AN EM ALGORITHM FOR FITTING A NEW CLASS OF MIXED EXPONENTIAL REGRESSION MODELS WITH VARYING DISPERSION

BY

GEORGE TZOUGAS AND DIMITRIS KARLIS

ABSTRACT

Regression modelling involving heavy-tailed response distributions, which have heavier tails than the exponential distribution, has become increasingly popular in many insurance settings including non-life insurance. Mixed Exponential models can be considered as a natural choice for the distribution of heavy-tailed claim sizes since their tails are not exponentially bounded. This paper is concerned with introducing a general family of mixed Exponential regression models with varying dispersion which can efficiently capture the tail behaviour of losses. Our main achievement is that we present an Expectation-Maximization (EM)-type algorithm which can facilitate maximum likelihood (ML) estimation for our class of mixed Exponential models which allows for regression specifications for both the mean and dispersion parameters. Finally, a real data application based on motor insurance data is given to illustrate the versatility of the proposed EM-type algorithm.

KEYWORDS

Mixed exponential distributions, EM algorithm, regression models for the mean and dispersion, parameters, non-life insurance, heavy-tailed losses.

1. INTRODUCTION

Over the last few decades, it has become increasingly clear that insurance losses are often characterized by asymmetry and thick tails. In particular, the claim size distribution is often non-negative, right skewed and leptokurtic, that is, it has a heavy tail. As far as the modelling of heavy-tailed insurance loss data in non-life insurance, which is the main focus of this study, is concerned, a few large size claims hitting a portfolio usually represent the greatest part of the indemnities paid by the company. Therefore, quantifying the risk

posed by those extreme losses has often been an imperative task for actuaries. The continuous probability distributions which have been traditionally employed in the actuarial literature for heavy tail analyses include, for instance, the Generalized Inverse Gaussian (GIG) distribution and many members of the Generalized Beta Type II (GB2) family of distributions. The GB2 family includes, for instance, the Burr distribution, the Pareto distribution, its generalization, namely the Generalized Pareto distribution (GP), which has been used in the context of Extreme Value Theory, see, for example, Embrechts *et al.* (1997), and also limiting distributions, such as the Generalized Gamma (GG) distribution. Furthermore, modelling the aforementioned heavy-tailed distributions in regression analysis has recently become a popular statistical tool for analyzing insurance data based on important explanatory variables for claim severities. In particular, the list of possible parametric candidates for modelling heavy-tailed data for the case with and the case without covariate information is not exhaustive. References include, for example, Beirlant *et al.* (1992), Kleiber and Kotz (2003, Section 4.11), Wills *et al.* (2006), Rosenberg *et al.* (2007), Frees and Valdez (2008), Frees *et al.* (2014a, Chapter 10; 2014b, Chapter 9). Additionally, Ahn *et al.* (2012) considered log phase-type distributions, Ramirez-Cobo *et al.* (2010), and Hürlimann (2014) Calderín-Ojeda *et al.* (2017) employed the Double-Pareto-Lognormal (DPLN) distribution for approximating large claim amounts for the case with and the case without covariate information respectively and Bladt and Rojas-Nandayapa (2018) used scale mixtures of phase-type distributions for modelling heavy-tailed data.

The aim of the present work is to introduce a general family of mixed Exponential regression models with varying dispersion for approximating heavy-tailed losses in non-life insurance. This class is based on a mixing between the Exponential distribution and a unit mean continuous prior, or mixing, distribution which belongs to a general distribution family. The proposed class of mixed Exponential distributions is defined so that its mean is an explicit parameter of the distribution. This allows easier interpretation when both the mean and dispersion parameters are modelled using explanatory variables and provides a more orthogonal parameterization to ease model fitting. Moreover, there is a wide range of practical applications in many non-life insurance settings for this type of models. In particular, the proposed modelling framework results in an improved risk evaluation as it can enable the actuary to fit more representative distributions for the data that can capture all their important stylized characteristics and thus more efficiently determine the appropriate level of premiums, reserves and reinsurance. Three members of this family are examined in depth, the Pareto, or Exponential-Inverse Gamma, Exponential-Inverse Gaussian (EIG) and Exponential-Lognormal (ELN) distributions with regression structures on both their mean and dispersion parameters.

In what follows, we discuss in detail our contributions and comment on how our approach can extend the current actuarial applications concerning positively skewed non-life insurance data with a larger upper tail.

Firstly, by introducing this general class of mixed Exponential distributions, we unified two mixed Exponential distributions which have been studied separately in the literature, namely the Pareto and the EIG distributions. The Pareto distribution has been excessively overused in the actuarial literature, see for instance, among many others, Klugman *et al.* (2012) for the case without covariate information, Frangos and Vrontos (2001) and Tzougas *et al.* (2014, 2018) who considered the case with covariate information. However, more limited literature exists for the EIG distribution. In fact, only Bhattacharya and Kumar (1986) and Hesselager *et al.* (1998) considered the EIG distribution and Frangos and Karlis (2004) added a regression component to the model. The advantage of the Pareto and EIG models is the conjugacy, in a Bayesian sense, between the Inverse Gamma and Inverse Gaussian (IG) prior distributions respectively and the Exponential distribution, which facilitates maximum likelihood (ML) estimation. However, apart from the two aforementioned cases, this family of models has not been studied in depth since. The main reason is because most of the members of the mixed Exponential family of models do not have their densities in closed form, and hence special numerical methods are needed for their ML estimation. Such an example is the ELN model with a regression component which was recently proposed by Tzougas *et al.* (2019) for deriving a priori and a posteriori ratemaking mechanisms for the costs of claims. In particular, compared to the two classical mixed Exponential models, there is no analytical form for the distribution of the cost of claims if the random effect term, which follows the Lognormal (LN) distribution, is marginalized out. The main contribution of this study is that it illustrates that ML estimation of mixed Exponential regression models with varying dispersion can be accomplished relatively easily via a novel Expectation-Maximization (EM) type algorithm which is computationally parsimonious and can remedy the computational issues which may occur by alternative estimation procedures, since ML estimation of mixed Exponential distributions is far from straightforward if regression specifications are allowed for both mean and the dispersion parameters and the computational complexity increases even further if the choice of the mixing density leads to algebraically intractable members of the mixed Exponential family. At this point, we would like to call attention to the fact that the development of ML estimation algorithms for modelling jointly all the parameters of discrete and continuous response distributions in terms of covariates remains a largely uncharted research field. Notable exceptions are the articles by Rigby and Stasinopoulos (2005) and Barreto-Souza and Simas (2016). Rigby and Stasinopoulos (2005) introduced the generalized additive models for location, scale and shape (GAMLSS) which allow modelling of every parameter of the distribution of the response, not only as parametric but also as additive nonparametric functions of covariates and/or random-effects terms. The general class of univariate regression models which they developed extends the setup of many well-known distributions including the Pareto but not other members of the mixed Exponential family. The model fitting of a GAMLSS model can

be achieved by either the RS algorithmic procedure which is based on the algorithm that was used for the fitting of the mean and dispersion additive models of Rigby and Stasinopoulos (1996a, 1996b) or the CG algorithm which is based on the algorithm by Cole and Green (1992). Barreto-Souza and Simas (2016) also proposed the use of the EM algorithm for estimating the parameters of the general class of mixed Poisson regression models with varying dispersion which they developed, extending the setup of Karlis (2001) who considered the case when only the mean is modelled in terms of explanatory variables. In their numerical illustration they gave special emphasis on the estimation of the Negative Binomial and Poisson-IG regression models with regression structures on both their mean and dispersion parameters. However, this is the first time that the EM algorithm is used, in a statistical or actuarial setting, for estimating mixed Exponential regression models with varying dispersion.

Secondly, as is well known, in the majority of insurance applications regarding two parameter continuous distributions employed for modelling positively skewed non-life insurance data with a larger upper tail, the commonly used specification is that only the mean claim severity is modelled as a function of risk factors. In this respect, the skewness of the response variable, which in general depends on both the mean and the dispersion parameters, is not modelled explicitly in terms of covariates but implicitly through their dependence on the mean parameter. Consequently, in practical situations where the assumption of constant dispersion is virtually never valid, modelling only the mean parameter in terms of covariates can play a central role in the misclassification of heavy-tailed risks. As a solution to the previously described problems we allow for regressors on both the mean and the dispersion parameters of mixed Exponential models. Moreover, with our general approach we can utilize all the available information for adequately modelling the tail of the actual claim severity distribution since we are able to model risk heterogeneity as the distribution of claim severity changes between various risk classes of policyholders by a function of the level of all risk factors underlying the analysed classes.

Finally, real non-life insurance data sets usually include a mix of moderate and large claim sizes. However, as empirical evidence has shown, moderate observations are those which constitute the largest proportion of the sample while large observations have very low frequencies. For instance, large claims are usually not frequent to appear, while a bigger number of relatively small claims is common in motor insurance, see Frees and Valdez (2008). In this respect, excessive losses can be caused either by a small number of claims with large amounts or by a large number of moderate claim costs. Nevertheless, while heavy-tailed models can be used to efficiently capture the characteristics of larger loss data with lower frequencies, moderate losses with higher frequencies are more effectively approximated by classical models such as, for instance, the Gamma, IG and LN models. For more details see, for example, Klugman *et al.* (2012) who provided several classification categories for the heaviness

of claim size models. On the other hand, light-tailed distributions cannot adequately model the risk posed by heavy-tailed phenomena since they have an inherent tendency to underestimate the magnitude of large claim amounts. Due to the reasons mentioned above, insurers tend to partition losses in their portfolios because in view of the unique features of the body and the tail of the actual claim size distribution, no standard parametric model seems to be able to provide an accurate fit for both moderate and large size claims. Regarding the three mixed Exponential distributions presented herein, the EIG and ELN distributions have just the appropriate level of generality for modelling non-life insurance data since they have a more sufficiently flexible kind of geometry than the Pareto distribution. In particular, the EIG and ELN distributions retain the heavy-tailed property, since the upper tail of the former is larger than that of the IG distribution and the latter shows a heavier tail relative to the LN density. Furthermore, the advantage that the EIG and ELN models enjoy over the Pareto is that they both have a more promising shape for moderate claim costs. In order to formally compare the tails of the Pareto, EIG and ELN models, we follow the setup of Wang (1998) who proposed the use of the right tail index for classifying claim severity distributions by their right tail behaviour.

The layout of the rest of the paper is as follows: Section 2 presents a detailed description of the proposed family of mixed Exponential regression models with varying dispersion. Also, we consider the Inverse Gamma, IG and LN mixing distributions, and we derive the probability density function (pdf) of the corresponding mixed Exponential regression models with varying dispersion. Section 3 fully describes the ML estimation through the EM algorithm. Additionally, detailed EM algorithms for the Pareto, EIG and ELN regression models with varying dispersion are provided. Section 4, contains an application to a data set concerning Motor Third Party Liability (MTPL) insurance claims at fault. Furthermore, a Monte Carlo simulation study is conducted to investigate the performance of our EM algorithm for estimating the parameters of the mixed Exponential model. In Section 5, we discuss some computational issues related to the implementation of the EM algorithm for the Pareto, EIG and ELN regression models with varying dispersion. Finally, an extension of the proposed class of models is discussed in Section 6 and concluding remarks can be found in Section 7.

2. THE MIXED EXPONENTIAL REGRESSION MODEL WITH VARYING DISPERSION

The general class of mixed Exponential regression models with varying dispersion, which is considered in this study, can be described as follows. Assume that Z_i are independent and identically distributed (i.i.d.) random variables and that the individual claim costs, $Y_i|Z_i$, arising from a policyholder $i, i = 1, \ldots, n,$

are i.i.d. random variables according to an Exponential distribution with pdf given by

$$f(y_i|z_i) = \frac{e^{-\frac{y_i}{\mu_i z_i}}}{\mu_i z_i}, \tag{2.1}$$

where $y_i, \mu_i, z_i > 0$ with $E(Y_i|z_i) = \mu_i z_i$ and $Var(Y_i|z_i) = (\mu_i z_i)^2$.

Let us now assume that Z_i are random variables from a continuous and at least twice differentiable mixing distribution with cumulative distribution function (cdf) $G(z_i; \phi_i)$, where in order for the model to be identifiable we assume that $E(Z_i) = 1$ and where $\phi_i > 0$ is the dispersion parameter. Considering the previous assumptions, it is easy to see that the unconditional distribution of Y_i is a mixed Exponential distribution with pdf given by

$$f(y_i) = \int_0^\infty f(y_i|z_i) g(z_i; \phi_i) \, dz_i, \tag{2.2}$$

where $g(z_i; \phi_i)$ is the pdf of Z. To allow for the mean and dispersion parameters to be modelled as functions of explanatory variables with parametric linear functional forms we assume that

$$\mu_i = \exp\left(\mathbf{x}_{1,i}^T \boldsymbol{\beta}_1\right) \text{ and} \tag{2.3}$$

$$\phi_i = \exp\left(\mathbf{x}_{2,i}^T \boldsymbol{\beta}_2\right), \tag{2.4}$$

where $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are covariate vectors[1] with dimensions $p_1 \times 1$ and $p_2 \times 1$ respectively, with $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, \ldots, \beta_{1,p_1}\right)^T$ and $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, \ldots, \beta_{2,p_2}\right)^T$ the corresponding parameter vectors and where it is assumed that the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, with rows given by $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ respectively, are of full rank.

Also, using the laws of total expectation and total variance and the moments of the Exponential distribution one can easily find that the mean and the variance of Y_i are given by

$$E(Y_i) = E_{Z_i}[E(Y_i|Z_i = z_i)] = \mu_i E_{Z_i}[Z_i] = \mu_i \tag{2.5}$$

and

$$\begin{aligned} Var(Y_i) &= E_{Z_i}[Var(Y_i|Z_i = z_i)] + Var_{Z_i}[E(Y_i|Z_i = z_i)] \\ &= \mu_i^2 \left[E_{Z_i}(Z_i^2) + Var_{Z_i}(Z_i)\right]. \end{aligned} \tag{2.6}$$

In what follows, different mixed Exponential distributions with regression structures on their mean and dispersion parameters are used to describe the

behaviour of the costs of claims as a function of the explanatory variables including the Pareto, EIG and ELN distributions.

2.1. Pareto regression model with varying dispersion

As previously mentioned, the Pareto distribution is the most well-known mixed Exponential distribution that has been traditionally used for modelling large size claims. Thus, a natural choice for the mixing distribution is to assume that the random effect term z_i is distributed according to an Inverse Gamma mixing distribution with pdf

$$g(z_i; \phi_i) = \frac{\frac{1}{(\phi_i - 1)} \exp\left(-\frac{(\phi_i - 1)}{z_i}\right)}{\left(\frac{z_i}{\phi_i - 1}\right)^{\phi_i + 1} \Gamma(\phi_i)}, \tag{2.7}$$

where $z_i > 0$ and $\phi_i > 2$, with mean $E(Z_i) = 1$ and variance $Var(Z_i) = \frac{1}{\phi_i - 2}$, for $i = 1, \ldots, n$.

Then, based on Equations (2.1) and (2.7) it is easy to see that the unconditional distribution is the Pareto distribution with pdf

$$f(y_i) = \frac{\phi_i \left[(\phi_i - 1)\mu_i\right]^{\phi_i}}{\left[y_i + (\phi_i - 1)\mu_i\right]^{\phi_i + 1}}, \tag{2.8}$$

where $\mu_i > 0$ and $\phi_i > 2$, with mean and variance given by

$$E(Y_i) = \mu_i \text{ and} \tag{2.9}$$

$$Var(Y_i) = \frac{\left[(\phi_i - 1)\mu_i\right]^2}{\phi_i - 1}\left(\frac{2}{\phi_i - 2} - \frac{1}{\phi_i - 1}\right). \tag{2.10}$$

2.2. EIG regression model with varying dispersion

The EIG is based on the IG mixing distribution with pdf given by

$$g(z_i; \phi_i) = \frac{\phi_i}{\sqrt{2\pi}} \exp\left(\phi_i^2\right) z_i^{-\frac{3}{2}} \exp\left[-\frac{1}{2}\left(\frac{\phi_i^2}{z_i} + \phi_i^2 z_i\right)\right], \tag{2.11}$$

where $z_i > 0$ and $\phi_i > 0$, with mean $E(Z_i) = 1$ and variance $Var(Z_i) = \frac{1}{\phi_i^2}$, for $i = 1, \ldots, n$. Note that, the overdispersion relative to the simple Exponential distribution is $\frac{1}{\phi_i^2}$. Thus, if ϕ_i tends to infinity, this distribution can be reduced to the Exponential distribution. More details about the IG distribution can be found in Jørgensen (1982). Note also that there are several different parameterizations of the IG distribution (see, Seshadri, 2012).

Considering the assumptions of the model, that is, Equations (2.1) and (2.11), it can be verified that the unconditional distribution of y_i is an EIG distribution with pdf

$$f(y_i) = \frac{\phi_i \exp\left[-\phi_i\left(\sqrt{\phi_i^2 + \frac{2y_i}{\mu_i}} - \phi_i\right)\right]\left(\phi_i\sqrt{\phi_i^2 + \frac{2y_i}{\mu_i}} + 1\right)}{\mu_i\left(\phi_i^2 + \frac{2y_i}{\mu_i}\right)^{\frac{3}{2}}}, \tag{2.12}$$

where $\mu_i > 0$ and $\phi_i > 0$, with mean and variance given by

$$E(Y_i) = \mu_i \text{ and} \tag{2.13}$$

$$Var(Y_i) = \mu_i^2\left(\frac{2}{\phi_i^2} + 1\right). \tag{2.14}$$

2.3. ELN regression model with varying dispersion

Let us now consider another plausible mixing distribution for constructing a heavy-tailed mixed Exponential model, namely the LN distribution with pdf

$$g(z_i; \phi_i) = \frac{1}{\sqrt{2\pi}\phi_i z_i} \exp\left[-\frac{\left(\log(z_i) + \frac{\phi_i^2}{2}\right)^2}{2\phi_i^2}\right], \tag{2.15}$$

where $z_i > 0$ and $\phi_i > 0$, with mean $E(Z_i) = 1$ and variance $Var(Z_i) = \exp\left(\phi_i^2\right) - 1$, for $i = 1, \ldots, n$.

Then, based on Equations (2.1) and (2.15) it is easy to see that the resulting distribution of y_i is the ELN distribution with pdf

$$f(y_i) = \int_0^\infty \frac{e^{-\frac{y_i}{\mu_i z_i}}}{\mu_i z_i} \frac{\exp\left[-\frac{\left(\log(z_i) + \frac{\phi_i^2}{2}\right)^2}{2\phi_i^2}\right]}{\sqrt{2\pi}\phi_i z_i} dz_i, \tag{2.16}$$

where $\mu_i > 0$ and $\phi_i > 0$. Unfortunately, the last integral cannot be simplified but it can be computed via numerical integration. Also, using Equations (2.5) and (2.6) we have that

$$E(Y_i) = \mu_i \text{ and} \tag{2.17}$$

$$Var(Y_i) = \mu_i^2\left[2\exp\left(\phi_i^2\right) - 1\right]. \tag{2.18}$$

3. THE EM ALGORITHM FOR ML ESTIMATION OF MIXED EXPONENTIAL MODELS WITH VARYING DISPERSION

Let $(y_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$, $i = 1, \ldots, n$, be a sample of independent observations, where y_i is the response variable and $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are the vectors of covariate information with dimensions $p_1 \times 1$ and $p_2 \times 1$ respectively. Also, consider that the data are produced according to the mixed Exponential model. Then, the log-likelihood of the model can be written as

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(f(y_i)), \qquad (3.1)$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T\right)^T$ is the vector of the parameters and where $f(y_i)$ is the pdf of the mixed Exponential model, which is given by Equation (2.2).

Direct maximization of the above function with respect to the vector of parameters $\boldsymbol{\theta}$ is not easy due to the complexity of the log-likelihood of the mixed Exponential model, which is not usually tractable. Moreover, when regression structures are allowed for both the mean and dispersion parameters this raises additional computational challenges.

Fortunately, ML estimation can be accomplished relatively easily via an EM-type algorithm[2] (see Dempster *et al.*, 1977 and McLachlan and Krishnan, 2007). In particular, if one augments the unobserved data z_i to the observed data $(y_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$, for $i = 1, \ldots, n$, then the complete data log-likelihood factorizes into two parts

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[-\frac{y_i}{\mu_i z_i} - \log(\mu_i) - \log(z_i) \right] + \sum_{i=1}^{n} \log(g(z_i; \phi_i)), \qquad (3.2)$$

where $g(z_i; \phi_i)$ is the pdf of the mixing distribution and where μ_i and ϕ_i are given by Equations (2.3) and (2.4) respectively. The EM-type algorithm for the mixed Exponential 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 in Equation (3.2), is given by

$$Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(r)}\right) \equiv E_{z_i}\left(l_c(\boldsymbol{\theta}) \,|\, \mathbf{Y}, \boldsymbol{\theta}^{(r)}\right) \propto$$

$$\sum_{i=1}^{n} \left[-\frac{y_i E_{z_i}\left[\frac{1}{z_i} | y_i; \boldsymbol{\theta}^{(r)}\right]}{\mu_i^{(r)}} - \log\left(\mu_i^{(r)}\right) \right]$$

$$+ \sum_{i=1}^{n} E_{z_i}\left[\log\left(g\left(z_i; \phi_i^{(r)}\right)\right)\right], \qquad (3.3)$$

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 $w_i = E_{Z_i}\left[\frac{1}{Z_i}|y_i;\boldsymbol{\theta}^{(r)}\right]$ and $\omega_{k,i} = E_{Z_i}[s_k(Z_i)|y_i;\boldsymbol{\theta}^{(r)}]$, for $i = 1,..,n$ and $k = 1,\ldots,v$, where $s_k(.)$ are certain functions[3] 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_i and $\omega_{k,i}$ from the E-Step and the Newton–Raphson algorithm twice,[4] 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:

$$h_1(\boldsymbol{\beta}_1) = \frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{1,j}} = \sum_{i=1}^{n}\left(\frac{y_i}{\mu_i^{(r)}}w_i - 1\right)x_{1,ij}, \qquad (3.4)$$

and

$$H_1(\boldsymbol{\beta}_1) = \frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{1,j}\partial \beta_{1,j}^T} = \sum_{i=1}^{n}\left(-\frac{y_i}{\mu_i^{(r)}}w_i\right)x_{1,ij}x_{1,ij}^T = \mathbf{X}_1^T\mathbf{W}_1\mathbf{X}_1, \qquad (3.5)$$

for $i = 1,\ldots,n$ and $j = 1,\ldots,p_1$ and where $\mathbf{W}_1 = diag\{-\frac{y_i}{\mu_i^{(r)}}w_i\}$.

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

$$\boldsymbol{\beta}_1^{(r+1)} \equiv \boldsymbol{\beta}_1^{(r)} - \left[H_1\left(\boldsymbol{\beta}_1^{(r)}\right)\right]^{-1}h_1\left(\boldsymbol{\beta}_1^{(r)}\right). \qquad (3.6)$$

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

$$h_2(\boldsymbol{\beta}_2) = \frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{2,j}} = \frac{\partial E_{z_i}\left[\log\left(g\left(z_i;\phi_i^{(r)}\right)\right)\right]}{\partial \beta_{2,j}} \qquad (3.7)$$

and

$$H_2(\boldsymbol{\beta}_2) = \frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{2,j}\partial \beta_{2,j}^T} = \frac{\partial E_{z_i}\left[\log\left(g\left(z_i;\phi_i^{(r)}\right)\right)\right]}{\partial \beta_{2,j}\partial \beta_{2,j}^T}, \qquad (3.8)$$

where for calculating $h_1(\boldsymbol{\beta}_2)$ and $H_2(\boldsymbol{\beta}_2)$ one needs to use the pseudo-values $\omega_{k,i}$ for $i = 1,\ldots,n$ and $k = 1,\ldots,v$, since 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:

$$\boldsymbol{\beta}_2^{(r+1)} \equiv \boldsymbol{\beta}_2^{(r)} - \left[H_2\left(\boldsymbol{\beta}_2^{(r)}\right)\right]^{-1}h_2\left(\boldsymbol{\beta}_2^{(r)}\right), \qquad (3.9)$$

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

- Finally, iterate between the E-step and the M-step until some convergence criterion is satisfied, for example, the relative change in log-likelihood between two successive iterations is smaller than 10^{-12}.

In what follows, we describe in detail the E-Step and the M-Step of our EM-type algorithm for the Pareto, EIG and ELN regression models with varying dispersion.

3.1. Pareto regression model with varying dispersion

In the case of the Inverse Gamma mixing distribution with pdf given by Equation (2.7), we have that the posterior distribution of $z_i|y_i; \boldsymbol{\theta}$ is an Inverse Gamma with parameters $\phi_i + 1$ and $\phi_i - 1 + \frac{y_i}{\mu_i}$, for $i = 1, \dots, n$. Then, the EM algorithm goes as follows:

- **E-step:**
  Calculate for all $i = 1, \dots, n$,

$$w_i = E_{Z_i}\left[\frac{1}{Z_i}|y_i; \boldsymbol{\theta}^{(r)}\right] = \frac{\phi_i^{(r)} + 1}{\phi_i^{(r)} - 1 + \frac{y_i}{\mu_i^{(r)}}} \tag{3.10}$$

  and

$$\omega_i = E_{Z_i}\left[\log(Z_i)|y_i; \boldsymbol{\theta}^{(r)}\right] = \log\left(\phi_i^{(r)} - 1 + \frac{y_i}{\mu_i^{(r)}}\right) - \Psi\left(\phi_i^{(r)} + 1\right), \tag{3.11}$$

  where $\Psi(.)$ is the digamma function and where $\mu_i^{(r)} = \exp(\mathbf{x}_{1,i}^T \boldsymbol{\beta}_1^{(r)})$ and $\phi_i^{(r)} = \exp(\mathbf{x}_{2,i}^T \boldsymbol{\beta}_2^{(r)})$ are the estimates obtained after r-th iteration.
- **M-step:**
  - Update the regression parameters $\boldsymbol{\beta}_1$ using the pseudo-values w_i, which are given by Equation (3.10), and the Newton–Raphson algorithm in Equations (3.4)–(3.6).
  - Update the regression parameters $\boldsymbol{\beta}_2$ using the pseudo-values w_i and ω_i, which are given by Equations (3.10) and (3.11) respectively, and the Newton–Raphson algorithm which, in the case of the Inverse Gamma mixing distribution, is as follows

$$h_2(\boldsymbol{\beta}_2) = \phi_i^{(r)}\left[\frac{\phi_i^{(r)}}{\phi_i^{(r)} - 1} + \log\left(\phi_i^{(r)} - 1\right) - w_i - \omega_i - \Psi\left(\phi_i^{(r)}\right)\right] x_{2,ij},$$
$$\tag{3.12}$$

$$H_2(\boldsymbol{\beta}_2) = \sum_{i=1}^{n} \phi_i^{(r)}\left[\frac{3\phi_i^{(r)}}{\phi_i^{(r)} - 1} + \log\left(\phi_i^{(r)} - 1\right) - w_i - \omega_i\right.$$

$$-\Psi\left(\phi_i^{(r)}\right) - \left(\frac{\phi_i^{(r)}}{\phi_i^{(r)}-1}\right)^2 - \phi_i^{(r)}\Psi_3\left(\phi_i^{(r)}\right)\Bigg] x_{2,ij}x_{2,ij}^T$$

$$= \mathbf{X}_2^T \mathbf{W}_2 \mathbf{X}_2, \tag{3.13}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p_2$, where $\Psi_3(.)$ is the trigamma function and where $\mathbf{W}_2 = diag\{\frac{2(\phi_i^2)^{(r)}}{\phi_i^{(r)}-1} + 1 + \phi_i^{(r)} \log(\phi_i^{(r)} - 1) - \phi_i^{(r)}w_i - \phi_i^{(r)}\omega_i - \phi_i^{(r)}\Psi(\phi_i^{(r)}) - (\phi_i^2)^{(r)}\Psi_3(\phi_i^{(r)})\}$.

Then, we can obtain the updated estimates of $\boldsymbol{\beta}_2^{(r)}$ using Equation (3.9).

3.2. EIG regression model with varying dispersion

In the case of the IG mixing distribution with pdf given by Equation (2.11), we have that the posterior distribution of $z_i|y_i;\boldsymbol{\theta}$ is a $\mathrm{GIG}\left(-3/2, \phi_i, \sqrt{\phi_i^2 + \frac{2y_i}{\mu_i}}\right)$ for $i = 1, \ldots, n$. Then, the EM algorithm is as follows:

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

$$w_i = E_{Z_i}\left[\frac{1}{Z_i}|y_i; \boldsymbol{\theta}^{(r)}\right] = \frac{3}{(\phi_i^2)^{(r)} + \frac{2y_i}{\mu_i^{(r)}}} + \frac{(\phi_i^2)^{(r)}}{\phi_i^{(r)}\sqrt{(\phi_i^2)^{(r)} + \frac{2y_i}{\mu_i^{(r)}}} + 1} \tag{3.14}$$

and

$$\omega_i = E_{z_i}\left[Z_i|y_i; \boldsymbol{\theta}^{(r)}\right] = \frac{(\phi_i^2)^{(r)} + \frac{2y_i}{\mu_i^{(r)}}}{\phi_i^{(r)}\sqrt{(\phi_i^2)^{(r)} + \frac{2y_i}{\mu_i^{(r)}}} + 1}, \tag{3.15}$$

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

- **M-step**:
  - Update the regression parameters $\boldsymbol{\beta}_1$ using the pseudo-values w_i, which are given by Equation (3.14), and the Newton–Raphson algorithm in Equations (3.4)–(3.6).
  - Update the regression parameters $\boldsymbol{\beta}_2$ using the pseudo-values w_i and ω_i, which are given by Equations (3.14) and (3.15) respectively, and the Newton–Raphson algorithm which, in the case of the IG mixing distribution, goes as follows:

$$h_2(\boldsymbol{\beta}_2) = \left[1 + 2\left(\phi_i^2\right)^{(r)} - \left(\phi_i^2\right)^{(r)} w_i - \left(\phi_i^2\right)^{(r)} \omega_i\right] x_{2,ij}, \tag{3.16}$$

and

$$H_2(\boldsymbol{\beta}_2) = \sum_{i=1}^{n} \left[4\left(\phi_i^2\right)^{(r)} - 2\left(\phi_i^2\right)^{(r)} w_i - 2\left(\phi_i^2\right)^{(r)} \omega_i \right] x_{2,ij} x_{2,ij}^T = \mathbf{X}_2^T \mathbf{W}_2 \mathbf{X}_2, \tag{3.17}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p_2$ and where $\mathbf{W}_2 = diag\{4(\phi_i^{(r)})^2 - 2(\phi_i^{(r)})^2 w_i - 2(\phi_i^{(r)})^2 \omega_i\}$.

Then, we can obtain the updated estimates of $\boldsymbol{\beta}_2^{(r)}$ using Equation (3.9).

3.3. ELN regression model with varying dispersion

The EM algorithm can also be employed to find the ML estimates of the ELN model which was defined in Equation (2.16). In this case, the complete data log-likelihood takes the form:

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[-\frac{y_i}{\mu_i z_i} - \log(\mu_i) - \log(z_i) \right]$$

$$+ \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi) - \log(\phi_i) - \log(z_i) - \frac{\left(\log(z_i) + \frac{\phi_i^2}{2}\right)^2}{2\phi_i^2} \right], \tag{3.18}$$

for $i = 1, \ldots, n$. Thus, the expectations needed for the M-step are $E_{z_i}[\frac{1}{z_i}]$ and $E_{z_i}[(\log(z_i))^2]$. Hence the algorithm can be written as follows:

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

$$w_i = E_{Z_i}\left[\frac{1}{Z_i} | y_i; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\displaystyle\int_0^{\infty} \frac{1}{z_i} \frac{e^{-\frac{y_i}{\mu_i^{(r)} z_i}}}{\mu_i^{(r)} z_i} \frac{\exp\left[-\frac{\left(\log(z_i) + \frac{\left(\phi_i^2\right)^{(r)}}{2}\right)^2}{2\left(\phi_i^2\right)^{(r)}}\right]}{\sqrt{2\pi}\phi_i^{(r)} z_i} \, dz_i}{\displaystyle\int_0^{\infty} \frac{e^{-\frac{y_i}{\mu_i^{(r)} z_i}}}{\mu_i^{(r)} z_i} \frac{\exp\left[-\frac{\left(\log(z_i) + \frac{\left(\phi_i^2\right)^{(r)}}{2}\right)^2}{2\left(\phi_i^2\right)^{(r)}}\right]}{\sqrt{2\pi}\phi_i^{(r)} z_i} \, dz_i} \tag{3.19}$$

and

$$\omega_i = E_{Z_i}\left[\left(\log\left(Z_i\right)\right)^2 \mid y_i; \boldsymbol{\theta}^{(r)}\right]$$

$$= \frac{\displaystyle\int_0^\infty \left(\log\left(z_i\right)\right)^2 \frac{e^{-\frac{y_i}{\mu_i^{(r)} z_i}}}{\mu_i^{(r)} z_i} \frac{\exp\left[-\frac{\left(\log(z_i)+\frac{\left(\phi_i^2\right)^{(r)}}{2}\right)^2}{2\left(\phi_i^2\right)^{(r)}}\right]}{\sqrt{2\pi}\,\phi_i^{(r)} z_i}\,dz_i}{\displaystyle\int_0^\infty \frac{e^{-\frac{y_i}{\mu_i^{(r)} z_i}}}{\mu_i^{(r)} z_i} \frac{\exp\left[-\frac{\left(\log(z_i)+\frac{\left(\phi_i^2\right)^{(r)}}{2}\right)^2}{2\left(\phi_i^2\right)^{(r)}}\right]}{\sqrt{2\pi}\,\phi_i^{(r)} z_i}\,dz_i}, \tag{3.20}$$

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

Clearly the expectations involved in the E-step of the algorithm do not have closed form expressions and thus numerical approximations are needed. Specifically, Equations (3.19) and (3.20) can be evaluated numerically. Alternatively, a Monte Carlo approach is also possible using a rejection algorithm. The latter case leads to variants of the EM algorithm such as the Monte Carlo EM algorithm (see, for instance, Booth and Hobert, 1999; Booth *et al.*, 2001; Karlis, 2001, 2005) which do not require knowledge of the pdf $f(y_i|z_i)$ but it suffices to be able to simulate from the posterior density $g\left(z_i|y_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i}\right)$.

- **M-step**:
  - Update the regression parameters $\boldsymbol{\beta}_1$ using the pseudo-values w_i, which are given by Equation (3.19), and the Newton–Raphson algorithm in Equations (3.4)–(3.6).
  - Update the regression parameters $\boldsymbol{\beta}_2$ using the pseudo-values w_i and ω_i, which are given by Equations (3.19) and (3.20) respectively, and the Newton–Raphson algorithm which, in the case of the LN mixing distribution, is as follows

$$h_2(\boldsymbol{\beta}_2) = \left[\frac{\omega_i}{\left(\phi_i^2\right)^{(r)}} - \frac{\left(\phi_i^2\right)^{(r)}}{4} - 1\right] x_{2,ij}, \tag{3.21}$$

and

$$H_2(\boldsymbol{\beta}_2) = \sum_{i=1}^n \left[\frac{-2\omega_i}{\left(\phi_i^2\right)^{(r)}} - \frac{\left(\phi_i^2\right)^{(r)}}{2}\right] x_{2,ij} x_{2,ij}^T = \mathbf{X}_2^T \mathbf{W}_2 \mathbf{X}_2, \tag{3.22}$$

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

Then, we can obtain the updated estimates of $\boldsymbol{\beta}_2^{(r)}$ using Equation (3.9).

4. NUMERICAL ILLUSTRATION

The data were kindly provided by a major insurance company operating in Greece and concern a MTPL insurance portfolio which was observed during the year 2017. Our interest lies in identifying factors that affect the amount paid for each claim and specifically the factors that correspond to the insured person and their characteristics, including the characteristics of their car. Only policyholders with complete records, that is, with availability of all the explanatory variables under consideration were included in the sample. Also, only policyholders with at least one reported accident were used for our analyses. There were 6993 observations that met our criteria. Additionally, an exploratory analysis was carried out in order to adequately select a subset of explanatory variables with the highest predictive power for claim sizes. The response variable is the costs of claims at fault registered for each insured vehicle in the data set and the a priori rating variables we employ are: the size of the city (CS) where the policyholders live, the age of their car (AC) and the horsepower (HP) of their car. Furthermore, in light of the heterogeneity which exists within the portfolio, consideration was given to grouping the levels of each explanatory variable with respect to risk profiles with similar claim amounts. This was done in order to achieve ratemaking accuracy and to balance homogeneity and sufficiency of the volume of data in each cell in order to provide credible patterns.

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

Table 1 contains some standard descriptive statistics for claim severities along with the number of observations in each category of the three explanatory variables.

In what follows, in our real data application, we will compare the Pareto, EIG and ELN regression models with varying dispersion with the traditional two parameter claims severity distributions Gamma (GA), IG and LN with regression components in their mean and dispersion parameters. Note that the GA, IG and LN regression models with regression structures on their mean and dispersion parameters can be fitted using the GAMLSS package. For more details on the GAMLSS package, see Stasinopoulos et al. (2008). Furthermore,

TABLE 1

DESCRIPTIVE STATISTICS OF CLAIM SEVERITIES—SIZE OF THE DIFFERENT
CATEGORIES OF THE EXPLANATORY VARIABLES.

Statistic	Claim severities	Size of the city (CS)		Age of the Car (AC)		Horsepower of the car (HP)	
Minimum	45	C1:	1755	C1:	1972	C1:	2586
Median	2289.4	C2:	1769	C2:	4090	C2:	2989
Mean	5578.8	C3:	3469	C3:	931	C3:	1418
Maximum	150,024		–		–		–

we will perform a brief study based on Monte Carlo simulations to assess the accuracy of the EM algorithm for the proposed class of mixed Exponential models.

Because we will be comparing the proposed mixed Exponential models with the GA, IG and LN models, we provide below expressions for the pdf, the mean and the variance of the latter three models.

- The pdf of the Gamma (GA) distribution is given by

$$f(y_i) = \frac{1}{\left(\phi_i^2 \mu_i\right)^{\frac{1}{\phi_i^2}}} \frac{y_i^{\frac{1}{\phi_i^2}-1} \exp\left(-\frac{y_i}{\phi_i^2 \mu_i}\right)}{\Gamma\left(\frac{1}{\phi_i^2}\right)}, \tag{4.1}$$

where $y_i > 0$, $\mu_i > 0$ and $\phi_i > 0$ for $i = 1, \ldots, n$, with mean and variance given by

$$E(Y_i) = \mu_i \tag{4.2}$$

and

$$Var(Y_i) = \phi_i^2 \mu_i^2. \tag{4.3}$$

- The pdf of the IG distribution is given by

$$f(y_i) = \frac{1}{\sqrt{2\pi \phi_i^2 y_i^3}} e^{\left[-\frac{(y_i - \mu_i)^2}{2\mu_i^2 \phi_i^2 y_i}\right]}, \tag{4.4}$$

where $y_i > 0$, $\mu_i > 0$ and $\phi_i > 0$ for $i = 1, \ldots, n$, with mean and variance given by

$$E(Y_i) = \mu_i \tag{4.5}$$

and

$$Var(Y_i) = \phi_i^2 \mu_i^3. \tag{4.6}$$

- The pdf of the LN distribution is given by

$$f(y_i) = \frac{1}{y_i \sqrt{2\pi} \phi_i} e^{\left[-\frac{[\log(y_i) - \mu_i]^2}{2\phi_i^2}\right]}, \tag{4.7}$$

where $y_i > 0$, $\mu_i > 0$ and $\phi_i > 0$ for $i = 1, \ldots, n$, with mean and variance given by

$$E(Y_i) = \sqrt{e^{\phi_i^2}} e^{\mu_i} \tag{4.8}$$

and

$$Var(Y_i) = e^{\phi_i^2} \left(e^{\phi_i^2} - 1\right) e^{2\mu_i}. \tag{4.9}$$

Finally, we assume that the mean and dispersion parameters of the GA, IG and LN distributions are modelled in terms of covariates

$$\mu_i = \exp\left(\mathbf{x}_{1,i}^T \boldsymbol{\beta}_1\right) \text{ and} \tag{4.10}$$

$$\phi_i = \exp\left(\mathbf{x}_{2,i}^T \boldsymbol{\beta}_2\right), \tag{4.11}$$

where $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ are covariate vectors with dimensions $p_1 \times 1$ and $p_2 \times 1$ respectively, with $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, \ldots, \beta_{1,p_1}\right)^T$ and $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, \ldots, \beta_{2,p_2}\right)^T$ the corresponding parameter vectors and where it is assumed that the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, with rows given by $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ respectively, are of full rank.

4.1. Modelling results

This subsection describes the modelling results of the Pareto, EIG, ELN, GA, IG and LN models with regression specifications on their mean and dispersion parameters. For illustrative purposes we considered that the location and the dispersion parameters of the aforementioned claim severity models are modelled using all three available explanatory variables. The ML estimates of the parameters and the corresponding standard errors in parentheses for the aforementioned claim severity models are presented in Table 2. However, it should be noted that for larger data sets variable selection can start with the examination of the mean parameter of the claim severity model. This can be achieved by adding all available covariates and testing whether the exclusion of each one lowers the Global Deviance (DEV), Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC) values. Then, after having selected the best predictor for the mean parameter, one can continue in determining the remaining predictors by testing which rating variable between those used in the mean parameter would lead to a further decrease of the DEV, AIC and SBC values when inserted in the dispersion parameter of the claim severity model. Additionally, if between the same claim severity distribution with different parameter specifications several models have similar DEV, AIC and SBC

TABLE 2

RESULTS OF THE FITTED PARETO, EIG, ELN, GA, IG AND LN MODELS.

| Pareto		EIG		ELN		GA		IG		LN	
	Coeff. β_1		Coeff. β_1		Coeff. β_1		Coeff. β_1		Coeff. β_1		Coeff. β_1
Intercept	8.684 (0.093)	Intercept	8.603(0.059)	Intercept	8.574(0.048)	Intercept	8.643 (0.041)	Intercept	8.642 (0.075)	Intercept	7.706 (0.044)
CS		CS		CS		CS		CS		CS	
C2	0.062 (0.088)	C2	0.064(0.061)	C2	0.064(0.056)	C2	0.072 (0.043)	C2	0.073 (0.077)	C2	0.060 (0.047)
C3	0.013 (0.076)	C3	0.010(0.051)	C3	0.015(0.024)	C3	0.005 (0.036)	C3	0.005 (0.065)	C3	−0.010 (0.040)
AC		AC		AC		AC		AC		AC	
C2	−0.082 (0.074)	C2	−0.054(0.049)	C2	−0.044(0.042)	C2	−0.067 (0.033)	C2	−0.067 (0.062)	C2	−0.013 (0.037)
C3	−0.034 (0.115)	C3	−0.038(0.071)	C3	−0.031(0.043)	C3	−0.057 (0.049)	C3	−0.051 (0.089)	C3	−0.030 (0.054)
HP		HP		HP		HP		HP		HP	
C2	0.064 (0.072)	C2	0.047(0.047)	C2	0.050(0.037)	C2	0.036 (0.033)	C2	0.036 (0.061)	C2	0.033 (0.037)
C3	−0.052 (0.082)	C3	−0.030(0.059)	C3	−0.023(0.049)	C3	−0.037 (0.042)	C3	−0.037 (0.075)	C3	0.008 (0.047)
	Coeff. β_2		Coeff. β_2		Coeff. β_2		Coeff. β_2		Coeff. β_2		Coeff. β_2
Intercept	0.702 (0.098)	Intercept	−0.074(0.015)	Intercept	−0.138(0.014)	Intercept	0.213 (0.019)	Intercept	−3.541 (0.023)	Intercept	0.283 (0.023)
CS		CS		CS		CS		CS		CS	
C2	−0.006 (0.097)	C2	−0.009(0.076)	C2	0.008(0.059)	C2	0.006 (0.021)	C2	−0.019 (0.025)	C2	0.004 (0.025)
C3	−0.004 (0.083)	C3	−0.009(0.064)	C3	0.015(0.047)	C3	0.008 (0.017)	C3	0.054 (0.021)	C3	0.030 (0.021)
AC		AC		AC		AC		AC		AC	
C2	0.101 (0.078)	C2	0.081(0.059)	C2	−0.053(0.052)	C2	−0.026 (0.016)	C2	0.025 (0.019)	C2	−0.002 (0.019)
C3	0.017 (0.118)	C3	0.021(0.086)	C3	−0.010(0.060)	C3	−0.012 (0.023)	C3	0.024 (0.028)	C3	0.007 (0.028)
HP		HP		HP		HP		HP		HP	
C2	−0.031 (0.078)	C2	−0.018(0.058)	C2	0.016(0.032)	C2	0.001 (0.016)	C2	−0.006 (0.019)	C2	0.013 (0.019)
C3	0.087 (0.099)	C3	0.065(0.077)	C3	−0.045(0.050)	C3	−0.022 (0.021)	C3	−0.041 (0.024)	C3	−0.017 (0.024)
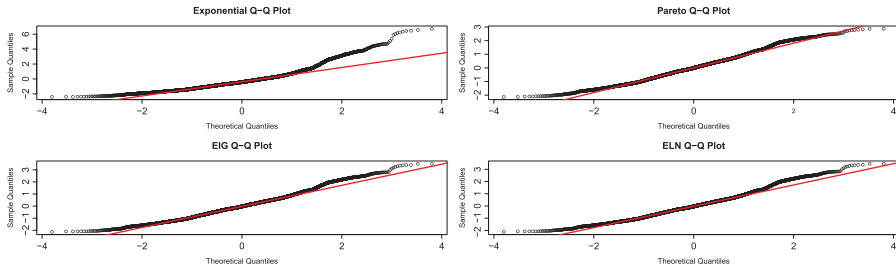
FIGURE 1: Normalized quantiles for the Pareto, EIG and ELN models.

values, the simpler model can be used in order to avoid overfitting. Therefore, in such cases, it should be expected that the dispersion parameters of the claim severity models will have fewer predictors than the mean parameters.

As we can see from Table 2, the same explanatory variables do not always have the same effect (positive and/or negative) on the parameters μ_i and ϕ_i, $i = 1, \ldots, n$, across all claim severity models. Furthermore, with the exception of the LN model, we observe that, as expected, the values of the estimated regression coefficients of explanatory variables are almost identical for μ_i but different for ϕ_i in the case of Pareto, EIG, ELN, GA and IG models which, unlike the case of the LN model, are defined so that the mean, μ_i, is an explicit parameter of each model. Hence, the traditional risk classification approach, which assumes that only the parameter μ_i of these models can be modelled in terms of covariates, would lead to mean claim severity values which will not differ much under different distributional assumptions. On the contrary, allowing both μ_i and ϕ_i to be modelled as functions of a priori rating variables breaks the nexus between the mean and variance implied by the standard procedure using claim severity regression models, leading to a more complete comparison of all the claim severity models through their variance values, see Equations (2.10), (2.14), (2.18), (4.3) and (4.9).

Finally, we rely on normalized quantile residuals (see, Dunn and Smyth, 1996) as an exploratory graphical device for investigating the adequacy of the fit of the competing mixed Exponential regression models with varying dispersion. Also, for comparison purposes, we fitted the simple Exponential regression model, which obviously has a thinner tail than the three mixed Exponential models. For these continuous response distributions, the normalized (randomized) quantile residuals are defined as $\hat{r}_i = \Phi^{-1}(u_i)$, where Φ^{-1} is the inverse cdf of a standard Normal distribution and $u_i = F_i(y_i|\boldsymbol{\theta}^{(r+1)})$, where F_i is the cdf estimated for the ith individual, $\boldsymbol{\theta}^{(r+1)}$ contains all estimated model parameters after the EM algorithm has reached the global maximum and y_i is the corresponding observation. The model fit can be evaluated by means of usual quantile–quantile plots. Specifically, if the data indeed follow the assumed distribution, then the residual on the quantile–quantile plot will fall approximately on a straight line. Figure 1 shows the normalized (random) quantiles for the Exponential, Pareto, EIG and ELN claim severity regression

PARETO, EIG, ELN, GA, IG AND LN MODELS COMPARISON.

Model	df	DEV	AIC	SBC
Pareto	14	132,295	132,323	132,419
EIG	14	132,353	132,381	132,477
ELN	14	132,389	132,417	132,512
GA	14	134,057	134,085	134,181
IG	14	133,490	133,518	133,614
LN	14	133,029	133,057	133,153

models with varying dispersion. From Figure 1 we observe that the residuals indicate that the Pareto, EIG and ELN models are better assumptions than the Exponential model since the residuals of the former three are close to the diagonal and indicate an adequately good fit to the distribution of the claim severity. Furthermore, it can be seen that the Pareto model yields a slightly better performance than the EIG and ELN models close to the tail of the claim size distribution. Overall, based on these plots, it is reasonable to suggest the employment of the Pareto, EIG and ELN models over the Exponential model for modelling claim severities in our data set.

4.1.1. *Models comparison*

In this subsection we compare the fit of the Pareto, EIG, ELN, GA, IG and LN regression models with varying dispersion for the observed claim sizes in the MTPL insurance portfolio we analyzed earlier employing the DEV, AIC and SBC. The DEV is defined as

$$DEV = -2\hat{l}\left(\hat{\boldsymbol{\theta}}\right),\tag{4.12}$$

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

$$AIC = DEV + 2 \times df\tag{4.13}$$

and the SBC is given by

$$SBC = DEV + \log(n) \times df,\tag{4.14}$$

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

The resulting DEV, AIC and SBC are given in Table 3 for the different claim severity fitted models.

As is well known, according to a commonly used rule of thumb, two models can be considered to be significantly different if the difference in the log-likelihoods exceeds five, corresponding to a difference in their respective
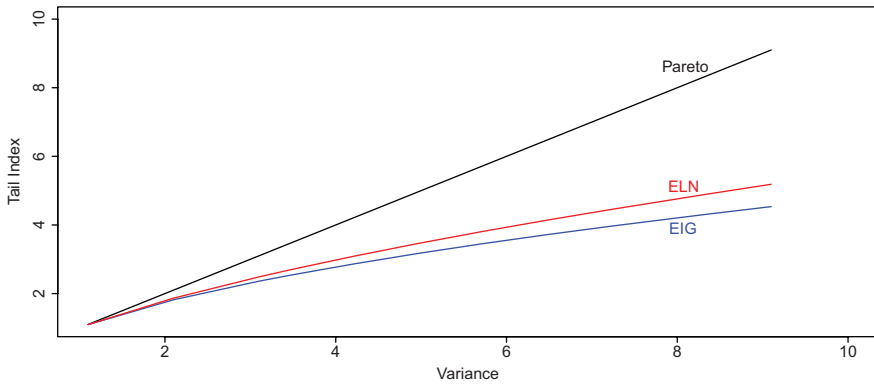
FIGURE 2: Plot of the right tail index as a function of the variance for the Pareto, ELN and EIG models with unit mean.

AIC values of more than ten and to a difference in their SBC values of more than five, see Burnham and Anderson (2002) and Raftery (1995) respectively. Thus, in this case we observe that overall, with respect to the Global DEV, AIC and SBC indices, the best fitting performances are provided by the mixed Exponential models. In particular, the Pareto model has the lowest DEV, AIC and SBC values followed by the EIG and ELN models respectively.

4.1.2. Estimation of the right tail index

Let us now present a ranking of the Pareto, EIG and ELN models by the right tail index[5] which is a risk measure for right tail deviation that was suggested by Wang (1998). The right tail index is defined as

$$d(Y) = \frac{\int\limits_0^\infty \sqrt{S_Y(t)}dt}{E(Y)} - 1, \qquad (4.15)$$

where $S_Y(t) = P(Y \geq t)$, is the survival function, or the decumulative distribution function, of Y.

Figure 2 displays three plots of the right tail index, $d(Y)$, as a function of the variance for the Pareto, ELN and EIG models respectively. The parameters were chosen so that the Pareto, ELN and EIG models have a unit mean, that is, $E(Y) = 1$, and varying variance, $Var(Y)$, taking on the same values for all densities. From Figure 2, we see that the right tail index ranks the Pareto distribution as having a fatter tail than the ELN distribution, which in turn has a fatter tail than the EIG distribution.

Finally, we will investigate the tail adequacy of the Pareto, EIG and ELN models for our data set. The empirical estimator of the right tail index, $\hat{d}(Y)$,

which was considered by Jones & Zitikis ([2003](#)) and can be calculated as

$$\hat{d}(Y) = \sum_{i=1}^{n} c_i \frac{Y_{(i)}}{\bar{Y}}, \tag{4.16}$$

where $Y_{(i)}$ is the i-th ordered observation of the sample $Y_1, \ldots, Y_n$ and where the coefficients c_i are given by

$$c_i = \sqrt{\left(\frac{n-i}{n}\right)} + \sqrt{\left(\frac{n-i+1}{n}\right)} - \frac{1}{n}, \tag{4.17}$$

for $i = 1, \ldots, n$.

The value which we obtained for the empirical estimator of the right tail index is $\hat{d}(Y) = 2.245$ while the Pareto, EIG and ELN distributions have right tail indices $d_1(Y) = 2.784$, $d_2(Y) = 2.373$ and $d_3(Y) = 2.492$ which are relatively close to the empirical result. However, since for smaller data sets the empirical approach can lead to an underestimation of $d(Y)$ it makes sense to build a parametric bootstrap two-sided confidence interval (CI) for $\hat{d}(Y)$. Given our data, we generated $B = 100{,}000$ bootstrap samples of size 6993 and we calculated the 95% bootstrap-based CI for $\hat{d}(Y)$ to be (2.058, 2.838). Thus, since the values of $d_1(Y)$, $d_2(Y)$ and $d_3(Y)$ are included in this CI, we can conclude that the Pareto, EIG and ELN distributions are able to effectively model the right tail of the data.

4.2. Simulation results

In this subsection a small Monte Carlo simulation study will be carried out to investigate the accuracy of the proposed EM-type algorithm for the class of mixed Exponential regression models with varying dispersion.

In particular, the EM-based ML estimators of the Pareto[6] regression model with varying dispersion will be compared against the ML estimators of the same model obtained via the GAMLSS fitting algorithm of Rigby and Stasinopoulos ([2005](#)), henceforth called the Pareto GAMLSS, by conducting the following experiment.

We generate 5000 data sets of sizes $n = 100, 150, 250, 500$ for the Pareto model based on the estimates $(\beta_{1,0}, \beta_{1,1}, \beta_{1,2}) = (8.684, -0.082, 0.064)$ and $(\beta_{2,0}, \beta_{2,1}, \beta_{2,2}) = (0.702, 0.101, -0.031)$ of the intercept, AC category 2 and HP category 2 for the mean and dispersion parameters, μ_i and ϕ_i respectively, given in Table [2](#) and the corresponding vectors of covariates $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$, for $i = 1, \ldots, n$. Then, we compute the mean of the Pareto firstly based on the EM algorithm and secondly by using the GAMLSS algorithm.

The results are shown in Table [4](#), where it is evident that the EM and GAMLSS approaches provide equally good estimates of the regression coefficients for μ_i. Furthermore, we observe that, especially for small and moderate

TABLE 4

RESULTS OF THE SIMULATION STUDY INVOLVING 5000 SIMULATIONS OF DATA SETS OF SIZE N.

Sample Size	Pareto EM Alg.	Pareto GAMLSS
$n = 100$		
$\beta_{1,0}$	8.688	8.687
$\beta_{1,1}$	-0.092	-0.094
$\beta_{1,2}$	0.020	0.019
$\beta_{2,0}$	1.663	1.818
$\beta_{2,1}$	0.333	0.368
$\beta_{2,2}$	-1.184	-1.366
$n = 150$		
$\beta_{1,0}$	8.661	8.661
$\beta_{1,1}$	-0.069	-0.067
$\beta_{1,2}$	0.069	0.070
$\beta_{2,0}$	1.057	1.065
$\beta_{2,1}$	0.253	0.264
$\beta_{2,2}$	-0.441	-0.462
$n = 250$		
$\beta_{1,0}$	8.673	8.672
$\beta_{1,1}$	-0.076	-0.073
$\beta_{1,2}$	0.066	0.068
$\beta_{2,0}$	0.793	0.807
$\beta_{2,1}$	0.186	0.197
$\beta_{2,2}$	-0.154	-0.166
$n = 500$		
$\beta_{1,0}$	8.676	8.675
$\beta_{1,1}$	-0.077	-0.075
$\beta_{1,2}$	0.065	0.067
$\beta_{2,0}$	0.745	0.758
$\beta_{2,1}$	0.121	0.138
$\beta_{2,2}$	-0.077	-0.086

sample sizes, the EM algorithm provides better estimates of the regression coefficients for ϕ_i than the GAMLSS algorithm in terms of bias.

5. COMPUTATIONAL ASPECTS

In this section, we point out some computational issues related to the implementation of the EM algorithm for the Pareto, EIG and ELN regression models with varying dispersion.

We should emphasize that for all three mixed Exponential models, the choice of initial values for the vectors of the regression coefficients β_1 and β_2 needed special attention because the M-step involves two Newton–Raphson

iterations and hence one may obtain inadmissible values if the starting values are bad. Good starting values for the regression parameters $\boldsymbol{\beta}_1$ were obtained by fitting the simple Exponential regression. Alternatively, the initial values for $\boldsymbol{\beta}_1$ can be obtained based on the data as follows: (i) calculate $E(y_i)$ for the 27 different risk classes, which can be formed by dividing the portfolio into clusters defined by the combinations of the characteristics of the policyholders and their cars based on all observations $i = 1, \ldots, n$ and (ii) assuming a log-link function for μ_i, see Equation (2.3), solve Equation (2.5) with respect to $\boldsymbol{\beta}_1$, since the parameterization we adopted considers that the mean is an explicit parameter of the Pareto, EIG and ELN models respectively. Furthermore, meaningful initial values for the regression parameters $\boldsymbol{\beta}_2$ were obtained by: (i) calculating $Var(y_i)$ for the 27 different risk classes based on all observations $i = 1, \ldots, n$ and (ii) calculating $E(y_i)$ for the 27 different risk classes (or alternatively calculating μ_i based on the initial values for $\boldsymbol{\beta}_1$ and the log-link function given by Equation (2.3)) and using the log-link function for ϕ_i, see Equation (2.4), so we solve the Equations (2.10), (2.14) and (2.18) with respect to $\boldsymbol{\beta}_2$ for the case of the Pareto, EIG and ELN models respectively.

Furthermore, the standard errors were obtained using the standard approach of Louis (1982) for the standard errors for the EM algorithm. Finally, all computing was made using the programming language R. The ELN model needed more time than the other two mixed Exponential models because the numerical evaluation of the integrals is computationally time consuming when regression structures are used for all the parameters of the model.

6. AN EXTENDED VERSION OF THE MIXED EXPONENTIAL MODEL

Finally, since two of the three examples of the mixing distributions considered in this paper, namely the Gamma and IG distributions, are limiting and special cases respectively of the more general family of GIG distributions, see Jørgensen (1982), it is interesting to explore how ML estimation of the Exponential-GIG (EGIG) model can be achieved when regression specifications are allowed to all of its parameters.

The EGIG distribution, whose pdf is given by Equation (6.1) below, see also, for instance, Frangos and Karlis (2004), can be considered as a candidate model for large claim sizes.

$$f(y_i) = \frac{K_{\nu_i-1}\left(\phi_i\sqrt{\mu_i^2 + 2y_i}\right)\phi_i\left(\mu_i^2 + 2y_i\right)^{\frac{\nu_i-1}{2}}}{K_{\nu_i}\left(\mu_i\phi_i\right)\mu_i^{\nu_i}}, \tag{6.1}$$

where $y_i > 0$, $\mu_i > 0$, $\phi_i > 0$ and $-\infty < \nu_i < \infty$, and where

$$K_\nu(\omega) = \int_0^\infty \chi^{\nu-1}\exp\left[-\frac{1}{2}\omega\left(\chi + \frac{1}{\chi}\right)\right]d\chi, \tag{6.2}$$

is the modified Bessel function of the third kind of order v with argument ω.

Under our general approach, the following regression structures can be chosen to ensure a valid range for the parameters of the EGIG distribution

$$\mu_i = \exp\left(\mathbf{x}_{1,i}^T \boldsymbol{\beta}_1\right), \tag{6.3}$$

$$\phi_i = \exp\left(\mathbf{x}_{2,i}^T \boldsymbol{\beta}_2\right) \text{ and} \tag{6.4}$$

$$v_i = \mathbf{x}_{3,i}^T \boldsymbol{\beta}_3 \tag{6.5}$$

where $\mathbf{x}_{1,i}$, $\mathbf{x}_{2,i}$ and $\mathbf{x}_{3,i}$ are covariate vectors with dimensions $p_1 \times 1$, $p_2 \times 1$ and $p_3 \times 1$ respectively, with $\boldsymbol{\beta}_1 = \left(\beta_{1,1}, \ldots, \beta_{1,p_1}\right)^T$, $\boldsymbol{\beta}_2 = \left(\beta_{2,1}, \ldots, \beta_{2,p_2}\right)^T$ and $\boldsymbol{\beta}_3 = \left(\beta_{3,1}, \ldots, \beta_{3,p_3}\right)^T$ the corresponding parameter vectors and where it is considered that the matrices $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ with rows given by $\mathbf{x}_{1,i}$, $\mathbf{x}_{2,i}$ and $\mathbf{x}_{3,i}$ respectively, are of full rank, for $i = 1, \ldots, n$.

At this point, it should be noted that ML estimation of the EGIG model is complicated to calculate when μ_i, ϕ_i and v_i vary through covariates. One of the main reasons is because the vector of regression coefficients β_3 is involved in the calculation of the order v_i of the modified Bessel function of the third kind and, as is well known, it is not always possible to obtain numerically reliable direct derivatives of this special function with respect to its order, see, for instance, Mencía and Sentana (2005).

Nevertheless, an extended version of the EM algorithm together with an accurate numerical approximation of the derivatives of the modified Bessel function of the third kind can reduce the computational burden for ML estimation in the EGIG model. The two steps of the EM procedure for estimating the parameters of the EGIG model can be briefly described as follows. At the E-step one needs to calculate the Q-function, which is the conditional expectation of the complete data log-likelihood of the EGIG model, and then the M-Step involves using the Newton–Raphson procedure three times maximizing the Q-function with respect to β_1, β_2 and β_3. Specifically, the Q-function should be updated by using the current estimates of β_1, β_2 and β_3 and then maximized again until some convergence criterion is satisfied.

7. CONCLUDING REMARKS

The main purpose of this paper was to introduce a general class of mixed Exponential regression models with varying dispersion which can efficiently capture the influence of tail risks. An EM scheme for estimating the parameters of the mixed Exponential models was also presented. In order to emphasize the utility and generality of our ML estimation framework a real data application provided a fresh look into the process of modelling and estimation of three mixed Exponential regression models with varying dispersion, namely the

Pareto, EIG and ELN distributions with regression structures for both their mean and dispersion parameters.

Finally, it is worth mentioning that the mixed Exponential models we considered in this study were parametric, and it would be interesting to explore a semiparametric approach when functional forms other than the linear are included in the mean and the dispersion parameters of the models, proceeding along similar lines as the GAMLSS approach of Rigby and Stasinopoulos (2005) and Stasinopoulos *et al.* (2017). Also see, for instance, Heller *et al.* (2007) and De Jong and Heller (2008) who used GAMLSS for the statistical analysis of the total amount of insurance paid out on a policy. Finally, the data augmentation which was used in the paper to derive the EM algorithm can be the basis for constructing Bayesian estimation methods as in the article by Klein *et al.* (2014) where Bayesian generalized additive models for location, scale and shape were employed for nonlife ratemaking and risk management.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/asb.2020.13.

NOTES

1. For instance, in MTPL insurance $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$ contain the characteristics of the policyholders and their cars.

2. The code for Section 3 is available online as Supplementary Material.

Please note that a toy data set is used in the code because we could n't make the data set which we used in the paper publicly available due to data use agreement with the company which provided the data.

3. Note that, as it will be demonstrated in what follows, if $s_k(z_i)$ is a linear function then the conditional posterior expectations can be easily and accurately obtained. Regarding more complicated functions, for which an exact solution is not available, one can resort to Taylor approximations, or numerical approximations, including numerical integration, and/or simulation-based approximations.

4. Note also that this procedure can be used for every continuous and at least twice differentiable mixing distribution, that is, similar to those we considered in this study. This approach works well in practice and hence we provide a complete estimation tool for our general class of mixed Exponential regression models with varying dispersion. However, for some other mixing

distributions a special iterative scheme may be appropriate and perhaps another EM algorithm inside the M-Step.

5. Note that the residuals and the right tail index for the ELN model which does not have a pdf in closed form can be efficiently computed using numerical integration methods.

6. Note that the reason why the Pareto model was chosen is because, as was previously mentioned, with the exception of this model, which is implemented into the GAMLSS package, ML estimation procedures for other mixed Exponential models have not been explored in the applied statistical literature so far in the case when regression specifications are allowed for both their mean and dispersion parameters.

REFERENCES

AHN, S., KIM, J.H. and RAMASWAMI, V. (2012) A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics*, **51**(1), 43–52.

BARRETO-SOUZA, W. and SIMAS, A.B. (2016) General mixed Poisson regression models with varying dispersion. *Statistics and Computing*, **26**(6), 1263–1280.

BEIRLANT, J., DERVEAUX, V., DE MEYER, A.M., GOOVAERTS, M.J., LABIE, E. and MAENHOUDT, B. (1992) Statistical risk evaluation applied to (Belgian) car insurance. *Insurance: Mathematics and Economics*, **10**(4), 289–302.

BHATTACHARYA, S.K. and KUMAR, S. (1986) E-IG model in life testing. *Calcutta Statistical Association Bulletin*, **35**(1–2), 85–90.

BLADT, M. and ROJAS-NANDAYAPA, L. (2018) Fitting phase–type scale mixtures to heavy–tailed data and distributions. *Extremes*, **21**(2), 285–313.

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

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

BURNHAM, K.P. and ANDERSON, D.R. (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York: Springer.

CALDERÍN-OJEDA, E., FERGUSSON, K. and WU, X. (2017) An EM algorithm for double-Pareto-lognormal generalized linear model applied to heavy-tailed insurance claims. *Risks*, **5**(4), 60.

COLE, T.J. and GREEN, P.J. (1992) Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, **11**(10), 1305–1319.

DE JONG, P. and HELLER, G.Z. (2008) *Generalized Linear Models for Insurance Data*, Vol. 10. Cambridge: Cambridge University Press.

DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **39**(1), 1–38.

DENUIT, M., MARECHAL, X., PITREBOIS, S. and WALHIN, J.F. (2007) *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems.* Chichester, West Sussex: John Wiley and Sons.

DUNN, P.K. and SMYTH, G.K. (1996) Randomized quantile residuals. *Computational and Graphical Statistics*, **5**(3), 236–245.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997) *Modelling Extremal Events*. Applications of Mathematics, Vol. 33. Berlin, Heidelberg: Springer-Verlag.

FRANGOS, N. and VRONTOS, S. (2001) Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin*, **31**(1) 1–22.

FRANGOS, N. and KARLIS, D. (2004) Modelling losses using an exponential-inverse Gaussian distribution. *Insurance: Mathematics and Economics*, **35**(1), 53–67.

FREES, E.W. and VALDEZ, E.A. (2008) Hierarchical insurance claims modeling *Journal of the American Statistical Association*, **103**(484), 1457–1469.

FREES, E.W., DERRIG, R.A. and MEYERS, G. (2014a) *Predictive Modeling Applications in Actuarial Science*, Vol. 1. New York: Cambridge University Press.

FREES, E.W., DERRIG, R.A. and MEYERS, G. (2014b) *Predictive Modeling Applications in Actuarial Science*, Vol. 2. New York: Cambridge University Press.

HELLER, G.Z., STASINOPOULOS, M.D., RIGBY R.A. and DE JONG, P. (2007) Mean and dispersion modeling for policy claims costs. *Scandinavian Actuarial Journal*, **4**, 281–292.

HESSELAGER, O., WANG, S. and WILLMOT, G. (1998) Exponential and scale mixtures and equilibrium distributions. *Scandinavian Actuarial Journal*, **2**, 125–142.

HÜRLIMANN, W. (2014) Pareto type distributions and excess-of-loss reinsurance. *International Journal of Research and Reviews in Applied Sciences*, **18**(3), 1.

Jones, B.L. and Zitikis, R. (2003) Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, **7**(4), 44–54.

JØRGENSEN, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture Notes in Statistics, Vol. 9. New York: Springer-Verlag.

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

KARLIS, D. (2005) EM algorithm for mixed Poisson and other discrete distributions. *ASTIN Bulletin*, **35**(1), 3–24.

KLEIBER, C. and KOTZ, S. (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, New Jersey: John Wiley and Sons.

KLEIN, N., DENUIT, M., LANG, S. and KNEIB, T. (2014) Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape. *Insurance: Mathematics and Economics*, **55**, 225–249.

KLUGMAN, S., PANJER, H. and WILLMOT, G. (2012) *Loss Models: From Data to Decisions*, Vol. 715. New York: John Wiley and Sons.

LOUIS, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **44**, 226–233.

MCLACHLAN, G. and KRISHNAN, T. (2007) *The EM Algorithm and Extensions*, Vol. 382. Hoboken, New Jersey: John Wiley and Sons.

MENCÍA, J. and SENTANA, E. (2005) *Estimation and testing of dynamic models with generalized hyperbolic innovations*. CEPR Discussion Papers, No. 5177.

RAFTERY, A.E. (1995) Bayesian model selection in social research. *Sociological Methodology*, **25**, 111–163.

RAMIREZ-COBO, P., LILLO, R.E., WILSON, S. and WIPER, M.P. (2010) Bayesian inference for double Pareto lognormal queues. *The Annals of Applied Statistics*, **4**(3), 1533–1557.

RIGBY, R.A. and STASINOPOULOS, D.M. (1996a) A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, **6**(1), 57–65.

RIGBY, R.A. and STASINOPOULOS, D.M. (1996b) Mean and dispersion additive models. In *Statistical Theory and Computational Aspects of Smoothing* (eds. W. Härdle and M.G. Schimek), pp. 215–230. Heidelberg: Physica.

RIGBY, R.A. and STASINOPOULOS, D.M. (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(3), 507–554.

STASINOPOULOS, D.M., RIGBY, B. and AKANTZILIOTOU, C. (2008) *Instructions on How to Use the Gamlss Package in R*, 2nd edn. http://www.gamlss.org.

ROSENBERG, M.A., FREES, E.W., SUN, J., JOHNSON, P.H. and ROBINSON, J.M. (2007) Predictive modeling with longitudinal data: A case study with Wisconsin nursing homes *North American Actuarial Journal*, **11**(3), 54–69.

SESHADRI, V. (2012) *The Inverse Gaussian Distribution: Statistical Theory and Applications*, Vol. 137. New York: Springer Science and Business Media.

STASINOPOULOS, M.D., RIGBY, R.A., HELLER, G.Z., VOUDOURIS, V. and DE BASTIANI, F. (2017) *Flexible Regression and Smoothing: Using GAMLSS in R*. Florida: Chapman and Hall/CRC.

TZOUGAS, G., VRONTOS, S. and FRANGOS, N. (2014) Optimal bonus-malus systems using finite mixture models. *ASTIN Bulletin*, **44**(2), 417–444.

TZOUGAS, G., VRONTOS, S. and FRANGOS, N. (2018) Bonus-malus systems with two component mixture models arising from different parametric families. *North American Actuarial Journal*, **22**(1), 55–91.

TZOUGAS, G., YIK, W.H. and MUSTAQEEM, M.W. (2019) Insurance ratemaking using the exponential-lognormal regression model. *Annals of Actuarial Science*, 1–30. doi: 10.1017/S1748499519000034.

WILLS, M., VALDEZ, E. and FREES, E. (2006) GB2 regression with insurance claim severities. *Conference Paper, UNSW Actuarial Research Symposium 2008, Sydney, New South Wales, Australia.*

WANG, S. (1998) An actuarial index of the right-tail risk. *North American Actuarial Journal*, **2**(2), 88–101.

GEORGE TZOUGAS (Corresponding author)
Department of Statistics
London School of Economics and Political Science
London WC2A 2AE, UK
E-mail: g.tzougas@lse.ac.uk

DIMITRIS KARLIS
Department of Statistics
Athens University of Economics and Business
Greece
E-mail: karlis@aueb.gr