Lisboa School of Economics & Management

Master in
Actuarial Science

Masters Final Work
Internship Report

Tariff Systems for Fleets of Vehicles: A Study on the Portfolio of Fidelidade

Tiago Marques Fardilha

October - 2015

LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

Master in Actuarial Science

Masters Final Work

Internship Report

Tariff Systems for Fleets of Vehicles: A Study on the Portfolio of Fidelidade

Tiago Marques Fardilha

Supervisors:

Prof. Maria de Lourdes Caraças Centeno

Director Rui Alexandre Silva Esteves

October - 2015

Acknowledgments

I would like to thank my parents for their support and encouragement, which have been a constant throughout my life. I would also like to thank Prof. Lourdes Centeno for precious guidance in the scientific part of the internship that made this work possible; to my supervisor in Fidelidade, Director Rui Esteves, for his remarks, always as helpful as accurate; to Febio Levezinho, who has enlightened me about many practical aspects of auto insurance in Portugal; to Catarina Jácomo, always available for my beginner doubts about SAS and about the databases of Fidelidade; to my internship colleague João Senra, for all our conversations about credibility theory and regression models. Finally, I would like to thank my wife and son. My decision of working and studying simultaneously for the last two years has naturally taken its toll on our family time and I thank them both for their love and endless understanding.

Abstract

In Portugal, insurance policies for fleets of vehicles are, in general, similar to policies for individual vehicles. In most cases, only an adjustment related to the standard industrial classification of the fleet owner and a discount related to the fleet size are applied. Such is the case of the insurance company "Fidelidade". 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. The insurer intends to substitute this tariff for a new one, under which the claim history of any vehicle would affect the fleet's premium as a whole.

We considered the credibility models for claim counts proposed by Desjardins, Dionne and Pinquet in 2001. The first model is based on the claims history at fleet level and the other is based also on the claims history at vehicle level. From the first model, it is easy to derive a system which assigns experience rating coefficients at fleet level. The latter assigns different experience rating coefficients to the different vehicles on the fleet.

We applied both models in order to calculate experience rating coefficients for the vehicles in the portfolio of fleets insured by Fidelidade, based on their observed claims history.

Keywords: Motor insurance, fleets, experience rating, *bonus-malus*, credibility.

Resumo

Em Portugal, as apólices de seguro automóvel para frotas são geralmente similares às apólices individuais. Na maioria dos casos, há apenas um ajustamento relativo ao ramo de actividade do utilizador da frota e um desconto em função do tamanho da mesma. É esse o caso da seguradora Fidelidade. O sistema de *bonus-malus* é praticamente o mesmo que nas apólices individuais e é aplicado a cada veículo separadamente, não tendo assim qualquer efeito no prémio relativo aos outros veículos da frota.

Este sistema é ineficiente, pois ignora os potenciais riscos específicos da frota na tarifação *a posteriori*. A seguradora pretende substituir esta tarifa por outra, em que o histórico de sinistralidade de cada veículo afectasse o prémio da frota como um todo.

Considerámos os modelos de credibilidade para o número de sinistros propostos por Desjardins, Dionne e Pinquet em 2001. O primeiro baseia-se no comportamento de sinistralidade ao nível da frota e o outro baseia-se também no comportamento de sinistralidade ao nível do veículo. A partir do primeiro modelo, derivam-se facilmente coeficientes de *bonus-malus* ao nível da frota. O segundo modelo atribui diferentes coeficientes de *bonus-malus* aos diversos veículos de cada frota.

Aplicámos ambos os modelos para atribuir coeficientes de *bonus-malus* aos veículos do portfolio de seguros de frotas da Fidelidade, a partir do seu histórico de sinistralidade.

Palavras-chave: Seguro automóvel, frotas, *bonus-malus*, credibilidade.

Contents

The truth lies somewhere between the two extremes.

Hans Bühlmann

Introduction

This work is the result of a curricular internship included in the Master of Actuarial Science of ISEG, which took place in the insurance company "Fidelidade", from February to June 2015. 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. In practice, this kind of policy for fleets is treated as a set of separate policies. Eventually it is applied a "gross discount" depending on the dimension of the fleet and, perhaps, the line of business of the client.

The insurer intended 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 Desjardins *et al.* (2001), which uses credibility theory in order to calculate theoretical *experience rating*[1] coefficients based on the observed claims history.

[1] Desjardins *et al.* (2001) use the expression "Bonus-Malus coefficients". We chose not to do so because we consider these credibility methods to be a different approach from the traditional Markov-chain based Bonus-Malus methodology.

A part of this methodology was already approached by Sobral (2008), but we apply it in a different context and broader scope. In that case, the data came from a vehicle leasing company, "Leaseplan", over a single time period (one year) and only the model which uses the claims experience at fleet level was applied. Also, the only software used was *R*. In our case, the data came from an insurer, covering several years of activity and the much bigger amount of data allowed a deeper analysis of the models. Moreover, we applied not only the experience rating scheme which uses the claims experience at fleet level ("Model 1"), but also another scheme which uses the claims experience at vehicle level ("Model 2"). In addition, 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 these two models proposed by Desjardins *et al.* (2001). The learning of both programming languages and software was included in the internship.

In Model 1 (where only the claims history at fleet level is taken into account), the experience rating varies inside each fleet, although not significantly. The reason is that, under this model, the estimated future claims of each vehicle influence its experience rating. This is not exactly what the insurer wanted. However, we may adapt the experience rating to be applied to each fleet as a whole simply by taking averages of the credibility coefficients. Also, we propose a slight modification on the estimators used for computing the experience rating coefficients.

Rather than the claims history aggregated by fleet, Model 2 considers the claims history of each vehicle. Hence, it affects the tariff in two layers: one at fleet level, influencing equally all vehicles, and another at vehicle level, affecting each vehicle independently. Naturally, this allows for greater variations in the *a posteriori* rating inside each fleet (provided that there are claims). The advantages for fleet management are evident, since this highlights which vehicles have a "bad" claim behavior, through more severe penalties. This experience rating scheme was not approached by Sobral (2008). In any case, the advantages of doing so in that context would be slim. Leaseplan, being ultimately a fleet renting company, would have little gain in differentiating the experience rating vehicle per vehicle. It has absolutely no control over who is driving them.

Both systems have potential use by the insurer. The first one could be the base for a new *a posteriori* tariff for fleets up to a certain dimension, which would treat each fleet as a unit. The second one could be used for building a tool for decision making when negotiating premiums for large fleets, for example, greater than 300 vehicles, or with aggregated effective premium greater than €50 000.

1. Fleets insured in Fidelidade

In the first section of this chapter, we describe the current situation relating to *a posteriori* rating of auto insurance in the Portuguese insurer Fidelidade, with particular focus on business clients. In the second section, we describe in detail the portfolio of policies for fleets of vehicles of Fidelidade.

1.1. Current fleet tariff system

Fidelidade currently sells policies for fleets of vehicles, divided essentially in two categories: Fleets of less than 10 vehicles, belonging either to families or small businesses, and fleets of 10 vehicles or more, belonging to bigger companies. The tariff is currently applied to each vehicle separately, although the whole fleet is aggregated in a single "client account" for convenience of the client. A discount will be made, depending on the fleet dimension. Also, there may be adjustments related to the additional subscription of other kinds of insurance. For ratemaking purposes, this is a set of independent insured vehicles to which a commercial discount is applied.

The variables taken into account in the *a priori* rating are listed in the beginning of Section A.2. With respect to the *a posteriori* tariff system, it will eventually modify only the fraction of the premium relative to the vehicle which originated the claim, leaving the remaining of the fleet untouched. For each vehicle, we consider the *Bonus-Malus* classes displayed in Table A.2 (page 40), which are exactly the same as in a single vehicle insurance. For example, a vehicle in class 17 will pay the base premium as prescribed by the *a priori* rating; one in class 14 will pay 140% of the base premium; one in class 25 will pay 60% of the base premium.

After determining the *a priori* tariff for all the vehicles in a fleet, the entry class for each one is determined based on two criteria: Number of previous years of insurance in Fidelidade and claims experience of the vehicle in the previous 5 years[1].

Table A.3 in page 41 applies, for entry classes. For example, a vehicle with three years of insurance and one claim two years ago would go to the 10% discount class (18). If the claim was last year, the driver would go to the 0% class (17). As for the rules of transition, the most important are the following:

- If one has the maximum bonus and no claims in the current year (classes 31 to 34), one keeps the bonus. In case of a single claim, one keeps the bonus and only claims in the following three years would increase the premium.
- If one is in any other class and no claims, one goes up one class (meaning a lower penalty or higher discount); in case of a claim, one goes down at least two classes, unless one already is in class 11 or 12.

For higher claim numbers, the transition rules are not as simple. We refer to Table A.2 in page 40 for further details.

1.2. Description of the portfolio

In this section we give a broad view of the portfolio of motor insurance for business clients of Fidelidade, including roughly one million vehicles and spanning 7 years of activity (from 2007 to 2013). For all rating variables, a more detailed description may be found in Section A.4 of the Appendix. We observe that IBNR claims are very unlikely, since this study was carried out in 2015. In this work, we only considered

[1] In Portugal, insurance policies are linked to the license plate of the vehicle and the mandatory available data from other insurers is only the claims experience per policy in the previous 5 years.

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. More details on this may also be found in the Appendix (Section A.3). 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 seven characteristics: legal category, age, fuel and capital insured, at vehicle level; fleet dimension, type of economic activity and geographical area, at fleet level.

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. The details are in Table A.4, on page 44. We summarize the distribution of the vehicles in Fig. 1.1.

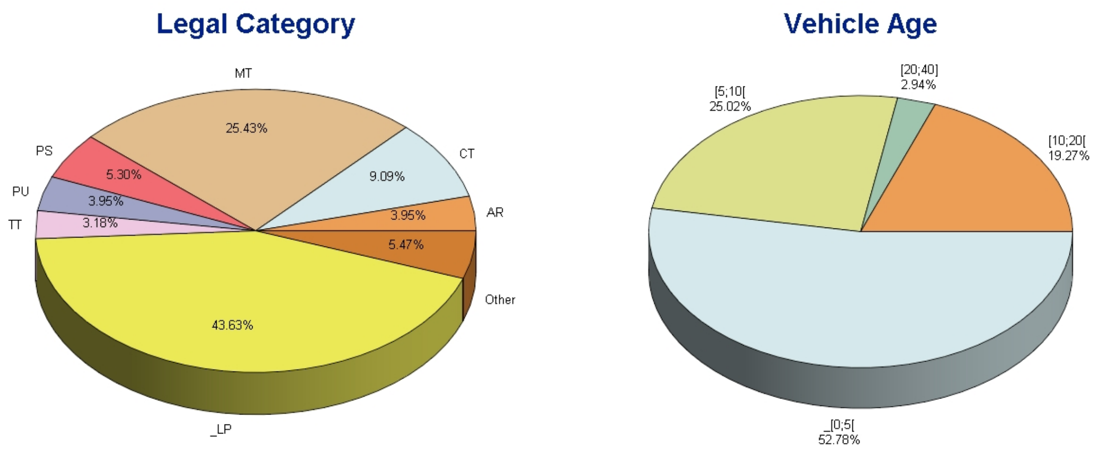

Figure 1.1.: Distribution of the vehicles by legal category and age.

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. 1.2): "Diesel", "Gasoline", "Others" and "No info".

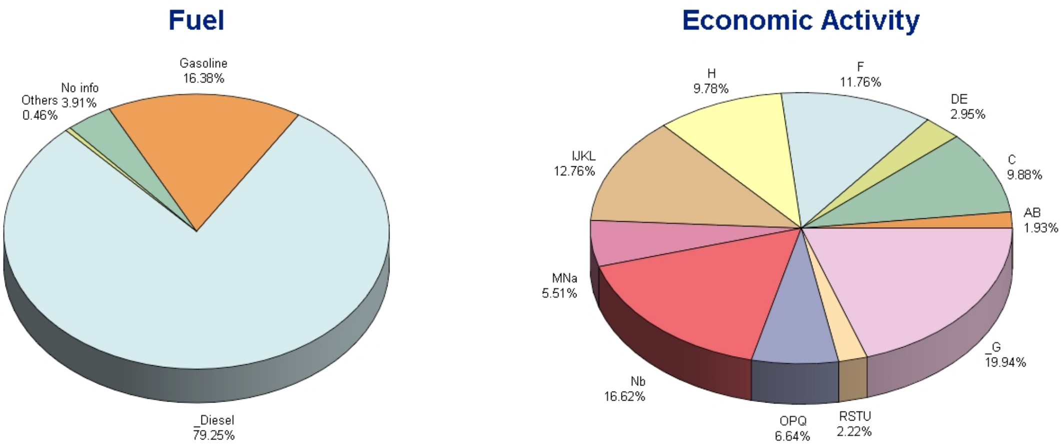

Figure 1.2.: Distribution of the vehicles by fuel and economic activity of their firm.

The next factor taken into account was the kind of **economic activity** practiced by the fleet owner. We based our own classification, displayed in Table 1.1, on the Portuguese economic classification code (CAE) by letters. For more details, please see page 44 of Section A.4. The results are summarized in Fig. 1.2. The biggest category is "G - Commerce and auto-repairs", followed by "Nb - Vehicle rentals".

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

Table 1.1.: Distribution of the vehicles by economic activity of their firm.

As for the **fleet dimension**, we had a few issues grouping the vehicles into fleets, which we describe in Section A.3. We ended up grouping them 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. 1.3 (in percentage over the total number of vehicles).

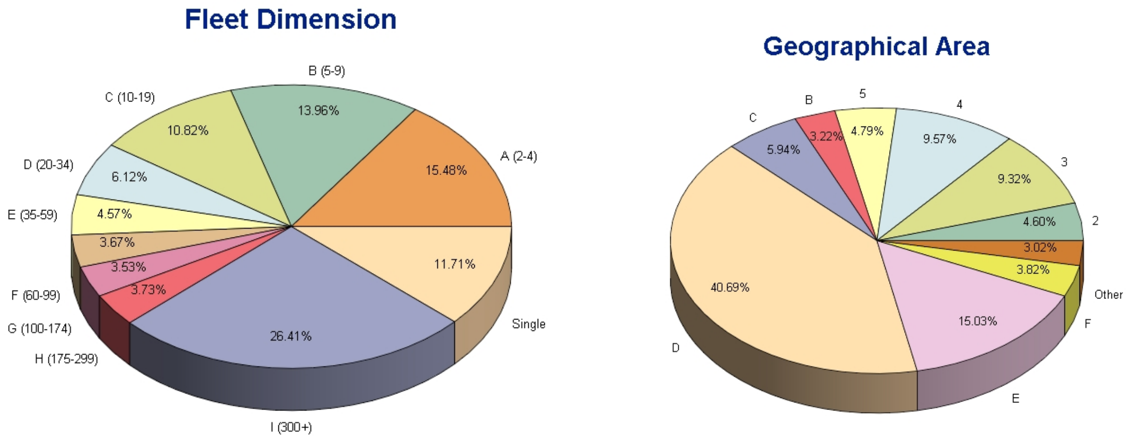

Figure 1.3.: Distribution of the vehicles by the size of their fleet and by the geographical area of their firm.

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: 1, A, 2, B, ..., E, 6 and F. The explanation for the curious labeling is given in the Appendix (page 45 and following). 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 1 to the "most" urban area F. That is not the case for fleets. Nevertheless, we considered that there was a fair chance of obtaining a coherent variable, after appropriate changes described in Section 3.1.

We chose not to include the **capital insured** of the vehicle as a rating factor because, in motor third party liability insurance, its relevance is slim. In Portugal, the legal requirements for compulsory liability limit differ almost exclusively according to the legal category of the vehicle and the kind of economic activity of the owner, which are factors already taken into account.

Before ending this chapter, a word on the exposure. We suspected that vehicles with low exposure time tended to have a higher claims rate. We decided to investigate, including this effect in a GLM model for the whole period 2007-2013. We added a new factor, **exposure of the vehicle**, bearing the levels "shorter" (up to a quarter, including 70 652 vehicles) and "longer" (between one quarter and seven years, including 801 229 vehicles). This factor turned out to be statistically significant and the inclusion of a vehicle in level "shorter" would mean a claims experience 69.3% higher, on average. Moreover, if one would consider the coverage related to collision damages (27 625 vehicles with "shorter" exposure and 371 330 with "higher" exposure) which, unlike the previous one, covers own damages, the increase would be higher than 500%.

Of course this factor cannot be used in *a priori* rating and so we did not use these results for further calculations involving the credibility models described in chapter 2. Nevertheless, this confirmed that the claims experience for short-term policies is much higher than the average. This could be related with the common effect of "fleeing after a claim", since in the very competitive context of motor insurance in Portugal, sometimes insurers disregard the claims history of potential clients in order to get new business. It could also be related with "total loss" claims, which naturally imply termination of contract. Another possibility is that this might also be an indicator of fraud. Fraudsters have a tendency for underwriting policies with monthly or quarterly installments and cancel them after participating the claim. We did not investigate further since this is not the core of this work, but we believe it is a good suggestion for future research to investigate this increased claims experience in low exposure policies.

2. Experience rating for fleets

In this chapter, we describe two models by Desjardins, Dionne and Pinquet, which provide experience rating coefficients for each vehicle of each fleet using credibility theory. These are hierarchical credibility models. We detail the results of Desjardins *et al.* (2001), which are not of easy deduction. In Section 2.1, we also took into account the work of Sobral (2008) and Nora (2004). The notations used are summarized on page 39. For a detailed exposition about experience rating systems, see for example (Centeno (2003)) or (Denuit et al. (2007)).

2.1. A Poisson model in a stratified portfolio

Portfolios of fleets 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 Jewell (1975)).

The system currently used in Fidelidade is a limit situation, in the sense that the history of a vehicle cannot be used to predict risk levels of other vehicles in the fleet. At the opposite end lie systems bearing only a fleet specific effect. In the Appendix (page 47) we summarize Bühlmann's model (see Bühlmann (1967)) and use it to build a simple example of such a system.

Now we present a model by Desjardins *et al*, 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

$$N_{fi} \sim \text{Poisson}\left(t_{fi}\lambda_{fi}u_{fi}\right), \; X_{fi} \stackrel{\text{def}}{=} \frac{N_{fi}}{t_{fi}}, \; f = 1, \ldots, F\,; i = 1, \ldots, m_f. \tag{2.1}$$

The random variables X_{fi} represent the average number of claims per year and are supposed to be independent (as well as N_{fi}). The *a priori* frequency risk of vehicle i of fleet f is $t_{fi}\lambda_{fi} \stackrel{\text{def}}{=} \mu_{fi}$ (whereas the *a priori* frequency risk per year is λ_{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:

$$\mu_{fi} = t_{fi} \exp\left(\underline{y}_f \underline{\gamma} + \underline{z}_{fi}\underline{\delta}\right), \qquad U_{fi} = R_f S_{fi} \tag{2.2}$$

where vectors $\hat{\gamma}$ and $\hat{\delta}$ are the maximum likelihood estimators. The *a priori* rating model is a Poisson model with neither fixed nor random effects, which we describe in detail in Section 3.1.

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 $\{R_f\}_{f=1,\ldots,F}$ and $\{S_{fi}\}_{f=1,\ldots,F;\, i=1,\ldots,m_f}$ 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 $\mathrm{E}(R) = \mathrm{E}(S) = 1$ and define $\mathrm{Var}(R) = V_{RR}$; $\mathrm{Var}(S) = V_{SS}$; $\mathrm{Var}(U) = V_{UU}$. In this semi-parametric approach, the distributions of the random effects will only be specified by the variances, because from $U = RS$, we get

$$\mathrm{E}(U) = \mathrm{E}(RS) = \mathrm{E}(R)\,\mathrm{E}(S) = 1. \tag{2.3}$$

The second part of equation (2.3) reflects the natural hypothesis that the *a priori* rating captures the mean of the risk, i.e. $t_{fi}\lambda_{fi} = \mathrm{E}(N_{fi})$. Also, we have

$$\begin{aligned} \mathrm{V}_{UU} = \mathrm{Var}(RS) &= \mathrm{E}\left(R^2S^2\right) - \left[\mathrm{E}(RS)\right]^2 = \mathrm{E}\left(R^2\right)\mathrm{E}\left(S^2\right) - 1 \\ &= (V_{RR}+1)(V_{SS}+1) - 1 \\ &= V_{RR} + V_{SS} + V_{RR}V_{SS} \end{aligned} \tag{2.4}$$

$$\begin{aligned} \mathrm{Var}(N_{fi}) &= \mathrm{E}\left(\mathrm{Var}(N_{fi}|U_{fi})\right) + \mathrm{Var}\left(\mathrm{E}(N_{fi}|U_{fi})\right) \\ &= \mathrm{E}(t_{fi}\lambda_{fi}U_{fi}) + \mathrm{Var}(t_{fi}\lambda_{fi}U_{fi}) \\ &= \mu_{fi} + \mu_{fi}^2 V_{UU}, \end{aligned} \tag{2.5}$$

$$\begin{aligned} \mathrm{Cov}(N_{fi}, N_{fj}) &= \mathrm{Cov}(t_{fi}\lambda_{fi}U_{fi}, t_{fi}\lambda_{fj}U_{fj}) \qquad (\text{for } i \neq j,) \\ &= \mu_{fi}\mu_{fj}\mathrm{Cov}(U_{fi}, U_{fj}) \\ &= \mu_{fi}\mu_{fj}V_{RR}. \end{aligned} \tag{2.6}$$

In (2.6) we used the assumed independence inside family $\{S_{fi}\}$. The portfolio is large, so we may use a frequentist approach, substituting the parameters by consistent estimators. From the moments computed above, we may deduce the following limits:

$$\hat{V}_{RR} = \frac{\sum\limits_{f}\sum\limits_{1\leq i\neq j\leq m_f}(n_{fi}-\hat{\mu}_{fi})(n_{fj}-\hat{\mu}_{fj})}{\sum\limits_{f}\sum\limits_{1\leq i\neq j\leq m_f}\hat{\mu}_{fi}\hat{\mu}_{fj}} = \frac{\sum\limits_{f,i\neq j} t_{fi}t_{fj}\left(x_{fi}-\hat{\lambda}_{fi}\right)\left(x_{fj}-\hat{\lambda}_{fj}\right)}{\sum\limits_{f,i\neq j} t_{fi}t_{fj}\hat{\lambda}_{fi}\hat{\lambda}_{fj}} \longrightarrow V_{RR}, \tag{2.7}$$

$$\hat{V}_{UU} = \frac{\sum\limits_{f,i}\left[(n_{fi}-\hat{\mu}_{fi})^2 - n_{fi}\right]}{\sum\limits_{f,i}\hat{\mu}_{fi}^2} = \frac{\sum\limits_{f,i}\left[t_{fi}^2\left(x_{fi}-\hat{\lambda}_{fi}\right)^2 - t_{fi}x_{fi}\right]}{\sum\limits_{f,i}t_{fi}^2\hat{\lambda}_{fi}^2} \longrightarrow V_{UU} \tag{2.8}$$

It is therefore possible to obtain consistent estimators of $\mathrm{V}(U)$ and $\mathrm{V}(R)$ based on the *a priori* model (the limits above considered convergence in probability). Also, from equation 2.4 we obtain a consistent estimator of $\mathrm{V}(S)$,

$$\hat{V}_{SS} = \frac{\hat{V}_{UU} - \hat{V}_{RR}}{1 + \hat{V}_{RR}}. \tag{2.9}$$

Estimator $\hat{V}_{RR}$ assesses observed contagion between the claims histories of vehicles within the same fleet. 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. Now we will deduce another formula for $\hat{V}_{RR}$ which is better for computational purposes. If we define

$$n_f = \sum_{i=1}^{m_f} n_{fi}; \quad \hat{\mu}_f = \sum_{i=1}^{m_f} \hat{\mu}_{fi}, \tag{2.10}$$

and consider the numerator in formula (2.7),

$$\sum_f \sum_{1 \le i \ne j \le m_f} (n_{fi} - \hat{\mu}_{fi})(n_{fj} - \hat{\mu}_{fj}), \tag{2.11}$$

we have

$$\begin{aligned}
(2.11) &= \sum_f \sum_i (n_{fi} - \hat{\mu}_{fi}) \sum_{j \ne i} (n_{fj} - \hat{\mu}_{fj}) \\
&= \sum_f \sum_i (n_{fi} - \hat{\mu}_{fi}) \left[(n_f - \hat{\mu}_f) - (n_{fi} - \hat{\mu}_{fi})\right] \\
&= \sum_f \sum_i \left[n_{fi} n_f - n_{fi} \hat{\mu}_f - \hat{\mu}_{fi} n_f + \hat{\mu}_{fi} \hat{\mu}_f - (n_{fi} - \hat{\mu}_{fi})^2\right] \\
&= \sum_f \left[n_f^2 - n_f \hat{\mu}_f - \hat{\mu}_f n_f + \hat{\mu}_f^2 - \sum_i (n_{fi} - \hat{\mu}_{fi})^2\right] \\
&= \sum_f (n_f - \hat{\mu}_f)^2 - \sum_{f,i} (n_{fi} - \hat{\mu}_{fi})^2 .
\end{aligned}$$

On the other hand, the denominator in formula (2.7) may be written as

$$\begin{aligned}\sum_f \sum_{1\leq i\neq j\leq m_f} \hat{\mu}_{fi}\hat{\mu}_{fj} &= \sum_f \sum_i \hat{\mu}_{fi} \sum_{j\neq i} \hat{\mu}_{fj} = \sum_f \sum_i \hat{\mu}_{fi} \left(\hat{\mu}_f - \hat{\mu}_{fi}\right) \\ &= \sum_f \sum_i \left(\hat{\mu}_{fi}\hat{\mu}_f - \hat{\mu}_{fi}^2\right) \\ &= \sum_f \hat{\mu}_f^2 - \sum_f \sum_i \hat{\mu}_{fi}^2.\end{aligned}$$

Hence, we have that

$$\hat{V}_{RR} = \frac{\sum\limits_f \left(n_f - \hat{\mu}_f\right)^2 - \sum\limits_{f,i} \left(n_{fi} - \hat{\mu}_{fi}\right)^2}{\sum\limits_f \hat{\mu}_f^2 - \sum\limits_{f,i} \hat{\mu}_{fi}^2}. \tag{2.12}$$

We remark that these estimators are not bounded. Although a variance is known to be non-negative, its estimation could be negative. If that would happen, i.e, 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.

Before ending this section, we would like to point out that the quadratic exposure terms in the numerators in formulas (2.8) and (2.7) seem counter-intuitive, since one could expect them to be linear, like in the Bühlmann-Straub method. In fact, these estimators are not the only ones possible to deduct from formulas (2.5) and (2.6) above. After equating V_{UU} or V_{RR} to a quotient, we may multiply the numerator and denominator by any constant, without affecting the consistency of the resulting estimators. Through multiplication by suitable constants in both quotients, we obtain

$$\hat{V}_{UU} = \frac{\sum\limits_{f,i} \left[t_{fi}^{-1} \left(n_{fi} - \hat{\mu}_{fi}\right)^2 - t_{fi}^{-1} n_{fi}\right]}{\sum\limits_{f,i} t_{fi}^{-1} \hat{\mu}_{fi}^2} = \frac{\sum\limits_{f,i} \left[t_{fi} \left(x_{fi} - \hat{\lambda}_{fi}\right)^2 - x_{fi}\right]}{\sum\limits_{f,i} t_{fi} \hat{\lambda}_{fi}^2} \longrightarrow V_{UU}, \tag{2.13}$$

$$\hat{V}_{RR} = \frac{\sum\limits_{f} \sum\limits_{1 \le i \ne j \le m_f} \sqrt{t_{fi} t_{fj}} \left(x_{fi} - \hat{\lambda}_{fi} \right) \left(x_{fj} - \hat{\lambda}_{fj} \right)}{\sum\limits_{f} \sum\limits_{1 \le i \ne j \le m_f} \sqrt{t_{fi} t_{fj}} \hat{\lambda}_{fi} \hat{\lambda}_{fj}} \longrightarrow V_{RR}. \tag{2.14}$$

If we define

$$n_f = \sum_{i=1}^{m_f} \left(\sqrt{t_{fi}} \right)^{-1} n_{fi}; \quad \hat{\mu}_f = \sum_{i=1}^{m_f} \left(\sqrt{t_{fi}} \right)^{-1} \hat{\mu}_{fi}, \tag{2.15}$$

we get

$$\hat{V}_{RR} = \frac{\sum\limits_{f} \left(n_f - \hat{\mu}_f \right)^2 - \sum\limits_{f,i} t_{fi}^{-1} \left(n_{fi} - \hat{\mu}_{fi} \right)^2}{\sum\limits_{f} \hat{\mu}_f^2 - \sum\limits_{f,i} t_{fi}^{-1} \hat{\mu}_{fi}^2}, \tag{2.16}$$

after performing similar computations to those described above. We applied these and the previous estimators to our data and commented on the results (see Section 3.2).

2.2. Experience rating using credibility

In this section we derive two experience rating schemes from the model described above. Let i_0 be a vehicle belonging to fleet f_0, which has m 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. Please see the notations on page 39.

2.2.1. 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 is of the form

$$\hat{a}_{i_0} + \hat{b}_{i_0} \left(\sum_{i=1}^{m} n_i \right). \tag{2.17}$$

We remark that no specific weight is being given to the history of vehicle i_0. The credibility coefficients $\hat{a}_{i_0}$ and $\hat{b}_{i_0}$ are such that the quadratic loss function below is minimized (and hence, they depend on the vehicle).

$$\mathrm{E}\left[\left(U_{i_0} - a - b \sum_{i=1}^{m} N_i \right)^2 \right] = \underbrace{\mathrm{Var}\left(U_{i_0} - b \sum_{i=1}^{m} N_i \right)}_{f(b)} + \underbrace{\left(\mathrm{E}\left[U_{i_0} - b \sum_{i=1}^{m} N_i \right] - a \right)^2}_{g(a,b)} \tag{2.18}$$

Taking the derivative of g with respect to a and equating to zero will yield

$$\hat{a}_{i_0} = 1 - \hat{b}_{i_0} \sum_{i=1}^{m} \hat{\mu}_i.$$

On the other hand, we have

$$f(b) = \mathrm{Var}(U_{i_0}) + b^2 \mathrm{Var}\left(\sum_{i=1}^{m} N_i \right) - 2b \sum_{i=1}^{m} \mathrm{Cov}(U_{i_0}, N_i).$$

Taking the derivative with respect to b and equating to zero will yield

$$\hat{b}_{i_0} = \frac{\sum_i \mathrm{Cov}(U_{i_0}, N_i)}{\mathrm{Var}\left(\sum_i N_i \right)}. \tag{2.19}$$

Since $\mathrm{E}(U_{i_0}) = 1$, the experience rating coefficient for vehicle i_0 may be written as

$$(1 - cred_{i_0}) + cred_{i_0} \frac{\sum_i n_i}{\sum_i \hat{\mu}_i} \tag{2.20}$$

where $cred_{i_0}$ is the credibility weight given to vehicle i_0 (bearing a fleet-specific and a vehicle-specific component).

Hence, we have

$$
\begin{aligned}
(2.20) = (2.17) \quad &\Leftrightarrow \quad (1 - cred_{i_0}) + cred_{i_0} \frac{\sum_i \hat{n}_i}{\sum_i \hat{\mu}_i} = 1 - \hat{b}_{i_0} \sum_i \hat{\mu}_i + \hat{b}_{i_0} \sum_i n_i \\
&\Leftrightarrow \quad 1 - cred_{i_0} \left(\frac{\sum_i \hat{n}_i}{\sum_i \hat{\mu}_i} - 1 \right) = 1 - \hat{b}_{i_0} \left(\sum_i \hat{\mu}_i - \sum_i n_i \right) \\
&\Leftrightarrow \quad cred_{i_0} = \hat{b}_{i_0} \sum_i \hat{\mu}_i .
\end{aligned}
$$

Consistent estimators for the individual moments are (see (2.5) and (2.6) on page 12):

$$
\begin{aligned}
\hat{\mathrm{Cov}}\left(U_{i_0}, N_i\right) &= \hat{\mathrm{Cov}}\left(U_{i_0}, \hat{\mu}_i U_i\right) = \hat{\mu}_i \hat{\mathrm{Cov}}\left(U_{i_0}, U_i\right) = \begin{cases} t_i \hat{\lambda}_i \hat{V}_{RR}, & i_0 \neq i \\ t_i \hat{\lambda}_i \hat{V}_{UU}, & i_0 = i \end{cases}, \qquad (2.21) \\
\hat{\mathrm{Var}}\left(N_i\right) &= \hat{\mathrm{E}}\left(\hat{\mu}_i U_i\right) + \hat{\mathrm{Var}}\left(\hat{\mu}_i U_i\right) = \hat{\mu}_i \hat{\mathrm{E}}\left(U_i\right) + \hat{\mu}_i^2 \hat{\mathrm{Var}}\left(U_i\right) \\
&= t_i \hat{\lambda}_i + t_i \hat{\lambda}_i^2 \hat{V}_{UU}, \qquad (2.22) \\
\hat{\mathrm{Cov}}\left(N_i, N_j\right) &= \hat{\mathrm{Cov}}\left(\hat{\mu}_i U_i, \hat{\mu}_j U_j\right) = \hat{\mu}_i \hat{\mu}_j \hat{\mathrm{Cov}}\left(U_i, U_j\right) \\
&= t_i t_j \hat{\lambda}_i \hat{\lambda}_j \hat{V}_{RR} \;\; (i \neq j) . \qquad (2.23)
\end{aligned}
$$

Considering formula (2.19), we will now express $\hat{b}_{i_0}$ in terms of the *a priori* estimates $\hat{\lambda}_i$ and the estimators just presented. We have

$$
\begin{aligned}
\sum_{\mathrm{i=1}}^{\mathrm{m}} \hat{\mathrm{Cov}}\left(U_{i_0}, N_i\right) &= \hat{V}_{UU} \mu_{i_0} + \sum_{i \neq i_0}^{m} \hat{V}_{RR} \mu_i , \qquad (2.24) \\
\hat{\mathrm{Var}}\left(\sum_{i=1}^{m} N_i\right) &= \sum_{i=1}^{m} \hat{\mathrm{Var}}\left(N_i\right) + \sum_{1 \leq j \neq i \leq m} \hat{\mathrm{Cov}}\left(N_i, N_j\right) \\
&= \sum_i \left(\hat{\mu}_i + \hat{\mu}_i^2 \hat{V}_{UU}\right) + \hat{V}_{RR} \sum_i \hat{\mu}_i \sum_{j \neq i} \hat{\mu}_j , \\
&= \sum_i \hat{\mu}_i + \hat{V}_{UU} \sum_i \hat{\mu}_i^2 + \hat{V}_{RR} \sum_i \hat{\mu}_i \left[\left(\sum_j \hat{\mu}_j \right) - \hat{\mu}_i \right] \\
&= \sum_i \hat{\mu}_i \left[1 + \hat{V}_{RR} \left(\sum_i \hat{\mu}_i \right) + \left(\hat{V}_{UU} - \hat{V}_{RR} \right) \frac{\sum_i \hat{\mu}_i^2}{\sum_i \hat{\mu}_i} \right] .
\end{aligned}
$$

We must consider two situations:

1. The vehicle was not in the fleet when the first period started, i.e., it was not observed: $i_0 \neq i$, $\forall i \in \{1, \ldots, m\}$. This means that (2.24) $= \sum_i \hat{V}_{RR}\mu_i$. Then

$$cred_{i_0} = \alpha = \hat{b}_{i_0} \sum_{i=1}^{m} t_i \hat{\lambda}_i = \sum_{i=1}^{m} t_i \hat{\lambda}_i = \frac{\hat{V}_{RR} \sum\limits_i t_i \hat{\lambda}_i}{1 + \hat{V}_{RR} \left(\sum\limits_i t_i \hat{\lambda}_i \right) + \left(\hat{V}_{UU} - \hat{V}_{RR} \right) \sum_{i=1}^{m} t_i \hat{\lambda}_i = \frac{\sum\limits_i t_i^2 \hat{\lambda}_i^2}{\sum\limits_i t_i \hat{\lambda}_i}}. \tag{2.25}$$

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:

$$cred_{i_0} = \alpha + \beta_{i_0}; \quad \beta_{i_0} = \frac{\left(\hat{V}_{UU} - \hat{V}_{RR} \right) t_{i_0} \hat{\lambda}_{i_0}}{1 + \hat{V}_{RR} \left(\sum\limits_i t_i \hat{\lambda}_i \right) + \left(\hat{V}_{UU} - \hat{V}_{RR} \right) \frac{\sum\limits_i t_i^2 \hat{\lambda}_i^2}{\sum\limits_i t_i \hat{\lambda}_i}}. \tag{2.26}$$

We may interpret $cred_{i_0}$ as the bonus granted to the firm if its fleet has no claims (see (2.20)). It 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 (2.9)). Our data satisfies this condition (see Section 3.2). 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 (again, see (2.20)).

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 $cred_{i_0} = \alpha + (1 - \rho)\,\bar{\beta}$, where $\bar{\beta}$ is the average of all vehicle-specific components β_i.

2.2.2. 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 Subsection 2.2.1 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 (see page 39) and all vectors are column vectors by default. We assume that the experience rating coefficient for vehicle i_0 has the form

$$\hat{a}_{i_0} + \underline{\hat{b}}_{i_0}^T \underline{n}; \qquad \underline{n} = (n_1, \ldots, n_m)^T, \ \underline{b}_{i_0} = (b_{i_0,1}, \ldots, b_{i_0,m})^T \tag{2.27}$$

where parameters $\hat{a}_{i_0}, \hat{b}_{i_0,1}, \ldots, \hat{b}_{i_0,m}$ minimize the quadratic loss function below:

$$\begin{aligned} \mathrm{E}\left[\left(U - a - \underline{b}^T \underline{N}\right)^2\right] &= \mathrm{Var}\left(U_{i_0} - a - \underline{b}^T \underline{N}\right) + \left(\mathrm{E}\left[U_{i_0} - a - \underline{b}^T \underline{N}\right]\right)^2 \\ &= \underbrace{\mathrm{Var}\left(U_{i_0} - \underline{b}^T \underline{N}\right)}_{f(\underline{b})} + \underbrace{\left(1 - a - \underline{b}^T \underline{\mu}\right)^2}_{g(a,\underline{b})}, \end{aligned} \tag{2.28}$$

where $\underline{\mu} = (\mu_1, \ldots, \mu_m)^T$ is the vector of frequency premiums, which were estimated through m.l.e. in the *a priori* rating. Let us consider g first. Taking the derivative with respect to a and equating to zero will yield $a = 1 - \underline{b}^T \underline{\mu}$, and then

$$\hat{a}_{i_0} + \underline{\hat{b}}_{i_0}^T \underline{n} = 1 + \underline{\hat{b}}_{i_0}^T \left(\underline{n} - \underline{\hat{\mu}}\right). \tag{2.29}$$

On the other hand,

$$\begin{aligned} f\left(\underline{b}_{i_0}\right) &= \mathrm{Var}\left(U_{i_0}\right) + \mathrm{Var}\left(\underline{b}_{i_0}^T \underline{N}\right) - 2\mathrm{Cov}\left(U_{i_0}, \underline{b}_{i_0}^T \underline{N}\right) \\ &= V_{UU} + \mathrm{Var}\left(\sum_{i=1}^m b_{i_0,i} N_i\right) - 2\sum_{i=1}^m b_{i_o,i} \mathrm{Cov}\left(U_{i_0}, N_i\right) \\ &= V_{UU} + \sum_i b_{i_0,i}^2 \mathrm{Var}\left(N_i\right) + \sum_i \sum_{j \neq i} b_{i_0,i} b_{i_0,j} \mathrm{Cov}\left(N_i, N_j\right) + 2\sum_i b_{i_0,i} \mu_i \mathrm{Cov}\left(U_{i_0}, U_i\right). \end{aligned}$$

Now, we take into account the derivation of the estimators for the individual moments given by (2.21), (2.22) and (2.23) on page 17, and proceed to minimize f. We must consider two cases:

1. Differentiating with respect to b_i, with $i \neq i_0$, and equating to zero yields

$$\begin{aligned} & 2b_{i_0,i}\left(\mu_i + \mu_i^2 V_{UU}\right) + 2V_{RR}\mu_i \sum_{j\neq i} b_{i_0,j}\mu_j - 2V_{RR}\mu_i = 0 \\ \Leftrightarrow \quad & b_{i_0,i}\left[\mu_i + \mu_i^2\left(V_{UU} - V_{RR}\right)\right] + V_{RR}\mu_i \sum_j \mu_j b_{i_0,j} - V_{RR}\mu_i = 0. \end{aligned}$$

2. Doing the same for $i = i_0$ yields

$$b_{i_0,i}\left[\mu_i + \mu_i^2\left(V_{UU} - V_{RR}\right)\right] + V_{RR}\mu_i \sum_j \mu_j b_{i_0,j} - V_{UU}\mu_{i_0} = 0.$$

We obtained a system of linear equations, which we now write in matrix form,

$$D\underline{b} + A\underline{b} = \underline{v}, \tag{2.30}$$

where matrices D, A and vector v are given by

$$D = \operatorname*{diag}_{1\leq i\leq m}\left[\mu_i + \mu_i^2\left(V_{UU} - V_{RR}\right)\right], \quad A = V_{RR}\underline{\mu}\underline{\mu}^T, \quad \underline{v} = V_{RR}\underline{\mu} + \left(V_{UU} - V_{RR}\right)\underline{\mu}\underline{e}_{i_0}$$

and $\underline{e}_{i_0}$is the i_0-th vector of the canonic $\mathbb{R}^m$ basis. The solution is

$$\underline{\hat{b}}_{i_0} = \left(\hat{D} + \hat{A}\right)^{-1}\underline{\hat{v}}. \tag{2.31}$$

Now we will compute $\left(\hat{D} + \hat{A}\right)^{-1}$. First, we define the M-norm of a vector $\underline{u}$,

$$\left\|\underline{u}\right\|_M \stackrel{\text{def}}{=} \underline{u}^T M \underline{u}\left(\in \mathbb{R}_0^+\right), \tag{2.32}$$

where M is any invertible matrix. We note that

$$\left(\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)^2 = \hat{D}^{-1}\underline{\hat{\mu}}\overbrace{\underline{\hat{\mu}}^T\hat{D}^{-1}\underline{\hat{\mu}}}^{\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\underline{\hat{\mu}}^T = \|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T.$$

This formula will be used below. Now, we have

$$\left(\hat{D}+\hat{A}\right)^{-1} = \left[\hat{D}\left(\mathbb{I}_m + \hat{V}_{RR}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)\right]^{-1}.$$

Assuming $\left(\mathbb{I}_m + \hat{V}_{RR}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)^{-1}$ of the form $\left(\mathbb{I}_m + \gamma\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)$, $\gamma\in\mathbb{R}$, we have

$$\begin{aligned}
&\left(\mathbb{I}_m + \hat{V}_{RR}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)\left(\mathbb{I}_m + \gamma\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right) = \mathbb{I}_m\\
\Leftrightarrow\ &\mathbb{I}_m + \hat{V}_{RR}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T + \gamma\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T + \gamma\hat{V}_{RR}\left(\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right)^2 = \mathbb{I}_m\\
\Leftrightarrow\ &\left(\hat{V}_{RR}+\gamma\right)\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T + \gamma\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T = \mathbb{O}_m\\
\Leftrightarrow\ &\hat{V}_{RR}+\gamma+\gamma\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}} = 0\\
\Leftrightarrow\ &\gamma = -\frac{\hat{V}_{RR}}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}.
\end{aligned}$$

Then, a first expression for (2.31) is

$$\begin{aligned}
\underline{\hat{b}}_{i_0} &= \left[\mathbb{I}_m - \frac{\hat{V}_{RR}}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\right]\hat{D}^{-1}\left[\hat{V}_{RR}\underline{\hat{\mu}} + \left(\hat{V}_{UU}-\hat{V}_{RR}\right)\underline{\hat{\mu}}\underline{e}_{i_0}\right]\\
&= \left[\hat{D}^{-1}\underline{\hat{\mu}} - \frac{\hat{V}_{RR}}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\hat{D}^{-1}\underline{\hat{\mu}}\underline{\hat{\mu}}^T\hat{D}^{-1}\underline{\hat{\mu}}\right]\left[\hat{V}_{RR} + \left(\hat{V}_{UU}-\hat{V}_{RR}\right)\underline{e}_{i_0}\right]
\end{aligned}\qquad(2.33)$$

Also, noting that $\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}} = \underline{\hat{\mu}}^T\hat{D}^{-1}\underline{\hat{\mu}}$, the first factor is equal to

$$\left(\hat{D}^{-1}\underline{\hat{\mu}} - \frac{\hat{V}_{RR}}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\hat{D}^{-1}\underline{\hat{\mu}}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}\right) = \left(1 - \frac{\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\right)\hat{D}^{-1}\underline{\hat{\mu}}.$$

Hence, we have

$$\underline{\hat{b}}_{i_0} = \frac{1}{1+\hat{V}_{RR}\|\underline{\hat{\mu}}\|_{\hat{D}^{-1}}}\left[\hat{V}_{RR}\hat{D}^{-1}\underline{\hat{\mu}} + \hat{\mu}_{i_0}\left(\hat{V}_{UU}-\hat{V}_{RR}\right)\hat{D}^{-1}\underline{e}_{i_0}\right].\qquad(2.34)$$

Now we return to the experience rating coefficient for vehicle i_0. We have

$$(2.29) = 1 + \hat{\underline{b}}_{i_0}^T \left(\underline{n} - \hat{\underline{\mu}}\right) = 1 + \sum_{i=1}^{m} \alpha_i \left(\frac{n_i}{\hat{\mu}_i} - 1\right) + \beta_{i_0}\left(\frac{n_{i_0}}{\hat{\mu}_{i_0}} - 1\right),$$

where

$$\alpha_i = \hat{\mu}_i \hat{b}_{i_0,i} = t_i \hat{\lambda}_i \hat{b}_{i_0,i} \ (i \neq i_0)\,; \quad \alpha_{i_0} + \beta_{i_0} = \hat{\mu}_{i_0} \hat{b}_{i_0,i_0} = t_{i_0} \hat{\lambda}_{i_0} \hat{b}_{i_0,i_0}. \tag{2.35}$$

We will express this experience rating coefficient as a function of the estimates $\hat{V}_{UU}$, $\hat{V}_{RR}$, $\hat{\mu}_i$, $(i = 1, \ldots, n)$. 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.

Unlike the previous model, now the credibility coefficient β_{i_0} is applied to the individual claims history of the vehicle. This will result in a greater within fleets dispersion. The next step is to deduce formulas for the α_i's and β_{i_0}. First we note that, because $\hat{D}$ is a diagonal matrix,

$$\left(\hat{D}^{-1}\hat{\mu}\right)_i = \sum_{j=1}^{m} \hat{D}_{ij}^{-1}\hat{\mu}_j = \sum_{j=1}^{m} \frac{\delta_{ij}\hat{\mu}_j}{\hat{\mu}_i + \hat{\mu}_i^2\left(\hat{V}_{UU} - \hat{V}_{RR}\right)} = \frac{1}{1 + \hat{\mu}_i\left(\hat{V}_{UU} - \hat{V}_{RR}\right)}, \tag{2.36}$$

where δ is the *Kronecker delta,* i.e., the indicator function. Consequently, we have

$$1 + \hat{V}_{RR}\left\|\underline{\hat{\mu}}\right\|_{\hat{D}^{-1}} = 1 + \hat{V}_{RR}\left(\underline{\hat{\mu}}^T \hat{D}^{-1}\underline{\hat{\mu}}\right) = 1 + \hat{V}_{RR}\sum_{j=1}^{m} \frac{\hat{\mu}_j}{1 + \hat{\mu}_j\left(\hat{V}_{UU} - \hat{V}_{RR}\right)}. \tag{2.37}$$

Then, from equations (2.34) and (2.35) we get, for new vehicles in the fleet $(1 \leq i \leq m)$,

$$\alpha_i = \hat{\mu}_i \hat{b}_{i_0,i} = \frac{\hat{\mu}_i}{1 + \hat{V}_{RR}\left\|\underline{\hat{\mu}}\right\|_{\hat{D}^{-1}}} \times \frac{\hat{V}_{RR}}{1 + \hat{\mu}_i\left(\hat{V}_{UU} - \hat{V}_{RR}\right)} = \frac{\frac{t_i\hat{\lambda}_i\hat{V}_{RR}}{1+t_i\hat{\lambda}_i\left(\hat{V}_{UU}-\hat{V}_{RR}\right)}}{1 + \sum_j \frac{t_j\hat{\lambda}_j\hat{V}_{RR}}{1+t_j\hat{\lambda}_j\left(\hat{V}_{UU}-\hat{V}_{RR}\right)}} \tag{2.38}$$

and, for vehicles already observed in the fleet ($1 \le i, i_0 \le m$),

$$\begin{aligned}
\alpha_{i_0} + \beta_{i_0} &= \hat{\mu}_{i_0} \hat{b}_{i_0,i_0} \qquad (2.39)\\
&= \frac{\hat{\mu}_i}{1 + \hat{V}_{RR} \left\| \underline{\hat{\mu}} \right\|_{\hat{D}^{-1}}} \times \left(\frac{\hat{V}_{RR}}{1 + \hat{\mu}_{i_0} \left(\hat{V}_{UU} - \hat{V}_{RR} \right)} + \frac{\hat{V}_{UU} - \hat{V}_{RR}}{1 + \hat{\mu}_{i_0} \left(\hat{V}_{UU} - \hat{V}_{RR} \right)} \right)\\
&\Rightarrow \beta_{i_0} = \frac{\frac{t_{i_0} \hat{\lambda}_{i_0} \left(\hat{V}_{UU} - \hat{V}_{RR} \right)}{1 + t_{i_0} \hat{\lambda}_{i_0} \left(\hat{V}_{UU} - \hat{V}_{RR} \right)}}{1 + \sum_j \frac{t_j \hat{\lambda}_j \hat{V}_{RR}}{1 + t_j \hat{\lambda}_j \left(\hat{V}_{UU} - \hat{V}_{RR} \right)}}. \qquad (2.40)
\end{aligned}$$

As in Subsection 2.2.1, 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 equation 2.9).

3. Application of the models

In this chapter we apply the models described in chapter 2 to the portfolio of fleets of vehicles of the insurer. In the first section, we estimate the number of claims per vehicle/year through generalized linear models (GLM), using the variables presented in Section 1.2. In the second section, we describe the results of applying the two experience rating schemes of Section 2.2 to our portfolio. In the last section, we present the main conclusions of our study.

3.1. 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 to 2013). More specifically, we used the GENMOD procedure in SAS. We performed the estimation for the time period 2007-2013. Before proceeding, we briefly describe the origin and scope of this kind of regression models. A detailed exposition may be found, for example, in (McCullagh and Nelder (1989)).

The generalized linear models are an extension of the classic linear models. Suppose we have n observations, $\underline{y} = (y_1, \ldots, y_n)^T$ of the variable of interest and also n observations of p covariates, or explanatory variables, organized in a $n \times p$ matrix X (whose n rows represent the n observations). The classic linear models are of the form

$$\underline{y} = X\gamma + \epsilon, \tag{3.1}$$

where $\gamma = (\gamma_1, \ldots, \gamma_n)^T$ is a set of unknown parameters, each one associated to a different covariate, and ϵ is the vector of random i.i.d. errors with distribution $N(0,1)$. The main assumptions of these models are that the expected value of the variable of interest is a linear function of the covariates, the so-called *linear predictor,*

$$\mathrm{E}(Y_i) = \mu_i = \sum_{j=1}^{p} x_{ij}\gamma_j, \quad i = 1, \ldots, n, \tag{3.2}$$

as well as normality of the residuals. The generalization of these models is made by:

1. Extending the scope of the distribution of the i.i.d. errors ϵ_i, which may then be any distribution in the exponential family.

2. Introducing the notion of *link function.* The relation between the expected value of the variable of interest and the linear predictor η is given by the function g:

$$\eta_i = \sum_{j=1}^{p} x_{ij}\gamma_j, \quad i = 1, \ldots, n; \qquad \eta_i = g(\mu_i) \tag{3.3}$$

We remark that Y belongs to the exponential family if its probability density function takes the form

$$f_Y(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) \tag{3.4}$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. It may be proved that

$$\mathrm{E}(Y) = \mu = b'(\theta); \qquad \mathrm{Var}(Y) = b''(\theta)\, a(\phi). \tag{3.5}$$

Among the members of the exponential family of distributions are, for example, the Normal and Gamma distributions, as well as the Binomial and the Poisson distributions.

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, i.e., a GLM where the variables Y_i are supposed to follow a Poisson distribution and the link is the logarithmic function (see Section 2.1). Also, we used the logarithm of the time exposure as an offset. The Poisson distribution is usually appropriate for modeling discrete data, like claim counts, and the logarithmic function is the canonical link for this distribution. This means that the link is such that the parameter θ of the exponential family of distributions, taken as a function of μ, satisfies $\eta(\mu) = \theta(\mu)$. Summing up, we assume that

$$\frac{\hat{\mu}_{fi}}{t_{fi}} = \hat{\lambda}_{fi} = \exp\left(\underline{y}_f \hat{\underline{\gamma}} + \underline{z}_{fi} \hat{\underline{\delta}}\right), \tag{3.6}$$

where vectors $\hat{\gamma}$ and $\hat{\delta}$ are the maximum likelihood estimators, obtained from equations

$$\sum_{f,i} (n_{fi} - \hat{\mu}_{fi})\, y_{f,k} = 0,\ k = 1, \ldots, K;\ \ \sum_{f,i} (n_{fi} - \hat{\mu}_{fi})\, z_{fi,l} = 0,\ l = 1, \ldots, L, \tag{3.7}$$

K being the number of explanatory variables relative to each fleet and L being the number of explanatory variables relative to each vehicle inside a fleet. Since, in our model with random effects, we have

$$\mathrm{E}(N_{fi}) = \mathrm{E}(\mathrm{E}(N_{fi}|U_{fi})) = \mu_{fi}\mathrm{E}(U) = \mu_{fi}, \tag{3.8}$$

the introduction of these factors does not modify the expected value for claim counts. Hence, the maximum likelihood estimators for the parameters of the *a priori* model are consistent estimators of the corresponding parameters in the model with random effects (see Gouriéroux et al. (1984)).

In our model, the dispersion parameter ϕ is 1 and $b(\theta) = \exp(\theta)$. Also, we specified no weights ω_{fi} (considering the common assumption $a(\phi) = \frac{\phi}{\omega_{fi}}$). Hence, the variance of each Y_i equals its mean (see equation 3.5). Again, for more details see for example (McCullagh and Nelder (1989)).

We point out that our approach is mathematically equivalent to replacing the claim count by the claim frequency (claims divided by time exposures - ${}^{N_{fi}}/{}_{t_{fi}}$) as our variable of interest, using time exposure as the weight and ruling out the offset. For a proof of this result please see (Yan et al. (2009)).

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 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 modeling 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 $]10;\infty]$, bearing 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.

Finally, we redesigned the **geographical area** factor. When including this factor in the GLM, the rural geographical areas 1 and A consistently got higher coefficients than the ”less rural“ regions 2 and B. We believe these results were due to two weaknesses in the design of the variable: First, it might happen that not all of the vehicles were allocated to the headquarters of the company, this being especially plausible for big fleets. Second, it might also happen that a significant number of vehicles circulated outside the company’s region frequently, or even daily.

In order to counter these adverse effects, we aggregated the 12 original geographical areas into 4 wider regions, labeled "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 vehicles, as well as smaller fleets whose economic activity implied long trips (for example, long distance transportation by road). A more accurate description may be found on pages 45 to 47 of the Appendix.

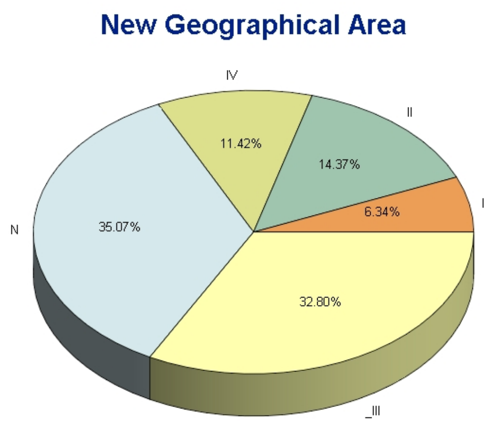

Figure 3.1.: Distribution of the vehicles by the new geographical areas.

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 favor 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.

3.2. Empirical results

In this section, we refer to the experience rating systems described in Subsection 2.2.1 and in Subsection 2.2.2 as Model 1 and Model 2, respectively. We start with a regard over the estimated variance of the random effect U (which is the product of the fleet-specific and vehicle-specific random effects R and S, respectively). 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 Section 2.1). The results are in Table 3.1.

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

Table 3.1.: Estimates according to Desjardins et al. (2001) and according to our modification.

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 by Desjardins *et al.* (2001) 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. 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 the estimators from the end of Section 2.1 are more adequate, especially when time exposure varies greatly between vehicles.

Also, from these results, we observe that the estimates for the variances display substantial variations over the years. That is probably due to the significant reduction in the number of clients that this portfolio endured from 2007 to 2013 (from 434 861 to 220 423 vehicles). As a consequence, the resulting experience rating coefficients for models 1 and 2 will also vary substantially over the years. This reduction has naturally changed fundamental characteristics of the portfolio, namely the variances of the random effects, both fleet-specific and vehicle-specific. Hence, we believe it is not advisable to rely on data relative to years which are too far apart, for developing an experience rating system.

Table 3.2 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 the middle of *Alentejo,* 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.

As for the column labels, they follow the notation described in page 39. Each column presents averages over all fleets in each category. Similar data for previous years may be found in the Appendix (Section A.6).

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 Desjardins *et al.* (2001).

λ_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 3.2.: Data respecting to Model 1 and Model 2 for 2013.

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 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. Again, the column labels follow the notation described in page 39. The labels "$\rho = \ldots$" represent the adaptation of Model 1 mentioned at the end of Subsection 2.2.1 for different values of the turnover (proportion of new cars in the fleet).

f	i	t_{fi}	μ_{fi}	n_{fi}	$\sum_i \mu_{fi}$	n_f	**ER1**	$\rho = 0$	$\rho = .3$	$\rho = .7$	**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

Table 3.3.: Data for the 17th fleet of the portfolio of 2012

In Table 3.3 we have a fleet of four vehicles without any claim. Nevertheless, the experience rating coefficients for Model 1 (labeled "ER1") vary according to the claims expectancy μ_{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.

f	i	t_{fi}	μ_{fi}	n_{fi}	$\sum_i \mu_{fi}$	n_f	**ER1**	$\rho = 0$	$\rho = .3$	$\rho = .7$	**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 3.4.: Data for the 174th fleet of the portfolio of 2012

In Table 3.4 we have a fleet of five vehicles with a claims experience far worse than expected (3 claims), given its expected claims $\sum_i \mu_{fi} = 0.5119$. 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.

f	i	t_{fi}	μ_{fi}	n_{fi}	$\sum_i \mu_{fi}$	n_f	**ER1**	$\rho = 0$	$\rho = .3$	$\rho = .7$	**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

Table 3.5.: Data for the 1115th fleet of the portfolio of 2012

In Table 3.5, 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. We remark that the higher standard deviation characteristic of Model 2 is due only partially to the penalized vehicle. There is more variation in the bonuses under Model 2 than under Model 1.

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 3.6.: Average experience rating coefficients for for fleets with $\mu_f \leq 0.125$ in 2013.

In tables Table 3.6, Table 3.7 and Table 3.8, we present the averages of the experience rating coefficients for Model 1 by Desjardins *et al.* (2001), for different values of the turnover of the fleet. For notation, please see page 39. The experience rating coefficients are organized by fleet *a priori* rating.

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 3.7.: Average experience rating coefficients for fleets with $1 \leq \mu_f \leq 1.5$ in 2013.

We point out that the coefficients in tables Table 3.6, Table 3.7 and Table 3.8 are averages over all fleets of the of the adaptation of Model 1, described in the end of Subsection 2.2.1. 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

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

Table 3.8.: Average experience rating coefficients for fleets with $4 \leq \mu_f \leq 8$ in 2013.

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 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%.

3.3. Conclusions

In this work, we focused on two experience rating schemes for fleets of vehicles by Desjardins *et al.* (2001), based on credibility, which we denominate "Model 1" and "Model 2". The main strength of Model 1 lies in its adaptation described at the end of Subsection 2.2.1, 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.

The main strength of Model 2 lies in the inherent variability of the rating coefficients inside each fleet. By construction, the coefficients of each vehicle are a sum of two terms: one related to the claims history and a priori rating of its fleet as a whole, and another influenced exclusively by the claims history and a priori rating of the vehicle at hand. The latter is of course zero for new vehicles. The main features of the variability of the coefficients are patent in the examples of Section 3.2 and have great practical interest, namely for fleet management purposes. The consistency of the experience rating coefficients of both models is remarkable, when we consider different claims experiences and fleet dimensions.

We applied the theory described in chapter 2 to the portfolio of Fidelidade, to each of the years 2007-2013. Only the mandatory coverage (third-party liability insurance) was taken into account. It is important to mention that the most time-consuming task during the internship was handling the raw data and presenting them in a manner that allowed reliable results.

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 (Desjardins *et al.* (2001)). 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.

When estimating the number of claims of each vehicle through GLM, some variables (or levels within them) that were statistically significant when considering the whole period 2007-2013, lose that property when considering each year separately. Nevertheless, from the variables (and levels within them) that remained statistically significant, 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.

The insurer will rely on the adaptation of the first model for constructing a bonus-malus system for fleets of vehicles whose yearly premia lie between certain bounds. Below these bounds, the current system will remain unchanged. Above them, the insurance contract is negotiated on a case-by-case basis, and the second model shows potential to be the basis for a simulator that will provide support in these negotiations. Our future work in the insurer will include the application of these models to other motor insurance coverages, allowing the development of a more general experience rating system.

Regarding the statistical software used (SAS Enterprise Guide and R), we emphasize the versatility and fast response displayed by both programs. SAS was used for raw data handling and a priori rating (through GLM), while R was used for computing the experience rating coefficients and handling the results. In the particular case of R, which was ran on a laptop with rather modest technical specifications, the lightness of the program and efficiency in computation were remarkable.

Bibliography

Bühlmann, H. (1967). The use of collateral data in credibility theory: A hierarchical model. *Astin Bulletin*, pages 199 – 207.

Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer, 1st edition.

Centeno, L. (2003). *Teoria do Risco na Actividade Seguradora*. Celta Editora.

Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley & Sons, Chichester, 1st edition.

Desjardins, D., Dionne, G., and Pinquet, J. (2001). Experience rating schemes for fleets of vehicles. *Astin Bulletin*, 31:81 – 105.

Gouriéroux, C., Monfort, A., and Trognon, A. (1984). Pseudo-likelihood methods: Application to poisson models. *Econometrica*, 52:701 – 720.

INE (2007). *Classificação Portuguesa das Actividades Económicas Rev.3*. Instituto Nacional de Estatística, I.P.

Jewell, W. S. (1975). The use of collateral data in credibility theory: A hierarchical model. *Giornale dell'Instituto degli Attuari*, 38:1 – 16.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2008). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., 3rd edition.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall, 2nd edition.

Nora, E. (2004). Sistema de Bonus-Malus para Frotas de Veículos. Master thesis, Instituto Superior de Economia e Gestão.

Sobral, P. (2008). Sistema de bonus-malus para frotas. Master thesis, Instituto Superior de Economia e Gestão.

Yan, J., Gusjcza, J., Flynn, M., and Wu, P. (2009). Applications of the offset in property-casualty predictive modelling. *Casualty Actuarial Society E-Forum*, Winter 2009:366 – 385.

A. Appendix

A.1. Notation

Notation	Description
f	a fleet
i	a vehicle in the fleet
i_0	a vehicle for which the ER1 or ER2 are being calculated
V_{UU}	Variance of the heterogeneity component of the models
V_{RR} (resp. V_{SS})	Variance of the fleet-specific (resp. vehicle-specific) effect
t_{fi}	time exposure of vehicle i of fleet f
μ_{fi}	predicted number of claims of vehicle i of fleet f
$\lambda_{fi} = {}^{\mu_{fi}}/_{t_{fi}}$	predicted number of claims of vehicle i of fleet f, per year
n_{fi}	observed number of claims of vehicle i of fleet f
$x_{fi} = {}^{n_{fi}}/_{t_{fi}}$	observed number of claims of vehicle i of fleet f, per year
m_f	total number of vehicles of fleet f
$\mu_f = \sum_i \mu_{fi}$	predicted number of claims of fleet f
$n_f = \sum_i n_{fi}$	observed number of claims of fleet f
α_f	credibility assigned to fleet f
$cred_{fi} = \alpha_f + \beta_{fi_0}$	credibility assigned to vehicle i_0 of fleet f
$\bar{\beta}_f$	average over fleet f of the credibilities respecting to its vehicles
ER1	experience rating coefficient for Model 1
σ_1	standard deviation of the ER coefficients for Model 1 inside a fleet
ρ	turnover coefficient (proportion of renewal of the fleet)
ER2	experience rating coefficient for Model 2
σ_2	standard deviation of the ER coefficients for Model 2 inside a fleet

Table A.1.: Notation related to the computation of the experience rating coefficients (subscript f dropped in chapter 2).

A.2. The rating system currently in force

With respect to a priori rating, the following variables are taken into account: Type of client (family or company) and, in case of a company, its line of business; age of the usual driver's license (optional for companies); legal category, age and fuel type; number of seats (only when the coverage "Passengers" is in force) and characteristics of the contract (subscribed options, insured capital and existence of deductibles and/or extensions of coverage). The horsepower, cylinder capacity, tare weight and gross weight[1] may also be required, depending on the legal category of the vehicle.

Situation in the previous year		Nr of claims in the previous year				
Class	Bonus/Malus	0	1	2	3	4 or more
11	130%	12	11	11	11	11
12	90%	13	11	11	11	11
13	60%	14	11	11	11	11
14	40%	15	12	11	11	11
15	20%	16	13	11	11	11
16	10%	17	14	12	11	11
17	0%	18	15	13	11	11
18	-10%	19	16	14	12	11
19	-20%	20	17	15	13	11
20	-25%	21	18	15	13	11
21	-30%	22	18	15	13	11
22	-33%	23	19	16	14	11
23	-35%	24	19	16	14	11
24	-38%	25	20	17	15	11
25	-40%	26	20	17	15	11
26	-42%	27	21	18	15	11
27	-44%	28	22	18	16	11
28	-46%	29	23	19	16	11
29	-48%	34	25	19	17	11
31	-50%	32	21	17	15	11
32	-50%	33	23	17	15	11
33	-50%	34	25	17	15	11
34	-50%	34	31	20	17	11

Table A.2.: Bonus-Malus classes and transition rules of the current tariff system.

[1]In Portuguese, "tara" (weight of the vehicle not counting cargo and passengers) and "peso bruto" (taking cargo and passengers into account).

Years	Nr. of claims in the last 5 years															
	0	1					2					3				
		Nr. of years with no claims														
insured		0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
New	0%	20%					60%					130%				
1	-10%	20%	10%	10%			60%	40%	40%			130%	90%	60%		
2	-20%	10%	10%	10%			60%	40%	40%			130%	90%	60%		
3	-25%	0%	0%	0%	-10%	-10%	40%	20%	20%	10%	10%	90%	90%	60%	40%	20%
4	-30%	0%	-10%	-10%	-10%	-10%	20%	20%	20%	10%	10%	60%	60%	60%	40%	20%
5	-32,5%	-10%	-20%	-20%	-20%	-20%	10%	10%	10%	0%	0%	60%	40%	40%	40%	20%
6	-35%	-10%	-20%	-20%	-25%	-25%	10%	0%	0%	-10%	-10%	60%	40%	40%	40%	20%
7	-37,5%	-10%	-20%	-20%	-25%	-25%	10%	0%	0%	-10%	-10%	40%	40%	20%	20%	10%
8	-37,5%	-20%	-25%	-25%	-30%	-30%	10%	0%	0%	-10%	-10%	40%	40%	20%	20%	10%
9	-40%	-20%	-25%	-25%	-30%	-30%	0%	0%	0%	-20%	-20%	40%	40%	20%	20%	10%
10	-42%	-20%	-25%	-25%	-30%	-30%	0%	0%	0%	-20%	-20%	40%	40%	20%	20%	10%
11	-42%	-20%	-25%	-25%	-30%	-30%	0%	0%	0%	-20%	-20%	40%	40%	20%	20%	10%
12	-44%	-20%	-25%	-25%	-30%	-30%	0%	0%	0%	-20%	-20%	40%	40%	20%	20%	10%
13	-46%	-25%	-30%	-30%	-32,5%	-32,5%	0%	0%	0%	-20%	-20%	40%	40%	20%	20%	10%
14	-48%	-25%	-30%	-30%	-32,5%	-32,5%	0%	-10%	-10%	-20%	-20%	20%	20%	20%	10%	0%
15	-50%	-30%	-30%	-30%	-32,5%	-32,5%	0%	-10%	-10%	-20%	-20%	20%	20%	20%	10%	0%
16	-50%	-32,5%	-32,5%	-32,5%	-35%	-35%	0%	-10%	-10%	-20%	-20%	20%	20%	20%	10%	0%
17	-50%	-32,5%	-32,5%	-32,5%	-35%	-35%	0%	-10%	-10%	-20%	-20%	20%	20%	20%	10%	0%
>17	-50%	-35%	-35%	-35%	-35%	-35%	0%	-10%	-10%	-20%	-20%	20%	20%	20%	10%	0%

Table A.3.: Entry classes in the *a posteriori* tariff system of Fidelidade, as a function of years insured and claim experience of previous 5 years.

A.3. Handling the data

Our data set is taken out from the portfolio of motor insurance for business clients of Fidelidade, including roughly one million vehicles and spanning 7 years of activity (from 2007 to 2013).

Before describing the portfolio, it is worth pointing out that Fidelidade is the result of a merger process which has ended in recent years. Because of that, the data were in two very different databases (each of the two being the result of previous mergers). The policies in these two databases had to be joined carefully, in order to preserve the information as much as possible. However, the amount of incomplete, wrongly filled or ambiguous data was considerable in both databases, despite all the efforts made by the insurer. The truth is that many business clients resist to correctly update the information given to the insurer about their own fleets.

Even an apparently straightforward task like separating the portfolio into its various fleets revealed to have its part of ambiguity. At first, we distributed the vehicles among the various policies in the database. However, later we discovered that many business clients (probably the majority) had several policies for their auto insurance, many times associated to different accounts in each of the "sister" companies of Fidelidade. For example, there was a company which had a fleet of 10 vehicles, distributed among three policies of, respectively, one, four and five vehicles. Having thousands of cases like this in our database, our distribution of the vehicles among the fleets was obviously not adequate. The insurer aggregates all policies of each business client under an account number, but applying it directly to determining the fleets was also not adequate in some cases. For example, the fleet of a vehicle leasing company is in reality an aggregation of the fleets of its clients and shouldn't be treated as a single fleet. We describe our solution on page 45, under the paragraph "fleet dimension".

Also, there were situations of apparent non-sense when handling the data, like for example obtaining exposures[2] far greater than 1 (values like 1.5, 1.8 or even 2) after

[2]We consider exposure to be the time that the policy was in force, measured in years.

handling some kinds of yearly policies. Those policies were not very numerous, but raised concern because they could be the visible part of greater mistakes. It turned out that some were related to freezer-trucks or trucks with trailer, which were entitled to two entries in one of the databases: one entry for the vehicle itself and another for the trailer or the freezer, both entries with the same policy number, license plate, exposure and number of claims. After getting acquainted with the idiosyncrasies of the records of Fidelidade and its "sister" insurance companies, much of this data proved to be recoverable.

Naturally, not all missing or faulty data could be successfully recovered. Summing up, more than 10% of the observations were ruled out, among missing or faulty data, non admissible age and legal category of the vehicles.

A.4. Methodology used and further details on the description of the portfolio

With respect to the **legal category** of the vehicle, the portfolio includes trucks, buses, passenger cars, commercial cars, vans, pickups, motorcycles, bikes, animal-drawn vehicles, quadrycicles and several kinds of machinery (agricultural, industrial, etc). Although all these are "vehicles" at the eyes of the law, our definition is more strict. We consider that a vehicle has an engine, requires a driver and is mainly used for transportation of people or cargo on public roads. Hence, we did not include categories AM, DP, EP, GR, M4, MI, TA, VC and OU (each category is identified by two letters, based on the legal definition currently in force in Portugal). Categories AB and GR were also ruled out. These are special kinds of insurance binded to the driver's license of some particular kinds of employees (e.g. car sellers or car mechanics), when at work.

As for the **age** of the vehicle, we were to consider only vehicles aged up to 10 years old at first, but there was a significant number of older vehicles and so we considered

Category	Description PT	Description EN	Nr. of vehicles
AB	Automobilista	Driver	-
AM	Tracção animal	Animal drawn	-
AR	Articulado	Articulated Truck	34 535
AU	Autocarro	Bus	13 872
CC	Ciclomotor	Moped	6 546
CT	Caminheta	Van	79 280
DP	*Dumper*	Dumper	-
EP	Empilhadora	Stacker	-
GR	Garagista	Garage worker	-
LP	Ligeiro de passageiros	Passenger car	381 253
M4	Moto quatro	Quad	-
MC	Motociclo	Motorcycle	8 072
MI	Máquina industrial	Industrial machinery	-
MT	Misto	Commercial car	222 001
MV	Mono-volume	MPV (Multipurpose vehicle)	17 214
PS	Pesado	Truck	46 339
PU	*Pick-up*	Pick-up	34 454
QD	Quadriciclo	Quadricycle	536
RB	Reboque	Trailer	-
TA	Tractor agrícola	Agricultural Tractor	-
TT	Todo-o-terreno	Jeep	27 779
VC	Velocípede s/motor	Bike (no engine)	-
OU	Outros	Others	-

Table A.4.: Description of the legal categories in the portfolio and number of vehicles included in this study for each one.

vehicles aged up to 40 years old.

For the **fuel**, at first we considered three categories: “Diesel”, “Gasoline” and “Others”, the latter including mainly natural gas or electric cars. This category is by far the smallest one (only 4 000 vehicles). However, it displays a significantly different claim behavior (see Section 3.1), so we chose to keep it. Fuel was the variable with the most missing or faulty data. Information regarding fuel lacked for 34 441 vehicles. We decided to include them in a new category, “No info”, in order not to lose the information regarding the other variables for such a significant number of vehicles.

The next factor taken into account was the kind of **economic activity** practiced by its owner. Here we used the CAE code[3], adapted to our own needs. The starting

[3]Every company in Portugal is assigned a 9-digit economic activity code which is used by the public

point was to consider the CAE classification by letters (from A to U), which is an aggregation of the CAE code based on the first two digits. For detailed information on the conversion from the CAE digits to the letters, please see page 39 of (INE, 2007).

While some of the "letters" included a large number of vehicles, others contained insignificant portions of our portfolio. In the latter case, we joined contiguous letters in the alphabet. On the other hand, we considered car rental companies separately from the rest of category "N", dividing it in "Na" and "Nb". Table 1.1 in page 7 includes a rough description of the final categories, as well as vehicle numbers for each of them.

With respect to the **fleet dimension**, our criteria for fleet formation was the following, in response to the problems described in Section A.3 (page 42): After defining a single client number for each company, we listed the vehicles associated to clients with five-digit CAE code[4] identified in Table A.5. The vehicles in this list were still grouped into fleets according to their policy number, because most of them aren't really being used by the company that legally owns them. All remaining vehicles were grouped into fleets according to the client number of the company.

This partition is probably not flawless, but is certainly closer to reality than the first one we considered. It is worth pointing out that this new partition of the fleets implied, for example, that the amount of "single-vehicle fleets" decreased from more than 400 000 to little over 100 000. The results are in Table A.5.

The **geographical area** factor considered in this study has to do with the location of the company headquarters. Fidelidade developed an aggregation of the Portuguese parishes, which we do not know in full detail. However, we know that each region resulting from this aggregation is geographically contiguous with the next in the ordered list: 1, 2, 3, 4, 5 and 6. In addition, these geographical areas are, by construction, increasingly urbanized (or decreasingly rural). Area 1 is rural, with the lowest frequency of claims, while 6 represents an urban area with the highest frequency of

administration for tax purposes, statistical purposes and others. There is an aggregation of this 9-digit code to broader classes identified by letters.

[4]For further details, please see pages 181, 223, 224 and 237 of INE (2007).

CAE	Description
45110	Car sales
45190	Truck sales
64910	Leasing of vehicles
64920	Other activities of financial institutions
77110	Car rentals
77120	Truck rentals

Fleet dim.	Vehicles
Single vehicle	102 132
2-4	134 955
5-9	121 734
10-19	94 330
20-34	53 317
35-59	39 873
60-99	32 035
100-174	30 775
175-299	32 492
300+	230 238

Table A.5.: On the left, the description of the economic activities whose vehicles were grouped into fleets by policy number. On the right, the distribution of the vehicles according to the dimension of their fleet.

claims. Later, the partition was refined, originating six additional categories (with respect to the expected claim frequency), preserving the geographical continuity property and ordered from the "lowest frequency region" to the "highest frequency region": 1, A, 2, B, ..., E, 6, F.

This variable is currently used as a tariff factor only for family clients. We redesigned it, in order to be coherent when applied to the fleets portfolio. First, we created a new level for this factor, level "N", including the companies we thought were more likely to have vehicles circulating outside its geographical area:

- Companies with 5-digit CAE 45110, 45190, 64910, 64921, 77110, 77120 (described in Table A.5 on this page), 49410 (transportation of goods by road) and 49391 (long distance passenger transportation by road).

- Companies with fleets greater than 300 vehicles, excluding CAE codes 49310 (urban and suburban transportation of passengers by road) and 84113 (public local administration).

After, we distributed the remaining fleets over 4 regions, each one aggregating three of the original areas (see Table A.6). The distribution of the vehicles is also in Table A.6.

This aggregation restored the property of increasing the expected number of claims

Area	Nr. of vehicles
1	8 727
A	13 929
2	40 155
B	28 103
3	81 296
C	51 959
4	83 533
D	353 958
5	41 833
E	131 270
6	3 753
F	33 365

Region	Nr. of vehicles
I (1+A+2)	55 357
II (B+3+C)	125 327
III (4+D+5)	284 663
IV (E+6+F)	99 674
N (no region)	306 860

Table A.6.: Distribution of the vehicles according to the original areas developed by Fidelidade (on the left) and according to our own aggregation (on the right).

when moving from more rural to more urban areas, even though losing some information in the process. Nevertheless, its explanatory power is still significant.

A.5. A simple Poisson model for fleets

Consider a Poisson model with a fixed effect u common to a fleet of m vehicles. We denote N_i as the number of claims reported by vehicle i and suppose that all the vehicles in the fleet have the same *a priori* frequency risk λ. Considering that the N_i are independent, we have

$$N_i \sim P\left(\lambda u\right), \; \forall i = 1, \ldots, m \Rightarrow N = \sum_i N_i \sim P\left(m\lambda u\right).$$

Naturally, the risk associated to the vehicles belonging to the same fleet presents a certain degree of homogeneity. Parameter u represents not only the characteristics or behavior of the company which might be observable and quantified, but also the heterogeneous risk inside the fleet, which is to say, the unobservable or hidden factors. Since u also represents this residual heterogeneity of the distribution of the number of claims, it is more natural to consider a random effect U, with $\mathrm{E}\left(U\right) = 1$, $\mathrm{V}\left(U\right) = \sigma^2$.

Now suppose that we wanted to use credibility to compute the premium in this context. Let us use Bühlmann's model (Bühlmann (1967)), which we briefly summarize. For a detailed exposition, see e.g. (Klugman et al. (2008)) or (Bühlmann and Gisler (2005)). Linear credibility basically consists on estimating the random variable X_{t+1} by finding a linear function of the observations $x_1, \dots, x_t$ of the random variables $X_1, \dots, X_t$ (indexed by time periods) that approximate the hypothetical mean

$$\mu_{t+1}(u) = \mathrm{E}\left[X_{t+1} | U = u\right].$$

So we restrict the search to estimators of the form $\beta_0 + \sum_{j=1}^{t} \beta_j X_j$. We minimize the squared error loss

$$\mathrm{Q} = \mathrm{E}\left\{\left[\mu_{t+1}(U) - \beta_0 - \sum_{j=1}^{t} \beta_j X_j\right]^2\right\},$$

with respect to parameters $\beta_0, \dots, \beta_t$, obtaining

$$\mathrm{E}\left(\mu_{t+1}(U)\right) = \beta_0 + \sum_{j=1}^{t} \beta_j \mathrm{E}\left(X_j\right); \qquad \mathrm{Cov}\left(X_i, X_j\right) = \sum_{j=1}^{t} \beta_j \mathrm{Cov}\left(X_i, X_j\right). \quad \text{(A.1)}$$

These $t+1$ equations are called the *normal equations.* If the covariance matrix of the X_j's is invertible, they have a unique solution $\tilde{\beta}_0, \dots, \tilde{\beta}_t$ and the credibility premium is $\tilde{\beta}_0 + \sum_{j=1}^{t} \tilde{\beta}_j X_j$. It can be proved that the credibility premium is also the best linear estimator for the hypothetical mean $\mathrm{E}\left(X_{t+1}|U\right)$, for the Bayesian premium $\mathrm{E}\left(X_{t+1}|X_1, \dots, X_t\right)$ and for X_t. If the $X_1, \dots, X_t$ have the same mean and variance and are i.i.d. conditional on U, we are in the conditions of the Bühlmann model. Define

$$\mu = \mathrm{E}\left[\mu(u)\right]; \; v = \mathrm{E}\left[\mathrm{Var}\left(X_{t+1}|U = u\right)\right]; \; a = \mathrm{Var}\left[\mu(U)\right].$$

Straightforward computations yield: $\mathrm{E}\left(X_j\right) = \mu$; $\mathrm{Var}\left(X_j\right) = v + a$; $\mathrm{Cov}\left(X_i, X_j\right) = a$.

After solving the normal equations (A.1), we find that the credibility premium is

$$\tilde{\beta}_0 + \sum_{j=1}^{m} \tilde{\beta}_j X_j = \alpha \bar{X} + (1-\alpha)\,\mu, \qquad \alpha = \frac{t}{t+\frac{v}{a}}.$$

Considering that we have $t = 1$ (corresponding to one time period), then

$$\mu = v = \mathrm{E}\,(m\lambda U) = m\lambda,\ a = \mathrm{Var}\,(m\lambda U) = m^2\lambda^2\sigma^2.$$

In the fixed effects model, the weight $(1-\alpha)$ assigned to the fleet and the experience rating coefficient would be

$$1-\alpha = 1 - \frac{1}{1-\frac{v}{a}} = \frac{m\lambda\sigma^2}{1+m\lambda\sigma^2}; \qquad ER = (1-\alpha)\,\frac{\sum_j n_j}{m\lambda} + \alpha u.$$

The weight α increases towards one when the fleet size goes to infinity. This means that $ER \overset{n\to\infty}{\longrightarrow} u$. In the random effects model, the only difference would be the substitution of u by $\mathrm{E}\,(u) = 1$ in the experience rating coefficient. Its variance increases with fleet size.

A.6. Data relative to applying the models

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0; 0.25]$	1.419	0.01430	0.04617	0.06047	NA	NA
$]0.25; 0.5]$	4.161	0.05310	0.06585	0.1190	NA	NA
$]0.5; 1]$	7.522	0.09868	0.06752	0.1662	0.02907	0.1401
$]1; 2]$	14.27	0.1787	0.06775	0.2465	0.02307	0.1607
$]2; 4]$	26.04	0.2966	0.06625	0.3629	0.01759	0.1609
$]4; 8]$	50.97	0.4553	0.05560	0.5110	0.01087	0.1393
$]8; 16]$	98.81	0.6199	0.04221	0.6621	0.006546	0.09532
$]16; 32]$	216.4	0.7608	0.02941	0.7902	0.004570	0.06319
$]32; 64]$	451.7	0.8680	0.01435	0.8824	0.001895	0.03426
> 64	1600	0.9432	0.006998	0.95022	0.000906	0.01545

Table A.7.: Data respecting to Model 1 and Model 2 for 2007.

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0;0.25]$	1.432	0.009020	0.04637	0.05539	NA	NA
$]0.25;0.5]$	4.200	0.03462	0.06786	0.1025	NA	NA
$]0.5;1]$	7.583	0.06523	0.07141	0.1366	NA	NA
$]1;2]$	14.91	0.1227	0.07119	0.1939	0.02389	0.1713
$]2;4]$	28.20	0.2167	0.07075	0.2875	0.01915	0.1810
$]4;8]$	57.31	0.3544	0.06323	0.4176	0.01300	0.1621
$]8;16]$	104.6	0.5147	0.05338	0.5681	0.008456	0.1278
$]16;32]$	195.1	0.6699	0.03805	0.7079	0.005792	0.08844
$]32;64]$	468.9	0.8074	0.02020	0.8276	0.002806	0.04833
> 64	2352	0.9339	0.009140	0.9431	0.001169	0.01867

Table A.8.: Data respecting to Model 1 and Model 2 for 2008.

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0;0.25]$	1.5782	0.01349	0.04400	0.05749	NA	NA
$]0.25;0.5]$	4.2502	0.04668	0.06093	0.1076	NA	NA
$]0.5;1]$	7.9584	0.08740	0.06241	0.1498	NA	NA
$]1;2]$	15.21	0.1612	0.06170	0.2229	0.01903	0.1530
$]2;4]$	28.74	0.2774	0.05962	0.3370	0.01526	0.1597
$]4;8]$	59.19	0.4261	0.04978	0.4759	0.01039	0.1257
$]8;16]$	113.1	0.6016	0.03680	0.6384	0.006586	0.09637
$]16;32]$	215.1	0.7409	0.02891	0.7698	0.003883	0.06318
$]32;64]$	427.2	0.8536	0.01601	0.8697	0.002197	0.03642
> 64	1995	0.9450	0.007174	0.9522	0.001178	0.01538

Table A.9.: Data respecting to Model 1 and Model 2 for 2009.

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0;0.25]$	1.451	0.008646	0.04524	0.05388	NA	NA
$]0.25;0.5]$	4.364	0.03114	0.05693	0.08807	NA	NA
$]0.5;1]$	8.029	0.05890	0.05967	0.1186	0.02331	0.1317
$]1;2]$	15.71	0.1124	0.06101	0.1734	0.017623	0.1551
$]2;4]$	29.49	0.2025	0.06155	0.2641	0.01407	0.1594
$]4;8]$	59.59	0.3314	0.05443	0.3858	0.008343	0.1457
$]8;16]$	99.34	0.4926	0.04745	0.5401	0.006826	0.1142
$]16;32]$	175.7	0.6460	0.03697	0.6830	0.004101	0.08500
$]32;64]$	408.8	0.7963	0.02257	0.8188	0.002217	0.05360
> 64	1846	0.9165	0.01043	0.9270	0.001270	0.02391

Table A.10.: Data respecting to Model 1 and Model 2 for 2010.

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0; 0.25]$	1.528	0.01057	0.05371	0.06428	NA	NA
$]0.25; 0.5]$	4.713	0.04084	0.07162	0.1125	NA	NA
$]0.5; 1]$	9.018	0.07703	0.07433	0.1514	0.02917	0.1660
$]1; 2]$	17.87	0.1416	0.07349	0.2150	0.02232	0.1818
$]2; 4]$	35.17	0.2482	0.06867	0.3168	0.01724	0.1933
$]4; 8]$	71.80	0.3955	0.05630	0.4518	0.01266	0.1619
$]8; 16]$	132.8	0.5592	0.04248	0.6016	0.009071	0.1237
$]16; 32]$	210.3	0.7039	0.03623	0.7401	0.004755	0.09573
$]32; 64]$	499.3	0.8327	0.01638	0.8490	0.006248	0.05722
> 64	1975	0.9396	0.006999	0.9466	0.0006417	0.01809

Table A.11.: Data respecting to Model 1 and Model 2 for 2011.

λ_f	m_f	α_f	$\bar{\beta}_f$	$cred_f$	σ_1	σ_2
$]0; 0.25]$	1.568	0.006979	0.04771	0.05469	NA	NA
$]0.25; 0.5]$	4.987	0.02813	0.06559	0.09372	NA	NA
$]0.5; 1]$	8.966	0.05337	0.07069	0.1241	0.02857	0.1591
$]1; 2]$	16.93	0.1009	0.07423	0.1751	0.02233	0.1820
$]2; 4]$	31.30	0.1836	0.07685	0.2604	0.01918	0.2059
$]4; 8]$	54.82	0.3043	0.07730	0.3816	0.01394	0.1964
$]8; 16]$	102.4	0.4544	0.06056	0.5150	0.008956	0.1606
$]16; 32]$	164.0	0.6186	0.05011	0.6687	0.008238	0.1379
$]32; 64]$	387.2	0.7692	0.02749	0.7967	0.004314	0.07753
> 64	2249	0.9080	0.009373	0.9173	0.000893	0.02587

Table A.12.: Data respecting to Model 1 and Model 2 for 2012.