Optimal Bonus-Malus Systems Using Finite Mixture Models

George Tzougas
Department of Statistics, Athens University of Economics and Business

Spyridon Vrontos
Department of Mathematical Sciences, University of Essex

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

Abstract

This paper presents the design of optimal Bonus-Malus Systems (BMS) using finite mixture models, extending the work of Lemaire (1995) and Frangos and Vrontos (2001). Specifically, for the frequency component we employ a finite Poisson, Delaporte and Negative Binomial mixture, while for the severity component we employ a finite Exponential, Gamma, Weibull and Generalized Beta Type II mixture, updating the posterior probability. We also consider the case of a finite Negative Binomial mixture and a finite Pareto mixture updating the posterior mean. The generalized BMS we propose, integrate risk classification and experience rating by taking into account both the a priori and a posteriori characteristics of each policyholder.

Keywords: Optimal BMS; Claim frequency; Claim severity; Mixtures of distributions; A priori classification criteria; A posteriori classification criteria.

Acknowledgements: We would like to thank the editor and the referees for their constructive comments and suggestions. We also would like to thank the participants at the 6th Conference in Actuarial Science & Finance, University of the Aegean, Samos, Greece. The usual disclaimer applies.

*Athens University of Economics and Business, 76 Patission Str., 104 34, Athens, Greece, Tel: +302108203579, Email: nef@aueb.gr

1 Introduction

A Bonus-Malus System (BMS) penalizes policyholders responsible for one or more claims by a premium surcharge (malus) and rewards the policyholders who had a claim-free year by awarding discount of the premium (bonus). An optimal BMS is financially balanced for the insurer and fair for the policyholder. Optimal BMSs can be broadly derived in two ways; based only on the a posteriori classification criteria and based on both the a priori and the a posteriori classification criteria. Typically, a posteriori classification criteria include the number and severity of individual claims, while a priori classification criteria include variables such as characteristics of the policyholder and the automobile.

Contributions to the literature of BMS include, among others, Lemaire (1995), Dionne and Vanasse (1989, 1992), Picech (1994), Pinquet (1997, 1998), Frangos and Vrontos (2001), Brouhns et al. (2003), Denuit et al. (2007) and Mahmoudvand and Hassani (2009). The literature more closely related to ours is Lemaire (1995), Dionne and Vanasse (1989, 1992), Pinquet (1997) and Frangos and Vrontos (2001). Lemaire (1995) considered the good risk/bad risk model employing a two component Poisson mixture distribution. Dionne and Vanasse (1989, 1992) proposed a BMS that integrates a priori and a posteriori information on an individual basis. Specifically, this generalized BMS is derived as a function of the years that the policyholder is in the portfolio, the number of accidents and their individual characteristics. Pinquet (1997) proposed the design of a BMS on the pure premium of insurance contracts from Poisson and Lognormal distributions on the number and cost of claims, with a joint distribution on the random effects. Frangos and Vrontos (2001) considered an optimal BMS based on both the number of accidents and their severity employing Negative Binomial and Pareto distributions/regression models by updating the posterior mean of the claim frequency and severity respectively.

Our first contribution is the development of an optimal BMS that takes into account the number of claims of each policyholder and the exact size of loss that these claims incurred using various finite mixtures of distributions. For the frequency component we assume that the number of claims is distributed according to a finite Poisson, Delaporte and Negative Binomial mixture, and for the severity component we consider that the losses are distributed according to a finite Exponential, Gamma, Weibull and Generalized Beta Type II (GB2) mixture. In this way we expand the setup that Lemaire (1995) used to design an optimal BMS based on the number of claims. Applying Bayes theorem we derive the posterior probability of the policyholder's classes of risk. Furthermore, we extend the setup of Frangos and Vrontos (2001) for Negative Binomial and Pareto mixtures and derive the posterior distributions of the mean claim frequency and the mean claim size given the information we have for the claim frequency and severity for each policyholder for the period they are in the portfolio. Our third contribution is the development of a generalized BMS that integrates the a priori and the a posteriori information on a individual basis, extending the framework developed by Dionne and Vanasse (1989, 1992) and Frangos and Vrontos (2001). This is achieved by using finite mixtures of generalized linear models. In this generalized BMS, the premium is a function of the years the policyholder is in the portfolio, the number and size of loss of their accidents, and the significant a priori rating variables for the number of accidents and their severity.

The layout of the paper is as follows. Section 2 introduces the finite mixture models we use. Section 3 describes the design of optimal BMS by updating the posterior probability and Section 4 by updating the posterior mean. Section 5 contains an application to a data set concerning car-insurance claims at fault. Finally, Section 6 concludes the paper.

2 Finite Mixture Models

Finite mixture models are a popular statistical modelling technique since they constitute a flexible and easily extensible model class for approximating general distribution functions in a semi-parametric way and accounting for unobserved heterogeneity. Finite mixture models have been widely applied in many areas, such as biology, biometrics, genetics, medicine and marketing. However, as we have mentioned, with the exception of Lemaire (1995), these models have not been extensively studied in BMS literature.

In what follows we present a short summary of the main characteristics of finite mixture models. McLachlan and Peel (2000) provide a detailed treatment of finite mixture models and their applications. We consider the random variable Y having probability density function

$$f_Y(y) = \sum_{z=1}^{n} \pi_z f_z(y), \tag{1}$$

where $0 \leq \pi_z \leq 1$ is the prior (or mixing) probability of component z and $\sum_{z=1}^{n} \pi_z = 1$, $z = 1, .., n$.

The simplest finite mixture models are finite mixtures of distributions which are used for model-based clustering. In this case the model is given by a convex combination of a finite number of different distributions where each of the distributions is referred to as component.

An extension is to estimate finite mixture models assuming that the n components $f_z(y)$ can be represented by generalized additive models for location scale and shape, GAMLSS (see Rigby and Stasinopoulos, 2005 and 2009). In this setup the probability density function $f_z(y)$ for component z depends on $\boldsymbol{\theta}_z$, a vector of the distribution parameters, each of which can be a function of the explanatory variables $\mathbf{x}_z$, i.e. $f_z(y) = f_z(y|\boldsymbol{\theta}_z, \mathbf{x}_z)$. Similarly, $f_Y(y)$ depends on parameters $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\pi})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and $\boldsymbol{\pi}^T = (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_n)$ and explanatory variables $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, i.e. $f_Y(y) = f_Y(y|\boldsymbol{\psi}, \mathbf{x})$, and

$$f_Y(y|\boldsymbol{\psi}, \mathbf{x}) = \sum_{z=1}^{n} \psi_z f_z(y|\boldsymbol{\theta}_z, \mathbf{x}_z). \tag{2}$$

Note that the prior probabilities may also depend on explanatory variables $\mathbf{x}_0$ and parameters through a multinomial logistic model (for more information, refer to Rigby and Stasinopoulos, 2009). In this study, we assume that all the component distributions, $f_z(y)$, arise from the same parametric distribution family and the covariates are only included in the linear predictor for the mean parameter (of claim frequency or severity). Using this formulation, the heterogeneity in the data can be accounted for in two ways. Firstly, the population heterogeneity is accounted for by choosing a finite number of 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 a finite number of support points. Secondly, depending on the choice of the $f_z(y)$ distribution, heterogeneity can also be accommodated within each component by including the explanatory variables in the mean function.

The log-likelihood function can be maximized with respect to $\boldsymbol{\psi}$, i.e. with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$, using the EM algorithm (for more details see Rigby and Stasinopoulos, 2009).

2.1 Frequency Component

Mixtures of count distributions can be considered as counterparts or generalizations to the simple count distribution (e.g. Poisson) for the description of non-homogeneous populations, consisting of a finite number of homogeneous sub-populations. In these cases the probability distribution (pdf) of the population can be regarded as a finite mixture of count distributions (e.g. finite mixture of Poisson distributions). The use of mixture distributions is justified when the portfolio is considered to be heterogeneous, consisting of n categories of policyholders classified according to their driving skills. In this respect, we have fractions of drivers π_z where the risk that each policyholder of category z is imposing to the pool, with respect to their claim frequency is denoted by λ_z, $z = 1, \cdots, n$. The distribution of the number of claims k in each category is denoted by $P_z(k)$. Thus, the structure function is an n-point discrete distribution and the unconditional distribution of the number of claims, denoted by $P(k)$, is given by

$$P(k) = \sum_{z=1}^{n} \pi_z P_z(k), \tag{3}$$

for $k = 0, 1, 2, 3, ..., \pi_z, \lambda_z > 0$, and $\sum_{z=1}^{n} \pi_z = 1$. The expected value of the number of claims is equal to $E(k) = \sum_{z=1}^{n} \pi_z \lambda_z$ and its variance is equal to $Var(k) = E[Var(k|\pi)] + Var[E(k|\pi)] = E_\pi\left[\sigma_k^2\right] + Var_\pi[\lambda]$.

In this study we model the claim frequency using a finite Poisson, Delaporte and Negative Binomial mixture.

- In the case of the finite Poisson mixture Eq.(3) is equal to

$$P(k) = \sum_{z=1}^{n} \pi_z \frac{e^{-\lambda_z} \lambda_z^k}{k!}. \tag{4}$$

- In the case of the finite Negative Binomial[1] mixture distribution Eq.(3) is equal to

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

$$P(k)=\sum_{z=1}^{n}\pi_z\binom{k+\frac{1}{\alpha_z}-1}{k}\left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{1}{\alpha_z}}\left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^{k},\alpha_z>0. \tag{5}$$

- In the case of the finite Delaporte[2] mixture distribution Eq.(3) is equal to

$$\begin{aligned} P(k) &= \sum_{z=1}^{n}\pi_z\frac{e^{-\lambda_z\nu_z}}{\Gamma\left(\frac{1}{\sigma_z}\right)}\left(1+\lambda_z\sigma_z\left(1-\nu_z\right)\right)^{-\frac{1}{\sigma_z}}S,\sigma_z>0,0<\nu_z<1,\\ S &= \sum_{m=0}^{k}\binom{k}{m}\frac{\left(\lambda_z\right)^k\left(\nu_z\right)^{k-m}}{k!}\left(\lambda_z+\frac{1}{\sigma_z\left(1-\nu_z\right)}\right)^{-m}\Gamma\left(\frac{1}{\sigma_z}+m\right). \end{aligned} \tag{6}$$

2.2 Severity Component

Let us consider now the severity component. The portfolio is considered to be heterogeneous and we have fractions of drivers ρ_z where the risk (with respect to the mean claim size) that each policyholder of category z is imposing to the pool, $z=1,\cdots,n$, is denoted by y_z. The pdf of the claim size x in each category is denoted by $f_z(x)$. Thus, the structure function is an n-point discrete distribution and the unconditional distribution of claim size, denoted by $f(x)$, is given by

$$f(x)=\sum_{z=1}^{n}\rho_z f_z(x) \tag{7}$$

for $x,\rho_z,y_z>0$ and $\sum_{z=1}^{n}\rho_z=1$. The expected value of the claim size is equal to $E(x)=\sum_{z=1}^{n}\rho_z y_z$ and its variance is equal to $Var(x)=E\left[Var(x|\rho)\right]+Var[E(x|\rho)]=E_\rho\left[\sigma_x^2\right]+Var_\rho[y]$.

In this paper we model the costs of claims using a finite mixture of Exponential, Gamma, Weibull and GB2 distributions.

- In the case of the finite mixture of Exponential Eq.(7) is equal to

$$f(x)=\sum_{z=1}^{n}\rho_z\frac{e^{-\frac{x}{y_z}}}{y_z}. \tag{8}$$

- In the case of the finite mixture of Gamma[3] Eq.(7) is equal to

$$f(x)=\sum_{z=1}^{n}\rho_z\frac{1}{\left[\theta_z^2y_z\right]^{\frac{1}{\theta_z^2}}}\frac{x^{\frac{1}{\theta_z^2}-1}e^{-\frac{x}{\theta_z^2y_z}}}{\Gamma\left(\frac{1}{\theta_z^2}\right)},\theta_z>0. \tag{9}$$

- In the case of the finite mixture of Weibull[4] Eq.(7) is equal to

$$f(x)=\sum_{z=1}^{n}\rho_z\frac{\theta_z}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\left[\frac{x}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z-1}e^{-\left[\frac{x}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z}},\theta_z>0. \tag{10}$$

- In the case of the finite mixture of GB2[5] Eq.(7) is equal to

$$f(x)=\sum_{z=1}^{n}\rho_z|\sigma_z|x^{\sigma_z\nu_z-1}\left\{y_z^{\sigma_z\nu_z}B(\nu_z,s_z)\left[1+\left(\frac{x}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1},\nu_z,s_z>0,-\infty<\sigma_z<\infty. \tag{11}$$

In this case the mean is given by

$$E(x)=\sum_{z=1}^{n}\rho_z y_z\frac{B(\nu_z+\frac{1}{\sigma_z},s_z-\frac{1}{\sigma_z})}{B(\nu_z,s_z)}.$$

[2] The Delaporte distribution can be alternatively employed for modeling the number of claims when we deal with overdispersed count data. We use the parameterization of Delaporte given by Rigby and Stasinopoulos (2009).

[3] The pdf is obtained by setting $\theta^2=\frac{1}{\alpha}$ and $y=\alpha\beta$ using the reparameterization of Johnson et al. (1994). See also Rigby and Stasinopoulos (2009).

[4] We use the reparameterization of Weibull III given by Rigby and Stasinopoulos (2009).

[5] We use the reparameterization of GB2 given by McDonald and Xu (1995), McDonald (1996) and Rigby and Stasinopoulos (2009).

3 The Optimal BMS Derived by Updating the Posterior Probability

We assume that the number of claims of each policyholder is independent from the severity of each claim so as to deal with the frequency and severity components separately. In Section 3.1 we consider the design of an optimal BMS based on the a posteriori criteria while in Section 3.2 we present a generalized BMS based on both the a priori and a the posteriori criteria.

3.1 A BMS Based on the a Posteriori Criteria

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).

3.1.1 Frequency Component

In what follows, we consider a policyholder with claims history $k_1, ..., k_t$, where k_j is the number of claims the policyholder had in year j, $j = 1, ..., t$. Let us denote by $K = \sum_{j=1}^{t} k_j$ the total number of claims they had in t years and with R_l the risk, imposed on the insurance company, associated by the lth category of policyholders. Moreover, the posterior probability of the policyholder belonging to the lth category is denoted by $\pi_l (k_1, \cdots, k_t)$. Applying Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by

$$\pi_l (k_1, ..., k_t) = \frac{P(k_1, ..., k_t|R_l)\pi_l}{\sum\limits_{z=1}^{n} P(k_1, ..., k_t|R_z)\pi_z}. \tag{12}$$

In this way, we update the posterior probability of belonging in category l given the information we have for the claim history of the policyholder. The setup we described above is applied to the finite Poisson, Negative Binomial and Delaporte mixture distributions and the following results are obtained:

- In the case of the finite Poisson mixture distribution Eq.(12) becomes

$$\pi_l (k_1, ..., k_t) = \frac{\lambda_l^K e^{-t\lambda_l}\pi_l}{\sum\limits_{z=1}^{n} \lambda_z^K e^{-t\lambda_z}\pi_z}. \tag{13}$$

- In the case of the finite NBI mixture distribution Eq.(12) becomes

$$\pi_l (k_1, ..., k_t) = \frac{\prod\limits_{j=1}^{t} \binom{k_j + \frac{1}{\alpha_l} - 1}{k_j} \left(\frac{1}{1+\alpha_l\lambda_l}\right)^{\frac{t}{\alpha_l}} \left(\frac{\alpha_l\lambda_l}{1+\alpha_l\lambda_l}\right)^K \pi_l}{\sum\limits_{z=1}^{n}\prod\limits_{j=1}^{t} \binom{k_j + \frac{1}{\alpha_z} - 1}{k_j} \left(\frac{1}{1+\alpha_z\lambda_z}\right)^{\frac{t}{\alpha_z}} \left(\frac{\alpha_z\lambda_z}{1+\alpha_z\lambda_z}\right)^K \pi_z}. \tag{14}$$

- In the case of the finite Delaporte mixture distribution Eq.(12) becomes

$$\pi_l (k_1, ..., k_t) = \frac{\left[\frac{e^{-t\lambda_l\nu_l}}{\left[\Gamma\left(\frac{1}{\sigma_l}\right)\right]^t} \left(1 + \lambda_l\sigma_l\left(1-\nu_l\right)\right)^{-\frac{t}{\sigma_l}} \prod\limits_{j=1}^{t} S_{j,l}\right]\pi_l}{\sum\limits_{z=1}^{n}\left[\frac{e^{-t\lambda_z\nu_z}}{\left[\Gamma\left(\frac{1}{\sigma_z}\right)\right]^t} \left(1 + \lambda_z\sigma_z\left(1-\nu_z\right)\right)^{-\frac{t}{\sigma_z}} \prod\limits_{j=1}^{t} S_{j,z}\right]\pi_z}, \tag{15}$$

$$S_{j,l} = \sum_{m=0}^{k_j} \binom{k_j}{m} \frac{(\lambda_l)^{k_j} (\nu_l)^{k_j - m}}{k_j!} \left(\lambda_l + \frac{1}{\sigma_l\left(1-\nu_l\right)}\right)^{-m} \Gamma\left(\frac{1}{\sigma_l} + m\right).$$

It should be noted that due to the existence of k_j in Eq.(14) and Eq.(15), the explicit claim frequency history determines the calculation of the posterior probabilities and thus of premium rates and not just the total number of claims as in the case of the two component Poisson mixture.

Under a quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}$ for a policyholder with claim history $k_1, ..., k_t$ is the mean of the posterior structure function given by

$$\hat{\lambda}_{t+1}\left(k_1, ..., k_t\right) = \sum_{z=1}^{n} \pi_z\left(k_1, \cdots, k_t\right) \lambda_z. \tag{16}$$

3.1.2 Severity Component

We assume again that a policyholder stays in the portfolio for t years and that the number of claims in year j is denoted by k_j, the total number of claims in t years is denoted, as before, by K and the claim amount for the kth claim is denoted by x_k. In such a case the information we have for their claim size history will be in the form of a vector $x_1, ..., x_K$. The risk imposed on the pool by the policyholder who belongs to the lth category of policyholders based on the severity of their claims is denoted by Q_l. Then the posterior probability of the policyholder belonging to the lth category is given by $\rho_l\left(x_1, ..., x_K\right)$. In order to design an optimal BMS that accounts for each claim amount, we must find the posterior probability of belonging in each risk class, given the information we have about the claim size history for each policyholder for the period they are in the portfolio. Applying Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by

$$\rho_l\left(x_1, ..., x_K\right) = \frac{f\left(x_1, ..., x_K | Q_l\right) \rho_l}{\sum_{z=1}^{n} f\left(x_1, ..., x_K | Q_z\right) \rho_z}. \tag{17}$$

Thus, we update the posterior probability of belonging in category l given the information we have for the claim size history of the policyholder. The framework we introduced above is applied to finite Exponential, Weibull, Gamma and GB2 mixture distributions.

- In the case of the finite mixture of Exponential Eq.(17) becomes

$$\rho_l\left(x_1, ..., x_K\right) = \rho_l \frac{e^{-\frac{\sum_{k=1}^{K} x_k}{y_l}}}{y_l^K} \left[\sum_{z=1}^{n} \rho_z \frac{e^{-\frac{\sum_{k=1}^{K} x_k}{y_z}}}{y_z^K}\right]^{-1}. \tag{18}$$

- In the case of the finite mixture of Gamma Eq.(17) becomes

$$\rho_l\left(x_1, ..., x_K\right) = \frac{\rho_l \left[\left(\theta_l^2 y_l\right)^{\frac{1}{\theta_l^2}} \Gamma\left(\frac{1}{\theta_l^2}\right)\right]^{-K} \left(\prod_{j=1}^{K} x_j\right)^{\frac{1}{\theta_l^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_l^2 y_l}}}{\sum_{z=1}^{n} \rho_z \left[\left(\theta_z^2 y_z\right)^{\frac{1}{\theta_z^2}} \Gamma\left(\frac{1}{\theta_z^2}\right)\right]^{-K} \left(\prod_{j=1}^{K} x_j\right)^{\frac{1}{\theta_z^2}-1} e^{-\frac{\sum_{j=1}^{K} x_j}{\theta_z^2 y_z}}}. \tag{19}$$

- In the case of the finite mixture of Weibull Eq.(17) becomes

$$\rho_l\left(x_1, ..., x_K\right) = \frac{\rho_l \theta_l^K \left[\frac{\Gamma\left(\frac{1}{\theta_l}+1\right)}{y_l}\right]^{K\theta_l} \prod_{j=1}^{K} x_j^{\theta_l - 1} e^{-\sum_{j=1}^{K}\left[\frac{x_j}{y_l}\Gamma\left(\frac{1}{\theta_l}+1\right)\right]^{\theta_l}}}{\sum_{z=1}^{n} \rho_z \theta_z^K \left[\frac{\Gamma\left(\frac{1}{\theta_z}+1\right)}{y_z}\right]^{K\theta_z} \prod_{j=1}^{K} x_j^{\theta_z - 1} e^{-\sum_{j=1}^{K}\left[\frac{x_j}{y_z}\Gamma\left(\frac{1}{\theta_z}+1\right)\right]^{\theta_z}}} \tag{20}$$

- In the case of the finite mixture of GB2 Eq.(17) becomes

$$\rho_l\left(x_1,...,x_K\right)=\frac{\left(\frac{|\sigma_l|}{y_l^{\sigma_l\nu_l}B(\nu_l,s_l)}\right)^K\left(\prod_{j=1}^{K}x_j\right)^{\sigma_l\nu_l-1}\left\{\prod_{j=1}^{K}\left[1+\left(\frac{x_j}{y_l}\right)^{\sigma_l}\right]^{\nu_l+s_l}\right\}^{-1}\rho_l}{\sum_{z=1}^{n}\left(\frac{|\sigma_z|}{y_z^{\sigma_z\nu_z}B(\nu_z,s_z)}\right)^K\left(\prod_{j=1}^{K}x_j\right)^{\sigma_z\nu_z-1}\left\{\prod_{j=1}^{K}\left[1+\left(\frac{x_j}{y_z}\right)^{\sigma_z}\right]^{\nu_z+s_z}\right\}^{-1}\rho_z}. \tag{21}$$

Using the quadratic error loss function the optimal choice of y_{t+1} for a policyholder with claim size history $x_1,...,x_K$, in t years is the mean of the posterior structure function, that is

$$\hat{y}_{t+1}\left(x_1,...,x_K\right)=\sum_{z=1}^{n}y_z\rho_z\left(x_1,...,x_K\right), \tag{22}$$

for the case of the finite mixture of Exponential, Gamma and Weibull, while for the case of GB2 it is

$$\hat{y}_{t+1}\left(x_1,...,x_K\right)=\sum_{z=1}^{n}\rho_z\left(x_1,...,x_K\right)y_z\frac{B(\nu_z+\frac{1}{\sigma_z},s_z-\frac{1}{\sigma_z})}{B(\nu_z,s_z)}. \tag{23}$$

3.2 A BMS Based Both on the a Priori and the a Posteriori Criteria

In the models discussed above, the characteristics of each policyholder and the insured vehicle are not taken into consideration for the calculation of $\hat{\lambda}_{t+1}$ and $\hat{y}_{t+1}$. In this way the premiums do not vary simultaneously with other variables that affect the claim frequency and/or the claim severity distributions. For this purpose in 3.2.1 and 3.2.2 we employ finite mixture generalized linear models. Specifically, using the exponential link function we allow the parameters of the claim frequency distribution, λ_z, and severity distribution, y_z, presented in Sections 2.1 and 2.2 respectively, to be modelled as functions of the significant a priori rating variables for the number and costs of claims.

3.2.1 Frequency Component

Consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j are independent. If we assume that the portfolio consists of n categories of policyholders classified with respect to the risk they impose on the pool and that the observations K_i^j follow an n-point discrete finite mixture of Poisson, Negative Binomial and Delaporte, given by Eq.(4), Eq.(5) and Eq.(6) respectively, then we have fractions of policyholders π_z, $z=1,...,n$, with mean claim frequency, denoted as $\lambda_{z,i}^j$. The vector parameter $\lambda_{z,i}^j$ can be modelled as $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$, where $c_{z,i}^j\left(c_{z,i,1}^j,...,c_{z,i,h}^j\right)$ is the vector of h individual characteristics[6], which correspond to different a priori rating variables and β_z^j is the vector of coefficients. The use of the exponential link function ensures the non-negativity of $\lambda_{z,i}^j$. Let us denote with $K=\sum_{j=1}^{t}K_i^j$ the total number of claims of policyholder i in t years and R_l the risk the policyholder imposes on the pool if we assume that she belongs to the lth category of policyholders. Applying Bayes theorem the posterior probability $\pi_l\left(K_i^1,...,K_i^t;c_{l,i}^1,...,c_{l,i}^{t+1}\right)$ that the policyholder belongs to the lth category is given by Eqs(13, 14 and 15) for the case of finite Poisson, Negative Binomial and Delaporte mixture regression models respectively, by letting $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$. Moreover, the optimal choice of $\hat{\lambda}_i^{t+1}\left(K_i^1,...,K_i^t;c_{z,i}^1,...,c_{z,i}^{t+1}\right)$ under the quadratic loss function will be the mean of the posterior structure function, given by letting $\lambda_{z,i}^j=\exp\left(c_{z,i}^j\beta_z^j\right)$ in Eq.(16), for the case of the Poisson, Negative Binomial and Delaporte respectively.

3.2.2 Severity Component

Let us consider now the severity component. Consider a policyholder i with an experience of t periods. whose number of claims for period j are independent and are denoted as $K_i^j=k$, her total number

[6] All the characteristics we consider are observable.

of claims over t periods is denoted as K and by $X_{i,k}^{j}$ is denoted the loss incurred from her claim k for the period j. Then, the information we have for her claim size history will be in the form of a vector $X_{i,1}, X_{i,2}, ..., X_{i,K}$ and the total claim amount for the specific policyholder over the t periods that she is in the portfolio will be equal to $\sum_{k=1}^{K} X_{i,k}$. We assume that the portfolio consists of n categories of drivers based on their claims severity and their losses follow an n-point discrete finite mixture of Exponential, Gamma, Weibull and GB2, given by Eqs(8, 9, 10 and 11) respectively. Furthermore, we consider that the expected claim severity, $y_{z,i}^{j}$, is a function of the vector $d_{z,i}^{j}\left(d_{z,i,1}^{j}, ..., d_{z,i,h}^{j}\right)$ of h individual characteristics which are represented using different a priori rating variables. Specifically, assume that $y_{z,i}^{j} = \exp\left(d_{z,i}^{j}\gamma_{z}^{j}\right)$, for $z = 1, ..., n$, non-negativity of $y_{z,i}^{j}$ is implied from the exponential function and γ_{z}^{j} is the vector of the coefficients. Let us denote as Q_l the risk that is imposed on the insurance company assuming that the policyholder belongs to the lth category of drivers sorted by the amount of loss that their accidents produce. In order to design an optimal BMS that will take into account the size of loss of each claim, we must estimate the posterior probability $\rho_l\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{l,i}^{1}, ..., d_{l,i}^{t+1}\right)$ of belonging in each risk class given the information we have about the claim size history for each policyholder for the period they are in the portfolio. Applying Bayes theorem, the posterior probability of the policyholder belonging to the lth category is given by Eqs(18, 19, 20 and 21) for the case of finite Exponential, Gamma, Weibull and GB2 mixture regression models respectively, by employing $y_{z,i}^{j} = \exp\left(d_{z,i}^{j}\gamma_{z}^{j}\right)$. The optimal choice of $\hat{y}_{i}^{t+1}\left(X_{i,1}, X_{i,2}, ..., X_{i,K}; d_{z,i}^{1}, ..., d_{z,i}^{t+1}\right)$ under the quadratic error loss function will be the mean of the posterior structure function, given by letting $y_{z,i}^{j} = \exp\left(d_{z,i}^{j}\gamma_{z}^{j}\right)$ in Eqs(22 and 23), for the case of the Exponential, Gamma and Weibull and GB2 respectively.

4 The Optimal BMS Derived by Updating the Posterior Mean

We model the number and cost of claims using the finite Negative Binomial and Pareto mixture models respectively. The optimal BMS we propose is derived by updating the mean claim frequency and the mean claim severity, assuming independence between the two components. Our analysis in Section 4.1 is based only on the a posteriori criteria while in Section 4.2 is based both on the a priori and a posteriori criteria.

4.1 A BMS Based on the a Posteriori Criteria

Generalizing the structure used by Lemaire (1995) and Frangos and Vrontos (2001) we present an optimal BMS derived by updating the posterior mean.

4.1.1 Finite Negative Binomial Mixture Distribution

As previously, the portfolio is considered to be heterogeneous and all policyholders have constant over time but unequal underlying risks of having an accident. We assume that the number of claims given the mean claim frequency, $k|\lambda$, is distributed as a Poisson(λ) random variable and that the structure function follows an n-component mixture of Gamma distributions which has a pdf of the form

$$u(\lambda) = \sum_{z=1}^{n} \pi_z \frac{\lambda^{\alpha_z - 1}\tau_z^{\alpha_z}\exp(-\tau_z\lambda)}{\Gamma(\alpha_z)},$$

$\lambda, \alpha_z, \tau_z > 0, \sum_{z=1}^{n} \pi_z = 1$, with mean $E(\lambda) = \sum_{z=1}^{n} \pi_z \frac{\alpha_z}{\tau_z}$. Then the unconditional distribution of the number of claims k is an n-component mixture of Negative Binomial distributions[7] with probability density function

[7] Note that the finite Negative Binomial(τ_z, α_z) mixture, derived by updating the posterior mean, is given from a reparameterization of the pdf of the finite NBI(λ_z, σ_z) mixture, derived by updating the posterior probability, if we let $\lambda_z = \frac{\alpha_z}{\tau_z}$ and $\sigma_z = \frac{1}{\alpha_z}$, for $z = 1, ..., n$.

$$P(k)=\sum_{z=1}^{n}\pi_z\binom{k+\alpha_z-1}{k}p_z^{\alpha_z}q_z^k,\ p_z=\left(\frac{\tau_z}{1+\tau_z}\right),q_z=\left(\frac{1}{1+\tau_z}\right). \tag{24}$$

Consider a policyholder with claim history $k_1,...,k_t$, where k_j is the number of claims that the policyholder had in year $j, j=1,...,t$. Let K denote the total number of claims in t years. Applying Bayes theorem, one can find that the posterior structure function, $u(\lambda|k_1,...,k_t)$, for a policyholder or a group of policyholders with claim history $k_1,...,k_t$ is given by

$$u(\lambda|k_1,...,k_t)=\sum_{z=1}^{n}\pi_z\frac{(\tau_z+t)^{K+\alpha_z}\lambda^{K+\alpha_z-1}e^{-(\tau_z+t)\lambda}}{\Gamma(\alpha_z+K)}, \tag{25}$$

which is the pdf of an n-component mixture of Gamma densities. Consequently, using the quadratic error loss function, the optimal choice of $\hat{\lambda}_{t+1}$ for a policyholder with claim history $k_1,...,k_t$ is the mean of the posterior structure function, that is

$$\hat{\lambda}_{t+1}(k_1,...,k_t)=\sum_{z=1}^{n}\pi_z\frac{K+\alpha_z}{\tau_z+t}. \tag{26}$$

4.1.2 Finite Pareto Mixture Distribution

We consider a heterogeneous portfolio with respect to the mean claim size of each policyholder. Assume that the claim severity given the mean claim severity, $x|y$, is distributed according to an Exponential(x) distribution and that the structure function follows an n-component mixture of Inverse Gamma distributions, with pdf given by

$$g(y)=\sum_{z=1}^{n}\rho_z\frac{\frac{1}{m_z}\exp\left(-\frac{m_z}{y}\right)}{\left(\frac{y}{m_z}\right)^{s_z+1}\Gamma(s_z)},$$

$y>0, s_z>0, m_z>0, \sum_{z=1}^{n}\rho_z=1$, with mean $E(y)=\sum_{z=1}^{n}\rho_z\frac{m_z}{s_z-1}$. Then the unconditional distribution of the claim severity x will be an n-component mixture of Pareto distributions with pdf

$$f(x)=\sum_{z=1}^{n}\rho_z s_z m_z^{s_z}(x+m_z)^{-s_z-1}. \tag{27}$$

Consider that a policyholder stays in the portfolio for t years with claims in year j, k_j, and their total number of claims in t years is denoted by K. As before by x_k is denoted the claim amount for the kth claim. Then, the information we have for their claim size history will be in the form of a vector $x_1,...,x_K$ and the total claim amount for that specific policyholder over the t years that they are in the portfolio will be equal to $\sum_{k=1}^{K}x_k$. Applying Bayes theorem, we find that the posterior structure function of the mean claim size y, given the policyholder's claim size history $x_1,...,x_K$, denoted as $g(y|x_1,...,x_K)$, is given by

$$g(y|x_1,...,x_K)=\sum_{z=1}^{n}\rho_z\frac{\left(m_z+\sum_{k=1}^{K}x_k\right)^{K+s_z}e^{-\frac{\left(m_z+\sum_{k=1}^{K}x_k\right)}{y}}}{y^{K+s_z+1}\Gamma(K+s_z)}, \tag{28}$$

which is the pdf of an n-component mixture of Inverse Gamma densities. Consequently, using the quadratic error loss function, the optimal choice of $\hat{y}_{t+1}$ for a policyholder with claim size history $x_1,...,x_K$ is the mean of the posterior structure function, that is

$$\hat{y}_{t+1}(x_1,...,x_K)=\sum_{z=1}^{n}\rho_z\frac{m_z+\sum_{k=1}^{K}x_k}{K+s_z-1}. \tag{29}$$

4.2 A BMS Based Both on the a Priori and the a Posteriori Criteria

In this case, the generalized BMS obtained for the frequency component will be derived as a generalization of the structure proposed by Dionne and Vanasse (1989, 1992) and the generalized system obtained for the severity component will be derived as a generalization of the setup used by Frangos and Vrontos (2001).

4.2.1 Finite Negative Binomial Mixture Regression Model

Consider a policyholder i with an experience of t periods whose number of claims for period j, denoted as K_i^j are independent. We assume that K_i^j follows Poisson distribution with parameter λ^j. We consider a heterogeneous portfolio of n categories of policyholders, with expected number of claims of the individual i who belongs to the zth category denoted as $\lambda_{z,i}^j$, $z = 1, ..., n$. We allow λ_z^j vary from one individual to another. Let $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$, where $c_{z,i}^j \left(c_{z,i,1}^j, ..., c_{z,i,h}^j\right)$ is the vector of h individual characteristics and β_z^j is the vector of the coefficients. The conditional to $c_{z,i}^j$ probability, $P\left(K_i^j = k | c_{z,i}^j\right)$, that policyholder i will be involved in k accidents during the period j is given by Eq.(4), for $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$.

For the determination of the expected number of claims in this model we assume that the h individual characteristics provide enough information. However, if one assumes that the a priori rating variables do not contain all the significant information for the expected number of claims then a random variable ε_i must be introduced into the regression component. According to Gourieroux, Montfort and Trognon (1984 a), (1984 b) we can write

$$\lambda_i^j = \exp\left(c_{z,i}^j \beta_z^j + \varepsilon_i\right) = \exp\left(c_{z,i}^j \beta_z^j\right) u_i,$$

where $u_i = \exp(\varepsilon_i)$, yielding a random λ_i^j. Assume that u_i follows an $n-$component gamma mixture distribution with pdf

$$\upsilon(u_i) = \sum_{z=1}^{n} \pi_z \frac{u_i^{\frac{1}{\alpha_z}-1} \frac{1}{\alpha_z}^{\frac{1}{\alpha_z}} \exp\left(-\frac{1}{\alpha_z} u_i\right)}{\Gamma\left(\frac{1}{\alpha_z}\right)},$$

$u_i > 0, \alpha_z > 0$ for $z = 1, ..., n, \sum_{z=1}^{n} \pi_z = 1$ with mean $E(u_i) = 1$. Under this assumption the conditional distribution of $K_i^j | c_{z,i}^j$ becomes an n-component Negative Binomial mixture distribution with pdf given by Eq.(5), for $\lambda_{z,i}^j = \exp\left(c_{z,i}^j \beta_z^j\right)$. Then the posterior distribution of the mean claim frequency λ_i^{t+1} for an individual i observed over $t+1$ periods with $K_i^1, ..., K_i^t$ claim history and $c_{i,z}^1, ..., c_{i,z}^{t+1}$ characteristics is obtained using Bayes theorem and is given by an n-component Gamma mixture with updated parameters $\frac{1}{\alpha_z} + K$ and $S_{i,z}^j$, with pdf

$$f\left(\lambda_i^{t+1} | K_i^1, ..., K_i^t; c_{i,z}^1, ..., c_{i,z}^t\right) = \sum_{z=1}^{n} \pi_z \frac{\left(S_{i,z}^j\right)^{K+\frac{1}{\alpha_z}} \left(\lambda_i^{t+1}\right)^{K+\frac{1}{\alpha_z}-1} \exp\left[-S_{i,z}^j \lambda_i^{t+1}\right]}{\Gamma\left(\frac{1}{\alpha_z} + K\right)},$$

where $S_{i,z}^j = \frac{\frac{1}{\alpha_z} + \sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}{\exp\left(c_{z,i}^{t+1} \beta_z^{t+1}\right)}$ with $u_i > 0, \alpha_z > 0$ and $\sum_{z=1}^{n} \pi_z = 1$. Using the quadratic loss function, one can find that the optimal estimator of λ_i^{t+1} is the mean of the posterior structure function given by

$$\hat{\lambda}_i^{t+1}\left(K_i^1, ..., K_i^t; c_{i,z}^1, ..., c_{i,z}^{t+1}\right) = \sum_{z=1}^{n} \pi_z \exp\left(c_{z,i}^{t+1} \beta_z^{t+1}\right) \left[\frac{\frac{1}{\alpha_z} + \sum_{j=1}^{t} K_i^j}{\frac{1}{\alpha_z} + \sum_{j=1}^{t} \exp\left(c_{z,i}^j \beta_z^j\right)}\right]. \quad (30)$$

4.2.2 Finite Pareto Mixture Regression Model

Consider a policyholder i with an experience of t periods. Assume that the number of claims of the individual i for period j are independent and is denoted by K_i^j and by $X_{i,k}^j$ is denoted the loss incurred from their claim k for the period j. We consider that $X_{i,k}^j$ follows the Exponential distribution with mean claim severity for period j, y^j. We allow the y^j parameter to vary from one individual to another, assuming that the expected claim severity $y_{z,i}^j$ is a function of the vector $d_{z,i}^j \left(d_{z,i,1}^j, ..., d_{z,i,h}^j\right)$ of h individual characteristics, which are represented using different a priori rating variables. Since policyholders have different mean claim severity, it is fair for each policyholder to pay a premium proportional to the risk they impose on the pool. Specifically, we assume that $y_{z,i}^j = \exp\left(d_{z,i}^j \gamma_z^j\right)$, where γ_z^j is the vector of the coefficients. Then, the conditional to $d_{z,i}^j$ pdf of the claim size x, $f\left(x|d_{z,i}^j\right)$, for a claim k of a policyholder i in period j, is given by Eq.(8), employing $y_{z,i}^j = \exp\left(d_{z,i}^j \gamma_z^j\right)$.

For the determination of the expected claim severity in this model we assume that the h individual characteristics provide enough information. Nevertheless, if one assumes that the a priori rating variables do not contain all the significant information for the mean claim severity then a random variable ξ_i must be introduced into the regression component. Thus we can write

$$y_i^j = \exp\left(d_{z,i}^j \gamma_z^j + \xi_i\right) = \exp\left(d_{z,i}^j \gamma_z^j\right) w_i,$$

where $w_i = \exp(\xi_i)$, yielding a random y_i^j. We assume that w_i follows an n-component Inverse Gamma mixture distribution with pdf

$$\omega(w_i) = \sum_{z=1}^{n} \rho_z \frac{\frac{1}{(s_z-1)} \exp\left(-\frac{(s_z-1)}{w_i}\right)}{\left(\frac{w_i}{s_z-1}\right)^{s_z+1} \Gamma(s_z)},$$

$w_i > 0, s_z > 0$ for $z = 1, ..., n, \sum_{z=1}^{n} \rho_z = 1$ with mean $E(w_i) = 1$. It can be shown that the above parameterization does not affect the results if there is a constant term in the regression. We chose $E(w_i) = 1$ in order to have $E(\xi_i) = 0$. Under this assumption the conditional distribution of $X_{i,k}^j | d_{z,i}^j$ becomes

$$P\left(X_{i,k}^j | d_{z,i}^j\right) = \sum_{z=1}^{n} \rho_z s_z \frac{\left((s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{s_z}}{\left(x + (s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)\right)^{s_z+1}},$$

which is an n-component Pareto mixture distribution with parameters s_z and $(s_z - 1)\exp\left(d_{z,i}^j \gamma_z^j\right)$ and has $E(X_{i,k}^j | d_{z,i}^j) = \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^j \gamma_z^j\right)$. The posterior pdf of the mean claim severity y_i^{t+1} for an individual i observed over $t+1$ periods, with $X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t$ claim size history and $d_i^1, ..., d_i^{t+1}$ characteristics, is obtained applying Bayes theorem and is an n-component inverse gamma mixture with updated parameters, given by

$$g\left(y_i^{t+1} | X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{i,z}^1, ..., d_{i,z}^{t+1}\right) = \sum_{z=1}^{n} \rho_z \frac{\frac{1}{C_{i,z}^j} \exp\left(-\frac{C_{i,z}^j}{y_i^j}\right)}{\left(\frac{y_i^j}{C_{i,z}^j}\right)^{K+s_z+1} \Gamma(s_z + K)},$$

$$\text{for } y_i^j > 0, s_z > 0 \text{ and } z = 1, ..., n, \sum_{z=1}^{n} \rho_z = 1, \text{ where} C_{i,z}^j = \left[(s_z - 1) + \sum_{j=1}^{t} \frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_{z,i}^j \gamma_z^j\right)}\right] \exp\left(d_{z,i}^{t+1} \gamma_z^{t+1}\right).$$

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

$$\hat{y}_i^{t+1}\left(X_{i,1}^1, X_{i,2}^2, ..., X_{i,K_i^t}^t; d_{i,z}^1, ..., d_{i,z}^{t+1}\right) = \sum_{z=1}^{n} \rho_z \exp\left(d_{z,i}^{t+1}\gamma_z^{t+1}\right) \left[\frac{(s_z-1)+\sum_{j=1}^{t}\frac{\sum_{k=1}^{K_i^j} X_{i,k}^j}{\exp\left(d_{z,i}^j\gamma_z^j\right)}}{s_z+K-1}\right]. \quad (31)$$

5 Numerical Illustration

The data were kindly provided by a Greek insurance company. In our application we fit the Poisson and Negative Binomial distributions and their two component mixtures on the number of claims and the Exponential, Gamma, Weibull, GB2 and Pareto and their two and three component mixtures on the claim sizes. Furthermore, we introduce a regression component in the above 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. We give emphasis on both the analysis of the claim frequency and severity using two and/or three component mixtures of distributions and generalized linear models (glm) as these methods have not been extensively studied in the BMS literature. The location and weight of these components are estimated from the data employing the EM algorithm. The number of components was chosen based on the information we had from the frequency and severity data respectively[8]. The distributions and regression models were estimated using the GAMLSS package in R.

5.1 Modelling Results

The available a priori rating variables we employ are the Bonus Malus (BM) class, the horsepower (HP) of the car and gender of the driver. The variable BM class divides the classes of the current Greek BMS into five categories of drivers, those who belong to BM classes: C1= "1-2", C2 = "3-5", C3 = "6-9", C4 = "10" and C5 = "11-20". The variable HP consists of eleven categories of cars, those with a HP between: C1 = "0-33", C2 = "34-44", C3 = "45-55", C4 = "56-66", C5 = "67-74", C6 = "75-82", C7 = "83-90", C8 = "91-99", C9 ="100-110", C10 = "111-121" and C11 = "122-132". Finally, the gender consists of three categories: M = "male", F = "female" and B = "both", since, data for fleet vehicles used by either male or female drivers were also available, i.e. shared use. As suggested by Rigby and Stasinopoulos (2005 and 2009), model selection was performed using the Generalized Akaike information criterion (GAIC) in order to find the variables that are considered as better predictors. The claim frequency and severity models we consider in our application were the best fitted models.

In what follows, for brevity, we present only the values of the estimated parameters of the claim severity models, i.e. the Exponential, Gamma, Weibull, GB2 and Pareto[9] and their two and three component mixture regression models for assessing claim severity. Tables 1 and 2 summarize our findings with respect to the one, two and three component mixture models respectively[10].

[8] In principle one could use more components regarding the data set examined and then select the best models.

[9] The GAMLSS package allows us to find the maximum likelihood estimators of the parameters of the regression model where the distribution of the response variable is the Pareto2 (m', s') distribution, with pdf given by $f(x) = s'm'^{s'}(x+m')^{-s'-1}$. The Pareto$(m,s)$ response distribution can be derived from a reparameterization of the pdf of the Pareto2 (m', s') distribution with $s' = s$ and $m' = (s'-1)m$. Thus $\hat{s} = \hat{s}'$ and $\hat{m} = \frac{\hat{m}'}{\hat{s}'-1}$.

[10] Note that a * in Tables 1 and 2 indicates the estimated values which are statistically significant at a 5% threshold.

Table 1: Results of the Fitted One and Two Component Severity Mixture Regression Models

	EXP	2C EXP		GA	2C GA		WEI	2C WEI		GB2	2C GB2		PA	2C PA	
		C1	C2		C1	C2									
		ρ_1	ρ_2		ρ_1	ρ_2		ρ_1	ρ_2		ρ_1	ρ_2		ρ_1	ρ_2
		0.509	0.491		0.486	0.514		0.456	0.544		0.619	0.381		0.526	0.474
Variable															
Intercept	5.746*	5.746*	5.746*	5.746*	5.865*	5.249*	5.743*	6.573*	5.678*	5.462*	5.282*	5.988*	8.679*	7.678*	7.677*
BM															
C2	-0.023	-0.0233	-0.023	-0.023	-0.053*	0.073*	-0.041*	0.073	-0.053*	0.020	0.071*	-0.050*	-0.022	-0.020	-0.020
C3	0.113*	0.1128	0.113	0.113*	0.071*	0.185*	0.093*	0.183*	0.059*	0.149*	0.183*	0.089*	0.115*	0.117	0.117*
C4	-0.702*	-0.703*	-0.701*	-0.702*	-0.552*	-1.678*	-0.502*	-1.659*	-0.433*	-1.279*	-1.672*	-0.628*	-0.727*	-0.765*	-0.765*
C5	0.411	0.411	0.411	0.411*	0.383*	0.318*	0.405*	0.303	0.407*	0.420*	0.310*	0.318*	0.411	0.412	0.412
HP															
C2	-0.210	-0.210	-0.210	-0.210*	-0.161	-0.009	-0.199*	0.608	-0.423*	-0.185*	-0.024	-0.135	-0.211	-0.211	-0.211
C3	-0.201	-0.201	-0.201	-0.201*	-0.185*	0.006	-0.234*	0.001	-0.207*	-0.139*	-0.008	-0.166*	-0.198	-0.192	-0.192
C4	-0.016	-0.016	-0.017	-0.016	-0.032	0.171*	-0.043	0.166	-0.005*	0.044	0.159*	-0.000	-0.013	-0.007	-0.007
C5	0.003	0.003	0.003	0.003	0.002	0.205*	-0.018	0.196	0.011*	0.058	0.191*	0.023	0.006	0.010	0.010
C6	0.138	0.138	0.138	0.138*	0.164*	0.278*	0.137*	0.276*	0.191*	0.160*	0.267*	0.178*	0.139	0.141	0.141
C7	0.157	0.158	0.157	0.157*	0.179*	0.319*	0.191*	0.302*	0.247*	0.197*	0.304*	0.166*	0.156	0.156	0.156
C8	0.339*	0.339	0.339*	0.339*	0.325*	0.364*	0.383*	0.371*	0.390*	0.295*	0.353*	0.314*	0.335*	0.331	0.331
C9	0.445*	0.445*	0.444*	0.445*	0.414*	0.416*	0.490*	0.434*	0.444*	0.384*	0.402*	0.400	0.441*	0.435*	0.435*
C10	0.656*	0.656*	0.656*	0.656*	0.626*	0.609*	0.756*	0.596*	0.753*	0.582*	0.602*	0.564	0.646*	0.633*	0.634*
C11	1.090*	1.090*	1.091*	1.090*	1.057*	0.689*	1.313*	1.522*	1.067*	0.859*	0.671*	0.860	1.070*	1.040*	1.039*
Gender															
M	-0.078*	-0.079	-0.078	-0.078*	0.007	0.003	-0.078*	-1.328	0.153*	-0.109*	0.001	0.021	-0.081	-0.085	-0.085
F	-0.023	-0.024	-0.023	-0.023	0.065	0.017	-0.026	-1.313	0.201*	-0.067*	0.015	0.066*	-0.026	-0.030	-0.030
Parameter	-	-	-	θ	θ_1	θ_2	θ	θ_1	θ_2	σ	σ_1	σ_2	s'	s'_1	s'_2
	-	-	-	0.427*	0.096*	0.457*	2.234*	12.441*	2.234*	6.155*	17.350*	5.424*	2.962*	2.007*	2.007*
Parameter	-	-	-	-	-	-	-	-	-	ν	ν_1	ν_2	-	-	-
	-	-	-	-	-	-	-	-	-	1.006*	0.878*	0.522*	-	-	-
Parameter	-	-	-	-	-	-	-	-	-	s	s_1	s_2	-	-	-
	-	-	-	-	-	-	-	-	-	0.511*	1.136*	0.935*	-	-	-

Table 2: Results of the Fitted Three Component Severity Mixture Regression Models

	3C EXP			3C GA			3C WEI			3C GB2			3C PA		
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3
	0.310	0.355	0.335	0.012	0.467	0.521	0.373	0.495	0.132	0.604	0.201	0.195	0.363	0.310	0.327
Variable															
Intercept	5.745*	5.745*	5.746*	5.094*	5.247*	5.890*	5.320*	5.723*	6.322*	5.283*	5.747*	6.448*	7.677*	7.678*	7.677*
BM															
C2	-0.023	-0.023	-0.023	0.095*	0.069*	-0.042*	0.071*	-0.049*	0.070*	0.068*	0.059*	-0.150	-0.020	-0.019	-0.020
C3	0.113	0.113	0.113	0.209	0.182*	0.082	0.181*	0.013	0.744*	0.182*	0.198*	-0.074	0.117	0.117	0.117
C4	-0.702*	-0.702*	-0.702*	-1.660*	-1.663*	-0.546*	0.647	-0.702*	-0.033	-0.359*	-2.251*	-0.370*	-0.764*	-0.767*	-0.765*
C5	0.411	0.411	0.411	0.248	0.334*	0.414	0.256	-0.930	0.319	0.318*	0.209*	0.608*	0.412	0.412	0.412
HP															
C2	-0.210	-0.210	-0.210	0.480	-0.020	-0.149*	0.513	-0.283	-0.577	-0.024	0.090	-0.364*	-0.211	-0.211	-0.211
C3	-0.201	-0.201	-0.201	0.391	-0.005*	-0.174*	0.523	-0.336*	-0.561	-0.003	0.088	-0.454*	-0.192	-0.192	-0.192
C4	-0.016	-0.016	-0.016	1.635	0.160*	-0.020	0.207*	0.304	-0.359	0.169*	0.253*	-0.270*	-0.007	-0.007	-0.007
C5	0.003	0.003	0.003	1.702	0.192*	0.012	0.103*	0.079	0.254	0.198*	0.263*	-0.220	0.010	0.010	0.010
C6	0.138	0.138	0.138	1.706	0.270*	0.169	0.207*	0.304	-0.359	0.277*	0.310*	0.084	0.141	0.142	0.141
C7	0.157	0.157	0.158	0.407	0.307*	0.183*	0.209*	0.157	0.010	0.314*	0.298*	0.104	0.156	0.157	0.156
C8	0.339	0.339	0.339	3.027	0.352*	0.328*	0.305*	0.327*	0.037	0.359*	0.350*	0.318*	0.331	0.332	0.331
C9	0.444	0.444	0.445	0.360*	0.403*	0.424*	0.305	0.516*	0.037	0.416*	0.408*	0.416*	0.435	0.435	0.435
C10	0.656*	0.656*	0.656*	1.071*	2.228*	0.356*	0.541*	0.868*	0.423*	0.631*	0.525*	0.674*	0.634*	0.633*	0.634*
C11	1.088*	1.088*	1.094*	0.779*	0.681*	1.053*	0.568*	1.482*	0.362*	0.800*	0.768*	1.176*	1.039*	1.040*	1.039*
Gender															
M	-0.078	-0.078	-0.079	-1.311	0.017*	-0.030*	0.018*	-0.012	-0.492	-0.017	0.040	-0.064	-0.084	-0.085	-0.085
F	-0.023	-0.023	-0.023	-1.307	0.031*	0.019	0.036*	0.056	-0.509	-0.001	0.023	0.054	-0.030	-0.031	-0.030
Parameter	-	-	-	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	σ_1	σ_2	σ_3	s'_1	s'_2	s'_3
	-	-	-	9.3e-08*	0.096*	0.453*	12.317*	2.156*	14.311*	15.100*	11.790*	4.482*	2.007*	2.007*	2.007*
Parameter	-	-	-	-	-	-	-	-	-	ν_1	ν_2	ν_3	-	-	
	-	-	-	-	-	-	-	-	-	0.932*	1.090*	0.398*	-	-	
Parameter	-	-	-	-	-	-	-	-	-	s_1	s_2	s_3	-	-	
	-	-	-	-	-	-	-	-	-	0.994*	0.915*	1.146*	-	-	

5.2 Models Comparison

So far, we have several competing models for the claim frequency and severity components. The differences between models produce different premiums. Consequently, to distinguish between these models, this section compares them so as to select the best for each case.

5.2.1 Claim Frequency Models

The distributions and regression models we have for the claim frequency component are all nested. In order to accept or reject some models, classical hypothesis/specification tests for nested models can be used (see Boucher et al., 2007, 2008). The three standard tests are the log-likelihood ratio (LR), Wald, and Score (or Lagrange Multiplier, LM) tests, which are all asymptotically equivalent. Another standard method of comparing nested models (and also non-nested models) is to use the information criteria, such as the Global Deviance, AIC or the SBC (see Rigby and Stasinopoulos, 2009). Table 3 reports our results with respect to these nested comparisons. Specifically, from Panel A and Panel B we observe the superiority of the two component mixture distributions/regression models vs those with one component and the superiority of Negative Binomial vs Poisson. Overall, the best fit is given by the two component Negative Binomial mixture distribution/regression model.

Table 3: Claim Frequency Models Comparison

Panel A: Based on Likelihood Ratio Test

		Distributions		Regression Models	
Null Hypothesis	Alternative Hypothesis	Value	Decision	Value	Decision
Poisson	NBI	1032.2*	Reject	649.2*	Reject
Poisson	Poisson ($C = 2$)	986.5*	Reject	784.2*	Reject
NBI	NBI ($C = 2$)	1043*	Reject	158.3*	Reject
Poisson ($C = 2$)	NBI ($C = 2$)	1088.7*	Reject	23.3*	Reject

Panel B: Based on Global Deviance, AIC, SBC

	Distributions			Regression Models			
Model	df	AIC	SBC	df	Global Deviance	AIC	SBC
Poisson	1	30368.8	30376.4	9	29067.1	29085.1	29154.0
NBI	2	29338.6	29353.9	10	28417.9	28437.9	28514.5
Poisson (C=2)	3	29386.3	29409.3	19	28282.9	28320.9	28466.4
NBI (C=2)	5	29307.3	29345.6	21	28259.6	28301.6	28462.4

5.2.2 Claim Severity Models

Regarding the claim severity component, there are both nested and non-nested distributions/regression models comparisons.

Table 4 reports our results with respect to the nested comparisons. We observe that there is a superiority of Gamma, Weibull and GB2 with two or three components vs Gamma, Weibull, and GB2 respectively, while for Exponential we do not reject the null hypothesis. In the case of Pareto distribution we do not reject the null hypothesis, whereas when a regression component is included, the model with two components is superior. When we compare the Exponential with one component vs the Gamma, Weibull, and GB2 with two or three components, we can conclude that mixture distributions/regression models are superior to the simpler ones. Also, the finite mixtures of GB2 employing two and three components provided better fitting performance compared to the Pareto with one component.

Table 4: Nested Severity Models Comparison Based on Likelihood Ratio Test

		Distributions			Regression Models		
Null Hypothesis	Alternative Hypothesis	Value	p-value	Decision	Value	p-value	Decision
Exponential	Gamma ($C = 1$)	4004.4	0.00	Reject	5544.9	0.00	Reject
Exponential	Gamma ($C = 2$)	5844.3	0.00	Reject	9011.0	0.00	Reject
Exponential	Gamma ($C = 3$)	6340.8	0.00	Reject	10283.5	0.00	Reject
Exponential	Weibull ($C = 1$)	2893.4	0.00	Reject	4639.8	0.00	Reject
Exponential	Weibull ($C = 2$)	5480.9	0.00	Reject	8215.6	0.00	Reject
Exponential	Weibull ($C = 3$)	69784.8	0.00	Reject	7785.0	0.00	Reject
Exponential	Exponential ($C = 2$)	0.00	1.00	No Reject	0.00	1.00	No Reject
Exponential	Exponential ($C = 3$)	0.00	1.00	No Reject	0.00	1.00	No Reject
Exponential ($C = 2$)	Exponential ($C = 3$)	0.00	1.00	No Reject	0.00	1.00	No Reject
Gamma	Gamma ($C = 2$)	1839.9	0.00	Reject	3466.1	0.00	Reject
Gamma	Gamma ($C = 3$)	2336.4	0.00	Reject	4738.6	0.00	Reject
Gamma ($C = 2$)	Gamma ($C = 3$)	496.5	0.00	Reject	1272.5	0.00	Reject
Weibull	Weibull ($C = 2$)	357.4	0.00	Reject	3575.8	0.00	Reject
Weibull	Weibull ($C = 3$)	3282.3	0.00	Reject	3145.2	0.00	Reject
Weibull ($C = 2$)	Weibull ($C = 3$)	694.8	0.00	Reject	430.6	0.00	Reject
GB2	GB2 ($C = 2$)	1095.5	0.00	Reject	2088	0.00	Reject
GB2	GB2 ($C = 3$)	1350.5	0.00	Reject	2807	0.00	Reject
GB2 ($C = 2$)	GB2 ($C = 3$)	255	0.00	Reject	719	0.00	Reject
Pareto	GB2 ($C = 1$)	5158.1	0.00	Reject	6925	0.00	Reject
Pareto	GB2 ($C = 2$)	6253.6	0.00	Reject	9013	0.00	Reject
Pareto	GB2 ($C = 3$)	6508.6	0.00	Reject	9732	0.00	Reject
Pareto	Pareto ($C = 2$)	4.9	0.17	No Reject	367.2	0.00	Reject
Pareto	Pareto ($C = 3$)	4.8	0.56	No Reject	232.3	0.00	Reject
Pareto ($C = 2$)	Pareto ($C = 3$)	0.1	0.99	No Reject	0.1	1.00	No Reject

In Table 5 (Panels A and B) we compare the non-nested severity distributions/regression models. Overall, with respect to Global Deviance, AIC and SBC criteria and the Vuong test, the best fit is given by the GB2 when one, two or three components are used.

Table 5: Non - Nested Severity Models Comparison

Panel A: Based on Global Deviance, AIC, SBC

	Distributions			Regression Models			
Model	df	AIC	SBC	df	Global Deviance	AIC	SBC
Exponential	1	75946.5	75953.1	17	75520.6	75554.6	75667.3
Gamma	2	71944.1	71957.3	18	69975.7	70011.7	70131.0
Weibull	2	73055.1	73068.3	18	70880.8	70916.8	71036.1
GB2	4	70835.2	70861.7	20	68821.9	68861.9	68994.5
Pareto	2	75989.3	76002.5	18	75746.9	75782.9	75902.2
Exp (C=2)	3	75950.5	75970.4	35	75520.6	75590.6	75822.6
Gamma (C=2)	5	70110.2	70143.4	37	66509.6	66583.6	66828.8
Weibull (C=2)	5	70473.6	70506.8	37	67305.0	67379.0	67624.3
GB2 (C=2)	9	69749.7	69809.4	41	66733.9	66815.9	67087.7
Pareto (C=2)	5	76000.2	76033.3	37	76114.1	76188.1	76433.4
Exp (C=3)	5	75954.5	75987.6	53	75520.6	75626.6	75978.0
Gamma (C=3)	8	69619.7	69672.7	56	65237.1	65349.1	65720.3
Weibull (C=3)	8	69784.8	69837.8	56	67735.6	67847.6	68218.8
GB2 (C=3)	14	69504.7	69597.5	62	66014.9	66138.9	66549.9
Pareto (C=3)	8	76006.1	76059.2	56	76114.2	76226.2	76597.4

Table 5 (continued)							
Panel B: Based on Vuong test							
		Distributions			Regression Models		
Model 1	Model 2	Vuong Test	p-value	Decision	Vuong Test	p-value	Decision
Exponential	GB2	-33.81	0.00	GB2	-35.49	0.00	GB2
Exponential	Pareto	17.38	0.00	Exp	99.87	0.00	Exp
Gamma	GB2	-9.05	0.00	GB2	-10.76	0.00	GB2
Gamma	Pareto	23.64	0.00	Gamma	36.79	0.00	Gamma
Weibull	Gamma	-14.38	0.00	Gamma	-10.71	0.00	Gamma
Weibull	GB2	-12.43	0.00	GB2	-11.80	0.00	GB2
Weibull	Pareto	14.73	0.00	Weibull	34.39	0.00	Weibull
Exp (C=2)	GB2 (C=2)	-43.96	0.00	GB2	-47.35	0.00	GB2
Exp (C=2)	Pareto (C=2)	16.56	0.00	Exp	110.19	0.00	Exp
Gamma (C=2)	GB2 (C=2)	-8.29	0.00	GB2	-0.45	0.32	None
Gamma (C=2)	Pareto (C=2)	38.48	0.00	Gamma	51.19	0.00	Gamma
Weibull (C=2)	Gamma (C=2)	-4.67	0.00	Gamma	-6.00	0.00	Gamma
Weibull (C=2)	GB2 (C=2)	-6.45	0.00	GB2	-6.42	0.00	GB2
Weibull (C=2)	Pareto (C=2)	32.39	0.00	Weibull	49.73	0.00	Weibull
Exp (C=3)	GB2 (C=3)	-47.99	0.00	GB2	-37.31	0.00	GB2
Exp (C=3)	Pareto (C=3)	16.65	0.00	Exp	110.12	0.00	Exp
Gamma (C=3)	GB2 (C=3)	-33.84	0.00	GB2	8.57	0.00	Gamma
Gamma (C=3)	Pareto (C=3)	47.59	0.00	Gamma	49.83	0.00	Gamma
Weibull (C=3)	Gamma (C=3)	-27.98	0.00	Gamma	-5.84	0.00	Gamma
Weibull (C=3)	GB2 (C=3)	13.79	0.00	Weibull	-3.18	0.00	Weibull
Weibull (C=3)	Pareto (C=3)	45.65	0.00	Weibull	51.04	0.00	Weibull

5.3 Optimal BMS

Based on the current methodology as presented in Sections 3 and 4, we compute an optimal BMS with a frequency and a severity component based on the a posteriori criteria and based both on the a priori and the a posteriori criteria. When both criteria are considered, we examine a group of policyholders who share the following common characteristics: We consider that the policyholder i is a woman, who belongs to the first BM category, and has a car with HP between 0-33. We calculate the premium rates using the net premium principle for the set of the distributions/regression models that were presented in these sections. These premium rates will be divided by the premium when $t = 0$, since we are interested in the differences between various classes and the results are presented so that the premium for a new policyholder is 100.

We consider first the optimal BMS resulting from the claim frequency distributions/regression models. For the two component Poisson mixture 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 the Table 6 (Panels A and B), we consider that the specific policyholder belongs to the second category[11]. If the policyholder i has a claim free year, the probability of being a bad risk and the premium rates reduce, whereas if she has one or more claims, the probability of being a bad risk and the premium rates increase, resulting in bonus or malus respectively. For example, from Panel B we observe that if the policyholder has one claim in the first year, she faces a malus of 72.44% in her premium.

Table 6: Optimal BMS, Two Component Poisson Mixture Model

Panel A: Optimal BMS - A Posteriori Criteria						Panel B: Optimal BMS - Both Criteria					
Number of Claims						Number of Claims					
Year	k					Year	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	89.02	165.00	276.11	328.50	340.90	1	88.27	172.44	322.10	395.03	410.54
2	81.00	141.13	254.10	321.76	339.53	2	80.33	142.15	288.65	385.29	408.81

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

As we mentioned previously, the optimal BMS resulting from the two component Negative Binomial mixture can be derived in two alternative (not equivalent) ways, either by updating the posterior probability or by updating the posterior mean. Table 7 (Panels A and B) summarizes our findings. Note that in the case of updating the posterior probability based on the a posteriori criteria (Panel A, A1) and in the case of updating the posterior probability (Panel B, B1) and posterior mean (see Panel B, B2) based on both criteria, the explicit claim frequency history determines the calculation of the premium rates and not just the total number of claims as in the case of the two component Poisson mixture. For this reason in Panel A, A1 and Panel B we specify the exact order of the claims history in order to calculate the premium rates. For instance, when we update the posterior probability (Panel B, B1) and consider a bad risk policyholder, we observe that if she has at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk increases and thus her premium increases from 100 to 222.03 while if she has $k_1 = 1, k_2 = 1$ claim frequency history (i.e. total number of claims $K = 2$ at $t = 2$) then the posterior probability of being a bad risk increases and subsequently her premium increases from 100 to 219.72. When we update the posterior mean based on both criteria (Panel B, B2) we see that if she has at $t = 2$ claim frequency history $k_1 = 0, k_2 = 2$ then her premium increases from 100 to 237.37, while if she has $k_1 = 1, k_2 = 1$ claim frequency history her premium reaches 233.12.

Table 7: Optimal BMS, Two Component Negative Binomial Mixture Model

Panel A: Optimal BMS - A Posteriori Criteria			Panel B: Optimal BMS - Both Criteria	
		A1: Update of the Posterior Probability	B1: Update of the Posterior Probability	B2: Update of the Posterior Mean
Year	Number of Claims k_t			
$t = 0$	$k_0 = 0$	100	100	100
	$k_1 = 0$	99.56	92.43	95.14
$t = 1$	$k_1 = 1$	102.94	146.16	208.54
	$k_1 = 2$	105.08	243.16	245.46
	$k_1 = 0, k_2 = 0$	99.13	86.67	90.99
$t = 2$	$k_1 = 0, k_2 = 1$	102.50	129.65	201.72
	$k_1 = 0, k_2 = 2$	104.64	222.03	237.37
	$k_1 = 1, k_2 = 0$	102.50	129.65	109.49
$t = 2$	$k_1 = 1, k_2 = 1$	105.85	219.72	233.12
	$k_1 = 1, k_2 = 2$	107.88	299.50	342.75
	$k_1 = 2, k_2 = 0$	104.64	222.03	233.12
$t = 2$	$k_1 = 2, k_2 = 1$	107.88	299.50	342.75
	$k_1 = 2, k_2 = 2$	109.77	329.81	388.13

A2: Update of the Posterior Mean

Year	Number of Claims k				
t	0	1	2	3	4
0	100.00	0.00	0.00	0.00	0.00
1	97.12	119.58	142.04	164.49	186.95
2	94.42	116.21	138.00	159.79	181.58

For the severity component, we consider the optimal BMS determined by the two and three component Exponential, Gamma, Weibull and GB2 mixture distributions/regression models for the case of updating the posterior probability and the system resulting from the two and three component Pareto mixture distribution/regression model for the case of updating the posterior mean. Table 8 (Panels A and B) displays the premium rates resulting from these models with respect to the a posteriori criteria (Panel A, column A1 and Panel B, column B1) and to both the a priori and the a posteriori criteria (Panel A, column A2 and Panel B, column B2). From Table 8 we observe that the premium is equal to 100, the basic premium, in the case of the two and three component Exponential mixture, revealing the unnecessity of the two and three components. As expected, in the case of the two and three component Gamma, Weibull, GB2 and Pareto mixtures the premium values increase proportionally to the claim severity. For example, from Panel A, column A2, and Panel B, column B2, we see that for one claim size of 400 in the first year the premium increases from 100 to 130.884, 126.621, 161.237 and 101.043, and to 129.747, 116.587, 140.785 and 101.105 for the case of the Gamma, Weibull, GB2 and Pareto mixture regression models with two and three components respectively.

Table 8: Optimal BMS, Two and Three Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Panel A: Two component Mixture Models										
Claim Size	Exponential		Gamma		Weibull		GB2		Pareto	
	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2
200	100		83.856	73.436	82.1181	75.316	83.634	69.528	99.488	92.945
300	100		92.995	130.883	107.106	126.621	94.622	161.049	99.883	96.994
400	100		120.098	130.884	122.941	126.621	118.517	161.237	100.279	101.043
Panel B: Three component Mixture Models										
Claim Size	Exponential		Gamma		Weibull		GB2		Pareto	
	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2
200	100		79.481	75.180	80.818	81.179	77.176	72.365	99.485	93.003
300	100		93.682	129.744	113.637	115.930	89.795	125.452	99.878	97.054
400	100		128.462	129.747	121.039	116.587	145.876	140.785	100.271	101.105

Finally, we present the optimal BMS with a frequency and severity component. The premiums resulting from this system are calculated via the product of the expected claim frequency and the expected claim severity with independence between the two components assumed. Table 9 (Panels A, B and C) summarizes our findings with respect to the a posteriori criteria (Panel A, column A1, Panel B, column B1 and Panel C, column C1) and to both the a priori and the a posteriori criteria (Panel A, column A2, Panel B, column B2 and Panel C, column C2). We observe that for one claim size of 400 in the first year the premium increases from 100 to 172.44, 225.696, 218.344, 278.038, 174.240 in the two component Poisson mixture model and the corresponding two component severity models (Panel A, column A2), to 146.16, 191.300, 185.070, 235.664, 174.240 in the two component Negative Binomial mixture model (updating the posterior probability) and the corresponding two component severity models (Panel B, column B2), to 208.54, 272.945, 264.054, 336.244, 210.716 in the two component Negative Binomial mixture model (updating the posterior mean) and the corresponding two component severity models (Panel C, column C2).

Table 9: Optimal BMS Based on the Alternative Two Component Poisson Mixture Models for Assessing Claim Frequency and the Various Two Component Mixture Models for Assessing Claim Severity, One Claim in the First Year of Observation

Panel A: Two Component Poisson Mixture Model (Update of the Posterior Probability)										
Claim Size	POIS-EXP		POIS-GA		POIS-WEI		POIS-GB2		POIS-PA	
	A1	A2	A1	A2	A1	A2	A1	A2	A1	A2
200	165.00	172.44	138.362	126.634	135.495	129.875	138.000	119.8955	164.155	160.274
300	165.00	172.44	153.442	225.693	176.726	218.344	156.127	277.713	164.808	167.257
400	165.00	172.44	198.162	225.696	202.853	218.344	195.553	278.038	165.460	174.240
Panel B: Two Component Negative Binomial Mixture Model (Update of the Posterior Probability)										
Claim Size	NB-EXP		NB-GA		NB-WEI		NB-GB2		NB-PA	
	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2
200	102.94	146.16	86.321	107.335	84.532	110.082	86.093	101.622	102.413	160.274
300	102.94	146.16	95.729	191.298	110.256	185.070	97.404	235.390	102.820	167.260
400	102.94	146.16	123.629	191.300	126.556	185.070	122.001	235.664	103.230	174.240
Panel C: Two Component Negative Binomial Mixture Model (Update of the Posterior Mean)										
Claim Size	NB-EXP		NB-GA		NB-WEI		NB-GB2		NB-PA	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
200	119.58	208.54	100.275	153.144	98.197	157.064	100.001	144.994	118.970	193.830
300	119.58	208.54	111.203	272.942	128.078	264.054	113.150	335.852	119.441	202.272
400	119.58	208.54	143.613	272.945	147.013	264.054	141.723	336.244	119.913	210.716

6 Conclusions

In this paper, we developed the design of an optimal BMS that takes into account the number of claims of each policyholder and the exact size of loss that these claims incurred using various finite mixtures of distributions. For the frequency component we considered that the number of claims is distributed according to a finite Poisson, Delaporte and Negative Binomial mixture, and for the severity component we consider that the losses are distributed according to a finite Exponential, Gamma, Weibull and GB2 mixture. These optimal BMS were obtained by updating the posterior probability of the policyholder's risk class. Furthermore, we extended the setup of Frangos and Vrontos (2001) for Negative Binomial and Pareto mixtures and designed an optimal BMS based on posterior distributions of the mean claim frequency and size, given the information we have about the claim frequency and size history for each policyholder. We have also developed a generalized BMS that integrates the a priori and a posteriori information on an individual basis, extending the framework developed by Dionne and Vanasse (1989, 1992) and Frangos and Vrontos (2001) using finite mixtures of regression models that allow us to account for unobserved heterogeneity. The optimal BMS obtained have all the attractive properties of the BMS developed by Lemaire (1995), Frangos and Vrontos (2001) and Dionne and Vanasse (1989, 1992). Extensions to other distributions/regression models can be obtained in a similar straightforward way.

A possible line of further research is to apply the same mixtures to all the contracts of the same insured so a dependence between the contracts can be modeled (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] 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.

[8] Gourieroux, C., A. Montfort and A. Trognon (1984 a). Pseudo maximum likelihood methods: theory. Econometrica, 52, 681-700.

[9] Gourieroux, C., A. Montfort and A. Trognon (1984 b). Pseudo maximum likelihood methods: applications to Poisson models. Econometrica, 52, 701-720.

[10] Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). Continuous Univariate Distributions. Wiley.

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

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

[13] Mahmoudvand, R. and H. Hassani (2009). Generalized bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. ASTIN Bulletin, 39, 307-315.

[14] McDonald, J. B. (1996). Probability Distributions for Financial Models. In Maddala, G. S. and C. R.Rao (eds.). Handbook of Statistics, 14, 427-460. Elsevier.

[15] McDonald, J. B. and Y. J. Xu (1995). A generalisation of the beta distribution with applications. Journal of Econometrics, 66, 133-152.

[16] McLachlan, G.and D. Peel (2000). Finite Mixture Models. John Wiley & Sons.

[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.