EM Estimation for the Poisson-Inverse Gamma Regression Model with Varying Dispersion: An Application to Insurance Ratemaking

George Tzougas

Department of Statistics, London School of Economics and Political Science, London WC2A 2AE, UK; g.tzougas@lse.ac.uk

Abstract: This article presents the Poisson-Inverse Gamma regression model with varying dispersion for approximating heavy-tailed and overdispersed claim counts. Our main contribution is that we develop an Expectation-Maximization (EM) type algorithm for maximum likelihood (ML) estimation of the Poisson-Inverse Gamma regression model with varying dispersion. The empirical analysis examines a portfolio of motor insurance data in order to investigate the efficiency of the proposed algorithm. Finally, both the a priori and a posteriori, or Bonus-Malus, premium rates that are determined by the Poisson-Inverse Gamma model are compared to those that result from the classic Negative Binomial Type I and the Poisson-Inverse Gaussian distributions with regression structures for their mean and dispersion parameters.

Keywords: poisson-inverse gamma distribution; em algorithm; regression models for mean and dispersion parameters; motor third party liability insurance; ratemaking

1. Introduction

Over the last few decades, it has become evident that the most important consequence of unobserved heterogeneity in the regression analysis of count data is overdispersion, i.e., a situation that violates the Poisson hallmark assumption of equidispersion, since the variance of the response variable exceeds the mean. The inadequacy of the Poisson regression model to accommodate overdispersion resulted in the proliferation of the use of mixed Poisson regression models for studying count regression type problems in many disciplines, such as, for example, sociology, econometrics, manufacturing, engineering, agriculture, biology, biometrics, genetics, medicine, sports, marketing, and insurance. The Negative Binomial Type I (NBI), or Poisson-Gamma, and Poisson-Inverse Gaussian (PIG) have been traditionally employed for modelling count data primarily due to the simplicity of their log-likelihood function that implies that the formality of parameter estimation via easy to implement Maximum Likelihood (ML) estimation procedures is straightforward. See, for instance, Lawless (1987), Cameron and Trivedi (1998) and Hilbe (2008) regarding the former and Ord and Whitmore (1986), Willmot (1987), and Dean et al. (1989) for the latter. Furthermore, alternative mixed Poisson regression models have been proposed for handling different levels of overdispersion, even if the literature on these models is not as abundant as for the NBI and PIG models due to algebraic intractability or because their densities involve special functions and appropriate numerical methods are required for their maximum likelihood (ML) estimation, such as, for example, the Poisson-Lognormal (PLN) regression model, see Denuit et al. (2007) and Boucher et al. (2007), the Poisson Exponential-Inverse Gaussian (PEIG) regression model, see Gómez-Déniz and Calderín-Ojeda (2016), the Poisson-mixed Inverse Gaussian (PMIG) distribution, where the mixed Inverse Gaussian distribution is a mixture of the Inverse Gaussian

distribution and its length-biased counterpart, see Gómez-Déniz et al. (2016), the Poisson-reciprocal Inverse Gaussian regression model, see Gómez-Déniz and Calderín-Ojeda (2018a), and the Poisson quasi-Lindley (PQL) regression model, see Altun (2019), among other models. Additionally, it should be noted that there are alternative approaches for enriching the classic model of claim counts, see, for instance, Majeske (2007), Giuricich and Burnecki (2019), and Romaniuk (2020).

Regarding the ratemaking process in Motor Third Party Liability (MTPL) insurance, which the present study is mainly concerned with, mixed Poisson regression models, with the NBI and PIG undoubtedly being the most popular choices, have been widely used for constructing a priori and a posteriori ratemaking schemes, or Bonus-Malus Systems (BMSs) for the frequency of claims. References for the former process can be found, for instance, in Haberman and Renshaw (1996), Denuit and Lang (2004), Boucher et al. (2007), De Jong and Heller (2008), Kaas et al. (2008), Frees (2010), and Tzougas et al. (2015). An excellent account of the latter process can be found in Lemaire (1995). The interested reader can also refer to the articles by, Dionne and Vanasse (1989, 1992), Trembley (1992), Picech (1994), Pinquet (1997, 1998), Frangos and Vrontos (2001), Brouhns et al. (2003), Gómez-Déniz and Vázquez-Polo (2003), Mert and Saykan (2005), Denuit et al. (2007), Boucher et al. (2008), Gómez-Déniz et al. (2008), Mahmoudvand and Hassani (2009), Schiegl (2010), Ni et al. (2014), Gómez-Déniz et al. (2014), Ni et al. (2014), Lemaire et al. (2015, 2016), Karlis et al. (2018), Gómez-Déniz and Calderín-Ojeda (2018b), Tzougas et al. (2014, 2018), and Tzougas et al. (2019).

In this work, the Poisson-Inverse Gamma (PIGA) regression model with varying dispersion is introduced for deriving ratemaking mechanisms for heavy-tailed and overdispersed claim counts. The probability mass function (pmf) of the model is parameterized in terms of its mean and its dispersion parameters. This results in an easier interpretation when both parameters are modelled while using covariate information and facilitates ML estimation due to its more orthogonal parameterization. A detailed discussion of our contribution relative to the existing literature concerning MTPL claim count data that are characterized by their long tails and overdispersion follows below.

Firstly, it is worth noting that the PIGA distribution was first considered by Willmot (1993) and a Bayesian estimation method for the PIGA regression model was proposed by Khazraee et al. (2018). However, unlike the Bayesian estimation approach, using the ML estimation procedure for the PIGA regression model within the traditional frequentist approach is far from straightforward and, to our knowledge, has not been explored in the literature so far. The main obstacle for finding the ML estimates of the parameters of the PIGA regression model is that it has a complicated likelihood function that is expressed in terms of the modified Bessel function of the third kind, see Abramowitz and Stegun (1965, p. 374) as well as Section 2, and hence its maximization needs a special effort. Moreover, under our general approach, where both the mean and dispersion parameters of the model are allowed to vary through explanatory variables, the computational complexity increases even further, since regression structures are incorporated in the order and argument of the modified Bessel function of the third kind. The main achievement of this work is that it demonstrates that ML estimates of the PIGA regression model with varying dispersion can be obtained in an easy manner by employing an Expectation-Maximization (EM) type algorithm that exploits the stochastic mixture representation of the PIGA model to reduce the problem of maximizing its complicated likelihood function to the simpler problem of maximizing the likelihood function of its mixing distribution. Moreover, the EM type scheme we propose is easily programmable and can remedy the computational issues, which may occur by alternative estimation procedures. At this point, we would like to emphasize that the development of ML estimation algorithms for modelling jointly all of the parameters of discrete and continuous response distributions in terms of covariates has not been thoroughly addressed both in the statistical and actuarial literature. Regarding the former, notable exceptions are the papers by Rigby and Stasinopoulos (2005) and Barreto-Souza and Simas (2015). In particular, Rigby and Stasinopoulos (2005) developed the generalized additive models for location, scale and shape (GAMLSS). Within the GAMLSS framework, every parameter of the discrete and continuous response distribution can be modelled as parametric and/or as additive nonparametric functions

of explanatory variables and/or random-effects terms. Moreover, the GAMLSS class of models extends the setup of many well known distributions, such as the NBI and PIG distributions, their generalizations, such as the Sichel, Delaporte and Poisson-shifted generalized inverse Gaussian (PSGIG) distributions, and also some of their zero-inflated versions for handling data sets that contain a large number of zeros, see Rigby et al. (2008a). The GAMLSS model can be fitted using either the RS algorithm, which is based on the algorithm of Rigby and Stasinopoulos (1996a, 1996b), or the CG algorithm, which is based on the algorithm of Cole and Green (1992). Additionally, Barreto-Souza and Simas (2015) implemented the EM algorithm for estimating the parameters of the general class of mixed Poisson regression models with varying dispersion, which they proposed extending the work of Karlis (2001) who considered the case when only the mean is modelled in terms of covariates. In their empirical illustration, they paid special attention to the estimation of the NBI and PIG regression models with regression specifications on their mean and dispersion parameters. Regarding the latter, an EM type scheme for estimating the parameters of mixed Exponential regression models with varying dispersion that can efficiently approximate heavy-tailed losses in non-life insurance was recently developed by Tzougas and Karlis (2020). However, this is the first time that the EM algorithm is used in a statistical or actuarial setting for estimating the PIGA regression model with varying dispersion.

Secondly, the ability of any claim count model for capturing the influence of the overdispersion phenomenon in real insurance data to a good approximation should always be investigated since there are many factors that may lead to extra variation occurring in the frequency of claims, which is an important measure of the risk exposure of a policy and, hence, directly affects how insurers price the policy. This can be clearly understood, since, in a real-world situation, the occurrence of an accident is a multifaceted event involving circumstances, such as, for instance, demographics, terrain, and exposure to weather conditions and differences among policyholders, which cannot be observed by the actuary, such as, for example, different perceptions and attitudes to compulsory MTPL insurance obligations and different driving skills and habits. Moreover, according to a recent report by Insurance Europe, as these factors differ widely from one country to another, vast differences between EU member states were observed in the frequency of MTPL claims, ranging from 2.4% in Finland to 7.5% in Turkey in 2016, see Insurance Europe (2019). Furthermore, as the empirical evidence has shown, and as it can also be verified by Shared's two crossings theorem for mixed Poisson models, see Shared (1980), the overdispersion phenomenon can be attributed to the excess of zeros and/or heavy upper tails in count data. The overdispersion in claim frequencies which is related to an excess of zeros can be handled well by zero-inflated models, see, for example, 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). On the other hand, it would not be realistic to assume that the overdispersion which is caused by a heavy tail in the claim count data will always be efficiently captured by a specific member of the mixed Poisson family of models. However, failing to account for overdispersion yields biased and inconsistent parameter estimates, which, in turn, can cause the actuary to make erroneous inferences from models and leads to inaccurate ratemaking. Nevertheless, because of the common structure of all members of the mixed Poisson family of models, there is a clear link between how a particular model is expected to perform for data sets with certain characteristics and the shape of their mixing densities. The Inverse Gamma distribution has a low probability in the vicinity of zero and a thick tail that is representative of greater probabilities for high values in its right tail. Thus, even if it is difficult to predict with certainty how a specified model will perform before it is fitted to the actual data, due to the thickness of the tail of the Inverse Gamma mixing distribution, the advantage that the PIGA model might enjoy over mixed Poisson models stemming from less heavy-tailed mixing distributions, such as the classic NBI and PIG models, is that it has a more promising shape for accommodating overdispersed claim counts with a long tail and, hence, it can improve ratemaking when dealing with this type of data.

Finally, as is well known, in most actuarial applications concerning two parameter mixed Poisson distributions employed for modelling claim counts it is commonly assumed that only the mean claim frequency is modelled in terms of covariates. Therefore, the skewness of the mixed Poisson distribution, which, in general, depends on both the mean and dispersion parameters, is not modelled explicitly as a function of explanatory variables, but implicitly through their dependence on the mean parameter. Nevertheless, since, in practical situations, the assumption of constant dispersion is not valid, only modelling the mean parameter in terms of risk factors can lead to a misclassification of policyholders with a high number of claims due to the unobserved heterogeneity changes with covariates. Moreover, because the posterior claim frequency distribution is expressed in terms of both the mean and dispersion parameters assuming that the latter parameter does not vary through covariates can have an impact on determining the appropriate level of the a posteriori, or Bonus-Malus, premiums, which, in turn, can lead to financial implications for the company, since, if the punishment of all insureds is not justified on a sound risk measuring basis, then policyholders may change to competing companies with a better risk adjusted pricing system. As a solution to the aforementioned problems, we allow for regressors on both the mean and the dispersion parameters of the heavy-tailed PIGA distribution. In particular, taking into consideration that the PIGA distribution can be viewed as an overdispersed Poisson random variable, with the extra variation in the count data controlled by the value of its dispersion parameter, allowing for the dispersion parameter to be modelled in terms covariates will result in adding the appropriate amount of weight to the right tail of its pmf which corresponds to high claim frequencies. As we will observe in our numerical illustration, for the data set used in this study, the employment of the PIGA distribution combined with the proposed modelling framework in an experience ratemaking scheme is beneficial for the company, as it will result in an improved risk evaluation of policyholders who are more likely to have accidents, establishing fair a posteriori, or Bonus-Malus, premiums, and mitigating adverse selection.

The remainder of this article is organized, as follows: Section 2 deals with the construction of the PIGA regression model with varying dispersion. In Section 3, we describe the ML estimation via the EM algorithm. A real data application based on MTPL data is presented in Section 4. In Section 5, we comment on the computational issues concerning the use of the EM algorithm for fitting the PIGA regression model with varying dispersion. Finally, concluding remarks can be found in Section 6.

2. The Poisson-Inverse Gamma Regression Model with Varying Dispersion

The Poisson-Inverse Gamma (PIGA) regression model with varying dispersion, which we present in this article, can be derived as follows. Consider that the individual claim frequencies, k_i, arising from a policyholder i, $i = 1, ..., n$, are independent and assume that given a continuous random variable $\lambda_i > 0$, $k_i|, \lambda_i$ follows a Poisson distribution with probability mass function (pmf) given by

$$P\left(k_i|\lambda_i\right) = \frac{e^{-\lambda_i \mu_i}\left(\lambda_i \mu_i\right)^{k_i}}{k_i!}, \tag{1}$$

for $k = 0, 1, 2, ...$, where $\mu_i > 0$, and where $E\left(k_i|\lambda_i\right) = Var\left(k_i|\lambda_i\right) = \lambda_i \mu_i$.

Furthermore, consider that λ_i is distributed according to an Inverse Gamma distribution with probability density function (pdf), given by

$$f\left(\lambda_i; \phi_i\right) = \frac{\phi_i^{\phi_i+1}}{\Gamma\left(\phi_i+1\right)}\lambda_i^{-\phi_i-2}\exp\left(-\frac{\phi_i}{\lambda_i}\right), \tag{2}$$

where $\lambda_i > 0$, where $E(\lambda_i) = 1$ and where $Var(\lambda_i) = \frac{1}{(\phi_i-1)}$, for $\phi_i > 1$ The Inverse Gamma prior distribution which is given by Equation (2) has to have a unit mean in order for the model to be identifiable. Note that the Inverse Gamma distribution nests some well known distributions such as the Inverse Exponential, Inverse Chi Squared and Scaled Inverse Chi Squared distributions. Furthermore,

it should be noted that the Inverse Gamma distribution is a limiting case of the Generalized Inverse Gaussian (GIG) family of distributions, see Johnson et al. (1994) and Jørgensen (2012).

When considering the assumptions in Equations (1) and (2), it is easy to see that the unconditional distribution of k_i will be a PIGA distribution with pmf given by

$$
\begin{aligned}
P\left(k_i\right) &= \int_0^\infty P\left(k_i | \lambda_i\right) f\left(\lambda_i\right) d\lambda_i \\
&= \frac{2}{k_i!} \frac{\left(\mu_i \phi_i\right)^{\frac{k_i+\phi_i+1}{2}}}{\Gamma\left(\phi_i+1\right)} K_{k_i-\phi_i-1}\left(2\sqrt{\mu_i \phi_i}\right),
\end{aligned}
\tag{3}
$$

where $K_\nu\left(\omega\right)$ is the modified Bessel function of the third kind of order ν with argument ω that has the following integral representation

$$
K_\nu\left(\omega\right) = \int_0^\infty z^{\nu-1} \exp\left[-\frac{1}{2}\omega\left(z+\frac{1}{z}\right)\right] dz.
\tag{4}
$$

Under our general setup, the mean and dispersion parameters of the PIGA distribution are modelled as functions of explanatory variables with parametric linear functional forms. In particular, we assume that

$$
\begin{aligned}
\mu_i &= \exp\left(x_{1,i}^T \boldsymbol{\beta}_1\right) \text{ and} \\
\end{aligned}
\tag{5}
$$
$$
\begin{aligned}
\phi_i &= \exp\left(x_{2,i}^T \boldsymbol{\beta}_2\right),
\end{aligned}
\tag{6}
$$

where $x_{1,i}$ and $x_{2,i}$ are the covariate vectors with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, regarding the policyholders and their vehicles, with $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, ..., \beta_{1,p_1}\right)^T$ and $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, ..., \beta_{2,p_2}\right)^T$ the corresponding vectors of regression coefficients and where it is considered that the matrices X_1 and X_2, with rows given by $x_{1,i}$ and $x_{2,i}$ respectively, are of full rank.

Finally, using the laws of total expectation and total variance, we obtain the mean and the variance of k_i, as given by

$$
E\left(k_i\right) = E_{\lambda_i}\left[E\left(k_i | \lambda_i\right)\right] = \mu_i E_{\lambda_i}\left[\lambda_i\right] = \mu_i
\tag{7}
$$

and

$$
\begin{aligned}
Var\left(k_i\right) &= E_{\lambda_i}\left[Var\left(k_i | \lambda_i\right)\right] + Var_{\lambda_i}\left[E\left(k_i | \lambda_i\right)\right] \\
&= E\left(k_i\right) + \frac{E^2\left(k_i\right)}{\left(\phi_i-1\right)} \\
&= \mu_i + \frac{\mu_i^2}{\left(\phi_i-1\right)}.
\end{aligned}
\tag{8}
$$

3. The EM Algorithm

In this section, an Expectation-Maximization (EM) algorithm, see Dempster et al. (1977) and McLachlan and Krishnan (2007), will be employed in order to facilitate the maximum likelihood (ML) estimation of the PIGA regression model with varying dispersion which was described in Section 2. Let $\left(k_i, x_{1,i}, x_{2,i}\right)$, $i = 1, ..., n$, be a sample of independent observations, where k_i is the response variable and $x_{1,i}$ and $x_{2,i}$ are the vectors of covariates with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively. Furthermore, assume that the data are produced according to the PIGA model and let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T\right)^T$ be the vector of the parameters. Subsequently, the log-likelihood of the PIGA model can be written as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left(\frac{2}{k_i!}\right) + \frac{k_i + \phi_i + 1}{2}\log\left(\mu_i\phi_i\right) - \log\left[\Gamma\left(\phi_i + 1\right)\right] + \log\left[K_{k_i-\phi_i-1}\left(2\sqrt{\mu_i\phi_i}\right)\right]. \quad (9)$$

Direct maximization of Equation (9) with respect to $\boldsymbol{\theta}$ is very cumbersome when both the mean and the dispersion parameters are modelled as functions of explanatory variables, since it would be required to differentiate the last term in Equation (9) with respect to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

Fortunately, the ML estimation of the model can be accomplished in a very easy manner through an EM type algorithm, which is specifically tailored to ML estimation for mixed Poisson models, see, for instance, Karlis (2001, 2005) and Barreto-Souza and Simas (2015), since their stochastic mixture representation involving a non-observable random variable, denoted by λ_i herein, can be considered to produce missing data. In our case, if one augments the unobserved data λ_i to the observed data $(k_i, \boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,i})$, for $i = 1, ..., n$, then the complete data log-likelihood factorizes into two parts

$$
\begin{aligned}
l_c(\boldsymbol{\theta}) \;=\; & \sum_{i=1}^{n} \left[k_i \log\left(\lambda_i\mu_i\right) - \lambda_i\mu_i - \log\left(k_i!\right)\right] + \\
& \sum_{i=1}^{n} \left[\left(\phi_i + 1\right)\log\left(\phi_i\right) - \left(\phi_i + 2\right)\log\left(\lambda_i\right) - \frac{\phi_i}{\lambda_i} - \log\left(\Gamma\left(\phi_i + 1\right)\right)\right],
\end{aligned}
\quad (10)
$$

for $i = 1, ..., n$. The regression coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are involved in the first and second terms of Equation (10), which correspond to the log-likelihoods of the Poisson and Inverse Gamma distributions that are given by Equations (1) and (2), respectively.

Subsequently, the $Q-$function is proportional to

$$
\begin{aligned}
Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}\right) \;\equiv\; & E_{z_i}\left(l_c(\boldsymbol{\theta})\,|k_i; \boldsymbol{\theta}^{(r)}\right) \propto \\
\propto\; & \sum_{i=1}^{n}\left[k_i \log\left(\mu_i^{(r)}\right) - \mu_i^{(r)} E_{\lambda_i}\left[\lambda_i|k_i; \boldsymbol{\theta}^{(r)}\right]\right] + \\
& \left[\sum_{i=1}^{n}\left(\phi_i^{(r)} + 1\right)\log\left(\phi_i^{(r)}\right) - \phi_i^{(r)} E_{\lambda_i}\left[\log\left(\lambda_i\right)|k_i; \boldsymbol{\theta}^{(r)}\right] - \\
& \phi^{(r)} E_{\lambda_i}\left[\frac{1}{\lambda_i}|k_i; \boldsymbol{\theta}^{(r)}\right] - \log\left(\Gamma\left(\phi_i^{(r)} + 1\right)\right)\right],
\end{aligned}
\quad (11)
$$

where $\boldsymbol{\theta}^{(r)}$ is the estimate of $\boldsymbol{\theta}$ at the rth iteration of our EM type algorithm and where $\mu_i^{(r)} = \exp\left(\boldsymbol{x}_{1,i}^T\boldsymbol{\beta}_1^{(r)}\right)$ and $\phi_i^{(r)} = \exp\left(\boldsymbol{x}_{2,i}^T\boldsymbol{\beta}_2^{(r)}\right)$.

At this point, it should be noted that if $k_i \sim \text{Poisson}(\lambda_i\mu_i)$ distribution and $\lambda_i \sim \text{Inverse Gamma}(\phi_i + 1, \phi_i)$ distribution then, applying Bayes theorem, one can find that the posterior distribution of $\lambda_i|k_i; \boldsymbol{\theta}$ is a Generalized Inverse Gaussian (GIG) distribution with pdf

$$f(\lambda_i|k_i; \boldsymbol{\theta}) = \frac{\left(\frac{\psi_i}{\chi_i}\right)^{\frac{\nu_i}{2}}}{2K_{\nu_i}\left(\sqrt{\psi_i\chi_i}\right)}\lambda_i^{\nu_i-1}\exp\left[-\frac{1}{2}\left(\frac{\chi_i}{\lambda_i} + \psi_i\lambda_i\right)\right], \quad (12)$$

where $\psi_i = 2\mu_i > 0, \chi_i = 2\phi_i > 0, \nu_i = k_i - \phi_i - 1 \in \mathcal{R}$ and where μ_i and ϕ_i are given by Equations (5) and (6), respectively.

In what follows, the above result will be useful for implementing the E-step of the EM algorithm, since it will enable us to compute the conditional sexpectations $E_{\lambda_i}\left[\lambda_i|k_i; \boldsymbol{\theta}^{(r)}\right]$, $E_{\lambda_i}\left[\frac{1}{\lambda_i}|k_i; \boldsymbol{\theta}^{(r)}\right]$ and $E_{\lambda_i}\left[\log\left(\lambda_i\right)|k_i; \boldsymbol{\theta}^{(r)}\right]$ in Equation (12). Furthermore, the M-step will consist in maximizing Equation (12) with respect to $\boldsymbol{\theta}$. The EM type algorithm procedure can be formally described, as follows.

- **E-step:** Given the estimates $\theta^{(r)}$, obtained from the *r*th iteration, compute for all $i = 1, ..., n$, the pseudo-values

$$w_{1,i} = E_{\lambda_i}\left[\lambda_i | k_i; \theta^{(r)}\right] = \sqrt{\frac{\phi_i^{(r)}}{\mu_i^{(r)}}} \frac{K_{k_i - \phi_i^{(r)}}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right)}{K_{k_i - \phi_i^{(r)} - 1}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right)}, \tag{13}$$

$$w_{2,i} = E_{\lambda_i}\left[\frac{1}{\lambda_i} | k_i; \theta^{(r)}\right] = \sqrt{\frac{\mu_i^{(r)}}{\phi_i^{(r)}}} \frac{K_{k_i - \phi_i^{(r)} - 2}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right)}{K_{k_i - \phi_i^{(r)} - 1}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right)} \tag{14}$$

and

$$w_{3,i} = E_{\lambda_i}\left[\log\left(\lambda_i\right) | k_i; \theta^{(r)}\right] = \frac{\log\left(\frac{\phi_i^{(r)}}{\mu_i^{(r)}}\right)}{2} + \frac{\partial K_{k_i - \phi_i^{(r)} - 1}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right) / \partial\left(k_i - \phi_i^{(r)} - 1\right)}{K_{k_i - \phi_i^{(r)} - 1}\left(2\sqrt{\mu_i^{(r)}\phi_i^{(r)}}\right)} \tag{15}$$

The Equations (13)–(15) involved in the E-step of the algorithm have closed form expressions. However, unlike the case with Equations (13) and (14), which can be easily evaluated, as is well known, see, for instance, Mencía and Sentana (2012), it is not always possible to obtain numerically reliable direct derivatives of the Bessel function with respect to its order, which is involved in the second term of Equation (15). In this study, in order to compute Equations (13)–(15) we rely on the function Egig within the R package ghyp, which was contributed by Weibel et al. (2020). Note that, in the case of Equation (15), Egig can provide an accurate numerical approximation of the first derivative of the modified Bessel function with respect to its order by using the function grad from the R package numDeriv, see Gilbert and Varadhan (2019).

- **M-step:** Using $w_{1,i}$, $w_{2,i}$ and $w_{3,i}$ from the E-step and the Newton-Raphson algorithm twice, find the maximum global point $\theta^{(r+1)}$ of the $Q-$function, i.e., obtain the updated estimates $\beta_1^{(r+1)}$ and $\beta_2^{(r+1)}$.

  - Firstly, taking the derivatives of the $Q-$function with respect to β_1 we obtain the following results:

$$h_1\left(\beta_1\right) = \frac{\partial Q\left(\theta; \theta^{(r)}\right)}{\partial\beta_{1,j}} = \sum_{i=1}^{n}\left(k_i - \mu_i^{(r)}w_{1,i}\right)x_{1,ij}, \tag{16}$$

  and

$$H_1\left(\beta_1\right) = \frac{\partial^2 Q\left(\theta; \theta^{(r)}\right)}{\partial\beta_{1,j}\partial\beta_{1,j}^T} = \sum_{i=1}^{n}\left(-\mu_i^{(r)}w_{1,i}\right)x_{1,ij}x_{1,ij}^T = X_1^T W_1 X_1, \tag{17}$$

  for $i = 1, ..., n$ and $j = 1, ..., p_1$ and where $W_1 = diag\{-\mu_i^{(r)}w_{1,i}\}$.

  Subsequently, the iterative procedure for the Newton–Raphson algorithm for β_1 goes, as follows:

$$\beta_1^{(r+1)} \equiv \beta_1^{(r)} - \left[H_1\left(\beta_1^{(r)}\right)\right]^{-1}h_1\left(\beta_1^{(r)}\right). \tag{18}$$

– Secondly, differentiating the Q−function with respect to $\boldsymbol{\beta}_2$ gives

$$h_2\left(\boldsymbol{\beta}_2\right) = \left[\phi_i^{(r)}\log\left(\phi_i^{(r)}\right) + \phi_i^{(r)} + 1 - \phi_i^{(r)}w_{2,i} - \phi_i^{(r)}w_{3,i} - \phi_i^{(r)}\Psi\left(\phi_i^{(r)} + 1\right)\right]x_{2,ij}, \quad (19)$$

$$
\begin{aligned}
H_2\left(\boldsymbol{\beta}_2\right) &= \sum_{i=1}^{n}\phi_i^{(r)}\left[\log\left(\phi_i^{(r)}\right) + 2 - w_{2,i} - w_{3,i} - \Psi\left(\phi_i^{(r)} + 1\right)\right. \\
&\quad \left. -\phi_i^{(r)}\Psi_3\left(\phi_i^{(r)} + 1\right)\right]x_{2,ij}x_{2,ij}^T \\
&= X_2^T W_2 X_2,
\end{aligned}
\quad (20)
$$

for $i = 1,...,n$ and $j = 1,...,p_2$, where $\Psi\left(.\right)$ and $\Psi_3\left(.\right)$ are the digamma and trigamma functions, and where
$$W_2 = diag\left\{\phi_i^{(r)}\log\left(\phi_i^{(r)}\right) + 2\phi_i^{(r)} - \phi_i^{(r)}w_{2,i} - \phi_i^{(r)}w_{3,i} - \phi_i^{(r)}\Psi\left(\phi_i^{(r)} + 1\right)\right.$$
$$\left. -\left(\phi_i^2\right)^{(r)}\Psi_3\left(\phi_i^{(r)} + 1\right)\right\}.$$
Then, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta}_2$ is as follows:

$$\boldsymbol{\beta}_2^{(r+1)} \equiv \boldsymbol{\beta}_2^{(r)} - \left[H_2\left(\boldsymbol{\beta}_2^{(r)}\right)\right]^{-1}h_2\left(\boldsymbol{\beta}_2^{(r)}\right), \quad (21)$$

for $i = 1,...,n$ and $j = 1,...,p_2$.

- Finally, it should be noted that when the regression structures for the mean and dispersion parameters of the model are limited to the constants $\beta_{1,0}$ and $\beta_{2,0}$ this EM type algorithm can be employed for the ML estimation of the 'univariate', without regression components, model.

4. Numerical Illustration

The study is based on subset of heavy-tailed and overdispersed claim frequency data from a pool of MTPL insurance policies observed for 3.5 years from a major Greek insurance company. The sample comprised of insureds with complete records, i.e., with availability of all the a priori rating variables under consideration. There were 14,143 observations that met our criteria. The response variable is the number of claims at fault registered for each insured vehicle in the data set and the explanatory variables we employ are: the age of the driver (AD), the horsepower (HP) of their car, and the age of their car (AC). Furthermore, an exploratory analysis was carried out in order to accurately select the subset of explanatory variables with the highest predictive power for the number of claims. Additionally, in light of the heterogeneity that exists within the portfolio, we grouped the levels of each a priori rating variable with respect to risk profiles with similar claim frequency. This will enable us to achieve ratemaking accuracy and balance homogeneity and sufficiency of the volume of data in each cell in order to provide credible patterns. This is necessary, since, under the proposed modelling framework, both the mean and dispersion parameters of the Poisson-Inverse Gamma (PIGA) distribution will be modelled in terms of covariate information.

- The variable AD consists of two categories of policyholders, those of age: C1 = "between 18 and 25 years" and C2 = "greater than 25 years".
- The variable HP consists of two categories of vehicles, those with a HP: C1 = "0–5000 cc" and C2 = "greater than 5000 cc".
- The variable AC consists of two categories of vehicles, those of age: C1 = "between 0 and 5 years" and C2 = "greater than 5 years".

Table 1 depicts some standard descriptive statistics for claim counts and the number of observations in each category of the three explanatory variables.

Table 1. Descriptive statistics of claim counts and the size of the different categories of the explanatory variables.

Statistic	Value	Age of the Driver (AD)		Horsepower of the Car (HP)		Age of the Car (AC)	
# Observations	14,143	C1:	3238	C1:	5042	C1:	4318
Minimum	0	C2:	10,905	C2:	9101	C2:	9825
Median	0		-		-		-
Mean	0.4827		-		-		-
Variance	0.6988		-		-		-
Maximum	12		-		-		-

In what follows, we will compare the fit of the PIGA model with the classic Negative Binomial Type I (NBI) and Poisson-Inverse Gaussian (PIG) models for the case without covariate information and for the case when the mean and dispersion parameters of these mixed Poisson models are allowed to be modelled as functions of covariates.

- The NBI and PIG claim frequency models can be constructed as follows. Consider a policyholder i, $i = 1, ..., n$, whose number of claims, denoted as k_i, with $k_i = 0, 1, 2, 3, ...$, are independent and suppose that given a continuous random variable $\lambda_i > 0$, with pdf $f(\lambda_i; \phi_i)$ defined on $\mathcal{R}^+$ and where $\phi_i > 0$, $k_i|, \lambda_i$ follows a Poisson distribution with pmf given by Equation (1). Additionally, we assume that that $E(\lambda_i) = 1$ as this ensures that the model is identifiable. The following results are very well known, see, for example, Dionne and Vanasse (1989, 1992) and Boucher et al. (2007, 2008).

  - Let λ_i follow a Gamma distribution with pdf given by

  $$f(\lambda_i; \phi_i) = \frac{\lambda_i^{\frac{1}{\phi_i} - 1} \frac{1}{\phi_i}^{\frac{1}{\phi_i}} \exp\left(-\frac{\lambda_i}{\phi_i}\right)}{\Gamma\left(\frac{1}{\phi_i}\right)}, \tag{22}$$

  Parameterization (22) ensures that $E(u_i) = 1$.

  Subsequently, the unconditional distribution of k_i becomes a NBI distribution, with pmf given by

  $$P(k_i) = \frac{\Gamma\left(k_i + \frac{1}{\phi_i}\right)}{k_i! \Gamma\left(\frac{1}{\phi_i}\right)} \left(\frac{\phi_i \mu_i}{1 + \phi_i \mu_i}\right)^{k_i} \left(\frac{1}{1 + \phi_i \mu_i}\right)^{\frac{1}{\phi_i}}. \tag{23}$$

  The mean and the variance of the NBI distribution are given by

  $$E(k_i) = \mu_i \tag{24}$$

  and

  $$Var(k_i) = \mu_i + \mu_i^2 \phi_i. \tag{25}$$

  - Let λ_i follow a Inverse Gaussian distribution with pdf given by

  $$f(\lambda_i; \phi_i) = \frac{1}{\sqrt{2\pi\phi_i\lambda_i^3}} \exp\left[-\frac{1}{2\phi_i\lambda_i}(\lambda_i - 1)^2\right]. \tag{26}$$

Parameterization (26) also ensures that $E(\lambda_i) = 1$. Then, the unconditional distribution of k_i becomes a PIG distribution, with pmf given by

$$P\left(k_i\right) = \left(\frac{2\sqrt{\phi_i^{-2} + \frac{2\mu_i}{\phi_i}}}{\pi}\right)^{\frac{1}{2}} \frac{\mu_i^{k_i} e^{\frac{1}{\phi_i}} K_{k_i - \frac{1}{2}}\left(\sqrt{\phi_i^{-2} + \frac{2\mu_i}{\phi_i}}\right)}{\left(\sqrt{\phi_i^{-2} + \frac{2\mu_i}{\phi_i}}\phi_i\right)^{k_i} k_i!}, \tag{27}$$

where $K_\nu\left(\omega\right)$ is the modified Bessel function of the third kind of order ν with argument ω with integral representation given by Equation (4).

The mean and the variance of the PIG distribution are given by

$$E(k_i) = \mu_i \tag{28}$$

and

$$Var(k_i) = \mu_i + \mu_i^2 \phi_i. \tag{29}$$

– We consider that the mean and dispersion parameters of the NBI and PIG distributions are modelled as functions of explanatory variables

$$\mu_i = \exp\left(x_{1,i}^T \beta_1\right) \text{ and} \tag{30}$$

$$\phi_i = \exp\left(x_{2,i}^T \beta_2\right), \tag{31}$$

where $x_{1,i}$ and $x_{2,i}$ are covariate vectors with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, with $\beta_1 = \left(\beta_{1,1}, ..., \beta_{1,p_1}\right)^T$ and $\beta_2 = \left(\beta_{2,1}, ..., \beta_{2,p_2}\right)^T$ the corresponding parameter vectors and where it is assumed that the matrices X_1 and X_2, with rows given by $x_{1,i}$ and $x_{2,i}$, respectively, are of full rank.

– Finally, it should be noted that when the regression components in each of the NBI and PIG models are limited to the constants $\beta_{1,0}$ and $\beta_{2,0}$, we obtain the univariate, without regression components, models.

4.1. Modelling Results

The ML estimates of the parameters and the corresponding standard errors in parentheses for the NBI, PIG, and PIGA distributions and regression models with varying dispersion are presented in Tables 2 and 3, respectively. Note that, for the case when the mean and dispersion parameters, μ_i and ϕ_i, $i = 1, ..., n$, of the NBI, PIG, and PIGA distributions are modelled in terms of covariates, variable selection should start by selecting the best predictor for parameter μ_i of each claim frequency model. This can be done by adding all available explanatory variables and testing whether the exclusion of each one will result in lower Global Deviance (DEV), Akaike information criterion (AIC), and Schwartz Bayesian criterion (SBC) values. Subsequently, we can continue by testing which explanatory variable between those used in parameter μ_i would lead to a further decrease of the DEV, AIC, and SBC values when inserted in parameter ϕ_i of each claim frequency model. Furthermore, if different parameter specifications for the same claim frequency model result in very close DEV, AIC, and SBC values, we should opt for the simpler model with the fewer predictors for the dispersion parameter ϕ_i in order to avoid overfitting. Regarding our data set, as we can observe from Table 3, the variables AD, HP, and AC are in the model equation for μ_i and the variable AD is in the model equation for ϕ_i. Additionally, we see that the values of the estimated regression coefficients of the variables AD, HP, and AC are almost identical for μ_i across all three claim frequency models. Additionally, it can be seen that the values of the estimated regression coefficients of the variable AD have a similar effect (positive and/or negative) on parameter ϕ_i in the case of the NBI and PIG models,

but have a different effect for ϕ_i in the case of the PIGA model. In what follows, we will see that, due to this discrepancy, the a posteriori, or Bonus-Malus, premium rates that result from the traditional NBI and PIG models will differ from those derived by the more heavy-tailed PIGA model.

Table 2. Negative Binomial Type I (NBI), Poisson-Inverse Gaussian (PIG), and Poisson-Inverse Gamma (PIGA) distributions.

NBI	PIG	PIGA
μ	μ	μ
0.4827 (0.0170)	0.4827 (0.0173)	0.4827 (0.0178)
ϕ	ϕ	ϕ
0.7107 (0.0678)	0.7787 (0.0741)	2.0107 (0.0712)

Table 3. NBI, PIG, and PIGA regression models with varying dispersion.

NBI		PIG		PIGA	
Coeff. $\boldsymbol{\beta}_1$		Coeff. $\boldsymbol{\beta}_1$		Coeff. $\boldsymbol{\beta}_1$	
Intercept AD	−0.4729 (0.0561)	Intercept CS	−0.4772 (0.0572)	Intercept CS	−0.4114 (0.0575)
C2 HP	−1.2390 (0.0390)	C2 AC	−1.2361 (0.0398)	C2 AC	−1.2654 (0.0422)
C2 AC	1.0378 (0.3206)	C2 HP	1.1135 (0.3267)	C2 HP	0.9752 (0.2644)
C2	−0.6481 (0.5225)	C2	−0.7196 (0.5275)	C2	−2.6608 (0.4652)
Coeff. $\boldsymbol{\beta}_2$		Coeff. $\boldsymbol{\beta}_2$		Coeff. $\boldsymbol{\beta}_2$	
Intercept AD	−2.4935 (0.2911)	Intercept CS	−2.1937 (0.2448)	Intercept CS	2.1639 (0.1607)
C2	0.8878 (0.1862)	C2	0.7366 (0.1016)	C2	−0.7704 (0.1036)

Finally, normalized randomized quantile residuals, see Dunn and Smyth (1996), are used as a graphical tool to help us assess the adequacy of the fit of the competing NBI, PIG, and PIGA regression models with varying dispersion. Additionally, the simple Poisson regression model was fitted for comparison purposes. The normalized randomized quantile residuals for these claim count regression models are defined as $\hat{r}_i = \Phi^{-1}(u_i)$, where Φ^{-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 $\left[F_i(k_i - 1|\boldsymbol{\theta}^{(r+1)}), F_i(k_i|\boldsymbol{\theta}^{(r+1)}) \right]$, where F_i is the cumulative distribution function estimated for the ith policyholder and where $\boldsymbol{\theta}^{(r+1)}$ is the vector of the estimated model parameters after the EM algorithm has reached the global maximum and k_i is the corresponding observation. The claim frequency model fit can be investigated via the usual quantile–quantile plots. In particular, if the data indeed follow the assumed claim frequency distribution, then the residual on the quantile-quantile plot will fall approximately on a straight line. Figure 1 depicts the normalized (random) quantiles for the Poisson, NBI, PIG, and PIGA models. From Figure 1, we see that, unlike the Poisson model, which has a light tail, and, hence, is not a good assumption, the residuals of the NBI, PIG, and PIGA models are close to the diagonal and indicate a good fit to the distribution of the claim frequency. Furthermore, we observe that the more heavy-tailed PIGA model yields a better performance than the NBI and PIG models close to the right tail of the claim frequency distribution. On the other hand, the PIGA model shows a worse fit than the NBI and PIG models in the lower tail. These were anticipated, since, as is well known, and as is was previously mentioned, the tails of mixed Poisson distributions are equivalent to the tails of their mixing distributions, also see, for example, Willmot (1998) and Perline (1990). Thus, as we move from the less heavy-tailed Gamma and Inverse Gaussian mixing distributions to the Inverse Gamma mixing distribution zero and near zero values in the left tail area become less likely and high values in the right tail area become more likely.

Therefore, it should be noted that, regarding our data set, it is reasonable to suggest the employment of the PIGA model, which, as we are going to see in what follows, performs better than the NBI and PIG models in terms of the DEV, AIC, and SBC values, for deriving a posteriori, or Bonus-Malus, ratemaking mechanisms for younger drivers who are more likely to have car accidents than older drivers and, hence, are more likely to make insurance claims. However, it should also be noted that, because, in view of the unique features of the body, the left and the right tail areas of the actual claim frequency distribution, for other data sets the NBI, PIG and/or a different mixed Poisson model may perform better than the PIGA model. Thus, because a particular model cannot represent all aspects of real insurance data, judging from a practical business standpoint, as an overall conclusion, it may be appropriate to use a combination of models that could provide alternative options to the insurer for carrying out different tasks, such as deciding on their pricing strategies, setting the appropriate level of reserves, and reinsurance.
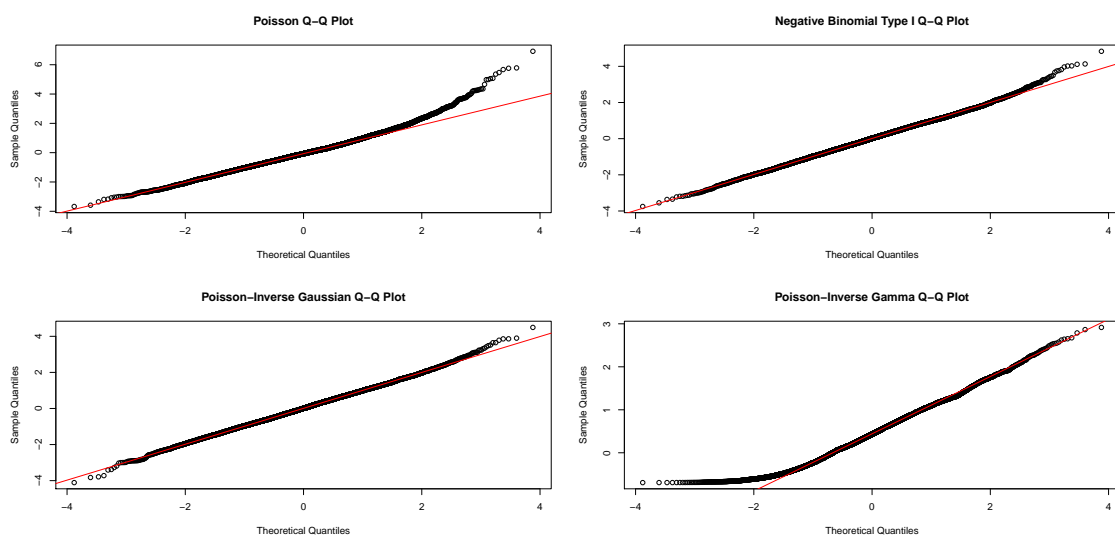

Figure 1. Normalized quantiles for the Poisson regression model & the NBI, PIG, and PIGA regression models with varying dispersion.

4.2. Models Comparison

In this subsection, the DEV, AIC, and the SBC, which are classic hypothesis/specification tests, will be used to compare the fit of the NBI, PIG, and PIGA distributions/regression models with varying dispersion

The DEV is given by

$$DEV = -2\hat{l}\left(\hat{\theta}\right), \tag{32}$$

where $\hat{l}$ is the maximum of the log-likelihood and $\hat{\theta}$ is the estimated parameter vector of the model. Additionally, the AIC and the SBC are defined as

$$AIC = DEV + 2 \times df \tag{33}$$

and

$$SBC = DEV + \log(n) \times df, \tag{34}$$

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

The resulting DEV, AIC, and SBC values for the competing distributions/regression models with varying dispersion are reported in Table 4 (Panels A and B). According to a very well known rule-of-thumb, a model noticeably outperforms its competitor if the difference in their log-likelihoods

is greater than five, corresponding to a difference in their AIC and SBC values of more than ten and greater than five, respectively, see Burnham and Anderson (2002) and Raftery (1995), respectively. Thus, as we can observe from Panels A and B, the PIGA distribution/regression model with varying dispersion model gives the best fit.

Table 4. Models comparison.

	Panel A: Distributions		**Panel B:**	**Regression Models with Varying Dispersion**		
Model	AIC	SBC	Model	DEV	AIC	SBC
NBI	17,829.1	17,843.5	NBI	15,885.1	15,897.1	15,940.1
PIG	17,799.2	17,813.5	PIG	15,867.3	15,879.4	15,922.3
PIGA	17,780.4	17,794.8	PIGA	15,848.6	15,860.6	15,903.6

4.3. Application to Ratemaking

In this subsection, the net premium principle is used to compute the a priori and a posteriori, or Bonus-Malus, premium rates resulting from the NBI, PIG, and PIGA distribution/regression models with varying dispersion.

Firstly, the differences between the claim frequency regression models with varying dispersion will be analyzed via the expected claim frequency of the insureds who belong to the eight different risk classes, which are determined by the relevant a priori characteristics. In particular, $E(k_i)$, for $i = 1, ..., n$, serves as a basis of the premium for each risk class. Table 5 presents the a priori premium rates resulting from the NBI, PIG, and PIGA models. From Table 5, we see that the group of policyholders with the lowest mean claim frequency are those who are older than 25 years and have a car with HP between 0 and 5000 cc and age greater than five years, i.e., risk class 6. Additionally, the group of insureds with the highest mean claim frequency are those who are aged between 18 and 25 years and have a car with HP between greater than 5000 cc and age between zero to five years, i.e., risk class 3. Overall, as expected, we observe that small discrepancies lie in the mean claim frequency values of the NBI, PIG, and PIGA models. However, when the a posteriori corrections will be computed in the following examples, we will see that allowing both the mean and dispersion parameters of the NBI, PIG, and PIGA models to be to modelled as functions of covariate information will affect the estimation of the Bonus-Malus premium rates. In particular, since, as was previously mentioned, the effect of the values of the estimated regression coefficients of the explanatory variable AD for the dispersion parameter is similar in the case of the NBI and PIG models but differs in the case of the PIGA model as a result the Bonus-Malus premiums determined by the NBI and PIG models will differ from the premiums rates that result from the PIGA model.

Table 5. A Priori premiums, regression models with varying dispersion.

Risk Class	**Explanatory Variables**			**A Priori Premiums**		
	AD	HP	AC	NBI	PIG	PIGA
1	C1	C1	C1	0.18	0.18	0.19
2	C1	C1	C2	0.09	0.09	0.10
3	C1	C2	C1	0.50	0.54	0.50
4	C1	C2	C2	0.26	0.26	0.26
5	C2	C1	C1	0.05	0.05	0.05
6	C2	C1	C2	0.03	0.03	0.03
7	C2	C2	C1	0.15	0.16	0.14
8	C2	C2	C2	0.08	0.08	0.07

Secondly, we investigate how the PIGA distribution/regression model with varying dispersion responds to claim experience. Consider an insured i with claim frequency history $k_i^1, ..., k_i^t$ and

$x_{1,i}^1, ..., x_{1,i}^{t+1}, x_{2,i}^1, ..., x_{2,i}^{t+1}$ characteristics and assume that $K = \sum_{l=1}^{t} k_i^l$ is the total number of claims that they had. In what follows, we determine at the renewal of the policy the expected claim frequency λ_i^{t+1} of the insured i for the period $t+1$ given the observation of the reported claims in the preceding t periods and the observable characteristics in the preceding $t+1$ periods and the current period. As was mentioned in Section 3, employing Bayes theorem, we can find that the posterior distribution of λ_i^{t+1} is a GIG. Thus, using the quadratic loss function and the net premium principle we can easily see that, in this case, the mean of the posterior structure function given by [1]

$$E\left(\lambda_i^{t+1}|k_i^1, ..., k_i^t; x_{1,i}^1, ..., x_{1,i}^{t+1}, x_{2,i}^1, ..., x_{2,i}^{t+1}\right) = \sqrt{\frac{\mu_i \phi_i}{t}} \frac{K_{K-\phi_i}\left(2\sqrt{t\mu_i\phi_i}\right)}{K_{K-\phi_i-1}\left(2\sqrt{t\mu_i\phi_i}\right)}, \tag{35}$$

where μ_i and ϕ_i are given by Equations (5) and (6), respectively.

Following this methodology, we calculate the Bonus-Malus premium rates determined by the PIGA model based only on the a posteriori criteria, i.e., the number of individual claims, and based both on the a posteriori and a priori criteria, i.e., the characteristics of the policyholders and their cars. When we consider both criteria, to illustrate the efficiency of the PIGA regression model with varying dispersion for deriving Bonus-Malus ratemaking mechanisms for heavy-tailed and overdispersed claim counts, we restrict our attention to young drivers aged between 18 and 25 years, because they reported significantly more claims when compared to older drivers. In what follows, we examine all four risk classes 1, 2, 3, and 4 of young policyholders who share common characteristics, i.e., which can be formed by all the possible combinations of category C1 of the variable AC with categories C1 and C2 of the variables HP and AC, see Table 5. Assuming that the number of claims ranges from 0 to 4 and the age of the policy is up to five years, we calculated comparable relative premiums for the NBI, PIG, and PIGA distributions/regression models with varying dispersion respectively. The results are presented in Tables 6–10 .

From all Tables 6–10, we see 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, we observe that the bonuses awarded to claim free policyholders are quite similar and moderate discrepancies lie in the premiums required to be paid by those insureds who have reported up to $K = 2$ claims or who have made more than $K = 2$ claims, but have some claim experience in the case of the NBI, PIG and PIGA distributions/regression models with varying dispersion. For example, for the case without covariates, as we can see from Table 6, the policyholders who are claim free will receive a bonus of 22.72%, 22.02% and 19.23% in year $t = 3$ in the case of the NBI, PIG and PIGA distributions, respectively. Additionally, the insureds who had $K = 3$ claims in year $t = 4$ will have to pay a malus of 125.00%, 140.63%, and 131.43% in the case of the NBI, PIG, and PIGA distributions, respectively. Similarly, for the case with covariates, we observe that claim free policyholders will receive bonuses of 4.23%, 5.45%, and 6.36%, see Table 7, 11.08%, 14.29%, and 14.00%, see Table 8, 2.26%, 2.77%, and 3.55%, see Table 9, and 6.12%, 7.78%, and 8.39%, see Table 10, in year $t = 3$ in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively. Additionally, the individuals who had $K = 3$ claims in year $t = 4$ will have to pay maluses of 17.85%, 24.50%, and 27.86%, see Table 7, 7.01%, 6.62%, and 7.65%, see Table 8, 21.06%, 30.41%, and 37.37%,

---

[1] Note that Schiegl (2010) used a different parameterization of the PIGA distribution to derive a Bonus-Malus system for the case without covariates, i.e., based only on the a posteriori criteria. Note also that, the Bonus-Malus premium functions determined by the classic NBI and PIG models were not included for the sake of brevity. Those functions can be found, for instance, in Dionne and Vanasse (1989, 1992), Frangos and Vrontos (2001), Mahmoudvand and Hassani (2009) and Tzougas et al. (2014, 2018) respectively. Note also that, the Bounus-Malus premium rates for the case when only on the a posteriori criteria are used can be obtained if the regression components of the NBI, PIG, and PIGA models are limited to constants.

see Table 9, and 14.81%, 19.55%, and 21.82%, see Table 10, in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively.

Furthermore, the more heavy-tailed PIGA distribution/regression model with varying dispersion model penalizes high risk policyholders who reported more than $K = 2$ claims in years $t = 1$ and $t = 2$ more severely than the NBI and PIG distribution/regression models with varying dispersion. Thus, the proposed model encourages good driving behavior more than the NBI and PIG models during the first two years of the policy. At this point, we would also like to call attention to the fact that, since, in our data, many of the young insureds who belong to risk classes 1, 2, 3, and 4 had just started to drive; this is in line with market practice when considering these years of the driving history as the most dangerous period. For instance, for the case without covariates, as we can see from Table 6, policyholders who had $K = 3$ claims will have to pay a malus of 161.87%, 205.03%, and 248.87% of the basic premium in year $t = 2$ in the case of the NBI, PIG, and PIGA distributions, respectively. Analogously, regarding the case with covariates, we observe, for example, that the insureds who had $K = 4$ claims and belong to risk class 1, see Table 7, will have to pay a malus of 29.25%, 43.48%, and 60.45% of the basic premium in year $t = 2$ in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively. Additionally, we see that the individuals who belong to risk class 3, see Table 9, will have to pay a malus of 31.03%, 47.29%, and 70.23% of the basic premium in year $t = 2$ in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively.

Additionally, the premiums required to be paid by a high risk policyholder who has reported more than $K = 2$ claims in different years are better distinguished under the PIGA distribution/regression model with varying dispersion rather than the NBI and PIG distributions/regression models with varying dispersion. Regarding the case without covariates, as we can see from Table 6, for example, an insured who had $K = 3$ claims in years $t = 3$ and $t = 5$ will have to pay maluses of 142.04% and 110.20%, 168.69% and 118.31%, and 173.91% and 103.42% in the case of the NBI, PIG, and PIGA distributions, respectively. Similarly, for the case with covariates, we observe, for instance, that a policyholder who had $K = 4$ claims in years $t = 3$ and $t = 5$ and belongs to risk class 2, see Table 8, will have to pay maluses of 18.31% and 10.17%, 22.69%, and 10.18% and 27.11% and 11.82% in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively. An insured who belongs to risk class 4, see Table 10, will have to pay maluses of 24.91% and 20.01%, 35.35%, and 26.97% and 44.71% and 31.07% in the case of the NBI, PIG, and PIGA regression models with varying dispersion, respectively.

Finally, it is worth noting that the Bonus-Malus premiums reported in Tables 7–10 are significantly lower than the Bonus-Malus premiums presented in Table 6. Therefore, allowing both the mean and the dispersion parameters of the three mixed Poisson models to vary through covariates is justified from a practical business standpoint since the MTPL market is very competitive and, hence, insurance companies can better attract clients by offering them lower penalties. Overall, for all the reasons listed above, it is reasonable to agree that, for the heavy-tailed and overdispersed MTPL data set used in this study, the employment of the PIGA model, which provided the best fitting performances, leads to a better tariffication than the classic NBI and PIG models, since, while it rewards claim free policyholders in a similar way to the latter, it also results in a more reasonable growth in the premium payments.

Table 6. A Posteriori, or Bonus-Malus, premiums, 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	91.07	155.80	220.53	285.25	349.98
2	83.61	143.03	202.45	261.87	321.30
3	77.28	132.20	187.12	242.04	296.96
4	71.84	122.89	173.94	225.00	276.05
5	67.11	114.81	162.50	210.20	257.89
PIG					
Year	**Number of Claims *k***				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	90.73	154.83	245.47	354.04	471.96
2	83.64	138.11	214.06	305.03	404.23
3	77.98	125.34	190.59	268.69	354.12
4	73.34	115.23	172.33	240.63	315.55
5	69.44	106.99	157.71	218.31	284.92
PIGA					
Year	**Number of Claims *k***				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	90.92	145.55	268.85	534.54	990.08
2	85.14	127.20	206.65	348.87	567.61
3	80.77	115.70	175.77	273.91	416.53
4	77.24	107.39	156.18	231.43	336.82
5	74.28	100.96	142.26	203.42	286.81

Table 7. A Posteriori, or Bonus-Malus, premiums for risk class 1, regression models with varying dispersion.

NBI					
Year	**Number of Claims *k***				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	98.55	106.69	114.83	122.98	131.12
2	97.14	105.17	113.19	121.22	129.25
3	95.77	103.69	111.60	119.51	127.42
4	94.44	102.25	110.05	117.85	125.65
5	93.15	100.84	108.54	116.24	123.93
PIG					
Year	**Number of Claims *k***				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	98.08	108.81	120.59	133.40	147.19
2	96.27	106.60	117.93	130.25	143.48
3	94.55	104.52	115.44	127.28	140.01
4	92.92	102.55	113.08	124.50	136.75
5	91.38	100.69	110.86	121.87	133.68

Table 7. *Cont.*

PIGA				
Year	**Number of Claims k**			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	97.67	109.62	124.74	144.40	170.76
2	95.57	106.66	120.45	137.90	160.45
3	93.64	104.03	116.73	132.49	152.34
4	91.87	101.64	113.44	127.86	145.66
5	90.24	99.47	110.51	123.82	140.00

Table 8. A Posteriori, or Bonus-Malus, premiums for risk class 2, regression models with varying dispersion.

NBI				
Year	**Number of Claims k**			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	96.01	103.95	111.88	119.81	127.74
2	92.33	99.96	107.59	115.22	122.84
3	88.92	96.27	103.61	110.96	118.31
4	85.75	92.84	99.92	107.01	114.10
5	82.81	89.65	96.49	103.33	110.17

PIG				
Year	**Number of Claims k**			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	94.47	104.43	115.33	127.15	139.86
2	89.77	98.76	108.57	119.17	130.54
3	85.71	93.90	102.81	112.42	122.69
4	82.16	89.68	97.84	106.62	115.99
5	79.01	85.97	93.49	101.57	110.18

PIGA				
Year	**Number of Claims k**			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	94.29	104.91	117.97	134.27	154.97
2	89.76	98.85	109.68	122.69	138.44
3	86.01	93.99	103.31	114.24	127.11
4	82.81	89.96	98.17	107.65	118.59
5	80.03	86.51	93.88	102.27	111.82

Table 9. A Posteriori, or Bonus-Malus, premiums for risk class 3, regression models with varying dispersion.

NBI				
Year	**Number of Claims *k***			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	99.24	107.44	115.63	123.83	132.03
2	98.48	106.62	114.76	122.89	131.03
3	97.74	105.82	113.89	121.97	130.05
4	97.01	105.03	113.04	121.06	129.08
5	96.30	104.25	112.21	120.16	128.12

PIG				
Year	**Number of Claims *k***			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	99.05	109.99	122.02	135.11	149.20
2	98.13	108.87	120.66	133.49	147.29
3	97.23	107.77	119.35	131.92	145.45
4	96.36	106.71	118.07	130.41	143.67
5	95.51	105.68	116.83	128.93	141.95

PIGA				
Year	**Number of Claims *k***			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	98.75	111.16	127.05	148.04	176.89
2	97.57	109.48	124.53	144.08	170.23
3	96.45	107.90	122.22	140.55	164.57
4	95.38	106.41	120.09	137.37	159.64
5	94.37	105.01	118.11	134.47	155.27

Table 10. A Posteriori, or Bonus-Malus, premiums for risk class 4, regression models with varying dispersion.

NBI				
Year	**Number of Claims *k***			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	97.87	105.96	114.05	122.13	130.22
2	95.84	103.75	111.67	119.59	127.51
3	93.88	101.64	109.40	117.15	124.91
4	92.01	99.61	107.21	114.81	122.41
5	90.20	97.66	105.11	112.56	120.01

PIG				
Year	**Number of Claims *k***			
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	97.19	107.72	119.29	131.85	145.37
2	94.61	104.59	115.52	127.38	140.13
3	92.22	101.70	112.07	123.30	135.35
4	90.00	99.03	108.89	119.55	130.98
5	87.94	96.56	105.95	116.10	126.97

Table 10. *Cont.*

	PIGA				
Year	**Number of Claims *k***				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	96.83	108.44	123.00	141.73	166.45
2	94.07	104.60	117.53	133.65	154.04
3	91.61	101.29	112.96	127.19	144.71
4	89.40	98.37	109.03	121.82	137.25
5	87.39	95.76	105.60	117.24	131.07

5. Computational Aspects

The PIGA distribution and regression model with varying dispersion were estimated using the EM algorithm, which is presented in Section 3. A rather strict criterion was used and it took the model, for both the cases with and without covariate information, a fairly large number of iterations converge. In particular, the algorithm iterated between the E- and the M-steps until the relative change in log-likelihood, which is given by Equation (9), between two successive iterations was smaller than 10^{-12}.

We also call attention to the fact that, because the M-step involves two Newton–Raphson iterations, the choice of meaningful initial values for the vectors β_1 and β_2 is important, as it can influence the speed of convergence of the EM algorithm and its ability to locate the global maximum. We obtained good starting values for β_1 by fitting the simple Poisson regression. Alternatively, we can obtain the initial values for β_1 based on the data, as follows: (i) compute $E(k_i)$ for the eight different risk classes in Table 5 based on all observations $i = 1, ..., n$ and (ii) using a log-link function for μ_i, see Equation (5), solve Equation (7) with respect to β_1. Furthermore, we obtained good starting values for β_2 by: (i) computing $Var(k_i)$ for the eight different risk classes based on all observations $i = 1, ..., n$ and (ii) computing $E(k_i)$ for the eight different risk classes (or alternatively computing μ_i based on the initial values for β_1 and the log-link function given by Equation (5)) and using the log-link function for ϕ_i, see Equation (6), so we solve the Equation (7) with respect to β_2. However, it should be noted that, in order to ensure that the global maximum had been obtained and the algorithm had not been trapped in a local maximum, we checked with many other initial values for β_2 and for all cases the algorithm converged to the same solution. All of the computing was done using the programming language R. The standard errors were numerically obtained by the estimating the Hessian matrix of the log-likelihood. Alternatively, one can obtain them by using the standard approach of Louis (1982).

Additionally, the NBI and PIG distributions and regression models with varying dispersion, which have a less complicated likelihood than the proposed model, were fitted using the generalized additive models for the location, scale, and shape (GAMLSS) package in R, see Stasinopoulos et al. (2008b).

Furthermore, the computational time requirements of the PIGA distribution/regression model with varying dispersion were compared to those of the NBI and PIG distributions/regression models with varying dispersion. As anticipated, the NBI and PIG distributions/regression models with varying dispersion compared significantly more favorably to the PIGA distribution/regression model with varying dispersion in terms of computing times required for ML estimation, since it took all four cases fewer than 30 s of CPU time. On the other hand, it took the PIGA distribution less 30 min. of CPU time, while the PIGA regression model with varying dispersion exceeded 1 h of CPU time. However, it should be taken into account that there were 14,143 observations in the sample of MTPL data that was examined in this article, which we used a rather strict stopping criterion for EM iterations and that the numerical approximation of the first derivative of the modified Bessel function with respect to its order in the E-step of the algorithm is very chronologically demanding, especially for the case when both parameters of the PIGA distribution are modelled in terms of covariates.

Finally, it is worth mentioning that, for larger data sets with more covariates, the computing effort can be substantially reduced if the E- and the M-steps are executed in parallel across multiple threads to take advantage of the processing power of modern-day multicore machines.

6. Concluding Remarks

This paper proposed an EM type algorithm for estimating the parameters of the PIGA regression model with varying dispersion. The thick tail of the distribution combined with the adopted modelling framework, which allows both of its parameters to be modelled as functions of important risk factors, can provide an advantage relative to previous approaches in the insurance ratemaking literature when the claim frequency data are heavy-tailed and overdispersed.

Furthermore, it is worth mentioning that a possible line of further research could be to include functional forms other than the linear in the mean and the dispersion parameters of the PIGA response distribution following the generalized additive models for the location, scale, and shape (GAMLSS) approach of Rigby and Stasinopoulos (2005). Additionally, see, for instance, De Jong and Heller (2008). Additionally, the data augmentation which was used in the article to derive the EM algorithm can be the basis for constructing estimation methods within the Bayesian framework proceeding along similar lines as Klein et al. (2014), who used Bayesian generalized additive models for the location, scale, and shape for nonlife ratemaking and risk management.

Finally, it would be interesting to design a priori and a posteriori, or Bonus-Malus, ratemaking mechanisms based on two component mixture models where the first component distribution is the PIGA and the second component distribution can be less heavy-tailed, such as the NBI or the PIG, and allowing regression structures on the mixing probabilities and all of the parameters of the two component models, see Tzougas et al. (2018). The ML estimates of these models can be easily obtained using standard techniques for finite mixtures, see Böhning (1999).

References

Abramowitz, Milton, and Irene A. Stegun. 1965. *Handbook of Mathematical Functions*. New York: Dover Publications, pp. 241–47.

Altun, Emrah. 2019. A new model for over-dispersed count data: Poisson quasi-Lindley regression model. *Mathematical Sciences* 13: 241–47. [CrossRef]

Barreto-Souza, Wagner, and Alexandre B. Simas. 2016. General mixed Poisson regression models with varying dispersion. *Statistics and Computing* 26: 1263–80. [CrossRef]

Böhning, Dankmar. 1999. *Computer Assisted Analysis of Mixtures and Applications in Meta-Analysis, Disease Mapping and Others*. Boca Raton CRC Press.

Boucher, Jean-Philippe, Michel Denuit, and Montserrat Guillén. 2007. Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal* 11: 110–31. [CrossRef]

Boucher, Jean-Philippe, Michel Denuit, and Montserrat Guillén. 2008. Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. *Variance* 2: 135–62.

Brouhns, Natacha, Montserrat Guillén, Michel Denuit, and Jean Pinquet. 2003. Bonus-Malus Scales in Segmented Tariffs With Stochastic Migration Between Segments.*Journal of Risk and Insurance* 70: 577–99. [CrossRef]

Burnham, Kenneth P., and David R. Anderson. 2002. *A Practical Information-Theoretic Approach. Model Selection and Multimodel Inference*, 2nd ed. New York: Springer.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, vol. 53.

Cohen, A. Clifford, Jr. 1966. A note on certain discrete mixed distributions. *Biometrics* 22: 566–72. [CrossRef]
[PubMed]

Cole, Timothy J., and Pamela J. Green. 1992. Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* 11: 1305–19. [CrossRef]

De Jong, Piet, and Gillian Z. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press, vol. 10.

Dean, Charmaine, J. F. Lawless, and G. E. Willmot. 1989. A mixed poisson–inverse-gaussian regression model. *Canadian Journal of Statistics* 17: 171–81. [CrossRef]

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39: 1–22.

Denuit, Michel, and Stefan Lang. 2004. Non-life rate-making with Bayesian GAMs. *Insurance: Mathematics and Economics* 35: 627–47. [CrossRef]

Denuit, Michel, Xavier Maréchal, Sandra Pitrebois, and Jean-François Walhin. 2007. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Hoboken: John Wiley & Sons.

Dionne, Georges, and Charles Vanasse. 1989. A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component. *ASTIN Bulletin* 19: 199–212. [CrossRef]

Dionne, Georges, and Charles Vanasse. 1992. Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics* 7: 149–65. [CrossRef]

Dunn, Peter K., and Gordon K. Smyth. 1996. Randomized quantile residuals. *Computational and Graphical Statistics* 5: 236–45.

Frangos, Nicholas E., and Spyridon D. 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: 1–22. [CrossRef]

Frees, Edward W. 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge: Cambridge University Press.

Gilbert, Paul, and Ravi Varadhan. 2019. Accurate Numerical Derivatives R Package Manual. Version 2016.8-1.1. Available online: https://cran.r-project.org/web/packages/numDeriv/numDeriv.pdf (accessed on 6 June 2019).

Giuricich, Mario Nicoló, and Krzysztof Burnecki. 2019. Modelling of left-truncated heavy-tailed data with application to catastrophe bond pricing. *Physica A: Statistical Mechanics and Its Applications* 525: 498–13. [CrossRef]

Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2016. The Mixture Poisson Exponential-Inverse Gaussian Regression Model: An application in Health Services. *Metodoloski Zvezki* 13: 71.

Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2018a. Properties and applications of the Poisson-reciprocal inverse gaussian distribution. *Journal of Statistical Computation and Simulation* 88: 269–89. [CrossRef]

Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2018b. Multivariate Credibility in Bonus-Malus Systems Distinguishing between Different Types of Claims. *Risks* 6: 34. [CrossRef]

Gómez-Déniz, Emilio, and F. J. Vázquez-Polo. 2003. Robustness in Bayesian model for Bonus-Malus systems. In *Intelligent And Other Computational Techniques In Insurance: Theory and Applications*. Toh Tuck: World Scientific Publishing, pp. 435–63.

Gómez-Déniz, Emilio, Agustín Hernández-Bastida, and M. Pilar Fernández-Sánchez. 2014. Computing credibility bonus-malus premiums using the aggregate claims distribution. *Hacettepe Journal of Mathematics and Statistics* 43: 1047–61.

Gómez-Déniz, Emilio, José María Sarabia, and Enrique Calderín-Ojeda. 2008. Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics* 42: 39–49. [CrossRef]

Gómez-Déniz, Emilio, M.E. Ghitany, and Ramesh C. Gupta. 2016. Poisson-mixed inverse gaussian regression model and its application. *Communications in Statistics - Simulation and Computation* 45: 2767–81. [CrossRef]

Haberman, Steven, and Arthur E. Renshaw. 1996. Generalized linear models and actuarial science. *Journal of the Royal Statistical Society: Series D (The Statistician)* 45: 407–36. [CrossRef]

Hilbe, Joseph M. 2008. *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Insurance Europe. European Motor Insurance Markets 2019. Available online: https://www.insuranceeurope.eu/european-motor-insurance-markets-2019 (accessed on 8 February 2019).

Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan. 1994. *Continuous Univariate Distributions*. Hoboken: John Wiley & Sons, vol. 1.

Jørgensen, Bent. 2012. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Cham: Springer Science & Business Media, vol. 9.

Kaas, Rob, Marc Goovaerts, Jan Dhaene, and Michel Denuit. 2008. *Modern Actuarial Risk Theory: Using R*. Cham: Springer Science & Business Media, vol. 128.

Karlis, Dimitris, George Tzougas, and Nicholas Frangos. 2018. Confidence intervals of the premiums of optimal bonus malus systems. *Scandinavian Actuarial Journal* 2: 129–44. [CrossRef]

Karlis, Dimitris. 2001. A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling* 1: 305–18. [CrossRef]

Karlis, Dimitris. 2005. EM algorithm for mixed Poisson and other discrete distributions. *ASTIN Bulletin* 35: 3–24. [CrossRef]

Khazraee, S. Hadi, Valen Johnson, and Dominique Lord. 2018. Bayesian Poisson hierarchical models for crash data analysis: Investigating the impact of model choice on site-specific predictions. *Accident Analysis & Prevention* 117: 181–95.

Klein, Nadja, Michel Denuit, Stefan Lang, and Thomas Kneib. 2014. Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape. *Insurance: Mathematics and Economics* 5: 225–49. [CrossRef]

Lambert, Diane. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34: 1–14. [CrossRef]

Lawless, Jerald F. 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15: 209–25. [CrossRef]

Lemaire, Jean, Sojung Carol Park, and Kili C. Wang. 2015. The Impact of Covariates on a Bonus-Malus System: An Application of Taylor's Model. *European Actuarial Journal* 5: 1–10. [CrossRef]

Lemaire, Jean, Sojung Carol Park, and Kili C. Wang. 2016. The Use of Annual Mileage as a Rating Variable. *ASTIN Bulletin* 46: 39–69. [CrossRef]

Lemaire, Jean. 1995. *Bonus-Malus Systems in Automobile Insurance*. New York: Kluwer Academic Publishers.

Louis, Thomas A. 1982. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 44: 226–33.

Mahmoudvand, Rahim, and Hossein Hassani. 2009. Generalized bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin* 39: 307–15. [CrossRef]

Majeske, Karl D. 2007. A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering & System Safety* 92: 243–51.

McLachlan, Geoffrey J., and Thriyambakam Krishnan. 2007. *The EM Algorithm and Extensions*. Hoboken: John Wiley & Sons, vol. 38.

Mencía, Javier, and Enrique Sentana. 2005. *Estimation and Testing of Dynamic Models with Generalized Hyperbolic Innovations*. CEPR Discussion Papers, No. 5177. London: Centre for Economic Policy Research (CEPR).

Mert, Mehmet, and Yasemin Saykan. 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.

Ni, Weihong, Bo Li, Corina Constantinescu, and Athanasios A. Pantelous. 2014. Bonus-Malus Systems with Hybrid Claim Severity Distributions. In *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*. Liverpool: American Society of Civil Engineers, pp. 1234–44.

Ni, Weihong, Corina Constantinescu, and Athanasios A. Pantelous. 2014. Bonus-Malus systems with Weibull distributed claim severities. *Annals of Actuarial Science* 8: 217–233. [CrossRef]

Ord, J. Keith, and G. Alex Whitmore. 1986. The Poisson-inverse gaussian distribution as a model for species abundance. *Communications in Statistics-theory and Methods* 15: 853–71. [CrossRef]

Perline, Richard. 1998. Mixed Poisson Distributions Tail Equivalent to their Mixing Distributions. *Statistics and Probability Letters* 38: 229–33. [CrossRef]

Picech, Liviana. 1994. The Merit-Rating Factor in a Multiplicating Rate-Making model. In *ASTIN Colloquium, Cannes*. Cannes: Cambridge University Press.

Pinquet, Jean. 1997. Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin* 27: 33–57. [CrossRef]

Pinquet, Jean. 1998. Designing Optimal Bonus-Malus Systems From Different Types of Claims. *ASTIN Bulletin* 28: 205–20. [CrossRef]

Raftery, Adrian E. 1995. Bayesian model selection in social research. *Sociological Methodology* 25: 111–63. [CrossRef]

Rigby, Robert A., and D. M. Stasinopoulos. 1996. A semi-parametric additive model for variance heterogeneity. *Statistics and Computing* 6: 57–65. [CrossRef]

Rigby, Robert A., and Dimitrios M. Stasinopoulos. 1996. Mean and dispersion additive models. In *Statistical Theory and Computational Aspects of Smoothing*. Edited by Wolfgang Härdle and Michael Schimek. Heidelberg: Physica, pp. 215–30.

Rigby, Robert A., and Dimitrios M. Stasinopoulos. 2005. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54: 507–54. [CrossRef]

Rigby, Robert A., Dimitrios M. Stasinopoulos, and Calliope Akantziliotou. 2008a. A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Computational Statistics & Data Analysis* 53: 381–93.

Rigby, Robert A., Dimitrios M. Stasinopoulos, and Calliope Akantziliotou. 2008b. Instructions on How to Use the Gamlss Package in R, 2nd Edition. Available online: http://www.gamlss.org (accessed on 11 January 2008).

Romaniuk, Maciej. 2020. Imprecise Approaches to Analysis of Insurance Portfolio with Catastrophe Bond. Paper presented at International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Lisbon, Portugal, 15–19 June 2020. Cham: Springer.

Schiegl, Magda. 2010. About the Justification of Experience Rating: Bonus Malus System and a new Poisson Mixture Model. *arXiv*, arXiv:1009.4142.

Shared, Moshe. 1980. On mixtures from exponential families. *Journal of the Royal Statistical Society: Series B (Methodological)* 42: 192–98. [CrossRef]

Trembley, Luc. 1992. Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bulletin* 22: 97–106. [CrossRef]

Tzougas, George, and Dimitris Karlis. 2020. An EM algorithm for fitting a new class of mixed exponential regression models with varying dispersion. *Astin Bulletin* 50: 555–83. [CrossRef]

Tzougas, George, Spyridon Vrontos, and Nicholas Frangos. 2014. Optimal Bonus-Malus Systems Using Finite Mixture Models. *ASTIN Bulletin* 44: 417–44. [CrossRef]

Tzougas, George, Spyridon Vrontos, and Nicholas Frangos. 2015. Risk Classification for Claim Counts and Losses Using Regression Models for Location, Scale and Shape. *Variance* 9: 140–57.

Tzougas, George, Spyridon Vrontos, and Nicholas Frangos. 2018. Bonus-Malus Systems with Two-Component Mixture Models Arising from Different Parametric Families. *North American Actuarial Journal* 22: 55–91. [CrossRef]

Tzougas, George, Wei Li Hoon, and Jun Ming Lim. 2019. The negative binomial-inverse Gaussian regression model with an application to insurance ratemaking. *European Actuarial Journal* 9: 323–44. [CrossRef]

Weibel, Mark, David Lóthi, and Wolfgang Breymann. 2020. ghyp: Generalized Hyperbolic Distribution and Its Special Cases. Available online: https://cran.r-project.org/web/packages/ghyp/ghyp.pdf (accessed on 27 April 2020).

Willmot, Gordon E. 1987. The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal* 3–4: 113–27. [CrossRef]

Willmot, Gordon E. 1990. Asymptotic tail behaviour of Poisson mixtures with applications. *Advances in Applied Probability* 22: 147–59. [CrossRef]

Willmot, Gordon E. 1993. On recursive evaluation of mixed Poisson probabilities and related quantities. *Scandinavian Actuarial Journal* 2: 114–33.

Yip, Karen C. H., and Kelvin K. W. Yau. 2005. On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics* 36: 153–63. [CrossRef]