


Tariff systems for fleets of vehicles: a study on the portfolio of Fidelidade

Tiago Fardilha¹ · Maria de Lourdes Centeno²  · Rui Esteves¹

Received: 25 February 2016 / Revised: 20 June 2016 / Accepted: 18 August 2016 /
Published online: 13 September 2016
© EAJ Association 2016

Abstract In Portugal, insurance policies for fleets of vehicles are, in general, similar to policies for individual vehicles. Such is the case of the insurance company at hand. The experience rating system is practically the same as in individual motor insurance and is applied independently to each vehicle, thus having no effect on the premium paid by other vehicles in the fleet. This experience rating system is inefficient since it ignores the potential fleet-specific risks in the *a posteriori* tariff. We considered two credibility-based experience rating schemes proposed by Desjardins, Dionne and Pinquet in 2001. One is based on the claims numbers at fleet level and the other is based on the claims numbers at vehicle level. We applied both models in order to calculate experience rating coefficients for the vehicles in the portfolio of fleets of the insurer. We propose different estimators for the structure parameters of the model, which in our opinion handle better the heterogeneity of the time exposures of our data set.

Keywords Motor insurance · Fleets · Experience rating · *bonus–malus* · Credibility

Maria de Lourdes Centeno was partially supported by the project CEMAPRE-MULTI/00491 financed by FCT/MEC through national funds and when applicable co-financed by FEDER, under the Partnership Agreement PT2020.

✉ Maria de Lourdes Centeno
lcenteno@iseg.ulisboa.pt

¹ Fidelidade Companhia de Seguros, Lisbon, Portugal

² CEMAPRE, ISEG, Universidade de Lisboa, Lisbon, Portugal

1 Introduction

This work is the result of a master curricular internship, which took place in a Portuguese insurance company. Its goal is to implement a new *a posteriori* tariff system for fleets of vehicles belonging to business clients.

Currently, the tariff for fleets is applied independently to each vehicle, although the fleet is typically aggregated in a single policy. This means that, in case of a claim, the *a posteriori* tariff system (experience rating based on the number of claims) will potentially increase only the part of the premium respecting to the vehicle at hand, leaving the rest of the premium untouched. It may be applied a “gross discount” depending on the dimension of the fleet and, perhaps, the line of business of the client.

The insurer intends to substitute this tariff. The purely commercial aggregation of policies would give place to a new tariff, under which the claims history of any vehicle would affect the fleet’s premium as a whole, i.e., an effective and mathematically justified aggregation of these risks in a single policy. With this idea in mind, we followed the methodology proposed by [2], which uses credibility theory in order to calculate theoretical experience rating coefficients based on the observed claims history. For a parametric approach see [1]. We propose different estimators for the structure parameters of the model. In Sect. 2 we provide a reason for such modification on the estimators.

The software used in data handling and statistical calibration of the estimation of the *a priori* tariff was the “statistical analysis system” (SAS). The software “R” was used in the application of the two models proposed by [2].

2 A Poisson model with random effects

Let us consider a portfolio of vehicles without stratification. Let N_i be the number of claims reported by vehicle i during the exposure period t_i (in years). Given $U_i = u_i$, we consider that N_i follows a Poisson distribution with mean $t_i\mu_i u_i$, where μ_i is the individual hazard rate, estimated *a priori* (using a regression model, for instance). Let u_i be the outcome of a random variable U_i and assume that $\{U_i\}$ are i.i.d. with $E(U_i) = 1$ and $V(U_i) = \sigma^2$. This implies that $E(N_i) = t_i\mu_i = \lambda_i$ and that $V(N_i) = \lambda_i + \lambda_i^2\sigma^2$. Then the usual semiparametric estimator of the variance σ^2 ($= (V(N_i) - \lambda_i)/\lambda_i^2$) is equal to

$$\widehat{\sigma^2} = \frac{\sum_i \left[(N_i - \hat{\lambda}_i)^2 - N_i \right]}{\sum_i \hat{\lambda}_i^2} = \frac{\sum_i \hat{\lambda}_i^2 \widehat{\sigma_i^2}}{\sum_i \hat{\lambda}_i^2}, \quad (1)$$

where $\hat{\lambda}_i$ is the priori estimates of λ_i and

$$\widehat{\sigma}_i^2 = \frac{(N_i - \hat{\lambda}_i)^2 - N_i}{\hat{\lambda}_i^2}. \tag{2}$$

If the time exposures have a wide variation in the portfolio, there is no reason, in our opinion, not to question the weights given to each observation in the numerator and the denominator of (1).

Assuming that the sample is large enough (which is our case) we can suppose that the only source of randomness in $\widehat{\sigma}_i^2$ is the N_i , so we replace $\hat{\lambda}_i$ by its limit $\lambda_i = E(N_i)$ in the a priori rating model. In that case any estimator of the form

$$\widetilde{\sigma}^2 = \frac{\sum_i \gamma_i \lambda_i^2 \widetilde{\sigma}_i^2}{\sum_i \gamma_i \lambda_i^2}, \tag{3}$$

with

$$\widetilde{\sigma}_i^2 = \frac{(N_i - \lambda_i)^2 - N_i}{\lambda_i^2} \tag{4}$$

will be unbiased for σ^2 , where $\gamma_i > 0$ is a positive weight given to observation i , because $E(\widetilde{\sigma}_i^2) = \sigma^2$. The choice of the γ_i 's that minimizes the variance of the estimator would be ideal, but also difficult to achieve in general. However we may note that if we minimize $V(\widetilde{\sigma}_i^2)$ w.r.t. the γ_i , we obtain

$$\text{Min}_{\gamma_i} V(\widetilde{\sigma}^2) = \text{Min}_{\gamma_i} \frac{\sum_i \gamma_i^2 \lambda_i^4 V(\widetilde{\sigma}_i^2)}{(\sum_i \gamma_i \lambda_i^2)^2}. \tag{5}$$

As multiplying $V(\widetilde{\sigma}^2)$ by a positive constant does not affect the minimizer we may constraint $\sum_i \gamma_i \lambda_i^2$ to be 1. Let $2v$ be the Lagrange multiplier associated to the constraint $\sum_i \gamma_i \lambda_i^2 = 1$. As the first order conditions are of the type

$$\gamma_i \lambda_i^2 V(\widetilde{\sigma}_i^2) = v, \tag{6}$$

the estimator given by (1), which has weights $\gamma_i = 1$, is related to the specification of the individual variance $V(\widetilde{\sigma}_i^2) = \frac{v}{\lambda_i^2}$. But we can think of other specifications for $V(\widetilde{\sigma}_i^2)$, that with other weights would satisfy (6). This is the case of $V(\widetilde{\sigma}_i^2) = \frac{v t_i}{\lambda_i^2}$ and $\gamma_i = \frac{1}{t_i}$. As $\lambda_i = t_i \mu_i$, the first specification means that the elasticity of $V(\widetilde{\sigma}_i^2)$ w.r.t. to the duration t_i and to the individual hazard function μ_i is equal to -2 , while for the second specification these values are equal to -1 and to -2 , respectively. As we will see in Sect. 5.3, the second specification provides more adequate results when time exposure varies greatly between vehicles, which is our case. Of course that the variance of $\widetilde{\sigma}_i^2$ depends on moments of the non-observable random variable

U . A frequent assumption in actuarial science is that the structure random variable U follows a Gamma distribution (in this case with mean equal to 1). Hence we calculate $V(\widetilde{\sigma}_i^2)$ as if the moments of U are equal to the moments of a Gamma with mean 1 and variance $\theta (= \sigma^2)$, i.e. when

$$E(U_i^3) = (1 + \theta)(1 + 2\theta) \tag{7}$$

and

$$E(U_i^4) = (1 + \theta)(1 + 2\theta)(1 + 3\theta). \tag{8}$$

After some tedious calculations we obtain

$$V(\widetilde{\sigma}_i^2) = 2\theta^2(1 + 3\theta) + \frac{4\theta(1 + 2\theta)}{\mu_i t_i} + \frac{2(1 + \theta)}{\mu_i^2 t_i^2}. \tag{9}$$

The elasticities of (9) with respect to the time exposure t_i or the hazard rate μ_i are both

$$e = -\frac{2\theta + 4\theta^2 \mu_i t_i + 2\theta \mu_i t_i + 2}{3\theta^3 \mu_i^2 t_i^2 + \theta^2 \mu_i^2 t_i^2 + 4\theta^2 \mu_i t_i + 2\theta \mu_i t_i + \theta + 1}, \tag{10}$$

which are increasing functions varying from -2 (when t_i or μ_i go to zero) to 0 (when t_i or μ_i go to infinity). Both specifications for $V(\widetilde{\sigma}_i^2)$ imply an elasticity with respect to the hazard rate of -2 , which seems reasonable because in practice the values of μ_i are small. Figures 1 and 2 show the elasticity of $V(\widetilde{\sigma}_i^2)$ with respect to t_i for values of θ equal to 1 and 2 and values of μ_i equal to 10 and 8 % respectively.

We can conclude that the choice of the specification to use depends on the portfolio that we have. If there are many old policies in the portfolio the second specification seems preferable. The choice depends also on the heterogeneity of the portfolio that is not explained by the *a priori* model and on the claim frequency.

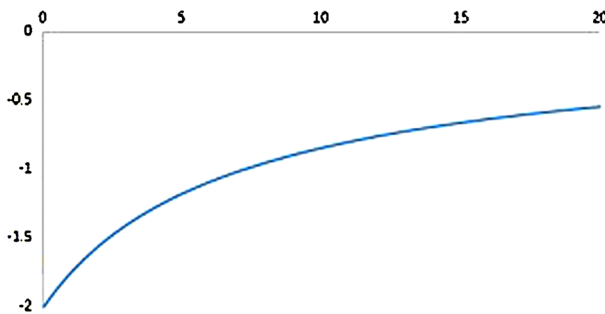


Fig. 1 Elasticity of the of the individual variance with respect to t_i , for $\theta = 1$ and $\mu_i = 10\%$

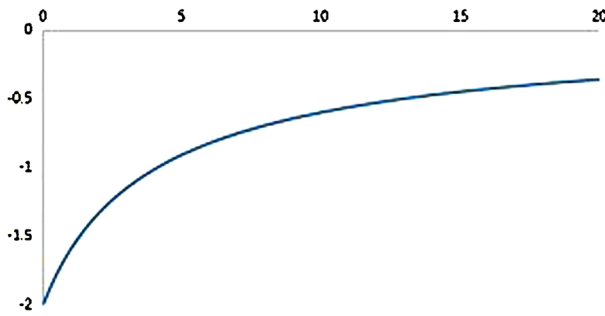


Fig. 2 Elasticity of the of the individual variance with respect to t_i , for $\theta = 2$ and $\mu_i = 8\%$

3 A Poisson model in a stratified portfolio

Portfolios of fleets of vehicles are a classical example of stratified portfolios. In the risk evaluation we take into account the individual characteristics of the vehicles and also the characteristics at fleet level. For example, the practices of a company regarding safety rules will influence the risk of its fleet. Hence, the effects introduced to design an optimal experience rating system must have a hierarchical structure (see [5]).

The system currently used in this insurance company Fidelidade is limited, in the sense that the history of a vehicle is not used to predict risk levels of other vehicles in the fleet. Now we briefly describe a model by [2], in which the hierarchical nature of the portfolio is taken into account by a double indexation.

Let N_{fi} be the number of claims reported by vehicle i of fleet f during period t_{fi} (in years). We consider that, given $U_{fi} = u_{fi}$, N_{fi} follows a Poisson distribution with mean $t_{fi}\mu_{fi}u_{fi}$. Let

$$X_{fi} \stackrel{\text{def}}{=} \frac{N_{fi}}{t_{fi}}, \quad f = 1, \dots, F; \quad i = 1, \dots, m_f. \tag{11}$$

The random variables X_{fi} represent the average number of claims per year and given $U_{fi} = u_{fi}$ are supposed to be independent (as well as N_{fi}). The *a priori* frequency risk of vehicle i of fleet f is $t_{fi}\mu_{fi} \stackrel{\text{def}}{=} \lambda_{fi}$ (whereas the *a priori* frequency risk per year is μ_{fi}). It is a function of the rating factors observed at fleet and at vehicle level and constitutes the regression component of the model. The random effect U is the heterogeneity component of the model. We distinguish firm-specific and vehicle-specific effects in both the regression and the heterogeneity components:

$$\lambda_{fi} = t_{fi} \exp\left(\underline{y}_f \widehat{\underline{\eta}} + \underline{z}_{fi} \widehat{\underline{\delta}}\right), \quad U_{fi} = R_f S_{fi} \tag{12}$$

where the vectors $\widehat{\underline{\eta}}$ and $\widehat{\underline{\delta}}$ are the maximum likelihood estimators. The *a priori* rating model is a Poisson model with neither fixed nor random effects.

As for the residual heterogeneity component U_{fi} , it represents the factors which are not observable or hard to quantify. It splits into a fleet-specific effect R_f and a

vehicle-specific effect S_{fi} . The random factors $\{R_f\}_{f=1,\dots,F}$ and $\{S_{fi}\}_{f=1,\dots,F; i=1,\dots,m_f}$ are families of i.i.d. random variables which are also mutually independent. If R and S are random variables with these distributions, we suppose that $E(R) = E(S) = 1$ and define $\text{Var}(R) = V_{RR}$; $\text{Var}(S) = V_{SS}$; $\text{Var}(U) = V_{UU}$. The distributions of the random effects will only be specified by the variances, because from $U = RS$, we get

$$E(U) = E(RS) = E(R)E(S) = 1. \tag{13}$$

Equation (13) reflects the natural hypothesis that the *a priori* rating captures the mean of the risk, i.e. $t_{fi}\mu_{fi} = E(N_{fi})$. Also, straightforward computations yield

$$V_{UU} = \frac{V(N_{fi}) - \lambda_{fi}}{\lambda_{fi}^2}, \tag{14}$$

$$V_{RR} = \frac{\text{Cov}(N_{fi}, N_{fj})}{\lambda_{fi}\lambda_{fj}}, \quad i \neq j \tag{15}$$

$$V_{SS} = \frac{V_{UU} - V_{RR}}{1 + V_{RR}}. \tag{16}$$

Given (14), (15) and (16), [2] proposed the following estimators for V_{RR} , V_{SS} and V_{UU} :

$$\hat{V}_{RR} = \frac{\sum_f \sum_{1 \leq i \neq j \leq m_f} (N_{fi} - \hat{\lambda}_{fi})(N_{fj} - \hat{\lambda}_{fj})}{\sum_f \sum_{1 \leq i \neq j \leq m_f} \hat{\lambda}_{fi}\hat{\lambda}_{fj}}, \tag{17}$$

$$\hat{V}_{UU} = \frac{\sum_{f,i} [(N_{fi} - \hat{\lambda}_{fi})^2 - N_{fi}]}{\sum_{f,i} \hat{\lambda}_{fi}^2} \tag{18}$$

$$\hat{V}_{SS} = \frac{\hat{V}_{UU} - \hat{V}_{RR}}{1 + \hat{V}_{RR}}. \tag{19}$$

These estimators can be written in terms of the X_{fi} 's as

$$\hat{V}_{RR} = \frac{\sum_f \sum_{1 \leq i \neq j \leq m_f} t_{fi}t_{fj} (X_{fi} - \hat{\mu}_{fi})(X_{fj} - \hat{\mu}_{fj})}{\sum_f \sum_{1 \leq i \neq j \leq m_f} t_{fi}t_{fj} \hat{\mu}_{fi}\hat{\mu}_{fj}}, \tag{20}$$

and

$$\hat{V}_{UU} = \frac{\sum_{f,i} [t_{fi}^2 (X_{fi} - \hat{\mu}_{fi})^2 - t_{fi} X_{fi}]}{\sum_{f,i} t_{fi}^2 \hat{\mu}_{fi}^2} \tag{21}$$

If \hat{V}_{RR} is greater than zero, then the history of a vehicle may reveal hidden features in the risk distribution of the other vehicles in the same fleet. We remark that, if a certain sample would generate $\hat{V}_{RR} < 0$, that would translate into a null estimator for V_{RR} . In that case, the fleet-specific effect brings no additional information and should be abandoned.

There is another formula for \hat{V}_{RR} which is better for computational purposes. If we define

$$N_f = \sum_{i=1}^{m_f} N_{fi}; \quad \hat{\lambda}_f = \sum_{i=1}^{m_f} \hat{\lambda}_{fi}, \tag{22}$$

it is possible to deduct that

$$\hat{V}_{RR} = \frac{\sum_f (N_f - \hat{\lambda}_f)^2 - \sum_{f,i} (N_{fi} - \hat{\lambda}_{fi})^2}{\sum_f \hat{\lambda}_f^2 - \sum_{f,i} \hat{\lambda}_{fi}^2}. \tag{23}$$

Although the estimators proposed by [2] are consistent, it is not obvious, as explained in Sect. 2, the weight that should be given to the claim frequency of each vehicle. When the exposures vary from a few days to several years, as in our portfolio, a different choice of weights could lead to different conclusions.

We propose to estimate those parameters using the second specification referred in Sect. 2, which lead in the stratified model to:

$$\hat{V}_{RR}^{mod} = \frac{\sum_f \sum_{1 \leq i \neq j \leq m_f} (\sqrt{t_{fi} t_{fj}})^{-1} (N_{fi} - \hat{\lambda}_{fi}) (N_{fj} - \hat{\lambda}_{fj})}{\sum_f \sum_{1 \leq i \neq j \leq m_f} (\sqrt{t_{fi} t_{fj}})^{-1} \hat{\lambda}_{fi} \hat{\lambda}_{fj}} \tag{24}$$

$$= \frac{\sum_f \sum_{1 \leq i \neq j \leq m_f} \sqrt{t_{fi} t_{fj}} (X_{fi} - \hat{\mu}_{fi}) (X_{fj} - \hat{\mu}_{fj})}{\sum_f \sum_{1 \leq i \neq j \leq m_f} \sqrt{t_{fi} t_{fj}} \hat{\mu}_{fi} \hat{\mu}_{fj}}. \tag{25}$$

$$\hat{V}_{UU}^{mod} = \frac{\sum_{f,i} [t_{fi}^{-1} (N_{fi} - \hat{\lambda}_{fi})^2 - t_{fi}^{-1} N_{fi}]}{\sum_{f,i} t_{fi}^{-1} \hat{\lambda}_{fi}^2} = \frac{\sum_{f,i} [t_{fi} (X_{fi} - \hat{\mu}_{fi})^2 - X_{fi}]}{\sum_{f,i} t_{fi} \hat{\mu}_{fi}^2}, \tag{26}$$

If we define

$$N_f^{\text{mod}} = \sum_{i=1}^{m_f} (\sqrt{t_{fi}})^{-1} N_{fi}; \quad \hat{\lambda}_f^{\text{mod}} = \sum_{i=1}^{m_f} (\sqrt{t_{fi}})^{-1} \hat{\lambda}_{fi}, \tag{27}$$

it is possible to deduct

$$\hat{V}_{RR}^{\text{mod}} = \frac{\sum_f (N_f^{\text{mod}} - \hat{\lambda}_f^{\text{mod}})^2 - \sum_{f,i} t_{fi}^{-1} (N_{fi} - \hat{\lambda}_{fi})^2}{\sum_f (\hat{\lambda}_f^{\text{mod}})^2 - \sum_{f,i} t_{fi}^{-1} \hat{\lambda}_{fi}^2}. \tag{28}$$

Note that if we substitute the $\hat{\lambda}_{fi}$'s by λ_{fi} 's on (24) and on (26), it is straightforward to show that the resulting estimators are unbiased, given (14) and (15). The introduction of weights different from 1, do not affect its expectation once the weight given is the same in the numerator and denominator of the estimators.

For further details, see [3]. In Sect. 5.3 we apply these and the previous estimators to our data and comment on the results.

4 Experience rating using credibility

In this Section we summarize two experience rating schemes from the model described above. Let i_0 be a vehicle belonging to fleet f , which has m_f vehicles at start. After a period of observation, an *experience rating* coefficient is computed for the next one. In order to allow for a turnover in the portfolio, vehicle i_0 may appear in the second period or not. For both systems, the linear predictors are obtained separately for each fleet, so we may drop the fleet index.

4.1 A system using the claims history at fleet level

In this system, we calculate linear credibility predictors based on the claims history at fleet level.

The experience rating coefficient for vehicle i_0 , ER_{i_0} , is given by the formulas below. For further details see [2] or [3]. We have

$$ER_{i_0} = (1 - cred_{i_0}) + cred_{i_0} \frac{\sum_i n_i}{\sum_i \hat{\lambda}_i} \tag{29}$$

where $cred_{i_0}$ is the credibility weight given to vehicle i_0 (bearing a fleet-specific and a vehicle-specific component). We must consider two situations:

1. The vehicle was not in the fleet when the first period started, i.e., it was not observed: $i_0 \neq i, \forall i \in \{1, \dots, m\}$. Then

$$cred_{i_0} = \alpha = \frac{\hat{V}_{RR} \sum_i \hat{\lambda}_i}{1 + \hat{V}_{RR} \left(\sum_i \hat{\lambda}_i \right) + (\hat{V}_{UU} - \hat{V}_{RR}) \frac{\sum_i \hat{\lambda}_i^2}{\sum_i \hat{\lambda}_i}} \tag{30}$$

2. The vehicle was in the fleet during the first period and then the credibility coefficient may be regarded as the sum of a component α , related to the fleet claim history, and a component β_{i_0} , related to the vehicle claim history:

$$cred_{i_0} = \alpha + \beta_{i_0}; \quad \beta_{i_0} = \frac{(\hat{V}_{UU} - \hat{V}_{RR}) \hat{\lambda}_{i_0}}{1 + \hat{V}_{RR} \left(\sum_i \hat{\lambda}_i \right) + (\hat{V}_{UU} - \hat{V}_{RR}) \frac{\sum_i \hat{\lambda}_i^2}{\sum_i \hat{\lambda}_i}} \tag{31}$$

These experience rating coefficients may be computed only if the estimated vehicle-specific variance \hat{V}_{SS} is greater than zero [or if $\hat{V}_{UU} > \hat{V}_{RR}$, from (19)]. Our data satisfies this condition (see 5.3). Also, we notice that this model generates experience rating coefficients which do not vary much inside fleets, because the credibility granted to the claims history of a vehicle is applied to a ratio computed at fleet level.

We consider an adaptation in order to accommodate the turnover (i.e., the proportion of new vehicles in the fleet). If ρ is the expected turnover, then we take $cred_{i_0} = \alpha + (1 - \rho)\beta$, where β is the average of all vehicle-specific components β_i .

4.2 A system which uses full information on claims history

At first glance, computing premiums at vehicle level may seem of little importance, since the firm will pay them jointly for the fleet. Under that point of view, a system like the one in Sect. 4.1 would be perfectly adequate. However, the information on claims at vehicle level may be important. For example, if there is an increase of the overall fleet premium, it may be of interest to know which vehicles are “responsible” in order to take specific measures.

The idea behind the following system is that a vehicle with claims should have a greater premium than the one prescribed by the previous system. The opposite should happen to vehicles without claims. The notations are the same as above.

We will express the experience rating coefficient for vehicle i_0 , $ER2_{i_0}$, as a function of the estimates \hat{V}_{UU} , \hat{V}_{RR} and the predicted claim frequencies for each vehicle in the fleet, $\hat{\lambda}_i$, ($i = 1, \dots, n$). It is presented as a sum of two terms:

1. The first, including the α_i 's and not depending on the vehicles within the fleet.
2. The second, β_{i_0} , related to past observation of the vehicle.

We have

$$ER2_{i_0} = 1 + \sum_{i=1}^m \alpha_i \left(\frac{n_i}{\hat{\lambda}_i} - 1 \right) + \beta_{i_0} \left(\frac{n_{i_0}}{\hat{\lambda}_{i_0}} - 1 \right).$$

Unlike the previous model, now the credibility coefficient β_{i_0} is applied to the individual claims history of the vehicle. For all vehicles in the fleet ($1 \leq i \leq m$), we have

$$\alpha_i = \frac{\frac{\hat{\lambda}_i \hat{V}_{RR}}{1 + \hat{\lambda}_i (\hat{V}_{UU} - \hat{V}_{RR})}}{1 + \sum_j \frac{\hat{\lambda}_j \hat{V}_{RR}}{1 + \hat{\lambda}_j (\hat{V}_{UU} - \hat{V}_{RR})}}. \tag{32}$$

For vehicles already observed in the fleet ($1 \leq i_0 \leq m$), we have

$$\beta_{i_0} = \frac{\frac{\hat{\lambda}_{i_0} (\hat{V}_{UU} - \hat{V}_{RR})}{1 + \hat{\lambda}_{i_0} (\hat{V}_{UU} - \hat{V}_{RR})}}{1 + \sum_j \frac{\hat{\lambda}_j \hat{V}_{RR}}{1 + \hat{\lambda}_j (\hat{V}_{UU} - \hat{V}_{RR})}}.$$

For new vehicles, we have $\beta_{i_0} = 0$. As in 4.1, this credibility system makes sense only if $\hat{V}_{UU} > \hat{V}_{RR}$, i.e., if the estimated variance \hat{V}_{SS} of the vehicle-specific effect is positive (see Eq. 19).

5 Fleets insured in Fidelidade

This section is divided in three subsections. In the first, we briefly describe the portfolio of motor insurance of business clients in the insurance company. In the second, we sketch the estimation of the expected number of claims per vehicle through GLMs—see [6]. Further discussion of GLM’s within an actuarial context may be found, for example, in [4]. Finally, in the last subsection we present the results of applying the two experience rating schemes of [2] to the portfolio of the insurer.

5.1 Description of the portfolio

First, we observe that IBNR claims are unlikely, since this study was carried out in 2015 with data spanning from 2007 to 2013. In this work, we only considered the mandatory coverage by law in Portugal, which is a third party liability insurance (although many of the vehicles were also covered for a broader range of risks).

Some of the data was ruled out due to missing or faulty information. In addition, we excluded some vehicles based on their age and legal category, leaving a total of 871, 881 vehicles, 102, 132 of whom are “single-vehicle fleets”. According to this broad definition, we have 182, 855 fleets.

Initially, we classified each vehicle according to six characteristics: legal category, age and fuel, at vehicle level; fleet dimension, type of economic activity

and geographical area, at fleet level. We chose not to include the capital insured because the legal requirements for compulsory liability limit currently in force in Portugal differ almost exclusively according to the legal category of the vehicle and the kind of economic activity of the firm.

With respect to the legal category of the vehicle, we identified each one with two letters, based on the legal definition currently in force in Portugal. In our study, we included categories AR and PS (trucks), AU (buses), LP (passenger cars), MT (commercial cars), CT (vans), PU (pickups) and others less significant in number (Fig. 3).

As for the age, we considered vehicles up to 40 years old when entering the portfolio. For the fuel, we considered four categories (see Fig. 4): “Diesel”, “Gasoline”, “Others” and “No info”.

The next factor taken into account was the kind of economic activity practised by the fleet owner. We based our own classification, displayed in Table 1, on the Portuguese economic classification code by letters (for more details see [3]).

As for the fleet dimension, we grouped the vehicles into fleets according to the client number of the owner, except in the case where the owner’s activity was related to leasing or long duration rental of vehicles. The results are in Fig. 5 (in percentage over the total number of vehicles).

The geographical area factor considered in this study has to do with the location of the company headquarters. The insurer developed an aggregation of the Portuguese parishes in 12 geographical areas. Each of these areas is geographically contiguous with the next and is, by construction, less urbanized than the next. Hence, in the portfolio of family clients, it is not surprising that the frequency of claims increases steadily throughout the scale, from the “most” rural area to the “most” urban area. That is not the case for fleets, essentially for two reasons: First, it might happen that not all of the vehicles were allocated to the headquarters of the company. Second, a significant part of the fleet might also circulate outside the company’s region frequently, or even daily.

Nevertheless, we considered that there was a fair chance of obtaining a coherent variable, after appropriate redesigning. We aggregated the original geographical

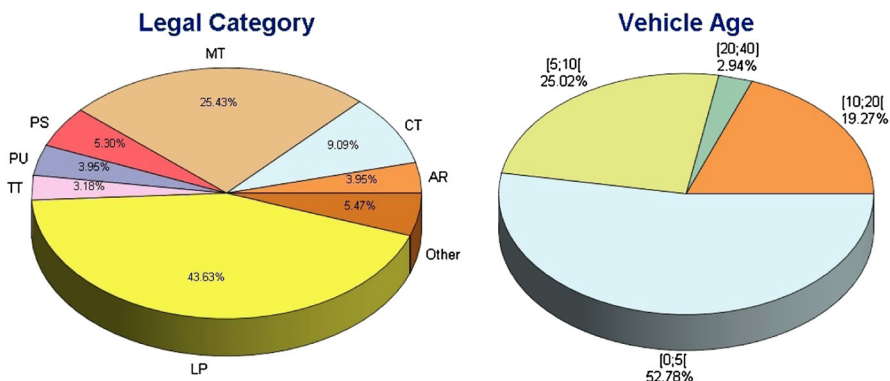
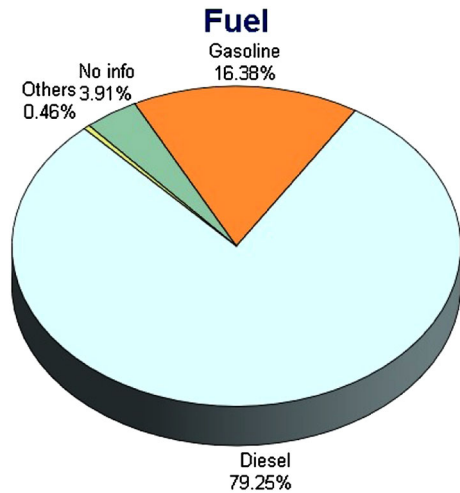


Fig. 3 Distribution of the vehicles by legal category and age

Fig. 4 Distribution of the vehicles by fuel



areas into 4 wider regions, labelled “I” , “II” , “III” and “IV” . In addition, we created a new level (“N”-no region) for the fleets we thought were likely to circulate throughout all regions or even internationally. We included in level N most of the fleets greater than 300 vehicles¹, as well as smaller fleets whose economic activity implied long trips (for example, long distance transportation by road).

5.2 A priori estimation of the number of claims

We used GLM in order to estimate the number of claims per vehicle, for the whole seven years (2007–2013). More specifically, we used the GENMOD procedure in SAS. We performed the estimation for the time period 2007–2013.

We recall that our variable of interest is N , the number of claims in the period of time each vehicle was exposed. The covariates are the geographical area, the economic activity and the dimension of the fleet, on the part of the company; the legal category, the type of fuel and the age, on the part of the vehicle.

We used a Poisson log-linear model, with the logarithm of the time exposure as offset. Also, we specified no weights ω_{fi} (considering the common assumption $a(\phi) = \frac{\phi}{\omega_{fi}}$). For more details see for example [6].

We point out that our approach is mathematically equivalent to replacing the claim count by the claim frequency (claims divided by time exposures $\frac{N_{fi}}{t_{fi}}$) as our variable of interest, using time exposure as the weight and ruling out the offset. For a proof of this result see [8].

After running some preliminary GLM models for the period 2007-2013, we reached some conclusions and made some choices, which are summarized below.

We chose to model the *age* of the vehicle as a continuous variable instead of a factor. The use of a linear variable, with a cap on the vehicles more than 15 years

¹ Except urban passenger transportation and public local administration.

Table 1 Distribution of the vehicles by economic activity of their firm

Code	Description of the economic activity	Vehicles
AB	Primary sector, including extractive industries	16, 865
C	Manufacturing industry	86, 388
DE	Energy, water, sanitation, residuals recycling or disposal	25, 851
F	Construction	102, 520
G	Commerce and auto-repairs	174, 180
H	Transportation and storage of goods	83, 773
IJKL	Tourism, media, finance, real estate	111, 530
MNa	Consulting, science and technology, administrative activities	48, 154
Nb	Vehicle rentals (three digit CAE 771)	145, 469
OPQ	Public administration, education, health, social services	57, 842
RSTU	Arts, sports, events, international organizations and others	19, 309

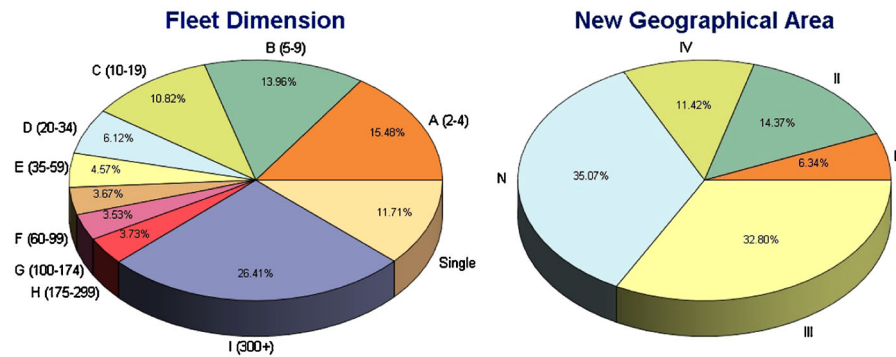


Fig. 5 Distribution of the vehicles by the size of their fleet and by the geographical area of their firm

old, revealed adequate. As for the factor *fuel*, the categories “No info” and “Others” were merged.

The factor *fleet dimension* did not seem to have much explanatory value. We decided to consider instead the *total exposure of the fleet* (measured in number of vehicles per year). We considered modelling it through a factor variable or, alternatively, a continuous variable. In the end, we chose a factor variable with only two levels, $]0; 10]$ and $]10; \infty]$, including 29.5 and 70.5 % of the vehicles, respectively.

With respect to the factors *economic activity* of the company, and *legal category* of the vehicle, some of the levels were aggregated. These two variables were those with greater explanatory power.

We also estimated GLM’s for each of the years 2007,..., 2013. Their features are similar to the ones described above. The main exception is the variable *total exposure of the fleet*, whose levels were barely significant or not significant at all, depending on the year considered. Since the gains in terms of deviance were also barely significant, we dropped the variable in order to favour the parsimony of the

models. With respect to the other factors (legal category, economic activity, fuel, geographical area), there were at times different aggregations of the different levels.

Several tendencies, coherent throughout the years, were uncovered. Many of them were already expected for being well-known in the industry, for example the high claim numbers for trucks and buses and low claim numbers for motorcycles; higher claim numbers in urban areas than in rural areas; or higher claim numbers from diesel-propelled vehicles than gasoline ones.

5.3 Empirical results

In this section, we refer to the experience rating systems described in Sects. 4.1 and 4.2 as Model 1 and Model 2, respectively. We start with a regard over the estimated variance of the random effect U . We computed estimates \hat{V}_{RR} , \hat{V}_{UU} and \hat{V}_{SS} at all years 2007, 2008, ..., 2013, as proposed by Desjardins *et al.* Also, we computed estimates \hat{V}_{RR}^{mod} , \hat{V}_{UU}^{mod} and \hat{V}_{SS}^{mod} as proposed by us (in the end of Sect. 3). The results are in Table 2.

We observe that the yearly estimates for V_{UU} can differ up to 17 % (in 2007). As for V_{RR} , the highest difference between estimates is 27 %, in 2009. In some years, however, the estimates are very close (2010 and 2011). On the contrary, if we take the period 2007–2013 as a whole, the estimates prescribed in [2] for V_{RR} differ more significantly from our own.

The smaller scale for the differences between the yearly estimates is probably due to the fact that the majority of the vehicles have exposure equal to or close to 1, in which case its contribution to both estimates is the same. In the case of the whole period 2007–2013, that does not happen. The policies that have exposure close to 7 years are a small minority (less than 5 % of the vehicles have exposure greater than 6 years). In addition, the estimates \hat{V}_{UU} and \hat{V}_{SS} for the whole period 2007–2013 are significantly different from all corresponding yearly estimates. This reflects the fact that the predictive ability of claims decreases with their seniority. This point is discussed for instance in [7]. In our case, the estimates \hat{V}_{UU}^{mod} , \hat{V}_{RR}^{mod} and \hat{V}_{SS}^{mod} for the period 2007–2013 lie between the minimum and maximum yearly estimates. These results seem to support the idea that our estimators are more adequate when time exposure varies greatly between vehicles.

Table 3 presents data respecting to Model 1 and Model 2 for 2013, organized by fleet *a priori* rating. Since we have a very heterogeneous portfolio, we consider that it is more natural to present the results organized by expected fleet claims rather than by number of vehicles. Under this point of view, the fleet of an agricultural business in a rural area, including a dozen vehicles like jeeps and pickups, does not belong together with a company owning a dozen cabs operating in Lisbon, nor with a transportation company operating at international level with the same number of large trucks. This makes perfect sense for us.

We remind that m_f is the number of vehicles of fleet f and that the credibility assigned to vehicle i_0 of fleet f is represented by

$$cred_{f_{i_0}} = \alpha_f + \beta_{f_{i_0}}$$

Table 2 Estimates according to [2] and according to our modification

Period	\hat{V}_{UU}	\hat{V}_{rr}	\hat{V}_{ss}	\hat{V}_{UU}^{mod}	\hat{V}_{rr}^{mod}	\hat{V}_{ss}^{mod}
2007	0.80310	0.16043	0.55382	0.93607	0.17640	0.64576
2008	0.83629	0.11667	0.64444	0.90361	0.11314	0.71012
2009	0.84049	0.12168	0.64084	0.89864	0.15409	0.64514
2010	0.78422	0.10218	0.61881	0.81184	0.10026	0.64673
2011	1.11964	0.13668	0.86477	1.13002	0.13587	0.87523
2012	1.0159	0.09019	0.84909	1.03414	0.09151	0.86360
2013	0.89149	0.14086	0.65796	0.93018	0.13250	0.70435
2007–2013	0.67735	0.11432	0.50527	0.83171	0.10542	0.65702

where α_f is relative to the fleet and β_{i_0} is relative to the vehicle. Also, σ_1 is the inside-fleet standard deviation of the ER1 coefficients and the same is valid for σ_2 and ER2. Each column presents averages over all fleets in each category. For data respecting to other years, see [3].

These data are compatible with the conjecture that the credibility assigned to an observed vehicle is greater if the dimension of its fleet is greater. A feature of interest regarding the inside-fleets standard deviations (in both models) is the decreasing to values close to zero of the standard deviation when the fleet size becomes very large. However, we remark that, in general, the inside-fleet standard deviation is not a monotone function of the size of the fleet. That fact is patent only for Model 2 in this example, but is true for both models—see [2].

Also, the inside-fleets standard deviation is, without surprise, always lesser under Model 1 than under Model 2. If the latter takes into account the claims experience of each vehicle in the fleet, penalizing specifically the vehicles responsible for the claims, it is natural that the resulting experience rating coefficients display wider variations inside each fleet than those of Model 1. We remark that, in Model 1, the

Table 3 Data respecting to Model 1 and Model 2 for 2013

μ_f	m_f	α_f	$bar\beta_f$	$cred_f$	σ_1	σ_2
]0; 0.25]	1.578	0.01034	0.04103	0.05138	NA	NA
]0.25; 0.5]	4.964	0.04067	0.05686	0.09753	NA	NA
]0.5; 1]	8.874	0.07687	0.06073	0.1376	0.02220	0.1350
]1; 2]	16.86	0.1416	0.06264	0.2043	0.01699	0.1511
]2; 4]	31.73	0.2488	0.06162	0.3104	0.01243	0.1663
]4; 8]	59.54	0.3918	0.05559	0.4474	0.008829	0.1434
]8; 16]	111.1	0.5605	0.03894	0.5995	0.006935	0.1103
]16; 32]	180.4	0.7043	0.03219	0.7365	0.004204	0.08378
]32; 64]	409.6	0.8279	0.01725	0.8452	0.003376	0.05099
>64	1841	0.9329	0.006509	0.9394	0.000565	0.01720

Table 4 Data for the 17th fleet of the portfolio of 2012

f	i	t_{fi}	λ_{fi}	n_{fi}	λ_f	n_f	ER1	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	ER2
17	1	1	0.0648	0	0.1995	0	0.9272	0.9401	0.9531	0.9703	0.9271
17	2	0.24	0.0191	0	0.1995	0	0.9667	0.9401	0.9531	0.9703	0.9660
17	3	0.25	0.0265	0	0.1995	0	0.9604	0.9401	0.9531	0.9703	0.9595
17	4	1	0.0891	0	0.1995	0	0.9062	0.9401	0.9531	0.9703	0.9077

variations between vehicles in the same fleet result exclusively from the variations in the *a priori* rating. We will illustrate this with a few examples of claims experiences of fleets. The labels “ $\rho = \dots$ ” represent the adaptation of Model 1 mentioned at the end of Sect. 4.1 for different values of the turnover (proportion of new cars in the fleet).

In Table 4 we have a fleet of four vehicles without any claim. Nevertheless, the experience rating coefficients for Model 1 (labelled “ER1”) vary according to the claims expectancy λ_{fi} of each vehicle. The same happens for Model 2, which generates almost the same coefficients as Model 1. When we take the turnover into account, we observe an increase of the experience rating coefficient for the fleet as the turnover increases.

In Table 5 we have a fleet of five vehicles with a claims experience far worse than expected (3 claims), given its expected claims $\sum_i \lambda_{fi} = 0.5119$. We remark that vehicles 3 and 5 have exactly the same expected claims and hence Model 1 assigns exactly the same penalty to both. On the contrary, Model 2 increases heavily the penalty of vehicle 3 because of its two claims in 2012, while reducing the penalty of vehicle 5, which had no claims. Also, we notice that the lowest penalty by Model 2 was assigned to vehicle 5, the one among those with no claims which had the highest expected claims. When we take the turnover into account, we observe a decrease of the experience rating coefficient for the fleet as the turnover increases.

In Table 6, we have a fleet of 13 vehicles with only one observed claim, which is a lower value than the expected claims of the fleet, 1.296. Hence, Model 1 assigns bonuses to all vehicles, with low variations between them. Model 2 assigns higher bonuses to all vehicles but the one with a claim, which has a severe penalty.

For simplicity in the result presentation, we aggregated the fleets in 10 classes, according to their *a priori* rating. For three of these classes of *a priori* rating, we present the averages of the experience rating coefficients for Model 1 in Tables 7, 8 and 9, considering different values for the turnover ρ of the fleet. We remind that n_f represents the number of claims of fleet f .

We point out that the coefficients in Tables 7, 8 and 9 are averages over all fleets of the adaptation of Model 1, described in the end of Sect. 4.1. We remind that this adaptation assigns a single experience rating coefficient to the whole fleet, depending on its turnover.

Finally, a word on the total risk premium of the portfolio. In 2013, the predicted total number of claims is 15, 778. After applying the models and assuming a “frozen” portfolio, it turns to 15, 853.6 (Model 1) or to 15, 848.4 (Model 2), a deviation of less than 0.5 % in both cases. We may say that both models are fairly

Table 5 Data for the 174th fleet of the portfolio of 2012

f	i	t_{fi}	λ_{fi}	n_{fi}	$\sum_i \lambda_{fi}$	n_f	ER1	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	ER2
174	1	1	0.1113	0	0.5119	3	1.6385	1.6030	1.4813	1.3190	1.1059
174	2	1	0.0682	0	0.5119	3	1.4676	1.6030	1.4813	1.3190	1.1388
174	3	1	0.1174	2	0.5119	3	1.6624	1.6030	1.4813	1.3190	2.7162
174	4	1	0.0976	1	0.5119	3	1.5840	1.6030	1.4813	1.3190	1.9371
174	5	1	0.1174	0	0.5119	3	1.6624	1.6030	1.4813	1.3190	1.1015

Table 6 Data for the 1115th fleet of the portfolio of 2012

f	i	t_{fi}	λ_{fi}	n_{fi}	$\sum_i \lambda_{fi}$	n_f	ER1	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	ER2
1115	1	1	0.0929	0	1.296	1	0.9617	0.961	0.966	0.973	0.906
1115	2	1	0.171	0	1.296	1	0.9481	0.961	0.966	0.973	0.854
1115	3	1	0.0772	0	1.296	1	0.9645	0.961	0.966	0.973	0.918
1115	4	1	0.0763	0	1.296	1	0.9646	0.961	0.966	0.973	0.918
1115	5	1	0.0954	0	1.296	1	0.9613	0.961	0.966	0.973	0.905
1115	6	1	0.0979	0	1.296	1	0.9608	0.961	0.966	0.973	0.903
1115	7	1	0.0979	0	1.296	1	0.9608	0.961	0.966	0.973	0.902
1115	8	1	0.101	1	1.296	1	0.9604	0.961	0.966	0.973	1.672
1115	9	1	0.101	0	1.296	1	0.9604	0.961	0.966	0.973	0.901
1115	10	1	0.148	0	1.296	1	0.9523	0.961	0.966	0.973	0.869
1115	11	1	0.101	0	1.296	1	0.9604	0.961	0.966	0.973	0.901
1115	12	1	0.101	0	1.296	1	0.9604	0.961	0.966	0.973	0.901
1115	13	0.4	0.0381	0	1.296	1	0.9713	0.961	0.966	0.973	0.947

neutral in this case, since they approximately redistribute the individual premia between the different policies. In all other years considered, the deviation does not exceed 1 %.

6 Conclusions

We applied two experience rating schemes, “Model 1” and “Model 2”, developed by Desjardins et al. (2001), see [2], to the portfolio of a Portuguese insurance company at each of the years 2007–2013. Only the mandatory coverage (third-party liability insurance) was taken into account.

The main strength of Model 1 lies in its adaptation described at the end of Sect. 4.1, which assigns a single experience rating coefficient to each fleet as a whole and, in addition, takes into account its turnover (i.e., the percentage of new vehicles in the fleet). It allows building a simple *a posteriori* rating system, adequate for small and medium-sized fleets. It may be the basis for a bonus-malus system for fleets up to a certain dimension, expressed for example in yearly amount of *premia*.

Table 7 Average experience rating coefficients for fleets with $\lambda_f \leq 0.125$ in 2013

n_f	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 1$
0	0.95517	0.96642	0.98143	0.99269
1	1.7113	1.5329	1.2950	1.1166
2	2.4079	2.0579	1.5911	1.2411
3	2.9862	2.5002	1.8521	1.3661
4	4.2786	3.4351	2.3105	1.4670

Table 8 Average experience rating coefficients for fleets with $1 \leq \lambda_f \leq 1.5$ in 2013

n_f	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 1$
0	0.80957	0.82859	0.85395	0.87298
1	0.96903	0.97203	0.97603	0.97903
2	1.1279	1.1152	1.0983	1.0856
3	1.2793	1.2517	1.2149	1.1873
4	1.4345	1.3927	1.3370	1.2952
> 4	1.6304	1.5901	1.5333	1.4935
Max (7)	1.7225	1.7021	1.6749	1.6545

Table 9 Average experience rating coefficients for fleets with $4 \leq \lambda_f \leq 8$ in 2013

n_f	$\rho = 0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 1$
0	0.64567	0.66324	0.68667	0.70424
1	0.75135	0.76474	0.78259	0.79598
2	0.85698	0.86440	0.87429	0.88171
3	0.95727	0.95962	0.96276	0.96511
4	1.0606	1.0577	1.0538	1.0509
5	1.1660	1.1559	1.1424	1.1323
6	1.2612	1.2473	1.2288	1.2149
7	1.3462	1.3295	1.3073	1.2907
8	1.4307	1.4119	1.3869	1.3681
>8	1.8942	1.8461	1.7818	1.7355
Max (16)	2.3207	2.2202	2.0863	1.9859

The main strength of Model 2 lies in the inherent variability of the rating coefficients inside each fleet. The main features of the variability of the coefficients are patent in the examples of Sect. 5.3 and have great practical interest, namely for fleet management purposes. For fleets that are large enough to justify case-by-case negotiation of the *premia*, Model 2 shows potential to be the basis for a simulator that will provide support in these negotiations.

Since both models rely entirely on the estimators for the variances of the fleet-specific random effect and of the vehicle-specific random effect, we derived slightly different estimators from those in [2]. Our yearly estimates seemed more coherent when compared to the 7-year estimates, so we used our estimators in the computation of the experience rating coefficients.

Acknowledgments We are thankful to the Portuguese insurance company “Fidelidade” for providing the data used in this article. We are indebted to an anonymous reviewer of an earlier version of the paper for providing insightful comments and providing directions for additional work which has resulted in this paper. Without the anonymous reviewers supportive work this paper would not have been possible.

References

1. Angers JF, Desjardins D, Dionne G, Guertin F (2006) Vehicle and fleet random effects in a model of insurance rating for fleets of vehicles. *ASTIN Bull* 36(1):25–77
2. Desjardins D, Dionne G, Pinquet J (2001) Experience rating schemes for fleets of vehicles. *ASTIN Bull* 31(1):81–105
3. Fardilha T (2015) Tariff systems for fleets of vehicles: a study on the portfolio of Fidelidade. ISEG-Lisbon School of Economics and Management, Master thesis, <https://www.iseg.ulisboa.pt/aquila/getFile.do?fileId=644862&method=getFile>
4. Frees EW (2010) Regression modeling with actuarial and financial applications. Cambridge University Press, Cambridge
5. Jewell WS (1975) The use of collateral data in credibility theory: a hierarchical model. *Giornale dell’Istituto degli Attuari* 38:1–16
6. McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall, Boca Raton
7. Pinquet J, Montserrat G, Bolancé C (2001) Allowance for the age of claims in Bonus-Malus systems. *ASTIN Bull* 31(2):337–348
8. Yan J, Guszcza J, Flynn M, Wu P (2009) Applications of the offset in property-casualty predictive modelling. *Casualty Actuarial Society E-Forum*, Winter 366:385