Article

Mixed Poisson Regression Models with Varying Dispersion Arising from Non-Conjugate Mixing Distributions

George Tzougas [1,2,*], Natalia Hong [1,†] and Ryan Ho [1,†]

1 Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh TD1 3HE, UK; natlhongg@gmail.com (N.H.); ryanhkm5.13@gmail.com (R.H.)
2 Department of Statistics, London School of Economics and Political Science, London TD1 3HE, UK
* Correspondence: George.Tzougas@hw.ac.uk
† Work undertaken whilst at the London School of Economics.

Abstract: In this article we present a class of mixed Poisson regression models with varying dispersion arising from non-conjugate to the Poisson mixing distributions for modelling overdispersed claim counts in non-life insurance. The proposed family of models combined with the adopted modelling framework can provide sufficient flexibility for dealing with different levels of overdispersion. For illustrative purposes, the Poisson-lognormal regression model with regression structures on both its mean and dispersion parameters is employed for modelling claim count data from a motor insurance portfolio. Maximum likelihood estimation is carried out via an expectation-maximization type algorithm, which is developed for the proposed family of models and is demonstrated to perform satisfactorily.

Keywords: EM algorithm; regression structures on the mean and dispersion parameters; non-life insurance; claim frequency

1. Introduction

During the last three decades, mixed Poisson regression models have been applied in various fields of studies, including non-life insurance for modelling overdispersed claim count data. The members of this family of models, which can be constructed based on a mixing distribution which is conjugate to the Poisson distribution, such as the negative binomial and the Poisson–inverse-Gaussian, have been the most popular choices due to the simplicity of their log-likelihood functions, which can be easily maximized using the standard maximum likelihood (ML) estimation approach. See, for example, Refs. [1–3] for the former and [4–6] for the latter, among many others. However, it should be noted that the assumption of conjugacy can be very restrictive for constructing a mixed Poisson model that will be able to efficiently capture different levels of overdispersion in real claim count data sets. In particular, as is well known, overdispersion is a direct consequence of unobserved heterogeneity due to systematic effects in the data. For example, in motor insurance, the driving skills, preferences, habits, and driving experience of policyholders, which differ, may lead to extra variation in the claim count data, the degree of which is controlled by the value of the dispersion parameter of the mixed Poisson model. Furthermore, overdispersion can either be caused by a large presence of zeros or a heavy tail in the data. Regarding the latter case, as is well known, the tails of mixed Poisson distributions in the case of continuous mixing distributions are similar to the tails of their mixing distributions (see, for instance, Ref. [7]). Therefore, restricting attention only to mixed Poisson models which are derived based on conjugate to the Poisson mixing densities may result in biased parameter estimates because of their inability to always efficiently model the tail of claim count distribution. This, in turn, may have a profound impact on many tasks which are carried out by the actuaries, such as risk management and pricing of (re-)insurance contracts. Thus, it becomes clear that an important task of actuaries is to be able to design more

representative probabilistic models for the number of claims with good prediction accuracy. This procedure depends on the reliability of the statistical method which will be used to construct them.

The aim of this article is to present a general class of mixed Poisson regression models with varying dispersion stemming from non-conjugate C^2 densities with a continuous first derivative and a continuous second derivative. The class of mixed Poisson models we consider is very wide and the flexibility it provides in (i) the distributional choice for the mixing density and (ii) modelling jointly the mean and dispersion parameters as parametric functions of risk factors allows us to add the required amount of weight to the right tail area of the claim count distribution for accommodating different levels of overdispersion, thus resulting in an improved risk evaluation. At this point, it should be noted that, with the exception of very few articles, such as those by [8–10], modelling jointly all the parameters of mixed Poisson models in terms of explanatory variables has not been explored in depth. Nevertheless, mean regression models often cannot adequately account for the heteroscedasticity of the claim count distribution or its possible dependence on risk factors. In addition, note that in [9], the exponential family distribution assumption for the univariate response variable is relaxed and replaced by a general distribution family, including distributions based on Box–Cox transformations (such as the Box–Cox t-distribution or the Box–Cox power exponential distribution) and zero adjusted-distributions. However, to the best of our knowledge, this is the first study to consider regression structures on the mean and dispersion parameters of univariate mixed Poisson models that have a probability mass function (pmf) which cannot be written in closed-form expressions. For demonstration purposes, the Poisson-lognormal (PLN) regression model with varying dispersion is fitted on a motor third-party liability (MTPL) insurance claim count data set using an expectation-maximization (EM) type algorithm, which takes advantage of the stochastic mixture representation of the proposed family of models which have a density that cannot be written in closed form in an easy and efficient manner. Moreover, it is worth noting that the development of stochastic algorithms, such as the EM and stochastic gradient descent algorithms, is of particular importance in machine learning and artificial intelligence applications, as they can be employed for efficiently calibrating various statistical models and deep networks. Two very interesting recent articles are those by [11,12].

The rest of this article is structured as follows. In Section 2, we present the derivation of the proposed mixed Poisson regression model with varying dispersion for claim counts. Section 3 deals with the ML estimation procedure for the PLN regression model with varying dispersion based on the proposed EM-type algorithm. Section 4 contains an application to the MTPL claim count data, and we fit the PLN claim count regression model with varying dispersion. In addition, the negative binomial regression model with varying dispersion and the zero-inflated (ZIP) Poisson regression model are used as benchmarks for comparison. Finally, concluding remarks are provided in Section 5.

2. Mixed Poisson Regression Model with Varying Dispersion

2.1. Modelling Framework

Consider a non-life insurance portfolio with n policyholder contracts each involving a particular claim type, and assume that the individual claim frequencies, k_i, arising from each insured i, for $i = 1, \ldots, n$, are independent. In addition, suppose that given a continuous random variable, $z_i > 0$, $k_i|z_i$ is distributed according to a Poisson distribution with probability mass function (pmf) given by

$$P(k_i|z_i) = \frac{\exp(-\mu_i z_i)(\mu_i z_i)^{k_i}}{k_i!},$$ (1)

for $k_i = 0, 1, 2, \ldots$, and where $\mu_i > 0$. The mean and variance of $k_i|z_i$ are $\mathrm{E}(\mathrm{k_i}|\mathrm{z_i}) = \mathrm{Var}(\mathrm{k_i}|\mathrm{z_i}) = \mu_i z_i$.

Furthermore, consider that z_i is distributed according to a C^2 mixing distribution, with probability density function (pdf) $g(z_i; \phi_i)$, with $\phi_i > 0$, which is not conjugate to the Poisson distribution given by Equation (1). Additionally, we assume that z_i has a unit mean, that is, $E(z_i) = 1$, to ensure that the model is identifiable.

Considering the previous assumptions, we see that the resulting distribution of k_i is a mixed Poisson distribution with pmf

$$P(k_i) = \int_0^\infty P(k_i|z_i) g(z_i; \phi_i) \mathrm{d}z_i, \tag{2}$$

where $g(z_i; \phi_i)$ is the pdf of z_i. In addition, the mean and the variance of k_i are given by

$$\mathrm{E}(k_i) = \mathrm{E}_{z_i}[\mathrm{E}(k_i|z_i)] = \mu_i \mathrm{E}_{z_i}[z_i] = \mu_i \tag{3}$$

and

$$\mathrm{Var}(k_i) = \mathrm{E}_{z_i}[\mathrm{Var}(k_i|z_i)] + \mathrm{Var}_{z_i}[\mathrm{E}(k_i|z_i)]. \tag{4}$$

Finally, under the proposed modelling framework, the mean and dispersion parameters, μ_i and ϕ_i, are modelled as functions of risk factors

$$\mu_i = \exp(\mathbf{x}_{1,i}^\top \boldsymbol{\beta}_1) \tag{5}$$

and

$$\phi_i = \exp(\mathbf{x}_{2,i}^\top \boldsymbol{\beta}_2), \tag{6}$$

where $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are the, potentially different, vectors of explanatory variables with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, and where $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, \ldots, \beta_{1,p_1}\right)^\top$ and $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, \ldots, \beta_{2,p_2}\right)^\top$ are vectors of regression coefficients, where we consider that the matrices $\mathbf{X_1}$ and $\mathbf{X_2}$ are of full rank and are composed of the rows given by $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$, respectively.

2.2. Model Specification: The Poisson-Lognormal Regression Model with Varying Dispersion

For expository purposes, we specify the lognormal distribution as the mixing distribution of z_i with the following pdf

$$g(z_i; \phi_i) = \frac{1}{\sqrt{2\pi}\phi_i z_i} \exp\left[-\frac{1}{2\phi_i^2} \left(\log(z_i) + \frac{\phi_i^2}{2} \right)^2 \right], \tag{7}$$

where $z_i > 0$ and $\phi_i > 0$, with mean $E(z_i) = 1$ and variance $Var(z_i) = \exp(\phi_i^2) - 1$.

Then, based on Equations (1) and (7), it is easy to see that the resulting distribution of k_i is the Poisson-lognormal (PLN) distribution with pmf

$$P(k_i) = \int_0^\infty \frac{\exp(-\mu_i z_i)(\mu_i z_i)^{k_i}}{k_i!} \frac{\exp\left[-\frac{1}{2\phi_i^2} \left(\log(z_i) + \frac{\phi_i^2}{2} \right)^2 \right]}{\sqrt{2\pi}\phi_i z_i} \mathrm{d}z_i, \tag{8}$$

where $\mu_i > 0$ and $\phi_i > 0$ are given by Equations (5) and (6), respectively. Unfortunately, since $g(z_i; \phi_i)$ is not conjugate to the Poisson, the integral in Equation (8) is mathematically intractable, but it can be easily calculated using numerical integration.

Using the results in Equations (3) and (4), we calculate the mean and the variance of the PLN regression model with varying dispersion

$$\mathrm{E}(k_i) = \mathrm{E}_{z_i}[\mathrm{E}(k_i|z_i)] = \mu_i \mathrm{E}_{z_i}[z_i] = \mu_i \tag{9}$$

and

$$\mathrm{Var}(k_i) = \mathrm{E}_{Z_i}[\mathrm{Var}(k_i|Z_i = z_i)] + \mathrm{Var}_{Z_i}[\mathrm{E}(k_i|Z_i = z_i)]$$
$$= \mu_i + \mu_i^2\left[\exp\left(\phi_i^2\right) - 1\right]. \tag{10}$$

3. Statistical Inference: The EM-Type Algorithm

Let $(k_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$, $i = 1, \ldots, n$, be a sample of independent observations, where k_i is the response variable, and $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are the vectors of explanatory variables with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively. In addition, suppose that the data are produced according to the mixed Poisson model with varying dispersion. Then, the log-likelihood of the model can be written as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(P(k_i)), \tag{11}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$ is the vector of the parameters, and where $P(k_i)$ is the pmf of the model which is given by Equation (2). It should be noted that the likelihood given by Equation (11) is cumbersome to maximize since it is not usually tractable. Moreover, when the mean and dispersion parameters are modelled in terms of risk factors, additional computational challenges can be encountered.

However, ML estimation can be accomplished relatively easily via an EM-type algorithm. In particular, if the unobserved data z_i are augmented to the observed data $(k_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$, for $i = 1, \ldots, n$, then the complete data log-likelihood factorizes into two parts

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[-\mu_i z_i + k_i \log(\mu_i) + k_i \log(z_i) - \log(k_i!)\right] + \sum_{i=1}^{n} \log(g(z_i; \phi_i)) \tag{12}$$

where $g(z_i; \phi i)$ is the pdf of the mixing distribution which is not conjugate to the Poisson, and where μ_i and ϕ_i are given by Equations (5) and (6) respectively.

The EM-type algorithm for the mixed Poisson regression model with varying dispersion can be described as follows

- **E-Step**: The Q-function, which is the conditional expectation of the complete data log-likelihood, is given by

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}) = E_{z_i}\left(l_c(\boldsymbol{\theta})|\boldsymbol{K}, \boldsymbol{\theta}^{(r)}\right) \propto \tag{13}$$

$$\sum_{i=1}^{n}[-\mu_i^{(r)} E_{z_i}(z_i|k_i; \boldsymbol{\theta}^{(r)}) + k_i \log(\mu_i^{(r)})] \tag{14}$$

$$+ \sum_{i=1}^{n} E_{z_i}[\log(g(z_i; \phi_i^{(r)}))] \tag{15}$$

where $\boldsymbol{\theta}^{(r)}$ is the estimate of $\boldsymbol{\theta}$ at the rth iteration in the E-step of our EM algorithm. Then, using the estimates $\boldsymbol{\theta}^{(r)}$, calculate the pseudo-values $w_{1_i} = E_{z_i}(z_i|k_i; \boldsymbol{\theta}^{(r)})$ and $w_{k_i} = E_{z_i}(s_k(Z_i)|k_i; \boldsymbol{\theta}^{(r)})$ for $i = 1, \ldots, n$ and $k = 1, \ldots, \nu$, where $s_k(.)$ are certain functions which are involved in the terms needed for maximizing the part of the Q-function which corresponds to the conditional expectation of the log-likelihood of the mixing distribution $g(z_i; \phi_i)$.

- **M-Step**: Using the pseudo-values w_{1_i} and w_{k_i} from the E-Step and the Newton–Raphson algorithm twice, find the maximum global point $\boldsymbol{\theta}^{(r+1)}$ of the Q-function, that is, obtain the updated estimates $\boldsymbol{\beta}_1^{(r+1)}$ and $\boldsymbol{\beta}_2^{(r+1)}$.

 –   Firstly, taking the necessary derivatives of the Q-function with respect to $\boldsymbol{\beta}_1$, we obtain the following results

$$h_1(\boldsymbol{\beta}_1) = \frac{\partial Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{1,j}} = \sum_{i=1}^{n} \left(-\mu_i^{(r)} w_{1_i} + k_i\right) x_{1,ij}, \tag{16}$$

and

$$H_1(\boldsymbol{\beta}_1) = \frac{\partial^2 Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{1,j} \partial \beta_{1,j}^{\top}} = \sum_{i=1}^{n} \left(-\mu_i^{(r)} w_{1_i}\right) x_{1,ij} x_{1,ij}^{\top} = \boldsymbol{X}_1^{\top} \boldsymbol{W}_1 \boldsymbol{X}_1, \tag{17}$$

for $i = 1, \dots, n$ and $j = 1, \dots, p_1$, and where $\boldsymbol{W}_1 = \text{diag}\left\{-\frac{k_i}{\mu_i^{(r)}} w_{1_i}\right\}$.

Then, the iterative procedure for the Newton–Raphson algorithm for $\boldsymbol{\beta}_1$ goes as follows

$$\boldsymbol{\beta}_1^{(r+1)} \equiv \boldsymbol{\beta}_1^{(r)} - \left[H_1\left(\boldsymbol{\beta}_1^{(r)}\right)\right]^{-1} h_1\left(\boldsymbol{\beta}_1^{(r)}\right). \tag{18}$$

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

$$h_2(\boldsymbol{\beta}_2) = \frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)})}{\partial \beta_{2,l}} = \frac{\partial \mathrm{E}_{z_i}[\log g(z_i; \phi_i^{(r)})]}{\partial \beta_{2,l}} \tag{19}$$

and

$$H_2(\boldsymbol{\beta}_2) = \frac{\partial^2 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)})}{\partial \beta_{2,l} \partial \beta_{2,l}^{\top}} = \frac{\partial \mathrm{E}_{z_i}[\log g(z_i; \phi_i^{(r)}]}{\partial \beta_{2,l} \partial \beta_{2,l}^{\top}}, \tag{20}$$

where for computing $h_1(\boldsymbol{\beta}_2)$ and $H_2(\boldsymbol{\beta}_2)$, we need to use the pseudo-values w_{k_i} for $i = 1, \dots, n$ and $k = 1, \dots, \nu$, because in this case, the maximization of the Q-function reduces to the maximization of the conditional expectation of the log-likelihood of the mixing distribution $g(z_i; \phi_i)$.

Then, the Newton–Raphson iterative algorithm for $\boldsymbol{\beta}_2$ is as follows

$$\boldsymbol{\beta}_2^{(r+1)} = \boldsymbol{\beta}_2^{(r)} - [H_2(\boldsymbol{\beta}_2^{(r)})]^{-1} h_2(\boldsymbol{\beta}_2^{(r)}), \tag{21}$$

for $i = 1, \dots, n$ and $l = 1, \dots, p_2$.

•   Finally, iterate between the E- and the M-Steps until some convergence criterion is satisfied, for instance

$$\left| \frac{l^{(r+1)} - l^{(r)}}{l^{(r)}} \right| < tol, \tag{22}$$

where $l^{(r)}$ is the value of the log-likelihood after the r-th iteration, and where tol is a small number usually of the form 10^{-m}, where $m \in \mathbb{Z}^+$. The stopping criterion refers to the progress of the likelihood function (i.e., its convergence). If the stopping criterion is satisfied, the EM algorithm stops iterating, and the estimate of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(r+1)}$. Otherwise, $\boldsymbol{\theta}$ is updated by $\boldsymbol{\theta}^{(r+1)}$, and the algorithm returns to the E-step.

EM Estimation for the PLN Regression Model with Varying Dispersion

In this section, we implement the EM algorithm for finding the ML estimates of the parameters of the PLN regression model with varying dispersion (Algorithm 1). The complete data log-likelihood of the model is given by

$$
l_c(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n}[-\mu_i z_i + k_i \log(\mu_i) + k_i \log(z_i) - \log(k_i!)] +
$$

$$
\sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi) - \log(\phi_i) - \log(z_i) - \frac{1}{2\phi_i^2}\left(\log(z_i) + \frac{\phi_i^2}{2} \right)^2 \right], \qquad (23)
$$

for $i = 1, \ldots, n$. Thus, the posterior expectations needed for the E-step are $\mathrm{E}_{z_i}\left[z_i | k_i; \boldsymbol{\theta}^{(r)} \right]$ and $\mathrm{E}_{z_i}\left[(\log(z_i))^2 | k_i; \boldsymbol{\theta}^{(r)} \right]$, while at the M-step one needs to maximize the expected value of $l_c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In particular, more formally, the EM-type algorithm can be written as follows.

Algorithm 1 EM Algorithm for the PLN Regression Model with Varying Dispersion

---

1. Provide initial values $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}_1 0, \boldsymbol{\beta}_2^0)$.
2. (E-step) Update the conditional expectations $w_{1_i} = \mathrm{E}\left[z_i | k_i; \boldsymbol{\theta}^{(r)} \right]$ and $w_{2_i} = \mathrm{E}\left[\log(z_i)^2 | k_i; \boldsymbol{\theta}^{(r)} \right]$ using $\boldsymbol{\theta}^{(r)}$, for $i = 1, \ldots, n$ from the rth iteration.
3. (M-step) Find the maximum global point, $\boldsymbol{\theta}^{(r+1)}$, of the log-likelihood function $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)})$.
4. If the criterion $\left| \frac{l^{(r+1)} - l^{(r)}}{l^{(r)}} \right| < tol$ is satisfied, the estimate of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(r+1)}$. Otherwise, update $\boldsymbol{\theta}^{(r)}$ by $\boldsymbol{\theta}^{(r+1)}$ and return to step 2.

- **E-Step:**
  Calculate, for all $i = 1, \ldots, n$,

$$
w_{1_i} = \mathrm{E}\left[z_i | k_i; \boldsymbol{\theta}^{(r)} \right]
$$

$$
= \frac{\displaystyle\int_0^\infty z_i^{k_i} \exp\left[-\frac{1}{2(\phi_i^2)^{(r)}}\left(\log(z_i) + \frac{(\phi_i^2)^{(r)}}{2} \right)^2 - \mu_i^{(r)} z_i \right] dz_i}{\displaystyle\int_0^\infty z_i^{k_i-1} \exp\left[-\frac{1}{2(\phi_i^2)^{(r)}}\left(\log(z_i) + \frac{(\phi_i^2)^{(r)}}{2} \right)^2 - \mu_i^{(r)} z_i \right] dz_i} \qquad (24)
$$

and

$$
w_{2_i} = \mathrm{E}\left[\log(z_i)^2 | k_i; \boldsymbol{\theta}^{(r)} \right]
$$

$$
= \frac{\displaystyle\int_0^\infty (\log(z_i))^2 z_i^{k_i-1} \exp\left[-\frac{1}{2(\phi_i^2)^{(r)}}\left(\log(z_i) + \frac{(\phi_i^2)^{(r)}}{2} \right)^2 - \mu_i^{(r)} z_i \right] dz_i}{\displaystyle\int_0^\infty z_i^{k_i-1} \exp\left[-\frac{1}{2(\phi_i^2)^{(r)}}\left(\log(z_i) + \frac{(\phi_i^2)^{(r)}}{2} \right)^2 - \mu_i^{(r)} z_i \right] dz_i} \qquad (25)
$$

where $\mu_i^{(r)} = \exp\left(\boldsymbol{x}_{1,i}^\top \boldsymbol{\beta}_1^{(r)} \right)$ and $\phi_i^{(r)} = \exp\left(\boldsymbol{x}_{2,i}^\top \boldsymbol{\beta}_2^{(r)} \right)$.

Note that the expectations in Equations (24) and (25) can be evaluated numerically. Alternatively, a Monte Carlo approach can be used based on a rejection algorithm, leading to variants of the EM algorithm, such as the Monte Carlo EM (MCEM) algorithm, which do not rely on the pdf $g(k_i|z_i)$, that cannot be written in closed form, but it is sufficent to simulate from the posterior distribution $g(z_i|k_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$.

- **M-Step**:

  - Firstly, the regression parameters $\boldsymbol{\beta}_1$ are updated using the pseudo-values w_{1_i}, which are given by Equation (24), and the Newton–Raphson algorithm, which is given in Equations (16)–(18).

  - Secondly, the regression parameters $\boldsymbol{\beta}_2$ are updated using the pseudo-values w_{1_i} and w_{2_i}, which are given by Equations (24) and (25), respectively, and the Newton–Raphson algorithm, which, in the case of the lognormal mixing distribution, is as follows

$$h_2(\boldsymbol{\beta}_2) = \left[\frac{w_{2_i}}{\left(\phi_i^2\right)^{(r)}} - \frac{\left(\phi_i^2\right)^{(r)}}{4} - 1 \right] x_{2,il}, \tag{26}$$

and

$$H_2(\boldsymbol{\beta}_2) = \sum_{i=1}^{n} \left[\frac{-2w_{2_i}}{\left(\phi_i^2\right)^{(r)}} - \frac{\left(\phi_i^2\right)^{(r)}}{2} \right] x_{2,il} x_{2,il}^\top = \mathbf{X}_2^\top \mathbf{W}_2 \mathbf{X}_2, \tag{27}$$

for $i = 1, \ldots, n$ and $l = 1, \ldots, p_2$, and where $\mathbf{W}_2 = \mathrm{diag}\left\{ \frac{-2\omega_{2,i}}{\left(\phi_i^2\right)^{(r)}} - \frac{\left(\phi_i^2\right)^{(r)}}{2} \right\}$.

Then, we can obtain the updated estimates of $\boldsymbol{\beta}_2^{(r)}$ as follows

$$\boldsymbol{\beta}_2^{(r+1)} \equiv \boldsymbol{\beta}_2^{(r)} - \left[H_2\left(\boldsymbol{\beta}_2^{(r)}\right) \right]^{-1} h_2\left(\boldsymbol{\beta}_2^{(r)}\right). \tag{28}$$

4. Numerical Illustration

This study was based on a subset of claim frequency data from a pool of MTPL insurance policies observed for 3.5 years from a major Greek insurance company. A total of 14,143 observations with complete records (i.e., with availability of all the explanatory variables) were taken for our analysis. The response variable is the number of claims at fault registered for each insured vehicle. In addition, a subset of explanatory variables with the highest predictive power for the response variable was chosen based on exploratory analysis. In particular, we considered the following covariates: the age of the driver (AD), the horsepower (HP) of their car, and the age of their car (AC). Additionally, we grouped the levels of each a priori rating variable with respect to risk profiles with similar claim frequency in order to balance homogeneity as well as sufficiency of the volume of data in each cell.

The summary of the explanatory variables and their corresponding groupings with the number of observations in each category along with the descriptive statistics for claim counts are shown in Table 1.

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 the following subsection, we fit the Poisson-lognormal (PLN) regression model on the number of claims. Moreover, we will compare its fit with those of the classic negative binomial type I (NBI) distribution, which has been used in an abundance of actuarial settings for approximating claim counts, for the case when regression components are introduced on its mean and dispersion parameters. Finally, the high presence of zeros in the MTPL data set 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. In this study, the zero-inflated Poisson (ZIP) regression model will be used as a benchmark for comparison.

- The NBI regression model with varying dispersion is derived as follows. Consider policyholder i, $i = 1, \ldots, n$, whose number of claims, denoted as k_i, with $k_i = 0, 1, 2, 3, \ldots$, are independent. In addition, assume that $k_i|, z_i$ follows a Poisson distribution with pmf given by Equation (1), and z_i follows a Gamma distribution with pdf given by

$$g(z_i; \phi_i) = \frac{z_i^{\frac{1}{\phi_i} - 1} \frac{1}{\phi_i}^{\frac{1}{\phi_i}} \exp\left(-\frac{z_i}{\phi_i}\right)}{\Gamma\left(\frac{1}{\phi_i}\right)}, \tag{29}$$

where $\phi_i > 0$. Parameterization (29) ensures that $E(z_i) = 1$, and hence the model is identifiable.

Then, the unconditional distribution of k_i becomes an NBI distribution, with pmf given by

$$P(k_i) = \frac{\Gamma\left(k_i + \frac{1}{\phi_i}\right)}{k_i! \Gamma\left(\frac{1}{\phi_i}\right)} \left(\frac{\phi_i \mu_i}{1 + \phi_i \mu_i}\right)^{k_i} \left(\frac{1}{1 + \phi_i \mu_i}\right)^{\frac{1}{\phi_i}}. \tag{30}$$

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

$$E(k_i) = \mu_i \tag{31}$$

and

$$\text{Var}(k_i) = \mu_i + \mu_i^2 \phi_i. \tag{32}$$

- The mean and dispersion parameters of the NBI distribution are modelled in terms of covariates

$$\mu_i = \exp\left(x_{1,i}^\top \boldsymbol{\beta}_1\right) \text{ and} \tag{33}$$

$$\phi_i = \exp\left(x_{2,i}^\top \boldsymbol{\beta}_2\right), \tag{34}$$

where $x_{1,i}$ and $x_{2,i}$ are covariate vectors with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, with $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, \ldots, \beta_{1,p_1}\right)^\top$ and $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, \ldots, \beta_{2,p_2}\right)^\top$ 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.
- The pmf of the ZIP regression model is given by

$$P(k_i) = \begin{cases} \pi + (1 - \pi)exp(-\mu_i), & \text{if } k_i = 0 \\ (1 - \pi)\frac{e^{-z_i \mu_i}(z_i \mu_i)^{k_i}}{k_i!}, & \text{if } k_i = 1, 2, 3 \ldots \end{cases} \tag{35}$$

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

$$E(k_i) = \mu_i(1 - \pi) \tag{36}$$

and

$$\text{Var}(k_i) = \mu_i(1 - \pi)[1 + \mu_i \pi], \tag{37}$$

where $\mu_i = exp(x_{1,i}^\top \beta_1)$, and where $x_{1,i}$ is a covariate vector with dimension $p_1 \times 1$, with $\beta_1 = (\beta_{1,1}, \ldots, \beta_{1,p_1})^\top$ the corresponding parameter vector, and where it is assumed that the matrix X_1 with rows given by $x_{1,i}$, respectively, are of full rank (note that π can also be modelled in terms of covariates using the logit link function. However, we refrain from doing this in this paper since this approach did not lead to better fitting performances for the ZIP model for the MTPL data).

4.1. Modelling Results

The ML estimates of the parameters (all the parameters were statistically significant at a 5% threshold) for the NBI and PLN regression models with varying dispersion and the ZIP regression model are presented in Table 2. Note that variable selection can be performed for all the models by selecting the best predictor for parameter μ_i using backward elimination. This can be done by including all available explanatory variables present in the data set and testing whether the exclusion of each variable will result in lower global deviance (DEV), Akaike information criterion (AIC), and Schwartz Bayesian criterion (SBC) values. Subsequently, in the case of the NBI and PLN models, we can take all the variables selected for the parameter μ_i and continue variable selection for the parameter ϕ_i by performing forward selection, where we can test which explanatory variable would lead to a further decrease of the DEV, AIC, and SBC values when added to parameter ϕ_i. Additionally, if different subsets of explanatory variables result in very similar values of DEV, AIC, and SBC, we should chose the simpler model with less predictors to avoid overfitting. Regarding our data, as we can see from Table 2, the explanatory variables AD, HP, and AC were chosen for μ_i, and only the variable AD was chosen for ϕ_i.

Table 2. Parameter estimates of the ZIP regression model and the NBI and PLN regression models with varying dispersion.

NBI		ZIP		PLN	
Coeff. β_1		**Coeff. β_1**		**Coeff. β_1**	
Intercept	−0.4729	Intercept	−0.1277	Intercept	−0.4709
AD		CS		CS	
C2	−1.2390	C2	−1.2454	C2	−1.2360
HP		HP		HP	
C2	1.0378	C2	0.9892	C2	1.0469
AC		AC		AC	
C2	−0.6481	C2	−0.6398	C2	−0.6586
Coeff. β_2				**Coeff. β_2**	
Intercept	−2.4935	Prob. π	0.3032	Intercept	−1.0969
AD				CS	
C2	0.8878			C2	0.3481

From the results in Table 2, we observe that the values of the estimated regression coefficients of the variables AD, HP, and AC have a similar effect (positive and/or negative) on parameter μ_i in the case of all the models, and the same observation can be made for parameter ϕ_i in the case of the NBI and PLN models.

Finally, we rely on normalized quantile residuals [13] as an exploratory graphical tool to help us evaluate the adequacy of the fit of the NBI, ZIP, and PLN models. For these discrete response distributions, the normalized (randomized) quantile residuals 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 $\left[F_i(k_i - 1|\theta^{(r+1)}), F_i(k_i|\theta^{(r+1)}) \right]$, where F_i is the cumulative distribution function estimated for the ith policyholder, and where $\theta^{(r+1)}$ contains the estimated model parameters after the EM algorithm has reached the global maximum, and k_i is the corresponding observation. The fit of the claim count model can be evaluated

by means of the 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 1 shows the normalized (random) quantiles for the ZIP regression model and the NBI and PLN claim frequency regression models with varying dispersion.
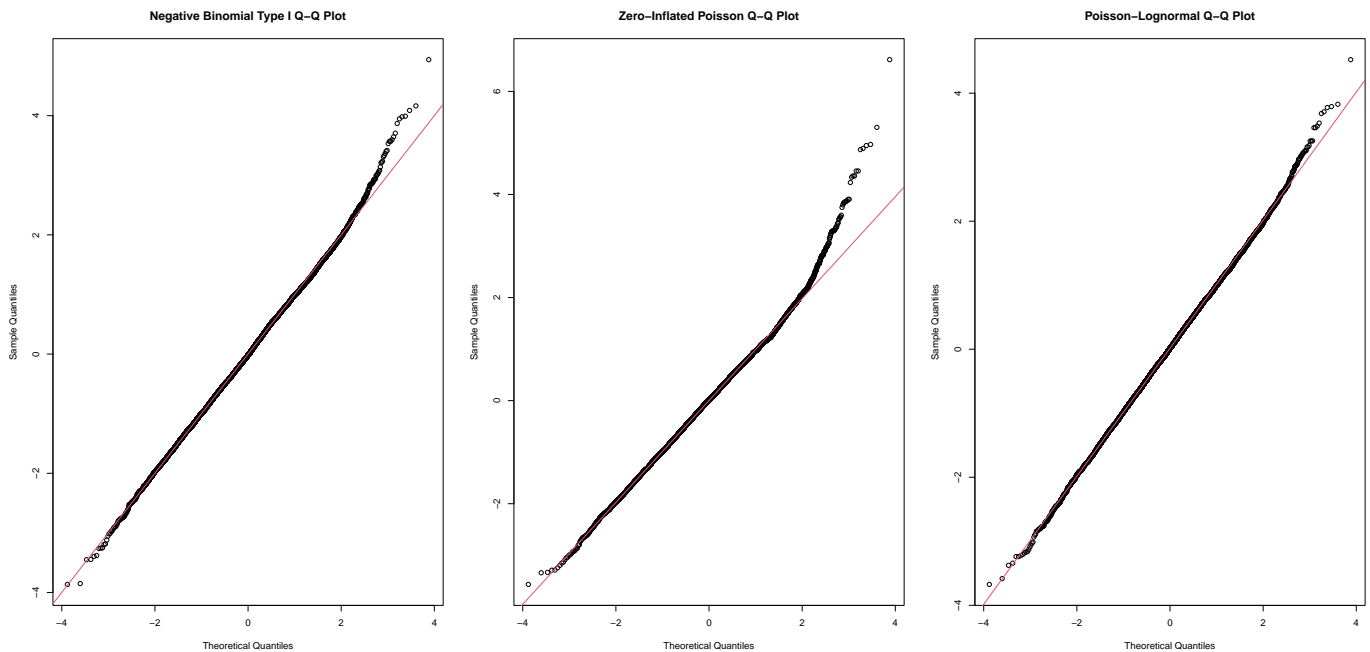

Figure 1. Normalized quantiles for the ZIP regression model and the NBI and PLN regression models with varying dispersion.

From Figure 1, we observe that the residuals indicate that the NBI and PLN are better assumptions than the ZIP model since the residuals of the former two are close to the right tail of the claim frequency distribution. Furthermore, the PLN model seems to fit the claim count data slightly better than the NBI model, since, as was previously mentioned, the tail of mixed Poisson models is equivalent to the tail of their mixing distributions [7], and in this case the lognormal mixing density has a thicker right tail than the Gamma mixing density. Therefore, overall it is reasonable to suggest the employment of the PLN model for modelling claim counts in our data set. As we are going to observe in what follows, the PLN model also provides better fitting performances than the NBI and PIG models in terms of the DEV, AIC, and SBC values.

4.2. Models Comparison

In this subsection, we compare the fit of the ZIP regression model and the NBI and PLN regression models with varying dispersion based on DEV, AIC, and SBC, which are classic hypothesis/specification criteria.

The DEV is defined as

$$DEV = -2\hat{l}(\hat{\boldsymbol{\theta}}),\tag{38}$$

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

$$AIC = DEV + 2 \times df\tag{39}$$

and

$$SBC = DEV + \log(n) \times df\,,\tag{40}$$

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

The resulting DEV, AIC, and SBC values for the competing models are presented in Table 3. We observe that the PLN regression model provides the best fit with respect to all three criteria.

Table 3. ZIP regression model and NBI and PLN regression models comparison.

	Specification Criteria Values		
	DEV	AIC	SBC
NBI	15,885.1	15,897.1	15,940.1
ZIP	16,052.2	16,062.2	16,098
PLN	15,859.4	15,871.4	15,914.4

4.3. Computational Aspects

All computing was made using the programming language R. The PLN regression model with varying dispersion was estimated using the EM algorithm, which was presented in Section 3. In addition, the ML estimates of the parameters of the NBI regression model with varying dispersion and the ZIP regression model were obtained using the generalized additive models for the location, scale, and shape (GAMLSS) package in R [14].

Note that a rather strict criterion was used, and it took the algorithm quite a large number of iterations to converge. In particular, the stopping criterion was set as $tol = 10^{-12}$. Note also that the M-step involves two Newton–Raphson iterations, and hence it is important to identify the choice of meaningful initial values for the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, as this can increase increase the computational time requirements for the EM algorithm and make it more difficult to locate the global maximum. We obtained good initial values for $\boldsymbol{\beta}_1$ by fitting the simple Poisson regression. Additionally, we obtained good initial values for $\boldsymbol{\beta}_2$ by (i) calculating $Var(k_i)$ for the eight different risk classes which can be formed using all available risk factors and the observations $i = 1, \ldots, n$ and (ii) calculating $E(k_i)$ for the eight different risk classes and using the log-link function for ϕ_i (see Equation (6)), so we solve Equation (4) with respect to $\boldsymbol{\beta}_2$. However, we also checked with many other starting values for $\boldsymbol{\beta}_2$ in order to ensure that the global maximum had been obtained. For all cases, the EM algorithm converged to a similar solution. The standard errors were computed by using the standard approach of [15].

Finally, as was anticipated, in terms of CPU time, it took the NBI regression model with varying dispersion and the ZIP regression model less than one minute, and they both compared significantly more favorably to the PLN regression model with varying dispersion, which exceeded 30 min of CPU time. However, it should be taken into account that the PLN model has a density which does not exist in closed form, and that there were 14,143 observations in the sample of MTPL data that was examined in this article. For larger data sets with more features, the computing effort can be reduced if the E- and M-steps are executed in parallel across multiple threads to exploit the processing power of modern-day multicore machines.

5. Conclusions, Limitations, and Future Research

In this article, we considered a family of mixed Poisson claim count regression models with varying dispersion and dependence parameters arising from non-conjugate mixing distributions for approximating overdispersed claim frequencies in non-life insurance. The flexibility in the choice of the mixing distribution combined with the proposed approach, which assumes that the mean and dispersion parameters of the model can be modelled in terms of risk factors, can provide an advantage relative to the majority of previous approaches in the literature, which have concentrated on mixing densities conjugate to the Poisson mixing and assumed that only the mean parameter can vary through covariates. From a practical business standpoint, the proposed modelling framework is beneficial for the insurance company, as it will result in an improved risk evaluation of policyholders who are more likely to have accidents, since the tail behaviour of mixed Poisson models is

similar to that of the mixing density and the majority of heavy-tailed mixing distributions are not conjugate to the Poisson.

The PLN regression model with regression specifications on its mean and dispersion parameters was considered for expository purposes. Furthermore, we developed an efficient EM algorithm for maximum likelihood estimation of the parameters of the model. The implementation of the algorithm was illustrated by fitting the model to a real MTPL insurance data set. An interesting line for further research would be to extend the model to the multivariate case to permit inferences about the dependence structure between different types of overdispersed claim counts from the same and/or different types of coverage. However, it should be noted that the PLN model becomes more complicated in the two-dimensional setting due to algebraic intractability, which is a problem that is inherited from the univariate case. Moreover, modelling all the parameters of the PLN model in terms of covariates can further increase the computational burden in the high-dimensional setting. Finally, another fruitful future research direction is to include time series components to take into account both cross-dependence between different types of claims and time dependence.

Abbreviations

The following abbreviations are used in this manuscript:

EM	Expectation-maximization
NBI	Negative binomial type I
MCEM	Monte Carlo expectation-maximization
ML	Maximum likelihood
MTPL	Motor third-party liability
pdf	Probability density function
PLN	Poisson log-normal
pmf	Probability mass function
ZIP	Zero-inflated Poisson

References

1. Lawless, J.F. Negative binomial and mixed Poisson regression. *Can. J. Stat.* **1987**, *15*, 209–225. [CrossRef]
2. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*, 1st ed.; Cambridge University Press: Cambridge, UK, 1998.
3. Hilbe, J.M. *Negative Binomial Regression*, 1st ed.; Cambridge University Press: Cambridge, UK, 2008.
4. Ord, J.K.; Whitmore, G.A. The Poisson-inverse Gaussian distribution as a model for species abundance. *Commun. Stat. Theory Methods* **1986**, *15*, 853–871. [CrossRef]
5. Willmot, G.E. The Poisson-Inverse Gaussian distribution as an alternative to the negative binomial. *Scand. Actuar. J.* **1987**, *3–4*, 113–127. [CrossRef]
6. Dean, C.; Lawless, J.F.; Willmot, G.E. A mixed Poisson–inverse-Gaussian regression model. *Can. J. Stat.* **1989**, *17*, 171–181. [CrossRef]
7. Perline, R. Mixed Poisson distributions tail equivalent to their mixing distributions. *Stat. Comput.* **1998**, *38*, 229–233. [CrossRef]
8. Rigby, R.A.; Stasinopoulos, D.M.; Akantziliotou, C. A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Comput. Stat. Data Anal.* **2008**, *53*, 381–393. [CrossRef]
9. Barreto-Souza, W.; Simas, A.B. General mixed Poisson regression models with varying dispersion. *Stat. Comput.* **2016**, *26*, 1263–1280. [CrossRef]
10. Tzougas, G. EM estimation for the Poisson–inverse Gamma regression model with varying dispersion: An application to insurance ratemaking. *Risks* **2020**, *8*, 97. [CrossRef]
11. Blueschke, D.; Blueschke-Nikolaeva, V.; Neck, R. Approximately Optimal Control of Nonlinear Dynamic Stochastic Problems with Learning: The OPTCON Algorithm. *Algorithms* **2021**, *14*, 181. [CrossRef]
12. Amirghasemi, M. An Effective Decomposition-Based Stochastic Algorithm for Solving the Permutation Flow-Shop Scheduling Problem. *Algorithms* **2021**, *14*, 112. [CrossRef]
13. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *Comput. Graph. Stat.* **1996**, *5*, 236–245.
14. Stasinopoulos, D.M.; Rigby, B.; Akantziliotou, C. *Instructions on How to Use the Gamlss Package in R*, 2nd ed. 2008. Available online: http://www.gamlss.org (accessed on 30 December 2021).
15. Louis, T.A. Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc.* **1982**, *44*, 226–233.