

---

# The Negative Binomial-Inverse Gaussian Regression Model with an Application to Insurance Ratemaking

G. Tzougas · W.L. Hoon · J.M. Lim

**Abstract** This paper presents the Negative Binomial-Inverse Gaussian regression model for approximating the number of claims as an alternative to mixed Poisson regression models that have been widely used in various disciplines including actuarial applications. The Negative Binomial-Inverse Gaussian regression model can be considered as a plausible model for highly dispersed claim count data and this is the first time that it is used in a statistical or actuarial context. The main achievement is that we propose a quite simple Expectation-Maximization type algorithm for maximum likelihood estimation of the model. Finally, a real data application using motor insurance data is examined and both the a priori and a posteriori, or Bonus-Malus, premium rates resulting from the Negative Binomial-Inverse Gaussian model are calculated via the net premium principle and compared to those determined by the Negative Binomial Type I and the Poisson-Inverse Gaussian regression models that have been traditionally used for a priori and a posteriori ratemaking.

**Keywords** Negative Binomial-Inverse Gaussian Regression Model · EM Algorithm · Motor Third Party Liability Insurance · Ratemaking

## 1 Introduction

In a highly competitive insurance market, in order to avoid lapses, a major challenge of the ratemaking process is the measurement and design of fair tariff structures that will match the premium rates to the risks as closely as possible and also to the rating structures used by competitors. Traditionally, a dual approach to ratemaking was adopted by actuaries who developed both a priori ratemaking schemes and a posteriori ratemaking mechanisms or Bonus-Malus Systems (BMSs). The former process relies on the use of claim frequency and severity generalized linear models (GLMs) for computing the a priori premiums. References for a priori ratemaking include, for example, Dean, Lawless and Willmot (1989), Haberman and Renshaw (1996), Denuit and Lang (2004), Yip and Yau (2005), Boucher et al. (2007), Denuit et al. (2007) de Jong and Heller (2008), Kaas et al. (2008), Frees (2010, Section 18.3) and Tzougas et al. (2015). The latter process uses additional information about the claim frequency history of the policyholders to calculate the a posteriori, or Bonus Malus, premium rates in way which readjusts the previous a priori premiums, making the price discrimination even more fair and reasonable. There is a large amount of literature on BMSs and for a detailed description of these systems the interested reader can refer to the seminal work of Lemaire (1995). Further references for BMSs include, among others, Trembley (1992), Picech (1994), Pinquet (1997 and 1998), Dionne and Vanasse (1989 and 1992), Frangos and Vrontos (2001), Brouhns et al. (2003), Gómez-Déniz et al. (2014), Denuit et al. (2007), Boucher et al. (2008), Lemaire et al. (2015 and 2016) and Tzougas et al. (2014 and 2018).

Let us now discuss our motivation behind this study. As is well known, mixed Poisson regression models have been widely used in Motor Third Party Liability (MTPL) insurance settings for approximating claim

---

G. Tzougas  
Department of Statistics, London School of Economics and Political Science, London, UK  
Tel.: +44 (0)20 7955 7453  
E-mail: g.tzougas@lse.ac.uk

W.L. Hoon  
Department of Statistics, London School of Economics and Political Science, London, UK

J.M. Lim  
Department of Statistics, London School of Economics and Political Science, London, UK

frequency primarily due to the lack of algebraic problems appearing when one attempts to maximize their nonlinear log-likelihood function. In particular, the Negative Binomial Type I (NBI) and Poisson-Inverse Gaussian (PIG) models have been the most popular choices because of the conjugacy, in a Bayesian sense, between the Gamma and Inverse Gaussian priors, or mixing, distributions respectively and the Poisson distribution, which facilitates maximum likelihood (ML) estimation and a Bayesian approach towards updating mean frequency estimates.

However, regardless of the statistical and mathematical convenience of mixed Poisson models, there is no guarantee that individual overdispersion and variation in claim propensity have precisely the distributional forms implied by the members of the mixed Poisson family. Moreover, taking into account that according to a 2016 report by Insurance Europe, motor insurance totalled 132bn Euros in premiums in 2015 (see, Insurance Europe, 2017), it becomes clear that an important task of actuaries is to be able to innovate in designing a priori and a posteriori ratemaking mechanisms by constructing more representative probabilistic models for the distribution of the number of claims based on a sound statistical basis that can capture important aspects of real insurance portfolios, enabling the actuary to apply premiums to insured risks in a fair and equitable manner. Otherwise, the concept of adverse selection could undermine the solvency of the company.

The aim of the present work is to propose the Negative Binomial-Inverse Gaussian (NBIG) regression model as a competitive alternative to mixed Poisson regression models. The NBIG distribution has thick tails and can be considered as a candidate model for highly dispersed claim count data. A first derivation of the Negative Binomial-Inverse Gaussian distribution can be found in Gómez-Déniz et al. (2008), who considered some estimation methods for both the univariate and multivariate case, gave an excellent account of statistical methods connected to both cases and also proposed their use in an actuarial context. Also, it should be noted that in the actuarial literature, so far, only very few mixed Negative Binomial distributions and regression models have been studied in depth. In particular, the Negative Binomial-Pareto distribution (see, Klugman et al., 1998, Shengwang et al., 1999 and Gómez-Déniz and Vázquez, 2003), the Negative Binomial- Beta regression model (see, Boucher et al. 2008), the Negative Binomial-Gamma (see, Gençtürk and Yiğiter, 2016) and the Negative Binomial-Lindley distribution (see, Zamani and Ismail, 2010 and Gómez-Déniz and Calderín, 2017) were used in an actuarial context. However, this is the first time that the Negative Binomial-Inverse Gaussian model with a regression component is used in a statistical or an actuarial setting. One of the main obstacles for using the Negative Binomial-Inverse Gaussian distribution for the case with covariates is that its log-likelihood is complicated and hence its maximization needs a special effort. In particular, compared to the classical Negative Binomial regression model, there is no analytical form for the distribution of the number of claims if the random effect term, which follows the Inverse Gaussian distribution, is marginalized out. As a result, ML estimation of the Negative Binomial-Inverse Gaussian regression model is less straightforward to calculate.

The contribution of this paper is that it illustrates that ML estimation of the NBIG regression model can be accomplished relatively easily via an Expectation Maximization (EM) type algorithm which can address situations where the mixing distribution, such as the Inverse Gaussian, is not conjugate to the Negative Binomial distribution. Furthermore, the a priori and the a posteriori, or Bonus-Malus, premium rates resulting from the new model, which does not have a probability mass function (pmf) in closed form, are calculated via the law of total expectation and the use of numerical approximation and compared to those resulting from the NBI and PIG models, that have been widely used for actuarial purposes.

The layout of the paper is as follows: Section 2 presents the derivation of the NBIG regression model. Section 3 fully describes the ML estimation through the EM algorithm. Section 4 contains an application to a data set concerning car insurance claims at fault. Finally, some further issues are examined in Section 5 and concluding remarks can be found in Section 6.

## 2 The Negative Binomial-Inverse Gaussian Regression Model

Consider a policyholder  $i$  whose number of claims, denoted as  $k_i$ , are independent and suppose that  $\mathbf{x}_i$  is the vector of individual characteristics or characteristics of the car related to the  $i$ th insured person,  $i = 1, \dots, n$ , which represent different a priori rating variables. The Negative Binomial-Inverse Gaussian regression model can be constructed as follows.

Assume that  $k_i|\mathbf{x}_i, \lambda_i$  follows the Negative Binomial (NB) distribution with probability mass function (pmf) given by

$$P(k_i|\mathbf{x}_i, \lambda_i) = \frac{\Gamma(k_i + \sigma)}{k_i! \Gamma(\sigma)} \left( \frac{\lambda_i \varepsilon_i}{\sigma + \lambda_i \varepsilon_i} \right)^{k_i} \left( \frac{\sigma}{\sigma + \lambda_i \varepsilon_i} \right)^\sigma, \quad (1)$$

for  $k_i = 0, 1, 2, 3, \dots$ , with  $\lambda_i > 0$ ,  $\sigma > 0$ , where  $\varepsilon_i = \exp(\mathbf{x}_i^T \beta)$  and where  $\beta$  is the vector of the regression coefficients. The mean and the variance of  $k_i|\mathbf{x}_i, \lambda_i$  are given by

$$E(k_i | \mathbf{x}_i, \lambda_i) = \exp(\mathbf{x}_i^T \beta + \log(\lambda_i)) \quad (2)$$

and

$$Var(k_i | \mathbf{x}_i, \lambda_i) = \exp(\mathbf{x}_i^T \beta + \log(\lambda_i)) \left[ 1 + \frac{\exp(\mathbf{x}_i^T \beta + \log(\lambda_i))}{\sigma} \right]. \quad (3)$$

Note that the scale parameter  $\sigma$  governs the responsiveness of overdispersion to the mean number of claims. The larger is  $\sigma$ , the smaller is the degree of overdispersion. In the limit, as  $\sigma$  approaches infinity,  $P(k_i | \mathbf{x}_i, \lambda_i)$  tends to a Poisson distribution with mean  $\exp(\mathbf{x}_i^T \beta + \log(\lambda_i))$ .

Let us now assume that  $\lambda_i$  follows an Inverse Gaussian (IG) distribution with probability density function (pdf) given by

$$f(\lambda_i) = \frac{\gamma}{\sqrt{2\pi}} \exp(\gamma^2) \lambda_i^{-\frac{3}{2}} \exp\left[-\frac{1}{2} \left(\frac{\gamma^2}{\lambda_i} + \gamma^2 \lambda_i\right)\right], \quad (4)$$

where  $\gamma > 0$  with mean  $E(\lambda_i) = 1$  and variance  $Var(\lambda_i) = \frac{1}{\gamma^2}$ . The Inverse Gaussian prior distribution which is given by Eq. (4) has to have a unit mean in order for the model to be estimable, otherwise identifiability issues may arise. Under this restriction, the overdispersion relative to the simple exponential distribution is  $\frac{1}{\gamma^2}$ . Thus, if  $\gamma$  tends to infinity, this distribution can be reduced to the exponential distribution. More details about the Inverse Gaussian distribution which is a special case of the more general family of generalized Inverse Gaussian distributions can be found in Jørgensen (1982). Note also that there are several different parameterizations of the Inverse Gaussian distribution (see, Seshadri, 1993).

Considering the assumptions of the model, i.e. Eq. (1) and Eq.(4), the distribution of  $k_i$  will be a Negative Binomial-Inverse Gaussian (NBIG) distribution with parameters  $\exp(\mathbf{x}_i^T \beta)$ ,  $\sigma$ ,  $\gamma$  and pmf<sup>1</sup>

$$P(k_i | \mathbf{x}_i) = \int_0^{\infty} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i. \quad (5)$$

Unfortunately, the last integral cannot be simplified but it can be computed via numerical integration. Also, using the laws of total expectation and total variance and the moments of the NB distribution one can easily find that the mean and the variance of  $k_i | \mathbf{x}_i$  are given by

$$E(k_i | \mathbf{x}_i) = E_{\lambda_i} [E(k_i | \mathbf{x}_i, \lambda_i)] = \exp(\mathbf{x}_i^T \beta) E_{\lambda_i} [\lambda_i] = \exp(\mathbf{x}_i^T \beta) \quad (6)$$

and

$$\begin{aligned} Var(k_i | \mathbf{x}_i) &= E_{\lambda_i} [Var(k_i | \mathbf{x}_i, \lambda_i)] + Var_{\lambda_i} [E(k_i | \mathbf{x}_i, \lambda_i)] \\ &= E(k_i | \mathbf{x}_i) + E^2(k_i | \mathbf{x}_i) \left( \frac{1 + \sigma + \gamma^2}{\sigma \gamma^2} \right). \end{aligned} \quad (7)$$

### 3 The EM Algorithm for the Negative Binomial-Inverse Gaussian Regression Model

In this section we describe how an EM type algorithm can be used to facilitate the ML estimation of the model. Let  $(k_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$ , be a sample of independent observations, where  $k_i$  is the response and  $\mathbf{x}_i$  is a vector of covariate information. Also, consider that the data are produced according to the Negative Binomial-Inverse Gaussian model. Then, the log-likelihood can be written as

$$l(\phi) = \sum_{i=1}^n \log(P(k_i | \mathbf{x}_i)), \quad (8)$$

where  $\phi = (\gamma, \sigma, \beta)$  is the vector of the parameters and where  $P(k_i | \mathbf{x}_i)$  is the pmf of Negative Binomial-Inverse Gaussian distribution, which is given by Eq. (5). The EM algorithm (see, Dempster et al. 1977 and McLachlan and Krishnan, 2007) is a very useful algorithm for ML estimation for data arising from mixtures, since the mixing operation can be considered as producing missing data. In our case the missing data are simply the realizations of the unobserved mixing parameter  $\lambda_i$  for the  $i$ th observation. Thus, if

<sup>1</sup> Note that the NBIG distribution given in Eq. (5) is different from the one used in Gómez-Deniz et al. (2008), who considered the case without covariate information.

one augments the unobserved data  $\lambda_i$  to the observed data  $(k_i, \mathbf{x}_i)$ , for  $i = 1, \dots, n$ , then the complete data log-likelihood takes the form

$$l_c(\phi) = \sum_{i=1}^n \log(P(k_i|\mathbf{x}_i, \lambda_i)) + \sum_{i=1}^n \log(f(\lambda_i|\gamma)), \quad (9)$$

where  $P(k_i|\mathbf{x}_i, \lambda_i)$  is the pmf of Negative Binomial distribution, which is given by Eq. (1), and where  $f(\lambda_i|\gamma)$  denotes the pdf of the mixing distribution, the Inverse Gaussian distribution, which is given by Eq. (4).

Maximization of the above function with respect to the vector of parameters  $\phi$  is not easy. Fortunately, the following mixture derivation of the model can be used to achieve ML estimation via an EM type algorithm

$$\begin{aligned} k_i &\sim \text{Poisson}(\vartheta_i), \\ \vartheta_i &\sim \text{Gamma}\left(\sigma, \frac{\sigma}{\lambda_i \varepsilon_i}\right) \\ \lambda_i &\sim \text{Inverse Gaussian}(\gamma). \end{aligned} \quad (10)$$

Also, let us assume that  $P(k_i|\vartheta_i) = e^{-\vartheta_i} \vartheta_i^{k_i} / k_i!$  is the pmf of the Poisson distribution,  $g(\vartheta_i|\beta, \sigma, \lambda_i) = \vartheta_i^{\sigma-1} \exp(-\frac{\sigma}{\lambda_i \varepsilon_i} \vartheta_i) \left(\frac{\sigma}{\lambda_i \varepsilon_i}\right)^\sigma / \Gamma(\sigma)$  is the pdf of the Gamma distribution and  $f(\lambda_i|\gamma)$  is the pdf of the Inverse Gaussian distribution with unit mean. Then, using the mixture representation in (10) the complete data log-likelihood can be expressed as

$$l_c(\phi) \propto \sum_{i=1}^n \log(g(\vartheta_i|\beta, \sigma, \lambda_i)) + \sum_{i=1}^n \log(f(\lambda_i|\gamma)). \quad (11)$$

The regression coefficients  $\beta$  and the parameter  $\sigma$  are involved in the first term and the parameter  $\gamma$  is involved in the second term of Eq. (11), which correspond to the log-likelihoods of the Gamma and Inverse Gaussian components respectively.

The conditional expectation of the complete data log-likelihood is given by

$$\begin{aligned} Q(\phi; \phi_{(r)}) &\equiv E_{\lambda_i}(l_c(\phi) | k_i, \mathbf{x}_i, \phi_{(r)}) \\ &\propto E_{\lambda_i} \left[ \sum_{i=1}^n \log(g(\vartheta_i|\beta_{(r)}, \sigma_{(r)}, \lambda_i)) \right] + E_{\lambda_i} \left[ \sum_{i=1}^n \log(f(\lambda_i|\gamma_{(r)})) \right], \end{aligned} \quad (12)$$

where  $\phi_{(r)} = (\gamma_{(r)}, \sigma_{(r)}, \beta_{(r)})$  is the estimate of  $\phi$  in the E-step of our EM algorithm. In what follows, in the E-step we present a result that gives us the conditional expectations of some functions of the unobserved data  $\lambda_i$  which are involved in Eq. (12), while the M-step, consists in maximizing the  $Q$ -function with respect to  $\phi$ . We call attention to the fact that for maximizing the log-likelihood of the Inverse Gaussian variate it is well known (see, for instance, Karlis, 2001) that we need to compute the conditional expectations  $E_{\lambda_i}(\lambda_i | k_i, \mathbf{x}_i, \phi)$  and  $E_{\lambda_i}\left(\frac{1}{\lambda_i} | k_i, \mathbf{x}_i, \phi\right)$  which correspond to the sufficient statistics for  $\gamma$ . On the other hand, it is easy to see that  $E_{\lambda_i}\left[E_{\vartheta_i}\left(\frac{\vartheta_i}{\lambda_i \varepsilon_i} | k_i, \mathbf{x}_i, \phi\right)\right]$  and  $E_{\lambda_i}\left[E_{\vartheta_i}\left(\log\left(\frac{\vartheta_i}{\lambda_i \varepsilon_i}\right) | k_i, \mathbf{x}_i, \phi\right)\right]$  have to be calculated for maximizing the log-likelihood of the Gamma variate with respect to  $\beta$  and  $\sigma$ . Note also that if  $k_i \sim \text{Poisson}(\vartheta_i)$  distribution and  $\vartheta_i \sim \text{Gamma}\left(\sigma, \frac{\sigma}{\lambda_i \varepsilon_i}\right)$  distribution then the posterior distribution of  $\vartheta_i | k_i, \mathbf{x}_i, \sigma, \beta$  is a  $\text{Gamma}\left(k_i + \sigma, \frac{\sigma}{\lambda_i \varepsilon_i} + 1\right)$  distribution. Thus, we have that

$$E_{\vartheta_i}\left(\frac{\vartheta_i}{\lambda_i \varepsilon_i} | k_i, \mathbf{x}_i, \phi\right) = \frac{k_i + \sigma}{\lambda_i \varepsilon_i + \sigma} \quad (13)$$

and

$$E_{\vartheta_i}\left(\log\left(\frac{\vartheta_i}{\lambda_i \varepsilon_i}\right) | k_i, \mathbf{x}_i, \phi\right) = \Psi(k_i + \sigma) - \log(\lambda_i \varepsilon_i + \sigma), \quad (14)$$

where  $\Psi(\cdot)$  denotes the digamma function. These conditional expectations will be very useful for implementing the E-step of the EM algorithm. The EM type algorithm for the NBIG regression model can be formally described as follows.

**E-Step:**

- Given the values of the parameters after the  $r$ th iteration, calculate for all  $i = 1, 2, \dots, n$ , the pseudo-values

$$w_{1,i} = E_{\lambda_i} (\lambda_i | k_i, \mathbf{x}_i, \phi_{(r)}) = \frac{\int_0^{\infty} \lambda_i P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}{\int_0^{\infty} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}, \quad (15)$$

$$w_{2,i} = E_{\lambda_i} \left( \frac{1}{\lambda_i} | k_i, \mathbf{x}_i, \phi_{(r)} \right) = \frac{\int_0^{\infty} \frac{1}{\lambda_i} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}{\int_0^{\infty} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}, \quad (16)$$

$$w_{3,i} = E_{\lambda_i} \left( \frac{1}{(\lambda_i \varepsilon_i + \sigma_{(r)})} | k_i, \mathbf{x}_i, \phi_{(r)} \right) = \frac{\int_0^{\infty} \frac{1}{(\lambda_i \varepsilon_i + \sigma_{(r)})} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}{\int_0^{\infty} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}, \quad (17)$$

$$w_{4,i} = E_{\lambda_i} (\log(\lambda_i \varepsilon_i + \sigma_{(r)}) | k_i, \mathbf{x}_i, \phi_{(r)}) = \frac{\int_0^{\infty} \log(\lambda_i \varepsilon_i + \sigma_{(r)}) P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}{\int_0^{\infty} P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i) d\lambda_i}. \quad (18)$$

- Using Eqs (13, 14, 17 and 18) we also have that

$$s_{1,i} = E_{\lambda_i} \left[ E_{\vartheta_i} \left( \frac{\vartheta_i}{\lambda_i \varepsilon_i} | k_i, \mathbf{x}_i, \phi_{(r)} \right) \right] = (k_i + \sigma_{(r)}) w_{3,i} \quad (19)$$

and

$$s_{2,i} = E_{\lambda_i} \left[ E_{\vartheta_i} \left( \log \left( \frac{\vartheta_i}{\lambda_i \varepsilon_i} \right) | k_i, \mathbf{x}_i, \phi_{(r)} \right) \right] = \Psi(k_i + \sigma_{(r)}) - w_{4,i}. \quad (20)$$

Clearly the expectations involved in the E-step of the algorithm do not have closed form expressions and thus numerical approximations are needed. Specifically, Eqs (15, 16, 17 and 18) can be evaluated numerically. Alternatively, a Monte Carlo approach is also possible using a rejection algorithm. The latter case leads to variants of the EM algorithm such as the Monte Carlo EM (MCEM) algorithm (see, for instance, Booth and Hobert, 1999, Booth et al., 2001 and Karlis, 2005) which do not require knowledge of the pmf  $P(k_i | \mathbf{x}_i)$  but it suffices to be able to simulate from the posterior density  $f(\lambda_i | k_i, \mathbf{x}_i, \phi)$ .

**M-step:**

- The Newton-Raphson algorithm is employed to obtain ML estimates of the elements of  $\beta$ . Taking the necessary derivatives of  $Q(\phi; \phi_{(r)})$  with respect to  $\beta$  we obtain the following results:

$$g(\beta) = E_{\lambda_i} \left( \frac{\partial l_c}{\partial \beta} | k_i, \mathbf{x}_i, \phi \right) = \sigma \sum_{i=1}^n (s_{1,i} - 1) \mathbf{x}_i \quad (21)$$

and

$$G(\beta) = E_{\lambda_i} \left( \frac{\partial^2 l_c}{\partial \beta \partial \beta^T} | k_i, \mathbf{x}_i, \phi \right) = -\sigma \sum_{i=1}^n s_{1,i} x_i x_i^T = -\sigma \mathbf{X}^T \mathbf{W} \mathbf{X}, \quad (22)$$

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

The iterative procedure for the Newton-Raphson algorithm goes as follows:

$$\beta_{(r+1)} \equiv \beta_{(r)} - [G(\beta_{(r)})]^{-1} g(\beta_{(r)}). \quad (23)$$

- Then, update  $\sigma$  with the one step ahead Newton iteration

$$\sigma_{(r+1)} = \sigma_{(r)} - \frac{\Psi(\sigma_{(r)}) + \bar{s}_1 - \bar{s}_2 - \log(\sigma_{(r)}) - 1}{\Psi_3(\sigma_{(r)}) - \frac{1}{\sigma_{(r)}}}, \quad (24)$$

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

- Finally, update  $\gamma$  with

$$\gamma_{(r+1)} = (\bar{w}_1 + \bar{w}_2 - 2)^{-\frac{1}{2}}. \quad (25)$$

- Note also that when the regression component of the model is limited to a constant  $\beta_0$  one obtains  $E(k_i | \mathbf{x}_i) = \exp(\beta_0) = \mu$  and thus this EM type algorithm can be employed for the ML estimation of the ‘univariate’, without regression component, model.

#### 4 Numerical Illustration

The data were kindly provided by a major insurance company operating in Greece and concern a MTPL insurance portfolio observed for 3.5 years (from April 2013 up to October 2016). Only policyholders with complete records, i.e. with availability of all the explanatory variables under consideration were considered. There were 146129 observations that met our criteria. Response variable is the number of claims at fault registered for each insured vehicle in the data set and the a priori rating variables we employ are: the size of the city (CS) where the policyholders live, the age of their car (AC) and the horsepower (HP) of their car.

- The variable CS consists of three categories of policyholders, those who live in a: C1= "small city", C2 = "middle sized city" and C3 = "large city".
- The variable AC consists of three categories of cars, those of age: C1 = "between 0 to 8 years", C2 = "between 8 to 16 years" and C3 = "greater than 16 years".
- The variable HP consists of three categories of cars, those with a HP: C1 = "0-1400 cc", C2 = "1400-1800 cc", C3 = "greater than 1800 cc".

Table 1 contains some standard descriptive statistics for claim counts along with the number of observations in each category of the three explanatory variables. Furthermore, Figure 1 displays the corresponding to the explanatory variables descriptive histograms, giving us a flavour of the range of their values.

**Table 1** Descriptive Statistics of Claim Counts - Size of the Different Categories of the Explanatory Variables

statistic	value	Age of the car (AC)	Horsepower of the car (HP)	Size of the city (CS)
# observations	146129	C1: 83927	C1: 55219	C1: 18106
Minimum	0	C2: 25188	C2: 29547	C2: 20021
Median	0	C3: 37014	C3: 61363	C3: 108002
Mean	0.4029	-	-	-
Variance	0.5811	-	-	-
Maximum	12	-	-	-

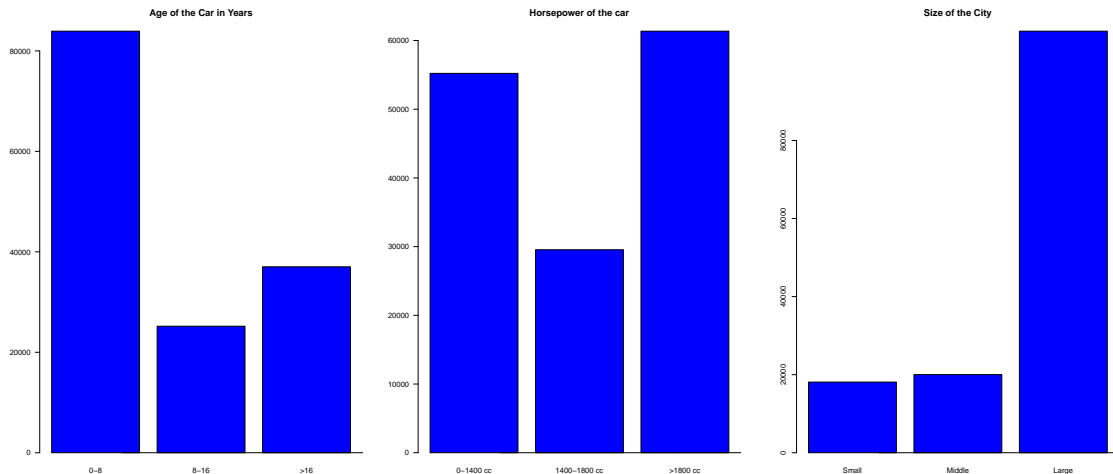


Figure 1. Descriptive Histograms for the Explanatory Variables

Finally, Table 2 presents a summary of the effects of the covariates on claim counts for 27 different risk classes, which can be formed by dividing the portfolio into clusters defined by the combinations of the characteristics of the policyholders and their cars based on all 146129 observations. In particular, Table 2 shows the percentage of observations with claim counts equal to 0, 1, 2, 3 and higher than or equal to 4 for each of the 27 groups of policyholders.

**Table 2** Summary Statistics of the Claim frequencies of the Different Risk Classes Determined by the Combination of the Explanatory Variables

Counts	Age of the car: C1			Age of the car: C1			Age of the car: C1		
	Horsepower of the Car: C1			Horsepower of the Car: C2			Horsepower of the Car: C3		
	Size of the City:			Size of the City:			Size of the City:		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
0	69.72	67.28	63.61	70.47	70.01	65.37	69.69	65.85	61.80
1	21.82	24.25	26.47	20.98	20.32	24.39	21.31	24.87	27.90
2	5.96	5.82	6.95	5.67	6.12	6.89	5.84	6.91	7.38
3	1.81	1.86	2.06	2.10	2.33	2.20	2.08	1.62	2.07
$\geq 4$	0.69	0.79	0.91	0.78	1.22	1.15	1.08	0.75	0.85

Counts	Age of the car: C2			Age of the car: C2			Age of the car: C2		
	Horsepower of the Car: C1			Horsepower of the Car: C2			Horsepower of the Car: C3		
	Size of the City:			Size of the City:			Size of the City:		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
0	78.49	76.79	74.51	81.68	79.88	76.46	78.31	77.61	74.70
1	17.31	19.22	20.29	13.35	14.79	17.33	17.05	16.82	19.19
2	3.23	2.84	4.02	2.48	3.55	4.43	3.55	4.16	4.65
3	0.79	0.74	0.89	1.55	1.18	1.38	0.68	0.97	1.05
$\geq 4$	0.18	0.41	0.29	0.94	0.60	0.40	0.41	0.44	0.41

Counts	Age of the car: C3			Age of the car: C3			Age of the car: C3		
	Horsepower of the Car: C1			Horsepower of the Car: C2			Horsepower of the Car: C3		
	Size of the City:			Size of the City:			Size of the City:		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
0	74.61	71.77	69.40	74.75	74.89	69.21	73.37	70.20	67.11
1	16.83	19.78	20.58	17.97	16.45	21.44	18.24	21.41	23.34
2	5.79	5.94	7.03	4.28	5.41	6.46	5.64	6.00	6.52
3	1.97	1.66	1.94	2.00	1.95	1.75	1.83	1.52	2.06
$\geq 4$	0.80	0.85	1.05	1.00	1.30	1.14	0.92	0.87	0.97

In what follows we compare the fit the Negative Binomial-Inverse Gaussian (NBIG) distribution/regression model with the traditional Negative Binomial Type I (NBI) and Poisson-Inverse Gaussian (PIG) distributions/regression models taken, for instance, from Dionne and Vanasse (1989 and 1992) and Trembley (1992) and Tzougas et al. (2017) respectively.

#### 4.1 Modelling Results

This subsection describes the modelling results of the NBI, PIG and NBIG distributions/regression models that have been applied to model claim frequency<sup>2</sup>. The ML estimators of the parameters for the frequency distributions are presented in Table 3.

**Table 3** Results of the Fitted NBI, PIG and NBIG Distributions

NBI	PIG	NBIG
$\mu$	$\mu$	$\mu$
0.4029	0.4029	0.4029
$\sigma$	$\sigma$	$\sigma$
1.0285	1.1045	1.9695
-	-	$\gamma$
-	-	1.5878

Let us now consider the NBI, PIG and NBIG regression models for approximating the number of claims. Table 4 reports our findings with respect to the aforementioned claim frequency regression models<sup>3</sup>.

<sup>2</sup> Note that the location and scale parameters of the NBI and PIG models are denoted by  $\mu$  and  $\sigma$  respectively.

<sup>3</sup> Note that all the explanatory variables and the parameters of the models are statistically significant at a 5% threshold.

**Table 4** Results of the Fitted NBI, PIG and NBIG Regression Models

NBI		PIG		NBIG	
Variable	Coeff. $\beta$	Variable	Coeff. $\beta$	Variable	Coeff. $\beta$
Intercept	-0.8602	Intercept	-0.8605	Intercept	-0.8600
CS		CS		CS	
C2	0.0659	C2	0.0672	C2	0.0674
C3	0.1784	C3	0.1782	C3	0.1796
AC		AC		AC	
C2	-0.4518	C2	-0.4537	C2	-0.4521
C3	-0.1131	C3	-0.1124	C3	-0.1127
HP		HP		HP	
C2	-0.0030	C2	-0.0024	C2	-0.0058
C3	0.0434	C3	0.0400	C3	0.0400
Variable		Variable		Variable	
$\sigma$	0.9867	$\sigma$	1.0579	$\sigma$	2.0659
-	-	-	-	Variable	
-	-	-	-	$\gamma$	1.6066

Finally, we rely on normalized quantile residuals (see, Dunn and Smyth, 1996) as an exploratory graphical device for investigating the adequacy of the fit of the competing NBI, PIG and NBIG distributions. For these discrete response distributions, the normalized randomized quantile residuals are defined as  $\hat{r}_i = \Phi^{-1}(u_i)$ , where  $\Phi^{-1}$  is the inverse cumulative distribution function of a standard Normal distribution and where  $u_i$  is defined as a random value from the uniform distribution on the interval  $[F_i(x_i - 1|\hat{\vartheta}), F_i(x_i|\hat{\vartheta})]$ , where  $F_i$  is the cumulative distribution function estimated for the  $i$ th individual and where  $\hat{\vartheta}$  contains all estimated model parameters and  $x_i$  is the corresponding observation. The model fit can be evaluated by means of usual quantile-quantile plots. Specifically, if the data indeed follow the assumed distribution, then the residual on the quantile-quantile plot will fall approximately on a straight line. Figure 2 shows the normalized (random) quantiles for the NBI, PIG and NBIG claim frequency regression models.

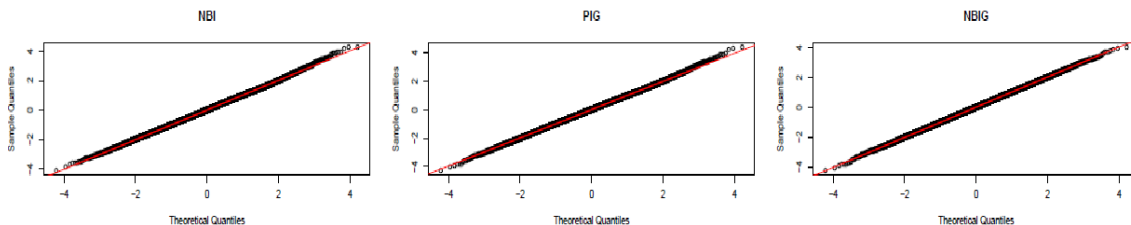


Figure 2. Normalized Quantiles for the NBI, PIG and NBIG Regression Models

From Figure 2 we observe that the residuals of the NBI, PIG and NBIG claim frequency models are very close to the diagonal and indicate a very good fit to the distribution of the claim frequencies.

#### 4.2 Models Comparison

In this subsection we compare the fit of the NBI, PIG and NBIG models for the observed claim frequencies in the MTPL insurance portfolio we analyzed earlier employing Global Deviance (DEV), Akaike information criterion (AIC)<sup>4</sup>, the Schwartz Bayesian criterion (SBC) and the Vuong test (see Vuong, 1989). The models were also calibrated with respect to the Consistent Akaike Information Criteria (CAIC), which was proposed by Bozdogan (1987) and can compensate for the overestimating nature of AIC. Table 5 (Panels A and B) reports our results for the different claim frequency distributions/regression models. Overall, from Panels A and B we observe that the best fit is given by the NBIG distribution/regression model.

<sup>4</sup> We also used we used three fourths of the data set to estimate the parameters of the models and the remaining one fourth was used to test the out-of-sample prediction accuracy of the models. As expected, our findings were consistent with those provided by the AIC criterion. For more details, refer to Stone (1977) who showed that AIC and leave-one out cross validation are asymptotically equivalent.



**Table 5** Models Comparison

Panel A: Based on Global Deviance, AIC, SBC and CAIC										
Distributions					Regression Models					
Model	df	AIC	SBC	CAIC	Model	df	Global Deviance	AIC	SBC	CAIC
NBI	2	232009	232029	232031	NBI	8	230966	230982	231061	231069
PIG	2	231935	231955	231957	PIG	8	230879	230895	230974	230982
NBIG	3	231795	231825	231828	NBIG	9	230735	230753	230842	230851

Panel B: Based on Vuong test									
Distributions					Regression Models				
Model 1	Model 2	Vuong Test Statistic	p-value	Decision	Model 1	Model 2	Vuong Test Statistic	p-value	Decision
PIG	NBI	-3.15	0.00	PIG	NBI	PIG	77.03	0.00	NBI
PIG	NBIG	-2.90	0.00	NBIG	NBIG	NBI	-6.12	0.00	NBIG

### 4.3 Calculation of the A Priori Premiums

In this subsection, based on the use of the net premium calculation principle, we analyze the differences between the claim frequency regression models through the mean of the number of claims of the policyholders who belong to the 27 different risk classes, which are determined by the relevant a priori characteristics. Note that in the case of the NBIG model, which does not have a pmf in closed form, the estimated expected annual claim frequency for each risk class was calculated in Eq. (6) using the law of total expectation. The a priori premium rates resulting from the NBI, PIG and NBIG models are presented in Table 6. For example, the expected claim frequency for a policyholder who lives in a small city and has a car that belongs to AC category 1 and HP category 1, i.e. for the reference class, is equal to 0.1209, 0.1208 and 0.1209, while the expected claim frequency for an insured who lives in a large city and has a car with similar characteristics, i.e. for risk class 3, is equal to 0.1445, 0.1444 and 0.1447 in the case of the NBI, PIG and NBIG model respectively. Moreover, we observe that the group of policyholders with the lowest expected claim frequency are those who live in a small city and have a car that belongs to AC category 2 and HP category 2, i.e. risk class 13. On the other hand, the group of policyholders with the highest expected claim frequency are those who live in a large city and have a car that belongs to AC category 1 and HP category 3, i.e. risk class 9.

**Table 6** A Priori Premium Rates, Claim Frequency Regression Models

	Risk Class	NBI	PIG	NBIG
1	<i>ACC1, HPC1, CSC1</i>	0.1209	0.1208	0.1209
2	<i>ACC1, HPC1, CSC2</i>	0.1291	0.1292	0.1293
3	<i>ACC1, HPC1, CSC3</i>	0.1445	0.1444	0.1447
4	<i>ACC1, HPC2, CSC1</i>	0.1205	0.1206	0.1202
5	<i>ACC1, HPC2, CSC2</i>	0.1287	0.1289	0.1286
6	<i>ACC1, HPC2, CSC3</i>	0.1441	0.1441	0.1438
7	<i>ACC1, HPC3, CSC1</i>	0.1262	0.1258	0.1258
8	<i>ACC1, HPC3, CSC2</i>	0.1348	0.1345	0.1346
9	<i>ACC1, HPC3, CSC3</i>	0.1509	0.1503	0.1506
10	<i>ACC2, HPC1, CSC1</i>	0.0769	0.0768	0.0769
11	<i>ACC2, HPC1, CSC2</i>	0.0822	0.0821	0.0823
12	<i>ACC2, HPC1, CSC3</i>	0.0920	0.0917	0.0921
13	<i>ACC2, HPC2, CSC1</i>	0.0767	0.0766	0.0765
14	<i>ACC2, HPC2, CSC2</i>	0.0819	0.0819	0.0818
15	<i>ACC2, HPC2, CSC3</i>	0.0917	0.0915	0.0915
16	<i>ACC2, HPC3, CSC1</i>	0.0804	0.0799	0.0801
17	<i>ACC2, HPC3, CSC2</i>	0.0858	0.0855	0.0857
18	<i>ACC2, HPC3, CSC3</i>	0.0960	0.0955	0.0958
19	<i>ACC3, HPC1, CSC1</i>	0.1080	0.1080	0.1080
20	<i>ACC3, HPC1, CSC2</i>	0.1153	0.1155	0.1156
21	<i>ACC3, HPC1, CSC3</i>	0.1291	0.1291	0.1293
22	<i>ACC3, HPC2, CSC1</i>	0.1076	0.1077	0.1074
23	<i>ACC3, HPC2, CSC2</i>	0.1150	0.1152	0.1149
24	<i>ACC3, HPC2, CSC3</i>	0.1287	0.1288	0.1285
25	<i>ACC3, HPC3, CSC1</i>	0.1127	0.1124	0.1124
26	<i>ACC3, HPC3, CSC2</i>	0.1204	0.1202	0.1203
27	<i>ACC3, HPC3, CSC3</i>	0.1348	0.1343	0.1345

Overall, as expected, Table 6 shows that small differences lie in the a priori premiums resulting from these models when the net premium principle is used because, as is well known, in this case only the location

parameters of the models, which are modelled using the same covariate information, affect the estimation of the premium rates. However, when the a posteriori correction is going to be calculated we will observe that, compared to the BMS provided by mixed Poisson regression models, the NBIG model will show much less extreme relative a posteriori, or Bonus-Malus premiums for policyholders with some claim experience. This can be explained by the fact that the NBIG regression model has the potential to capture more efficiently the stylized characteristics of the data. In particular, given the estimate of the parameter  $\sigma$  of the NBIG model, which is the same parameter that appears in Eq. (1) of the Negative Binomial distribution, we can assess the extent of overdispersion for the claim distribution of an individual policyholder with any given mean claim rate. The NBIG model assumes that individual claim experience will be overdispersed relative to a Poisson model and the overdispersion is larger for policyholders with larger mean claim rates. Consequently, this situation affects the calculation of the a posteriori or Bonus-Malus premium rates. Thus, the employment of the new model, which also provided a better fit than the other two competing mixed Poisson models, is beneficial for the insurance company because it can provide the actuary with alternative pricing strategies in addition to those already existing in ratemaking literature.

#### 4.4 Calculation of the A Posteriori Premiums

In this subsection we examine how the NBIG model responds to claim experience. Consider a policyholder  $i$  with claim history  $k_i^1, \dots, k_i^t$  and  $x_i^1, \dots, x_i^{t+1}$  characteristics and denote by  $K = \sum_{j=1}^t k_i^j$  the total number of claims that they had. The problem is to determine at the renewal of the policy the expected claim frequency  $\lambda_i^{t+1}$  of the policyholder  $i$  for the period  $t+1$  given the observation of the reported accidents in the preceding  $t$  periods and observable characteristics in the preceding  $t+1$  periods and the current period. We use Bayes theorem to derive the posterior distribution of  $\lambda_i^{t+1}$ , given those claim and characteristics records, as follows

$$f\left(\lambda_i^{t+1} | k_i^1, \dots, k_i^t; x_i^1, \dots, x_i^{t+1}\right) = \frac{\prod_{j=1}^t P(k_i^j | \mathbf{x}_i, \lambda_i) f(\lambda_i^{t+1})}{\int_0^\infty P(k_i | \mathbf{x}_i, \lambda_i) f(\lambda_i^{t+1}) d\lambda_i^{t+1}}, \quad (26)$$

where  $P(k_i | \mathbf{x}_i, \lambda_i)$  is the pmf of the Negative Binomial distribution, which is given by Eq. (1), and where  $f(\lambda_i^{t+1})$  is the pdf of the Inverse Gaussian prior distribution which is given by Eq. (4). Using the quadratic loss function and the net premium principle, one can find that the optimal estimator of  $\lambda_i^{t+1}$  is the mean of the posterior structure function given by

$$E\left(\lambda_i^{t+1} | k_i^1, \dots, k_i^t; x_i^1, \dots, x_i^{t+1}\right) = \int_0^\infty \lambda_i^{t+1} f\left(\lambda_i^{t+1} | k_i^1, \dots, k_i^t; x_i^1, \dots, x_i^{t+1}\right) d\lambda_i^{t+1}, \quad (27)$$

The expectation in Eq. (27) does not have a closed form expression. However, it can be easily computed based on either numerical integration or a Monte Carlo approach since both schemes do not require knowledge of the pdf of the posterior distribution of  $\lambda_i^{t+1}$ .

Based on the aforementioned methodology, we compute the Bonus-Malus premium rates resulting from the NBIG model based only on the number of individual claims, i.e. the a posteriori criteria, and based both on the number of individual claims and the characteristics of the policyholder and the automobile, i.e. the a priori criteria. When both criteria are considered, we examine a group of policyholders who share the following common characteristics: We consider that the policyholder  $i$  lives in a large city and has a car between 0 to 8 years old with HP between 0-1400 cc. 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. Tables 7 and 8 show comparable relative premiums for the NBI, PIG and NBIG distributions/regression models respectively, assuming that the number of claims ranges from 0 to 4 and the age of the policy is up to 5 years. From both of tables 7 and 8 we observe that if the policyholder  $i$  has a claim free year, the premium rates reduce, whereas if they have one or more claims, the premium rates increase, resulting in bonus or malus respectively. Furthermore, as expected, both of these tables show that for policyholders with some claim experience the NBIG model provides lower relative premiums rates than those resulting from both mixed Poisson models. As was previously mentioned, this difference between the NBIG model and the NBI and PIG models can be attributed to the fact the NBIG model assumes that the individual claims follow the overdispersed Negative Binomial distribution whereas, in contrast, both the NBI and PIG models assume

that the individual claims follow a Poisson distribution, which does not allow for overdispersion. Thus, lower relative Bonus-Malus premiums are more likely under the NBIG distribution/regression model than those under one based on a mixed Poisson distribution/regression model. For example, from Table 7 we see that policyholders who had four claims over the third year of observation will have to pay a malus of 277.37%, 355.37% and 136.94% of the basic premium, while those who had four claims over the fifth year of observation will have to pay a malus of 221.24%, 256.88% and 115.19% of the basic premium in the case of the NBI, PIG and NBIG distributions respectively. Furthermore, from Table 8 when both the a priori and the a posteriori criteria are considered, we see, for instance, that policyholders had four claims over the third year of observation will have to pay a malus of 246.48%, 303.69% and 122.90% of the basic premium and those who had four claims over the fifth year of observation will have to pay a malus of 188.80%, 209.91% and 99.77% of the basic premium in the case of the NBI, PIG and NBIG regression models respectively. Overall, it is worth mentioning that the Bonus-Malus premium rates which are presented in Table 8 provide a more complete picture to the actuary since since they consider all the important a priori and a posteriori information for the number of claims of each policyholder in order to estimate their risk of having an accident and thus they permit the differentiation of the premiums for various number of claims based on the expected claim frequency of each policyholder as this is estimated both from the a priori and the a posteriori classification criteria.

**Table 7** A Posteriori, or Bonus-Malus, Premium Rates, Claim Frequency Distributions

NBI					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	89.4145	181.3762	273.3380	365.2998	457.2615
2	80.8555	164.0145	247.1734	330.3324	413.4914
3	73.7920	149.6862	225.5804	301.4745	377.3687
4	67.8634	137.6602	207.4569	277.2536	347.0504
5	62.8166	127.4228	192.0290	256.6352	321.2414
PIG					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	89.2901	177.3503	309.1353	466.0915	633.5270
2	81.4178	154.6349	262.5190	391.3363	529.4585
3	75.3173	137.9733	229.0824	338.0425	455.3728
4	70.4102	125.1677	203.8802	298.1039	399.9333
5	66.3521	114.9796	184.1729	267.0425	356.8794
NBIG					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	96.0443	129.5972	170.1744	215.6181	263.6407
2	92.4986	123.8452	161.7446	204.3245	249.5761
3	89.2996	118.6957	154.2130	194.2249	236.9355
4	86.3963	114.0578	147.4467	185.1409	225.5256
5	83.7473	109.8579	141.3367	176.9333	215.1931

**Table 8** A Posteriori, or Bonus-Malus, Premium Rates, Claim Frequency Regression Models

NBI					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	87.5220	173.8774	260.2327	346.5881	432.9434
2	77.8126	154.5880	231.3633	308.1387	384.9140
3	70.0423	139.1510	208.2596	277.3683	346.4769
4	63.6830	126.5171	189.3512	252.1853	315.0194
5	58.3823	115.9864	173.5905	231.1946	288.7987

PIG					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	87.5187	168.5516	288.5418	431.7100	584.9724
2	78.7834	144.4478	239.9623	354.1874	477.1743
3	72.2310	127.4270	206.5314	301.2413	403.6909
4	67.0815	114.6879	182.0555	262.7493	350.3711
5	62.8966	104.7483	163.3217	233.4806	309.9055

NBIG					
Year	Number of Claims $k$				
$t$	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	95.2674	127.0983	165.1884	207.5591	252.1951
2	91.1096	120.5022	155.6831	194.9866	236.6477
3	87.4240	114.7025	147.3403	183.9234	222.8985
4	84.1309	109.5625	139.9645	174.1277	210.6809
5	81.1677	104.9742	133.4002	165.4031	199.7682

## 5 Other Topics

### 5.1 Computational Aspects

The EM algorithm described in Section 3 was used to estimate both the NBIG distribution and regression model. The model, both for the case with and the case without covariate information, converged after quite a few iterations using a rather strict criterion. In particular, we iterated between the E-step and the M-step until the relative change in log-likelihood, which is given by Eq. (8), between two successive iterations was smaller than  $10^{-12}$ . We also emphasize that for this model the choice of initial values for both the vector of the regression coefficients  $\beta$  and the scale parameter  $\sigma$  needed special attention because the M-step is in fact a Newton-Raphson iteration and one may obtain inadmissible values if the starting values are bad. Good starting values for  $\beta$  and  $\sigma$  were obtained by fitting the simple negative binomial regression. Additionally, to ensure that the global maximum had been obtained and the algorithm had not been trapped in a local maximum, we checked with many initial values for the parameter  $\gamma$  of the Inverse Gaussian mixing distribution, but for all cases we converged on the same solution. Furthermore, standard errors were obtained relatively easily using the standard approach of Louis (1982). All computing was made using the statistical computing environment language R. The NBIG regression model needed more time than the NBIG distribution because the numerical evaluation of the integrals for the case with covariates is more computationally time consuming than for the case without covariates. However, it took less than 3 minutes on an Intel Duo 2Ghz processor. Finally, ML estimation of both NBI and PIG distributions/regression models, for which the definition of a density function in closed form is feasible, was straightforward by using standard statistical packages in R.

### 5.2 An Extended Version of the Model

Finally, as is well known, claim count data sets, often include a high presence of zeros. The high percentage of zero values motivates the use of zero-inflated models which can provide a parsimonious yet powerful way to handle data sets that contain a large number of zeros. See, for instance, Cohen (1966), Lambert (1992), Yip and Yau (2005), Boucher et al. (2007), Denuit et al. (2007), Tzougas et al. (2015) and Gómez-Déniz and Calderín-Ojeda (2016). Thus, an interesting possible line of further research would be to consider the Zero-inflated version of the proposed model, namely the Negative Binomial-Inverse Gaussian (ZINBIG) regression model. The ZINBIG distribution can arise as follows:

$$P(k_i = k | \mathbf{x}_i) = \begin{cases} \pi + (1 - \pi) P(k = 0 | \mathbf{x}_i) \\ (1 - \pi) P(k | \mathbf{x}_i), \text{ if } k > 0, \end{cases} \quad (28)$$

where  $0 < \pi \leq 1$ , for  $k_i, i = 1, \dots, n$ , is the response, where  $P(k_i | \mathbf{x}_i)$  is the pmf of Negative Binomial-Inverse Gaussian distribution, which is given by Eq. (5), and where  $\mathbf{x}_i$  is a vector of covariate information. Note also that the ZINBIG model is in fact a finite mixture model since assumes that the data are a mixture of two distributions: a degenerate distribution for the zero case and a standard count distribution. Thus, one can fit this general model without special effort using standard techniques for finite mixtures (see, Bohning, 1999).

## 6 Conclusions

The main purpose of this paper was to propose an EM scheme that reduces the computational burden for ML estimation in the NBIG regression model. The NBIG regression model extends the commonly used specification that assumes that the number of claims is distributed according to a mixed Poisson regression model, which was widely accepted for a priori and a posteriori ratemaking.

The NBIG distribution can be considered as a plausible model for overdispersed claim count data and thus is suitable for application, not only in insurance settings, but also in other fields where this phenomenon is present, as it is likely to capture its influence to a good approximation. Also, we compared the a priori and a posteriori premium rates determined by the NBIG model to those derived by the NBI and PIG models respectively that have been widely used for modelling claim count data.

Furthermore, it should be noted that the novel EM type algorithm we developed was based on the Poisson-Gamma-Inverse Gaussian triple mixture representation of the NBIG model, and did not require knowledge of its pmf, which could not be written in closed form. Thus, it is obvious that the ML estimation framework we proposed has the considerable mathematical flexibility for fitting other mixed Negative Binomial regression models stemming from several other mixing distributions. Moreover, the data augmentation which was used in the paper to derive the EM algorithm can be the basis for constructing Bayesian estimation methods proceeding along similar lines as Zhou et al. (2012) and Geedipally et al. (2012) who considered the Lognormal and Gamma mixed Negative Binomial and the Negative Binomial-Lindley regression models respectively for analyzing count data.

Finally, in a forthcoming paper bivariate mixed Negative Binomial regression models will be employed to design merit rating plans in accordance with the a priori ratemaking structure of the insurance company, extending the setup of Bermúdez and Karlis (2011 and 2017) who used bivariate mixed Poisson regression models to make an a priori and a posteriori ratemaking.

**Acknowledgements** The research reported here has been supported by LSE LIFE and LSE Teaching and Learning Centre. We would like to thank the two anonymous referees for their constructive comments and suggestions that have greatly improved the paper. We also would like to thank the participants at the 10th International Conference of the ERCIM WG on Computational and Methodological Statistics. The usual disclaimer applies.

## References

1. Bermúdez, L., and D. Karlis (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, 48(2), 226-236.
2. Bermúdez, L., and D. Karlis (2017). A posteriori ratemaking using bivariate Poisson models. *Scandinavian Actuarial Journal*, 2017(2), 148-158.
3. Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC). The general theory and its analytical extensions. *Psychometrika*, 52, 3, 345-370.
4. Bohning, D. (1999). *Computer Assisted Analysis of Mixtures and Applications in Meta-analysis, Disease Mapping and Others*. CRC Press, New York.
5. Booth, J. G., and J.P. Hobert (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.
6. Booth, J. G., J.P. Hobert and W. Jank. (2001). A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modelling*, 1(4), 333-349.
7. Boucher, J. P., M. Denuit and M. Guillen (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. *North American Actuarial Journal*, 11, 4, 110-131.
8. Boucher, J. P., M. Denuit and M. Guillen (2008). Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. *Variance*, 2, 1, 135-162.
9. Brouhns, N., M. Guillen, M. Denuit and J. Pinquet (2003). Bonus-malus scales in segmented tariffs with stochastic migration between segments. *Journal of Risk and Insurance*, 70, 577-599.
10. Cohen, A. C. (1966). A note on certain discrete mixed distributions. *Biometrics*, 22, 3, 566-572.

11. Dean, C., J.F. Lawless and G.E. Willmot (1989). A mixed Poisson-inverse-Gaussian regression model. *Canadian Journal of Statistics* 17 (2), 171-181.
12. De Jong, P. and G. Z. Heller (2008). *Generalized linear models for insurance data* (Vol. 10). Cambridge: Cambridge University Press.
13. Dempster, A. P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.
14. Denuit, M. and S. Lang (2004). Nonlife Ratemaking with Bayesian GAM's. *Insurance, Mathematics and Economics* 35: 627-47.
15. Denuit, M., X. Marechal, S. Pitrebois and J. F. Walhin (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Wiley.
16. Dionne, G. and C. Vanasse (1989). A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component. *ASTIN Bulletin*, 19, 199-212.
17. Dionne, G. and C. Vanasse (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics*, 7, 149-165.
18. Dunn, P.K. and G.K. Smyth (1996). Randomized quantile residuals. *Computational and Graphical Statistics* 5, 236-245.
19. Frangos, N. and S. Vrontos (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin*, 31, 1, 1-22.
20. Frees, E. W. (2010). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press: Cambridge, U.K.
21. Geedipally, S. R., D. Lord and S.S. Dhavala (2012). The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. *Accident Analysis & Prevention*, 45, 258-265.
22. Gençtürk, Y. and A. Yiğiter (2016). Modelling claim number using a new mixture model: negative binomial gamma distribution. *Journal of Statistical Computation and Simulation*, 86(10), 1829-1839.
23. Gómez-Déniz, E. and F. Vázquez (2003). Robustness in Bayesian model for Bonus-Malus systems. In: *Intelligent and Other Computation Techniques in Insurance. Theory and Applications*. World Scientific.
24. Gómez-Déniz, E., J.M. Sarabia and E. Calderín-Ojeda (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, 42(1), 39-49.
25. Gómez-Déniz, E., A. Hernández-Bastida and M.P. Fernández-Sánchez (2014). Computing credibility bonus-malus premiums using the aggregate claims distribution. *Hacetatepe Journal of Mathematics and Statistics*, 43, 6, 1047-1061.
26. Gómez-Déniz, E., and E. Calderín-Ojeda (2016). "The Mixture Poisson Exponential-Inverse Gaussian Regression Model: An application in Health Services." *Metodoloski Zvezki* 13.2 : 71.
27. Gómez-Déniz, E. and E. Calderín-Ojeda (2017). "An alternative representation of the negative binomial-Lindley distribution. New results and applications." arXiv preprint arXiv:1703.04812.
28. Haberman, S. and A. E. Renshaw (1996). Generalized linear models and actuarial science. *The Statistician*, 407-436.
29. Insurance Europe (2016). *European Insurance-Key Facts*. Link: <https://www.insuranceeurope.eu/sites/default/files/attachments/European%20Insurance%20-%20Key%20Facts%20-%20August%202016.pdf>
30. Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. In *Lecture Notes in Statistics*, 9. Springer-Verlag, New York.
31. Kaas, R., M. Goovaerts, J. Dhaene and M. Denuit (2008). *Modern Actuarial Risk Theory: Using R*. Springer-Verlag: Berlin Heidelberg, Second edition.
32. Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1(4), 305-318.
33. Karlis, D. (2005). EM algorithm for mixed Poisson and other discrete distributions. *ASTIN Bulletin: The Journal of the IAA*, 35(1), 3-24.
34. Klugman, S., H. Panjer and G. Willmot (1998). *Loss Models. From Data to Decisions*. John Wiley and Sons, New York.
35. Lambert, D. (1992). "Zero-inflated Poisson regression, with an application to defects in manufacturing." *Technometrics* 34.1: 1-14.
36. Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers.
37. Lemaire, J. S. C. Park and K. C. Wang (2015). The Impact of Covariates on a Bonus-Malus System: An Application of Taylor's Model. *European Actuarial Journal*, 5 (1), pp. 1-10.
38. Lemaire, J. S. C. Park and K. C. Wang (2016). The Use of Annual Mileage as a Rating Variable. *ASTIN Bulletin*, 46 (1), pp. 39-69.
39. Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society B*, 44, 226-233.
40. McLachlan, G. and T. Krishnan (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.
41. Picech, L. (1994). The Merit-Rating Factor in a Multiplicating Rate-Making model. *ASTIN Colloquium*, Cannes.
42. Pinquet, J. (1997). Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin*, 27, 33-57.
43. Pinquet, J. (1998). Designing Optimal Bonus-Malus Systems From Different Types of Claims. *ASTIN Bulletin*, 28, 205-220.
44. Seshardi, V. (1993). *The Inverse Gaussian Distribution*. Oxford Science Publications.
45. Shengwang, M., Y. Wei and G. A. Whitmore (1999). Accounting for individual overdispersion in a bonus-malus automobile insurance system. *ASTIN Bulletin: The Journal of the IAA*, 29(2), 327-337.
46. Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society Series B*. 39, 44-7.
47. Trembley, L. (1992). Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bulletin*, 22(1):97-106.
48. Tzougas, G., S. Vrontos and Frangos, N. (2014). Optimal Bonus-Malus Systems Using Finite Mixture Models. *ASTIN Bulletin*, Volume 44, Issue 02, May 2014, pp. 417-444.
49. Tzougas, G., S. Vrontos and Frangos, N. (2015). Risk Classification for Claim Counts and Losses Using Regression Models for Location, Scale and Shape. *Variance* 9:1, 2015, pp. 140-157.
50. Tzougas, G., S. Vrontos and Frangos, N. (2018). Bonus-Malus Systems with Two-Component Mixture Models Arising from Different Parametric Families. *North American Actuarial Journal*, 22:1, 55-91
51. Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-333.
52. Yip, K. and K. Yau (2005). On modeling Claim Frequency Data in General Insurance with Extra Zeros. *Insurance, Mathematics and Economics* 36: 153-63.

- 
53. Zamani, H. and N. Ismail (2010). Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics*, 6(1), 4-9.
  54. Zhou, M., L. Li, D. Dunson and L. Carin (2012). Lognormal and gamma mixed negative binomial regression. In *Proceedings of the... International Conference on Machine Learning*. International Conference on Machine Learning (Vol. 2012, p. 1343). NIH Public Access.