Bonus-Malus Systems with Two Component Mixture Models Arising from Different Parametric Families

George Tzougas
Department of Statistics, London School of Economics, UK

Spyridon Vrontos
Department of Mathematical Sciences, University of Essex, UK

Nicholas Frangos
Department of Statistics, Athens University of Economics and Business, Greece

June 6, 2017

Abstract

Two component mixture distributions defined so that the component distributions do not necessarily arise from the same parametric family are employed for the construction of Optimal Bonus-malus Systems (BMS) with frequency and severity components. The proposed modelling framework is used for the first time in actuarial literature research and includes an abundance of alternative model choices to be considered by insurance companies when deciding on their Bonus-Malus pricing strategies. Furthermore, we advance one step further by assuming that all the parameters and mixing probabilities of the two component mixture distributions are modelled in terms of covariates, extending our previous work in Tzougas, Vrontos and Frangos (2014). Applying Bayes theorem we derive optimal BMS either by updating the posterior probability of the policyholders' classes of risk or by updating the posterior mean and the posterior variance. The resulting tailor-made premiums are calculated via the expected value and variance principles and are compared to those based only on the a posteriori criteria. The use of the variance principle in a Bonus-Malus ratemaking scheme in a way that takes into consideration both the number and the costs of claims based on both the a priori and the a posterior classification criteria has not yet been proposed and can alter the resulting premiums significantly, providing the actuary with useful alternative tariff structures.

Keywords: Optimal BMS; Claim frequency; Claim severity; Two component mixture regression models for location, scale, shape and prior probabilities; Expected value premium calculation principle; Variance premium calculation principle.

1 Introduction

Bonus-Malus Systems, BMS in short, are experience rating mechanisms which impose penalties on policyholders responsible for one or more accidents by premium surcharges (or maluses) and reward discounts (or bonuses) to policyholders who had a claim-free year. In view of the economic importance of motor third party liability (MTPL) insurance in developed countries a basic interest of recent actuarial literature research is their optimal design that takes into account both the number and the cost of claims reported by policyholders. Optimal BMS are defined as systems obtained through Bayesian analysis and are financially balanced for the insurer. For a detailed description of optimal BMS the interested reader can refer to the seminal work of Lemaire (1995). Further references for BMS include, among others, Picech (1994), Pinquet (1997, 1998) and Brouhns et al. (2003). Furthermore, the construction of such systems based on the inclusion of important a priori rating variables for the number and/or costs of claims plays a major role, see for example Dionne and Vanasse (1989, 1992), Denuit et al. (2007), Boucher, Denuit and Guillen (2008), Frangos and Vrontos (2001), Tzougas and Frangos (2014) and Tzougas, Vrontos and Frangos (2014). The aforementioned systems were constructed by assuming that the claim frequency and severity components are independent. Gómez et al. (2014) presented a BMS which takes into account of some kind of dependence between the two components by compounding the claim frequency and severity distributions in order to obtain the distribution of the aggregated losses.

The main contributions of the present study are the following: a) We present a new methodology for the design of optimal BMS which pioneers the allowance of both the number and costs of claims through the use of two component mixture distributions, without necessarily assuming that the component frequency/severity distributions arise from the same parametric family. In this respect, more flexible systems are designed to include a large number of alternative possible model choices, which enlarges substantially the pricing toolbox of general insurance companies. b) We extend the framework of our previous work in Tzougas, Vrontos and Frangos (2014) by assuming that all the parameters and mixing probabilities of the claim frequency/severity distributions can be modelled as functions of explanatory variables with parametric linear functional forms, enabling the actuary to fit more representative distributions of the data that capture all their important stylized characteristics. c) We propose the use of the variance principle, as an alternative to the expected value principle for calculating the premiums derived by BMS, in a way that incorporates all the important a priori information from the individual characteristics of the policyholders, both for the frequency and the severity components. This principle provides a more complete picture to the actuary since it takes into account an additional characteristic of the distribution, i.e. the variance of the number of claims and losses.

In what follows, we discuss in detail our motivation for proposing the aforementioned frameworks and comment on how these extend current BMS literature research. Regarding our first contribution, two component mixture models, which do not necessarily have all of their parameters in common, are considered for the first time in an actuarial context, and we suggest their employment for designing optimal BMS with frequency and severity components for the following academic and practical reasons. Firstly, with respect to the frequency component, this modelling framework allows for a rich, flexible and easily extensible family of claim frequency models instead of restricting attention to particular mixed Poisson laws that have been widely applied for the construction of optimal BMS. Secondly, regarding the severity component, it is common knowledge that in a competitive market an insurance company has to design tariff structures that will fairly distribute the burden of large and small claim sizes among policyholders. In other words, it is required that policyholders with large size claims or frequent smaller claims should pay higher premiums and vice versa. Otherwise, the bonus-hunger phenomenon may arise, i.e. the tendency of policyholders not to report low cost accidents to avoid premium surcharges. However, when dealing with real insurance data sets insurers tend to partition losses in their portfolios and innovate in designing new BMS because it is difficult to find a simple model that fits all claim sizes. Specifically, heavy-tailed distributions are used for modelling large size claims while those with a lighter tail are usually preferred for modelling small size claims. In this respect, a unified approach for providing alternative options to the insurer when they are deciding on their Bonus-Malus pricing strategies does not exist. Two component mixture models with no parameters in common is a very rational solution to this problem as they provide the actuary with an abundance of alternative convex combinations of heavy-tailed and light-tailed distributions which can generate tailor-made Bonus-Malus premiums that fairly punish more for large size claims and less for small size claims, alleviating the bonus hunger phenomenon. Furthermore, with respect to our second contribution, it should be noted that until now the commonly used specification for the design of optimal BMS was that only the mean frequency and/or severity is modelled as a function of risk factors. In this respect, any model for the mean in terms of a priori risk factors indirectly yields a model for scale, shape and prior (mixing) probabilities in the case of two component mixture models. Thus, even if the mean is the most commonly used measure of the expected claim frequency and expected claim severity it fails to describe the scale and shape parameters of a distributions as well as prior probabilities due to the unobserved heterogeneity changes with covariates. Consequently, this situation affects the construction of optimal BMS with frequency and severity components since the posterior frequency/severity distributions are used to calculate premiums. Joint modelling of all the parameters in an experience ratemaking scheme enables us to use all the available information in the estimation of the claim frequency/severity distribution in order to group risks with similar risk characteristics and establish fair Bonus-Malus premiums employing the expected value and variance principles. Moreover, using this formulation, the risk heterogeneity in the data is modelled as the distribution of frequency and/or severity of claims changes between and within two sub-populations in the following ways. Firstly, the population heterogeneity is accounted for by choosing two unobserved latent components, each of which may be regarded as a sub-population. This is a discrete representation of heterogeneity since the mean is approximated by two support points which are modelled in terms of a priori rating variables by using the multinomial logit link function. Secondly, depending on the choice of the component frequency/severity distribution, heterogeneity can also be accommodated within each component through the use of known monotonic link functions chosen to ensure a valid range for the distribution parameters, see Rigby and Stasinopoulos (2005 and 2009). Specifically, in this paper, for the frequency component we assume that the number of claims is distributed according to a two component (2C) Poisson mixture, 2C Negative Binomial mixture, 2C Sichel mixture (and 2C Poisson Inverse Gaussian mixture and 2C Sichel-Poisson Inverse Gaussian mixture as special cases), 2C Poisson-Negative Binomial mixture (i.e., in this case, the first component follows the Poisson distribution and the second component follows the Negative Binomial distribution), 2C Poisson-Sichel mixture (2C Poisson-Poisson Inverse Gaussian mixture as a special case) and 2C Negative Binomial-Sichel mixture (2C Negative Binomial-Poisson Inverse Gaussian mixture as a special case) distributions. For the severity component, we consider that the losses are distributed according to a 2C Exponential mixture, 2C Pareto mixture, 2C Lognormal mixture, 2C Exponential-Pareto mixture (i.e., in this case, the first component follows the Exponential distribution and the second component follows the Pareto distribution), 2C Exponential-Lognormal mixture and 2C Lognormal-Pareto mixture distributions. Also, the Negative Binomial, Sichel, Poisson-Inverse Gaussian and Pareto distributions are considered as special cases of the aforementioned distributions. Within the adopted framework all the parameters and mixing probabilities of these distributions are modelled in terms of covariates. Applying Bayes theorem, we derive optimal BMS either by updating the posterior probability of the policyholders' classes of risk or by updating the posterior mean and the posterior variance. The aforementioned models are compared on the basis of a sample of the automobile portfolio of a major insurance company employing the Generalized Akaike Information Criterion (GAIC), which is valid for both nested or non-nested model comparisons (as suggested by Rigby and Stasinopoulos, 2005 and 2009). Finally, regarding our third contribution, it should be mentioned that traditionally the expected value principle was used with BMS by the majority of authors, while the variance principle was recommended by, for example, Lemaire (1995), Heilmann (1989) and Gómez et al. (2000 and 2002) in the construction of BMS with a frequency component based only on the a posteriori criteria. However, the latter principle, as mentioned in Gómez et al. (2002), is much more robust than the expected value principle when BMS is used. Furthermore, this is the first time the variance principle is used with BMS with frequency and severity components that integrate a priori information, thus our work expands on this setup also. The variance principle is more applicable for an insurance company which would like to adopt a more conservative pricing profile in cases where this is considered necessary. Overall, in the generalized systems we propose, the premiums calculated by either principle are functions of the years that the policyholder is in the portfolio, the number and costs of accidents and all the available information for the policyholder and the automobile taken into consideration by assuming that every parameter of the response frequency/severity distribution as well as the mixing probabilities are modelled in terms of covariates.

The rest of this paper proceeds as follows. Section 2 introduces the alternative models we employ for modelling claim frequency and severity. Section 3 presents the optimal BMS derived by updating the posterior probabilities and those determined by updating the posterior mean and the posterior variance. Section 4 contains an application to a data set concerning car-insurance claims at fault. Finally, Section 5 concludes the paper.

2 Two Component Mixture Regression Models for Location, Scale, Shape and Prior Probabilities

This section summarizes the characteristics of the alternative models used in this study for assessing claim frequency and severity respectively. In what follows, each model will be given by a convex combination of two frequency and/or severity distributions where each will be referred to as frequency and/or severity component distributions defined so they do not necessarily have their parameters in common.

2.1 Claim Frequency Models

Suppose that the portfolio is considered to be heterogeneous, consisting of two homogeneous subpopulations. In this respect, we have two fractions of drivers π_z, $z = 1, 2$, and the probability that a policyholder has reported k claims to the insurer, $k = 0, 1, 2, ...$, in each category is denoted by $P_z(k)$. Henceforth, $P_z(k)$ will be referred to as frequency component distributions. Thus, the structure function is a 2-point discrete distribution and the unconditional distribution of the number of claims, denoted by $P(k)$, is given by

$$P(k) = \sum_{z=1}^{2} \pi_z P_z(k), \tag{1}$$

for $k = 0, 1, 2, 3, ..., \pi_z > 0$, for $z = 1, 2$, and $\sum_{z=1}^{2} \pi_z = 1$. Let us denote by $E_z(k)$ and $Var_z(k)$ the mean and the variance of the component frequency distributions. The expected value of the number of claims is equal to $E(k) = \sum_{z=1}^{2} \pi_z E_z(k)$ and its variance is equal to $Var(k) = \sum_{z=1}^{2} \pi_z Var_z(k) + \pi_1 \pi_2 [E_1(k) - E_2(k)]^2$. Furthermore, it is assumed that the component distributions $P_z(k)$ belong to a family of mixed Poisson models defined so that $E_z(k) = \lambda_z$, where $\lambda_z > 0$, $z = 1, 2$, is an explicit parameter of them. Thus, we have that mean and the variance of Eq.(1) are simplified to $E(k) = \sum_{z=1}^{2} \pi_z \lambda_z$, is common for all the alternative models, and $Var(k) = \sum_{z=1}^{2} \pi_z Var_z(k) + \pi_1 \pi_2 (\lambda_1 - \lambda_2)^2$. In this respect, in what follows, we only report the probability density functions (pdf's) of the component distributions, i.e. $P_z(K_i = k)$, and the variances, $Var_z(k)$ for $z = 1, 2$ for each of the two component mixture models we consider for modelling the number of claims.

- In the case of the 2C Poisson mixture distribution we have that

$$P_z(k) = \frac{e^{-\lambda_z} \lambda_z^k}{k!}, z = 1, 2. \tag{2}$$

  The variance of the Poisson component distributions is given by

$$Var_z(k) = \lambda_z, z = 1, 2. \tag{3}$$

- In the case of the 2C Negative Binomial Type I[1] (NBI) mixture distribution we have that

$$P_z(k) = \binom{k + \frac{1}{\sigma_z} - 1}{k} \left(\frac{\sigma_z \lambda_z}{1 + \sigma_z \lambda_z} \right)^k \left(\frac{1}{1 + \sigma_z \lambda_z} \right)^{\frac{1}{\sigma_z}}, \sigma_z > 0, z = 1, 2. \tag{4}$$

  The variance of the Negative Binomial Type I component distributions is given by

$$Var_z(k) = \lambda_z + \lambda_z^2 \sigma_z, z = 1, 2. \tag{5}$$

- In the case of the 2C Sichel[2] mixture distribution we have that

$$P_z(k) = \frac{\left(\frac{\lambda_z}{c_z}\right)^k B_{k+\nu_z}(a_z)}{k! (a_z \sigma_z)^{k+\nu_z} B_{\nu_z}\left(\frac{1}{\sigma_z}\right)}, \tag{6}$$

[1] We use the parameterization of Negative Binomial Type I given by Johnson et al. (2005) and Rigby and Stasinopoulos (2009).

[2] The construction of optimal BMS based on the use of the Sichel distribution for modelling claim frequency where regression is only performed on the mean parameter has been recommended by Tzougas and Frangos (2014).

$z=1,2$, where $\sigma_z>0$ and $-\infty<\nu_z<\infty$, with $a_z^2=\sigma_z^{-2}+2\lambda_z\left(c_z\sigma_z\right)^{-1}$ and where $c_z=\frac{B_{\nu_z+1}\left(\frac{1}{\sigma_z}\right)}{B_{\nu_z}\left(\frac{1}{\sigma_z}\right)}$, where

$$B_{\nu_z}\left(\omega\right)=\frac{1}{2}\int_0^\infty x^{\nu-1}\exp\left[-\frac{1}{2}\omega\left(x+\frac{1}{x}\right)\right]dx, \tag{7}$$

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

- The variance of the Sichel component distributions is given by

$$Var_z\left(k\right)=\lambda_z+\lambda_z^2\left(\frac{2\sigma_z\left(\nu_z+1\right)}{c_z}+\frac{1}{c_z^2}-1\right), z=1,2. \tag{8}$$

- In the case of the 2C Poisson Inverse Gaussian (PIG) mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(6 and 8) if we let $\nu_z=-0.5$ for $z=1,2$ respectively.
- In the case of the 2C Poisson-Negative Binomial Type I mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(2, 4, 3 and 5) for $z=1$ and $z=2$ respectively.
- In the case of the 2C Poisson-Sichel mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(2, 6, 3 and 8) for $z=1$ and $z=2$ respectively.
- In the case of the 2C Poisson-Poisson Inverse Gaussian mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(2, 6, 3 and 8) for $z=1$ and $z=2$ when $\nu_2=-0.5$ respectively.
- In the case of the 2C Negative Binomial Type I-Sichel mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(4, 6, 5 and 8) for $z=1$ and $z=2$ respectively.
- In the case of the 2C Negative Binomial Type I-Poisson Inverse Gaussian mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(4, 6, 5 and 8) for $z=1$ and $z=2$ when $\nu_2=-0.5$ respectively.
- In the case of the 2C Poisson Inverse Gaussian-Sichel mixture distribution we have that $P_z\left(K_i=k\right)$ and $Var_z\left(k\right)$ are given by Eqs(6 and 8) for $\nu_1=-0.5$ and $z=1,2$ respectively.

2.2 Claim Severity Models

In this section we need to address the severity component. The portfolio is considered to be heterogeneous, consisting of two fractions of drivers ρ_z, $z=1,2$, and the pdf of the claim size x in each category is denoted by $f_z\left(x\right)$. In what follows $f_z\left(x\right)$ will be known as the severity component distributions. Thus, the structure function is a 2-point discrete distribution and the unconditional distribution of claim size, denoted by $f\left(x\right)$, is given by

$$f\left(x\right)=\sum_{z=1}^{2}\rho_z f_z\left(x\right), \tag{9}$$

for $x,\rho_z>0$ and $\sum_{z=1}^{2}\rho_z=1$. Let $E_z\left(x\right)$ and $Var_z\left(x\right)$ represent the mean and the variance of the severity component distributions. The expected value of the claim size is equal to $E\left(x\right)=\sum_{z=1}^{2}\rho_z E_z\left(x\right)$ and its variance is equal to $Var(x)=\sum_{z=1}^{2}\rho_z Var_z\left(x\right)+\rho_1\rho_2\left[E_1\left(x\right)-E_2\left(x\right)\right]^2$. In what follows, we present the probability density functions (pdf's) of the component distributions, i.e. $f_z\left(x\right)$, and the variances, $Var_z\left(x\right)$ for $z=1,2$ for each of the models we consider for approximating claim severity.

- In the case of the 2C mixture Exponential distribution we have that

$$f_z\left(x\right)=\frac{e^{-\frac{x}{y_z}}}{y_z}, y_z>0, z=1,2. \tag{10}$$

- The mean and the variance of the Exponential component distributions are given by

$$E_z(x) = y_z \text{ and } Var_z(x) = y_z^2, z = 1, 2. \tag{11}$$

- In the case of the 2C Lognormal distribution we have that

$$f_z(x) = \frac{1}{\sqrt{2\pi s_z^2}} \frac{1}{x} e^{\left\{-\frac{[\log(x)-y_z]^2}{2s_z^2}\right\}}, y_z > 0, s_z > 0, z = 1, 2. \tag{12}$$

The mean and the variance of the Lognormal component distributions are given by

$$E_z(x) = \sqrt{e^{s_z^2}} e^{y_z} \text{ and } Var_z(x) = e^{s_z^2}\left(e^{s_z^2} - 1\right) e^{2y_z}, z = 1, 2. \tag{13}$$

- In the case of the 2C mixture Pareto distribution we have that

$$f_z(x) = s_z \frac{[(s_z - 1) y_z]^{s_z}}{[x + (s_z - 1) y_z]^{s_z+1}}, y_z > 0, s_z > 2, z = 1, 2. \tag{14}$$

The mean and the variance of the Pareto component distributions are given by

$$E_z(x) = y_z \text{ and } Var_z(x) = \frac{[(s_z - 1) y_z]^2}{s_z - 1}\left(\frac{2}{s_z - 2} - \frac{1}{s_z - 1}\right), z = 1, 2. \tag{15}$$

- In the case of the 2C Exponential-Lognormal mixture distribution we have that $f_z(x)$, $E_z(x)$ and $Var_z(k)$ are given by Eqs(10, 12, 11 and 13) for $z = 1$ and $z = 2$ respectively.
- In the case of the 2C Exponential-Pareto mixture distribution we have that $f_z(x)$, $E_z(x)$ and $Var_z(k)$ are given by Eqs(10, 14, 11 and 15) for $z = 1$ and $z = 2$ respectively.
- In the case of the 2C Lognormal-Pareto mixture distribution we have that $f_z(x)$, $E_z(x)$ and $Var_z(k)$ are given by Eqs(12, 14, 13 and 15) for $z = 1$ and $z = 2$ respectively.

3 An Optimal Bonus-Malus System

It is assumed that the number of claims of each policyholder is independent from the severity of each claim in order to deal with the frequency and severity components separately. The framework we develop for both the claim frequency and the severity components is a generalization of the good risk/bad risk model proposed by Lemaire (1995) and our previous work in Tzougas, Vrontos and Frangos (2014).

3.1 The Optimal Bonus-Malus System Derived by Updating the Posterior Probability

3.1.1 Frequency Component

Consider a policyholder i with $K_i^1, ..., K_i^t$ claim history for $i = 1, ..., n$. Also, denote as $K = \sum_{j=1}^{t} K_i^j$ the total number of claims that they had, where K_i^j is the number of claims of this individual in period j. Following the framework of Rigby and Stasinopoulos (2005, 2009) we can model the parameters and mixing probabilities of the claim frequency distributions presented in Section 2.1 as

$$\begin{aligned}
\lambda_{z,i}^j &= \exp\left(c_{1z,i}^j \beta_{1z}^j\right), && (16)\\
\sigma_{z,i}^j &= \exp\left(c_{2z,i}^j \beta_{2z}^j\right), && (17)\\
\nu_{z,i}^j &= c_{3z,i}^j \beta_{3z}^j \text{ and} && (18)\\
\pi_{z,i}^j &= \frac{\exp\left(c_{4z,i}^j \beta_{4z}^j\right)}{1 + \exp\left(c_{4z,i}^j \beta_{4z}^j\right)}, && (19)
\end{aligned}$$

where $c^j_{\xi z,i}\left(c^j_{\xi z,i,1},...,c^j_{\xi z,i,\xi'_\xi}\right)$ are covariate vectors of individual characteristics[3] of length $1\times\phi_\xi$ and $\beta^{jT}_\xi\left(\beta^j_{\xi z,1},...,\beta^j_{\xi z,\xi'_\xi}\right)$ are the corresponding parameter vectors of length $1\times\phi_\xi$, where $\xi=1,2,3,4$, where $i=1,...,n$ and where $z=1,2$.

Let us denote with R_2 the risk, imposed on the insurance company, associated with the second category of policyholders. Moreover, the posterior probability of the policyholder i belonging to the second category is denoted by $\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)$ for $\xi=1,2,3,4$. Applying Bayes theorem, the posterior probability of the individual i belonging to the second category is given by

$$\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=\frac{P(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}|R_2)\pi^j_{2,i}}{\sum_{z=1}^{2}P(K^1_i,...,K^t_i;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}|R_z)\pi^j_{z,i}}. \quad (20)$$

Also, $\pi_1\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=1-\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)$. The setup we described previously is applied to the models presented in Section 2.1.

- In the case of the 2C Poisson mixture distribution Eq.(20) becomes

$$\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=\frac{\left(\lambda^j_{2,i}\right)^K e^{-t\lambda^j_{2,i}}\pi^j_{2,i}}{\sum_{z=1}^{2}\left(\lambda^j_{z,i}\right)^K e^{-t\left(\lambda^j_{z,i}\right)^K}\pi^j_{z,i}}. \quad (21)$$

- In the case of the 2C Negative Binomial Type I mixture distribution Eq.(20) becomes

$$\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=\frac{\prod_{j=1}^{t}\binom{K^j_i+\frac{1}{\sigma^j_{2,i}}-1}{K^j_i}\left(\frac{1}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^{\frac{t}{\sigma^j_{2,i}}}\left(\frac{\sigma^j_{2,i}\lambda^j_{2,i}}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^K\pi^j_{2,i}}{\sum_{z=1}^{2}\prod_{j=1}^{t}\binom{K^j_i+\frac{1}{\sigma^j_{z,i}}-1}{K^j_i}\left(\frac{1}{1+\sigma^j_{z,i}\lambda^j_{z,i}}\right)^{\frac{t}{\sigma^j_{z,i}}}\left(\frac{\sigma^j_{z,i}\lambda^j_{z,i}}{1+\sigma^j_{z,i}\lambda^j_{z,i}}\right)^K\pi^j_{z,i}}. \quad (22)$$

- In the case of the 2C Sichel mixture distribution Eq.(20) becomes

$$\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=\frac{\frac{\left(\frac{\lambda^j_{2,i}}{c^j_{2,i}}\right)^K\prod_{j=1}^{t}B_{K^j_i+\nu^j_{2,i}}\left(a^j_{2,i}\right)^t}{\left(a^j_{2,i}\sigma^j_{2,i}\right)^{K+t\nu^j_{2,i}}\left[B_{\nu^j_{2,i}}\left(\frac{1}{\sigma^j_{2,i}}\right)\right]^t}\pi^j_{2,i}}{\sum_{z=1}^{2}\frac{\left(\frac{\lambda^j_{z,i}}{c^j_{z,i}}\right)^K\prod_{j=1}^{t}B_{K^j_i+\nu^j_{z,i}}\left(a^j_{z,i}\right)^t}{\left(a^j_{z,i}\sigma^j_{z,i}\right)^{K+t\nu^j_{z,i}}\left[B_{\nu^j_{z,i}}\left(\frac{1}{\sigma^j_{z,i}}\right)\right]^t}\pi^j_{z,i}}, \quad (23)$$

where $\left(a^j_{z,i}\right)^2=\left(\sigma^j_{z,i}\right)^{-2}+2\lambda^j_{2,i}\left(c^j_{z,i}\sigma^j_{z,i}\right)^{-1}$ and where $c^j_{z,i}=\frac{B_{\nu^j_{z,i}+1}\left(\frac{1}{\sigma^j_{z,i}}\right)}{B_{\nu^j_{z,i}}\left(\frac{1}{\sigma^j_{z,i}}\right)}$ for $z=1,2$.

- In the case of the 2C Poisson-Negative Binomial Type I mixture distribution Eq.(20) becomes

$$\pi_2\left(K^1_i,...,K^t_i;c^1_{\xi 2,i},...,c^{t+1}_{\xi 2,i}\right)=$$
$$\frac{\prod_{j=1}^{t}\binom{K^j_i+\frac{1}{\sigma^j_{2,i}}-1}{K^j_i}\left(\frac{1}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^{\frac{t}{\sigma^j_{2,i}}}\left(\frac{\sigma^j_{2,i}\lambda^j_{2,i}}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^K\pi^j_{2,i}}{\left(\lambda^j_{1,i}\right)^K e^{-t\lambda^j_{1,i}}\pi^j_{1,i}+\prod_{j=1}^{t}\binom{K^j_i+\frac{1}{\sigma^j_{2,i}}-1}{K^j_i}\left(\frac{1}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^{\frac{t}{\sigma^j_{2,i}}}\left(\frac{\sigma^j_{2,i}\lambda^j_{2,i}}{1+\sigma^j_{2,i}\lambda^j_{2,i}}\right)^K\pi^j_{2,i}}. \quad (24)$$

[3] All the characteristics we consider are observable.

- In the case of the 2C Poisson-Sichel mixture distribution Eq.(20) becomes

$$\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)=$$

$$\frac{\dfrac{\left(\frac{\lambda_{2,i}^j}{c_{2,i}^j}\right)^K \prod_{j=1}^{t} B_{K_i^j+\nu_{2,i}^j}\left(a_{2,i}^j\right)^t}{\left(a_{2,i}^j\sigma_{2,i}^j\right)^{K+t\nu_{2,i}^j}\left[B_{\nu_{2,i}^j}\left(\frac{1}{\sigma_{2,i}^j}\right)\right]^t}\pi_{2,i}^j}{\left(\lambda_{1,i}^j\right)^K e^{-t\lambda_{1,i}^j}\pi_{1,i}^j+\dfrac{\left(\frac{\lambda_{2,i}^j}{c_{2,i}^j}\right)^K \prod_{j=1}^{t} B_{K_i^j+\nu_{2,i}^j}\left(a_{2,i}^j\right)^t}{\left(a_{2,i}^j\sigma_{2,i}^j\right)^{K+t\nu_{2,i}^j}\left[B_{\nu_{2,i}^j}\left(\frac{1}{\sigma_{2,i}^j}\right)\right]^t}\pi_{2,i}^j}. \tag{25}$$

- In the case of the 2C Negative Binomial Type I-Sichel mixture distribution Eq.(20) becomes

$$\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)=$$

$$\frac{\dfrac{\left(\frac{\lambda_{2,i}^j}{c_{2,i}^j}\right)^K \prod_{j=1}^{t} B_{K_i^j+\nu_{2,i}^j}\left(a_{2,i}^j\right)^t}{\left(a_{2,i}^j\sigma_{2,i}^j\right)^{K+t\nu_{2,i}^j}\left[B_{\nu_{2,i}^j}\left(\frac{1}{\sigma_{2,i}^j}\right)\right]^t}\pi_{2,i}^j}{\prod_{j=1}^{t}\binom{K_i^j+\frac{1}{\sigma_{1,i}^j}-1}{K_i^j}\left(\frac{1}{1+\sigma_{1,i}^j\lambda_{1,i}^j}\right)^{\frac{t}{\sigma_{1,i}^j}}\left(\frac{\sigma_{1,i}^j\lambda_{1,i}^j}{1+\sigma_{1,i}^j\lambda_{1,i}^j}\right)^K\pi_{1,i}^j+\dfrac{\left(\frac{\lambda_{2,i}^j}{c_{2,i}^j}\right)^K \prod_{j=1}^{t} B_{K_i^j+\nu_{2,i}^j}\left(a_{2,i}^j\right)^t}{\left(a_{2,i}^j\sigma_{2,i}^j\right)^{K+t\nu_{2,i}^j}\left[B_{\nu_{2,i}^j}\left(\frac{1}{\sigma_{2,i}^j}\right)\right]^t}\pi_{2,i}^j}. \tag{26}$$

- In the case of the 2C Poisson Inverse Gaussian mixture distribution $\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)$ is given by Eq.(23) if we let $\nu_{z,i}^j=-0.5$ for $z=1,2$ respectively.
- In the case of the 2C Poisson-Poisson Inverse Gaussian mixture distribution $\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)$ is given by Eq.(25) if we let $\nu_{2,i}^j=-0.5$.
- In the case of the 2C Negative Binomial Type I-Poisson Inverse Gaussian mixture distribution $\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)$ is given by Eq.(26) if we let $\nu_{2,i}^j=-0.5$.
- In the case of the 2C Poisson–Inverse Gaussian-Sichel mixture distribution $\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)$ is given by Eq.(23) if we let $\nu_{1,i}^j=-0.5$.

Note that due to the existence of K_i^j in Eqs(22, 23, 24, 25 and 26), the explicit claim frequency history determines the calculation of the posterior probabilities and thus of premium rates to be calculated with the expected value and variance principles and not just the total number of claims as in the case of the 2C Poisson mixture.

Calculation of the Premiums According to the Expected Value and Variance Principles Under a quadratic error loss function, the optimal estimate of λ_i^{t+1},the mean claim frequency of the individual i at $t+1$, is the mean of the posterior structure function given by

$$E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)=\sum_{z=1}^{n}\pi_z\left(K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)\lambda_{z,i}^j \tag{27}$$

and the variance of the posterior structure function is given by

$$
\begin{aligned}
& Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right) \\
= & \sum_{z=1}^{2}\pi_z\left(K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)Var_z\left(K_i^j\right)+ \\
& \pi_1\left(K_i^1,...,K_i^t;c_{\xi 1,i}^1,...,c_{\xi 1,i}^{t+1}\right)\pi_2\left(K_i^1,...,K_i^t;c_{\xi 2,i}^1,...,c_{\xi 2,i}^{t+1}\right)\left[\lambda_{1,i}^j-\lambda_{2,i}^j\right]^2. \quad (28)
\end{aligned}
$$

The premium rates calculated according to the expected value principle are given by

$$
P_1=(1+w_1)\,E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right), \quad (29)
$$

where $w_1>0$ is a risk load.

The premium rates calculated according to the variance principle are given by

$$
P_2=E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)+w_2Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right), \quad (30)
$$

where $w_2>0$ is a risk load.

Note that the premium rates calculated according to the expected value and variance premium principles based only on the a posteriori criteria are obtained if the regression components are limited to constants.

3.1.2 Severity Component

Similarly to the case of the frequency component, we assume that a policyholder stays in the portfolio for t years, the number of claims in year j is denoted by $K_i^j=k$. Denote by $X_{i,k}^j$ the loss incurred from their claim k for the period j. Then, the information we have for their claim size history will be in the form of a vector $X_{i,1}^1,...,X_{i,k}^t$ and the total claim amount will be equal to $\sum_{k=1}^{K}X_{i,k}^j$. Following the framework of Rigby and Stasinopoulos (2005, 2009), we can model the parameters and mixing proportions of the 2C Exponential, 2C Pareto and 2C Exponential-Pareto mixture models as

$$
\begin{aligned}
y_{z,i}^j &= \exp\left(d_{1z,i}^j\gamma_{1z}^j\right), && (31)\\
s_{z,i}^j &= \exp\left(d_{2z,i}^j\gamma_{2z}^j\right), && (32)\\
\rho_{z,i}^j &= \frac{\exp\left(d_{3z,i}^j\gamma_{3z}^j\right)}{1+\exp\left(d_{3z,i}^j\gamma_{3z}^j\right)}, && (33)
\end{aligned}
$$

while in the case when one or both of the component distributions is the Lognormal, i.e. in the case of the 2C Lognormal mixture, 2C Exponential-Lognormal mixture and 2C Pareto-Lognormal mixture models, we can model the location parameter as

$$
y_{z,i}^j=\exp d_{1z,i}^j\gamma_{1z}^j, \quad (34)
$$

where the scale parameters and mixing probabilities are again given by Eqs(32 and 33) and where $d_{\xi z,i}^j\left(d_{\xi z,i,1}^j,...,d_{\xi z,i,\xi_\xi^{/}}^j\right)$ are covariate vectors of individual characteristics[4] of length $1\times\phi_\xi$, where $\gamma_\xi^{jT}\left(\gamma_{\xi z,1}^j,...,\gamma_{\xi z,\xi_\xi^{/}}^j\right)$ are the corresponding parameter vectors of length $1\times\phi_\xi$, where $\xi=1,2,3$ and where $i=1,...,n$ and $z=1,2$.

Let us denote as Q_2 the risk that it is imposed on the insurance company if we assume that a policyholder i belongs to the second category of drivers based on the severity of their claims. Moreover, the posterior probability of the policyholder i belonging to the second category is denoted by $\rho_2\left(X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$ for $\xi=1,2,3$. Applying Bayes theorem, the posterior probability of the individual i belonging to the second category is given by

[4] All the characteristics we consider are observable.

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)=\frac{f\left(X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}|Q_2\right)\rho_{2,i}^j}{\sum_{z=1}^{2}f\left(X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi z,i}^1,...,d_{\xi z,i}^{t+1}|Q_z\right)\rho_{z,i}^j}. \tag{35}$$

Also, $\rho_2\left(X_{i,1}^1,...,X_{i,K}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)=1-\rho_1\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$. The setup we described above is applied to the models presented in Section 2.2.

- In the case of the 2C Exponential mixture distribution Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$
$$=\frac{e^{-\frac{\sum_{k=1}^{K}X_{i,k}^j}{y_{2,i}^j}}}{\left(y_{2,i}^j\right)^K}\rho_{2,i}^j\left\{\sum_{z=1}^{n}\frac{e^{-\frac{\sum_{k=1}^{K}X_{i,k}^j}{y_{z,i}^j}}}{\left(y_{z,i}^j\right)^K}\rho_{z,i}^j\right\}^{-1}. \tag{36}$$

.

- In the case of the 2C Lognormal mixture distribution Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$
$$=\frac{\left[\frac{1}{\sqrt{2\pi\left(s_{2,i}^j\right)^2}}\right]^K\prod_{j=1}^{K}\frac{1}{X_{i,k}^j}e^{\left\{-\frac{\sum_{k=1}^{K}\left[\log(X_{i,k}^j)-y_{2,i}^j\right]^2}{2\left(s_{2,i}^j\right)^2}\right\}}\rho_{2,i}^j}{\sum_{z=1}^{2}\left[\frac{1}{\sqrt{2\pi\left(s_{z,i}^j\right)^2}}\right]^K\prod_{j=1}^{K}\frac{1}{X_{i,k}^j}e^{\left\{-\frac{\sum_{k=1}^{K}\left[\log(X_{i,k}^j)-y_{z,i}^j\right]^2}{2\left(s_{z,i}^j\right)^2}\right\}}\rho_{z,i}^j}. \tag{37}$$

- In the case of the 2C Pareto mixture distribution Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$
$$=\frac{\left(s_{2,i}^j\right)^K\left\{\left[\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j}\right\}^K}{\prod_{j=1}^{K}\left[X_{i,k}^j+\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j+1}}\rho_{2,i}^j\left\{\sum_{z=1}^{2}\frac{\left(s_{z,i}^j\right)^K\left\{\left[\left(s_{z,i}^j-1\right)y_z\right]^{s_{z,i}^j}\right\}^K}{\prod_{j=1}^{K}\left[X_{i,k}^j+\left(s_{z,i}^j-1\right)y_z\right]^{s_{z,i}^j+1}}\rho_{z,i}^j\right\}^{-1}. \tag{38}$$

- In the case of the 2C mixture of Exponential-Lognormal Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$

$$= \frac{\left[\frac{1}{\sqrt{2\pi\left(s_{2,i}^j\right)^2}}\right]^K \prod_{j=1}^K \frac{1}{X_{i,k}^j} e^{\left\{-\frac{\sum_{k=1}^K\left[\log(X_{i,k}^j)-y_{2,i}^j\right]^2}{2\left(s_{2,i}^j\right)^2}\right\}} \rho_{2,i}^j}{\frac{e^{-\frac{\sum_{k=1}^K X_{i,k}^j}{y_{1,i}^j}}}{\left(y_{1,i}^j\right)^K}\rho_{1,i}^j + \left[\frac{1}{\sqrt{2\pi\left(s_{2,i}^j\right)^2}}\right]^K \prod_{j=1}^K \frac{1}{X_{i,k}^j} e^{\left\{-\frac{\sum_{k=1}^K\left[\log(X_{i,k}^j)-y_{2,i}^j\right]^2}{2\left(s_{2,i}^j\right)^2}\right\}} \rho_{2,i}^j}. \tag{39}$$

- In the case of the 2C mixture of Exponential-Pareto Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$

$$= \frac{\frac{\left(s_{2,i}^j\right)^K\left\{\left[\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j}\right\}^K}{\prod_{j=1}^K\left[X_{i,k}^j+\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j+1}}\rho_{2,i}^j}{\frac{e^{-\frac{\sum_{k=1}^K X_{i,k}^j}{y_{1,i}^j}}}{\left(y_{1,i}^j\right)^K}\rho_{1,i}^j + \frac{\left(s_{2,i}^j\right)^K\left\{\left[\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j}\right\}^K}{\prod_{j=1}^K\left[X_{i,k}^j+\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j+1}}\rho_{2,i}^j}. \tag{40}$$

- In the case of the 2C mixture of Lognormal-Pareto Eq.(35) becomes

$$\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)$$

$$= \frac{\frac{\left(s_{2,i}^j\right)^K\left\{\left[\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j}\right\}^K}{\prod_{j=1}^K\left[X_{i,k}^j+\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j+1}}\rho_{2,i}^j}{\left[\frac{1}{\sqrt{2\pi\left(s_{1,i}^j\right)^2}}\right]^K \prod_{j=1}^K \frac{1}{X_{i,k}^j} e^{\left\{-\frac{\sum_{k=1}^K\left[\log(X_{i,k}^j)-y_{1,i}^j\right]^2}{2\left(s_{1,i}^j\right)^2}\right\}} \rho_{1,i}^j + \frac{\left(s_{2,i}^j\right)^K\left\{\left[\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j}\right\}^K}{\prod_{j=1}^K\left[X_{i,k}^j+\left(s_{2,i}^j-1\right)y_2\right]^{s_{2,i}^j+1}}\rho_{2,i}^j}. \tag{41}$$

Calculation of the Premiums According to the Expected Value and Variance Principles Using a quadratic error loss function, the optimal estimate of y_i^{t+1},the mean claim severity of the individual i at $t+1$, is the mean of the posterior structure function given by

$$E\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right) = \sum_{z=1}^n \rho_z\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right) y_{z,i}^j \tag{42}$$

and the variance of the posterior structure function is given by

$$\begin{aligned}
&Var\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\\
= &\sum_{z=1}^2 \rho_z\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right) Var_z\left(y_i^{t+1}\right) +\\
&\rho_1\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\rho_2\left(X_{i,1}^1,...,X_{i,K_i^t}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\left[E(y_{1,i}^j)-E(y_{2,i}^j)\right]^2.
\end{aligned} \tag{43}$$

The premium rates calculated according to the expected value principle are given by

$$P_1 = (1+\omega_1)\, E\left(y_i^{t+1}|X_{i,1}^1, ..., X_{i,K_i^j}^t; d_{\xi 2,i}^1, ..., d_{\xi 2,i}^{t+1}\right), \tag{44}$$

where $\omega_1 > 0$ is a risk load.

The premium rates calculated according to the variance principle are given by

$$P_2 = E\left(y_i^{t+1}|X_{i,1}^1, ..., X_{i,K_i^j}^t; d_{\xi 2,i}^1, ..., d_{\xi 2,i}^{t+1}\right) + \omega_2 Var\left(y_i^{t+1}|X_{i,1}^1, ..., X_{i,K_i^j}^t; d_{\xi 2,i}^1, ..., d_{\xi 2,i}^{t+1}\right), \tag{45}$$

where $\omega_2 > 0$ is a risk load.

The premium rates calculated according to these principles based only on the a posteriori criteria are obtained if the regression components are limited to constants.

3.2 The Optimal Bonus-Malus System Derived by Updating the Posterior Mean and the Posterior Variance

3.2.1 Frequency Component

Assume that given a continuous random variable $u > 0$ with probability density function $\upsilon(u)$ defined on $\mathcal{R}^+$, K_i^j follows the Poisson distribution with parameter λu, where $\lambda > 0$. Then, the marginal distribution of K_i^j is a mixed Poisson distribution. The following are well-known results applied to the above situation (see, for example, Dionne and Vanasse, 1989 and 1992, Lemaire, 1995, and Boucher et al., 2007, 2008). We consider that $E(u) = 1$. Depending on the chosen parametric form of u, the mixed Poisson distribution will lead to different distributions. In what follows we consider the optimal BMS derived by updating the posterior mean and the posterior variance in the case of the 2C Negative Binomial Type I mixture, 2C Sichel mixture and 2C Negative Binomial Type I-Sichel mixture models. Note that the systems determined by the 2C Poisson Inverse Gaussian mixture, 2C Sichel-Poisson Inverse Gaussian mixture, 2C Negative Binomial-Poisson Inverse Gaussian mixture, Negative Binomial, Sichel and Poisson Inverse Gaussian models can be obtained as special cases of those for the case of the aforementioned models.

- Let u follow a 2C Gamma mixture distribution with pdf

$$\upsilon(u) = \sum_{z=1}^{2} \pi_z \frac{u^{\frac{1}{\sigma_z}-1} \frac{1}{\sigma_z}^{\frac{1}{\sigma_z}} \exp\left(-\frac{1}{\sigma_z}u\right)}{\Gamma\left(\frac{1}{\sigma_z}\right)},$$

for $z = 1, 2, \sum_{z=1}^{2} \pi_z = 1$, where $\sigma_z > 0$. Under this assumption the unconditional distribution of K_i^j becomes a 2C Negative Binomial Type I mixture distribution, where the frequency component distributions, $P_z(K_i = k)$, are given by Eq.(4) for $z = 1, 2$. We can allow the parameters and the mixing probabilities of this model to vary from one individual to another. Let λ_i^j, $\sigma_{z,i}^j$ and $\pi_{z,i}^j$ be given by Eqs(16, 17 and 19). Then, the posterior distribution of $\lambda_i^{t+1}$is obtained by employing a fully Bayesian approach (i.e. by updating both the parameters and the mixing proportions of the mixing distribution) and is given by a 2C Gamma mixture with updated parameters $w_{1,z,i}^j = \frac{1}{\sigma_{z,i}^j} + K$ and $w_{2,z,i}^j = \frac{\frac{1}{\sigma_{z,i}^j} + \sum_{j=1}^{t} \lambda_{z,i}^j}{\lambda_{z,i}^j}$, for $z = 1, 2$, and updated mixing probabilities $\acute{\pi}_{z,i}^j = \pi_{z,i}^j \frac{P_z\left(K; \lambda_i^j; \sigma_{z,i}^j\right)}{\sum_{z=1}^{2} \pi_{z,i}^j P_z\left(K; \lambda_i^j; \sigma_{z,i}^j\right)}$, where $P_z\left(K; \lambda_i^j; \sigma_{z,i}^j\right)$ are given by Eq.(4), for $z = 1, 2$.

Using a quadratic error loss function, the optimal estimate of $\lambda_i^{t+1}$is the mean of the posterior structure function given by

$$E\left(\lambda_i^{t+1}|K_i^1, ..., K_i^t; c_{\xi z,i}^1, ..., c_{\xi z,i}^{t+1}\right) = \sum_{z=1}^{2} \acute{\pi}_{z,i}^j \frac{w_{1,z,i}^j}{w_{2,z,i}^j} \tag{46}$$

and the variance of the posterior structure function is given by

$$
\begin{aligned}
&Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right)= \\
&\sum_{z=1}^{2}\acute{\pi}^j_{z,i}\frac{w^j_{1,z,i}}{\left(w^j_{2,z,i}\right)^2}+\acute{\pi}^j_{1,i}\acute{\pi}^j_{2,i}\left[\frac{w^j_{1,1,i}}{w^j_{2,1,i}}-\frac{w^j_{1,2,i}}{w^j_{2,2,i}}\right]^2.
\end{aligned}
\tag{47}
$$

- Now let u be distributed according to a 2C Generalized Inverse Gaussian, GIG, mixture distribution with probability density function given by

$$
\upsilon(u)=\sum_{z=1}^{2}\pi_z\frac{(c_z)^{\nu_z}u^{\nu_z-1}\exp\left[-\frac{1}{2\sigma_z}\left(c_zu+\frac{1}{c_zu}\right)\right]}{2B_{\nu_z}\left(\frac{1}{\sigma_z}\right)},
\tag{48}
$$

for $z=1,2,\sum_{z=1}^{2}\pi_z=1$, where $\sigma_z>0$, where $-\infty<\nu_z<\infty$ and where $c_z=\frac{B_{\nu_z+1}\left(\frac{1}{\sigma_z}\right)}{B_{\nu_z}\left(\frac{1}{\sigma_z}\right)}$, where B_{ν_z} is the modified Bessel function of the third kind of order ν_z with argument ω given by Eq.(7). Then, K_i^j follows a 2C Sichel mixture distribution, where the frequency component distributions, $P_z\left(K_i=k\right)$, are given by Eq.(6) for $z=1,2$. We assume that the parameters and the mixing probabilities of this model are modelled in terms of a priori rating variables. Specifically, let λ_i^j, $\sigma^j_{z,i}$, $\nu^j_{z,i}$ and $\pi^j_{z,i}$ be given by Eqs(16, 17, 18 and 19). The posterior distribution of $\lambda_i^{t+1}$is obtained by employing a fully Bayesian approach and is given by a 2C GIG$\left(t^j_{1,z,i},t^j_{2,z,i},K+\nu^j_{z,i}\right)$ mixture, with updated parameters $t^j_{1,z,i}=\frac{c^j_{z,i}+2\sigma^j_{z,i}\sum_{j=1}^{t}\lambda_i^j}{\sigma^j_{z,i}\lambda_i^j}$, $t^j_{2,z,i}=\frac{\lambda_i^j}{\sigma^j_{z,i}c^j_{z,i}}$ and $K+\nu^j_{z,i}$, with $c^j_{z,i}=\frac{B_{\nu^j_{z,i}+1}\left(\frac{1}{\sigma^j_{z,i}}\right)}{B_{\nu^j_{z,i}}\left(\frac{1}{\sigma^j_{z,i}}\right)}$, for $z=1,2$, and updated mixing probabilities $\acute{\pi}^j_{z,i}=\pi^j_{z,i}\frac{P_z\left(K;\lambda_i^j;\sigma^j_{z,i};\nu^j_{z,i}\right)}{\sum_{z=1}^{2}\pi^j_{z,i}P_z\left(K;\lambda_i^j;\sigma^j_{z,i};\nu^j_{z,i}\right)}$, where $P_z\left(K;\lambda_i^j;\sigma^j_{z,i};\nu^j_{z,i}\right)$ are given by Eq.(6), for $z=1,2$.

Under a quadratic error loss function, the optimal estimate of $\lambda_i^{t+1}$is the mean of the posterior structure function given by

$$
E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right)=\sum_{z=1}^{2}\acute{\pi}^j_{z,i}\sqrt{\frac{t^j_{2,z,i}}{t^j_{1,z,i}}}\frac{B_{K+\nu^j_{z,i}+1}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}{B_{K+\nu^j_{z,i}}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}
\tag{49}
$$

and the variance of the posterior structure function is given by

$$
\begin{aligned}
&Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right)= \\
&\sum_{z=1}^{2}\acute{\pi}^j_{z,i}\frac{t^j_{2,z,i}}{t^j_{1,z,i}}\left[\frac{B_{K+\nu^j_{z,i}+2}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}{B_{K+\nu^j_{z,i}}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}-\left(\frac{B_{K+\nu^j_{z,i}+1}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}{B_{K+\nu^j_{z,i}}\left(\sqrt{t^j_{1,z,i}t^j_{2,z,i}}\right)}\right)^2\right]+ \\
&\acute{\pi}^j_{1,i}\acute{\pi}^j_{2,i}\left[\sqrt{\frac{t^j_{2,1,i}}{t^j_{1,1,i}}}\frac{B_{K+\nu^j_{1,i}+1}\left(\sqrt{t^j_{1,1,i}t^j_{2,1,i}}\right)}{B_{K+\nu^j_{1,i}}\left(\sqrt{t^j_{1,z,i}t^j_{2,1,i}}\right)}-\sqrt{\frac{t^j_{2,2,i}}{t^j_{1,2,i}}}\frac{B_{K+\nu^j_{2,i}+1}\left(\sqrt{t^j_{1,2,i}t^j_{2,2,i}}\right)}{B_{K+\nu^j_{2,i}}\left(\sqrt{t^j_{1,2,i}t^j_{2,2,i}}\right)}\right]^2.
\end{aligned}
\tag{50}
$$

- Finally, let u be distributed according to a 2C Gamma-Generalized Inverse Gaussian mixture distribution with probability density function given by

$$
\upsilon(u)=\pi_1\frac{u^{\frac{1}{\sigma_1}-1}\frac{1}{\sigma_1}^{\frac{1}{\sigma_1}}\exp\left(-\frac{1}{\sigma_1}u\right)}{\Gamma\left(\frac{1}{\sigma_1}\right)}+\pi_2\frac{(c_2)^{\nu_2}u^{\nu_2-1}\exp\left[-\frac{1}{2\sigma_2}\left(c_2u+\frac{1}{c_2u}\right)\right]}{2B_{\nu_2}\left(\frac{1}{\sigma_2}\right)},
\tag{51}
$$

for $z=1,2, \sum_{z=1}^{2}\pi_z=1$, where $\sigma_z>0$, where $-\infty<\nu_2<\infty$ and where $c_2=\frac{B_{\nu_2+1}\left(\frac{1}{\sigma_2}\right)}{B_{\nu_2}\left(\frac{1}{\sigma_2}\right)}$, where $B_{\nu 2}$ is the modified Bessel function of the third kind of order ν_2 with argument ω. Then, K_i^j follows a 2C Negative Binomial Type I-Sichel mixture distribution where the frequency component distributions, $P_z\left(K_i=k\right)$, are given by Eqs(4 and 6) for $z=1$ and $z=2$ respectively. We assume that the parameters and the mixing probabilities of this model are modelled in terms of a priori rating variables. Specifically, let λ_i^j, $\sigma_{z,i}^j$, $\nu_{2,i}^j$ and $\pi_{z,i}^j$ be given by Eqs(16, 17, 18 and 19). The posterior distribution of λ_i^{t+1} is obtained by employing a fully Bayesian approach and is given by a 2C Gamma-Generalized Inverse Gaussian mixture $\left(w_{1,1,i}^j, w_{2,1,i}^j, t_{1,2,i}^j, t_{2,2,i}^j, K+\nu_{2,i}^j\right)$ (i.e. the first component follows the Gamma distribution and the second component follows the Generalized Inverse Gaussian distribution), with updated parameters $w_{1,1,i}^j=\frac{1}{\sigma_{1,i}^j}+K$, $w_{2,1,i}^j=\frac{\frac{1}{\sigma_{1,i}^j}+\sum_{j=1}^{t}\lambda_{1,i}^j}{\lambda_{1,i}^j}$, $t_{1,2,i}^j=\frac{c_{2,i}^j+2\sigma_{2,i}^j\sum_{j=1}^{t}\lambda_i^j}{\sigma_{2,i}^j\lambda_i^j}$, $t_{2,2,i}^j=\frac{\lambda_i^j}{\sigma_{2,i}^j c_{2,i}^j}$ and $K+\nu_{2,i}^j$, with $c_{2,i}^j=\frac{B_{\nu_{2,i}^j+1}\left(\frac{1}{\sigma_{2,i}^j}\right)}{B_{\nu_{2,i}^j}\left(\frac{1}{\sigma_{2,i}^j}\right)}$, and updated mixing probabilities $\acute{\pi}_{1,i}^j=\pi_{1,i}^j\frac{P_1\left(K;\lambda_i^j;\sigma_{1,i}^j\right)}{\pi_{1,i}^jP_1\left(K;\lambda_i^j;\sigma_{1,i}^j\right)+\pi_{2,i}^jP_2\left(K;\lambda_i^j;\sigma_{2,i}^j;\nu_{2,i}^j\right)}$ and $\acute{\pi}_{2,i}^j=\pi_{2,i}^j\frac{P_2\left(K;\lambda_i^j;\sigma_{2,i}^j;\nu_{2,i}^j\right)}{\pi_{1,i}^jP_1\left(K;\lambda_i^j;\sigma_{1,i}^j\right)+\pi_{2,i}^jP_2\left(K;\lambda_i^j;\sigma_{2,i}^j;\nu_{2,i}^j\right)}$, where $P_1\left(K;\lambda_i^j;\sigma_{1,i}^j\right)$ is given by Eq.(4) and $P_2\left(K;\lambda_i^j;\sigma_{2,i}^j;\nu_{2,i}^j\right)$ is given by Eq.(6).

Under a quadratic error loss function, the optimal estimate of $\lambda_i^{t+1}$is the mean of the posterior structure function given by

$$E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)=\acute{\pi}_{1,i}^j\frac{w_{1,1,i}^j}{w_{2,1,i}^j}+\acute{\pi}_{2,i}^j\sqrt{\frac{t_{2,2,i}^j}{t_{1,2,i}^j}}\frac{B_{K+\nu_{2,i}^j+1}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}{B_{K+\nu_{2,i}^j}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)} \tag{52}$$

and the variance of the posterior structure function is given by

$$\begin{aligned}&Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c_{\xi z,i}^1,...,c_{\xi z,i}^{t+1}\right)=\\&\acute{\pi}_{1,i}^j\frac{w_{1,1,i}^j}{\left(w_{2,1,i}^j\right)^2}+\acute{\pi}_{2,i}^j\frac{t_{2,2,i}^j}{t_{1,2,i}^j}\left[\frac{B_{K+\nu_{2,i}^j+2}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}{B_{K+\nu_{2,i}^j}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}-\left(\frac{B_{K+\nu_{2,i}^j+1}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}{B_{K+\nu_{2,i}^j}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}\right)^2\right]+\\&\acute{\pi}_{1,i}^j\acute{\pi}_{2,i}^j\left[\frac{w_{1,1,i}^j}{w_{2,1,i}^j}-\sqrt{\frac{t_{2,2,i}^j}{t_{1,2,i}^j}}\frac{B_{K+\nu_{2,i}^j+1}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}{B_{K+\nu_{2,i}^j}\left(\sqrt{t_{1,2,i}^jt_{2,2,i}^j}\right)}\right]^2.\end{aligned} \tag{53}$$

- The posterior mean and posterior variance of the 2C Poisson Inverse Gaussian mixture distribution are given by Eqs(49 and 50) if we let $\nu_z=-0.5$ for $z=1,2$ respectively.
- The posterior mean and posterior variance of the 2C Negative Binomial Type I-Poisson Inverse Gaussian mixture distribution are given by Eqs(52 and 53) for $z=1$ and $z=2$ when $\nu_2=-0.5$ respectively.
- The posterior mean and posterior variance of the 2C Poisson Inverse Gaussian-Sichel mixture distribution are given by Eqs(49 and 50) for $\nu_1=-0.5$ and $z=1,2$ respectively.
- The posterior mean and the posterior variance of the Negative Binomial Type I, Sichel and Poisson Inverse Gaussian distributions can be obtained as special cases of those for the case of the two component mixtures of these distributions.

Calculation of the Premiums According to the Expected Value and Variance Principles

The premium rates calculated according to the expected value principle are given by

$$P_1 = (1+\omega_1)\, E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right), \tag{54}$$

where $w_1 > 0$ is a risk load.

The premium rates calculated according to the variance principle are given by

$$P_2 = (1+w_2)\, E\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right) + w_2\left[Var\left(\lambda_i^{t+1}|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right)\right], \tag{55}$$

where $w_2 > 0$ is a risk load[5].

Note that the premiums derived by Eqs(54 and 55) in the case when only the a posteriori criteria is considered are obtained by assuming that the regression components are limited to constants.

3.2.2 Severity Component

Let us consider now the severity component. In what follows we construct an optimal BMS derived by updating the posterior mean and the posterior variance in the case of the 2C Pareto mixture model. Note that the system resulting from the Pareto model can be obtained as special cases of the one for the case of the 2C Pareto mixture model.

Assume that $X^j_{i,k}$ follows the Exponential distribution with mean yw, where $y > 0$ and where $w > 0$ is a continuous random variable distributed according to a 2C Inverse Gamma mixture distribution with pdf

$$\omega(w) = \sum_{z=1}^{2}\rho_z \frac{\frac{1}{(s_z-1)}\exp\left(-\frac{(s_z-1)}{w}\right)}{\left(\frac{w}{s_z-1}\right)^{s_z+1}\Gamma(s_z)}, \tag{56}$$

for $i = 1,...,n$ and $s > 0$, with mean $E(w) = 1$. Then, the unconditional distribution of $X^j_{i,k}$ is a Pareto distribution where the severity component distributions are given by Eq.(14). We can allow the parameters and the mixing probabilities of this model to vary from one individual to another. Let $y^j_{z,i}$, $s^j_{z,i}$ and $\rho^j_{z,i}$ be given by Eqs(31, 32 and 33). The posterior distribution of $y_i^{t+1}$is obtained by employing a fully Bayesian approach (i.e. by updating both the parameters and the mixing proportions of the mixing distribution) and is given by a 2C Inverse Gamma mixture $\left(v^j_{1,z,i}, v^j_{2,z,i}\right)$, with updated parameters $v^j_{1,z,i} = s^j_{z,i} + K$ and $v^j_{2,z,i} = \left(s^j_{z,i}-1\right)y^j_{z,i} + X$, for $z = 1,2$, with $X = \sum_{k=1}^{K} X^j_{i,k}$, and updated mixing probabilities $\acute{\rho}^j_{z,i} = \rho^j_{z,i}\frac{f_z\left(X;y^j_{z,i};s^j_{z,i}\right)}{\sum_{z=1}^{2}\rho^j_{z,i}f_z\left(X;y^j_{z,i};s^j_{z,i}\right)}$, where $f_z\left(X;y^j_{z,i};s^j_{z,i}\right)$ are given by Eq.(14), for $z = 1,2$.

Using the quadratic error loss function, the optimal estimator of y_i^{t+1} will be the mean of the posterior structure function and is given by

$$E\left(y_i^{t+1}|X^1_{i,1},...,X^t_{i,K_i^j};d^1_{\xi 2,i},...,d^{t+1}_{\xi 2,i}\right) = \sum_{z=1}^{2}\acute{\rho}^j_{z,i}\frac{v^j_{2,z,i}}{v^j_{1,z,i}-1} \tag{57}$$

and the variance of the posterior structure function is given by

[5]Notice the difference between Eq.(30) and Eq.(55). The alternative mixed Poisson models we consider in this Section were derived based on the assumption that their structure functions follow two component mixtures of alternative continuous distributions (rather than a two point discrete distributions). Thus, with the variance principle the premium is consequently given by

$$\begin{aligned} P_2 &= E\left(\mu\left(\lambda_i^{t+1}\right)|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right) + \\ &\quad w_2\left[E\left(\sigma^2\left(\lambda_i^{t+1}\right)|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right) + Var\left(\mu\left(\lambda_i^{t+1}\right)|K_i^1,...,K_i^t;c^1_{\xi z,i},...,c^{t+1}_{\xi z,i}\right)\right], \end{aligned}$$

where $\mu\left(\lambda_i^{t+1}\right) = \sigma^2\left(\lambda_i^{t+1}\right) = \lambda_i^{t+1}$ are the mean and the variance of the Poisson distribution. For more details the **interested** reader can refer to Lemaire(1995).

$$
\begin{aligned}
& Var\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right) \\
= & \sum_{z=1}^{2}\acute{\rho}_{z,i}^j\frac{\left(v_{2,z,i}^j\right)^2}{\left(v_{1,z,i}^j-1\right)^2\left(v_{1,z,i}^j-2\right)}+\acute{\rho}_{z,1}^j\acute{\rho}_{z,2}^j\left[\frac{v_{2,1,i}^j}{v_{1,1,i}^j-1}-\frac{v_{2,2,i}^j}{v_{1,2,i}^j-1}\right]^2. \qquad (58)
\end{aligned}
$$

Note that the posterior mean and the posterior variance of the Pareto distribution are obtained as special cases of those for the case of the 2C Pareto mixture distribution.

Calculation of the Premiums According to the Expected Value and Variance Principles The premium rates calculated according to the expected value principle are given by

$$
P_1=(1+\omega_1)\,E\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right), \qquad (59)
$$

where $\omega_1>0$ is a risk load.

The premium rates calculated according to the variance principle are given by

$$
\begin{aligned}
P_2 \;=\; & E\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)+\omega_2\left[E^2\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\right. \\
& \left.+2Var\left(y_i^{t+1}|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\right], \qquad (60)
\end{aligned}
$$

where $\omega_2>0$ is a risk load[6].

Note also that in the case when only the a posteriori criteria is considered the premiums rates determined by Eqs(59 and 59) are obtained by assuming that the regression components are limited to constants.

4 Numerical Illustration

The data were kindly provided by a major insurance company operating in Greece and concern a motor third party liability (MTPL) insurance portfolio observed over 3 years. The data set comprises 146129 policies. In our application, for the sake of brevity, we analyze the six best fitted claim frequency models from those presented in Section 2.1 and their special cases and all the seven claim severity models presented in Section 2.2. Specifically, the Negative Binomial Type I (NBI), the Poisson Inverse Gaussian (PIG), the Sichel (SICH), the two component Poisson mixture (2C POIS), the two component Negative Binomial Type I mixture (2C NBI) and the two component Poisson-Negative Binomial Type I mixture (2C POIS-NBI) distribution on the number of claims and the Pareto (PAR), the two component Exponential mixture (2C EXP), the two component Pareto mixture (2C PAR), the two component Lognormal mixture (2C LNO), the two component Exponential-Pareto mixture (2C EXP-PAR), the two component Exponential-Lognormal mixture (2C EXP-LNO) and the two component Lognormal-Pareto mixture (2C LNO-PAR) distribution[7] on the claim sizes. Furthermore, regression components

[6]Notice the difference between Eq.(45) and Eq.(60). The two component Pareto mixture we consider in this Section was derived by assuming that the structure function follows a two component Inverse Gamma mixture distribution (rather than a two point discrete distribution). Thus, with the variance principle the premium is consequently given by

$$
\begin{aligned}
P_2 \;=\; & E\left(\mu\left(y_i^{t+1}\right)|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)+ \\
& w_2\left[E\left(\sigma^2\left(y_i^{t+1}\right)|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)+Var\left(\mu\left(y_i^{t+1}\right)|X_{i,1}^1,...,X_{i,K_i^j}^t;d_{\xi 2,i}^1,...,d_{\xi 2,i}^{t+1}\right)\right],
\end{aligned}
$$

where $\mu\left(y_i^{t+1}\right)=y_i^{t+1}$ and $\sigma^2\left(\lambda_{i,t+1}\right)=\left(y_i^{t+1}\right)^2$ are the mean and the variance of the Exponential distribution.

[7]Note that the in the case of the Pareto, 2C Pareto mixture, 2C Exponential -Pareto mixture and 2C Lognormal-Pareto mixture models the GAMLSS package allows us to find the maximum likelihood estimators of the parameters of the Pareto2o (y',s') distribution, with pdf given by $f(x)=s'y'^{s'}(x+y')^{-s'-1}$. The Pareto$(y,s)$ distribution can be derived from a reparameterization of the pdf of the Pareto2o (y',s') distribution with $s'=s$ and $y'=(s'-1)y$. Thus $\hat{s}=\hat{s}'$ and $\hat{y}=\frac{\hat{y}'}{\hat{s}'-1}$.

are introduced in all the parameters and the mixing proportions of the aforementioned models and we include risk classifying characteristics so as to use all the available information in the estimation of the claim frequency and severity distributions. The log-likelihood function of these models is maximized with respect to their parameters and mixing probabilities, using the EM algorithm (for more details see Rigby and Stasinopoulos, 2009). In what follows, the aforementioned distributions/regression models for location, scale, shape and mixing probabilities will be used to construct optimal BMS either by updating the posterior probability of the policyholders' classes of risk or by updating the posterior mean and the posterior variance. The Bonus-Malus premium rates resulting from these systems will be calculated via the expected value and variance principles with independence between the claim frequency and severity components assumed.

4.1 Modelling Results

This subsection describes the modelling results of the distributions and regression models for location scale, shape and mixing probabilities that have been applied to model claim frequency and claim severity respectively.

The maximum likelihood estimators of the parameters and the mixing probabilities for the frequency and severity distributions are presented in Table 1 and Table 2 respectively.

Table 1: Results of the Fitted Claim Frequency Distributions

NBI	PIG	SICH	2C POIS		2C NBI		2C POIS-NBI	
λ	λ	λ	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
0.4029	0.4029	0.4029	0.0852	0.8118	0.2256	0.6328	0.1919	0.6189
σ	σ	σ	π_1		σ_1	σ_2	-	σ_2
1.0285	1.1045	1.1649	0.5627		1.9054	0.3070	-	0.6850
-	-	ν	-	-	π_1		π_1	
-	-	-0.2407	-	-	0.5646		0.5058	

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture distributions respectively.
λ, σ and ν are the location, scale and shape parameters, λ_i and σ_i are the location and shape parameters of the first, if $i = 1$ and the second, if $i = 2$, component distributions respectively and π_1 and $\pi_2 = (1 - \pi_1)$ are the mixing probabilities.

Table 2: Results of the Fitted Claim Severity Distributions

PAR	2C EXP		2C LNO		2C PAR		2C EXP-LNO		2C EXP-PAR		2C LNO-PAR	
y'	y_1	y_2	y_1	y_2	y_1'	y_2'	y_1	y_2	y_1	y_2'	y_1	y_2'
3676.44	1025.69	6815.81	6.9950	7.7481	979.46	6491.23	1514.89	8.28	841.34	2429.55	7.1592	3962.92
s'	ρ_1		s_1	s_2	s_1'	s_2'	-	s_2	-	s_2'	s_1	s_2'
2.7605	0.8165		0.2554	1.3629	2.9359	1.9224	-	0.2741	-	1.5646	0.2749	1.7877
-	-	-	ρ_1		ρ_1		ρ_1		ρ_1		ρ_1	
-	-	-	0.7972		0.7577		0.7763		0.6399		0.7963	

Note: PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture and the two component Lognormal-Pareto mixture distributions respectively.
y' and s' are the location and shape parameters, y_i, y_i' are the location parameters and s_i, s_i' are the shape parameters of the first, if $i = 1$ and the second, if $i = 2$, component distributions respectively and ρ_1 and $\rho_2 = (1 - \rho_1)$ are the mixing probabilities.

Let us now consider the regression models for approximating the number and the costs of claims respectively. The available a priori rating variables we employ are the Bonus Malus (BM) class, the horsepower (HP) of the car and the age of the car (AC). Only policyholders with complete records, i.e. where all of the variables under consideration were available, were considered. This BMS has 20 classes and the transition rules are described as follows: Each claim free year is rewarded by one class discount and each accident in a given year is penalized by one class. The variable BM class divides the classes of the current Greek BMS into four categories of drivers, those who belong to BM classes: C1= "1-2", C2 = "3-5", C3 = "6-9" and C4 = "10-20". The variable HP consists of three categories of cars, those with a HP: C1 = "0-1400 cc", C2 = "1400-1800 cc", C3 = "greater than 1800 cc". Finally, the variable AC consists of three categories of cars, those of age: C1 = "between 0 to 8 years", C2 = "between 8 to 16 years" and C3 = "greater than 16 years".

As suggested by Rigby and Stasinopoulos (2005, 2009) the claim frequency and severity regression models have been calibrated with respect to GAIC goodness of fit index. The Generalized Akaike Information Criterion (GAIC) is defined as

$$GAIC = \hat{D} + \kappa \times df, \tag{61}$$

where $\hat{D} = -2\hat{l}$ is the fitted Global deviance (DEV), $\hat{l}$ is the fitted log-likelihood, df is the degrees of freedom used in the model (i.e. the sum of the degrees of freedom used for the location, scale, shape parameters and mixing probabilities) and κ is a constant. The Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC) are special cases of the GAIC. Specifically, if we let $\kappa = 2$ we have the AIC, while if we let $\kappa = \log(n)$ we have the SBC, where n is the number of the independent observations assumed by a regression model. We followed a model selection technique close to that presented in Heller et al. (2007)[8]. Specifically, our variable selection began by examining the mean parameter of each frequency/severity model. This was achieved by adding all available explanatory variables and testing whether the exclusion of each lowered the GAIC, AIC and SBC values. After selecting the best predictor for the mean parameter, we proceeded in determining the remaining predictors by testing which rating variable of those used in the mean parameter would result in a further decrease of the GAIC when inserted in the scale and shape parameters and mixing proportions of the claim frequency and severity models respectively. Furthermore, if between the same frequency/severity distributions with different parameter specifications several models have similar AIC and BIC values, we preferred the simpler model so as to avoid overfitting. Therefore, the scale and shape parameters and the mixing probabilities of the models have fewer predictors than the mean parameter (see Tables 3 and 4). With regard to this, the final claim frequency and severity models we selected are those that yield the lowest GAIC, AIC, and SBC values. Also, every explanatory variable they contain is statistically significant at a 5% threshold[9].

[8] Heller et al. (2007) used generalized additive models for location scale and shape (GAMLSS) for the statistical analysis of the total amount of insurance paid out on a policy.

[9] Note that, as we have already mentioned, the location, scale, shape and mixing proportions of the alternative **claim** frequency models can be modelled according to Eqs(16, 17, 18 and 19) and the location and scale parameters and the mixing proportions of the various claim severity models can be modelled according to Eqs(31, 34, 32 and 33).

Table 3: Results of the Fitted Claim Frequency Regression Models for Location, Scale, Shape and Mixing Probabilities

NBI		PIG		SICH		2C POIS			2C NBI			2C POIS-NBI		
Variable	λ	Variable	λ	Variable	λ	Variable	λ_1	λ_2	Variable	λ_1	λ_2	Variable	λ_1	λ_2
Intercept	−0.8028	Intercept	−0.8025	Intercept	−0.8026	Intercept	−1.4317	0.3477	Intercept	−1.3293	−0.3474	Intercept	−1.4878	−0.41220
BM		BM		BM		BM			BM			BM		
C2	−0.0092	C2	−0.0100	C2	−0.0098	C2	−0.0165	−0.0359	C2	0.0104	−0.0352	C2	0.0256	−0.0328
C3	0.0749	C3	0.0755	C3	0.0754	C3	0.0392	0.0051	C3	0.1344	−0.0049	C3	0.0912	0.0780
C4	−0.0011	C4	−0.0033	C4	−0.0028	C4	−0.0091	0.0284	C4	0.0180	−0.0429	C4	0.0045	0.0128
HP		HP		HP		HP			HP			HP		
C2	0.0169	C2	0.0172	C2	0.0172	C2	0.0598	−0.0241	C2	0.7034	−0.7384	C2	−0.1235	0.1162
C3	0.0674	C3	0.0675	C3	0.0675	C3	0.0579	0.0847	C3	0.8670	−0.9212	C3	0.0252	0.0939
AC		AC		AC		AC			AC			AC		
C2	−0.1177	C2	−0.1180	C2	−0.1179	C2	−0.0803	−0.1005	C2	−0.1440	−0.0784	C2	−0.0971	−0.1593
C3	−0.4567	C3	−0.4572	C3	−0.4571	C3	1.1446	−2.5354	C3	−0.3686	−0.6020	C3	−2.2101	0.0588
Variable	σ	Variable	σ	Variable	σ	Variable	π_1		Variable	π_1		Variable	π_1	
Intercept	−0.0197	Intercept	0.0480	Intercept	0.1077	Intercept	1.5480		Intercept	−0.0691		Intercept	0.3815	
HP		HP		HP		BM			BM			BM	0.0026	
C2	−0.0398	C2	−0.0360	C2	−0.0409	C2	−0.0765		C2	0.1412		C2		
C3	0.1373	C3	0.1610	C3	0.1701	C3	−0.1965		C3	−0.1488		C3	0.0526	
AC		AC		AC		C4	0.0381		C4	0.4412		C4	0.0629	
C2	−0.0340	C2	−0.0402	C2	−0.0422	AC			AC			AC		
C3	0.1520	C3	0.1388	C3	0.1554	C2	0.0681		C2	0.4703		C2	−0.1477	
-	-	-	-	-	-	C3	−2.5297		C3	0.6296		C3	−0.1704	
-	-	-	-	-	ν	-	-		-	σ_1	σ_2	-	-	σ_2
-	-	-	-	-	−0.2057	-	-		-	−0.0334	−0.6001	-	-	0.9037

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture regression models for location, scale, shape and mixing probabilities respectively.
λ, σ and ν are the location, scale and shape parameters, λ_i and σ_i are the location and shape parameters of the first, if $i = 1$ and the second, if $i = 2$, component distributions respectively and π_1 and $\pi_2 = (1 - \pi_1)$ are the mixing probabilities.
BM, HP and AC are the variables Bonus-Malus class, horsepower of the car and age of the car respectively.
C1, C2, C3 and C4 are the categories 1, 2, 3 and 4 respectively.

Table 4: Results of the Fitted Claim Severity Regression Models for Location, Scale, Shape and Mixing Probabilities

PAR		2C EXP			2C LNO			2C PAR			2C EXP-LNO			2C EXP-PAR			2C LNO-PAR		
Variable	y'	Variable	y_1	y_2	Variable	y_1	y_2	Variable	y_1'	y_2'	Variable	y_1	y_2	Variable	y_1	y_2'	Variable	y_1	y_2'
Intercept	8.0617	Intercept	6.9002	8.5861	Intercept	6.8807	7.4369	Intercept	6.8462	8.9700	Intercept	7.1704	8.0779	Intercept	6.9971	7.6318	Intercept	6.9859	8.0508
BM		BM			BM			BM			BM			BM			BM		
C2	0.0066	C2	0.0278	−0.0382	C2	−0.0030	0.0190	C2	−0.0053	0.0357	C2	0.0120	0.0028	C2	0.0131	−0.0190	C2	0.0028	0.0119
C3	−0.0138	C3	0.0074	−0.0030	C3	0.0107	−0.0252	C3	−0.0142	−0.0248	C3	−0.0204	−0.0007	C3	−0.0183	−0.0137	C3	−0.0007	−0.0204
C4	−0.0295	C4	−0.0287	0.0085	C4	−0.0102	−0.0293	C4	0.0014	−0.0559	C4	−0.0275	−0.0177	C4	0.0063	−0.1148	C4	−0.0176	−0.0275
HP		HP			HP			HP			HP			HP			HP		
C2	−0.0697	C2	−0.0238	0.0207	C2	0.0046	−0.0097	C2	−0.0997	−0.0028	C2	−0.0602	0.0160	C2	0.0003	−0.1630	C2	0.0159	−0.0602
C3	−0.0335	C3	0.0950	0.0516	C3	0.0298	0.1569	C3	0.0253	−0.0065	C3	0.0797	0.0378	C3	0.0381	0.43144	C3	0.0378	0.0797
AC		AC			AC			AC			AC			AC			AC		
C2	0.0901	C2	−0.0853	0.1313	C2	−0.0593	−0.0982	C2	−0.0179	0.0422	C2	0.0276	−0.0607	C2	−0.0560	0.0212	C2	−0.0607	0.0275
C3	0.1984	C3	3.7863	−2.6951	C3	−1.1978	0.1638	C3	0.3164	−0.5636	C3	0.1084	−0.0826	C3	−0.0283	−0.0345	C3	−0.0827	0.1084
Variable	s'	Variable	ρ_1		Variable	ρ_1		Variable	ρ_1		Variable	ρ_1		Variable	ρ_1		Variable	ρ_1	
Intercept	1.0358	Intercept	1.5700		Intercept	1.3872		Intercept	1.1752		Intercept	1.2863		Intercept	0.5908		Intercept	1.4275	
HP		BM			BM			BM			BM			BM			BM		
C2	−0.0473	C2	0.1979		C2	0.0405		C2	−0.4504		C2	−0.3729		C2	−0.0047		C2	0.0252	
C3	−0.1280	C3	0.1262		C3	−1.1094		C3	0.1412		C3	−0.4275		C3	−0.0468		C3	0.0256	
AC		C4	0.1360		C4	0.0336		C4	−0.1686		C4	0.3027		C4	0.0340		C4	0.1242	
C2	0.1705	AC			AC			AC			AC			AC			AC		
C3	0.3010	C2	0.4892		C2	0.3448		C2	−0.4365		C2	0.4411		C2	0.7132		C2	−0.3103	
-	-	C3	−7.4517		C3	0.6958		C3	4.5289		C3	0.5670		C3	2.0575		C3	−0.6988	
-	-	-	-		Variable	s_1	s_2	Variable	s_1'	s_2'	Variable	-	s_2	Variable	-	s_2'	Variable	s_1	s_2'
-	-	-	-		Intercept	−1.2392	0.3519	Intercept	−0.7937	−0.9596	Intercept	-	−1.2753	Intercept	-	−0.5138	Intercept	−1.2753	−0.5448
-	-	-	-		HP			HP			HP	-		HP	-		HP		
-	-	-	-		C2	0.0054	−0.0014	C2	0.0874	−0.1186	C2	-	0.0378	C2	-	0.0788	C2	0.0379	0.0284
-	-	-	-		C3	0.1142	−0.0073	C3	0.0802	1.6276	C3	-	0.1701	C3	-	−0.0740	C3	0.1701	0.0500
-	-	-	-		AC			AC			AC	-		AC	-		AC		
-	-	-	-		C2	−0.1864	−0.0563	C2	−0.0653	−0.0266	C2	-	−0.1941	C2	-	0.0353	C2	−0.1941	−0.1135
-	-	-	-		C3	1.7550	−0.4078	C3	−0.4854	0.1057	C3	-	−0.2782	C3	-	0.4217	C3	−0.2781	−0.2116

Note: PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture and the two component Lognormal-Pareto mixture regression models for location, scale, shape and mixing probabilities respectively.
y_i, y_i' are the location parameters and s_i, s_i' are the shape parameters of the first, if $i = 1$, and the second, if $i = 2$, component distributions respectively and ρ_1 and $\rho_2 = (1 - \rho_1)$ are the mixing probabilities.
BM, HP and AC are the variables Bonus-Malus class, horsepower of the car and age of the car respectively.
C1, C2, C3 and C4 are the categories 1, 2, 3 and 4 respectively.

The models presented in Tables 3 and 4 extend the commonly used specification that assumes that only the mean claim frequency/severity is modelled in terms of risk factors, which was widely accepted for experience ratemaking. Moreover, the results for the location parameter of the claim frequency/severity models correspond with the existing results, based on the examination of the relative data sets, in recent Bonus-Malus literature research. Specifically, as expected, the values of the estimated regression coefficients of the explanatory variables for this parameter will lead to Bonus-Malus premiums calculated with the expected value principle which vary little under different distributional assumptions regarding a group of individuals that share the same characteristics. In the setup we consider, the systematic part of these models was extended to permit modelling of all the parameters and/or the mixing proportions of the claim frequency/severity distribution as functions of a priori rating variables enabling us to produce tailor-made premiums. Furthermore, in a Bonus-Malus ratemaking scheme that incorporates a priori risk characteristics, joint modelling of all the parameters breaks the nexus between the mean and variance implied by the standard procedure using GLM models. In this respect, the differences in the variance values of the posterior frequency/severity distributions alter significantly the premiums calculated through the variance principle since it is understood that in this case the loading is related to the variability of the loss. Moreover, our analysis shows that the employment of two component mixture models with no parameters in common captures the stylized characteristics of the data and is beneficial for the insurance company as it can provide the actuary with alternative pricing strategies in addition to those already existing in the Bonus-Malus literature.

Finally, as suggested by Stasinopoulos et al. (2008), 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 response distributions for the claim frequency and severity component. For continuous 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 $u_i = F_i(x_i|\hat{\vartheta})$, where F_i is the cumulative distribution function estimated for the ith individual, $\hat{\vartheta}$ contains all estimated model parameters and x_i is the corresponding observation. For discrete response distributions, the aforementioned definition is extended and u_i is defined as a random value from the uniform distribution on the interval $\left[F_i(x_i - 1|\hat{\vartheta}), F_i(x_i|\hat{\vartheta})\right]$. In both cases, 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 Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, 2C Poisson mixture, 2C Negative Binomial Type I mixture and 2C Poisson-Negative Binomial Type I mixture claim frequency regression models for location, scale, shape and mixing proportions.

Figure 1. Normalized quantiles for the claim frequency models

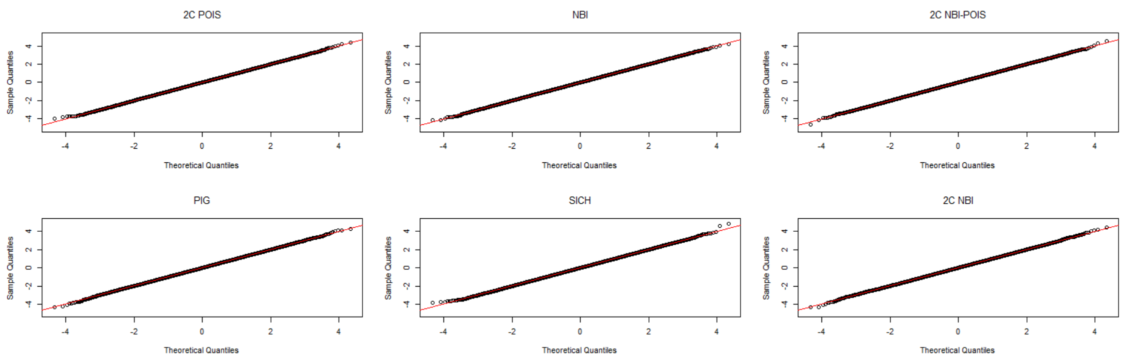
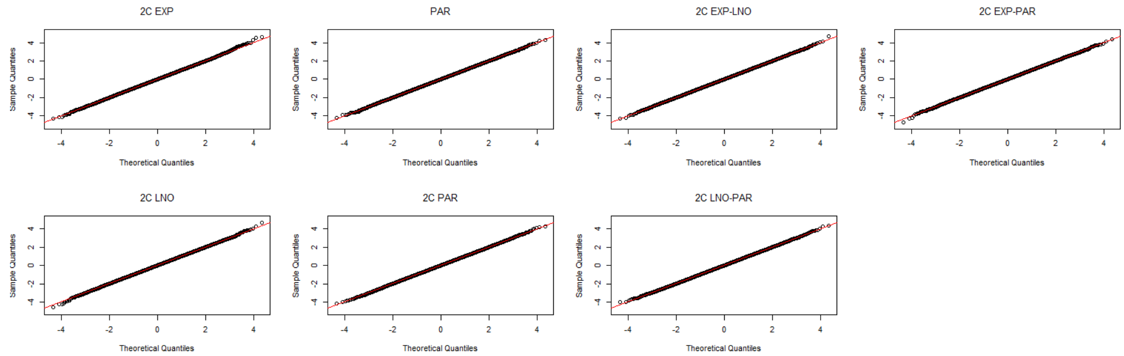


Figure 2 shows the normalized (random) quantiles for the Pareto, 2C Exponential mixture, 2C Pareto mixture, 2C Lognormal mixture, 2C Exponential-Pareto mixture, 2C Exponential-Lognormal mixture and the 2C Lognormal-Pareto mixture regression models for location, scale, shape and mixing probabilities.

Figure 2. Normalized quantiles for the claim severity models



From Figures 1 and 2 we see that the residuals of the claim frequency and severity models are very close to the diagonal and indicate a very good fit to the distribution of the claim frequencies and claim severities respectively.

4.2 Models Comparison

Thus far, we have several competing models for the claim frequency and severity components. The differences between models produce different premiums calculated according to the expected value and variance principles. Consequently, to differentiate between these models, this section compares them so as to select the best for each case. Following Rigby and Stasinopoulos (2009), we resort to the information criteria, such as the Global Deviance, AIC or the SBC which are valid for both nested or non-nested model comparisons. The resulting Global Deviance, AIC and SBC are given in Table 5 for the different claim frequency (Panel A) and claim severity (Panel B) fitted distributions and regression models for location, scale, shape and mixing probabilities.

Table 5: Models Comparison

Panel A: Frequency Component

	Distributions			Regression Models for Location, Scale, Shape and Mixing Probabilities			
Model	df	AIC	SBC	df	Global Deviance	AIC	SBC
NBI	2	245841	245861	13	244897	244922	245051
PIG	2	245767	245787	13	244830	244856	244984
SICH	3	245755	245775	14	244817	244845	244970
2C POIS	3	245862	245882	22	244851	244875	245013
2C NBI	5	245749	245768	24	244721	244743	244889
2C NBI-POIS	4	245792	245815	23	244789	244810	244929

Panel B: Severity Component

	Distributions			Regression Models for Location, Scale, Shape and Mixing Probabilities			
Model	df	AIC	SBC	df	Global Deviance	AIC	SBC
PAR	2	691872	691889	13	681649	681687	681953
2C EXP	3	691925	691951	22	681592	681632	681906
2C LNO	5	688101	688145	32	677529	677573	677943
2C PAR	5	686911	686954	32	676308	676356	676726
2C EXP-LNO	4	690557	690592	27	680120	680153	680426
2C EXP-PAR	4	690300	690335	27	679839	679876	680150
2C LNO-PAR	5	685929	685972	32	675411	675463	675842

Note: df is the degrees of freedom, AIC is the Akaike information criterion and SBC is Schwartz Bayesian criterion.
NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture models respectively.
PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture, two component Lognormal-Pareto mixture models respectively.

Overall, from Panel A we observe that the best fit is given by the 2C Negative Binomial Type I mixture distribution/regression model for location, scale, shape and prior probabilities. From Panel B, we see that the best fit is given by the 2C Lognormal-Pareto mixture distribution/regression model for location, scale, shape and prior probabilities.

4.3 Optimal Bonus-Malus Premiums Calculated Via the Expected Value and Variance Principles

Following the current methodology, as presented in sections 3.1 and 3.2, we derive optimal BMS with a frequency and a severity component both by updating the posterior probability of the policyholders' classes of risk and by updating the posterior mean and the posterior variance based on the a posteriori criteria and based both on the a priori and the a posteriori criteria. For the case of updating the posterior probability we assume that a policyholder who belongs to the first category is a good risk while one who belongs to the second category is a bad risk. In our application we consider that the specific policyholder belongs to the second category[10]. Furthermore, when both criteria are considered, we examine a group of policyholders who share the following common characteristics: We consider that the policyholder i belongs to the first BM class, and has a car between 0 to 8 years old with HP between 0-1400 cc. In (Section 1 and Section 2) the Bonus- Malus premiums rates will be calculated via the expected value and

[10] The analogous procedure can be applied for a policyholder who belongs in the first category.

the variance premium principle respectively. These premium rates will be divided by the premium when $t = 0$, since we are interested in the differences between various classes. The results are presented so that the premium for a new policyholder is 100. Thus, in what follows, when the expected value principle is used note the disappearance of the factors $(1 + w_1)$ and $(1 + \omega_1)$from Eqs(29, 44, 54 and 59). Also, when the variance principle is used, following and extending the framework of Lemaire (1995) for two component mixtures with no parameters in common of frequency and severity distributions/regression models for location, scale, shape and prior probabilities we assume that $w_2 = \omega_2 = 0.235$ in Eqs(30, 45, 55 and 60) which corresponds to a safety loading of 25% of the net premium.

4.3.1 Expected Value Premium Calculation Principle

We consider first the optimal BMS resulting from the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, 2C Poisson mixture, 2C Negative Binomial Type I mixture and 2C Poisson-Negative Binomial Type I mixture claim frequency distributions/regression models for location, scale, shape and mixing proportions. The results are presented in Table 6 and Table 7 respectively.

As we mentioned previously, for the optimal BMS derived by updating the posterior probability in the case of the 2C Negative Binomial Type I mixture and the 2C Poisson-Negative Binomial Type I mixture distributions/regression models, the explicit claim frequency history determines the calculation of the posterior probabilities and thus of premium rates to be calculated with the expected value principle, and not just the total number of claims as in the case of the 2C Poisson mixture distribution/regression model. Also, for the system resulting from updating the posterior mean in the case of the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel and 2C Negative Binomial Type I mixture regression models the explanatory variable Bonus-Malus class varies substantially depending on the number of claims of policyholder i for period j. Thus, in this case also, the explicit claim frequency history determines the calculation of the premium rates. Due to the aforementioned reasons, in Tables 6 and 7 we specify the exact order of the claims history in order to derive the scaled premiums that must be paid by the specific group of policyholders that we consider, assuming that the age of the policy is up to 2 years. From both of these tables we observe that if the policyholder i has a claim free year, the premium rates reduce, whereas if they have one or more claims, the premium rates increase, resulting in bonus or malus respectively. For example, from Table 6 we see that policyholders who had two claims over the second year of observation will have to pay a malus of 144.78%,158.61%,157.31%,130.43% and 97.59% of the basic premium in the case of the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, 2C Negative Binomial Type I mixture distributions derived by updating the posterior mean and the 2C Poisson mixture distribution derived by updating the posterior probability respectively. Also, we see that policyholders who had at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ (i.e. total number of claims $K = 2$ at $t = 2$) will have to pay a malus of 27.67% and 36.87% of the basic premium and those who had $k_1 = 1, k_2 = 1$ claim frequency history (i.e. total number of claims $K = 2$ at $t = 2$) will have to pay a malus of 41.32% and 39.15% of the basic premium in the case of the 2C NBI mixture and 2C Poisson-NBI mixture distributions derived by updating the posterior probability. Furthermore, from Table 7 when both the a priori and the a posteriori criteria are considered, we see, for instance, that policyholders who had at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ will have to pay a malus of 132.14%,113.49%,125.16%, 89.80%, 26.54% and 36.87% and those who had $k_1 = 1, k_2 = 1$ claim frequency history will have to pay a malus of 132.36%,114.00%,125.62%, 90.16%, 29.55% and 39.15% in the case of the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel and 2C Negative Binomial Type I mixture regression models derived by updating the posterior mean and the 2C Negative Binomial Type I mixture and 2C Poisson-Negative Binomial Type I mixture models derived by updating the posterior probability respectively. Also, we observe that a group of policyholders who had two claims over the second year of observation will have to pay a malus of 161.51% in the case of the 2C Poisson mixture model derived by updating the posterior probability.

Table 6: Optimal BMS, Expected Value Principle, Distributions for Assessing Claim Frequency

NBI						PIG					
Year	Number of Claims k					Year	Number of Claims k				
t	0	1	2	3	4	t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00	0	100.00	0.00	0.00	0.00	0.00
1	88.93	180.40	271.87	363.34	454.80	1	88.83	176.00	306.31	461.55	627.20
2	80.07	162.43	244.78	327.14	409.49	2	80.73	152.70	258.61	385.09	520.77
SICH						**2C POIS**					
Year	Number of Claims k					Year	Number of Claims k				
t	0	1	2	3	4	t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00	0	100.00	0.00	0.00	0.00	0.00
1	88.82	177.01	300.76	442.09	590.18	1	90.49	175.52	198.33	201.12	201.42
2	80.57	154.45	257.31	375.28	499.42	2	81.44	170.28	197.59	201.04	201.41

2C NBI (Post. Mean)					
Year	Number of Claims k				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	92.46	183.03	246.61	306.97	370.47
2	86.10	171.01	230.43	286.34	344.52

Year	Number of Claims k_t	2C NBI (Post. Prob.)	2C POIS-NBI
t=0	$k_0 = 0$	100	100
	$k_1 = 0$	97.02	96.9
t=1	$k_1 = 1$	123.87	123.21
	$k_1 = 2$	130.13	138.46
	$k_1 = 0, k_2 = 0$	94.11	93.82
t=2	$k_1 = 0, k_2 = 1$	121.12	120.61
	$k_1 = 0, k_2 = 2$	127.67	136.87
	$k_1 = 1, k_2 = 0$	121.12	120.61
t=2	$k_1 = 1, k_2 = 1$	141.32	139.15
	$k_1 = 1, k_2 = 2$	144.88	147.10
	$k_1 = 2, k_2 = 0$	127.67	136.87
t=2	$k_1 = 2, k_2 = 1$	144.88	147.10
	$k_1 = 2, k_2 = 2$	147.72	150.81

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture distributions respectively.

Table 7: Optimal BMS, Expected Value Principle, Regression Models for Location, Scale, Shape and Mixing Probabilities for Assessing Claim Frequency

Year	Number of Claims k_t	NBI	PIG	SICH
$t=0$	$k_0=0$	100	100	100
	$k_1=0$	88.34	88.30	88.29
$t=1$	$k_1=1$	173.37	162.47	169.74
	$k_1=2$	259.20	272.18	283.69
	$k_1=0,k_2=0$	79.12	79.94	78.70
$t=2$	$k_1=0,k_2=1$	155.27	134.82	140.25
	$k_1=0,k_2=2$	232.14	213.49	225.16
	$k_1=1,k_2=0$	156.85	136.40	141.87
$t=2$	$k_1=1,k_2=1$	232.36	214.00	225.62
	$k_1=1,k_2=2$	336.45	336.00	352.18
	$k_1=2,k_2=0$	232.36	214.00	225.62
$t=2$	$k_1=2,k_2=1$	336.45	336.00	352.18
	$k_1=2,k_2=2$	420.14	447.84	464.83

Year	Number of Claims k_t	2C NBI (Post. Mean)	2C NBI (Post. Prob.)	2C POIS-NBI
$t=0$	$k_0=0$	100	100	100
	$k_1=0$	90.44	97.22	96.9
$t=1$	$k_1=1$	150.36	117.18	123.21
	$k_1=2$	207.40	128.17	138.46
	$k_1=0,k_2=0$	82.59	94.45	93.82
$t=2$	$k_1=0,k_2=1$	137.44	114.83	120.61
	$k_1=0,k_2=2$	189.80	126.54	136.87
	$k_1=1,k_2=0$	141.11	114.83	120.61
$t=2$	$k_1=1,k_2=1$	190.16	129.55	139.15
	$k_1=1,k_2=2$	250.75	135.92	147.10
	$k_1=2,k_2=0$	190.16	126.54	136.87
$t=2$	$k_1=2,k_2=1$	250.75	135.92	147.10
	$k_1=2,k_2=2$	301.66	139.48	150.81

2C POIS

Year	Number of Claims k				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	88.00	177.74	275.62	309.70	316.51
2	78.74	154.96	261.51	306.36	315.91

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture regression models for location, scale, shape and mixing probabilities respectively.

Let us now consider the severity component and the optimal BMS derived by updating the posterior mean in the case of the Pareto, and the systems resulting from updating the posterior probability in the case of the 2C Exponential mixture, 2C Pareto mixture, 2C Lognormal mixture, 2C Exponential-Pareto mixture, 2C Exponential-Lognormal mixture and the 2C Lognormal-Pareto mixture distributions/regression models for location, scale, shape and mixing probabilities. Table 8 (Panels A and B) displays the premium rates resulting from these models with respect to the a posteriori criteria (Panel A) and to both the a priori and the a posteriori criteria (Panel B). From Table 8 we observe that the premium values increase proportionally to the claim costs. For example, from Panel A we see that for one claim size of 3500 in the first year the premium increases from 100 to 124.49, 154.59, 280.72, 268.32, 149.39, 150.49 and 236.18 in the case of the Pareto, 2C Exponential mixture, 2C Lognormal mixture, 2C Pareto mixture, 2C Exponential-Lognormal mixture, 2C Exponential-Pareto mixture and 2C Lognormal-Pareto mixture distributions respectively. Furthermore, from Panel B we observe that for one claim size of 3500 in the first year the premium increases from 100 to 136.61, 158.36, 267.13, 192.23, 153.82, 117.73 and 247.57 in the case of the Pareto, 2C Exponential mixture, 2C Lognormal mixture, 2C Pareto mixture, 2C Exponential-Lognormal mixture, 2C Exponential-Pareto mixture and 2C Lognormal-Pareto mixture regression models respectively.

Table 8: Optimal BMS, Expected Value Principle, One Claim in the First Year of Observation

Panel A: Distributions for Assessing Claim Severity							
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	89.79	78.20	74.32	149.23	72.84	97.61	73.50
2500	107.14	107.75	258.73	224.88	103.75	121.91	150.01
3500	124.49	154.59	280.72	268.32	149.39	150.49	236.18
4500	141.83	211.13	280.87	291.95	159.51	175.16	240.81

Panel B: Regression Models for Location, Scale, Shape and Mixing Probabilities for Assessing Claim Severity							
Claim Size	PAR	2C	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	95.66	86.10	86.20	122.53	77.49	95.39	80.37
2500	116.14	114.66	248.15	164.26	131.16	104.29	212.95
3500	136.61	158.36	267.13	192.23	153.82	117.73	247.57
4500	157.09	208.99	267.39	210.36	153.82	133.50	248.19

Note: PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture and the two component Lognormal-Pareto mixture models respectively.

Finally, we compute the optimal BMS with a frequency and a severity component using the expected value premium calculation principle. The premiums resulting from this system are calculated via the product of the premiums calculated for frequency component and those calculated for severity component with independence between the two components assumed. Table 9 (Panels A, B, C, D, E, F and G) summarizes our findings with respect to the a posteriori criteria and Table 10 (Panels A, B, C, D, E, F and G) presents our findings with respect to both criteria.

Table 9: Optimal BMS with a Frequency and a Severity Component, A Posteriori Criteria, Expected Value Principle, One Claim in the First Year of Observation

Panel A: NBI								Panel B: PIG						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	161.98	141.07	134.07	269.21	131.40	176.09	132.59	158.03	137.63	130.80	262.64	128.20	171.79	129.36
2500	193.28	194.38	466.75	405.68	187.16	219.93	270.62	188.57	189.64	455.36	395.79	182.60	214.56	264.02
3500	224.58	278.88	506.42	484.05	269.5	271.48	426.07	219.10	272.08	494.07	472.24	262.93	264.86	415.68
4500	255.86	380.88	506.69	526.68	287.76	315.99	434.42	249.62	371.59	494.33	513.83	280.74	308.28	423.83
Panel C: SICH								Panel D: 2C POIS						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	158.94	138.42	131.55	264.15	128.93	172.78	130.10	157.60	137.26	130.45	261.93	127.85	171.33	129.01
2500	189.65	190.73	457.98	398.06	183.65	215.79	265.53	188.05	189.12	454.12	394.71	182.10	213.98	263.30
3500	220.36	273.64	496.90	474.95	264.44	266.38	418.06	218.50	271.34	492.72	470.96	262.21	264.14	414.54
4500	251.05	373.72	497.17	516.78	282.35	310.05	426.26	248.94	370.58	492.98	512.43	279.97	307.44	422.67
Panel E: 2C NBI (Post. Mean)								Panel F: 2C NBI (Post. Prob.)						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	164.34	143.13	136.03	273.14	133.32	178.66	134.53	111.22	96.87	92.06	184.85	90.23	120.91	91.04
2500	196.10	197.21	473.55	411.60	189.89	223.13	274.56	132.71	133.47	320.49	278.56	128.52	151.01	185.82
3500	227.85	282.95	513.80	491.11	273.43	275.44	432.28	154.21	191.49	347.73	332.37	185.05	186.41	292.56
4500	259.59	386.43	514.08	534.36	291.95	320.60	440.75	175.68	261.53	347.91	361.64	197.59	216.97	298.29

Panel G: 2C POIS-NBI							
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	110.63	96.35	91.57	183.87	89.75	120.27	90.56
2500	132.01	132.76	318.78	277.07	127.83	150.21	184.83
3500	153.38	190.47	345.88	330.60	184.06	185.42	291.00
4500	174.75	260.13	346.06	359.71	196.53	215.81	296.70

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture distributions respectively.
PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture, two component Lognormal-Pareto mixture distributions respectively.

Table 10: Optimal BMS with a Frequency and a Severity Component, A Priori and A Posteriori Criteria, Expected Value Principle, One Claim in the First Year of Observation

	Panel A: NBI							Panel B: PIG						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	165.85	149.27	149.44	212.43	134.34	165.38	139.34	155.42	139.89	140.05	199.07	125.90	154.98	130.58
2500	201.35	198.79	430.22	284.78	227.39	180.81	369.19	188.69	186.29	403.17	266.87	213.10	169.44	345.98
3500	236.84	274.55	463.12	333.27	266.68	204.11	429.21	221.95	257.29	434.01	312.32	249.91	191.28	402.23
4500	272.35	362.33	463.57	364.70	266.68	231.45	430.29	255.22	339.55	434.43	341.77	249.91	216.90	403.23

	Panel C: SICH							Panel D: 2C POIS						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	162.37	146.15	146.32	207.98	131.53	161.91	136.42	170.03	153.03	153.21	217.78	137.73	169.55	142.85
2500	197.14	194.62	421.21	278.81	222.63	177.02	361.46	206.43	203.80	441.06	291.96	233.12	185.37	378.50
3500	231.88	268.80	453.43	326.29	261.09	199.83	420.23	242.81	281.47	474.80	341.67	273.40	209.25	440.03
4500	266.64	354.74	453.87	357.07	261.09	226.60	421.28	279.21	371.46	475.26	373.89	273.40	237.28	441.13

	Panel E: 2C NBI (Post. Mean)							Panel F: 2C NBI (Post. Prob.)						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	143.83	129.46	129.61	184.24	116.51	143.43	120.84	112.09	100.89	101.01	143.58	90.80	111.78	94.18
2500	174.63	172.40	373.12	246.98	197.21	156.81	320.19	136.09	134.36	290.78	192.48	153.69	122.21	249.53
3500	205.41	238.11	401.66	289.04	231.28	177.02	372.25	160.08	185.57	313.02	225.26	180.25	137.96	290.10
4500	236.20	314.24	402.05	316.30	231.28	200.73	373.18	184.08	244.89	313.33	246.50	180.25	156.44	290.83

	Panel G: 2C POIS-NBI						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	117.86	106.08	106.21	150.97	95.48	117.53	99.02
2500	143.10	141.27	305.75	202.38	161.60	128.50	262.38
3500	168.32	195.12	329.13	236.85	189.52	145.06	305.03
4500	193.55	257.50	329.45	259.18	189.52	164.49	305.79

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture regression models for location scale shape and mixing probabilities respectively.
PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture, two component Lognormal-Pareto mixture mixture regression models for location scale shape and mixing probabilities respectively.

4.3.2 Variance Premium Calculation Principle

In this case as well we consider first the optimal BMS resulting from the claim frequency distributions/regression models for location, scale, shape and prior probabilities. The results are shown in Table 11 and Table 12 respectively. Note that similarly to the results shown in the previous section, in the case of the optimal BMS derived by updating the posterior probability when the number of claims follow a 2C Negative Binomial Type I mixture and a 2C Poisson-Negative Binomial Type I mixture distribution/regression model, the explicit claim frequency history determines the calculation of the posterior probabilities and therefore of premium rates to be calculated with the variance principle, and not only the total number of claims as with the 2C Poisson mixture. Also, in the case of the systems derived by updating the posterior mean and variance when the number of accidents is approximated by the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel and 2C Negative Binomial Type I mixture regression models, the explicit claim frequency history determines the calculation of the premium rates.

Overall, from Tables 11 and 12 we observe that these seven systems are fair since if the policyholder has a claim free year the premium is reduced, while if the policyholder has one or more claims the premium is increased. For instance, from Table 11 we see that policyholders who had two claims over the second year of observation will have to pay a malus of 143.65%, 159.54%, 157.82%, 132.33% and 94.17% of the basic premium in the case of the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel and 2C Negative Binomial Type I mixture distributions derived by updating the posterior mean and the posterior variance and the 2C Poisson mixture distribution derived by updating the posterior probability respectively. Also, we see that policyholders who had at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ will have to pay a malus of 27.11% and 37.00% of the basic premium and those who had $k_1 = 1, k_2 = 1$ claim frequency history will have to pay a malus of 40.35% and 39.21% of the basic premium in the case of the 2C Negative Binomial Type I mixture and 2C Poisson-Negative Binomial Type I mixture distributions derived by updating the posterior probability. When both the a priori and a posteriori criteria are considered, from Table 12 one can see that, for example, policyholders who had at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ will have to pay a malus of 130.69%, 114.88%, 122.46%, 107.05%, 26.35% and 44.43% and those who had $k_1 = 1, k_2 = 1$ claim frequency history will have to pay a malus of 130.91%,115.35%, 122.92%, 107.48%, 29.31%, 32.39% in the case of the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel and 2C Negative Binomial Type I mixture regression models derived by updating the posterior mean and the posterior variance and the 2C Negative Binomial Type I mixture and 2C Poisson-Negative Binomial Type I mixture models derived by updating the posterior probability respectively. Also, we observe that a group of policyholders who had two claims over the second year of observation will have to pay a malus of 157.87% in the case of the 2C Poisson mixture model derived by updating the posterior probability.

Table 11: Optimal BMS, Variance Principle, Distributions for Assessing Claim Frequency

NBI						PIG					
Year	Number of Claims k					Year	Number of Claims k				
t	0	1	2	3	4	t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00	0	100.00	0.00	0.00	0.00	0.00
1	88.71	179.94	271.17	362.41	453.64	1	88.37	176.77	309.39	467.27	635.56
2	79.70	161.68	243.65	325.63	407.60	2	80.03	152.55	259.54	387.27	524.17

SICH						2C POIS					
Year	Number of Claims k					Year	Number of Claims k				
t	0	1	2	3	4	t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00	0	100.00	0.00	0.00	0.00	0.00
1	88.40	177.59	302.93	446.00	595.68	1	90.59	173.25	194.87	197.50	197.78
2	79.93	154.23	257.82	376.54	501.38	2	81.60	168.26	194.17	197.42	197.76

2C NBI (Post. Mean & Post. Var.)

Year	Number of Claims k				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	92.26	183.84	249.53	313.99	385.10
2	85.77	171.33	232.33	291.43	355.63

Year	Number of Claims k_t	2C NBI (Post. Prob.)	2C POIS-NBI
t=0	$k_0 = 0$	100	100
	$k_1 = 0$	97.06	96.86
t=1	$k_1 = 1$	123.40	123.33
	$k_1 = 2$	129.50	138.53
	$k_1 = 0, k_2 = 0$	94.19	93.75
t=2	$k_1 = 0, k_2 = 1$	120.72	120.72
	$k_1 = 0, k_2 = 2$	127.11	137.00
	$k_1 = 1, k_2 = 0$	120.72	120.72
t=2	$k_1 = 1, k_2 = 1$	140.35	139.21
	$k_1 = 1, k_2 = 2$	143.80	147.08
	$k_1 = 2, k_2 = 0$	127.11	137.00
t=2	$k_1 = 2, k_2 = 1$	143.80	147.08
	$k_1 = 2, k_2 = 2$	146.53	150.75

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture distributions respectively.

\

Table 12: Optimal BMS, Variance Principle, Regression Models for Location, Scale, Shape and Mixing Probabilities for Assessing Claim Frequency

Year	Number of Claims k_t	NBI	PIG	SICH
$t=0$	$k_0=0$	100	100	100
	$k_1=0$	88.01	87.80	86.53
$t=1$	$k_1=1$	172.67	163.26	167.63
	$k_1=2$	258.15	275.35	281.28
	$k_1=0,k_2=0$	78.64	79.18	76.88
$t=2$	$k_1=0,k_2=1$	154.30	134.80	138.00
	$k_1=0,k_2=2$	230.69	214.88	222.46
	$k_1=1,k_2=0$	155.90	136.42	139.63
$t=2$	$k_1=1,k_2=1$	230.91	215.35	222.92
	$k_1=1,k_2=2$	334.91	340.43	349.53
	$k_1=2,k_2=0$	230.91	215.35	222.92
$t=2$	$k_1=2,k_2=1$	334.91	340.43	349.53
	$k_1=2,k_2=2$	418.22	454.48	461.72

Year	Number of Claims k_t	2C NBI (Post. Mean & Post. Var.)	2C NBI (Post. Prob.)	2C POIS-NBI
$t=0$	$k_0=0$	100	100	100
	$k_1=0$	96.31	97.22	97.39
$t=1$	$k_1=1$	162.47	117.10	116.70
	$k_1=2$	227.61	128.00	146.15
	$k_1=0,k_2=0$	87.46	94.44	94.84
$t=2$	$k_1=0,k_2=1$	147.81	114.77	114.01
	$k_1=0,k_2=2$	207.05	126.35	144.43
	$k_1=1,k_2=0$	151.83	114.77	114.01
$t=2$	$k_1=1,k_2=1$	207.48	129.31	132.39
	$k_1=1,k_2=2$	281.43	135.57	154.73
	$k_1=2,k_2=0$	207.48	126.35	144.43
$t=2$	$k_1=2,k_2=1$	281.43	135.57	154.73
	$k_1=2,k_2=2$	342.62	139.05	163.85

2C POIS

Year	Number of Claims k				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	87.75	177.60	271.04	302.46	308.70
2	78.26	155.17	257.87	299.41	308.13

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture regression models for location, scale, shape and mixing probabilities respectively.

Then, for the severity component we consider the optimal BMS derived by updating the posterior mean and the posterior variance in the case of the Pareto, and the BMS resulting from updating the posterior probability in the case of the 2C Exponential, 2C Lognormal, 2C Pareto, 2C Exponential-Lognormal, 2C Exponential-Pareto and 2C Lognormal-Pareto mixture distributions/regression models. Table 13 (Panels A and B) shows the premium rates calculated according to the variance principle when the a posteriori criteria are taken into account (Panel A) and when both the a priori and the a posteriori criteria are considered (Panel B). Similarly to the results obtained when the expected value principle was used, from Table 13 we can see that the premium values calculated according to the variance principle increase proportionally to the claim severities. For instance, from Panel A we observe that for one claim size of 3500 in the first year the premium increases from 100 to 138.04,182.03, 448.77, 332.27, 102.31, 253.10 and 521.61 in the case of the Pareto, 2C Exponential mixture, 2C Lognormal mixture, 2C Pareto mixture, 2C Exponential-Lognormal mixture, 2C Exponential-Pareto mixture and 2C Lognormal-Pareto mixture distributions respectively. Also, from Panel B we can see that for one claim size of 3500 in the first year the premium increases from 100 to 113.75,198.39,463.86,239.20,105.89,154.44 and 599.99 in the case of the Pareto, 2C Exponential mixture, 2C Lognormal mixture, 2C Pareto mixture, 2C Exponential-Lognormal mixture, 2C Exponential-Pareto mixture and 2C Lognormal-Pareto mixture regression models respectively.

Table 13: Optimal BMS, Variance Principle, One Claim in the First Year of Observation

Panel A: Distributions for Assessing Claim Severity							
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	88.23	62.20	44.87	166.88	71.88	94.01	25.60
2500	111.74	112.75	410.00	271.36	102.31	160.27	247.19
3500	138.04	182.03	448.77	332.27	102.31	253.10	521.61
4500	167.10	247.96	449.03	365.69	102.31	346.24	537.10
Panel B: Regression Models for Location, Scale, Shape and Mixing Probabilities for Assessing Claim Severity							
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	55.78	72.95	67.48	134.67	79.35	86.35	43.28
2500	82.21	127.02	425.32	197.75	105.89	112.89	471.05
3500	113.75	198.39	463.86	239.20	105.89	154.44	599.99
4500	150.39	263.95	464.39	265.73	105.89	205.54	602.37

Note: PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture and the two component Lognormal-Pareto mixture models respectively.

Let us finally present the optimal BMS with a frequency and severity component when the variance principle is used. The premiums determined by this system are calculated via the product of the premiums calculated for frequency component and those calculated for severity component assuming that the frequency and severity components are independent. Table 14 (Panels A, B, C, D, E, F and G) summarizes our findings with respect to the a posteriori criteria and Table 15 (Panels A, B, C, D, E, F and G) presents our findings with respect to both criteria.

Table 14: Optimal BMS with a Frequency and a Severity Component, A Priori Criteria, Variance Principle, One Claim in the First Year of Observation

	Panel A: NBI							Panel B: PIG						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	158.76	111.92	80.74	300.28	129.34	169.16	46.06	155.96	109.95	79.32	294.99	127.06	166.18	45.25
2500	201.06	202.88	737.75	488.29	184.10	288.39	444.79	197.52	199.31	724.76	479.68	180.85	283.31	436.96
3500	248.39	327.54	807.52	597.89	184.10	455.43	938.59	244.01	321.77	793.29	587.35	180.85	447.40	922.05
4500	300.68	446.18	807.98	658.02	184.10	623.02	966.46	295.38	438.32	793.75	646.43	180.85	612.05	949.43

	Panel C: SICH							Panel D: 2C POIS						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	156.69	110.46	79.68	296.36	127.65	166.95	45.46	152.86	107.76	77.74	289.12	124.53	162.87	44.35
2500	198.44	200.23	728.12	481.91	181.69	284.62	438.98	193.59	195.34	710.33	470.13	177.25	277.67	428.26
3500	245.15	323.27	796.97	590.08	181.69	449.48	926.33	239.15	315.37	777.49	575.66	177.25	438.50	903.69
4500	296.75	440.35	797.43	649.43	181.69	614.89	953.84	289.50	429.59	777.94	633.56	177.25	599.86	930.53

	Panel E: 2C NBI (Post. Mean & Post. Variance)							Panel F: 2C NBI (Post. Prob.)						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	162.20	114.35	82.49	306.79	132.14	172.83	47.06	108.88	76.75	55.37	205.93	88.70	116.01	31.59
2500	205.42	207.28	753.74	498.87	188.09	294.64	454.43	137.89	139.13	505.94	334.86	126.25	197.77	305.03
3500	253.77	334.64	825.02	610.85	188.09	465.30	958.93	170.34	224.63	553.78	410.02	126.25	312.33	643.67
4500	307.20	455.85	825.50	672.28	188.09	636.53	987.40	206.20	305.98	554.10	451.26	126.25	427.26	662.78

	Panel G: 2C POIS-NBI													
Claim Size	PAR		2C EXP		2C LNO		2C PAR		2C EXP-LNO		2C EXP-PAR		2C LNO-PAR	
1500	108.81	65.10	76.71	85.13	55.34	78.75	205.81	157.16	88.65	92.60	115.94	100.77	31.57	50.51
2500	137.81	95.94	139.05	148.23	505.65	496.35	334.67	230.77	126.18	123.57	197.66	131.74	304.86	549.72
3500	170.24	132.75	224.50	231.52	553.47	541.32	409.79	279.15	126.18	123.57	312.15	180.23	643.30	700.19
4500	206.08	175.51	305.81	308.03	553.79	541.94	451.01	310.11	126.18	123.57	427.02	239.87	662.41	702.97

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture distributions respectively.
PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture, two component Lognormal-Pareto mixture distributions respectively.

Table 15: Optimal BMS with a Frequency and a Severity Component, A Priori and A Posteriori Criteria, Variance Principle, One Claim in the First Year of Observation

Panel A: NBI								Panel B: PIG						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	96.32	125.96	116.52	232.53	137.01	149.10	74.73	91.07	119.10	110.17	219.86	129.55	140.98	70.66
2500	141.95	219.33	734.40	341.45	182.84	194.93	813.36	134.22	207.37	694.38	322.85	172.88	184.30	769.04
3500	196.41	342.56	800.95	413.03	182.84	266.67	1036.00	185.71	323.89	757.30	390.52	172.88	252.14	979.54
4500	259.68	455.76	801.86	458.84	182.84	354.91	1040.11	245.53	430.92	758.16	433.83	172.88	335.56	983.43

Panel C: SICH								Panel D: 2C POIS						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	93.50	122.29	113.12	225.75	133.01	144.75	72.55	99.07	129.56	119.84	239.17	140.93	153.36	76.87
2500	137.81 1	212.92	712.96	331.49	177.50	189.24	789.62	146.00	225.59	755.37	351.20	188.06	200.49	836.58
3500	90.68	332.56	777.57	400.97	177.50	258.89	1005.76	202.02	352.34	823.82	424.82	188.06	274.29	1065.58
4500	252.10	442.46	778.46	445.44	177.50	344.55	1009.75	267.09	468.78	824.76	471.94	188.06	365.04	1069.81

Panel E: 2C NBI (Post. Mean & Post. Variance)								Panel F: 2C NBI (Post. Prob.)						
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	90.63	118.52	109.63	218.80	128.92	140.29	70.32	65.32	85.42	79.02	157.70	92.92	101.12	50.68
2500	133.57	206.37	691.02	321.28	172.04	183.41	765.31	96.27	148.74	498.05	231.57	124.00	132.19	551.60
3500	184.81	322.32	753.63	388.63	172.04	250.92	974.80	133.20	232.31	543.18	280.10	124.00	180.85	702.59
4500	244.34	428.84	754.49	431.73	172.04	333.94	978.67	176.11	309.09	543.80	311.17	124.00	240.69	705.38

Panel G: 2C POIS-NBI							
Claim Size	PAR	2C EXP	2C LNO	2C PAR	2C EXP-LNO	2C EXP-PAR	2C LNO-PAR
1500	65.10	85.13	78.75	157.16	92.60	100.77	50.51
2500	95.94	148.23	496.35	230.77	123.57	131.74	549.72
3500	132.75	231.52	541.32	279.15	123.57	180.23	700.19
4500	175.51	308.03	541.94	310.11	123.57	239.87	702.97

Note: NBI, PIG, SICH, 2C POIS, 2C NBI and 2C POIS-NBI are the Negative Binomial Type I, Poisson Inverse Gaussian, Sichel, two component Poisson mixture, two component Negative Binomial Type I mixture and two component Poisson-Negative Binomial Type I mixture regression models for location scale shape and mixing probabilities respectively.
PAR, 2C EXP, 2C LNO, 2C PAR, 2C EXP-LNO, 2C EXP-PAR and 2C LNO-PAR are the Pareto, the two component Exponential mixture, the two component Pareto mixture, the two component Lognormal mixture, the two component Exponential-Pareto mixture, the two component Exponential-Lognormal mixture, two component Lognormal-Pareto mixture mixture regression models for location scale shape and mixing probabilities respectively.

5 Conclusions

This paper was mainly concerned with the construction of optimal BMS using two component mixture distributions defined so that the component distributions do not necessarily arise from the same parametric family. Based on this newly proposed framework we were able to present an abundance of model choices that account for unobserved heterogeneity in alternative ways and can be employed by an insurer when deciding on their Bonus-Malus pricing strategies. Specifically, claim frequency was modelled using a 2C Poisson mixture, 2C Negative Binomial Type I mixture, 2C Sichel mixture (2C Poisson Inverse Gaussian mixture and 2C Sichel-Poisson Inverse Gaussian mixture as special cases), 2C Poisson-Negative Binomial Type I mixture, 2C Poisson-Sichel mixture (2C Poisson-Poisson Inverse Gaussian mixture as a special case) and 2C Negative Binomial Type I -Sichel mixture (2C Negative Binomial Type I-Poisson Inverse Gaussian mixture as a special case) distributions. Claim severity was approximated by employing a 2C Exponential mixture, 2C Pareto mixture, 2C Lognormal mixture, 2C Exponential-Pareto mixture, 2C Exponential-Lognormal mixture and 2C Lognormal-Pareto mixture distributions. Also, the Negative Binomial Type I, Sichel, Poisson Inverse Gaussian and Pareto distributions were considered as special cases of the previously mentioned distributions. Extending the framework used by Tzougas, Vrontos and Frangos (2014), all the parameters and mixing probabilities of these models were modelled in terms of risk factors. These models were calibrated employing a Generalized Akaike Information Criterion (GAIC), which is valid for both nested or non-nested model comparisons (see Rigby and Stasinopoulos, 2005 and 2009). On the path towards actuarial relevance the Bayesian view was taken and BMS were derived by updating the posterior probability of policyholders' classes of risk and by updating the posterior mean and the posterior variance. The premium rates were calculated via the expected value and variance principles with independence between the claim frequency and severity components assumed. Extensions to other frequency/severity regression models for location scale, shape and mixing probabilities can be obtained in a similar straightforward way.

A potentially interesting line of further research would be to go through the Bonus-Malus ratemaking exercise when functional forms other than the linear are included, based on the generalized additive models for location scale and shape and prior probabilities approach of Rigby and Stasinopoulos (2005 and 2009). Also see, for example, a recent paper by Klein et al. (2014) in which Bayesian generalized additive models for location, scale and shape claim frequency models are employed for nonlife ratemaking and risk management. Moreover, the proposed modelling framework could be employed with longitudinal data, see, for instance, Boucher et al. (2007).

References

[1] Boucher, J. P., M. Denuit and M. Guillen (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. North American Actuarial Journal, 11, 4, 110-131.

[2] Boucher, J. P., M. Denuit and M. Guillen (2008). Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. Variance, 2, 1, 135-162.

[3] Brouhns, N., M. Guillen, M. Denuit and J. Pinquet (2003). Bonus-malus scales in segmented tariffs with stochastic migration between segments. Journal of Risk and Insurance, 70, 577-599.

[4] Denuit, M., X. Marechal, S. Pitrebois and J. F. Walhin (2007). Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. Wiley.

[5] Dionne, G. and C. Vanasse (1989). A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component. ASTIN Bulletin, 19, 199-212.

[6] Dionne, G. and C. Vanasse (1992). Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics, 7, 149-165.

[7] Dunn, P.K. and G.K. Smyth, (1996). Randomized quantile residuals. Computational and Graphical Statistics 5, 236-245.

[8] Frangos, N. and S. Vrontos (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. ASTIN Bulletin, 31, 1, 1-22.

[9] Gómez, E., A. Hernández and F. Vázquez-Polo (2000). Robust Bayesian premium principles in Actuarial Science. Journal of the Royal Statistical Society, 49, 241-252.

[10] Gómez, E., J. Pérez, A. Hernández and F. Vázquez-Polo (2002). Measuring sensitivity in a bonus–malus system. Insurance: Mathematics & Economics, 31, 105–113.

[11] Gómez-Déniz, E., A. Hernández-Bastida and M.P. Fernández-Sánchez (2014). Computing credibility bonus-malus premiums using the aggregate claims distribution. Hacettepe Journal of Mathematics and Statistics, 43, 6, 1047-1061.

[12] Heilmann, W. (1989). Decision theoretic foundations of credibility theory. Insurance: Mathematics & Economics, 8, 77-95.

[13] Heller, G. Z., M. D. Stasinopoulos, R. A. Rigby and P. de Jong (2007). Mean and dispersion modeling for policy claims costs. Scandinavian Actuarial Journal, 4, 281-292.

[14] Johnson, N. L., S. Kotz, and A. W. Kemp (2005). Univariate Discrete Distributions. Wiley.

[15] Klein, N. , M. Denuit, S. Lang and K. Thomas (2014). Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape. In: Insurance: Mathematics and Economics, 55, 225-249.

[16] Lemaire, J. (1995). Bonus-Malus Systems in Automobile Insurance. Kluwer Academic Publishers.

[17] Picech, L. (1994). The Merit-Rating Factor in a Multiplicating Rate-Making model. ASTIN Colloquium, Cannes.

[18] Pinquet, J. (1997). Allowance for cost of claims in bonus-malus systems, ASTIN Bulletin, 27, 33-57.

[19] Pinquet, J. (1998). Designing Optimal Bonus-Malus Systems From Different Types of Claims. ASTIN Bulletin, 28, 205-220.

[20] Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape, (with discussion). Applied Statistics, 54, 507-554.

[21] Rigby, R. A., and D. M. Stasinopoulos (2009). A flexible regression approach using GAMLSS in R.

[22] Stasinopoulos, D.M., B. Rigby and C. Akantziliotou (2008). Instructions on how to use the gamlss package in R, Second Edition.

[23] Tzougas, G.and N. Frangos (2014). The Design of an Optimal Bonus-Malus System Based on the Sichel Distribution. Collective book: Modern Problems in Insurance Mathematics, Springer Verlag.

[24] Tzougas, G., S. Vrontos and N. Frangos (2014): Optimal Bonus-Malus Systems Using Finite Mixture Models. ASTIN Bulletin, Volume 44, Issue 02, May 2014, pp 417-444.