0.37	42.07
	ρ	0.50	0.50	2.71	−0.20	1.21	0.74	0.05	18.38
	ν	0.50	0.50	1.24	0.00	0.55	0.15	0.01	40.33
12	α_1	1.00	1.00	6.06	−0.26	2.71	3.68	0.21	16.45
	α_2	1.00	1.00	10.32	0.06	4.62	10.66	0.62	9.69
	β_1	−0.50	−0.50	5.74	−0.19	2.57	3.30	0.20	8.75
	β_2	3.00	3.00	6.97	0.06	3.12	4.86	0.29	43.06
	ρ	0.50	0.50	2.74	0.11	1.23	0.75	0.05	18.27
	ν	1.00	1.00	3.12	−0.03	1.39	0.97	0.06	32.09

Table B.3: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.5$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
13	α_1	1.00	1.06	3.43	6.04	1.53	4.82	0.19	30.90
	α_2	3.00	2.94	5.80	−6.37	2.59	7.42	0.37	50.64
	β_1	−1.00	−0.95	8.67	5.11	3.88	10.12	0.59	10.95
	β_2	3.00	2.96	10.98	−4.34	4.91	13.94	0.85	26.93
	ρ	0.50	0.50	2.90	−0.44	1.30	0.86	0.06	17.10
	ν	0.50	0.50	1.21	−0.10	0.54	0.15	0.01	41.37
14	α_1	1.00	1.00	6.25	0.06	2.79	3.90	0.24	16.01
	α_2	3.00	3.00	10.75	0.23	4.81	11.57	0.67	27.92
	β_1	−1.00	−1.00	8.25	0.13	3.69	6.81	0.43	12.11
	β_2	3.00	3.00	10.29	−0.04	4.60	10.58	0.65	29.16
	ρ	0.50	0.50	3.12	0.04	1.39	0.97	0.06	16.06
	ν	1.00	1.00	3.37	−0.29	1.51	1.14	0.07	29.61
15	α_1	1.00	1.03	3.03	3.34	1.36	2.04	0.10	34.10
	α_2	3.00	2.96	5.33	−3.72	2.38	4.23	0.23	55.57
	β_1	−0.50	−0.48	5.54	2.10	2.48	3.51	0.22	8.64
	β_2	3.00	2.98	7.07	−1.74	3.16	5.30	0.33	42.19
	ρ	0.50	0.50	2.58	−0.22	1.15	0.67	0.04	19.33
	ν	0.50	0.50	1.25	−0.02	0.56	0.16	0.01	40.04
16	α_1	1.00	1.01	5.81	0.64	2.60	3.42	0.20	17.31
	α_2	3.00	2.99	10.31	−1.14	4.61	10.77	0.62	28.98
	β_1	−0.50	−0.50	5.95	0.27	2.66	3.54	0.21	8.36
	β_2	3.00	3.00	7.64	−0.34	3.42	5.84	0.34	39.24
	ρ	0.50	0.50	2.85	0.02	1.28	0.82	0.05	17.52
	ν	1.00	1.00	3.28	0.05	1.46	1.07	0.07	30.55

Table B.4: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.5$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
17	α_1	1.00	1.06	2.22	5.88	0.99	3.95	0.12	47.67
	α_2	1.00	0.99	3.47	−0.54	1.55	1.23	0.07	28.68
	β_1	−1.00	−0.86	4.92	14.08	2.20	22.25	0.63	17.47
	β_2	3.00	2.89	5.66	−11.32	2.53	16.01	0.60	51.01
	ρ	0.90	0.90	0.70	0.24	0.31	0.05	0.00	129.47
	ν	0.50	0.50	1.23	−0.06	0.55	0.15	0.01	40.64
18	α_1	1.00	0.90	6.31	−9.64	2.82	13.27	0.62	14.32
	α_2	1.00	1.01	6.49	1.19	2.90	4.35	0.29	15.60
	β_1	−1.00	−1.11	8.06	−11.43	3.60	19.57	0.93	13.83
	β_2	3.00	3.10	8.68	9.71	3.88	16.95	0.81	35.69
	ρ	0.90	0.90	0.67	−0.39	0.30	0.06	0.00	133.77
	ν	1.00	1.00	3.14	−0.10	1.40	0.99	0.06	31.84
19	α_1	1.00	0.99	2.30	−0.89	1.03	0.61	0.04	43.09
	α_2	1.00	1.00	3.51	−0.04	1.57	1.23	0.08	28.49
	β_1	−0.50	−0.52	4.44	−1.95	1.99	2.35	0.15	11.70
	β_2	3.00	3.02	5.49	1.63	2.46	3.28	0.21	54.91
	ρ	0.90	0.90	0.67	0.02	0.30	0.05	0.00	133.61
	ν	0.50	0.50	1.28	0.08	0.57	0.16	0.01	39.17
20	α_1	1.00	0.84	5.60	−15.54	2.51	27.30	0.82	15.07
	α_2	1.00	1.00	6.02	0.27	2.69	3.63	0.22	16.66
	β_1	−0.50	−0.67	6.10	−16.99	2.73	32.61	0.96	10.98
	β_2	3.00	3.16	6.64	16.08	2.97	30.26	0.97	47.58
	ρ	0.90	0.89	0.75	−0.61	0.33	0.09	0.01	119.62
	ν	1.00	1.00	3.24	−0.15	1.45	1.05	0.07	30.81

Table B.5: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.9$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
21	α_1	1.00	1.06	2.35	5.68	1.05	3.79	0.12	44.90
	α_2	3.00	2.95	4.41	−5.49	1.97	4.95	0.25	66.84
	β_1	−1.00	−0.93	5.91	7.32	2.64	8.85	0.45	15.69
	β_2	3.00	2.94	6.69	−5.81	2.99	7.86	0.46	43.96
	ρ	0.90	0.90	0.76	0.10	0.34	0.06	0.00	118.64
	ν	0.50	0.50	1.19	−0.03	0.53	0.14	0.01	41.98
22	α_1	1.00	0.95	5.91	−5.50	2.64	6.52	0.47	15.98
	α_2	3.00	3.05	7.37	4.69	3.30	7.64	0.50	41.32
	β_1	−1.00	−1.02	7.82	−2.01	3.50	6.52	0.46	13.05
	β_2	3.00	3.01	9.15	1.26	4.09	8.53	0.62	32.93
	ρ	0.90	0.90	0.73	−0.25	0.33	0.06	0.00	122.63
	ν	1.00	1.00	3.12	−0.23	1.40	0.98	0.07	31.93
23	α_1	1.00	1.00	2.51	−0.35	1.12	0.64	0.04	39.65
	α_2	3.00	3.01	4.52	0.70	2.02	2.10	0.14	66.46
	β_1	−0.50	−0.50	4.17	0.22	1.86	1.74	0.11	11.94
	β_2	3.00	3.00	5.24	−0.35	2.34	2.76	0.17	57.20
	ρ	0.90	0.90	0.71	0.03	0.32	0.05	0.00	126.89
	ν	0.50	0.50	1.26	0.06	0.57	0.16	0.01	39.61
24	α_1	1.00	0.91	5.21	−8.91	2.33	10.66	0.47	17.48
	α_2	3.00	3.10	7.20	9.56	3.22	14.33	0.70	42.98
	β_1	−0.50	−0.55	5.40	−4.69	2.41	5.12	0.31	10.13
	β_2	3.00	3.04	6.23	4.40	2.79	5.82	0.35	48.83
	ρ	0.90	0.90	0.74	−0.20	0.33	0.06	0.00	120.82
	ν	1.00	1.00	3.17	−0.17	1.42	1.01	0.07	31.47

Table B.6: Simulation study results using the MBP without ρ_w-update based on 500 replications when $\rho = 0.9$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

B.2 Results using the MBP Algorithm with ρ_w-update

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
1	α_1	1.00	1.01	3.51	1.18	1.57	1.37	0.09	28.81
	α_2	1.00	1.00	6.21	−0.17	2.78	3.86	0.24	16.08
	β_1	−1.00	−0.99	9.99	0.66	4.47	10.01	0.62	9.95
	β_2	3.00	2.99	12.32	−0.76	5.51	15.25	0.97	24.28
	ρ	0.10	0.09	3.88	−0.67	1.73	1.55	0.10	2.41
	ν	0.50	0.50	1.27	0.06	0.57	0.16	0.01	39.29
2	α_1	1.00	0.99	6.90	−0.71	3.09	4.81	0.30	14.39
	α_2	1.00	1.01	11.53	1.22	5.16	13.44	0.89	8.78
	β_1	−1.00	−1.00	9.38	0.22	4.20	8.81	0.57	10.63
	β_2	3.00	3.00	11.95	−0.25	5.34	14.28	0.90	25.09
	ρ	0.10	0.10	4.23	0.17	1.89	1.80	0.11	2.40
	ν	1.00	1.00	3.04	−0.14	1.36	0.93	0.06	32.80
3	α_1	1.00	1.01	3.48	0.84	1.55	1.28	0.08	29.01
	α_2	1.00	1.00	5.91	−0.40	2.64	3.51	0.23	16.84
	β_1	−0.50	−0.51	6.62	−0.54	2.96	4.42	0.30	7.63
	β_2	3.00	3.01	8.58	0.70	3.84	7.42	0.50	35.03
	ρ	0.10	0.10	3.69	−0.07	1.65	1.36	0.09	2.69
	ν	0.50	0.50	1.28	0.04	0.57	0.16	0.01	39.04
4	α_1	1.00	1.00	6.67	0.14	2.98	4.45	0.29	15.02
	α_2	1.00	1.00	11.59	−0.09	5.18	13.43	0.91	8.62
	β_1	−0.50	−0.50	6.39	−0.03	2.86	4.09	0.26	7.83
	β_2	3.00	3.00	8.40	−0.08	3.76	7.05	0.45	35.71
	ρ	0.10	0.10	3.71	−0.12	1.66	1.38	0.09	2.66
	ν	1.00	1.00	3.12	−0.05	1.40	0.98	0.07	32.01

Table B.7: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.1$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
5	α_1	1.00	1.01	3.38	1.30	1.51	1.31	0.09	29.93
	α_2	3.00	2.99	5.66	−1.28	2.53	3.37	0.22	52.77
	β_1	−1.00	−0.99	9.99	0.55	4.47	10.00	0.62	9.96
	β_2	3.00	2.99	12.33	−0.66	5.51	15.25	0.97	24.28
	ρ	0.10	0.10	3.85	−0.38	1.72	1.50	0.09	2.50
	ν	0.50	0.50	1.27	0.06	0.57	0.16	0.01	39.29
6	α_1	1.00	0.99	6.34	−0.66	2.83	4.06	0.27	15.68
	α_2	3.00	3.01	10.80	1.34	4.83	11.83	0.75	27.91
	β_1	−1.00	−1.00	9.32	0.26	4.17	8.69	0.57	10.70
	β_2	3.00	3.00	11.86	−0.28	5.30	14.07	0.89	25.27
	ρ	0.10	0.10	4.06	0.15	1.81	1.65	0.10	2.50
	ν	1.00	1.00	3.04	−0.14	1.36	0.93	0.06	32.80
7	α_1	1.00	1.01	3.36	0.73	1.50	1.18	0.07	29.98
	α_2	3.00	2.99	5.69	−0.85	2.54	3.31	0.23	52.57
	β_1	−0.50	−0.51	6.62	−0.61	2.96	4.41	0.30	7.65
	β_2	3.00	3.01	8.58	0.76	3.84	7.42	0.50	35.04
	ρ	0.10	0.10	3.67	0.06	1.64	1.35	0.09	2.74
	ν	0.50	0.50	1.28	0.04	0.57	0.16	0.01	39.04
8	α_1	1.00	1.00	6.32	0.22	2.83	4.00	0.24	15.86
	α_2	3.00	3.00	11.20	−0.17	5.01	12.54	0.78	26.78
	β_1	−0.50	−0.50	6.34	−0.02	2.83	4.02	0.26	7.89
	β_2	3.00	3.00	8.33	−0.08	3.72	6.94	0.44	36.01
	ρ	0.10	0.10	3.67	−0.13	1.64	1.35	0.09	2.69
	ν	1.00	1.00	3.12	−0.05	1.40	0.98	0.07	32.01

Table B.8: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.1$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
9	α_1	1.00	1.06	3.29	5.92	1.47	4.59	0.19	32.19
	α_2	1.00	0.99	5.74	−1.43	2.57	3.50	0.22	17.17
	β_1	−1.00	−0.91	8.47	8.73	3.79	14.79	0.76	10.78
	β_2	3.00	2.93	10.48	−7.47	4.69	16.56	0.95	27.92
	ρ	0.50	0.49	3.16	−0.78	1.41	1.06	0.08	15.57
	ν	0.50	0.50	1.27	−0.02	0.57	0.16	0.01	39.22
10	α_1	1.00	1.00	6.31	−0.16	2.82	3.99	0.26	15.81
	α_2	1.00	1.00	10.85	0.35	4.85	11.79	0.76	9.25
	β_1	−1.00	−1.00	8.96	−0.44	4.01	8.05	0.58	11.21
	β_2	3.00	3.01	10.73	0.53	4.80	11.55	0.80	28.00
	ρ	0.50	0.50	3.04	−0.07	1.36	0.92	0.06	16.42
	ν	1.00	1.00	3.17	−0.11	1.42	1.01	0.06	31.52
11	α_1	1.00	1.04	3.06	3.91	1.37	2.46	0.12	34.01
	α_2	1.00	0.98	5.48	−1.75	2.45	3.31	0.21	17.93
	β_1	−0.50	−0.46	5.82	4.03	2.60	5.01	0.30	7.90
	β_2	3.00	2.96	7.43	−3.65	3.32	6.85	0.46	39.90
	ρ	0.50	0.50	2.76	−0.24	1.23	0.77	0.05	18.05
	ν	0.50	0.50	1.27	0.06	0.57	0.16	0.01	39.41
12	α_1	1.00	1.00	5.89	−0.10	2.63	3.47	0.22	16.96
	α_2	1.00	1.00	10.07	−0.41	4.51	10.17	0.66	9.89
	β_1	−0.50	−0.50	5.84	−0.37	2.61	3.43	0.22	8.62
	β_2	3.00	3.00	7.37	0.44	3.30	5.45	0.36	40.76
	ρ	0.50	0.50	2.92	−0.07	1.31	0.85	0.06	17.08
	ν	1.00	1.00	3.14	−0.22	1.40	0.99	0.06	31.78

Table B.9: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.5$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
13	α_1	1.00	1.06	3.26	6.17	1.46	4.87	0.19	32.55
	α_2	3.00	2.94	5.51	−6.41	2.46	7.15	0.38	53.29
	β_1	−1.00	−0.95	8.36	5.23	3.74	9.72	0.56	11.34
	β_2	3.00	2.96	10.44	−4.34	4.67	12.78	0.77	28.32
	ρ	0.50	0.49	3.09	−0.53	1.38	0.99	0.07	15.99
	ν	0.50	0.50	1.27	−0.02	0.57	0.16	0.01	39.22
14	α_1	1.00	1.00	6.31	0.04	2.82	3.98	0.25	15.86
	α_2	3.00	3.00	11.34	−0.01	5.07	12.86	0.82	26.46
	β_1	−1.00	−1.00	8.69	−0.41	3.89	7.56	0.54	11.56
	β_2	3.00	3.01	10.49	0.50	4.69	11.03	0.76	28.64
	ρ	0.50	0.50	3.04	−0.10	1.36	0.92	0.06	16.43
	ν	1.00	1.00	3.17	−0.11	1.42	1.01	0.06	31.52
15	α_1	1.00	1.03	3.05	3.39	1.36	2.08	0.11	33.90
	α_2	3.00	2.96	5.54	−3.72	2.48	4.46	0.28	53.45
	β_1	−0.50	−0.48	5.86	2.27	2.62	3.95	0.26	8.15
	β_2	3.00	2.98	7.49	−1.99	3.35	6.00	0.42	39.80
	ρ	0.50	0.50	2.73	−0.19	1.22	0.75	0.05	18.24
	ν	0.50	0.50	1.27	0.06	0.57	0.16	0.01	39.41
16	α_1	1.00	1.00	5.93	−0.02	2.65	3.51	0.23	16.87
	α_2	3.00	2.99	10.63	−0.51	4.75	11.33	0.72	28.17
	β_1	−0.50	−0.50	5.73	−0.34	2.56	3.30	0.21	8.78
	β_2	3.00	3.00	7.28	0.41	3.26	5.32	0.35	41.24
	ρ	0.50	0.50	2.93	−0.10	1.31	0.86	0.06	17.04
	ν	1.00	1.00	3.14	−0.22	1.40	0.99	0.06	31.78

Table B.10: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.5$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
17	α_1	1.00	1.06	2.28	6.16	1.02	4.32	0.13	46.47
	α_2	1.00	1.00	3.74	0.41	1.67	1.41	0.10	26.87
	β_1	−1.00	−0.84	5.22	15.74	2.34	27.52	0.74	16.13
	β_2	3.00	2.87	5.98	−12.72	2.68	19.76	0.71	48.02
	ρ	0.90	0.90	0.71	0.27	0.32	0.06	0.00	126.44
	ν	0.50	0.50	1.24	0.01	0.56	0.15	0.01	40.18
18	α_1	1.00	0.94	5.85	−6.44	2.62	7.57	0.42	15.98
	α_2	1.00	1.04	6.08	3.65	2.72	5.03	0.28	17.04
	β_1	−1.00	−1.07	8.02	−6.79	3.59	11.05	0.73	13.31
	β_2	3.00	3.06	7.81	5.85	3.49	9.52	0.64	39.16
	ρ	0.90	0.90	0.78	−0.20	0.35	0.06	0.00	115.65
	ν	1.00	1.00	3.10	−0.15	1.38	0.96	0.06	32.25
19	α_1	1.00	0.99	2.47	−1.04	1.11	0.72	0.04	40.01
	α_2	1.00	1.02	3.52	2.44	1.57	1.83	0.11	29.09
	β_1	−0.50	−0.50	4.55	−0.04	2.03	2.07	0.14	11.01
	β_2	3.00	3.00	5.55	0.15	2.48	3.08	0.22	54.09
	ρ	0.90	0.90	0.69	−0.00	0.31	0.05	0.00	131.02
	ν	0.50	0.50	1.20	−0.15	0.54	0.15	0.01	41.41
20	α_1	1.00	0.87	6.20	−13.27	2.77	21.44	0.86	14.00
	α_2	1.00	1.04	6.72	4.30	3.01	6.37	0.38	15.52
	β_1	−0.50	−0.63	6.75	−12.53	3.02	20.27	0.85	9.27
	β_2	3.00	3.12	6.88	11.62	3.08	18.23	0.80	45.29
	ρ	0.90	0.90	0.76	−0.44	0.34	0.08	0.00	117.94
	ν	1.00	1.00	3.33	−0.16	1.49	1.11	0.10	30.01

Table B.11: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.9$ and $\boldsymbol{\alpha} = (1, 1)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

Scen.		$\boldsymbol{\theta}$	$\overline{\boldsymbol{\theta}}$	$s_{\overline{\theta}} \cdot 10^2$	$\widehat{b} \cdot 10^2$	$s_{\widehat{b}} \cdot 10^3$	$\widehat{mse} \cdot 10^3$	$s_{\widehat{mse}} \cdot 10^3$	$\overline{\theta}/s_{\overline{\theta}}$
21	α_1	1.00	1.07	2.52	6.77	1.13	5.22	0.16	42.43
	α_2	3.00	2.94	4.87	−6.39	2.18	6.46	0.31	60.32
	β_1	−1.00	−0.91	5.58	9.35	2.50	11.85	0.50	16.24
	β_2	3.00	2.92	6.30	−7.77	2.82	10.00	0.51	46.40
	ρ	0.90	0.90	0.74	0.07	0.33	0.05	0.00	122.35
	ν	0.50	0.50	1.24	0.01	0.56	0.15	0.01	40.18
22	α_1	1.00	1.01	5.13	1.38	2.30	2.83	0.18	19.75
	α_2	3.00	2.99	8.48	−0.56	3.79	7.22	0.44	35.32
	β_1	−1.00	−0.99	6.39	1.34	2.86	4.27	0.29	15.43
	β_2	3.00	2.99	7.02	−1.32	3.14	5.10	0.34	42.57
	ρ	0.90	0.90	0.79	−0.00	0.35	0.06	0.00	113.84
	ν	1.00	1.00	3.10	−0.15	1.38	0.96	0.06	32.25
23	α_1	1.00	1.03	2.58	3.29	1.15	1.74	0.09	40.10
	α_2	3.00	2.99	4.19	−1.40	1.88	1.95	0.12	71.22
	β_1	−0.50	−0.45	4.59	4.93	2.05	4.54	0.23	9.82
	β_2	3.00	2.96	5.59	−4.34	2.50	5.01	0.28	52.86
	ρ	0.90	0.90	0.69	0.06	0.31	0.05	0.00	130.87
	ν	0.50	0.50	1.20	−0.15	0.54	0.15	0.01	41.41
24	α_1	1.00	1.00	5.53	−0.09	2.47	3.06	0.20	18.05
	α_2	3.00	3.02	7.34	1.78	3.28	5.71	0.37	41.11
	β_1	−0.50	−0.49	5.30	0.67	2.37	2.85	0.17	9.32
	β_2	3.00	2.99	6.20	−0.90	2.77	3.93	0.24	48.23
	ρ	0.90	0.90	0.74	−0.01	0.33	0.05	0.00	121.60
	ν	1.00	1.00	3.33	−0.16	1.49	1.11	0.10	30.01

Table B.12: Simulation study results using the MBP with ρ_w-update based on 500 replications when $\rho = 0.9$ and $\boldsymbol{\alpha} = (1, 3)$: Parameter estimates ($\overline{\boldsymbol{\theta}}$), scaled estimated std. error of estimates ($s_{\overline{\theta}} \cdot 10^2$), scaled estimated bias ($\widehat{b} \cdot 10^2$), scaled estimated std. error of bias ($s_{\widehat{b}} \cdot 10^3$), scaled estimated MSE ($\widehat{mse} \cdot 10^3$), estimated std. error of MSE ($s_{\widehat{mse}} \cdot 10^3$), ratio between estimate and estimated std. error of estimate ($\overline{\theta}/s_{\overline{\theta}}$).

C R-Functions

C.1 R-Functions for the Zero-truncated Poisson GLM

Main Function: MLE Calculation of Regression Parameter

Input:

- **Yein** response vector of dimension `n`
- **Xein** design matrix of dimension `n` × `k`
- **Offset** offset in the linear predictor
- **bound** if `True` the the log-likelihood is maximized with the R-function `optim` using the gradient otherwise without using the gradient

Output:

- **Coefficients** calculated values of regression coefficients
- **Log.Likelihood** value of the log-likelihood
- **Iterations** number of iterations
- **Time** needed time to calculate the maxima
- **Message** message of the optimization function
- **Hessematrix** calculated hesse matrix of the log-likelihood
- **Range.mu** range of the fitted means
- **RSS**
- **Residuals** residuals of the regrssion
- **Null.Df** degree of freedom of the null model
- **Residual.Df** residual degree of freedom
- **Residual.Deviance** residual deviance
- **AIC** approximative Akaike Information Criterium
- **Fitted.Values** fitted means
- **Response** response vector
- **Design** design matrix

x

```
mle.tuncpois <- function(Yein, Xein, Offset = rep(0,length(Yein)),
    bound = TRUE){
```

```
# variable initialisation
X <<- Xein
Y <<- Yein
if(is.matrix(X)) {
n <<- dim(X)[1]
k <<- dim(X)[2]
}
else {
n <<- length(X)
k <<- 1
}

    hesse <- 0
    it <- 1

# initial value
beta <- summary(glm(Y ~ offset(Offset) + X, family =
                        poisson(link=log)))$coefficients[, 1]

# maximization of the log-likelihood function
start <- proc.time()[2]

if(bound) {
  grenze <- rep( - Inf, k)
opt <- optim(beta, log.likelihood.truncpois, method = "L-BFGS-B", gr
                            =gradient, lower = grenze, hessian = TRUE)
      beta <- opt$par
hesse <- opt$hessian
it <- opt$counts
log.likelihood <-  - opt$value
message <- opt$message
}
else {
   opt <- optim(beta, log.likelihood.truncpois, hessian = TRUE)
   beta <- opt$par
   hesse <- opt$hessian
   it <- opt$counts
   log.likelihood <-  - opt$value
   message <- opt$message
 }

zeit <- proc.time()[2] - start

  mu <- fit.truncpois(beta)

res <- double(n)
RSS <- 0
for(i in 1:n) {
```

```
res[i] <- Y[i] - mu[i]
RSS <- RSS + res[i]^2
}

  AIC <- -2 * log.likelihood + 2 * k

  deviance.truncpois <- log.likelihood.truncpois.deviance(beta)

ausgabe <<- list(Coefficients = beta, Log.Likelihood = log.likelihood,
    Iterations = it, Time = zeit,
Message = message, Hessematrix = hesse,  Range.mu = c(min(mu), max(mu)),
    RSS = RSS, Residuals = res,
Null.Df = n-1, Residual.Df = n-k, Residual.Deviance=deviance.truncpois, AIC
    = AIC, Fitted.Values = mu, Response = Y, Design = X)

    return(ausgabe)

}
```

Sub Function: Log-likelihood Calculation

```
log.likelihood.truncpois <- function(beta){

  if(k == 1) {
      lambda <- exp(Offset + X * beta)
    }
    else {
      lambda <- exp(Offset + X %*% beta)
    }

  l <- sum(-lambda + Y * log(lambda) - log(1-exp(-lanbda)))

  return(-l)

}
```

Sub Function: Approximated Deviance Calculation

```
log.likelihood.truncpois.deviance <- function(beta)
{

  a <- Y -1 +0.0001
  t.Y <- a * (1 -(1-exp(-a))/(1-exp(-a)-a*exp(-a))) + (1-exp(-a))^2/(1-exp(-a)
  -a*exp(-a)) * Y

  if(k == 1) {
      lambda <- exp(Offset + X * beta)
    }
    else {
```

```
      lambda <- exp(Offset + X %*% beta)
    }

  dev <- -2 * sum( Y*(log(lambda)-log(t.Y)) -lambda + t.Y -log(1-exp(-mu))
  +log(1-exp(-t.Y)) )

  return(dev)

}
```

Sub Function: Log-likelihood Gradient Calculation

```
gradient <- function(beta)
{
  grad <- double(k)

    if(k == 1) {
      lambda <- exp(Offset + X * beta)
    }
    else {
      lambda <- exp(Offset + X %*% beta)
    }

# Derivative for beta
  for (j in 1:k){

    grad[j] <- sum( X[,j] * (-lambda + Y - (lambda * exp(-lambda))
                    /(1-exp(-lambda))) )

  }

  return(-grad)
}
```

Sub Function: Mean Fits Calculation

```
"fit.truncpois" <- function(beta)
{

  lambbda <- double(n)
  if(k == 1) {
    labbda <- exp(Offset + X * beta)
  }
  else {
    lambda <- exp(Offset + X %*% beta)
  }

  mu <- lambda/(1-exp(-lambda))
```

```
  return(mu)

}
```

C.2 The MBP Algorithm for the Gamma-Poisson Regression Model

Main Function

Input:

- **alpha** intitial value for the regression parameter of the marginal Gamma GLM
- **beta** intitial value for the regression parameter of the marginal Poisson GLM
- **Y1** response vector of the marginal Gamma GLM
- **Y2** response vector of the marginal Poisson GLM. The values has to be > 0 as we use the conditional log-likelihood of the Gamma-Poisson regression model
- **X** design matrix for the Gamma regression
- **Z** design matrix for the Poisson regression
- **offset1** offset for the marginal Gamma GLM. If there is no offset, the parameter has been set to "NO" (default value).
- **offset2** offset for the marginal Poisson GLM. If there is no offset, the parameter has been set to "NO" (default value).
- **dispersion** dispersion parameter of marginal Gamma GLM
- **rho_w** prespecified ρ_w of the expanded conditional likelihood. Value has to be in the range (-1,1). For running the algorithm with the rho_w-update set rho_w = "Update" (default value)
- **fd_cl_theta1** R-function to evaluate the score function of the conditional likelihood for theta 1
- **fd_cld_gamma** R-function to evaluate the score function of the conditional likelihood for gamma
- **FI_eclm** R-function to evaluate the Fisher Information matrix of the marginal part of the expanded conditional likelihood
- **tol.cov.theta1** stop criterium for theta1
- **tol.cov.rho** stop criterium for rho
- **max.it** maximal number of iterations

Output list containing the following items

- **alpha** MLEs of the Gamma regression parameter
- **beta** MLEs of the Poisson regression parameter
- **rho** MLE of the correlation parameter of the Gaussian copula
- **dispersion** dispersion parameter of the marginal Gamma distribution (is identical with the input parameter dispersion)]
- **used.data** list of the following items

Y1 response vector of the marginal Gamma GLM

offset1 offset for the marginal Gamma GLM

X design matrix of the marginal Gamma GLM

Y2 response vector of the marginal Poisson GLM. The values has to be ¿0 as we use the conditional log-likelihood of the Gamma-Poisson regression model

offset2 offset for the marginal Poisson GLM

Z design matrix of the marginal Poisson GLM

num.iterations number of iteration of the algorithm

all.theta1 values of theta1 in each iteration

all.rho values of rho in each iteration

all.Fgamma values of Fisher z-transformation of rho in each iteration

all.theta1.dif change of theta1 in each iteration

all.rho.dif change of rho in each iteration

```
MBP <- function(alpha, beta, Y1, Y2, X, Z, offset1="NO", offset2="NO",
        dispersion, rho_w="UPDATE", fd_cl_theta1, fd_cld_gamma, FI_eclm,
        tol.cov.theta1 = 0.001, tol.cov.rho = 0.001, max.it = 300) {

  rho_wupdate <- 0
  if(rho_w=="UPDATE"){ rho_wupdate <- 1 }

  #initial values

  theta1 <- c(alpha, beta)

  p1 <- length(alpha)
  p2 <- length(beta)

  if(offset1[1] == "NO") offset1 <- rep(0,length(Y1))
  if(offset2[1] == "NO") offset2 <- rep(0,length(Y2))

  mu1 = as.vector(exp(offset1 + X %*% alpha))
  mu2 = as.vector(exp(offset2 + Z %*% beta))

  nu <- sqrt(dispersion)

  bisect_fn <- function(x) {

  fd_cld_gamma(Y1, mu1, Y2, mu2, x, nu)

    }

  Fgamma <- bisect(bisect_fn, -3, 3)     #inital Fgamma

  if(rho_wupdate==1){ rho_w <- (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1) }
```

```
##########################################################################

  i <- 0
  theta1.dif <- 1
  all.theta1.dif <- 1
  rho.dif <- 1
  all.rho.dif <- 1
  all.theta1 <- theta1
  all.Fgamma <- Fgamma
  all.rho <- (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1)

#########################################################################


  #Step 1, 2, 3, ...
  while((theta1.dif > tol.cov.theta1 | rho.dif > tol.cov.rho) & i < max.it) {

    i = i + 1
    print(i)

    theta1.old <- theta1
    Fgamma.old <- Fgamma

    theta1 <- theta1.old + as.vector(solve(FI_eclm(Y1, mu1, X, Y2, mu2, Z,
        Fgamma, nu, rho_w)) %*% fd_cl_theta1(Y1, mu1, X, Y2, mu2, Z,
        Fgamma, nu))

    bisect_fn <- function(x) {

      fd_cld_gamma(Y1, mu1, Y2, mu2, x, nu)

    }

    Fgamma <- bisect(bisect_fn, -3, 3)

    rho.old <- (exp(2*Fgamma.old) - 1)/(exp(2*Fgamma.old) + 1)
    rho <- (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1)

    if(rho_wupdate==1){ rho_w <- rho }

    mu1 = as.vector(exp(offset1 + X %*% theta1[1:p1]))
    mu2 = as.vector(exp(offset2 + Z %*% theta1[(p1+1):(p1+p2)]))

    theta1.dif <- max(abs(theta1 - theta1.old))
    rho.dif <- abs(rho - rho.old)

    all.theta1 <- rbind(all.theta1, theta1)
    all.theta1.dif <- c(all.theta1.dif, theta1.dif)
```

```
    all.rho <- c(all.rho, rho)
    all.rho.dif <- c(all.rho.dif, rho.dif)
    all.Fgamma <- c(all.Fgamma, Fgamma)

  }

  list(alpha = theta1[1:p1], beta = theta1[(p1+1):(p1+p2)], rho = rho, Fgamma=
  Fgamma, dispersion = dispersion, used.data=list(Y1=Y1, offset1=offset1, X=X,
  Y2=Y2, offset2=offset2, Z=Z), num.iterations = i, all.theta1 = all.theta1,
  all.rho = all.rho, all.Fgamma=all.Fgamma, all.theta1.dif = all.theta1.dif,
  all.rho.dif = all.rho.dif )

}
```

Sub Function: Bisection Method for one dimensional root search

Input:

- **fn** number of observations
- **a** left bound of the start interval
- **b** right bound of the start interval
- **tol** stop criterium of the algorithm
- **max.it** maximal number of iterations

Output root of the function fn

```
 bisect <- function(fn, a=-3, b=3, tol=1e-6 , max.it=100){

  i<-0
  cur<-(a+b)/2

  if ((fn(a)*fn(b)>0)|(a>b)) {
    return('Improper start values') }
  else {
    while (abs(fn(cur))>tol & i < max.it) {
      i<-i+1
      if(i==max.it) print("max.it reached")
      if (fn(a)*fn(cur)>0) {
        a<-cur
        cur<-(a+b)/2
      }
      else {
        b<-cur
        cur<-(a+b)/2
      }
      #res<-rbind(res,c(i,cur,fn(cur)))
      res <- cur
```

```
    }
  }

  return(res)
}
```

Sub Function: fd_cl_theta1

Function to evaluate the score function $\frac{\partial l^c(\boldsymbol{\theta}_1,\gamma)}{\partial \boldsymbol{\theta}_1}$ for fix γ.

Input:

- **y1** response vector of the marginal Gamma GLM
- **mu1** fitted mean of the marginal Gamma distribution
- **X** design matrix of the Gamma GLM
- **y2** response vector of the marginal Poisson GLM
- **mu2** fitted mean of the marginal Poisson distribution
- **Z** design matrix of the Poisson GLM
- **Fgamma** Fisher z-transformation of ρ
- **nu** constant coefficient of variation ν of the Gamma distribution

Output value of $\frac{\partial l^c(\boldsymbol{\theta}_1,\gamma)}{\partial \boldsymbol{\theta}_1}$

```
fd_cl_theta1 <- function(y1, mu1, X, y2, mu2, Z, Fgamma, nu) {

  n <- length(y1)

  rho = (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1)

u1 = pgamma(y1, shape=1/nu^2, rate = 1/(mu1*nu^2))
u2 = ppois(y2, mu2)
u2m = ppois(y2-1, mu2)

D_rho_dif <- (D_rho(u1,u2,rho)- D_rho(u1,u2m,rho))
  for (i in 1:length(D_rho_dif)){
    if (D_rho_dif[i] == 0) D_rho_dif[i] = 1e-16
  }

  term1 <- (y1/mu1-1)/nu^2
  term2 <- ((d_rho(u1,u2,rho) - d_rho(u1,u2m,rho))/D_rho_dif)* ((-rho)/
    (sqrt(1-rho^2)) * (pgamma(y1, shape=(1/nu^2)+1, rate = 1/(mu1*nu^2))
    - pgamma(y1, shape=1/nu^2, rate = 1/(mu1*nu^2)))/(dnorm(qnorm(u1))))
  term3 <-  -mu2/(exp(mu2)-1)
  trem4 <- (mu2/D_rho_dif) * (1/sqrt(1-rho^2) * (d_rho(u1,u2m,rho)
            * dpois(y2-1, mu2)/(dnorm(qnorm(u2m))) - d_rho(u1,u2,rho)
```

```
          * dpois(y2, mu2)/(dnorm(qnorm(u2)))) )

  fd_l1 <- 0
  fd_l2 <- 0

  for(i in 1:n){

    fd_l1 <- fd_l1 + term1[i] * X[i,] + term2[i] * X[i,]
    fd_l2 <- fd_l2 + term3[i] * Z[i,] + trem4[i] * Z[i,]

  }

  return(c(fd_l1,fd_l2))
}
```

Sub Function: fd_cld_gamma

Function to evaluate the score function $\frac{\partial l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial \gamma}$.

Input:

- **y1** response vector of the marginal Gamma GLM
- **mu1** fitted mean of the marginal Gamma distribution
- **y2** response vector of the marginal Poisson GLM
- **mu2** fitted mean of the marginal Poisson distribution
- **Fgamma** Fisher z-transformation of ρ
- **nu** constant coefficient of variation ν of the Gamma distribution

Output value of $\frac{\partial l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial \gamma}$

```
fd_cld_gamma  <- function(y1, mu1, y2, mu2, Fgamma, nu) {

  rho = (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1)

u1 = pgamma(y1, shape=1/nu^2, rate = 1/(mu1*nu^2))
u2 = ppois(y2, mu2)
u2m = ppois(y2-1, mu2)

D_rho_dif <- (D_rho(u1,u2,rho)- D_rho(u1,u2m,rho))
  for (i in 1:length(D_rho_dif)){
    if (D_rho_dif[i] == 0) D_rho_dif[i] = 1e-16
  }

   1/sqrt(1-rho^2) * (1/D_rho_dif) %*% (d_rho(u1,u2,rho) * (rho
    * qnorm(u2) - qnorm(u1)) - d_rho(u1,u2m,rho) * (rho
    * qnorm(u2m) - qnorm(u1)))
}
```

Sub Function: FI_eclm

Function to calculate the Fisher information matrix $\mathcal{I}^*_m$ given in equation 4.34.

Input:

y1 response vector of the marginal Gamma GLM

mu1 fitted mean of the marginal Gamma distribution

X design matrix of the Gamma GLM

y2 response vector of the marginal Poisson GLM

mu2 fitted mean of the marginal Poisson distribution

Z design matrix of the Poisson GLM

Fgamma Fisher z-transformation of ρ

nu constant coefficient of variation ν of the Gamma distribution

rho_w prespecified rho_w of the expanded conditional likelihood. Value has to be in the range (-1,1). For running the algorithm with the rho_w-update set rho_w = "Update" (default value)

Output value of $\frac{\partial l_d^c(\boldsymbol{\theta}_1,\gamma)}{\partial\gamma}$

```
FI_eclm  <- function(y1, mu1, X, y2, mu2, Z, Fgamma, nu, rho_w) {

  n <- length(y1)
  p1 <- dim(X)[2]
  p2 <- dim(Z)[2]

  rho = (exp(2*Fgamma) - 1)/(exp(2*Fgamma) + 1)

Gamma_inv = solve(matrix(c(1, rho_w, rho_w, 1), nrow=2))

  fi_lf1 <- 0
  fi_lf2 <- 0

  fi_cor1 <- 0
  fi_cor2 <- 0

  term1 <- (mu2*(exp(mu2)-1-mu2*exp(mu2))/((exp(mu2)-1)^2))
  term2 <- mu1^2
  term3 <- mu2^2
  term4 <- mu1*mu2

  for(i in 1:n){

    fi_lf1 <- fi_lf1 + 1/nu^2 * X[i,] %*% t(X[i,])
   fi_lf2 <- fi_lf2 +  term1[i] * Z[i,] %*% t(Z[i,]) + mu2[i] * Z[i,] %*% t(Z[i,])
```

```
    fi_cor1 <- fi_cor1 * cbind(Gamma_inv[1,1] * term2[i] * X[i,] %*%
        t(X[i,]), Gamma_inv[1,2] * term4[i] * X[i,] %*% t(Z[i,]))
    fi_cor2 <- fi_cor2 * cbind(Gamma_inv[2,1] * term4[i] * Z[i,] %*%
        t(X[i,]), Gamma_inv[2,2] * term3[i] * Z[i,] %*% t(Z[i,]))

  }

  fi_lf <- matrix(0, p1+p2, p1+p2)
  fi_lf[1:p1,1:p1] <- fi_lf1
  fi_lf[(p1+1):(p1+p2),(p1+1):(p1+p2)] <- fi_lf2

  fi_cor <- matrix(0, p1+p2, p1+p2)
  fi_cor[1:p1,1:(p1+p2)] <- fi_cor1
  fi_cor[(p1+1):(p1+p2),1:(p1+p2)] <- fi_cor2


  return(fi_lf + fi_cor)

 }
```

C.3 R-Function to Estimate the Standard Error

Function to estimate the standard error of the MLEs resulting of the MBP algorithm.

Input:

- **alpha** intitial value for the regression parameter of the marginal Gamma GLM
- **beta** intitial value for the regression parameter of the marginal Poisson GLM
- **Y1** response vector of the marginal Gamma GLM
- **Y2** response vector of the marginal Poisson GLM. The values has to be > 0 as we use the conditional log-likelihood of the Gamma-Poisson regression model
- **X** design matrix for the Gamma regression
- **Z** design matrix for the Poisson regression
- **offset1** offset for the marginal Gamma GLM. If there is no offset, the parameter has been set to "NO" (default value).
- **offset2** offset for the marginal Poisson GLM. If there is no offset, the parameter has been set to "NO" (default value).
- **nu** constant coefficient of variation ν of the Gamma distribution

Output standard errors for the MLEs of the Gamma-Poisson model parameter

```
ste <- function(alpha, beta, rho, y1, y2, X, Z, offset1, offset2, nu){

  mu1 = as.vector(exp(offset1 + X %*% alpha))
  mu2 = as.vector(exp(offset2 + Z %*% beta))
```

```
  p1 <- length(alpha)
  p2 <- length(beta)

  n <- length(y1)

  FI <- matrix(0,p1+p2+1, p1+p2+1)     #average Fisher Information


  FI_lM <- matrix(0, p1+p2+1, p1+p2+1)  #average Fisher Information of L_M

  frac <- mu2*(exp(mu2)-1-mu2*exp(mu2))/((exp(mu2)-1)^2)

  u1 = pgamma(y1, shape=1/nu^2, rate = 1/(mu1*nu^2))
u2 = ppois(y2, mu2)
u2m = ppois(y2-1, mu2)

D_rho_dif <- (D_rho(u1,u2,rho)- D_rho(u1,u2m,rho))
  for (i in 1:length(D_rho_dif)){
    if (D_rho_dif[i] == 0) D_rho_dif[i] = 1e-16
  }

#est. Fisher Information for each obsrvation

  fd_lD_alpha <- X * ((d_rho(u1,u2,rho) - d_rho(u1,u2m,rho))/D_rho_dif)
    * ((-rho)/(sqrt(1-rho^2))) * ((pgamma(y1, shape=(1/nu^2)+1, rate =
    1/(mu1*nu^2)) - pgamma(y1, shape=1/nu^2, rate = 1/(mu1*nu^2)))
    /(dnorm(qnorm(u1))))
  fd_lD_beta <- Z * (mu2/D_rho_dif) * (1/sqrt(1-rho^2) *(d_rho(u1,u2m,rho)
  * dpois(y2-1, mu2)/(dnorm(qnorm(u2m))) - d_rho(u1,u2,rho)
  * dpois(y2, mu2)/(dnorm(qnorm(u2)))) )
  fd_lD_rho <- 1/((1-rho^2)^(2/3)) * (1/D_rho_dif) * (d_rho(u1,u2,rho)
  * (rho * qnorm(u2) - qnorm(u1)) - d_rho(u1,u2m,rho)
  * (rho * qnorm(u2m) - qnorm(u1)))

  fd_lD <- cbind(fd_lD_alpha, fd_lD_beta, fd_lD_rho)

  fi_lM_alpha <- 0
  fi_lM_beta <- 0
  sum.FI_lD <- matrix(0,p1+p2+1, p1+p2+1)

  for(i in 1:n){

    fi_lM_alpha <- fi_lM_alpha + 1/nu^2 * X[i,] %*% t(X[i,])
    fi_lM_beta <- fi_lM_beta + frac[i] * Z[i,] %*% t(Z[i,]) + mu2[i]
    * Z[i,] %*% t(Z[i,])

    sum.FI_lD = sum.FI_lD + fd_lD[i,] %*% t(fd_lD[i,])
```

```
  }

#average Fisher Information of L_M

  FI_lM[1:p1,1:p1] <- fi_lM_alpha/n
  FI_lM[(p1+1):(p1+p2),(p1+1):(p1+p2)] <- fi_lM_beta/n

#average Fisher Information of L_D

  FI_lD <- sum.FI_lD/n

#average Fisher Information of L

  FI <- FI_lM + FI_lD

  Covar <- solve(FI)/n         #asymptotic covariance

  ste <- sqrt(diag(Covar))     #standard error

  return(list(ste.theta=ste, ste.alpha = ste[1:p1], ste.beta=ste[(p1+1)
    :(p1+p2)], ste.rho = ste[p1+p2+1]))
}
```

C.4 Gamma-Poisson Sampler

Function to generate a data set containing observation of Gamma-Poison distribution with density (3.8)

Main Function

Input:

N number of observations

alpha regression coefficients for the gamma variable

beta regression coefficients for the poisson variable

rho correlation parameter of the Gaussian copula

dispersion dispersion parameter of the gamma distribution

X design matrix for the gamma regression

Z design matrix for the gamma regression

gam.offset offset for the gamma regression

pois.offset offset for the poisson regression

Zero.trunc if TRUE sample from gamma-poisson distribution with condition poisson observation Y2>0

eps define when the tail of the Gamma-Poisson distribution is truncated

Output data frame containing observation pairs ,Y1, Y2, of the Gamma-Poisson distribution, the observed covariates X of the marginal Gamma GLMs, the observed covariates Z of the marginal Poisson GLM, the exposure = $exp($gam.offset$)$ of the marginal Gamma GLM and exposure = $exp($pois.offset$)$ of the marginal Poisson GLM

```
gam.pois.sampler <- function(N, beta1, beta2, rho, dispersion, X, Z, gam.offset
    =rep(0,N),  pois.offset=rep(0,N), Zero.trunc=FALSE, eps=1e-10){

  mu1 <- exp(gam.offset + X %*% beta1)
  mu2 <- exp(pois.offset + Z %*% beta2)

  nu <- sqrt(dispersion)

Y1 <- rgamma(N, shape=1/nu^2, rate=1/(mu1*nu^2))
u1 <- rep(0,N)
  Y2 <- rep(0,N)
u2 <- rep(0,N)

if(Zero.trunc) {

    for(i in 1:N){
   temp <- rctruncpois(size=1, y1=Y1[i], nu=nu, mu1=mu1[i], mu2=mu2[i], rho)
   Y2[i] <- temp$y2
    }

}
else {

   for(i in 1:N){
   temp <- rcpois(size=1, y1=Y1[i], nu=nu, mu1=mu1[i], mu2=mu2[i], rho)
   Y2[i] <- temp$y2
    }

   }

Data <- data.frame(Y1 = Y1  , Y2 = Y2  , X = X, Z = Z, exposure1
            = exp(gam.offset), exposure2= exp(pois.offset))

return(Data)
}
```

Gamma-Poisson observation Generator

Function to generate poisson variables correlated to a gamma variable

```
rcpois <- function(size, y1, mu1, nu, mu2, rho, eps=1e-10){

  shape = 1/nu^2
rate = 1/(mu1*nu^2)
```

```
u1 <- pgamma(y1, shape, rate)
u20 <- ppois(0, mu2)

cprob = D_rho(u1, u20, rho)

  for(i in 1:1000){

    u2 <- ppois(i, mu2)
u2m <- ppois(i-1, mu2)
    p <- D_rho(u1,u2,rho) - D_rho(u1,u2m,rho)

    if(p<eps) break
    else cprob <- c(cprob, p)

}

y2 <- sample(c(0:(length(cprob)-1)), size, replace=TRUE, prob=cprob)

return(list(y1=y1, y2 = y2, u2 = cprob[y2+1], u1 = u1))
}
```

Gamma-Zero-truncated Poisson observation Generator

Function to generate poisson variables > 0 correlated to a gamma variable

```
rctruncpois <- function(size, y1, mu1, nu, mu2, rho, eps=1e-10){

  shape = 1/nu^2
rate = 1/(mu1*nu^2)

u1 <- pgamma(y1, shape, rate)
u20 <- ppois(0, mu2)
  u2 <- ppois(1, mu2)

  #probability for y2=1|y2>0
cprob = (D_rho(u1,u2,rho) - D_rho(u1,u20,rho)) / (1-D_rho(u1,u20,rho))

  #probability for y2=2,3,...|y2>0
  for(i in 2:1000){

    u2 <- ppois(i, mu2)
u2m <- ppois(i-1, mu2)
    p <- (D_rho(u1,u2,rho) - D_rho(u1,u2m,rho)) / (1-D_rho(u1,u20,rho))

    if(p<eps) break
    else cprob <- c(cprob, p)

}
```

```
y2 <- sample(c(1:(length(cprob))), size, replace=TRUE, prob=cprob)

return(list(y1=y1, y2 = y2, u2 = cprob[y2], u1 = u1))
}
```

C.5 R-Functions for the Explorative Data Analysis

The R-funcions gamma.main, poisson.main and gamma.inter are based on the R-functions used in the lecture 'Moderne Regressionsverfahren' (WS 2006/2007) of Prof. Dr. Claudia Czado.

Gamma EDA

The function plots the empirical log-means of a Gamma-distributed response for a covariate.

```
gamma.main <- function(y = Cost, x = Merit, xlabel = "Merit Class", label = " ",
                         ps = F, conf = T, link ="log")
{
    if(ps == T) {
        ps.options(colors = ps.colors.rgb[c("black", "cyan", "magenta","green",
            "MediumBlue", "red"),  ], horizontal = F)
        postscript(file = "gamma.main.ps")
    }
    ns <- tapply(rep(1, length(y)), x, sum)
    sy <- y
    temp <- tapply(sy, x, mean)
    if(link == "log") {
        upper <- log(temp + ((qnorm(0.95) * temp)/sqrt(ns)))
        lower <- log(temp - ((qnorm(0.95) * temp)/sqrt(ns)))
        r <- round(range(log(temp))[2] -  range(log(temp))[1],2)
        if(conf == T) {
            ry <- range(c(upper, lower))
            plot(log(temp), ylim = ry, axes = F, pch = 3, xlab = xlabel,
            ylab = "empirical log mean", sub = paste("(Range", r, ")"),
            type = "b", lty = 2, col = 3, lwd = 3)
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
            lines(1:length(temp), upper, lty = 2, col = 4, lwd = 2)
            lines(1:length(temp), lower, lty = 2, col = 4, lwd = 2)
            par(new = T)
            plot(log(temp), ylim = ry, axes = F, pch = 3, xlab = "",
            ylab = "empirical log mean", sub = paste("(Range", r, ")"),
            type = "n", lty = 2, col = 1, lwd = 3, main = label)
        }
        else {
            plot(log(temp), axes = F, pch = 3, xlab = "",
            ylab =  "empirical log mean", type = "b", lty = 1,
```

```
            col = 1, lwd = 3, sub = xlabel)
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
        }
    }
    if(link == "inverse") {
        upper <- (temp + ((qnorm(0.95) * temp)/sqrt(ns)))^(-1)
        lower <- (temp - ((qnorm(0.95) * temp)/sqrt(ns)))^(-1)
        if(conf == T) {
            ry <- range(c(upper, lower))
            plot(1/temp, ylim = ry, axes = F, pch = 3, xlab = "",
                ylab = "empirical inverse mean", sub = xlabel,
                type = "b", lty = 2, col = 3, lwd = 3)
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
            lines(1:length(temp), upper, lty = 2, col = 4, lwd = 2)
            lines(1:length(temp), lower, lty = 2, col = 4, lwd = 2)
            par(new = T)
            plot(log(temp), ylim = ry, axes = F, pch = 3, xlab = "",
                ylab = "empirical inverse mean", sub = xlabel,
                type = "n", lty = 2, col = 1, lwd = 3, main =
                label)
        }
        else {
            plot(1/temp, axes = F, pch = 3, xlab = "", ylab =
                "empirical inverse mean", type = "b", lty = 1,
                col = 1, lwd = 3, sub = xlabel)
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
        }
    }
}
```

The function plots a graphic to identify possible interactions between covariates for a Gamma-distributed response.

```
gamma.inter<-function(y = Claims, x1 = Merit, x2 = Class, label1 = "Merit",
    label2 = "Class",ps = F, conf = F, link = "log"){
    if(ps == T) {
        ps.options(colors = ps.colors.rgb[c("black", "cyan", "magenta",  "green",
            "MediumBlue", "red"),  ], horizontal = F)
            postscript(file = "gamma.inter.ps")
        }
        ns <- tapply(rep(1, length(y)), list(x1, x2), sum)
        sy <- y
        temp <- tapply(sy, list(x1, x2), mean)
        if(link == "log") {
            if(conf == F) {
                ry <- range(log(temp),na.rm = TRUE)
```

```
            plot(log(temp[1,  ]), ylim = ry, axes = F, pch = 3,
            xlab = "", ylab = "empirical log mean", sub =
            label2, type = "b", lty = 2, col = 1, lwd = 3) #,type = "n")
            title(paste(label1, "and", label2))
            axis(2)
            axis(1, at = 1:length(temp[1,  ]), labels = levels(x2))
            for(i in 1:length(levels(x1))) {
                par(new = T)
                plot(log(temp[i,  ]), axes = F, xlab = "",
                ylab = "", ylim = ry, lty = i + 1,
                lwd = 3, col = i + 1, type = "b")
                text(1 + 0.25, log(temp[i, 1]), paste(levels(
                x1)[i]), col = i + 1)
            }
        }
    }
    if(link == "inverse") {
        if(conf == F) {
            ry <- range(1/temp,na.rm = TRUE)
            plot(1/temp[1,  ], ylim = ry, axes = F, pch = 3, xlab
            = "", ylab = "empirical inverse mean", sub =
            label2, type = "b", lty = 2, col = 1, lwd = 3) # ,type = "n")
            title(paste(label1, "and", label2))
            axis(2)
            axis(1, at = 1:length(temp[1,  ]), labels = levels(x2))
            for(i in 1:length(levels(x1))) {
                par(new = T)
                plot(1/temp[i,  ], axes = F, xlab = "", ylab = "",
                ylim = ry, lty = i + 1, lwd = 3, col = i + 1, type = "b")
                text(1 + 0.25, log(temp[i, 1]), paste(levels(x1)[i]),
                    col = i + 1)
            }
        }
    }
  }
}
```

Poisson or Zero-truncated Poisson EDA

The function plot the empirical log-means of a Gamma-distributed response for a covariate.

```
poisson.main<-function(y = Claims, n = Insured, x = Merit,
            xlabel = "Merit Class", label = " ", ps = F, conf = T)
{
    if(ps == T) {
        ps.options(colors = ps.colors.rgb[c("black", "cyan", "magenta", "green",
            "MediumBlue", "red"), ], horizontal = F)
        postscript(file = "poisson.main.ps")
    }
```

```
    ns <- tapply(rep(1, length(y)), x, sum)
    sy <- y/n
    temp <- tapply(sy, x, mean)
    upper <- log(temp + ((qnorm(0.95) * temp)/sqrt(ns)))
    lower <- log(temp - ((qnorm(0.95) * temp)/sqrt(ns)))
    r <- round(range(log(temp))[2] -  range(log(temp))[1],2)
    if(conf == T) {
            ry <- range(c(upper, lower))
            plot(log(temp), ylim = ry, axes = F, pch = 3, xlab = xlabel,
                ylab = "empirical log mean", sub = paste("(Range", r, ")"),
                type = "b", lty = 2, col = 3, lwd = 3)
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
            lines(1:length(temp), upper, lty = 2, col = 4, lwd = 2)
            lines(1:length(temp), lower, lty = 2, col = 4, lwd = 2)
            par(new = T)
            plot(log(temp), ylim = ry, axes = F, pch = 3, xlab = "",
                ylab = "empirical log mean", sub = paste("(Range", r, ")"),
                type = "n", lty = 2, col = 1, lwd = 3, main = label)
    }
    else {
            plot(log(temp), axes = F, pch = 3, xlab = "", ylab =
                    "empirical log mean", type = "b", lty = 1, col = 1, lwd
                     = 3, sub = " ")
            axis(2)
            axis(1, at = 1:length(temp), labels = levels(x))
    }
}
```