Author/s:
Gomez-Deniz, E;Calderin-Ojeda, E

Title:
A Survey of the Individual Claim Size and Other Risk Factors Using Credibility Bonus-Malus Premiums

Date:
2020-02-01

Citation:
Gomez-Deniz, E. & Calderin-Ojeda, E. (2020). A Survey of the Individual Claim Size and Other Risk Factors Using Credibility Bonus-Malus Premiums. Risks, 8 (1), https://doi.org/10.3390/risks8010020.

Persistent Link:
https://hdl.handle.net/11343/274586

risks

MDPI

Article

A Survey of the Individual Claim Size and Other Risk Factors Using Credibility Bonus-Malus Premiums

Emilio Gómez-Déniz [1,*] and Enrique Calderín-Ojeda [2]

[1] Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain
[2] Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Melbourne, VIC 3010, Australia; enrique.calderin@unimelb.edu.au
* Correspondence: emilio.gomez-deniz@ulpgc.es

Received: 24 December 2019; Accepted: 18 February 2020; Published: 21 February 2020

Abstract: In this paper, a flexible count regression model based on a bivariate compound Poisson distribution is introduced in order to distinguish between different types of claims according to the claim size. Furthermore, it allows us to analyse the factors that affect the number of claims above and below a given claim size threshold in an automobile insurance portfolio. Relevant properties of this model are given. Next, a mixed regression model is derived to compute credibility bonus-malus premiums based on the individual claim size and other risk factors such as gender, type of vehicle, driving area, or age of the vehicle. Results are illustrated by using a well-known automobile insurance portfolio dataset.

Keywords: aggregate claims; auto insurance; Bayesian; bonus-malus; compound distribution

MSC: 60E05; 62P05; 62E99

1. Introduction

In a recent work, a modification in the bonus-malus systems was proposed Gómez-Déniz (2016), which are commonly applied in automobile insurance, that differentiated between two different types of claims by including a bivariate model based on the assumption of dependence. The aforementioned work studied the impact on the bonus-malus premium in a general setting without involving individual's risk factors, such as gender, type of vehicle, area of circulation, etc.

It is well known that under the traditional bonus-malus system, the premium charged to each insured is based solely on the number of claims made. Therefore, an insured who has had an accident that causes a relatively small loss amount is penalised to the same extent as one who has experienced a more expensive accident. This event would seem to be unfair by the insureds. In the mentioned work a bivariate prior model, conjugated with respect to the likelihood, was also proposed, and as a result of this, simple credibility bonus-malus premiums that satisfy appropriate transition rules were obtained. These expressions were used to compute credibility bonus-malus premiums by considering two different types of claims: those ones above and below a threshold claim size, say $\psi > 0$.

Similar related works have been proposed in the actuarial literature. In this sense, the work in Pinquet (1998) computed bonus-malus rates in a multi-equation Poisson model with random effects. The work in Ragulina (2011) introduced a bonus-malus system with different claim types and varying deductibles. The work in Walhin and Paris (2001) showed how to set up a practical bonus-malus system with a finite number of classes using both the actual claim amount and claim frequency distribution. The work in Bonsdorff (2005) also incorporated the claim number and the severity in the bonus-malus system literature by using Markov chains. The work in Bermúdez (2009) examined,

in automobile insurance claims, an a priori ratemaking procedure that included two different types of claim, i.e., with and without bodily injuries. See also Bermúdez and Karlis (2017).

The main objective of this work is to develop a reparametrization of the bivariate distribution proposed in the previous work with the purpose of incorporating individual information in the model to adjust the premiums charged to each policyholder. Additionally, some statistical properties of the proposed parametrization that were not addressed in the previous work will be shown. Furthermore, an extensive set of a priori classification variables such as age, gender, type and age of car, etc., will be used to incorporate, depending on the heterogeneity of the insured's behaviour, prior distributions assigned to the parameters of the model to build up a posteriori credibility, bonus-malus premiums.

The rest of this paper is organised as follows. The main model and some of its properties are presented in Section 2. In Section 3, the regression model is introduced, and maximum likelihood estimation methods are illustrated. We will show that the estimation procedure is simply derived, and Fisher's information matrix associated with this regression model is obtained in closed-form. Credibility premiums related to the regression models are provided in Section 4. Numerical illustrations and results connected with the compound model are shown in Section 5, and finally, Section 6 concludes the work.

2. The Model

As pointed out by Dionne and Vanasse (1989), the classical Poisson distribution is generally employed for the characterization of random and independent events such as automobile accidents. Thus, we assume that the number of claims in an automobile insurance portfolio follows a Poisson distribution with parameter $\mu_1 > 0$. When an insured declares a claim, it might be for an amount exceeding ψ monetary units. In order to accommodate this characteristic into the model, we incorporate a second random variable, thus giving rise to the consideration of two separate sub-events (claims worth more or less than ψ), in the following way. Let Z_i be the variable that takes the value one if the i^{th} claim corresponds to a claim size larger than ψ and the value zero otherwise. Thus, the $Z_i's$ variables are modelled as independent and identically distributed with the following Bernoulli probability density function:

$$f(z_i|p) = \begin{cases} \mu_2/\mu_1, & \text{if} \quad z_i = 1, \\ 1 - \mu_2/\mu_1, & \text{if} \quad z_i = 0, \end{cases}$$

where $p = \mu_2/\mu_1$ is the probability of declaring a claim larger than ψ with $0 < \mu_2 < \mu_1$.

Let us now assume that $X_2 = \sum_{i=1}^{X_1} Z_i$ is the total claim number with a claim amount larger than ψ. Thus, if the Z_i $(i = 1, \ldots, x_1)$ are assumed to be mutually independent, then the conditional probability function of X_2, given that $X_1 = x_1$, is binomial with parameters x_1 and μ_2/μ_1. Therefore, the joint distribution of the total claim number (X_1) and the corresponding claim number with claim amount exceeding ψ, X_2, has this probability function:

$$\Pr(X_1 = x_1, X_2 = x_2) = \frac{\mu_2^{x_2}(\mu_1 - \mu_2)^{x_1 - x_2} \exp(-\mu_1)}{(x_1 - x_2)! x_2!}, \tag{1}$$

for $x_1 = 0, 1, \ldots, x_2 = 0, 1, \ldots, x_1$, $\mu_1 > 0$, and $0 < \mu_2 < \mu_1$.

Observe that the probability function (1) can be written as:

$$\Pr(X_1 = x_1, X_2 = x_2) = h(\boldsymbol{x}) \exp\left[\sum_{i=1}^{2} x_i R_i(\boldsymbol{\Theta}) - Q(\boldsymbol{\Theta})\right],$$

where $\boldsymbol{x} = (x_1, x_2)$, $\boldsymbol{\Theta} = (\mu_1, \mu_2)'$, $R_1(\boldsymbol{\Theta}) = \log(\mu_1 - \mu_2)$, $R_2(\boldsymbol{\Theta}) = \log(\mu_2/(\mu_1 - \mu_2))$, $Q(\boldsymbol{\Theta}) = \mu_1$, and $h(\boldsymbol{x}) = [(x_1 - x_2)! x_2!]^{-1}$. Thus, (1) belongs to the multivariate exponential family of distributions provided in Khatri (1983a). See also Khatri (1983b) and Johnson et al. (1997, chp. 34). This family

includes also the multivariate Lagrangian distributions and the multivariate power series distributions; (see Khatri 1983b).

Properties of the Distribution

The marginal means are given by $E(X_i) = \mu_i, i = 1,2$. The cross moment, the covariance, and the correlation are given by:

$$\begin{aligned} E(X_1X_2|\mu_1,\mu_2) &= \mu_2(1+\mu_1), \\ cov(X_1,X_2|\mu_1,\mu_2) &= \mu_2, \\ \rho(X_1,X_2|\mu_1,\mu_2) &= \sqrt{\mu_2/\mu_1}, \end{aligned} \tag{2}$$

respectively. Thus, the model admits only positive correlation.

The probabilities for different values of (x_1, x_2) were calculated, and graphs were plotted for different values of these two parameters. They are shown in Figure 1. It is observable that for larger values of μ_1 and μ_2, the modal value increases in x_1 and x_2, illustrating that the new model is very versatile.

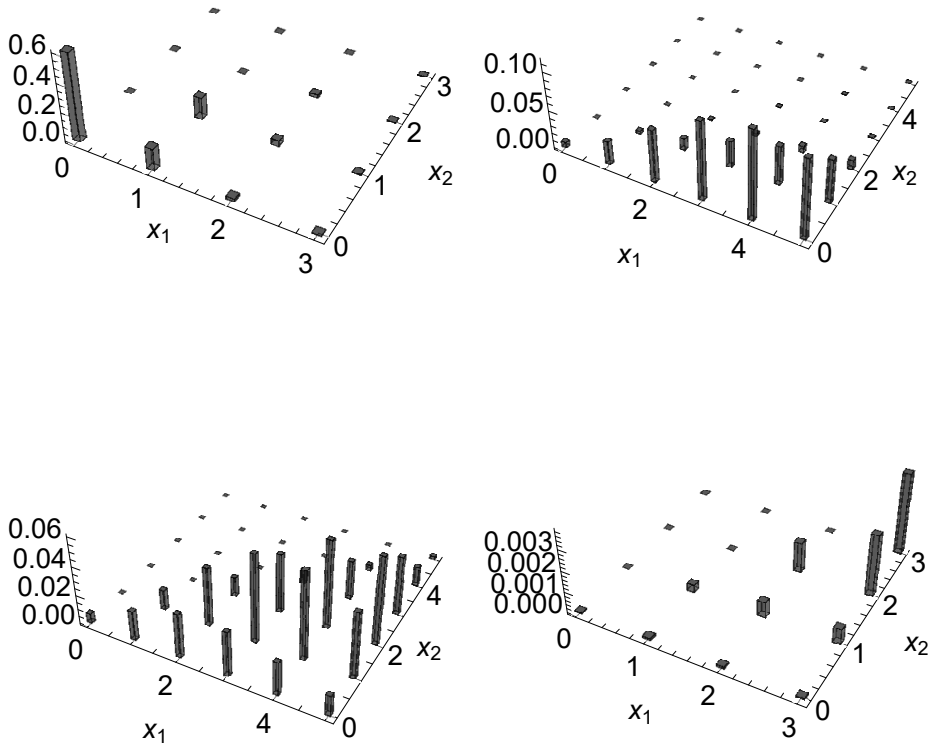

Figure 1. Joint probability mass functions of the bivariate discrete distribution proposed for selected values of the parameters. From top to bottom and left to right, we have: $(\mu_1, \mu_2) = (0.5, 0.25)$, $(\mu_1, \mu_2) = (5, 0.5)$, $(\mu_1, \mu_2) = (5, 2)$, and $(\mu_1, \mu_2) = (10, 8)$.

The expression provided in (1) can also be obtained differently as follows: Let us consider an automobile insurance portfolio in which X_1 is a random variable that represents the number of claims in a given period and X_2 yields the number of claims with a size above a threshold $\psi > 0$ over the same period of time. If each policyholder has a probability μ_2/μ_1 of having a claim with a claim size above ψ, then $\Pr(X_2 = x_2)$ and $\Pr(X_1 = x_1)$ are related as follows:

$$\Pr(X_2 = x_2) = \sum_{x_1=x_2}^{\infty} \binom{x_1}{x_2} \left(\frac{\mu_2}{\mu_1}\right)^{x_2} \left(1 - \frac{\mu_2}{\mu_1}\right)^{x_1-x_2} \Pr(X_1 = x_1). \tag{3}$$

Obviously, (3) represents a map from the probability function to the probability function. That is, $\sum_{x_2=0}^{\infty} \Pr(X_2 = x_2) = 1$ with $\Pr(X_2 = x_2) \geq 0$, $x_2 = 0, 1, \ldots$

Although other distributions, i.e., negative binomial, could be chosen to model count data, for the sake of simplicity, let us suppose that X_1 follows a Poisson distribution with parameter $\mu_1 > 0$. Then, we have:

$$\begin{aligned}
\Pr(X_2 = x_2) &= \sum_{x_1=x_2}^{\infty} \binom{x_1}{x_2} \left(\frac{\mu_2}{\mu_1}\right)^{x_2} \left(1 - \frac{\mu_2}{\mu_1}\right)^{x_1 - x_2} \frac{\mu_1^{x_1} \exp(-\mu_1)}{x_1!} \\
&= \frac{\exp(-\mu_1)}{x_2!} \left(\frac{\mu_2}{\mu_1 - \mu_2}\right)^{x_2} \sum_{x_1=x_2}^{\infty} \frac{(\mu_1 - \mu_2)^{x_1}}{(x_1 - x_2)!} \\
&= \frac{\mu_2^{x_2} \exp(-\mu_1)}{x_2!} \sum_{j=0}^{\infty} \frac{(\mu_1 - \mu_2)^j}{j!} \\
&= \frac{\mu_2^{x_2}}{x_2!} \exp(-\mu_2), \quad x_2 = 0, 1, \ldots
\end{aligned}$$

Expression (3) can be viewed as a weighted sum of binomial probabilities where the weights are given by the probability that the policyholder declares a certain number of claims. More specifically, it is the mean of the total number of claims with a threshold conditional on the fact that $X_1 = x_1$ claims and assuming the existence of a heterogeneity factor that causes different claims of different amounts. Hence, Expression (3) can be viewed as a mixture distribution. From this standpoint, the model provides a framework in which random effects are incorporated into the Poisson assumption. In this case, the bivariate distribution provided in (1) can be obtained by multiplying the conditional and the marginal distributions in the usual way.

Numerical simulation of the bivariate distribution can be simply obtained by following the approach explained in Kocherlakota and Kocherlakota (1992, chp. 1). In this regard, both the marginal distribution $f(x_1)$ and the conditional distribution $f(x_2|x_1)$ will be used. The former is a Poisson distribution with parameter μ_1 and the latter a binomial distribution with parameters x and μ_2/μ_1. Thus, for specific values of x_1, a realization of x_2 from $f(x_2|x_1)$ can be generated, and therefore, the pairs (x_1, x_2) are observations from the joint distribution given in (1). This procedure can be repeated n times in order to obtain a random sample of size n.

The joint probability generating function is given by:

$$G_{X_1,X_2}(s_1, s_2) = \exp[\mu_1(s_1 - 1) + \mu_2(s_2 - 1)s_1], \quad |s_1| \leq 1, |s_2| \leq 1. \tag{4}$$

Note that (4) is the limiting case of the bivariate Poisson distribution with parameters $\theta_1 = \mu_1 - \mu_2$, $\theta_2 \to 0$, and $\theta_{12} = \mu_2$ (see for instance this expression in (Johnson et al. 1997, chp. 37) and also Hesselager (1996) for more details of recursions for bivariate discrete distributions). Thus, the following recursions are valid:

$$\begin{aligned}
p_{x_1,x_2} &= \frac{\mu_1 - \mu_2}{x_1} p_{x_1-1,x_2} + \frac{\mu_2}{x_1} p_{x_1-1,x_2-1}, \\
p_{x_1,x_2} &= \frac{\mu_2}{x_2} p_{x_1-1,x_2-1},
\end{aligned}$$

with:

$$\begin{aligned}
p_{0,0} &= \exp(-\mu_1), \\
p_{x_1,0} &= \frac{(\mu_1 - \mu_2)^{x_1} \exp(-\mu_1)}{x_1!},
\end{aligned}$$

and zero otherwise.

3. The Role of the Covariates

Clearly, the number of claims below and above ψ may be influenced by different characteristics and factors; likewise, explanatory variables may be useful to explain the individual premium to be charged. As (1) satisfies that the marginal means are given by $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$, then covariates can be simply implemented in the model.

We now investigate the effect of including covariates to account for the total number of claims and the claims above the threshold ψ. Obviously, some factors are crucial when explaining the endogenous variables (X_{1i}, X_{2i}). Two appropriate links are needed to connect the explanatory variables with the marginal means. A natural way to proceed is to assume that (X_{1i}, X_{2i}) for $i = 1, \ldots, n$ follows the probability function (1) and:

$$\begin{aligned} \log \mu_{1i} &= \omega_{1i}\beta_1, \\ \mu_{2i} &= \frac{\mu_{1i} \exp(\eta_{2i}\beta_2)}{1 + \exp(\eta_{2i}\beta_2)}, \end{aligned}$$

where ω_{1i} and η_{2i} denote vectors of m explanatory variables for the i^{th} observation, i.e., with components ω_{ji} and η_{ji}, $(j = 1, \ldots, m)$, used to model μ_{1i} and μ_{2i}, respectively, and where $\beta_k = (\beta_{k1}, \ldots, \beta_{km})^\top$, $(k = 1, 2)$ designates the corresponding vector of regression coefficients. The log-linear specification for μ_{1i} is widely used, while the link function for μ_{2i} was chosen in this way to ensure that the latter one would not be larger than μ_{1i}, and thus, it would be compatible with $X_2 \leq X_1$.

These mean values may be influenced by several characteristics and variables, and the explanatory variables that are used to model each parameter μ_{1i} and μ_{2i} may not be the same in practice. In this respect, the work in Cameron and Trivedi (1998) provided good insight into standard count regression models.

The marginal effect reflects the variation of the conditional mean of X_1 and X_2 due to a one-unit change in the j^{th} covariate, and it is calculated as:

$$\begin{aligned} \frac{\partial \mu_{1i}}{\partial \beta_{1j}} &= \omega_{ji} \exp(\omega_{1i}\beta_1) = \omega_{ji}\mu_{1i}, \\ \frac{\partial \mu_{2i}}{\partial \beta_{2j}} &= \eta_{ji}\mu_{2i}\left(1 - \frac{\mu_{2i}}{\mu_{1i}}\right), \end{aligned} \tag{5}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Thus, the marginal effect indicates that a one-unit change in the j^{th} regressor increases or decreases the expectation of the total number of claims and the number of claims above the given threshold depending on the sign, positive or negative, of the regressor for each mean. For indicator variables such as ω_{ik}, which takes only the value zero or one, the marginal effect in terms of the odds-ratio is $\exp(\beta_{1j})$ for μ_{i1} and $\exp(\beta_{2j})$ for μ_{i2}. Therefore, for μ_{i1}, the conditional mean is $\exp(\beta_{1j})$ times larger if the indicator is one rather than zero. A similar conclusion is drawn for μ_{i2}. Certainly, if μ_{1i} and μ_{2i} share the same covariates, then (5) does not correspond to the marginal effect of the j^{th} covariate since μ_{1i} may also change in response to the changes of this covariate.

3.1. Estimation

In this section, we derive estimators based on the maximum likelihood for the model with and without covariates, and we also provide closed-form expressions for Fisher's information matrix.

3.1.1. Model without Covariates

Let $\boldsymbol{\Theta} = (\mu_1, \mu_2)$ and a random sample consisting of n observations $\boldsymbol{x} = \{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$, taken from the probability function (1). The log-likelihood is proportional to:

$$\ell(\boldsymbol{\Theta}; \boldsymbol{x}) \propto n\bar{x}_2 \log \mu_2 + n(\bar{x}_1 - \bar{x}_2) \log(\mu_1 - \mu_2) - n\mu_1,$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means of X_1 and X_2, respectively. The normal equations to be solved are:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\Theta}; \boldsymbol{x})}{\partial \mu_1} &= \frac{n(\bar{x}_1 - \bar{x}_2)}{\mu_1 - \mu_2} - n = 0, \\ \frac{\partial \ell(\boldsymbol{\Theta}; \boldsymbol{x})}{\partial \mu_2} &= \frac{n\bar{x}_2}{\mu_2} + \frac{n(\bar{x}_2 - \bar{x}_1)}{\mu_1 - \mu_2} = 0, \end{aligned}$$

from which it is easy to obtain the solution to obtain the maximum likelihood estimators $\widehat{\mu}_1 = \bar{x}_1$ and $\widehat{\mu}_2 = \bar{x}_2$ which coincide with the moment estimators. The second partial derivatives are:

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\Theta}; \boldsymbol{x})}{\partial \mu_1^2} &= -\frac{n(\bar{x}_1 - \bar{x}_2)}{(\mu_1 - \mu_2)^2}, \\ \frac{\partial^2 \ell(\boldsymbol{\Theta}; \boldsymbol{x})}{\partial \mu_2^2} &= -\frac{n\bar{x}_2}{\mu_2^2} + \frac{n(\bar{x}_2 - \bar{x})}{(\mu_1 - \mu_2)^2}, \\ \frac{\partial^2 \ell(\boldsymbol{\Theta}; \boldsymbol{x})}{\partial \mu_1 \partial \mu_2} &= \frac{n(\bar{x}_1 - \bar{x}_2)}{(\mu_1 - \mu_2)^2}. \end{aligned}$$

The expectation of the negative of the second partial derivative yields Fisher's information matrix:

$$\mathcal{J}(\widehat{\boldsymbol{\Theta}}) = \begin{bmatrix} \frac{n}{\widehat{\mu}_1 - \widehat{\mu}_2} & \frac{n\widehat{\mu}_1}{\widehat{\mu}_2(\widehat{\mu}_1 - \widehat{\mu}_2)} \\ \frac{n\widehat{\mu}_1}{\widehat{\mu}_2(\widehat{\mu}_1 - \widehat{\mu}_2)} & \frac{n}{\widehat{\mu}_2 - \widehat{\mu}_1} \end{bmatrix}.$$

The asymptotic variance-covariance matrix of $(\widehat{\mu}_1, \widehat{\mu}_2)$ is obtained by inverting this information matrix.

3.1.2. Model with Covariates

When covariates are considered, the log-likelihood is proportional to:

$$\ell(\boldsymbol{\beta}; \boldsymbol{x}) \propto \sum_{i=1}^{n} \left[x_{2i} \log \mu_{2i} + (x_{1i} - x_{2i}) \log(\mu_{1i} - \mu_{2i}) - \mu_{1i} \right], \tag{6}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)$.

Observe now that $\mu_{1i} = \mu_{1i}(\beta_1)$ and $\mu_{2i} = \mu_{2i}(\beta_1, \beta_2)$, to emphasize that the first expression depends only on β_1 and the second on both β_1 and β_2. Thus,

$$\frac{\partial \mu_{1i}}{\partial \beta_{1j}} = \omega_{ji} \mu_{1i}, \quad \frac{\partial \mu_{2i}}{\partial \beta_{1j}} = \omega_{ji} \mu_{2i}, \quad \frac{\partial \mu_{2i}}{\partial \beta_{2j}} = \frac{\mu_{2i} \eta_{ji}}{1 + \exp(\eta_{2i})},$$

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

Then, after some algebra, we obtain the normal equations,

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\beta};\boldsymbol{x})}{\partial \beta_{1j}} &= \sum_{i=1}^{n} \omega_{ji}(x_{1i}-\mu_{1i}) = 0, \quad j=1,\ldots,m,\\ \frac{\partial \ell(\boldsymbol{\beta};\boldsymbol{x})}{\partial \beta_{2j}} &= \sum_{i=1}^{n} \frac{\eta_{ji}\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i})}{1+\exp(\eta_{2i}\beta_2)} = 0, \quad j=1,\ldots,m,\end{aligned}$$

where:

$$\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i}) = \frac{x_{2i}\mu_{1i}-x_{1i}\mu_{2i}}{\mu_{1i}-\mu_{2i}}.$$

These equations provide the maximum likelihood estimates for the vector of parameters $\widehat{\beta}_1 = (\widehat{\beta}_{11},\ldots,\widehat{\beta}_{1m})^\top$ and $\widehat{\beta}_2 = (\widehat{\beta}_{21},\ldots,\widehat{\beta}_{2m})^\top$. Similarly to the previous case, Fisher's information matrix can be obtained in closed-form. See the details in Appendix A.

The normal equations illustrated above can be used to estimate model parameters with and without covariates. The Newton–Raphson method provides solutions in a non-prohibitive time, obviously depending on the number of regressors used.

4. Credibility Regression Premiums

Briefly speaking, a bonus-malus system is an experience rating system that is based on the insured's claim experience frequency rather than the claim size. Let us now assume some kind of heterogeneity between policyholders, by allowing that the parameters μ_i, $i = 1,2$ follow some probability functions. For μ_1, a gamma prior distribution will be assumed $\pi_1(\mu_1)$ with a shape hyperparameter $\alpha_1 > 0$ and a scale hyperparameter $\gamma_1 > 0$, whereas a type beta prior distribution will be considered for μ_2 with the probability density function given by:

$$\pi_2(\mu_2) = \frac{\mu_2^{\alpha_2-1}(\mu_1-\mu_2)^{\gamma_2-1}}{\mu_1^{\alpha_2+\gamma_2-1}B(\alpha_2,\gamma_2)}, \quad 0<\mu_2<\mu_1.$$

Here, $\alpha_2 > 0$, $\gamma_2 > 0$, and $B(a,b)$ is the beta function given by $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ where $\Gamma(\cdot)$ is the Euler gamma function.

The main benefit of selecting these prior distributions is that they are conjugate with respect to the likelihoods, and for that reason, they are common choices in Bayesian and actuarial statistics; see for instance Heilmann (1989), Denuit et al. (2009), and Klugman et al. (2008), among others.

Since μ_1 and μ_2 are dependent, we can choose the prior distribution given by:

$$\pi(\mu_1,\mu_2) = \pi_1(\mu_1)\pi_2(\mu_2)[1+\omega\phi_1(\mu_1)\phi_2(\mu_2)], \tag{7}$$

which corresponds to the copula proposed by Lee (1996). Here, $\phi_i(\mu_i)$, $i = 1,2$, are bounded non-constant functions such that $\int \pi_i(\mu_i)\phi_i(\mu_i)\,d\mu_i = 0$, and ω a real number, which satisfies that $1+\omega\phi_i(\mu_i) \geq 0$, $i = 1,2$. Now, given a sample $\boldsymbol{x} = (\tilde{x}_1,\tilde{x}_2) = \{(x_{11},x_{21}),\ldots,(x_{1t},x_{2t})\}$, where t is the sample size, the posterior distribution of (μ_1,μ_2) given the sample information is computed according to Bayes' theorem, and it is proportional to the product of the likelihood and the prior distribution. Thus, the posterior distribution is almost conjugated with respect to the likelihood and similar to the product of a gamma and a beta distribution and where the updated parameters are given by:

$$\alpha_1^* = \alpha_1 + t\bar{x}_1, \tag{8}$$

$$\alpha_2^* = \alpha_2 + t\bar{x}_2, \tag{9}$$

$$\gamma_1^* = \gamma_1 + t, \tag{10}$$

$$\gamma_2^* = \gamma_2 + t(\bar{x}_1-\bar{x}_2). \tag{11}$$

In practise, it is shown that μ_2 is near zero, then in this case, $\omega \to 0$, and the prior distribution reduces to $\pi(\mu_1, \mu_2) = \pi_1(\mu_1)\pi_2(\mu_2)$, which is the case considered here.

Now, the unconditional means and cross moment are given by:

$$\begin{aligned} E(X_1) &= \frac{\alpha_1}{\gamma_1}, \\ E(X_2) &= \frac{\alpha_1}{\gamma_1}\frac{\alpha_2}{\alpha_2+\gamma_2}, \\ E(X_1X_2) &= \frac{\alpha_1\alpha_2(\alpha_1+\gamma_1+1)}{\gamma_1^2(\alpha_2+\gamma_2)}. \end{aligned}$$

Finally, the unconditional bivariate distribution is:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = x_2) &= \frac{\gamma_1^{\alpha_1}}{(1+\gamma_1)^{x_1+\alpha_1}} \\ &\quad \times \frac{\Gamma(x_1+\alpha_1)\Gamma(x_2+\alpha_2)\Gamma(x_1-x_2+\gamma_2)}{(x_1-x_2)!x_2!B(\alpha_2,\gamma_2)\Gamma(\alpha_1)\Gamma(\alpha_2+\gamma_2+x_1)}. \end{aligned} \tag{12}$$

For computational reasons, sometimes, it is more convenient to work with the parametrization $\alpha_1 = \gamma_1\mu_1$ and $\alpha_2 = \gamma_2\mu_2/(\mu_1 - \mu_2)$.

The maximum likelihood estimates for this mixture regression model can be simply obtained by means of the EM algorithm. This method is a powerful technique that provides an iterative procedure to compute maximum likelihood estimation when data contain missing information. Details on the derivation of the EM algorithm can be found in Appendix B. The standard errors of the estimates $\hat{\Omega} = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\gamma}_1, \widehat{\gamma}_2)$ can be computed by using the method given by Louis (1982). Here, we use Fisher's information matrix found in Appendix A and replace the missing values by the corresponding pseudo-values calculated in the last iteration of the EM algorithm. Direct maximization of the likelihood surface is also possible to compute the maximum likelihood estimates of the mixture regression model.

By following the same arguments as those ones provided in Gómez-Déniz (2016) and also based on the ideas in Heilmann (1989) (see also Gerber 1979, Rolski et al. 1999, Bühlmann and Gisler 2005, and Gómez-Déniz 2008; among others), a premium calculation principle assigns to each risk vector of parameters $\boldsymbol{\Theta}$ a premium within the set $P \in \mathbb{R}$, the action space. Let $L(\boldsymbol{\Theta}, P) = (\boldsymbol{\Theta} - P)^2$ be the squared-error loss function sustained by a decision-maker who takes the action P and is faced with the outcome $\boldsymbol{\Theta}$ of a random experience. The premium must be determined in a way such that the expected loss is minimised. The unknown premium $P(\boldsymbol{\Theta})$, called the risk premium, can be obtained by minimising $(g(x_1, x_2) - P)^2$, where $g(x_1, x_2)$ is an appropriate function of the number of claims with a claim size below ψ and above ψ, respectively. It seems reasonable to take $g(x_1, x_2)$ as:

$$g(x_1, x_2) = p_l x_2 + p_s(x_1 - x_2), \tag{13}$$

where p_s, p_l are appropriate weights assigned to the number of claims for claim sizes above and below the critical value, respectively, with $p_s < p_l$. Now, simple algebra provides the risk premium given by,

$$P(\boldsymbol{\Theta}) = E[g(x_1, x_2)] = (p_l + p_s)\mu_1 - p_s\mu_2, \tag{14}$$

where the expectation is taken on (1). By taking $p_l = p_s = 1$ in (14), this reduces to $P(\boldsymbol{\Theta}) = \mu_1$, that is the risk premium depends only on the number of claims, irrespective of their size.

In the absence of experience, the actuary computes the collective premium,

$$P = E_{\pi(\boldsymbol{\Theta})}[P(\boldsymbol{\Theta})] = \frac{\alpha_1(p_s\gamma_2 + p_l(\alpha_2+\gamma_2))}{\gamma_1(\alpha_2+\gamma_2)}. \tag{15}$$

Again, by inserting $p_l = p_s = 1$ into (15), we obtain the collective premium computed under the traditional model, $P = \alpha_1/\gamma_1$. On the other hand, if experience is available, the actuary takes a sample $(\tilde{x}_1, \tilde{x}_2)$ from the random variables (X_1, X_2) and uses this information to estimate the unknown risk premium $P(\boldsymbol{\Theta})$, through the Bayes premium $P^*(\tilde{x}_1, \tilde{x}_2) = E_{\pi(\boldsymbol{\Theta}|(\tilde{x}_1,\tilde{x}_2))}[P(\boldsymbol{\Theta})]$. Due to the fact that the posterior distribution is conjugated with the prior, the Bayes premium can be derived from (15) by simply switching the parameters α_i and γ_i $(i = 1, 2)$ with the updated parameters by using (8)–(11). Furthermore, the Bayesian premium can be rewritten as a credibility expression, i.e., a linear function of the data and the collective premium.

Obviously, the Bayesian premium based on (15) does not depend on the individual's risk factors, and it is only based on the accumulated past claims. Individual's risk factors can be incorporated into the premium by computing $P_i^*(\tilde{x}_1, \tilde{x}_2, \beta_{1i}, \beta_{2i})$, for $i = 1, \ldots, n$. This general pricing formula is a function of the number of accumulated claims and the individual's significant characteristics in the regression component.

Finally, the Bayesian bonus-malus premium is computed as the ratio between the Bayesian premium and the collective premium. This bonus-malus premium is usually normalised by multiplying this ratio by 100.

5. Empirical Results

We will now analyse a dataset that includes information based on one-year vehicle insurance policies taken out in 2004 or 2005. This dataset is available on the website of the Faculty of Business and Economics, Macquarie University (Sydney, Australia) (see also de Jong and Heller 2008). The total portfolio contained 67,856 policies, of which 4624 have at least one claim. With respect to the number of claims, the minimum and maximum were zero and four, respectively. The mean was 0.072, and standard deviation was 0.278. On the other hand, regarding the claim size, the minimum and maximum were zero and 55,922.10, respectively. The mean was 137.27, and the standard deviation was 1056.30. This value was very large for the severity of claims, which meant that a premium based only on the mean claim size was not adequate for computing the bonus-malus premiums. As this portfolio only included the aggregate value of the claims' severity, we followed the approach provided in Gómez-Déniz (2016) to determine the exact value of all claims randomly. Since this portfolio only included the aggregate value of the claim amount for all of the claims in the portfolio, a simulation was performed to determine the exact amount corresponding to each claim. This simulation was carried out by using the Mathematica commands `Permute, RandomChoice, IntegerPartitions, IntegerPart and RandomPermutation`, as shown in the Appendix provided in Gómez-Déniz (2016). It is convenient to note that the partition obtained only provided the integer part, and this did not seem very relevant in the analysis. Furthermore, due to the `RandomChoice` command, the partition was different every time the program was run. The results obtained for the claim amounts via simulation are not shown in this work, but they are available from the authors upon request.

Below in Table 1, the observed (in bold) and expected frequencies with the threshold value for the claims assumed to be $\psi = \$1000$ are shown. For each entry, observed frequencies (top row in bold), expected frequency under the basic model (given by using (1) in the middle row), and mixture model (bottom row), obtained by using (12), are illustrated. Furthermore, the marginal observed and expected frequencies are in the far right column and in the bottom row for X_1 and X_2, respectively. The cells in this table are grouped to comply with the rule of five when applying the χ^2 test.

Similarly, Table 2 exhibits the observed and expected frequencies when the threshold amount was $\psi = \$3000$. Again, the cells are combined to comply with the rule of five. As can be seen, the fitting values obtained by using the mixture model were much more flexible since it incorporated heterogeneity among policyholders via the prior distributions, and it also provided a better fit to the data than those ones computed under the basic model for both thresholds.

Table 1. Observed (in bold) and expected frequencies for threshold value $\psi = \$1000$.

X_1 \ X_2	0	1	2	3	4	Total
0	**63,232**					**63,232**
	63,098.00					63,098.00
	63,279.50					63,279.50
1	**2551**	**1782**				**4333**
	2713.21	1874.01				4587.22
	2518.34	1768.19				4286.53
2	**109**	**114**	**48**			**271**
	58.33	80.58	27.83			166.74
	101.75	116.10	54.15			272
3	**5**	**6**	**6**	**1**		**18**
	0.83	1.73	1.20	0.27		4.03
	4.26	6.14	4.66	1.81		16.87
4	**1**	**0**	**0**	**1**	**0**	**2**
	0.01	0.02	0.01	0.02	0.01	0.07
	0.18	0.31	0.29	0.18	0.06	1.02
Total	**65,110**	**1902**	**54**	**2**	**0**	**67,856**
	65,870.38	1956.34	29.04	0.29	0.01	67,856.06
	65,904.03	1890.74	59.10	1.99	0.06	67,856.00

Table 2. Observed (in bold) and expected frequencies for threshold value $\psi = \$3000$.

X_1 \ X_2	0	1	2	3	4	Total
0	**63,232**					**63,232**
	63,098.00					63,098.00
	63,279.50					63,279.50
1	**3576**	**757**				**4333**
	3817.42	769.79				4587.21
	3554.25	732.28				4286.53
2	**216**	**44**	**11**			**271**
	115.48	46.57	4.69			166.74
	198.16	54.75	19.09			272
3	**12**	**4**	**2**	**0**		**18**
	2.33	1.41	0.28	0.01		4.03
	11.13	3.47	1.62	0.64		16.86
4	**2**	**0**	**0**	**0**	**0**	**2**
	0.03	0.03	0.01	0.00	0.00	0.07
	0.63	0.21	0.11	0.06	0.02	1.03
Total	**67038**	**805**	**13**	**0**	**0**	**67,856**
	67,033.26	815.80	4.98	0.01	0.00	67,856.05
	67,043.67	790.71	20.82	0.70	0.02	67,856.00

Maximum likelihood estimation was used in both cases. It is convenient to point out that in the case of the mixture model, it was proven that directly maximizing the logarithm of the log-likelihood function provided, as expected, the same results as using the EM algorithm shown in Appendix B of this work. Mathematica and WinRaTs were the two packages used in this case.

Parameter estimates, standard errors (in brackets), the maximum of the log-likelihood function, figures of the chi-squared test statistics, degrees of freedom (d.f.), and the p-value are exhibited in Table 3 for the basic and mixture models. Results under the threshold value first $\psi = \$1000$ are shown in the second and third columns and $\psi = \$3000$ in the last two columns. Virtually, the same estimates were obtained for parameters μ_1 and μ_2 under the basic and mixture models. Similarly, no changes were discernible in the estimates between the estimates for the two thresholds with the exemption of the estimate of parameter γ_2. In this case, it was observable that the estimate decreased when the threshold increased. By incrementing the threshold value, the fit to the data improved. The mixture

model provided the best fit to the data in terms of the χ^2 test statistic and the negative of the maximum of the likelihood function $\ell_{\max}$. Note that the mixture model was not rejected at the 5% significance level for the two thresholds previously considered. It is important to note that, although the gain in terms of maximum of the log-likelihood function did not seem significant, the mixture model was preferable in terms of the χ^2 test statistics since, unlike the basic one, it was not rejected at the 5% significance level (see the corresponding p-values) in either of the two thresholds mentioned above.

Table 3. Parameter estimates (in brackets) and measures of model selection for the basic and mixture models without covariates.

	$\psi =$ \$1000		$\psi =$ \$3000	
	Basic Model	**Mixture Model**	**Basic Model**	**Mixture Model**
$\widehat{\mu}_1$	0.0727	0.0727	0.0727	0.0727
	(0.001)	(0.000)	(0.001)	(0.000)
$\widehat{\mu}_2$	0.0297	0.0297	0.0122	0.0123
	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\gamma}_1$		15.900		15.900
		(0.000)		(0.000)
$\widehat{\gamma}_2$		4.334		2.035
		(0.000)		(0.000)
$\ell_{\max}$	−21,346.561	−21,292.395	−20,301.926	−20,242.391
χ^2	>100	5.16	>100	2.09
d.f.	4	2	3	1
p-value	0.00%	7.58%	0.00 %	14.83%

We now implement explanatory variables in our analysis. The following covariates were considered: gender of driver, vehicle body, driver's area of residence, age of vehicle, and driver's age category. In addition, an intercept was also included in the study. Details about the codification of these variables can be found on the same website. Moreover, an offset variable (exposure, log of the time exposed to risk) was included in the linear predictor associated with the first variable.

Table 4 illustrates the estimates of the regressors for the mixture model associated with the random variables X_1 and X_2 again for a threshold of $\psi =$ \$1000 and $\psi =$ \$3000. In the first case, the explanatory variables hardtop (HDTOP), motorized caravan (MCARA), driver's area of residence C (AREAC), age of Vehicles 1 and 2 (VAGE1 and VAGE2), and driver's Age Category 1 (AGE 1) were statistically significant at the 5% significance level for the random variable total number of claims given that the claim size exceeded $\psi =$ \$1000. Among these variables, only HDTOP, MCARA, VAGE1, VAGE2, and AGE1 were significant for both response variables. However, it is important to note that all these variables except for the regressors associated with AGE1 and AGE2, the sign of the estimates changed from positive to negative for claims above the threshold. Furthermore, the estimate of parameter γ_1 was statistically significant at the same nominal level. When the threshold value was increased up to $\psi =$ \$3000, the number of significative variables above the threshold considerably grew since now, the intercept (CONSTANT), gender of driver (GENDER), HDTOP, SEDAN, station wagon (STNWG), TRUCK, AREAA, AREAB, AREAC, AREAD, VAGE1, VAGE2, and AGE1, were relevant. However, only CONSTANT, HDTOP, STNWG, AREAD, VAGE1, VAGE2, and AGE1 were significant for both dependent variables at the same nominal level. The regressors associated with the explanatory variables CONSTANT, AREAD, and the AGE1 had the same sign for claims below and above $\psi =$ \$3000. The first two regressors were negatively correlated and the latter one positively correlated to the response variables, respectively. For the other regressors, once again, the sign of the estimates changed from positive to negative for claims above the threshold. Among the common statistically significant estimates for both threshold values, i.e., HDTOP, VAGE1, VAGE2, and AGE1, the same sign of the estimates in the variables X_1 and X_2 was observable. For the non-significant estimates, different signs were observed in the regressors. Furthermore, the estimates of parameters γ_1 and γ_2 were statistically significant at the same nominal level.

Table 4. Parameter estimates and p-values associated with the Wald test for the mixture model including covariates.

	$\psi = 1000$				$\psi = 3000$			
	Variable X_1		**Variable X_2**		**Variable X_1**		**Variable X_2**	
Parameter	**Estimate**	**p-Value**	**Estimate**	**p-Value**	**Estimate**	**p-Value**	**Estimate**	**p-Value**
GENDER	−0.015	0.613	0.105	0.090	−0.022	0.467	0.220	0.007
BUS	0.244	0.610	−0.384	0.558	1.005	0.002	−1.386	0.208
CONVT	−0.562	0.342	−0.406	0.738	−0.525	0.364	0.890	0.461
COUPE	0.503	0.000	0.204	0.401	0.489	0.000	0.007	0.982
HDTOP	0.208	0.024	−0.427	0.026	0.181	0.049	−0.682	0.011
MCARA	0.766	0.003	−1.291	0.050	0.668	0.011	−1.495	0.152
MIBUS	0.098	0.514	0.292	0.342	0.018	0.905	−0.501	0.234
PANVN	0.124	0.335	−0.286	0.272	0.132	0.299	−0.415	0.223
RDSTR	0.131	0.856	−0.278	0.823	0.318	0.624	−1.143	0.672
SEDAN	0.063	0.098	−0.148	0.055	0.058	0.128	−0.348	0.001
STNWG	0.124	0.002	−0.150	0.076	0.107	0.010	−0.471	0.000
TRUCK	0.055	0.570	−0.040	0.835	0.056	0.560	−0.506	0.050
UTE	−0.100	0.152	−0.054	0.699	−0.111	0.110	−0.271	0.126
AREAA	−0.010	0.885	−0.194	0.152	−0.064	0.343	−0.108	0.001
AREAB	0.050	0.472	−0.207	0.132	−0.005	0.938	−0.571	0.004
AREAC	0.007	0.920	−0.293	0.027	−0.053	0.421	−0.496	0.035
AREAD	−0.110	0.144	−0.139	0.352	−0.171	0.021	−0.345	0.003
AREAE	−0.037	0.641	−0.125	0.420	−0.093	0.228	−0.572	0.293
VAGE1	0.187	0.000	−0.388	0.000	0.168	0.000	−0.271	0.000
VAGE2	0.219	0.000	−0.259	0.001	0.207	0.000	−0.619	0.009
VAGE3	0.098	0.013	−0.010	0.208	0.083	0.035	−0.275	0.283
AGE1	0.512	0.000	0.291	0.034	0.464	0.000	0.746	0.000
AGE2	0.328	0.000	0.032	0.795	0.286	0.000	0.274	0.118
AGE3	0.275	0.000	0.039	0.746	0.229	0.000	0.273	0.111
AGE4	0.243	0.000	−0.043	0.723	0.202	0.001	0.196	0.253
AGE5	0.030	0.656	−0.044	0.740	−0.013	0.843	−0.002	0.990
CONSTANT	−2.273	0.000	0.027	0.880	−2.156	0.000	−1.045	0.000
$\widehat{\gamma}_1$	21.602	0.000			30.718	0.000		
$\widehat{\gamma}_2$	5.903	0.185			2.205	0.014		

Similarly to the case previously considered, the fit to the data improved when covariates were incorporated in the model and when the threshold value enlarged. Table 5 exhibits the negative of the maximum of the likelihood function ($-\ell_{\max}$), Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and the consistent Akaike's information criterion (CAIC) for the basic and mixture regression models. A lower value of these measures of model selection was desirable. It was observable that the latter model was preferable to the former one.

Table 5. Parameter estimates (in brackets) and measures of model selection for the basic and mixture models with covariates.

	$\psi =$ \$1000		$\psi =$ \$3000	
	Basic Model	**Mixture Model**	**Basic Model**	**Mixture Model**
$-\ell_{\max}$	20,604.355	20,588.936	19,545.212	19,511.783
AIC	41,312.710	41,289.872	39,198.422	39,135.565
BIC	41,809.468	41,800.880	39,691.180	39,646.573
CAIC	41,863.468	41,856.880	39,745.180	39,702.573

We plot the QQ-plots of the randomized quantile residuals to check for normality in Figure 2. The residuals for the basic regression models are shown in the top row and for the mixture regression model in the bottom row. Furthermore, models that use $\psi =$ \$1000 as the threshold value are exhibited in the left column and $\psi =$ \$3000 in the right-hand column. A perfect alignment with the 45° line

implies that the residuals are normally distributed. It was observable that the residuals for the larger threshold values adhered a little bit closer to the line, but these differences were not significant.

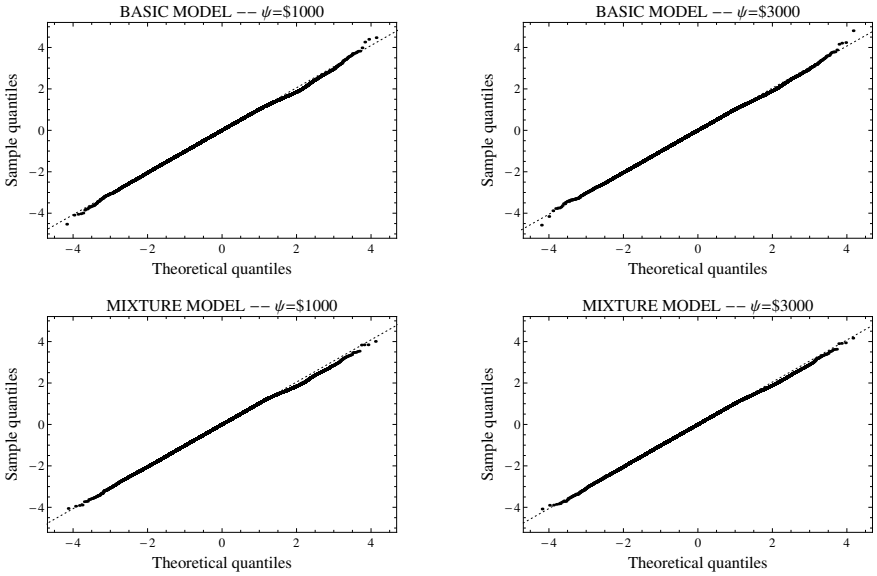

Figure 2. QQ-plots of the randomized quantile residuals for the basic (top) and mixture (bottom) regression models for $\psi = \$1000$ (left) and $\psi = \$3000$ (right) threshold values.

Figure 3 exhibits the bonus-malus premiums (BMP) for the mixture model without covariates. Here, x_1 is the total number of claims when x_2 claims out of x_1 have a size larger than ψ. In each chart, the thick line represents $\psi = \$1000$, and the thin line denotes $\psi = \$3000$. It was noticeable that the BMP decreased with the time period when the observed pair x_1 and x_2 was fixed for the two thresholds considered. The BMP was consistently lower when the threshold ψ decreased. Although for both values of ψ, the premium charged increased when x_1 and x_2 grew, the premium paid also increased with x_2 when x_1 was fixed.

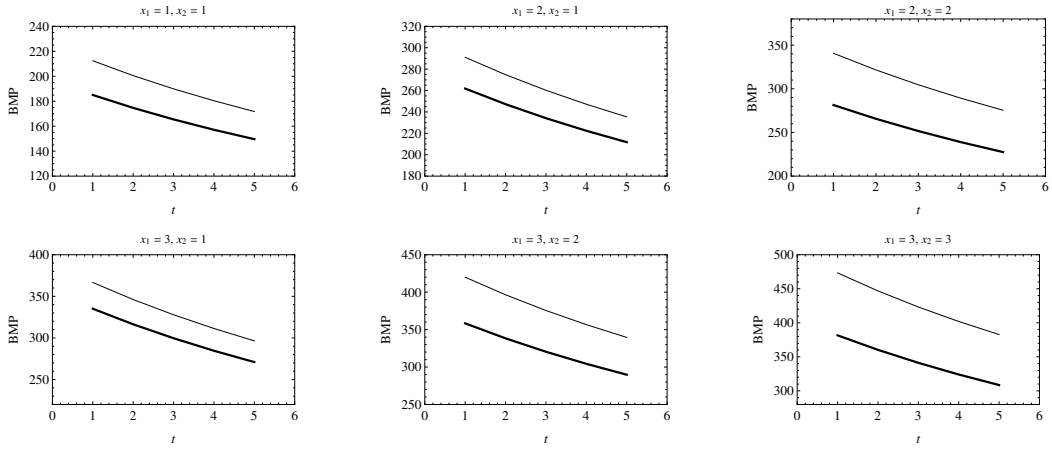

Figure 3. Bayesian bonus-malus premiums under the mixture model without covariates for x_1 claims when there are x_2 claims with a claim size larger than ψ. The thick line represents $\psi = \$1000$, and the thin line represents $\psi = \$3000$. BMP, bonus-malus premiums.

Figure 4 illustrates the bonus-malus premiums (BMP) to be charged to the subgroup of policyholders with SEDAN and AREAA. In this case, we used the mixture regression model including the rest of the explanatory variables and the exposure. Similar conclusions could be drawn from this

set of graphs. Again, the BMP was persistently lower when the threshold ψ decreased. The premium charged increased when x_1 and x_2 grew for either value of ψ; moreover, the premium paid rose with x_2 when x_1 was held fixed. As compared to the premiums obtained under this regression model were way higher than those ones derived before, this could be surely explained by the small sample size used to estimate regressors and also for the incorporation of the offset variable that without any doubt affected the individual average number of claims and the probability of making a claim higher than the threshold. Other different subgroups of policyholders could also be used for tarification purposes; however, for some of these classes, non-reliable estimates were obtained due to the very low sample size.

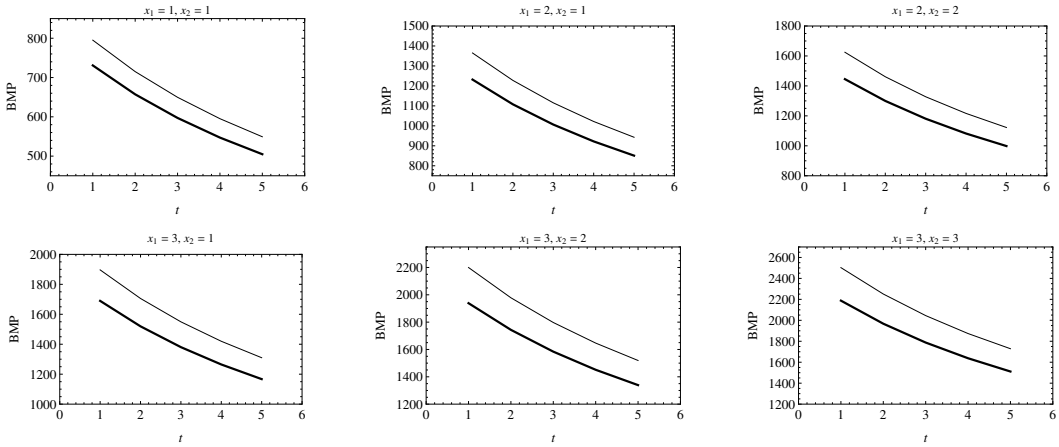

Figure 4. Bayesian bonus-malus premiums under the mixture model with covariates for x_1 claims when there are x_2 claims with a claim size larger than ψ. The thick line represents $\psi = \$1000$, and the thin line represents $\psi = \$3000$. This chart corresponds to the the subgroup of policyholders with SEDAN and AREAA.

Computations in the Compound Model

Although it is customary to calculate the bonus-malus premium based on the variable number of claims (it is usually considered that once a loss has occurred, the company does not have the ability to model the amount corresponding to the loss), some attempts have been made to implement the severity in the calculation of the premium. Some works related to this topic are Frangos and Vrontos (2001), Pinquet (1998), and Gómez-Déniz et al. (2014), among others. As the practitioner wishes to calculate the premium using both variables, it is useful to rely on the composite collective model. Similarly to the univariate case, the bivariate compound distributions for the aggregate claim size random variable can be simply derived as follows:

$$g(y_1, y_2) = \sum_{x_1, x_2 = 0}^{\infty} p_{x_1, x_2} \, f_1^{*\, x_1}(y_1) \, f_2^{*\, x_2}(y_2), \tag{16}$$

and this is the the joint probability density function of $(Y_1, Y_2) = (S_1, S_2,)$, where $S_1 = \sum_{i=0}^{X_1} Y_{1i}$, $S_2 = \sum_{i=0}^{X_2} Y_{2i}$ are the aggregate severities, Y_1 and Y_2 being mutually independent and also independent of (X_1, X_2) with probability functions (discrete or continuous) $f_1(y_1)$, $f_2(y_2)$, respectively, with x_1 and x_2-fold convolutions $f_1^{*\, x_1}(y_1)$ and $f_2^{*\, x_2}(y_2)$, respectively. General expressions for $E(S)$, $var(S)$ and $cov(S_1, S_2)$, where $S = S_1 + S_2$, were provided in Partrat (1994).

Recursion for bivariate count distributions and their compound distributions given in the form (16) have been previously considered in the actuarial literature; see Theorem 2.1. in Hesselager (1996). Other similar recursions can be found in Vernic (1997), Walhin and Paris (2000), Walhin and Paris (2001), Sundt (2002), and Sundt and Vernic (2009), among others. Moreover, bivariate recursions are

useful in prediction problems involving the conditional $g(y|x)$ of Y, given $X = x$; see Hesselager (1996) for more details.

Let us now assume that the random variables X_1 and X_2 represent two kinds of claims, for instance bodily injury and material damage, or as in our study, claims below and above a threshold ψ.

The fact that the probability generating function of (1) is analytically obtained helps us to derive the probability generating function of the joint random variable $(X_1(d_1), X_2(d_2))$ for d_i, which can be deduced in type i $(i = 1, 2)$ claim amounts. Here, $X_i(d_i)$ is the random variable corresponding to the yearly frequency of type i claims exceeding d_i. The work in Partrat (1994) then showed that the probability generating function of the random variable $(X_1(d_1), X_2(d_2))$ is given by:

$$\begin{aligned} G_{X_1(d_1),X_2(d_2)}(s_1, s_2) &= G_{X_1,X_2}((1 - F_1(d_1))s_1 \\ &\quad + F_1(d_1), (1 - F_2(d_2))s_2 + F_2(d_2)), \end{aligned}$$

where F_1 and F_2 are the cumulative distribution functions of the random variables Y_1 and Y_2, respectively; while the probability generating function of the random variable $X(d_1, d_2)$, with $X = X_1 + X_2$, is given by:

$$G_{X(d_1,d_2)}(s_1, s_2) = G_{X_1,X_2}((1 - F_1(d_1))s_1 + F_1(d_1), (1 - F_2(d_2))s_2 + F_2(d_2)).$$

6. Final Comments

In this paper, a flexible bivariate count data regression model that let us distinguish between different types of claims according to the claim size was introduced. Besides, it allowed us to examine the factors that affect the number of claims above and below a given claim size threshold. By means of a mixture regression model, the individual claim size and other risk factors such as gender, type of vehicle, driving area, or age of the vehicle could be used to compute credibility bonus-malus premiums. Extensions of this work includes a simple modification of this model to differentiate between more than two claims in the line of the work provided in Gómez-Déniz and Calderín-Ojeda (2018). Besides, a similar model can be simply implemented when the number of claims is distributed according to a negative binomial distribution. A study of this nature would be a possible extension of this work.

Author Contributions: E.G.-D. and E.C.-O. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: E.G.D. was partially funded by grant ECO2017–85577–P (Ministerio de Economía, Industria y Competitividad. Agencia Estatal de Investigación).

Acknowledgments: The authors wish to acknowledge the Associate Editor and three anonymous referees for the constructive comments that helped to improve the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The second partial derivatives are provided by:

$$\begin{aligned} \frac{\partial^2 \ell(\beta_1, \beta_2; \boldsymbol{x})}{\partial \beta_{1j}^2} &= -\sum_{i=1}^{n} \omega_{ji}^2 \mu_{1i}, \quad j = 1, \ldots, m, \\ \frac{\partial^2 \ell(\beta_1, \beta_2; \boldsymbol{x})}{\partial \beta_{1j} \partial \beta_{1k}} &= -\sum_{i=1}^{n} \omega_{ji} \omega_{ki} \mu_{1i}, \quad j \neq k, \\ \frac{\partial^2 \ell(\beta_1, \beta_2; \boldsymbol{x})}{\partial \beta_{1j} \beta_{2j}} &= 0, \quad j = 1, \ldots, m, \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell(\beta_1,\beta_2;\boldsymbol{x})}{\partial \beta_{2j}^2} &= -\sum_{i=1}^{n}\left(\frac{\eta_{ji}}{1+\exp(\eta_{2i}\beta_2)}\right)^2 [\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i})\exp(\eta_{2i}\beta_2) \\
&\quad + \frac{(x_{1i}-\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i}))\mu_{2i}}{\mu_{1i}-\mu_{2i}}\Bigg], \quad j=1,\ldots,m. \\
\frac{\partial^2 \ell(\beta_1,\beta_2;\boldsymbol{x})}{\partial \beta_{2j}\partial \beta_{2k}} &= -\sum_{i=1}^{n}\frac{\eta_{ji}\eta_{ki}}{(1+\exp(\eta_{2i}\beta_2))^2} [\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i})\exp(\eta_{2i}\beta_2) \\
&\quad + \frac{(x_i-\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i}))\mu_{2i}}{\mu_{1i}-\mu_{2i}}\Bigg], \quad j=1,\ldots,m.
\end{aligned}$$

Now, the entries of Fisher's information matrix (with dimension $m \times m$) are given by:

$$\begin{aligned}
E\left(-\frac{\partial \ell(\widehat{\beta}_1,\widehat{\beta}_2;\boldsymbol{x})}{\partial \widehat{\beta}_{1j}^2}\right) &= \sum_{i=1}^{n}\omega_{ji}^2\widehat{\mu}_{1i}, \\
E\left(-\frac{\partial^2 \ell(\widehat{\beta}_1,\widehat{\beta}_2;\boldsymbol{x})}{\partial \widehat{\beta}_{1j}\partial \widehat{\beta}_{1k}}\right) &= \sum_{i=1}^{n}\omega_{ji}\omega_{ki}\widehat{\mu}_{1i}, \quad j\neq k, \\
E\left(-\frac{\partial \ell(\widehat{\beta}_1,\widehat{\beta}_2;\boldsymbol{x})}{\partial \widehat{\beta}_{1j}\partial \widehat{\beta}_{2j}}\right) &= 0, \\
E\left(-\frac{\partial \ell(\widehat{\beta}_1,\widehat{\beta}_2;\boldsymbol{x})}{\partial \beta_{2j}^2}\right) &= \sum_{i=1}^{n}\frac{\widehat{\mu}_{1i}\widehat{\mu}_{2i}}{\widehat{\mu}_{1i}-\widehat{\mu}_{2i}}\left(\frac{\eta_{ji}}{1+\exp(\eta_{2i}\widehat{\beta}_2)}\right)^2, \\
E\left(-\frac{\partial \ell(\widehat{\beta}_1,\widehat{\beta}_2;\boldsymbol{x})}{\partial \beta_{2j}\partial \beta_{2k}}\right) &= \sum_{i=1}^{n}\frac{\widehat{\mu}_{1i}\widehat{\mu}_{2i}}{\widehat{\mu}_{1i}-\widehat{\mu}_{2i}}\frac{\eta_{ji}\eta_{ki}}{(1+\exp(\eta_{2i}\widehat{\beta}_2))^2}, \quad j\neq k,
\end{aligned}$$

for $j = 1,\ldots,m$, where we have taken into account that $E(\phi(\mu_{1i},\mu_{2i},x_{1i},x_{2i})) = 0$. Again, the asymptotic variance-covariance matrix of $(\widehat{\beta}_1,\widehat{\beta}_2)$ is obtained by inverting this information matrix.

Appendix B

Given the vector of complete data $\boldsymbol{x}$ and the vector of missing observations $(\tilde{\delta}_1,\tilde{\delta}_2) = \{(\tilde{\delta}_{11},\tilde{\delta}_{21}),\ldots,(\tilde{\delta}_{1n},\tilde{\delta}_{2n})\}$, then the complete data log-likelihood takes the form:

$$\begin{aligned}
\ell(\beta_1,\beta_2,\gamma_1,\gamma_2) &\propto \sum_{i=1}^{n} x_{2i}\log\delta_{2i}\mu_{2i} - (x_{1i}-x_{2i})\log(\delta_{1i}\mu_{1i}-\delta_{2i}\mu_{2i}) - \delta_{1i}\mu_{1i} \\
&+ n\gamma_1\log\gamma_1 + (\gamma_1-1)\sum_{i=1}^{n}\log\delta_{1i} - \gamma_1\sum_{i=1}^{n}\delta_{1i} \\
&+ (\gamma_2-1)\sum_{i=1}^{n}\log\delta_{2i} + (\gamma_2-1)\sum_{i=1}^{n}\log(\delta_{1i}-\delta_{2i}) \\
&- (2\gamma_2-1)\sum_{i=1}^{n}\log\delta_{1i} - n\log B(\gamma_2,\gamma_2). \qquad \text{(A1)}
\end{aligned}$$

Expression (A1) can be divided into two parts; the regressors are included in the first part, and the mixing distributions appear only in the second part (i.e., parameters γ_1 and γ_1). Furthermore, we assume, without loss of generality, that to make the model identifiable, $E_{\pi_1}(\delta_1) = 1$ and $E_{\pi_2}(\delta_2) = 1/2$. The EM algorithm is based on two steps. The E-step, i.e., expectation, fills in the missing data. Once the missing data are built-in, the parameters are estimated in the M-step, i.e., maximization. The regressors are estimated using the pseudo-values, $E(\delta_{1i}|\tilde{x}_1,\tilde{x}_2)$ and $E(\delta_{2i}|\tilde{x}_1,\tilde{x}_2)$ as offset variables and then

fitting the regression model given in (6). Then, to estimate the parameters γ_1 and γ_2, we maximize the log-likelihood of the mixing distributions, replacing the missing observations with their expectations. Next, if some terminating condition is achieved, then stop iterating, otherwise move back to the E-step for more iterations.

From the current estimates after the k^{th} iteration, the new estimates $(\widehat{\beta_1}^{(k)}, \widehat{\beta_2}^{(k)}, \widehat{\gamma}_1^{(k)}, \widehat{\gamma}_2^{(k)})$ are obtained as follows:

E-step: Consider:

$$n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i}) = \frac{m(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})}{\int_0^\infty \int_0^{\delta_{1i}} m(\delta_{1i}, \delta_{2i}, x_{1i}, x_{2i})\, d\delta_{2i}\, d\delta_{1i}},$$

where:

$$m(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i}) = (\delta_{2i}\mu_{2i})^{x_{2i}} (\delta_{1i}\mu_{1i} - \delta_{2i}\mu_{2i})^{x_{1i}-x_{2i}} \exp(-\delta_{1i}\mu_{1i}) \pi_1(\delta_{1i}) \pi_2(\delta_{2i}).$$

For all $i = 1, 2, \ldots, n$, we calculate:

$$\begin{aligned}
c_i &= E(\delta_{1i}|\boldsymbol{x}) = \int_0^\infty \int_0^{\delta_{1i}} \delta_{1i}\, n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})\, d\delta_{2i}\, d\delta_{1i}, \\
d_i &= E(\log \delta_{1i}|\boldsymbol{x}) = \int_0^\infty \int_0^{\delta_{1i}} \log(\delta_{1i})\, n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})\, d\delta_{2i}\, d\delta_{1i}, \\
m_i &= E(\delta_{2i}|\boldsymbol{x}) = \int_0^\infty \int_0^{\delta_{1i}} \delta_{2i}\, n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})\, d\delta_{2i}\, d\delta_{1i}, \\
n_i &= E(\log \delta_{2i}|\boldsymbol{x}) = \int_0^\infty \int_0^{\delta_{1i}} \log(\delta_{2i})\, n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})\, d\delta_{2i}\, d\delta_{1i}, \\
s_i &= E(\log(\delta_{1i} - \delta_{2i})|\boldsymbol{x}) = \int_0^\infty \int_0^{\delta_{1i}} \log(\delta_{1i} - \delta_{2i})\, n(\delta_{1i}, \delta_{2i}, \mu_{1i}, \mu_{2i})\, d\delta_{2i}\, d\delta_{1i}.
\end{aligned}$$

M-step: This step works as follows:

- Update the regressors $\widehat{\beta}_j^{(k+1)}$, $j = 1, 2$, using the pseudo-values c_i and m_i as offset variables by fitting a the regression model given in (6), and then,
- Update the estimate of the parameters $\widehat{\gamma}_1^{(k+1)}$ and $\widehat{\gamma}_2^{(k+1)}$ by using:

$$\begin{aligned}
\widehat{\gamma}_1^{(k+1)} &= \exp\left(\frac{1}{n}\sum_{i=1}^n c_i + \psi\left(\widehat{\gamma}_1^{(k)}\right) - 1 - \frac{1}{n}\sum_{i=1}^n d_i\right) \\
\widehat{\gamma}_2^{(k+1)} &= \frac{1}{2}\psi^{-1}\left(\frac{1}{n}\sum_{i=1}^n n_i + \frac{1}{n}\sum_{i=1}^n s_i - 2\frac{1}{n}\sum_{i=1}^n d_i\right),
\end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

Stop iterating if some terminating condition is satisfied.

The following result concerns the concept of multivariate log-concavity, which was introduced by Bapat (1988). See also Johnson et al. (1997).

Proposition A1. *The probability function given in (1) is generalized log-concave.*

Proof. To see this, observe that (1) can be rewritten as:

$$\Pr(X_1 = x_1, X_2 = x_2) = m(\boldsymbol{x}, \boldsymbol{\Theta}) \prod_{i=1}^2 f_i(x_i),$$

where:

$$\begin{aligned} m(\boldsymbol{x},\boldsymbol{\Theta}) &= \frac{x_1!(\mu_1-\mu_2)^{x_1-x_2}\exp(-\mu_1)}{\mu_1^{x_1}(x_1-x_2)!}, \\ f_i(x_i) &= \frac{\mu_i^{x_i}}{x_i!}. \end{aligned}$$

Since $f_i(x_i)$ are log-concave functions $(f_i(x_i)^2 \geq f_i(x_i-1)f_i(x_i+1),\ i=1,2,\ x_i=1,2,\ldots)$, then the result follows by applying Theorem 3 in Bapat (1988). □

The next result shows that the proposed distribution is strongly unimodal (see Barndorff-Nielsen 1973 and Pedersen 1975).

Proposition A2. *The probability function given in (1) is strongly unimodal.*

Proof. Taking into account that for $x_1 = 1,2,\ldots,$, $x_2 = 1,\ldots,x_1$, it is verified that:

$$\begin{aligned} \frac{p_{x_1,x_2}p_{x_1-1,x_2-1}}{p_{x_1-1,x_2}p_{x_1,x_2-1}} &= 1+\frac{1}{x_1-x_2} \geq 1, \\ \frac{p_{x_1,x_2}p_{x_1-1,x_2}}{p_{x_1,x_2+1}p_{x_1-1,x_2-1}} &= 1+\frac{1}{x_2} \geq 1, \\ \frac{p_{x_1,x_2}p_{x_1,x_2-1}}{p_{x_1+1,x_2}p_{x_1-1,x_2-1}} &= 1+\frac{1}{x_1-x_2} \geq 1, \end{aligned}$$

being $p_{x_1,x_2} = \Pr(X_1 = x_1, X_2 = x_2)$, and we get the result after applying Condition (b) in Theorem 1 in Pedersen (1975). □

References

Bapat, Ravindra B. 1988. Discrete multivariate distributions and generalized log-concavity. *Sankhyā: The Indian Journal of Statistics, Series A* 1: 98–110.

Barndorff-Nielsen, Ole. 1973. Unimodality and exponential families. *Communications in Statistics-Theory and Methods* 1: 189–216.

Bermúdez, Lluís. 2009. A priori ratemaking using bivariate Poisson regression models. *Insurance: Mathematics and Economics* 44: 135–41. [CrossRef]

Bermúdez, Lluís, and Dimitris Karlis. 2017. A priori ratemaking using bivariate Poisson models. *Scandinavian Actuarial Journal* 2: 148–58. [CrossRef]

Bonsdorff, Heikki. 2005. On asymptotic properties of Bonus-Malus systems based on the number and on the size of the claims. *Scandinavian Actuarial Journal* 4: 309–20. [CrossRef]

Bühlmann, Hans, and Alois Gisler. 2005. *A Course in Credibility Theory and Its Applications*. Berlin: Springer.

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

de Jong, lPiet, and Gillian H. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.

Denuit, Michel, Xavier Marèchal, Sandra Pitrebois, and Jean F. Walhin. 2009. *Actuarial Modelling of Claim Counts Risk Classification, Credibility and Bonus-Malus Systems*. New York: John Wiley & Sons.

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

Frangos, Nikolaos, and Spyridon Vrontos. 2001. Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin* 31: 1–22. [CrossRef]

Gerber, Hans U. 1979. *An Introduction to Mathematical Risk Theory*. Homewood: Huebner Foundation Monograph.

Gómez-Déniz, Emilio. 2008. A generalization of the credibility theory obtained by using the weighted balanced loss function. *Insurance: Mathematics and Economics* 42: 850–54. [CrossRef]

Gómez-Déniz, Emilio. 2016. Bivariate credibility bonus-malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics* 70: 117–24. [CrossRef]

Gómez-Déniz, Emilio, and Enrique Calderín-Ojeda. 2018. Multivariate credibility in bonus-malus systems distinguishing between different types of vlaims. *Risks* 6: 34. [CrossRef]

Gómez-Déniz, Emilio, Agustín Hernández, and María P. Fernández. 2014. Computing credibility bonus-malus premiums using the total claim amount distribution. *Hacettepe Journal of Mathematics and Statistics* 43: 1047–61.

Heilmann, Wolf R. 1989. Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* 8: 75–95. [CrossRef]

Hesselager, Ole. 1996. Recursions for certain bivariate counting distributions and their compound distributions. *ASTIN Bulletin* 26: 35–52. [CrossRef]

Johnson, Norman, Samuel Kotz, and Narayanaswamy Balakrishnan. 1997. *Discrete Multivariate Distributions*. New York: Wiley.

Khatri, Chinubhai G. 1983a. Multivariate discrete exponential distributions and their characterization by Rao-Rubin condition for additive damage model. *South African Statistical Journal* 17: 13–32.

Khatri, Chinubhai G. 1983b. Multivariate discrete exponential family of distributions. *Communications in Statistics-Theory and Methods* 12: 877–93. [CrossRef]

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2008. *Loss Models: From Data to Decisions*, 3rd ed. New York: Wiley.

Kocherlakota, Subrahmaniam, and Kathleen Kocherlakota. 1992. *Bivariate Discrete Distributions*. New York: Marcel Dekker.

Lee, Mei-Ling T. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods* 25: 1207–22.

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

Partrat, Christian. 1994. Compound model for two dependent kinds of claims. *Insurance: Mathematics and Economics* 15: 219–31. [CrossRef]

Pedersen, Jean G. 1975. On strong unimodality of two-dimensional discrete distributions with applications to M-Ancillarity. *Scandinavian Journal of Statistics* 2: 99–102.

Pinquet, Jean. 1998. Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin* 28: 205–20. [CrossRef]

Ragulina, Olena. 2011. Bonus-malus systems with different claim types and varying deductibles. *Modern Stochastics: Theory and Applications* 4: 141–59. [CrossRef]

Rolski, Tomasz, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugel. 1999. *Stochastic Processes for Insurance and Finance*. Hoboken: John Wiley & Sons.

Sundt, Bjoern. 2002. Recursive evaluation of aggregate claims distributions. *Insurance: Mathematics and Economics* 30: 297–322. [CrossRef]

Sundt, Bjoern, and Raluca Vernic. 2009. *Recursions for Convolutions and Compound Distributions with Insurance Applications*. New York: Springer.

Vernic, Raluca. 1997. On the bivariate generalized Poisson distribution. *ASTIN Bulletin* 27: 23–31. [CrossRef]

Walhin, Jean F., and John Paris. 2000. Recurs1ve formulae for some bivariate counting distributions obtained by the trivariate reduction method. *ASTIN Bulletin* 30: 141–55. [CrossRef]

Walhin, Jean F., and John Paris. 2001. The mixed bivariate Hofmann distribution. *ASTIN Bulletin* 31: 127–42. [CrossRef]