ARTICLE IN PRESS INSUMA:2825

Insurance: Mathematics and Economics ••• (••••) •••–•••

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

www.elsevier.com/locate/ime

# The multivariate mixed Negative Binomial regression model with an application to insurance a posteriori ratemaking

George Tzougas [a,*], Alice Pignatelli di Cerchiara [b]

[a] Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK
[b] Department of Statistics, London School of Economics and Political Science, UK

ARTICLE INFO

*Article history:*
Received June 2020
Received in revised form August 2021
Accepted 8 October 2021
Available online xxxx

*JEL classification:*
C13
C63
G22

*Keywords:*
Multivariate mixed Negative Binomial regression model
EM algorithm
A posteriori ratemaking
Nonlife insurance
Bodily injury and property damage MTPL claim counts

ABSTRACT

This paper is concerned with introducing a family of multivariate mixed Negative Binomial regression models in the context of a posteriori ratemaking. The multivariate mixed Negative Binomial regression model can be considered as a candidate model for capturing overdispersion and positive dependencies in multi-dimensional claim count data settings, which all recent studies suggest are the norm when the ratemaking consists of pricing different types of claim counts arising from the same policy. For expository purposes, we consider the bivariate Negative Binomial-Gamma and Negative Binomial-Inverse Gaussian regression models. An Expectation-Maximization type algorithm is developed for maximum likelihood estimation of the parameters of the models for which the definition of a joint probability mass function in closed form is not feasible when the marginal means are modelled in terms of covariates. In order to illustrate the versatility of the proposed estimation procedure a numerical illustration is performed on motor insurance data on the number of claims from third party liability bodily injury and property damage. Finally, the a posteriori, or Bonus-Malus, premium rates resulting from the bivariate Negative Binomial-Gamma and Negative Binomial-Inverse Gaussian regression model are compared to those determined by the bivariate Negative Binomial and Poisson-Inverse Gaussian regression models.



## 1. Introduction

Over the last few decades, univariate mixed Poisson regression models, with the Negative Binomial (NB) and Poisson-Inverse Gaussian (PIG) models being the most traditional choices, have been established by various previous studies as the appropriate statistical formalism in the a priori ratemaking process for Motor Third Party Liability (MTPL) insurance due to their efficiency for quantifying the relation between the overdispersed claim counts and the characteristics of the policyholders and their cars. Furthermore, such models can be used for deriving a posteriori ratemaking mechanisms, or Bonus-Malus Systems (BMSs), which can take into account both the a priori and a posteriori criteria, i.e. all the factors that could not be identified, measured and introduced in the a priori tariff. An excellent account of BMSs can be found in Lemaire (1995). Further references for BMSs include, among many others, Tremblay (1992), Picech (1994), Pinquet (1997), Pinquet (1998), Brouhns et al. (2003), Mert and Saykan (2005), Denuit et al. (2007), Boucher and Denuit (2008), Gómez-Déniz et al. (2008), Gómez-Déniz et al. (2014), Ni et al. (2014a), Ni et al. (2014b), Tzougas and Frangos (2014), Santi et al. (2016), Gómez-Déniz and Calderín-Ojeda (2018), Karlis et al. (2018), Tzougas et al. (2018), Tzougas et al. (2019) and Tzougas et al. (2020). However, by adopting the univariate mixed Poisson count regression modelling approach, the actuary can only specify a separate model for different claim types. Nevertheless, it is not uncommon for an insurer to find the need in non-life insurance practice to model the positive association between claim counts of two (and/or multiple) types. In fact, various studies have reported evidence of a positive correlation between different types of claims, see, for instance, Bermúdez (2009), Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014), Abdallah et al. (2016), Bermúdez and Karlis (2017), Bermúdez et al. (2018) and Gómez-Déniz and Calderín-Ojeda (2021). As far as MTPL insurance is concerned, which refers to a person's legal liability for the bodily injury and property damage sustained by another as the result of an accident, modelling the two types of claims, which are conceivably positively correlated with each other, and their

* Corresponding author.
*E-mail address:* George.Tzougas@hw.ac.uk (G. Tzougas).

https://doi.org/10.1016/j.insmatheco.2021.10.001


Please cite this article as: G. Tzougas and A. Pignatelli di Cerchiara, The multivariate mixed Negative Binomial regression model with an application to insurance a posteriori ratemaking, Insurance: Mathematics and Economics, https://doi.org/10.1016/j.insmatheco.2021.10.001

associated claim counts, is required for making the Bonus-Malus price discrimination even more fair and reasonable when the a posteriori correction is going to be calculated. Nevertheless, the Bayesian approach for calculating Bonus-Malus premiums in the bivariate setting has only been addressed very recently by Bermúdez and Karlis (2017). The contribution of the latter article can be regarded as a significant improvement over prior ratemaking literature, which only focused on bivariate experience rating models that were derived based on the credibility approach.

In the present paper, the multivariate mixed Negative Binomial (MVMNB) regression model is considered for joint modelling of claim counts from multiple types of claims in terms of covariates. For illustrative purposes, two members of this family, namely the bivariate Negative Binomial-Gamma (BNBGA) and bivariate Negative Binomial-Inverse Gaussian (BNBIG) regression models will be employed for examining the relation between the frequency of the positively correlated claims from MTPL bodily injury and property damage and the characteristics of the policyholders and their cars. Furthermore, motivated by the paper of Bermúdez and Karlis (2017), the models will be employed for calculating Bonus-Malus premium rates in a way which integrates a priori and a posteriori information on an individual basis. In what follows we provide a thorough discussion about how our contribution extends both the statistical and actuarial literature concerning multivariate claim count regression models, putting special emphasis on the probabilistic predictive modelling, computational ML estimation and practical insurance pricing perspectives in the context of MTPL insurance.

Firstly, even if, as previously mentioned, a plenitude of books and scholarly articles consistent with the standard probabilistic predictive modelling and MTPL ratemaking practice have been devoted to the use of univariate mixed Poisson regression models, since there are systematic effects in the data, there is no guarantee that overdispersion and variation in claim propensity have precisely the distributional forms implied by mixed Poisson models. Moreover, regarding the multivariate setting, due to the complexity of claim count data, modelling the relationship between different claim types and a set of explanatory variables is challenging. Therefore, unless the assumption that the count data are distributed according to a particular member of the mixed Poisson family is valid, then an inappropriate imposition of the mixed Poisson model may lead to huge financial impacts for the insurance company. As far as MTPL insurance is concerned, due to its economic importance,[1] very accurate predictions are required by the actuary for pricing risks. Furthermore, it should be noted that alternative mixed Poisson models usually lead to Bonus-Malus premium rates which are not substantially different for policyholders with some claim experience and hence there is in principle no reason why attention should be confined to this family of models. Thus, given the increasing interdisciplinary demand for data driven predictive models and maximum likelihood (ML) estimation methods, a very important aspect of the actuary's job is to be able to construct viable alternatives to the traditional mixed Poisson models that can capture the stylized characteristics of the data, since very accurate predictions are required for pricing, reserving, estimating future company liabilities and understanding the implications of these claims for the solvency of the company. Mixed Negative Binomial models have thick tails and can be considered as candidate models for analyzing highly overdispersed count data in numerous univariate and multivariate domains. Nevertheless, even if the literature on mixed Poisson models is abundant, only very few mixed Negative Binomial models have been studied in depth because their log-likelihoods are complicated and hence their maximization needs a special effort. In particular, the Negative Binomial-Pareto distribution (see Shengwang et al. (1999) and Gómez-Déniz and Vázquez-Polo (2003)), the Negative Binomial-Beta regression model (see Boucher et al. (2008)), the Negative Binomial-Gamma distribution (see Gençtürk and Yiğiter (2016)), the Negative Binomial-Lindley distribution (see Zamani and Ismail (2010) and Gómez-Déniz and Calderín-Ojeda (2017)), the Negative Binomial-Inverse Gaussian distribution/ regression model (see Gómez-Déniz et al. (2008) and Tzougas et al. (2019), who considered the cases without and with covariate information respectively) and the Negative Binomial-Reciprocal Inverse Gaussian distribution (see Ahmad et al. (2019)) have been considered in the univariate setting. Moreover, the literature on the multivariate extensions of mixed Negative Binomial models is even smaller since computational complexity increases even further when considering jointly two or more count variables. In fact, the only noTable exceptions so far are the articles by Gómez-Déniz et al. (2008) and Calderín-Ojeda and Gómez-Déniz (2019) who introduced the multivariate versions of the Negative Binomial-Inverse Gaussian and the Negative Binomial-Lindley distributions, considered ML estimation methods and gave a very detailed description of statistical methods connected to both models. However, regarding the case with covariates, this is the first time that the MVMNB regression model is used in a statistical or actuarial context because, due to algebraic intractability, direct maximization of its log-likelihood is difficult and has not been addressed in the literature so far. The model is constructed based on a mixing between multiple marginal Negative Binomial distributions and a unit mean continuous prior, or mixing, distribution belonging to a general distribution family including, but not limited to, members of the natural exponential family. At this point, we would like to call attention to the fact that the construction of multivariate count regression models that can efficiently model overdispersed two-dimensional positively correlated count data has so far only focused on multivariate Poisson (MVP) models and multivariate mixed Poisson (MVMP) models which can accommodate overdispersion. The literature on MVP and MVMP models includes, for instance, the works of Stein and Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Munkin and Trivedi (1999), Gurmu and Elder (2000), Park and Lord (2007), Ma et al. (2008), El-Basyouny and Sayed (2009), Aguero-Valverde and Jovanis (2009) and Zhan et al. (2015), Silva et al. (2019) and Chiquet et al. (2020) among many others. The MVMNB model can be regarded as a prominent candidate for modelling multivariate positively correlated count data when marginal overdispersion is observed. This situation is quite common in the field of MTPL insurance since bodily injury and property damage claim counts often exhibit a variance that noticeably exceeds their mean.

Secondly, from a ML estimation point of view, the main contribution of the present study is that we develop an EM type algorithm that reduces the computational burden when maximizing the likelihood surface of the BNBGA and BNBIG regression models which are fitted to MTPL bodily injury and property damage claim count data. In particular, the EM scheme we present does not require knowledge of their joint probability mass functions (jpmfs), which cannot be written in closed forms, and can be easily implemented by taking advantage of the mixture representation of each model and is demonstrated to perform well.

Finally, to examine the suitability of the proposed family of MVMNB models for experience rating purposes, the a posteriori, or Bonus-Malus, premium rates resulting from the BNBGA and BNBIG models will be calculated via the net premium principle and compared to those determined by the bivariate Negative Binomial (BNB) and bivariate Poisson-Inverse Gaussian (BPIG) models, which can be regarded as natural extensions of the NB and PIG models that have been routinely used by actuaries for pricing risks in the univariate setting. The

[1] For instance, according to the most recent report by Insurance Europe, MTPL insurance accounted for almost one third of non-life business in the European Union, see Insurance Europe (2015).

main finding is that the BNBGA and BNBIG models show much less extreme a posteriori, or Bonus-Malus, premiums for policyholders with some MTPL bodily injury and property damage claims experience than those produced by their bivariate mixed Poisson counterparts. Therefore, the work presented herein can be viewed as complementary to the articles of Shengwang et al. (1999), Gómez-Déniz et al. (2008) and Tzougas et al. (2019) who reported similar findings regarding the comparison of the mixed Negative Binomial models, which they developed with the traditional NB and PIG models in a univariate a posteriori ratemaking context. Also, another striking difference between the BNBGA and BNBIG and the BNB and BPIG models is that, for a given total number of claims, the former models can enable the actuary to differentiate the premium rates based on the exact frequencies of MTPL bodily injury and property damage claims, whereas the latter models do not allow to price discriminate by taking into account the difference in the numbers of the two types of claims. Overall, from a practical business perspective, since MTPL remains the most widely purchased non-life product in world markets with policyholders shopping around for the best deals, due to the aforementioned reasons, the employment of the BNBGA and BNBIG models is beneficial for insurance companies, since compared to the two bivariate mixed Poisson models, it can enable them to better refine their a priori risk classification and restore fairness by designing merit rating plans in accordance with the a priori ratemaking structure of the company

The rest of this paper proceeds as follows: In Section 2 we present the derivation of the MVMNB regression model and we provide some of its desirable properties. Also, we consider the Gamma and Inverse-Gaussian mixing distributions and we derive the jpmfs of the corresponding BNBGA and BNBGIG regression models. Section 3 fully describes the ML estimation of the BNBGA and BNBIG models through the EM algorithm. Section 4 briefly presents the BNB and BPIG models, to which the BNBGA and BNBIG models are compared and defended as suiTable alternatives for computing a posteriori or Bonus-Malus premiums. Section 5 contains an application to a data set concerning MTPL insurance bodily injury and property damage claim counts. Finally, concluding remarks can be found in Section 6.

## 2. The multivariate mixed Negative Binomial regression model

### 2.1. The general setting

Consider that $k_{i,j}$ is the number of claims of type $i$ from policyholder contract $j$, where $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. Furthermore, let $k_i = \sum_{j=1}^{n} k_{i,j}$ denote the total number of claims of type $i$. Also, suppose that $\mathbf{x_{i,j}}$ denotes the (potentially different) covariate vectors related to the response variables $k_{i,j}$.

The claims of multiple types from policyholder contract $j$, which are collectively denoted as $k_{1,j}, k_{2,j}, ..., k_{m,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, ..., \boldsymbol{x_{m,j}}$, are assumed to follow a multivariate mixed Negative Binomial (MVMNB) distribution with a generic joint probability mass function (jpmf) which can be constructed as follows.

Consider that $k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j$, per claim type $i$, follows a Negative Binomial (NB) distribution with density

$$P\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \frac{\Gamma\left(k_{i,j} + \sigma_i\right)}{k_{i,j}!\Gamma\left(\sigma_i\right)} \left(\frac{\lambda_j \boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{k_{i,j}} \left(\frac{\sigma_i}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{\sigma_i}, \tag{1}$$

with $k_{i,j} = 0, 1, 2, 3, ...,$ $\lambda_j > 0$, $\sigma_i > 0$, where $\boldsymbol{\varepsilon_{i,j}} = \exp\left(\boldsymbol{x_{i,j}^T \beta_i}\right)$ and where $\boldsymbol{\beta_i}$ are the vectors of the regression coefficients per claim type $i$. The mean and the variance of $k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j$ are given by

$$\mathbb{E}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \boldsymbol{\varepsilon_{i,j}}\lambda_j \tag{2}$$

and

$$\text{Var}\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right) = \boldsymbol{\varepsilon_{i,j}}\lambda_j \left[1 + \frac{\boldsymbol{\varepsilon_{i,j}}\lambda_j}{\sigma_i}\right]. \tag{3}$$

It is worth noting that the scale parameter $\sigma_i$ controls the responsiveness of overdispersion to the mean number of claims, with the degree of overdispersion decreasing when $\sigma_i$ increases per claim type $i$. Note also that, in the limit, when $\sigma_i$ approaches infinity, $P\left(k_{i,j}|\boldsymbol{x_{i,j}}, \lambda_j\right)$ tends to the Poisson distribution with mean equal to $\boldsymbol{\varepsilon_{i,j}}\lambda_j$ per claim type $i$.

Let us now assume that $\lambda_j$ are independent and identically distributed (i.i.d) random variables from a continuous and at least twice differentiable mixing distribution with probability density function (pdf) given by

$$\lambda_j \sim f(\lambda_j; \gamma), \tag{4}$$

with $\gamma > 0$, mean $\mathbb{E}(\lambda_j) = 1$ and variance $\text{Var}(\lambda_j) = \gamma^2$. The prior, or mixing, distribution given by Eq. (4) has to have unit mean, in order for the model to be estimable.[2]

Under the assumptions in Eqs (1) and (4), the jpmf of the MVMNB distribution[3]

$$P(k_{1,j}, ..., k_{m,j}|\mathbf{x_{1,j}}, ..., \mathbf{x_{m,j}}) = \int_0^\infty \prod_{i=1}^{m} P(k_{i,j}|\mathbf{x_{i,j}}, \lambda_j) f(\lambda_j; \gamma) d\lambda_j. \tag{5}$$

The last integral cannot be solved in closed form but can be calculated through numerical integration or based on a Monte Carlo approach.

[2] Note that the unit mean requirement for the mixing density is essential for constructing mixture models with regression structures, otherwise identifiability issues can ensue see, for example, Karlis (2001), Rigby et al. (2008), Barreto-Souza and Simas (2016), Ghitany et al. (2012) and Tzougas (2020). among many others.

[3] Note also that, due to its quintuple mixture decomposition which is given in Section 3, the jpmf of the MVMNB model can be written in the form of the jpmf of a multivariate mixed Poisson (MVMP) distribution. Thus, the model has all the desirable theoretical properties of mixed Poisson models.

### 2.2. Model properties

In this subsection, we provide some desirable properties associated with the proposed model which ensure its flexibility for modelling overdispersed and positively correlated claim counts in non-life insurance.

1. The marginal distribution of $k_{i,j}|\boldsymbol{x_{i,j}}$, with $i=1,2,...,m$ and $j=1,2,...,n$, is a mixed Negative Binomial distribution with the same mixing distribution $f(\lambda_j;\gamma)$.

$$
\begin{aligned}
P(k_{i,j}|\mathbf{x_{i,j}}) &= \textstyle\sum_{k_{1,j}=0}^{\infty}\cdots\sum_{k_{i-1,j}=0}^{\infty}\sum_{k_{i+1,j}=0}^{\infty}\cdots\sum_{k_{m,j}=0}^{\infty} P(k_{1,j},...,k_{m,j}|\mathbf{x_{1,j}},...,\mathbf{x_{m,j}})\\
&= \int\left[\textstyle\prod_{l\neq i}^{m}\sum_{k_{l,j}=0}^{\infty} P\left(k_{l,j}|\boldsymbol{x_{l,j}},\lambda_j\right)\right] P(k_{i,j}|\mathbf{x_{i,j}},\lambda_j) f(\lambda_j;\gamma)d\lambda_j\\
&= \int P(k_{i,j}|\mathbf{x_{i,j}},\lambda_j) f(\lambda_j;\gamma)d\lambda_j.
\end{aligned}
\tag{6}
$$

This property can enable the actuary price contracts for which insureds may participate in some but not all lines of businesses.

2. Using the laws of total expectation and total variance and the moments of the Negative Binomial distribution, one can find that the mean and variance of $k_{i,j}|\mathbf{x_{i,j}}$ are as follows

$$
\mathbb{E}(k_{i,j}|\mathbf{x_{i,j}}) = \mathbb{E}_{\lambda_j}\left[\mathbb{E}(k_{i,j}|\mathbf{x_{i,j}},\lambda_j)\right] = \boldsymbol{\varepsilon_{i,j}}\mathbb{E}_{\lambda_j}(\lambda_j) = \boldsymbol{\varepsilon_{i,j}}. \tag{7}
$$

and

$$
\begin{aligned}
\text{Var}(k_{i,j}|\mathbf{x_{i,j}}) &= \mathbb{E}_{\lambda_j}\left[\text{Var}(k_{i,j}|\mathbf{x_{i,j}},\lambda_j)\right] + \text{Var}_{\lambda_j}\left[\mathbb{E}(k_{i,j}|\mathbf{x_{i,j}},\lambda_j)\right]\\
&= \boldsymbol{\varepsilon_{i,j}} + \boldsymbol{\varepsilon_{i,j}^2}\Big(\frac{1}{\sigma_i} + \frac{\gamma^2}{\sigma_i} + \gamma^2\Big).
\end{aligned}
\tag{8}
$$

3. Based on the laws of total variance and total covariance, we can see that the covariance between $k_{i,j}|\mathbf{x_{i,j}}$ and $k_{l,j}|\mathbf{x_{l,j}}$, for $i,l\in\{1,..,m\}$ and $i\neq l$, is given by

$$
\begin{aligned}
\text{Cov}(k_{i,j},k_{l,j}|\mathbf{x_{i,j}},\mathbf{x_{l,j}}) &= \mathbb{E}\left[\text{Cov}(k_{i,j},k_{l,j}|\mathbf{x_{i,j}},\mathbf{x_{l,j}},\lambda_j)\right] + \text{Cov}\left[\mathbb{E}(k_{i,j}|\mathbf{x_{i,j}},\lambda_j),\mathbb{E}(k_{l,j}|\mathbf{x_{l,j}},\lambda_j)\right]\\
&= 0 + \text{Cov}\left[\boldsymbol{\varepsilon_{i,j}}\lambda_j \times \epsilon_{l,j}\lambda_j\right]\\
&= \boldsymbol{\varepsilon_{i,j}}\epsilon_{l,j}\text{Var}(\lambda_j) = \boldsymbol{\varepsilon_{i,j}}\epsilon_{l,j}\gamma^2.
\end{aligned}
\tag{9}
$$

Thus, the correlation between $k_{i,j}|\mathbf{x_{i,j}}$ and $k_{l,j}|\mathbf{x_{l,j}}$, for $i,l\in\{1,..,m\}$ and $i\neq l$, is given by

$$
\text{Corr}(k_{i,j},k_{l,j}|\mathbf{x_{i,j}},\mathbf{x_{l,j}}) = \frac{\boldsymbol{\varepsilon_{i,j}}\epsilon_{l,j}\gamma^2}{\sqrt{\boldsymbol{\varepsilon_{i,j}} + \boldsymbol{\varepsilon_{i,j}^2}\left(\frac{1}{\sigma_i} + \frac{\gamma^2}{\sigma_i} + \gamma^2\right)}\sqrt{\epsilon_{l,j} + \epsilon_{l,j}^2\left(\frac{1}{\sigma_l} + \frac{\gamma^2}{\sigma_l} + \gamma^2\right)}}. \tag{10}
$$

4. The generalized variance ratio (GVR) between the MVMNB model, as defined in Eq. (5), and a simple NB model, i.e. $y_{i,j} \sim \text{NB}\left(\sigma_i, \frac{\boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \boldsymbol{\varepsilon_{i,j}}}\right)$ is given by

$$
\begin{aligned}
\text{GVR}(k_{i,j},k_{l,j}|\mathbf{x_{i,j}},\mathbf{x_{l,j}}) &= \frac{\text{Var}(k_{i,j}|\mathbf{x_{i,j}}) + \text{Var}(k_{l,j}|\mathbf{x_{l,j}}) + 2\text{Cov}(k_{i,j},k_{l,j}|\mathbf{x_{i,j}},\mathbf{x_{l,j}})}{\text{Var}(k_{i,j}|\mathbf{x_{i,j}},\lambda=1) + \text{Var}(k_{l,j}|\mathbf{x_{l,j}},\lambda=1)}\\
&= \frac{\boldsymbol{\varepsilon_{i,j}} + \boldsymbol{\varepsilon_{i,j}^2}\left(\frac{1}{\sigma_i} + \frac{\gamma^2}{\sigma_i} + \gamma^2\right) + \epsilon_{l,j} + \epsilon_{l,j}^2\left(\frac{1}{\sigma_l} + \frac{\gamma^2}{\sigma_l} + \gamma^2\right) + 2\boldsymbol{\varepsilon_{i,j}}\epsilon_{l,j}\gamma^2}{\boldsymbol{\varepsilon_{i,j}} + \frac{\epsilon_{i,j}^2}{\sigma_i} + \epsilon_{l,j} + \frac{\epsilon_{l,j}^2}{\sigma_l}}\\
&= 1 + \gamma^2 \times \frac{\frac{\boldsymbol{\varepsilon_{i,j}^2}}{\sigma_i} + \frac{\epsilon_{l,j}^2}{\sigma_l} + (\boldsymbol{\varepsilon_{i,j}} + \epsilon_{l,j})^2}{\frac{\boldsymbol{\varepsilon_{i,j}^2}}{\sigma_i} + \frac{\epsilon_{l,j}^2}{\sigma_l} + (\boldsymbol{\varepsilon_{i,j}} + \epsilon_{l,j})}.
\end{aligned}
\tag{11}
$$

### 2.3. Model specifications

For demonstration purposes, from now on we focus on the bivariate case $i=2$. Furthermore, motivated by the characteristics of the MTPL insurance data which we will analyze in Section 5, we propose the use of the Gamma (GA) and Inverse Gaussian (IG) distributions as mixing densities. The resulting bivariate Negative Binomial-Gamma (BNBGA) and Negative Binomial-Inverse Gaussian (BNBIG) regression models allow for the positive correlation between bodily injury and property damage claims since $\text{Corr}\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}}\right) > 0$, see Eq. (10), and can accommodate the bivariate overdispersion in the data since GVR $> 1$, see Eq. (11). Thus, the BNBGA and BNBIG models are ideally suited for capturing the characteristics of two-dimensional MTPL insurance data since, as was previously mentioned, positive correlation and overdispersion are the two phenomena that have been most commonly reported in the pricing literature in the bivariate setting. Furthermore, as we will observe in Section 5, the BNBGA and BNBIG models will provide better fitting performances compared to bivariate mixed Poisson benchmarks which can be derived using the same mixing distributions.

At this point, it is worth noting that different mixing densities might be more appropriate for different MTPL data sets. In particular, as it can be clearly understood, there are many factors which cannot be directly observed by the actuary but can simultaneously affect the joint dynamics of MTPL bodily injury and property damage claims, leading to extra variation occurring in their associated claim counts. Thus, the choice of the mixing density, which measures the level of unobservable risk associated with each of the policies, is crucial since a potential distribution misspecification can result in biased and unreliable parameter estimates, which, in turn, can have an impact on how insurers price the policy, leading to financial implications for the company, since, if the punishment of all policyholders is not justified on a sound risk measuring basis, then they may switch to competing companies. The EM type algorithm which we will present in Section 3 for fitting the BNBGA and BNBIG models has sufficient flexibility to estimate alternative bivariate (and/or multivariate) Negative Binomial mixture models stemming from several other continuous and at least twice differentiable mixing distributions with a unit mean. However, for some mixing densities a special iterative scheme and potentially another EM algorithm within the M-step may be necessary.

#### 2.3.1. The bivariate Negative Binomial-Gamma regression model

Let $\lambda_j$, for $j = 1, 2, \ldots, n$, follow a GA distribution with a pdf of the form

$$f(\lambda_j; \gamma) = \frac{\gamma^{\gamma}}{\Gamma(\gamma)} \exp(-\gamma\lambda_j)\, \lambda_j^{\gamma-1}, \tag{12}$$

where $\gamma > 0$, with mean and variance $\mathbb{E}(\lambda_j) = 1$ and $\mathrm{Var}(\lambda_j) = 1/\gamma$.

Thus, based on Eqs (1) and (12) it is easy to see that the resulting distribution is the BNBGA distribution with jpmf

$$\begin{aligned} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \int_0^{\infty} \prod_{i=1}^{2} \frac{\Gamma\left(k_{i,j} + \sigma_i\right)}{k_{i,j}!\,\Gamma(\sigma_i)} \left(\frac{\lambda_j \boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{k_{i,j}} \left(\frac{\sigma_i}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{\sigma_i} \\ \times \frac{\gamma^{\gamma}}{\Gamma(\gamma)} \exp(-\gamma\lambda_j)\, \lambda_j^{\gamma-1} d\lambda_j. \end{aligned} \tag{13}$$

Note that the mean and the variance for $k_{i,j}|\mathbf{x_{i,j}}$ and the covariance and correlation for $k_{i,j}|\mathbf{x_{i,j}}$ and $k_{l,j}|\mathbf{x_{l,j}}$, for $i, l \in \{1, .., m\}$ and $i \neq l$, in the case of the BNBGA distribution are given by Eqs (7), (8), (9) and (10) but in the latter three equations $\gamma^2$ should be replaced with $1/\gamma$.

#### 2.3.2. The bivariate Negative Binomial-Inverse Gaussian regression model

Let $\lambda_j$, for $j = 1, 2, \ldots, n$, follow an IG distribution with pdf given by

$$f(\lambda_j; \gamma) = \frac{\gamma}{\sqrt{2\pi}} \exp\left(\gamma^2\right) \lambda_j^{-\frac{3}{2}} \exp\left[-\frac{1}{2}\left(\frac{\gamma^2}{\lambda_j} + \gamma^2 \lambda_j\right)\right], \tag{14}$$

with $\gamma > 0$, mean $\mathbb{E}(\lambda_j) = 1$ and variance $\mathrm{Var}(\lambda_j) = \frac{1}{\gamma^2}$. Because of the unit mean restriction, it follows that the overdispersion linked to the simple exponential distribution is $\frac{1}{\gamma^2}$ and hence the IG will reduce to the Exponential if $\gamma$ tends to infinity. For more information about the IG distribution, which is a special case of the generalized Inverse Gaussian (GIG) distribution, the interested reader can refer to Jørgensen (1982). Note also that several other parameterizations of the IG can be found in Seshadri (1993).

Thus, considering the assumptions in Eqs (1) and (14), it can be verified that the resulting distribution is the BNBIG distribution with jpmf[4]

$$\begin{aligned} P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \int_0^{\infty} \prod_{i=1}^{2} \frac{\Gamma\left(k_{i,j} + \sigma_i\right)}{k_{i,j}!\,\Gamma(\sigma_i)} \left(\frac{\lambda_j \boldsymbol{\varepsilon_{i,j}}}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{k_{i,j}} \left(\frac{\sigma_i}{\sigma_i + \lambda_j \boldsymbol{\varepsilon_{i,j}}}\right)^{\sigma_i} \\ \times \frac{\gamma}{\sqrt{2\pi}} \exp\left(\gamma^2\right) \lambda_j^{-\frac{3}{2}} \exp\left[-\frac{1}{2}\left(\frac{\gamma^2}{\lambda_j} + \gamma^2 \lambda_j\right)\right] d\lambda_j. \end{aligned} \tag{15}$$

Note that the mean and the variance for $k_{i,j}|\mathbf{x_{i,j}}$ and the covariance and correlation for $k_{i,j}|\mathbf{x_{i,j}}$ and $k_{l,j}|\mathbf{x_{l,j}}$, for $i, l \in \{1, .., m\}$ and $i \neq l$, in the case of the BNBIG distribution are given by Eqs (7), (8), (9) and (10) but in the latter three equations $\gamma^2$ should be replaced with $1/\gamma^2$.

## 3. The EM algorithm for ML estimation of the BNBGA and BNBIG regression models

Let $(k_{1,j}, k_{2,j}; \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}})$ be a sample of observations, $j = 1, \ldots, n$, where the responses $k_{i,j}$ are the number of claims for the policyholder $j$ and where $\boldsymbol{x_{i,j}}$ are the vectors of covariate information per claim type $i = 1, 2$. Considering that the data are produced according to a bivariate mixed NB model, its log-likelihood can be expressed as

$$l(\phi) = \sum_{j=1}^{n} \log\left(P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right)\right), \tag{16}$$

where $\phi = \left(\gamma, \sigma_1, \sigma_2, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}\right)$ is the vector of the parameters and where $P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right)$ is the jpmf of the bivariate mixed NB model which is given by Eqs (13) and (15) in the case of the BNBGA and BNBIG distributions respectively.

[4] Note that Gómez-Déniz et al. (2008) gave an excellent account of statistical methods connected to both the univariate and multivariate versions of the NBIG distribution. The BNBIG distribution as in Eq. (15) may be distinguished from the one by Gómez-Déniz et al. (2008) as the latter does not allow to include covariate information.

The log-likelihood given by Eq. (16) does not exist in closed form and hence $\phi$ cannot be estimated via traditional numerical maximization methods. In such cases, it is necessary to resort to the EM algorithm (see Dempster et al. (1977) and McLachlan and Krishnan (2007)). In particular, if one augments the unobserved data, denoted by $\lambda_j$ herein, to $(k_{1,j}, k_{2,j}; \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}})$ for $j = 1, \ldots, n$, then the complete data log-likelihood of the bivariate mixed NB regression model factorises into two parts:

$$l_c(\phi) = \sum_{j=1}^{n} \sum_{i=1}^{2} \log\left(P\left(k_{i,j} | \boldsymbol{x_{i,j}}, \lambda_j\right)\right) + \sum_{j=1}^{n} \log\left(f\left(\lambda_j; \gamma\right)\right), \tag{17}$$

where $P\left(k_{i,j} | \boldsymbol{x_{i,j}}, \lambda_j\right)$ is the pmf of each of the two NB distributions, which are given by Eq. (1) for $i = 1, 2$, and where $f(\lambda_j; \gamma)$ represents the pdf of the mixing distribution. Direct maximization of the complete data log-likelihood, as given by Eq. (17), with respect to $\phi$ is cumbersome. Fortunately, its ML estimation can be easily achieved via the EM algorithm, if one takes advantage of the following quintuple mixture derivation of the model. In particular, the number of claims $k_{1,j}$ is distributed as:

$$\begin{aligned} &k_{1,j} \sim \text{Poisson}\left(\boldsymbol{\vartheta_{1,j}}\right) \\ &\text{with } \boldsymbol{\vartheta_{1,j}} \sim \text{Gamma}\left(\sigma_1, \frac{\sigma_1}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}\right), \end{aligned} \tag{18}$$

the number of claims $k_{2,j}$ is distributed as:

$$\begin{aligned} &k_{2,j} \sim \text{Poisson}\left(\boldsymbol{\vartheta_{2,j}}\right) \\ &\text{with } \boldsymbol{\vartheta_{2,j}} \sim \text{Gamma}\left(\sigma_2, \frac{\sigma_2}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}\right) \end{aligned} \tag{19}$$

and both $k_{1,j}$ and $k_{2,j}$ share the same unobserved heterogeneity variable $\lambda_j$ which is distributed as

$$\lambda_j \sim f(\lambda_j; \gamma), \tag{20}$$

which is given by Eqs (12) and (14) in the case of the Gamma and Inverse Gaussian mixing distributions respectively.

Also, let us denote

$$P\left(k_{i,j} | \boldsymbol{\vartheta_{i,j}}\right) = e^{-\boldsymbol{\vartheta}_{i,j}} \frac{\boldsymbol{\vartheta_{i,j}}^{k_{i,j}}}{k_{i,j}!}, \tag{21}$$

for $i = 1, 2$, to be the two pmfs of $k_{1,j}$ and $k_{2,j}$ respectively, and

$$g_1\left(\boldsymbol{\vartheta_{1,j}} | \boldsymbol{\beta_1}, \sigma_1, \lambda_j\right) = \boldsymbol{\vartheta_{1,j}}^{\sigma_1 - 1} \exp\left(-\frac{\sigma_1}{\lambda_j \boldsymbol{\varepsilon_{1,j}}} \boldsymbol{\vartheta_{1,j}}\right) \left(\frac{\sigma_1}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}\right)^{\sigma_1} / \Gamma(\sigma_1) \tag{22}$$

and

$$g_2\left(\boldsymbol{\vartheta_{2,j}} | \boldsymbol{\beta_2}, \sigma_2, \lambda_j\right) = \boldsymbol{\vartheta_{2,j}}^{\sigma_2 - 1} \exp\left(-\frac{\sigma_2}{\lambda_j \boldsymbol{\varepsilon_{2,j}}} \boldsymbol{\vartheta_{2,j}}\right) \left(\frac{\sigma_2}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}\right)^{\sigma_2} / \Gamma(\sigma_2) \tag{23}$$

to be the pdfs of the two Gamma distributions.

Then, using the mixture representation in Eqs (18), (19) and (20) the complete data log-likelihood is proportional to

$$\begin{aligned} l_c(\phi) \propto &\sum_{j=1}^{n} \log\left(g_1\left(\boldsymbol{\vartheta_{1,j}} | \boldsymbol{\beta_1}, \sigma_1, \lambda_j\right)\right) + \\ &\sum_{j=1}^{n} \log\left(g_2\left(\boldsymbol{\vartheta_{2,j}} | \boldsymbol{\beta_2}, \sigma_2, \lambda_j\right)\right) + \sum_{j=1}^{n} \log\left(f\left(\lambda_j; \gamma\right)\right). \end{aligned} \tag{24}$$

The regression coefficients $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ and the parameters $\sigma_1$ and $\sigma_2$ are involved in the first and second terms and the parameter $\gamma$ is involved in the third term of Eq. (24), which correspond to the log-likelihoods of the two Gamma components and the Inverse Gaussian component respectively.

Therefore, $Q$-function, which is the conditional expectation of the complete data log-likelihood, is proportional to

$$\begin{aligned} Q\left(\phi;\phi_{(r)}\right) &\equiv \mathbb{E}_{\lambda_j}\left(l_c\left(\phi\right)|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right) \\ &\propto \mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(g_1\left(\boldsymbol{\vartheta_{1,j}}|\boldsymbol{\beta_{1,(r)}},\sigma_{1,(r)},\lambda_j\right)\right)\right] \\ &\quad+\mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(g_2\left(\boldsymbol{\vartheta_{2,j}}|\boldsymbol{\beta_{2,(r)}},\sigma_{2,(r)},\lambda_j\right)\right)\right] \\ &\quad+\mathbb{E}_{\lambda_j}\left[\sum_{j=1}^{n}\log\left(f\left(\lambda_j;\gamma_{(r)}\right)\right)\right], \end{aligned} \tag{25}$$

where $\phi_{(r)}=\left(\gamma_{(r)},\sigma_{1,(r)},\sigma_{2,(r)},\boldsymbol{\beta_{1,(r)}},\boldsymbol{\beta_{2,(r)}}\right)$ is the estimate of $\phi$ in the E-step of our EM algorithm.

In what follows, some functions of the unobserved data $\lambda_j$ which are involved in Eq. (25) will be calculated for implementing the E-step of the algorithm, while the M-step, involves maximizing the $Q$-function with respect to $\phi$. Also, the following posterior distributions will be needed in the E-step of the EM algorithm:

$$\boldsymbol{\vartheta_{1,j}}|k_{1,j},\boldsymbol{x_{1,j}},\sigma_1,\boldsymbol{\beta_1}\sim\text{Gamma}\left(k_{1,j}+\sigma_1,\frac{\sigma_1}{\lambda_j\boldsymbol{\varepsilon_{1,j}}}+1\right) \tag{26}$$

and

$$\boldsymbol{\vartheta_{2,j}}|k_{2,j},\boldsymbol{x_{2,j}},\sigma_2,\boldsymbol{\beta_2}\sim\text{Gamma}\left(k_{2,j}+\sigma_2,\frac{\sigma_2}{\lambda_j\boldsymbol{\varepsilon_{2,j}}}+1\right). \tag{27}$$

The EM algorithm can now be formally described as follows.

**E-Step:**

- Compute the pseudo-values for $j=1,2,...,n$, using the parameters' values after the $r$-th iteration

$$w_{1,j}=\mathbb{E}_{\lambda_j}\left(\lambda_j|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right)=\frac{\int_0^\infty\lambda_jP\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}, \tag{28}$$

$$w_{2,j}=\mathbb{E}_{\lambda_j}\left(\frac{1}{\lambda_j}|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right)=\frac{\int_0^\infty\frac{1}{\lambda_j}P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}, \tag{29}$$

$$\begin{aligned} w_{3,j}&=\mathbb{E}_{\lambda_j}\left(\frac{1}{\left(\lambda_j\boldsymbol{\varepsilon_{1,j}}+\sigma_{1,(r)}\right)}|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right) \\ &=\frac{\int_0^\infty\frac{1}{\left(\lambda_j\boldsymbol{\varepsilon_{1,j}}+\boldsymbol{\sigma_{1,(r)}}\right)}P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}, \end{aligned} \tag{30}$$

$$\begin{aligned} w_{4,j}&=\mathbb{E}_{\lambda_j}\left(\log\left(\lambda_j\boldsymbol{\varepsilon_{1,j}}+\sigma_{1,(r)}\right)|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right) \\ &=\frac{\int_0^\infty\log\left(\lambda_j\boldsymbol{\varepsilon_{1,j}}+\sigma_{1,(r)}\right)P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}, \end{aligned} \tag{31}$$

$$\begin{aligned} w_{5,j}&=\mathbb{E}_{\lambda_j}\left(\frac{1}{\left(\lambda_j\boldsymbol{\varepsilon_{2,j}}+\sigma_{2,(r)}\right)}|k_{1,j},k_{2,j},\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\phi_{(r)}\right) \\ &=\frac{\int_0^\infty\frac{1}{\left(\lambda_j\boldsymbol{\varepsilon_{2,j}}+\sigma_{2,(r)}\right)}P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}{\int_0^\infty P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_j;\gamma\right)d\lambda_j}, \end{aligned} \tag{32}$$

$$w_{6,j} = \mathbb{E}_{\lambda_j}\left(\log\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right) = \frac{\int_0^\infty \log\left(\lambda_j \boldsymbol{\varepsilon_{2,j}} + \sigma_{2,(r)}\right) P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}{\int_0^\infty P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \lambda_j\right) f\left(\lambda_j; \gamma\right) d\lambda_j}, \tag{33}$$

which is given by Eqs (12) and (14) in the case of the Gamma and Inverse Gaussian mixing distributions respectively.

- Using Eqs (26), (27), (30), (31), (32) and (33) we obtain that

$$s_{1,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{1,j}}}\left(\frac{\boldsymbol{\vartheta_{1,j}}}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \left(k_{1,j} + \sigma_{1,(r)}\right) w_{3,j}, \tag{34}$$

$$s_{2,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{1,j}}}\left(\log\left(\frac{\boldsymbol{\vartheta_{1,j}}}{\lambda_j \boldsymbol{\varepsilon_{1,j}}}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \Psi\left(k_{1,j} + \sigma_{1,(r)}\right) - w_{4,j} \tag{35}$$

and

$$s_{3,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{2,j}}}\left(\frac{\boldsymbol{\vartheta_{2,j}}}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \left(k_{2,j} + \sigma_{2,(r)}\right) w_{5,j}, \tag{36}$$

$$s_{4,j} = \mathbb{E}_{\lambda_j}\left[\mathbb{E}_{\boldsymbol{\vartheta_{2,j}}}\left(\log\left(\frac{\boldsymbol{\vartheta_{2,j}}}{\lambda_j \boldsymbol{\varepsilon_{2,j}}}\right)|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi_{(r)}\right)\right] = \Psi\left(k_{2,j} + \sigma_{2,(r)}\right) - w_{6,j}, \tag{37}$$

with $\Psi(\cdot)$ representing the digamma function. Clearly the expectations given by Eqs (28), (29), (30), (31), (32) and (33) cannot be written in closed form and thus they can be evaluated numerically. Alternatively, a Monte Carlo approach can also be used based on a rejection algorithm. The latter case leads to variants of the EM algorithm such as the Monte Carlo EM algorithm, see, for example, Booth and Hobert (1999), Booth et al. (2001) and Karlis (2005).

**M-step:**

In the M-Step, the pseudo-values from the E-Step can be used to maximize the $Q$-function.

- Firstly, the Newton-Raphson algorithm is employed to obtain ML estimates of the two vectors of regression coefficients $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$. Differentiating $Q\left(\phi; \phi_{(r)}\right)$ with respect to $\boldsymbol{\beta_1}$, we find:

$$g_1\left(\boldsymbol{\beta_1}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial l_c}{\partial \boldsymbol{\beta_1}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = \sigma_1 \sum_{j=1}^{n}\left(s_{1,i} - 1\right)\boldsymbol{x_{1,j}} \tag{38}$$

and

$$G_1\left(\boldsymbol{\beta_1}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial^2 l_c}{\partial \boldsymbol{\beta_1}\partial \boldsymbol{\beta_1}^T}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = -\sigma_1 \sum_{j=1}^{n} s_{1,i}\boldsymbol{x_{1,j}}\boldsymbol{x_{1,j}}^T = -\sigma_1 \mathbf{X_1}^T \mathbf{W_1} \mathbf{X_1}, \tag{39}$$

where $\mathbf{W_1} = diag\{s_{1,i}\}$.

Therefore, the Newton-Raphson iterative procedure for obtaining ML estimates of the elements of $\beta_1$ goes as follows:

$$\boldsymbol{\beta_{1,(r+1)}} \equiv \boldsymbol{\beta_{1,(r)}} - \left[G_1\left(\boldsymbol{\beta_{1,(r)}}\right)\right]^{-1} g_1\left(\boldsymbol{\beta_{1,(r)}}\right). \tag{40}$$

Then, following the same procedure for $\boldsymbol{\beta_2}$, we obtain:

$$g_2\left(\boldsymbol{\beta_2}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial l_c}{\partial \boldsymbol{\beta_2}}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = \sigma_2 \sum_{j=1}^{n}\left(s_{3,i} - 1\right)\boldsymbol{x_{2,j}} \tag{41}$$

and

$$G_2\left(\boldsymbol{\beta_2}\right) = \mathbb{E}_{\lambda_j}\left(\frac{\partial^2 l_c}{\partial \boldsymbol{\beta_2}\partial \boldsymbol{\beta_2}^T}|k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \phi\right) = -\sigma_2 \sum_{j=1}^{n} s_{3,i}\boldsymbol{x_{2,j}}, \boldsymbol{x_{2,j}}^T = -\sigma_2 \mathbf{X_2}^T \mathbf{W_2} \mathbf{X_2}, \tag{42}$$

where $\mathbf{W_2} = diag\{s_{3,i}\}$.

Therefore the iterated $\boldsymbol{\beta_2}$ is:

$$\boldsymbol{\beta_{2,(r+1)}} \equiv \boldsymbol{\beta_{2,(r)}} - \left[G_2\left(\boldsymbol{\beta_{2,(r)}}\right)\right]^{-1} g_2\left(\boldsymbol{\beta_{2,(r)}}\right). \tag{43}$$

- Secondly, the one step ahead Newton iteration is used twice for updating $\sigma_1$ and $\sigma_2$:

$$\sigma_{1,(r+1)} = \sigma_{1,(r)} - \frac{\Psi\left(\sigma_{1,(r)}\right) + \bar{s}_1 - \bar{s}_2 - \log\left(\sigma_{1,(r)}\right) - 1}{\Psi_3\left(\sigma_{1,(r)}\right) - \frac{1}{\sigma_{1,(r)}}}, \tag{44}$$

$$\sigma_{2,(r+1)} = \sigma_{2,(r)} - \frac{\Psi\left(\sigma_{2,(r)}\right) + \bar{s}_3 - \bar{s}_4 - \log\left(\sigma_{2,(r)}\right) - 1}{\Psi_3\left(\sigma_{2,(r)}\right) - \frac{1}{\sigma_{2,(r)}}}, \tag{45}$$

where $\Psi_3(\cdot)$ denotes the trigamma function.

- Finally, update $\gamma$ with

$$\gamma_{(r+1)} = \gamma_{(r)} - \frac{\Psi\left(\gamma_{(r)}\right) + \bar{w}_1 + \bar{w}_2 - \log\left(\gamma_{(r)}\right) - 1}{\Psi_3\left(\gamma_{(r)}\right) - \frac{1}{\gamma_{(r)}}} \tag{46}$$

and

$$\gamma_{(r+1)} = (\bar{w}_1 + \bar{w}_2 - 2)^{-\frac{1}{2}}, \tag{47}$$

in the case of the Gamma and Inverse Gaussian mixing distributions respectively.

- Note also that if the regression components of the model for the two responses $k_{1,j}$ and $k_{2,j}$ are limited to the constants $\beta_{1,0}$ and $\beta_{2,0}$, then we have that $\mathbb{E}\left(k_{1,j}|\boldsymbol{x_{1,j}}\right) = \exp\left(\beta_{1,0}\right) = \mu_1$ and $\mathbb{E}\left(k_{2,j}|\boldsymbol{x_{2,j}}\right) = \exp\left(\beta_{2,0}\right) = \mu_2$; and hence the ML estimation for the bivariate distribution, i.e. without regression components, can be computed via the EM type algorithm.

## 4. The BNB and BPIG regression models

In our numerical illustration the a posteriori, or Bonus-Malus, premium rates resulting from the bivariate Negative Binomial (BNB) and bivariate Poisson-Inverse Gaussian (BPIG) models will be compared to those determined by the BNBGA and BNBIG model. Therefore, we give below some rudimentary facts concerning the BNB and the BPIG models.

Consider that $k_{i,j} = 0, 1, 2, 3, \ldots$ are the number of bodily injury and the number of property damage claims, when $i = 1, 2$ respectively, for the policyholder $j$, with $j = 1, \ldots, n$ and suppose that $\boldsymbol{x_{i,j}}$ are the vectors of individual characteristics and/or characteristics of the car related to the $j$-th insured person per claim type $i = 1, 2$. Also, let $\boldsymbol{\varepsilon_{i,j}} = \exp\left(\boldsymbol{x_{i,j}^T \beta_i}\right)$, where $\boldsymbol{\beta_i}$ are the two vectors of the regression coefficients for $i = 1, 2$.

- The jpmf of the BNB model is given by

$$P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \frac{\Gamma\left(\sum_{i=1}^{2} k_{i,j} + \gamma\right)}{\Gamma(\gamma)\prod_{i=1}^{2} k_{i,j}!} \frac{\gamma^{\gamma} \prod_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}^{k_{i,j}}}{(\gamma + \sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}})^{\gamma + \sum_{i=1}^{2} k_{i,j}}}, \tag{48}$$

for $\gamma > 0$.

- The jpmf of the BPIG is given by

$$P\left(k_{1,j}, k_{2,j}|\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \frac{2\gamma e^{\gamma^2}}{\sqrt{2\pi}} K_{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}}(\gamma\omega) \left(\frac{\gamma}{\Delta}\right)^{\sum_{i=1}^{2} k_{i,j} - \frac{1}{2}} \prod_{i=1}^{2} \frac{\boldsymbol{\varepsilon_{i,j}}^{k_{i,j}}}{k_{i,j}!}, \tag{49}$$

where $\gamma > 0$, $\omega = \sqrt{\gamma^2 + 2\sum_{i=1}^{2} \boldsymbol{\varepsilon_{i,j}}}$ and where $K_r(\cdot)$ denotes the modified Bessel function of the third kind of order $r$.

Similarly to the BNBGA and BNBIG models, the BNB and BPIG are plausible models for overdispersed two-dimensional positively correlated MTPL claim count data. For more details about the BNB and BPIG models, the interested reader can refer to Stein and Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Munkin and Trivedi (1999), Gurmu and Elder (2000) and Ghitany et al. (2012).

## 5. Numerical illustration

The data were kindly provided by a major European insurance company and concern a MTPL insurance portfolio which was observed during the year 2017. We model two types of claims and their associated claim counts are recorded as $k_{1,j}$ and $k_{2,j}$, for the policyholder $j$, with $j = 1, \ldots, n$, which, as was previously mentioned, represent MTPL bodily injury and property damage claims respectively. The sample comprised insured parties with complete records; i.e., with the availability of all the explanatory variables which affect both $k_{1,j}$ and $k_{2,j}$. Additionally, an exploratory analysis was carried out in order to adequately select the subset of explanatory variables with the highest predictive power for both $k_{1,j}$ and $k_{2,j}$. There were $n = 5186$ observations that met our criteria. These explanatory variables[5] are summarized in Table 1.

[5] Note that it would be interesting to fit the same models to larger data sets in order to study the effect of other categorical and continuous explanatory variables such as age of driver, driving experience or driving zone, which have been traditionally used in MTPL insurance.

**Table 1**
The explanatory variables and their description.

| Variables | Categories | | |
|---|---|---|---|
| | C1 | C2 | C3 |
| City population (v1) | $\leq$ 1,000,000 | 1,000,001-2,000,000 | $\geq$ 2,000,001 |
| Number of years that the policyholder has been registered with the insurance company (v2) | $<$ 5 years | $>$ 5 years | - |
| Horsepower of the vehicle (v3) | 0-1400 cc | 1400-1800 cc | $\geq$ 1800 cc |

**Table 2**
Summary statistics for claim frequencies as classified by the explanatory variables.

| | | $k_1$ | | | $k_2$ | | |
|---|---|---|---|---|---|---|---|
| Covariates | Total | Count=0 (%) | Count=1 (%) | Count $\geq$2 (%) | Count=0 (%) | Count=1 (%) | Count$\geq$2 (%) |
| v1 C1 | 2203 | 92.81 | 5.28 | 1.91 | 94.68 | 5.04 | 0.28 |
| v1 C2 | 2386 | 92.98 | 4.65 | 2.37 | 93.76 | 5.94 | 0.30 |
| v1 C3 | 597 | 92.34 | 4.86 | 2.80 | 93.46 | 6.35 | 0.19 |
| v2 C1 | 4491 | 93.00 | 4.68 | 2.32 | 94.25 | 5.48 | 0.27 |
| v2 C2 | 695 | 91.69 | 6.81 | 1.50 | 93.19 | 6.48 | 0.33 |
| v3 C1 | 2372 | 92.77 | 5.02 | 2.21 | 94.28 | 5.50 | 0.22 |
| v3 C2 | 1815 | 93.84 | 3.99 | 2.17 | 94.35 | 5.36 | 0.29 |
| v3 C3 | 999 | 91.14 | 6.51 | 2.35 | 93.27 | 6.30 | 0.43 |

**Table 3**
Descriptive statistics for the two responses.

| $k_1$ | | $k_2$ | |
|---|---|---|---|
| statistic | value | statistic | value |
| Minimum | 0 | Minimum | 0 |
| Median | 0 | Median | 0 |
| Mean | 0.0954 | Mean | 0.0618 |
| Variance | 0.1375263 | Variance | 0.06439364 |
| Maximum | 4 | Maximum | 3 |
| Kendall's $\tau$: 0.17595 | | | |

Table 2 presents a summary of the effects of the covariates on the MTPL bodily injury and property damage claim counts $k_{1,j}$ and $k_{2,j}$ based on all 5186 observations. In the first column there is a list of all explanatory variables, all broken down in their respective categories. The second column represents how many policyholders, out of our data set of 5186 observations, fall into each subgroup/category for every covariate. The rest of the Table shows, conditionally on being included in a certain category per explanatory variable, the percentage of policies with claim frequencies equal to $0, 1, \geq 2$ for $k_{1,j}$ and $k_{2,j}$ respectively. For example, from Table 2, we can make the following observations. Firstly, in the case of the variable city population (v1) regarding the 2203 policyholders who live in a small city (C1), 92.81% of them have not made bodily injury claims and 94.68% of them have had no property damage claims. On the other hand, the 597 individuals who live in a large city (C3) seem to make more claims per both types, since the percentage that has resulted claim-free dropped to 92.34% and 93.46% for bodily injury and property damage claims respectively. Secondly, as far as the variable number of years that the policyholder has been registered with the insurance company (v2) is concerned, we see that the longer a policyholder has been with the company (C2), the bigger is the probability of its getting involved in an accident. Thirdly, regarding the variable horsepower of the car (v3), we observe that a high horsepower (C3) seems to be a risky category which corresponds to a lower number of claim-free policyholders, for both $k_{1,j}$ and $k_{2,j}$.

Table 3 shows some standard descriptive statistics for the bodily injury and property damage claims $k_{1,j}$ and $k_{2,j}$ respectively, along with the value of the Kendall's $\tau$ correlation coefficient.

### 5.1. Modelling results

This subsection presents the modelling results of the BNB, BPIG, BNBGA and BNBIG distributions/regression models. All computing was done using the statistical computing environment language **R**. The BNBGA and BNBIG distributions/regression models converged after a few iterations using a rather strict stopping criterion. In particular, we iterated between the E-step and the M-step until the relative change in the log-likelihood of each bivariate mixed NB model between two successive iterations was smaller than $10^{-12}$. We also emphasise that for both bivariate mixed NB distributions/regression models the choice of initial values for the vectors of the regression coefficients $\boldsymbol{\beta_i}$ and the parameters $\sigma_i$, with $i = 1, 2$, and the parameter $\gamma$ of the Gamma and Inverse Gaussian mixing densities respectively needed special attention because one may obtain inadmissible values if the starting values are bad. Good starting values for $\boldsymbol{\beta_i}$ and $\sigma_i$, with $i = 1, 2$, were obtained by fitting simple univariate Negative Binomial distributions/regression models. Also, a good initial value for $\gamma$, which relates to the correlation and overdispersion in the data, was feasible by equating the overdispersion of each model to the average of the observed overdispersion. Furthermore, standard errors for the parameters of the BNBGA and BNBIG distributions/regression models were obtained using the standard approach of Louis (1982). Additionally, ML estimation of the BNB and BPIG distributions/regression models was accomplished via the EM algorithms which were presented in Ghitany et al. (2012). As expected, ML estimation of the

**Table 4**
Parameter estimates and standard errors of the fitted BNB, BPIG, BNBGA and BNBIG distributions.

| BNB | | BPIG | | BNBGA | | BNBIG | |
|---|---|---|---|---|---|---|---|
| $\mu_1$ | $\mu_2$ | $\mu_1$ | $\mu_2$ | $\mu_1$ | $\mu_2$ | $\mu_1$ | $\mu_2$ |
| 0.0954 | 0.0618 | 0.0954 | 0.0618 | 0.0954 | 0.0618 | 0.0954 | 0.0618 |
| (0.0542) | (0.0639) | (0.0535) | (0.0633) | (0.0532) | (0.0634) | (0.0526) | (0.0629) |
| $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ |
| - | - | - | - | 0.7774 | 11.5401 | 0.785 | 11.6814 |
| - | - | - | - | (0.0752) | (3.4877) | (0.0744) | (3.4871) |
| $\gamma$ | | $\gamma$ | | $\gamma$ | | $\gamma$ | |
| 0.2612 | | 0.4866 | | 0.3523 | | 0.6028 | |
| (0.1037) | | (0.0554) | | (0.0364) | | (0.0671) | |

**Table 5**
Parameter estimate and standard errors of the fitted BNB, BPIG, BNBGA and BNBIG regression models.

| Variable | BNB | | BPIG | | BNBGA | | BNBIG | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. $\boldsymbol{\beta_1}$ | Coeff. $\boldsymbol{\beta_2}$ | Coeff. $\boldsymbol{\beta_1}$ | Coeff. $\boldsymbol{\beta_2}$ | Coeff. $\boldsymbol{\beta_1}$ | Coeff. $\boldsymbol{\beta_2}$ | Coeff. $\boldsymbol{\beta_1}$ | Coeff. $\boldsymbol{\beta_2}$ |
| Intercept | -2.3916 | -2.9249 | -2.3815 | -2.9148 | -2.3750 | -2.9264 | -2.3651 | -2.9163 |
| | (0.0998) | (0.1134) | (0.1132) | (0.1141) | (0.0852) | (0.1106) | (0.0966) | (0.1113) |
| v1 C2 | 0.0529 | 0.1511 | 0.0422 | 0.1407 | 0.0381 | 0.1563 | 0.0275 | 0.1459 |
| | (0.0243) | (0.0718) | (0.0119) | (0.0587) | (0.0178) | (0.0692) | (0.0087) | (0.0566) |
| v1 C3 | 0.1543 | 0.1760 | 0.1435 | 0.1645 | 0.1444 | 0.1813 | 0.1337 | 0.1698 |
| | (0.0775) | (0.0892) | (0.0652) | (0.0791) | (0.0737) | (0.0868) | (0.0620) | (0.0770) |
| v2 C2 | 0.0403 | 0.1733 | 0.0618 | 0.1944 | 0.0523 | 0.1597 | 0.0739 | 0.1808 |
| | (0.0120) | (0.0672) | (0.0273) | (0.0858) | (0.0039) | (0.0510) | (0.0088) | (0.0651) |
| v3 C2 | -0.1232 | -0.0216 | -0.1390 | -0.0376 | -0.1377 | -0.0151 | -0.1537 | -0.0311 |
| | (0.0526) | (0.0080) | (0.0557) | (0.0056) | (0.0467) | (0.0170) | (0.0494) | (0.0119) |
| v3 C3 | 0.1733 | 0.1686 | 0.1683 | 0.1636 | 0.1749 | 0.1660 | 0.1700 | 0.1610 |
| | (0.0759) | (0.0760) | (0.0686) | (0.0681) | (0.0694) | (0.0737) | (0.0627) | (0.0660) |
| | $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ | $\sigma_1$ | $\sigma_2$ |
| | - | - | - | - | 0.7756 | 11.0534 | 0.7834 | 11.1952 |
| | - | - | - | - | (0.0726) | (3.4848) | (0.0721) | (3.4842) |
| | $\gamma$ | | $\gamma$ | | $\gamma$ | | $\gamma$ | |
| | 0.2643 | | 0.4892 | | 0.3541 | | 0.6068 | |
| | (0.1011) | | (0.0538) | | (0.0337) | | (0.0644) | |

BNBGA and BNBIG distributions/regression models, which do not have closed form densities, was more chronologically demanding than that of the BNB and BPIG models both for the cases without and with covariate information. However, as it will become clear in what follows, the trade-off between CPU time requirements and the efficiency of the BNBGA and BNBIG distributions/regressions model for approximating claim frequencies of MTPL bodily injury and property damage claims in our sample and for deriving a posteriori, or Bonus-Malus, ratemaking mechanisms in a bivariate context is sifted in favour of the latter two.

The ML estimates of the parameters the corresponding standard errors in parentheses of the BNB, BPIG, BNBGA and BNBIG models[6] are reported in Table 4 for the case without covariates[7] and in Table 5 for the case with covariates. As we observe from Table 5 the values of the estimated regression coefficients of the variables v1, v2 and v3 are almost identical across all four bivariate claim frequency models Therefore, the a priori premiums per claim type resulting from these models would be almost identical when the net premium principle is used. However, as we are going to see in subsection 5.3, due to the discrepancies in the values of parameters $\sigma_1$ and $\sigma_2$ and $\gamma$, the a posteriori, or Bonus-Malus, premium rates resulting from the BNB, BPIG, BNBGA and BNBIG models will differ with this difference being more noticeable in the case of the premiums determined by latter model.

Finally, it should be noted that regarding larger MTPL bodilly injury and property damage count data with more covariate information a drawback of the likelihood-based estimation of the BNB, BPIG, BNBGA and BNBIG regression models is that we may collect a large amount of policyholders' information where not all is useful. Moreover, including too many explanatory variables in a regression model may result in overfitting and poor out-of-sample performance. Therefore, in such cases it is important to have a variable selection methodology for these regression models. To incorporate variable selection, we could maximize a penalized log-likelihood function using penalty functions such as the Lasso regression, which is a popular method that performs variable selection, see Chapter 6 in James et al. (2013). In what follows, we briefly describe the Lasso shrinkage method but because the main focus of this paper is to demonstrate how to construct a posteriori, or Bonus-Malus, ratemaking mechanisms based on the proposed mixed NB model class, and since our MTPL data set is small and does not contain many features, we will not further investigate variable selection in this numerical study. Nevertheless, we consider this topic as an important future research area.

[6] The parameters of the models are statistically significant at a 5% threshold.

[7] Note that the mean parameters of the BNB, BPIG, BNBGA and BNBIG distributions are denoted by $\mu_1$ and $\mu_2$ and the dispersion parameter is denoted by $\phi$.

**Table 6**
Models comparison based on Global Deviance, AIC and SBC.

| Distributions | | | | Regression Models | | | | |
|---|---|---|---|---|---|---|---|---|
| Model | *df* | AIC | SBC | Model | *df* | DEV | AIC | SBC |
| BNB | 3 | 5615 | 5635 | BNB | 13 | 4520 | 4546 | 4631 |
| BPIG | 3 | 5541 | 5561 | BPIG | 13 | 4433 | 4459 | 4544 |
| BNBGA | 5 | 5497 | 5530 | BNBGA | 15 | 4367 | 4397 | 4495 |
| BNBIG | 5 | 5432 | 5465 | BNBIG | 15 | 4334 | 4364 | 4462 |

Let $l$ be the log-likelihood function for a given bivariate mixed Poisson or mixed NB model. The new function to be maximised under the Lasso regression approach is given below:

$$\underset{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\in\mathbb{R}^{2\times n}}{\arg\max}\left\{l-\omega\sum_{i=1}^{2}|\beta_i|\right\}, \tag{50}$$

where $\boldsymbol{\beta_i}$ are the vectors of regression coefficients per claim type $i=1,2$ and where $\omega$ is the shrinkage parameter.

As $i$ takes values 1 and 2, the shrinkage considers regression coefficients for both the MTPL bodily injury and property damage claim count response variables $k_{1,j}$ and $k_{2,j}$. Thus, the above equations specify the shrinkage model in the context of a bivariate model. Note that, we have excluded the intercept coefficient since it constitutes a measure of the mean value of $k_{1,j}$ and $k_{2,j}$ when other explanatory variables are set to zero. Similarly, every other parameter of the model is excluded from shrinkage in order to focus on the shrinkage penalty for the estimated association of each explanatory variable with each response variable. Note also that, when $\omega$ is equal to zero the Lasso regression reduces to the full model. Furthermore, a larger $\omega$ gives greater emphasis to model simplicity, which might force coefficient values to deviate the MLE counterparts, resulting in loss of in-sample goodness-of-fit. This simplicity, however, might improve out-of-sample performance. Finally, it is clear that different values of $\omega$ will lead to different optimisation problems and hence different estimates for the model coefficients and ultimately different out-of-sample prediction results. Therefore, we must make the optimal choice of $\omega$ based on only the sample data to achieve the best possible out-of-sample prediction accuracy. For this purpose one can use $k$-fold cross-validation, where usually $k$ is set to be 5 or 10.

### 5.2. Model comparison

In this subsection we examine the model fit of the BNB, BPIG, BNBGA and BNBIG distributions/regression models employing the Global Deviance (DEV), Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) which are classic hypothesis/specification tests.[8] The (fitted) Global Deviance is defined as

$$\mathrm{DEV}=-2\hat{l}\left(\hat{\boldsymbol{\theta}}\right), \tag{51}$$

where $\hat{l}$ is the maximum of the log-likelihood and $\hat{\boldsymbol{\theta}}$ is the estimated parameter vector of the model. Furthermore, the AIC is given by

$$\mathrm{AIC}=\mathrm{DEV}+2\times df \tag{52}$$

and the SBC is given by

$$\mathrm{SBC}=\mathrm{DEV}+\log(n)\times df, \tag{53}$$

where $df$ are the degrees of freedom, that is, the number of fitted parameters in the model and $n$ is the number of observations in the sample.

As is well known, a commonly used rule-of-thumb states that a model significantly outperforms a competitor if the difference in their log-likelihoods exceeds 5, corresponding to a difference in their AIC values of more than 10 and to a difference in their SBC values of more than 5, see Burnham and Anderson (2003) and Raftery (1995) respectively. This means here that, as can be seen from Table 6, the best fit is given by the BNBIG distribution/regression model.

At this point, it should be noted that, overall regarding MTPL bodily injury and property damage claim count data sets, such as the one we consider in this study, the bivariare mixed NB model enjoys the following advantages over the bivariate mixed Poisson model which is constructed in a similar way, i.e. by using of a shared random effect that introduces overdispersion and positive correlation between the responses, since, in this case, as we can observe from Table 3, Kendall's tau is computed at 0.17595. Firstly, the bivariate mixed NB model class may provide better fitting performances than a bivariate mixed Poisson class, as the latter is a particular case of the former. More specifically, when the dispersion parameters $\sigma_1$ and $\sigma_2$ approach $\infty$ each individual claim distribution approaches a Poisson distribution and hence the bivariate mixed Poisson model is a special case of the bivariate mixed NB model. Indeed, as we can see from Table 6, the BNBGA and BNBPIG models provide superior fitting performances than the BNB and BPIG models. Secondly, the bivariate mixed NB model is more flexible than its bivarite mixed Poisson counterpart in capturing overdispersion since the dispersion parameters $\sigma_1$ and $\sigma_2$ control the extent of overdispersion of the individual bodily injury and property damage claim count distributions.

On the other hand, when modelling different types of claims from different types of coverage other model classes, some of which we briefly describe below, are more appropriate in terms of their flexibility in modelling data sets with more complex structures than the one used herein. For instance, the MVMNB models only allow for positive correlations between claim counts but in some other

[8] Note that for other data sets with more complex features it is good to compare the fitting results of the models both in terms of in- and out-of-sample validation to investigate the quality of the proposed models, as we do for the simulated data below.

**Table 7**
The explanatory variables of the simulated data set and their description.

| Variables | C1 | C2 |
|---|---|---|
| Region (v1) | Rural | Urban |
| Car fuel (v2) | Diesel | Gasoline |
| Car brand (v3) | Class A | Class B |
| Age of the policyholder (v4) | Continuous, takes integer values between 18 and 73 | |
| Age of the car (v5) | Continuous, takes integer values between 0 and 10 | |

cases negative correlations or strong right tail correlations may be of interest as well. In such cases, copula-based models can be used to more accurately model the correlation structure of discrete count variables, see, for example, Cameron et al. (2004), Nikoloulopoulos and Karlis (2009 and 2010), Nikoloulopoulos et al. (2011) and Shi and Valdez (2014) among many others. On the other hand, when pairing a continuous copula with discrete marginal distributions identifiability issues may be encountered, see Genest and Nešlehová (2007). Moreover, the jpmf of a copula-based model with discrete margins involves both summation and integration which leads to a computationally demanding estimation procedure in a high-dimensional setting because of multivariate numerical integration, see Oh et al. (2020). As was previously mentioned, the proposed class of models is identifiable and computational issues are not encountered since, as it can be seen from Eq. (5), the jpmf of the MVMNB model involves numerical integration of a single integral for any vector size positively correlated response variables. However, for large data sets with several explanatory variables computational speed is a handicap since most of the computational costs in the case of a multidimensional response variable will come from evaluating the expectations involved at the E-Step of the EM algorithm which is not tracTable. In such cases, due to the structure of the EM algorithm for mixed NB models, the E- and M-steps can be executed in parallel across multiple threads to take advantage of the processing power available in modern-day multicore machines. Additionally, another limitation of the MVMNB model is that it does not cover underdispersion and this is definitely encountered in modeling claim counts in other insurance settings. Overall, from a model fitting perspective, even if a particular data set has highly complex characteristics, it is desirable that a model class is flexible enough to capture any distribution, dependence and regression patterns. Such a model class is the Logit-weighted Reduced Mixture of Experts (LRMoE) class which was recently proposed by Fung et al. (2019a) and can fit all underlying data sets. The LRMoE model with mixing weights, called the gating function, which depend on the covariates, can take into account both over-and-under-dispersion and positive and negative correlations of different strength and magnitude between the responses. The advantage that the LRMoE class enjoyes over other model classes can be theoretically justified due to the denseness property, see Fung et al. (2019b), which guarantees that one can always obtain a model within this class that can adequately represent the data making it suiTable for many insurance applications. Furthermore, the LRMoE model is identifiable, closed under marginalization and mathematically tracTable in terms of premium and risk measure calculations.

Therefore, it is interesting to compare the in-sample and out-of-sample performance of the MBMNB model class to that of the LRMoE model class which can capture very complicated data characteristics. Because the MTPL data set is small and contains a few risk factors, we considered a simulated bivariate count data set which is much larger and has more features so that the training data set, which will be used to fit the models, and the test set, which will be used to evaluate their predictive performance, are sufficiently large. Furthermore, we assumed that the number of claims per claim type is larger in the simulated data since, as it can be verified by Shared's two crossings theorem, see Shared (1980), the overdispersion phenomenon can be attributed to the excess of zeros and/or heavy upper tails in count data and hence it is interesting to investigate the performance of the models in such scenarios. In particular, we randomly generated a data set of size $n = 10000$ from the bivariate normal copula and two NB regressions where the two response variables $k_{1,j}$ and $k_{2,j}$ represent the number of claims.[9] Furthermore, we chose three categorical covariates v1, v2 and v3, with two categories each, that represent the region (V1: rural and urban), car fuel (V2: diesel and gasoline) and car brand class (V3: classes A and B, where A corresponds to "good" brands with less claims and vice versa for B) respectively and two continuous covariates that represent the age of the policyholder (v4) and the age of the car (v5). These explanatory variables are summarized in Table 7.

We split the data at random into a training set and a validation set containing 80% and 20%0 of the data, i.e. 8000 and 2000 observations respectively, in order evaluate the fitting results of the BNBGA and BNBIG models, which belong to the bivariate mixed NB model class, and the bivariate LRMoE model. Regarding the latter, we consider the following choices of expert functions: Erlang count (EC-LRMoE), see Fung et al. (2019a), and the Poisson (PO-LRMoE), Negative Binomial (NB-LRMoE), Gamma count (GC-LRMoE) and their zero-inflated versions which are supported by the **LRMoE** package that was developed by Tseung et al. (2020). The parameters of the LRMoE models are estimated via Expectation-Conditional-Maximization (ECM) algorithm.[10] The log-likelihood, AIC and SBC values are used to compare the fit of the models on the training data. Subsequently, once the models are fitted on the training set, their prediction performances are assessed via out-of-sample validation using the root-mean squared error (RMSE) and their log-likelihood values for the test data.

In what follows, for brevity of presentation, we consider the BNBIG and bivariate zero-inflated NB-LRMoE (ZI-NB-LRMoE) regression models which provided the best fitting results compared to other models in the same class both in terms of in-sample estimation and out-of-sample validation.

- **In-sample estimation.** The estimation results for the BNBIG regression model are provided in Table 8.
  Let us now describe in detail the implementation of the bivariate ZI-NB-LRMoE model including the choice of the number of latent classes, or components $G$ and parameter initialization.[11]

[9] A similar numerical example in which bivariate count data are generated using the Normal copula can be found in Nikoloulopoulos et al. (2011).

[10] For more details regarding the ECM algorithm the interested reader can refer to Fung et al. (2019a, 2019b). Furthermore, the ECM algorithm is implemented in the **LRMoE** package.

[11] Note that we followed a similar procedure for all the aforementioned LRMoE models which are determined by alternative to the NB expert functions, their zero-inflated versions and their combinations.

**Table 8**
Parameters estimate of the fitted BNBIG regression model for each covariate on the simulated data set.

| | BNBIG | |
|---|---|---|
| Variable | Coeff. $\boldsymbol{\beta_1}$ | Coeff. $\boldsymbol{\beta_2}$ |
| Intercept | -0.871629352 | -1.804815354 |
| v1 C2 | 0.247094683 | 0.260463680 |
| v2 C2 | -0.082661311 | -0.024245947 |
| v3 C2 | 0.165582127 | 0.342735596 |
| v4 | -0.005138798 | -0.008163438 |
| v5 | 0.009070358 | 0.037774363 |
| | $\sigma_1$ | $\sigma_2$ |
| | 0.04114404 | 0.01803028 |
| | $\gamma$ | |
| | 6.628101 | |
| Log-likelihood$= -6518.336$ | | |
| AIC $= 13067$ | | |
| SBC $= 13171$ | | |

**Table 9**
Specification criteria to choose the number of components $G$ for the bivariate ZI-NB-LRMoE model.

| | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|
| Log-likelihood | -7830.742 | -6225.308 | -6217.729 | -6220.743 |
| AIC | 15667.97 | 12478.48 | **12473.15** | 12495.43 |
| SBC | 15797.75 | **12688.09** | 12766.62 | 12872.74 |

– Parameter initialization. For a given number of components $G$ of the bivariate ZI-NB-LRMoE model, which we explain how it can be determined below, a good initialization of parameters will lead to a faster convergence of the ECM algorithm. Following the setup of Gui et al. (2018) and Fung et al. (2019a), we used an initialization procedure which involves k-means clustering and clusterized method of moments (CMM) initializations that provide meaningful starting values for the parameters.[12]
– Number of components $G$. The number of components $G$ of the bivariate ZI-NB-LRMoE model which optimizes the fitting result can be determined by the AIC or the SBC information criteria.[13] As it can be observed from Table 9, three and four components are detected via the SBC and AIC criteria respectively. As is well known, see Kuha (2004), the SBC is more suiTable for detecting the true model. On the other hand, since the true model is never known for any real data set, as is mentioned in Fung et al. (2019a), the AIC is preferable from a practical standpoint as it can identify the model which can generalize better to future data, which is unseen policies in this case.
The fitting results for the four component $(G = 4)$ bivariate ZI-NB-LRMoE model which is given by Eqs (54), (55) and (56) are presented Tables 10 and 11. Note that $\boldsymbol{\alpha_{g=4}} = (0, 0, \dots, 0)^T$ in Table 10 ensures that the model is identifiable.

$$P\left(k_{1,j}, k_{2,j} | \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}\right) = \prod_{i=1}^{2} \sum_{g=1}^{4} \frac{\exp\left(\boldsymbol{x_{i,j}^T \alpha_{i,g}}\right)}{\sum_{g=1}^{4} \exp\left(\boldsymbol{x_{i,j}^T \alpha_{i,g}}\right)} h_{i,g}(k_{i,j}; \delta_{i,g}, \psi_{i,g}), \tag{54}$$

where $h_{i,g}(k_{i,j}; \delta_{i,g}, \psi_{i,g})$ is the expert function given by

$$h_{i,g}(k_{i,j}; \delta_{i,g}, \psi_{i,g}) = \delta_{i,g} \mathbb{1}\{k_{i,j} = 0\} + (1 - \delta_{i,g}) f_{i,g}(k_{i,j}; \psi_{i,g}) \tag{55}$$

with $0 < \delta_{i,g} < 1$, where $\delta_{i,g}$ is a probability mass at zero and $f_{i,g}(k_{i,j}; \psi_{i,g})$ is the density of some commonly used distribution with parameters $\psi_{i,g}$, which, in the case of the bivariate ZI-NB-LRMoE model, is a Negative Binomial with parameters $\psi_{i,g} = (n_{i,g}, p_{i,g})$ and pmf:

$$f_{i,g}(k_{i,j}; \psi_{i,g}) = \binom{k_{i,j} + n_{i,g} - 1}{n_{i,g} - 1} p_{i,g}^{n_{i,g}} (1 - p_{i,g})^{k_{i,j}}. \tag{56}$$

- **Out-of-sample validation.** The RMSE and log-likelihood values for the test data are given Table 12 in the case of the BNBIG and the four component bivariate zero-inflated NB-LRMoE regression models respectively.

Overall, as we can see from Tables 8 and 9 and Table 12, the four component bivariate zero-inflated NB-LRMoE regression model significantly outperforms the other models in terms of in-sample estimation and out-of-sample validation results. An interesting future

[12] Note that the **LRMoE** package contains the CMMFrequency initialization function which returns a list in which each element contains the cluster proportion, zero inflation and parameter initialization for the component distribution.
[13] Note also that as is mentioned in Fung et al. (2019b), due to the denseness property, five to thirty six components are sufficient for data with very complicated features, including over- and under-dispersion, positive and negative correlation structures across business lines and non-linear regression patterns with covariate interactions.

**Table 10**
Fitted logit regression coefficients $\hat{\boldsymbol{\alpha}}_{\boldsymbol{i},\boldsymbol{g}}$ for the four components bivariate ZI-NB-LRMoE model.

| | Intercept | v1 | v2 | v3 | v4 | v5 |
|---|---|---|---|---|---|---|
| $g=1$ | 0.6673705 | 0.5238889 | 0.1765627 | -0.1067959 | -0.021932367 | 0.04898917 |
| $g=2$ | -0.6531821 | 0.4899900 | -0.6944596 | 0.8423504 | 0.001824091 | 0.05927358 |
| $g=3$ | -0.5984935 | -0.9286111 | 0.2280733 | -1.2469809 | 0.025443073 | -0.00244666 |
| $g=4$ | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 11**
Component distributions for the four components bivariate ZI-NB-LRMoE model.

| | $i=1$ | | | $i=2$ | | |
|---|---|---|---|---|---|---|
| | comp.dist | $\hat{\delta}_{i,g}$ | $\hat{\psi}_{i,g}$ | comp.dist | $\hat{\delta}_{i,g}$ | $\hat{\psi}_{i,g}$ |
| $g=1$ | ZI-NB | 0.8467553 | $\hat{n}_{i,g}=8,\ \hat{p}_{i,g}=0.6449669$ | ZI-NB | 0.9448393 | $\hat{n}_{i,g}=2,\ \hat{p}_{i,g}=0.2802326$ |
| $g=2$ | ZI-NB | 0.9147926 | $\hat{n}_{i,g}=5,\ \hat{p}_{i,g}=0.5144303$ | ZI-NB | 0.9178029 | $\hat{n}_{i,g}=3,\ \hat{p}_{i,g}=0.4417814$ |
| $g=3$ | ZI-NB | 0.934662 | $\hat{n}_{i,g}=3,\ \hat{p}_{i,g}=0.5557306$ | ZI-NB | 0.956946 | $\hat{n}_{i,g}=2,\ \hat{p}_{i,g}=0.7106334$ |
| $g=4$ | ZI-NB | 0.9279882 | $\hat{n}_{i,g}=5,\ \hat{p}_{i,g}=0.5354817$ | ZI-NB | 0.9883822 | $\hat{n}_{i,g}=2,\ \hat{p}_{i,g}=0.3616542$ |

**Table 12**
Out-of-sample validation results.

| | BNBIG | 4C Biv. ZI-NB LRMoE |
|---|---|---|
| RMSE | 3.1838 | **3.1621** |
| Log-Likelihood | -1663.6542 | **-1632.2031** |

research direction would be to combine both approaches and consider an LRMoE model with mixed NB expert functions. Note that in this case, the mixed NB expert function will have a closed form density since it will be parameterized in terms of the probability rather than the mean parameter since in the LRMoE model the gating functions depend on covariates.

In what follows we will restrict our attention to the BNB, BPIG, BNBGA and BNBIG distributions/regression models that can be used within the Bayesian paradigm for deriving generalized Bonus-Malus ratemaking schemes by updating the posterior mean which (i) are functions of the years that the policyholder is in the portfolio and the risk factors which influence the number of their MTPL bodily injury and property damage claims which may not necessarily be exactly the same and (ii) can take into account unobserved heterogeneity, such as reaction times and aggressive driving behaviour, etc., which to a great extent reveal the riskiness of the policyholder since the occurrence of both claim types is under the control of the same individual.

### 5.3. *Calculation of the a posteriori premiums*

In this subsection, we examine the response of the BNBGA and BNBIG distribution/regression models to claim experience and we compare it to those of the two bivariate mixed Poisson distributions/regression models which were presented in Section 4.

Consider the policyholder $j$, $j=1,\ldots n$, with number of bodily injury and property damage claims $k_{1,j,l}$ and $k_{2,j,l}$ respectively, for the year of coverage $l$, with $l=1,\ldots,t$. Assume that the cumulative number of claims per type $i=1,2$ for all the years that the individual $j$ has been registered with the insurance company is denoted as $K_{i,j}=\sum_{l=1}^{t}k_{i,j,l}$. Also, let the unobserved Gamma and Inverse Gaussian random variables for the BNBGA and BNBIG respectively take into account individual characteristics. On the path towards actuarial relevance, the Bayesian view is taken[14] to compute the posterior distribution of $\lambda_{j,t+1}$ for the period $t+1$ given the observations of the reported accidents in the preceding $t$ periods and observable characteristics in the preceding $t+1$ periods and the current period. In particular, the posterior distribution of $\lambda_{j,t+1}$ can be derived as follows:

$$f\left(\lambda_{j,t+1}|k_{1,j,1},\ldots,k_{1,j,t};k_{2,j,1},\ldots,k_{2,j,t};\boldsymbol{x_{1,j,1}},\ldots,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},\ldots,\boldsymbol{x_{2,j,t+1}}\right)$$
$$=\frac{\prod_{l=1}^{t}P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_{j,t+1}\right)}{\int_0^{\infty}P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)f\left(\lambda_{j,t+1}\right)d\lambda_{j,t+1}},\qquad(57)$$

where $P\left(k_{1,j},k_{2,j}|\boldsymbol{x_{1,j}},\boldsymbol{x_{2,j}},\lambda_j\right)$ is the bivariate Poisson distribution in the case of the BNB and BPIG models, while it takes the form of the bivariate Negative Binomial in the case of the BNBGA and BNBIG models and where $f(\lambda_{j,t+1})$ is the pdf of the Gamma distribution in the case of the BNB and BNBGA models and the pdf of the Inverse Gaussian distribution in the case of the BPIG and BNBIG models respectively.

Using the net premium principle and the quadratic loss function, one can easily see that the optimal estimator of $\lambda_{j,t+1}$ is the mean of the posterior distribution in Eq. (57), given by

[14] For more details regarding the Bayesian interpretation of Bonus-Malus systems the interested reader can refer, for instance, to Dionne and Vanasse (1992) and Lemaire (1995).

**Table 13**
Comparison of the a posteriori, or Bonus-Malus, premium rates for $t=1,2,3$, bivariate claim frequency distributions.

| | $t=1$ BNB distribution | | | $t=1$ BPIG distribution | | | $t=1$ BNBGA distribution | | | $t=1$ BNBIG distribution | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 72.69 | 350.98 | 629.27 | 79.80 | 256.06 | 553.66 | 77.43 | 305.10 | 541.50 | 82.97 | 233.17 | 493.12 |
| 1 | 350.98 | 629.27 | 907.56 | 256.06 | 553.66 | 892.83 | 266.53 | 478.66 | 700.26 | 200.62 | 419.84 | 696.16 |
| 2 | 629.27 | 907.56 | 1185.85 | 553.66 | 892.83 | 1240.99 | 425.88 | 628.17 | 839.80 | 362.47 | 603.31 | 876.47 |
| | $t=2$ BNB distribution | | | $t=2$ BPIG distribution | | | $t=2$ BNBGA distribution | | | $t=2$ BNBIG distribution | | |
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 57.10 | 275.69 | 494.28 | 68.35 | 197.65 | 411.53 | 62.80 | 248.73 | 444.00 | 72.17 | 186.15 | 380.62 |
| 1 | 275.69 | 494.28 | 712.87 | 197.65 | 411.53 | 657.85 | 221.14 | 397.85 | 583.64 | 165.03 | 332.97 | 546.14 |
| 2 | 494.28 | 712.87 | 931.46 | 411.53 | 657.85 | 912.20 | 358.66 | 529.14 | 708.25 | 294.61 | 483.14 | 697.83 |
| | $t=3$ BNB distribution | | | $t=3$ BPIG distribution | | | $t=3$ BNBGA distribution | | | $t=3$ BNBIG distribution | | |
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 47.01 | 226.99 | 406.97 | 60.73 | 162.83 | 328.94 | 52.66 | 208.89 | 373.84 | 64.61 | 155.89 | 308.90 |
| 1 | 226.99 | 406.97 | 586.96 | 162.83 | 328.94 | 521.70 | 188.37 | 338.89 | 497.60 | 141.10 | 275.81 | 447.09 |
| 2 | 406.97 | 586.96 | 766.94 | 328.94 | 521.70 | 721.75 | 308.87 | 455.33 | 609.47 | 248.54 | 402.02 | 576.93 |

**Table 14**
Results of the fitted BNB, BPIG, BNBGA and BNBIG regression models for each risk class profile.

| Regression model | Profile | $\boldsymbol{\varepsilon_{1,j}}$ | $\mathrm{Var}\left(k_{1,j}\lvert\boldsymbol{x_{1,j}}\right)$ | $\boldsymbol{\varepsilon_{2,j}}$ | $\mathrm{Var}\left(k_{2,j}\lvert\boldsymbol{x_{2,j}}\right)$ |
|---|---|---|---|---|---|
| BNB | Best | 0.091483 | 0.123149 | 0.053670 | 0.064569 |
| | Average | 0.088779 | 0.118601 | 0.072650 | 0.092620 |
| | Worst | 0.132166 | 0.198256 | 0.090085 | 0.120790 |
| BPIG | Best | 0.092412 | 0.128097 | 0.054215 | 0.066497 |
| | Average | 0.089233 | 0.122506 | 0.072100 | 0.095267 |
| | Worst | 0.134270 | 0.209604 | 0.091419 | 0.126341 |
| BNBGA | Best | 0.093014 | 0.160104 | 0.053590 | 0.062693 |
| | Average | 0.088717 | 0.149751 | 0.072404 | 0.089022 |
| | Worst | 0.134876 | 0.275942 | 0.088975 | 0.114071 |
| BNBIG | Best | 0.093940 | 0.159764 | 0.054134 | 0.063065 |
| | Average | 0.089153 | 0.148440 | 0.072752 | 0.088883 |
| | Worst | 0.137038 | 0.277115 | 0.090293 | 0.115140 |

$$\begin{aligned}&\mathbb{E}\left(\lambda_{j,t+1}\lvert k_{1,j,1},\ldots,k_{1,j,t};k_{2,j,1},\ldots,k_{2,j,t};\boldsymbol{x_{1,j,1}},\ldots,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},\ldots,\boldsymbol{x_{2,j,t+1}}\right)\\&=\int_0^\infty \lambda_{j,t+1} f\left(\lambda_{j,t+1}\lvert k_{1,j,1},\ldots,k_{1,j,t};k_{2,j,1},\ldots,k_{2,j,t};\boldsymbol{x_{1,j,1}},\ldots,\boldsymbol{x_{1,j,t+1}};\right.\\&\left.\qquad\boldsymbol{x_{2,j,1}},\ldots,\boldsymbol{x_{2,j,t+1}}\right)d\lambda_{j,t+1}.\end{aligned}\tag{58}$$

- In the case of the BNB model, Eq. (57) is a Gamma distribution with parameters $\gamma+\sum_{i=1}^{2}k_{i,j}$ and $\gamma+\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}$, and hence Eq. (58) takes the form:

$$\begin{aligned}&\mathbb{E}\left(\lambda_{j,t+1}\lvert k_{1,j,1},\ldots,k_{1,j,t};k_{2,j,1},\ldots,k_{2,j,t};\boldsymbol{x_{1,j,1}},\ldots,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},\ldots,\boldsymbol{x_{2,j,t+1}}\right)\\&=\frac{\gamma+\sum_{i=1}^{2}k_{i,j}}{\gamma+\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}}.\end{aligned}\tag{59}$$

- In the case of the BPIG model, Eq. (57) is a Generalized Inverse Gaussian (GIG) distribution with parameters $\sqrt{\gamma^2+2\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}}$, $\gamma$ and $\sum_{i=1}^{2}k_{i,j}-\frac{1}{2}$ and thus Eq. (58) is given by:

$$\begin{aligned}&\mathbb{E}\left(\lambda_{j,t+1}\lvert k_{1,j,1},\ldots,k_{1,j,t};k_{2,j,1},\ldots,k_{2,j,t};\boldsymbol{x_{1,j,1}},\ldots,\boldsymbol{x_{1,j,t+1}};\boldsymbol{x_{2,j,1}},\ldots,\boldsymbol{x_{2,j,t+1}}\right)\\&=\frac{\gamma K_{\sum_{i=1}^{2}k_{i,j}+\frac{1}{2}}\left(\gamma\sqrt{\gamma^2+2\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}}\right)}{\sqrt{\gamma^2+2\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}}K_{\sum_{i=1}^{2}k_{i,j}-\frac{1}{2}}\left(\gamma\sqrt{\gamma^2+2\sum_{l=1}^{t}\sum_{i=1}^{2}\boldsymbol{\varepsilon_{i,j}}}\right)}.\end{aligned}\tag{60}$$

**Table 15**
A Posteriori, or Bonus-Malus, premium rates for $t = 1, 2, 3$ for the three risk profiles, bivariate claim frequency regression models.

| | $t=1$ BNB regression model Best profile | | | $t=1$ BPIG regression model Best profile | | | $t=1$ BNBGA regression model Best profile | | | $t=1$ BNBIG regression model Best profile | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 73.83 | 353.16 | 632.49 | 80.31 | 260.95 | 566.63 | 78.46 | 308.15 | 546.87 | 83.57 | 236.80 | 503.08 |
| 1 | 353.16 | 632.49 | 911.83 | 260.95 | 566.63 | 914.58 | 268.86 | 482.85 | 706.60 | 203.03 | 426.77 | 709.44 |
| 2 | 632.49 | 911.83 | 1191.16 | 566.63 | 914.58 | 1271.53 | 429.15 | 633.22 | 846.90 | 367.29 | 612.97 | 892.18 |
| | $t=2$ BNB regression model Average profile | | | $t=2$ BPIG regression model Average profile | | | $t=2$ BNBGA regression model Average profile | | | $t=2$ BNBIG regression model Average profile | | |
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 56.87 | 272.05 | 487.22 | 68.18 | 194.75 | 403.56 | 62.29 | 244.24 | 433.45 | 71.95 | 181.71 | 365.84 |
| 1 | 272.05 | 487.22 | 702.40 | 194.75 | 403.56 | 644.34 | 219.58 | 392.44 | 572.52 | 163.14 | 324.61 | 526.88 |
| 2 | 487.22 | 702.40 | 917.57 | 403.56 | 644.34 | 893.16 | 357.13 | 523.71 | 697.26 | 290.75 | 471.92 | 675.62 |
| | $t=3$ BNB regression model Worst profile | | | $t=3$ BPIG regression model Worst profile | | | $t=3$ BNBGA regression model Worst profile | | | $t=3$ BNBIG regression model Worst profile | | |
| $k_{1,j}/k_{2,j}$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| 0 | 42.19 | 201.81 | 361.43 | 58.12 | 141.20 | 273.16 | 46.51 | 185.29 | 333.70 | 60.81 | 135.37 | 258.38 |
| 1 | 201.81 | 361.43 | 521.05 | 141.20 | 273.16 | 427.77 | 165.14 | 299.20 | 442.14 | 122.04 | 229.40 | 367.95 |
| 2 | 361.43 | 521.05 | 680.67 | 273.16 | 427.77 | 589.46 | 269.98 | 400.76 | 539.76 | 205.75 | 328.89 | 471.60 |

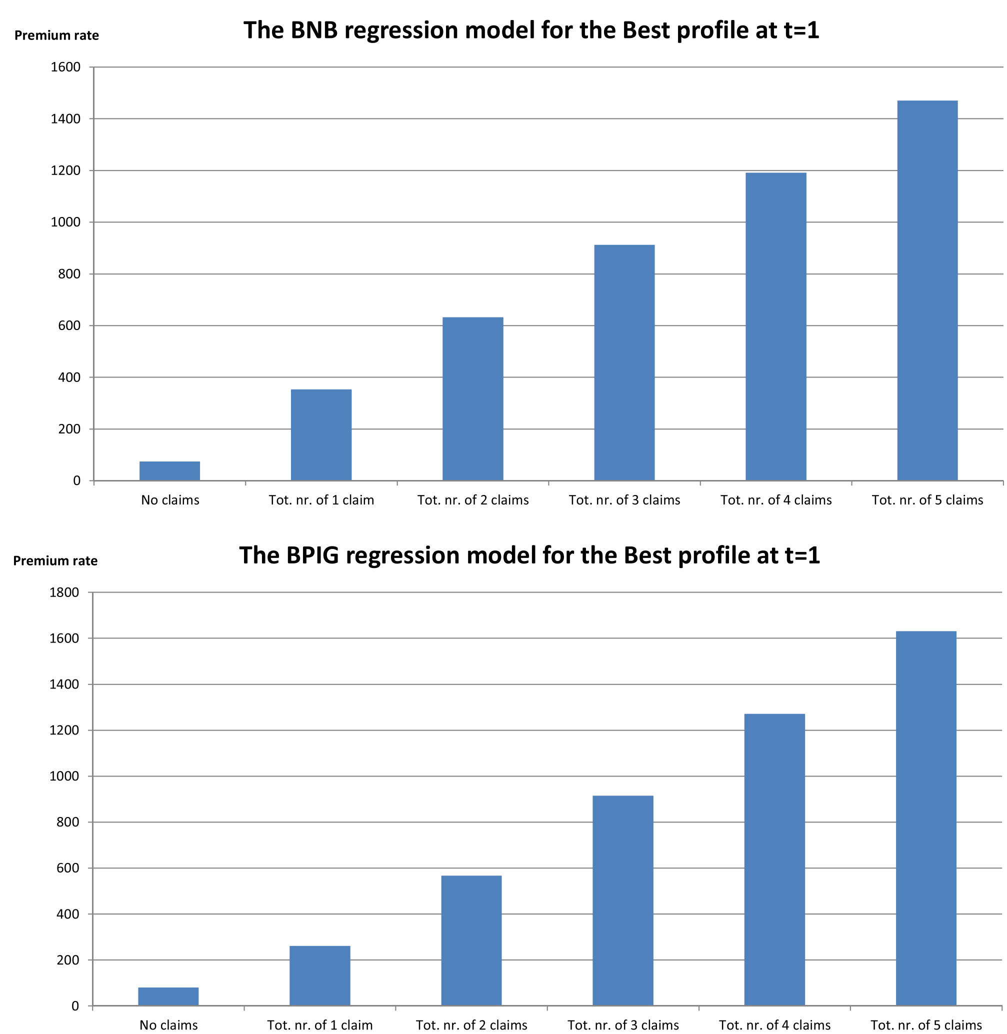



**Fig. 1.** Premium rates for the Best profile at $t = 1$, bivariate mixed Poisson regression models.

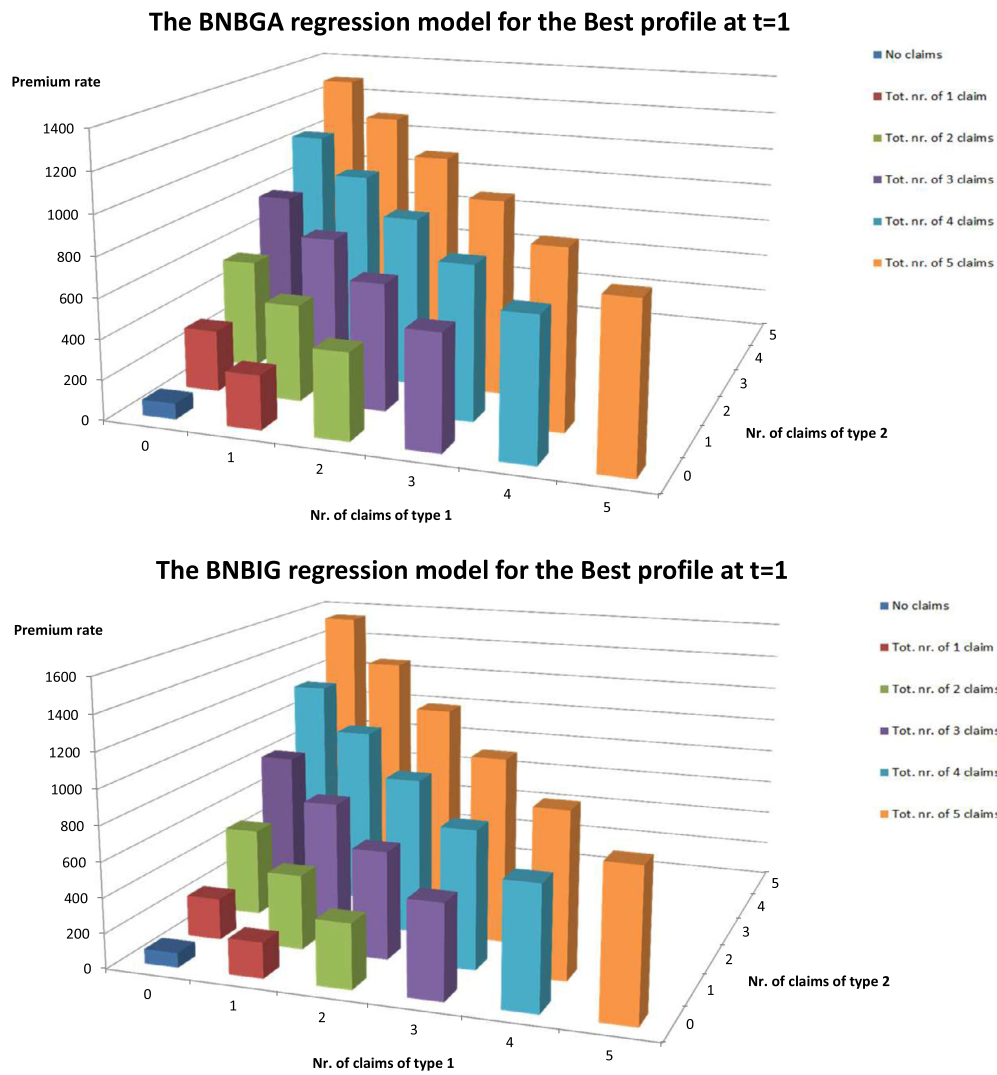



**Fig. 2.** Premium rates for the Best profile at $t = 1$, bivariate mixed Negative Binomial regression models.

- In the case of the BNBGA and BNBIG models, the expectation in Eq. (58) cannot be computed in closed form. However, it can be computed based on either numerical integration or a Monte Carlo approach since both schemes do not rely on the knowledge of the pdf given by Eq. (57).

Following the aforementioned methodology, we calculate the Bonus-Malus premiums resulting from the BNBGA and BNBIG models and we compare them to those derived by the BNB and BPIG models based only on the number of individual bodily injury and property damage claims, i.e. the a posteriori criteria, and based on the characteristics of the policyholders and their cars, i.e. the a priori criteria. The premium rates will be divided by the premium when $t = 0$, i.e. we calculate the relative premiums, since we are interested in the differences between various classes and the results are presented so that the premium for a new policyholder is 100.

Firstly, Table 13 depicts comparable relative premiums for the BNB, BPIG, BNBGA and BNBIG distributions, assuming that the number of claims $k_{1,j}$ and $k_{2,j}$ ranges from 0 to 2 for each claim type and the age of the policy is $t = 1$, $t = 2$ and $t = 3$ years.

Secondly, when both criteria are considered, we examine three risk class profiles that can be classified as Best, Average and Worst according to the mean claim frequencies $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$, with $j = 1, \ldots, n$, based on the same set of explanatory variables per claim type $i = 1, 2$. Specifically, the Best, Average and Worst profiles, for our data, are determined as such based on category C1 for all three explanatory variables v1, v2 and v3 in the case of the first, category C2 for v1, v2 and v3 in the case of the second, and category C3 for v1 and v3 and C2 for v2 in the case of the third. The results for all three profiles per claim type are presented in Table 14 in the case of the BNB, BPIG, BNBGA and BNBIG models respectively.

We observe from Table 14 that, as expected, for all three risk profiles small discrepancies lie in the mean values $\boldsymbol{\varepsilon_{1,j}}$ and $\boldsymbol{\varepsilon_{2,j}}$ in the case of the BNB, BPIG, BNBGA and BNBIG regression models respectively. However, when the a posteriori correction will be calculated, we will see that compared to the relative Bonus-Malus premiums provided by the two bivariate mixed Poisson models, the premiums derived from the BNBGA and BNBIG models will be much less extreme for policyholders with some bodily injury and property damage claim experience. This can be clearly justified since given the estimates of $\sigma_1$ and $\sigma_2$ of the BNBGA and BNBIG models, which also appear in the marginal Negative Binomial distributions, see Eq. (1), we can assess the extent of marginal overdispersion for the bodily injury and property damage claim distributions of an individual policyholder with any given mean frequency rates per claim type. Therefore, this situation affects the calculation of the Bonus-Malus premium rates. Table 15 shows some premiums for the three risk profiles during the

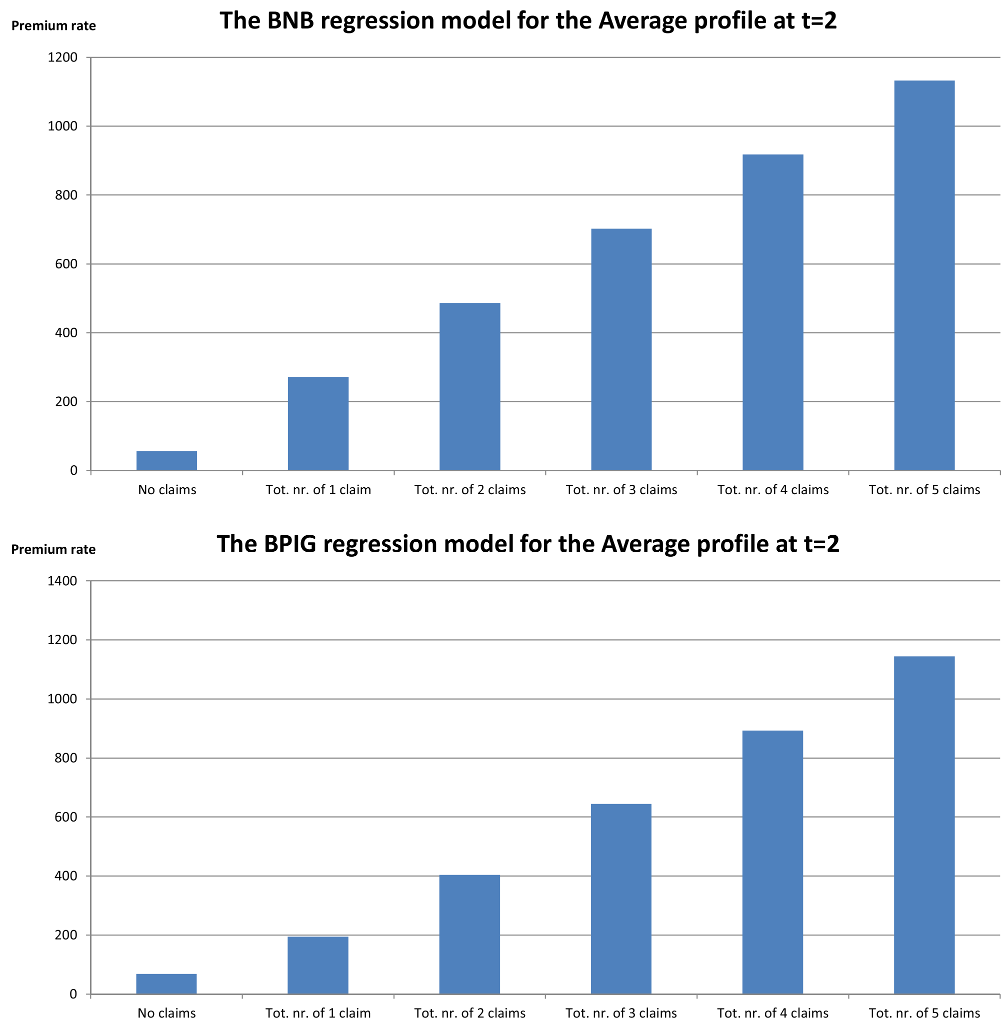



**Fig. 3.** Premium rates for the Average profile at $t = 2$, bivariate mixed Poisson regression models.

years $t = 1$, $t = 2$ and $t = 3$ respectively. Such Table can provide a more complete picture to the actuary than Table 13, where only the a posteriori criteria were considered, as they include all available information on the level of riskiness of the individual, as assessed by the insurance company.

Overall, from Tables[15] 13 and 15, we see that the BNB, BPIG, BNBGA and BNBIG distributions/regression models result in a noticeable decrement in the premiums that must be paid by the policyholder $j$ when it has a claim free year for both types of claims $i = 1, 2$, whereas if it has one or more claims of type $i = 1, 2$ the premium rates increase, hence resulting in bonus or malus in the former and latter case respectively. Furthermore, as was previously mentioned, we observe that two bivariate mixed Negative Binomial distributions/regression models return lower premiums for individuals with some claim history per claim type $i = 1, 2$ than the two bivariate mixed Poisson distributions/regression models. For example, under the BNBGA and BNBIG distributions, for a policyholder who had $k_{1,j} = 2$ and $k_{2,j} = 2$ claims, Table 13 shows premiums of only 839.80 and 876.47 respectively for the year of coverage $t = 1$. Meanwhile, for the same number of claims per claim type $i = 1, 2$, we observe that the BNB and the BPIG distributions result in higher premiums of 1185.85 and 1240.99 respectively. Also, similar discrepancies are observed if we incorporate the a priori information from Table 14. For example, still for the case when $k_{1,j} = 2$ and $k_{2,j} = 2$, according to Table 15, an individual with the Best profile is expected to pay a premium of 846.90 and 892.18 under the BNBGA and BNBIG models respectively as opposed to the higher premiums of 1191.16 and 1271.53 under the BNB and the BPIG regression models respectively for the year of coverage $t = 1$. This characteristic of the BNBGA and BNBIG models can be explained by the fact that these models are constructed by starting with two Negative Binomial models which assume that the individual bodily injury and property damage claim experience will be overdispersed as opposed to the two Poisson models in the BNB and BPIG models. The overdispersion is larger for policyholders with larger mean claim rates per claim type. Therefore, extreme individual bodily injury and property damage claim counts are more likely under the bivariate Negative Binomial mixtures resulting in more moderate relative premiums than under those models based on the bivariate Poisson mixtures.[16]

The second noticeable difference between the bivariate mixed Negative Binomial distributions/regression models and the two bivariate mixed Poisson distributions/regression models in the calculation of Bonus-Malus premiums is that the latter can only take into account

[15] Note that the symmetry of each Table for the bivariate mixed Poisson models is a logical consequence of the common random effects assumption, whereas the premiums are distinguishable per claim type under the BNBGA and BNBIG models which is due to its quantiple Poisson-Gamma-Poisson-Gamma-mixture decomposition.

[16] Similar findings were reported by Shengwang et al. (1999) and Gómez-Déniz et al. (2008) and Tzougas et al. (2019).

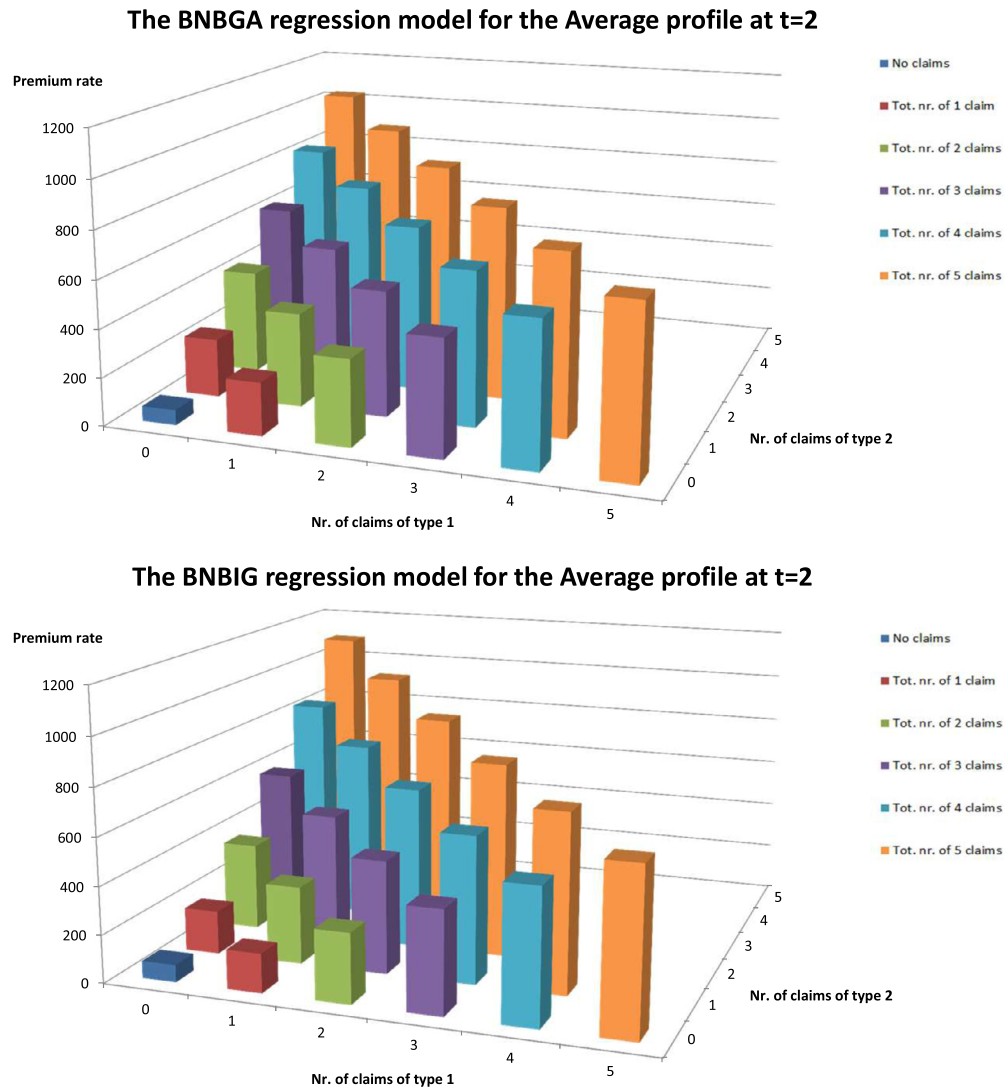
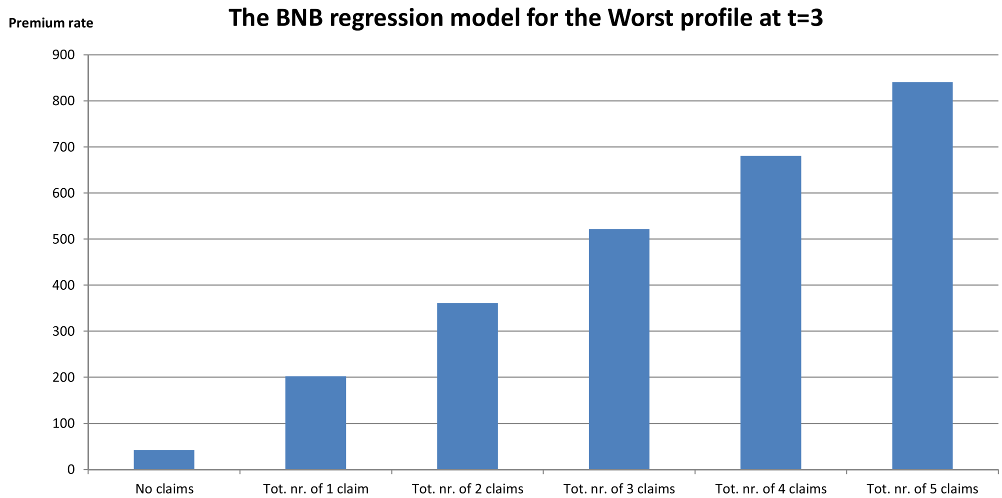



**Fig. 4.** Premium rates for the Average profile at $t = 2$, bivariate mixed Negative Binomial regression models.

the policyholder's total number of claims, which is computed by aggregating the bodily injury and property damage claims, but are unable to distinguish between the two types of claims. For instance, for the case without covariates, we see from Table 13 that for $\sum_{i=1}^{2} k_{i,j} = 2$ the policyholder has to pay premiums of 629.27 and 553.66 for $t = 1$ under the BNB and BPIG distributions respectively, regardless of the exact frequencies of bodily injury and property damage claims. Similarly, for the case with covariates, we see from Table 15 that for $\sum_{i=1}^{2} k_{i,j} = 2$, an individual with the Best profile has to pay premiums of 632.49 and 566.63 for $t = 1$ under the BNB and BPIG regression models respectively, regardless of the exact composition per type of claim. Overall, if we consider all the three cases for $k_{1,j} + k_{2,j} = 2$ in Tables 13 and 15:

1. $k_{1,j} = 2$ and $k_{2,j} = 0$,
2. $k_{1,j} = 1$ and $k_{2,j} = 1$,
3. $k_{1,j} = 0$ and $k_{2,j} = 2$,

for the same year of coverage $t$ and the same type of risk-profile, the premium rates do not vary per claim type in the case of the two bivariate mixed Poisson models. In particular, in Tables 13 and 15, the values on the diagonals are always the same for a certain profile and for a certain year of insurance in the case of the BNB and BPIG models. However, these premium rates ought to be evaluated differently than under the two bivariate mixed Poisson models since the two types of claims have different frequencies and hence different means (as seen in Table 3, $\mathbb{E}(k_1) = 0.0954$ and $\mathbb{E}(k_2) = 0.0618$). Consequently, the probability of resulting claim free for the first type of claim is not the same as for the second type of claim. Therefore, from a practical business standpoint, these discrepancies in the two responses $k_{1,j}$ and $k_{2,j}$ should be taken into consideration for constructing a bivariate claim frequency model that will be the building block for the a posteriori ratemaking process. Fortunately, as was previously mentioned, the BNBGA and BNBIG distribution/regression model result in varying premium rates depending on the total number of claims and the composition of claims. This feature of the bivariate mixed Negative Binomial model enhances its validity as a model for the claim numbers $k_{1,j}$ and $k_{2,j}$ as it leads to a premium structure that can be sufficiently explained to policyholders and regulators. In particular, the findings in Tables 13 and 15 indicate that under the BNBGA and BNBIG distribution/regression models, a high number of claims per type two is discouraged more than per type one in all three



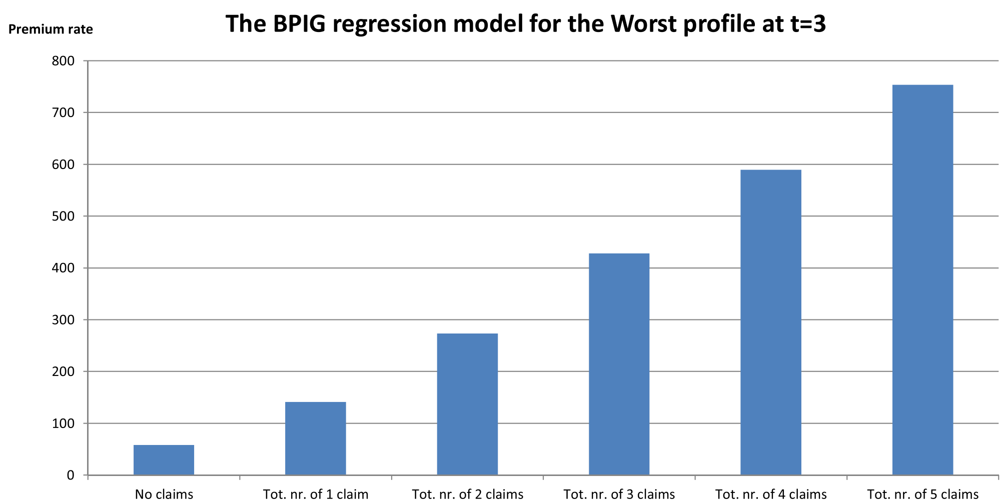




**Fig. 5.** Premium rates for the Worst profile at $t = 3$, bivariate mixed Poisson regression models.

of the above mentioned cases. Thus, since for our data the second response has a smaller mean than the first, this result is consistent with the core principle underpinning the design of Bonus-Malus systems which is to promote careful driving. For instance, for the case without covariates, we see from Table 13 that for $t = 1$ the policyholder has to pay premiums of 425.88, 478.66 and 541.50 under the BNBGA distribution and premiums of 362.47, 419.84 and 493.12 under the BNBIG distribution for the cases 1, 2 and 3 respectively. The same holds if we include the a priori information, in fact in Table 15 we see the same diversification in the premium rates depending on the type of claim. If the policyholder has the Best profile, then it has to pay premiums of 429.15, 482.85 and 546.87 under the BNBGA regression model and premiums of 367.29, 426.77 and 503.08 under the BNBIG regression model for the cases 1, 2 and 3 respectively for $t = 1$. A graphical representation of Table 15 is depicted in Figs. 1, 2, 3, 4, 5 and 6. In Figs. 1 and 2 we illustrate the case of the Best profile in $t = 1$ for the bivariate mixed Poisson models and the bivariate mixed Negative Binomial models respectively. In Figs. 3 and 4 we present the case of the Average profile in $t = 2$ for the bivariate mixed Poisson models and the bivariate mixed Negative Binomial models respectively. Finally, in Figs. 5 and 6 we show the case of the Worst profile in $t = 3$ for the bivariate mixed Poisson models and the bivariate mixed Negative Binomial models respectively.

## 6. Concluding remarks

In this article we demonstrated how to construct the MVMNB claim count regression model which can be used for ratemaking purposes in non-life insurance. The MVMNB model accommodates overdispersion and allows for positive correlation structures in high-dimensional count valued data. For demonstration purposes, we considered the BNBGA and BNBIG regression models which are suiTable for addressing the a posteriori, or Bonus-Malus, ratemaking problem of pricing an automobile insurance contract in the bivariate setting where the dynamics for the premium determination are governed by the interactions of third party bodily injury claims and property damage claims, which are conceivably positively correlated with each other. Furthermore, an EM type scheme was proposed for ML estimation of the parameters of both models which does not have their jpmf in closed form in a computationally parsimonious manner. The ML estimation procedure we developed avoids overflow issues that may be plausible via alternative numerical maximization algorithms.

In our numerical illustration, special consideration was put on the comparison of the a posteriori premium rates derived from the BNBIG distribution/regression model with those determined by the BNB and BPIG distributions/regression models. The reason for this is that, in contrast to the numerous studies that have been devoted to univariate experience rating models, the extent to which the a posteriori tariff system is affected when the claim frequency experience consists of detailed information on two different types of

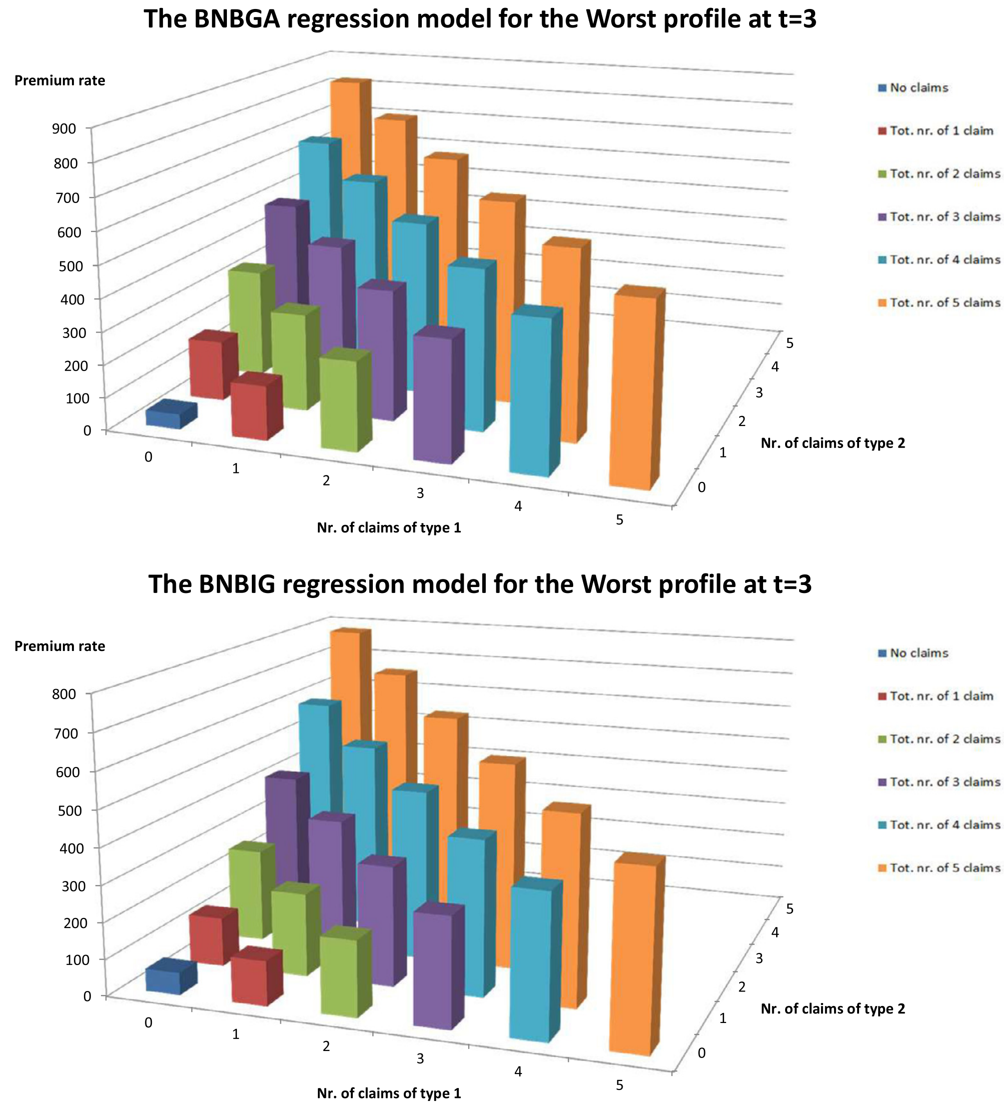



**Fig. 6.** Premium rates for the Worst profile at $t = 3$, bivariate mixed Negative Binomial regression models.

insurance claims arising from the same policy has not been fully elucidated in the MTPL pricing literature thus far. The results indicated that the employment of the BNBIG regression which presents the most superior fit for our data is beneficial for the insurance company, since it can enable them to adopt a milder a posteriori pricing strategy for policyholders with some claim experience and it can provide a more complete picture about the extent to which the premiums vary according to the frequency of each type of claim.

Finally, an interesting line of further research would be to consider generalizations of the MVMNB regression model. For instance, multiple random effects which are distributed according to different mixing densities can be added for modelling the unobserved heterogeneity when dealing with different types of claims from different types of coverage, see Bermúdez and Karlis (2017) who followed this approach in the bivairate setting. Also, for example, the MVMNB regression model can be adapted to take into account both the positive correlation between the different types of claims and the serial correlation between the observations of the same insured observed over time. This could be done proceeding along similar lines as Bermúdez et al. (2018), who were the first to consider a bivariate INAR(1) regression model which can provide an integrated framework that can take into account both time dependence and cross dependence, which have been commonly treated as separate entities in the ratemaking literature.

## Declaration of competing interest

There is no competing interest.

## Acknowledgements

The authors would like to thank the participants at the 23rd International Congress on "Insurance: Mathematics and Economics" (IME) in Munich, 2019. We are also deeply grateful to Dr Tsz Chai Fung for his very useful feedback. Finally, we would like to thank the editor and the two anonymous referees for their very helpful comments and suggestions which have significantly improved this article.

## References


Abdallah, A., Boucher, J.-P., Cossette, H., 2016. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insurance. Mathematics & Economics 68, 120–133.

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson-lognormal models for crash severity modeling and site ranking. Transportation Research Record 2136 (1), 82–91.

Ahmad, I.S., Hassan, A., Ahmad, P.B., 2019. Negative binomial-reciprocal inverse Gaussian distribution: statistical properties with applications. Preprint arXiv:1906.03224.

Barreto-Souza, W., Simas, A.B., 2016. General mixed Poisson regression models with varying dispersion. Statistics and Computing 26 (6), 1263–1280.

Bermúdez, L., 2009. A priori ratemaking using bivariate Poisson regression models. Insurance. Mathematics & Economics 44 (1), 135–141.

Bermúdez, L., Karlis, D., 2011. Bayesian multivariate Poisson models for insurance ratemaking. Insurance. Mathematics & Economics 48 (2), 226–236.

Bermúdez, L., Karlis, D., 2012. A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. Computational Statistics & Data Analysis 56 (12), 3988–3999.

Bermúdez, L., Karlis, D., 2017. A posteriori ratemaking using bivariate Poisson models. Scandinavian Actuarial Journal 2017 (2), 148–158.

Bermúdez, L., Guillén, M., Karlis, D., 2018. Allowing for time and cross dependence assumptions between claim counts in ratemaking models. Insurance. Mathematics & Economics 83, 161–169.

Booth, J.G., Hobert, J.P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society, Series B, Statistical Methodology 61 (1), 265–285.

Booth, J.G., Hobert, J.P., Jank, W., 2001. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. Statistical Modelling 1 (4), 333–349.

Boucher, J.-P., Denuit, M., 2008. Credibility premiums for the zero-inflated Poisson model and new hunger for bonus interpretation. Insurance. Mathematics & Economics 42 (2), 727–735.

Boucher, J.-P., Denuit, M., Guillén, M., 2008. Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. Variance 2 (1), 135–162.

Brouhns, N., Guillén, M., Denuit, M., Pinquet, J., 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. The Journal of Risk and Insurance 70 (4), 577–599.

Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York, USA.

Calderín-Ojeda, E., Gómez-Déniz, E., 2019. The multivariate negative binomial-Lindley distribution. Properties and new representation for the univariate case. Journal of Computational and Applied Mathematics 347, 36–48.

Cameron, A.C., Li, T., Trivedi, P.K., Zimmer, D.M., 2004. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. Econometrics Journal 7 (2), 566–584.

Chiquet, J., Mariadassou, M., Robin, S., 2020. The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances. BioRxiv.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, Methodological 39 (1), 1–22.

Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J.-F., 2007. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. John Wiley & Sons.

Dionne, G., Vanasse, C., 1992. Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics 7 (2), 149–165.

El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.

Fung, T.C., Badescu, A.L., Lin, X.S., 2019a. A class of mixture of experts models for general insurance: application to correlated claim frequencies. ASTIN Bulletin: The Journal of the IAA 49 (3), 647–688.

Fung, T.C., Badescu, A.L., Lin, X.S., 2019b. A class of mixture of experts models for general insurance: theoretical developments. Insurance. Mathematics & Economics 89, 111–127.

Gençtürk, Y., Yiğiter, A., 2016. Modelling claim number using a new mixture model: negative binomial gamma distribution. Journal of Statistical Computation and Simulation 86 (10), 1829–1839.

Genest, C., Nešlehová, J., 2007. A primer on copulas for count data. ASTIN Bulletin: The Journal of the IAA 37 (2), 475–515.

Ghitany, M., Karlis, D., Al-Mutairi, D., Al-Awadhi, F., 2012. An EM algorithm for multivariate mixed Poisson regression models and its application. Applied Mathematical Sciences 6 (137), 6843–6856.

Gómez-Déniz, E., Calderín-Ojeda, E., 2017. An alternative representation of the negative binomial-Lindley distribution. New results and applications. Preprint arXiv:1703.04812.

Gómez-Déniz, E., Calderín-Ojeda, E., 2018. Multivariate credibility in bonus-malus systems distinguishing between different types of claims. Risks 6 (2), 34.

Gómez-Déniz, E., Hernández-Bastida, A., Fernández-Sánchez, M.P., 2014. Computing credibility bonus-malus premiums using the total claim amount distribution. Hacettepe Journal of Mathematics and Statistics.

Gómez-Déniz, E., Sarabia, J.M., Calderín-Ojeda, E., 2008. Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. Insurance. Mathematics & Economics 42 (1), 39–49.

Gómez-Déniz, E., Vázquez-Polo, F., 2003. Robustness in Bayesian Models for Bonus-Malus Systems. Intelligent and Other Computational Techniques in Insurance: Theory and Applications. World Scientific, pp. 435–463.

Gómez-Déniz, E., Calderín-Ojeda, E., 2021. A priori ratemaking selection using multivariate regression models allowing different coverages in auto insurance. Risks 9 (7), 137.

Gui, W., Huang, R., Lin, X.S., 2018. Fitting the Erlang mixture model to data via a GEM-CMM algorithm. Journal of Computational and Applied Mathematics 343, 189–205.

Gurmu, S., Elder, J., 2000. Generalized bivariate count data regression models. Economics Letters 68 (1), 31–36.

Insurance Europe, 2015. European motor insurance markets, addendum, Accessed 15, 2017.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, vol. 112. Springer.

Jørgensen, B., 1982. Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics, vol. 9.

Karlis, D., 2001. A general EM approach for maximum likelihood estimation in mixed Poisson regression models. Statistical Modelling 1 (4), 305–318.

Karlis, D., 2005. EM algorithm for mixed Poisson and other discrete distributions. ASTIN Bulletin: The Journal of the IAA 35 (1), 3–24.

Karlis, D., Tzougas, G., Frangos, N., 2018. Confidence intervals of the premiums of optimal bonus malus systems. Scandinavian Actuarial Journal 2018 (2), 129–144.

Kocherlakota, S., 1988. On the compounded bivariate Poisson distribution: a unified treatment. Annals of the Institute of Statistical Mathematics 40 (1), 61–76.

Lemaire, J., 1995. Bonus-Malus Systems in Automobile Insurance. Kluwer Academic Publishers.

Kuha, J., 2004. AIC and BIC: comparisons of assumptions and performance. Sociological Methods & Research 33 (2), 188–229.

Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B, Methodological 44 (2), 226–233.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40 (3), 964–975.

McLachlan, G.J., Krishnan, T., 2007. The EM Algorithm and Extensions, vol. 382. John Wiley & Sons.

Mert, M., Saykan, Y., 2005. On a bonus-malus system where the claim frequency distribution is geometric and the claim severity distribution is Pareto. Hacettepe Journal of Mathematics and Statistics 34, 75–81.

Munkin, M.K., Trivedi, P.K., 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. Econometrics Journal 2 (1), 29–48.

Ni, W., Constantinescu, C., Pantelous, A.A., 2014a. Bonus-Malus systems with Weibull distributed claim severities. Annals of Actuarial Science 8 (2), 217–233.

Ni, W., Li, B., Constantinescu, C., Pantelous, A.A., 2014b. Bonus-Malus systems with hybrid claim severity distributions. Vulnerability, Uncertainty, and Risk. Quantification, Mitigation, and Management, 1234–1244.

Nikoloulopoulos, A.K., Karlis, D., 2009. Finite normal mixture copulas for multivariate discrete data modelling. Journal of Statistical Planning and Inference 139, 3878–3890.

Nikoloulopoulos, A.K., Karlis, D., 2010. Regression in a copula model for bivariate count data. Journal of Applied Statistics 37 (9), 1555–1568.

Nikoloulopoulos, A.K., Joe, H., Chaganty, N.R., 2011. Weighted scores method for regression models with dependent data. Biostatistics 12 (4), 653–665.

Oh, H., Jeong, J., Ahn, Y., Valdez, E.A., 2020. On a multi-year microlevel collective risk model. Preprint arXiv:2006.06151.

Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019 (1), 1–6.
Picech, L., 1994. The Merit Rating Factor in a Multiplicative Rate Making Model. ASTIN Colloquium, Cannes.
Pinquet, J., 1997. Allowance for cost of claims in bonus-malus systems. ASTIN Bulletin: The Journal of the IAA 27 (1), 33–57.
Pinquet, J., 1998. Designing optimal bonus-malus systems from different types of claims. ASTIN Bulletin: The Journal of the IAA 28 (2), 205–220.
Raftery, A.E., 1995. Bayesian model selection in social research. Sociological Methodology 25, 111–163.
Rigby, R.A., Stasinopoulos, D.M., Akantziliotou, C., 2008. A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. Computational Statistics & Data Analysis 53 (2), 381–393.
Santi, D., Purnaba, I., Mangku, I., 2016. Bonus-Malus System with the Claim Frequency Distribution is Geometric and the Severity Distribution is Truncated Weibull. IOP Conference Series: Earth and Environmental Science, vol. 31. IOP Publishing, p. 012006.
Seshadri, V., 1993. The Inverse Gaussian Distribution: Statistical Theory and Applications. Oxford Science Publications.
Shared, M., 1980. On mixtures from exponential families. Journal of the Royal Statistical Society, Series B, Methodological 42 (2), 192–198.
Shengwang, M., Wei, Y., Whitmore, G., 1999. Accounting for individual over-dispersion in a bonus-malus automobile insurance system. ASTIN Bulletin: The Journal of the IAA 29 (2), 327–337.
Shi, P., Valdez, E.A., 2014. Multivariate negative binomial models for insurance claim counts. Insurance. Mathematics & Economics 55, 18–29.
Silva, A., Rothstein, S.J., McNicholas, P.D., Subedi, S., 2019. A multivariate Poisson-lognormal mixture model for clustering transcriptome sequencing data. BMC Bioinformatics 20 (1), 1–11.
Stein, G.Z., Juritz, J.M., 1987. Bivariate compound Poisson distributions. Communications in Statistics. Theory and Methods 16 (12), 3591–3607.
Stein, G.Z., Zucchini, W., Juritz, J.M., 1987. Parameter estimation for the Sichel distribution and its multivariate extension. Journal of the American Statistical Association 82 (399), 938–944.
Tremblay, L., 1992. Using the Poisson inverse Gaussian in bonus-malus systems. ASTIN Bulletin: The Journal of the IAA 22 (1), 97–106.
Tseung, S.C., Badescu, A.L., Fung, T.C., Lin, X.S., 2020. LRMoE: An R Package for Flexible Actuarial Loss Modelling Using Mixture of Experts Regression Model. Available at SSRN 3740215.
Tzougas, G., 2020. EM estimation for the Poisson-inverse gamma regression model with varying dispersion: an application to insurance ratemaking. Risks 8 (3), 97.
Tzougas, G., Frangos, N., 2014. The design of an optimal bonus-malus system based on the Sichel distribution. In: Modern Problems in Insurance Mathematics. Springer, pp. 239–260.
Tzougas, G., Hoon, W., Lim, J., 2019. The negative binomial-inverse Gaussian regression model with an application to insurance ratemaking. European Actuarial Journal 9 (1), 323–344.
Tzougas, G., Vrontos, S., Frangos, N., 2018. Bonus-malus systems with two-component mixture models arising from different parametric families. North American Actuarial Journal 22 (1), 55–91.
Tzougas, G., Yik, W., Mustaqeem, M., 2020. Insurance ratemaking using the exponential-lognormal regression model. Annals of Actuarial Science 14 (1), 42–71. https://doi.org/10.1017/S1748499519000034.
Zhan, X., Aziz, H.A., Ukkusuri, S.V., 2015. An efficient parallel sampling technique for multivariate Poisson-lognormal model: analysis with two crash count datasets. Analytic Methods in Accident Research 8, 45–60.
Zamani, H., Ismail, N., 2010. Negative binomial-Lindley distribution and its application. Journal of Mathematics and Statistics 6 (1), 4–9
.