EXPONENTIAL BONUS-MALUS SYSTEMS INTEGRATING A PRIORI RISK CLASSIFICATION

Lluís Bermúdez
Departament de Matemàtica Econòmica i Empreserial
Facultat de Econòmiques, Universitat de Barcelona
Avda. Diagonal, 690
S-08034 Barcelona, Spain
lbermu@riscd2.eco.ub.es

Michel Denuit
Institut de Statistique, Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
denuit@stat.ucl.ac.be

Jan Dhaene
Departement Toegepaste Economische Wetenschappen
Faculteit Economische and Toegepaste Economische Wetenschappen
Katholieke Universiteit Leuven
Minderbroedersstraat, 5
B-3000 Leuven, Belgium
Jan.Dhaene@econ.kuleuven.ac.be

January 2001, Revised June 2001

Abstract

This paper examines an integrated ratemaking scheme including *a priori* risk classification and *a posteriori* experience rating. In order to avoid the high penalties implied by the quadratic loss function, the symmetry between the overcharges and the undercharges is broken by introducing parametric loss functions of exponential type.

Key words and phrases: Bonus-Malus system, quadratic loss function, exponential loss function, credibility estimation, explanatory variables, experience rating, risk classification

1 Introduction and Motivation

One of the main tasks of the actuary is to design a tariff structure that will fairly distribute the burden of claims among policyholders. If the risks in the portfolio are not all equal to each other (which means that they have different distribution functions), it is fair to partition all policies into homogeneous classes with all policyholders belonging to the same class paying the same premium. The classification variables introduced to partition risks into cells are called *a priori* variables (as their values can be determined before the policyholder starts to drive); in automobile third-party liability insurance, they commonly include the age, gender and occupation of the policyholders, the type and use of their car, the place where they reside and sometimes even the number of cars in the household, marital status, smoking behavior or the color of the vehicle. It is convenient to achieve *a priori* classification by resorting on generalized linear models; see e.g. Renshaw (1994) or Pinquet (1997,1999) for applications in actuarial sciences, and Mc Cullagh and Nelder (1989), Dobson (1990) or Fahrmeir and Tutz (1994) for a general overview of the statistical theory.

However, many important factors cannot be taken into account at this stage; think for instance of swiftness of reflexes, aggressiveness behind the wheel or knowledge of the highway code. Consequently, tariff cells are still quite heterogeneous despite of the use of many *a priori* variables. However, these hidden features are usually impossible to measure and to incorporate in a price list. But it is reasonable to believe that these characteristics are revealed by the number and sizes of claims reported by the policyholders over the successive insurance periods. Hence the adjustment of the premium based on the individual claims experience in order to restore fairness among policyholders.

It is interesting to mention that in North America, emphasis has traditionally been laid on *a priori* ratings using many classifying variables whereas in continental Europe just a few *a priori* variables were chosen and much importance was placed on the *a posteriori* evaluation of drivers. However, European directives have introduced complete rating freedom since July 1994. Insurance companies operating in EU countries are now (theoretically) free to set up their own rates, select their own classification variables and design their own Bonus-Malus system. In most European countries, companies have taken advantage of this freedom by introducing more rating variables. Indeed, in a competitive environment, the trend is towards a portfolio segmentation spiral: because of commercial pressure, insurers tend to make use of all available relevant information to match the premium as closely as the rating structure used by competitors. Since the only item of interest is in fact the distribution function of the claim amount produced by the driver during the year, which is impossible to measure, it seems fair to correct the inadequacies of the *a priori* system by using an adequate Bonus-Malus system. Such an experience rating system should be better accepted by policyholders than arbitrary *a priori* classifications.

Rating systems that penalize insureds responsible for one or more accidents by premium surcharges (or *maluses*), and that reward claim-free policyholders by awarding them discounts (or *bonuses*) are now in force in many developed countries. This *a posteriori* ratemaking is a very efficient way of classifying policyholders into cells according to their risk. As pointed out by Lemaire (1995), if insurers were allowed to use only one rating variable, it should be some form of merit rating: the best predictor of the number of claims incurred by a driver in the future is not age or car but past claims behavior. Besides encouraging policyholders to drive carefully (i.e. counteracting moral hazard), they aim to better assess individual risks, so that everyone will pay in the long run a premium corresponding to his own claim frequency. Such systems are called no-claim discounts, experience rating, merit rating, or Bonus-Malus systems. We will adopt here the latter terminology. For a thorough presentation of the techniques relating to Bonus-Malus systems, we refer the interested reader to Lemaire (1995).

The discussion in the remainder of this section is inspired from the paper by De Wit and Van Eeghen (1984). Consider a portfolio of n policies from automobile third-party liability insurance. The random variable Y models a quantity of actuarial interest for a policy taken at random from the portfolio (for instance the amount of a claim, the aggregate claims in one period, or the number of accidents at fault reported by the policyholder during one period). In order to explain the outcomes of Y, the actuary has observable covariates $\mathbf{X} = \{X_1, X_2, \dots\}$ at his disposal (e.g. age, gender and occupation of the policyholder, the place where he resides, type and use of his car, or even his marital status, smoking behaviour or color of the car). However, Y also depends on a sequence of unknown characteristics $\mathbf{Z} = \{Z_1, Z_2, \dots\}$ (e.g. annual mileage (that is, risk exposure), accuracy of judgment, aggressiveness behind the wheel, drinking behaviour, etc.). Some of these quantities are unobservable, others cannot be measured in a cost efficient way. The risk factors for this policyholder are

$$\mathbf{\Omega} = \mathbf{X} \cup \mathbf{Z}.$$

The "true" premium for this policyholder is $\mathbb{E}[Y|\mathbf{\Omega}]$; it is worth mentioning that this premium is a random variable but much less dispersed than Y itself, making insurance policies worth to be bought. The situation can be summarized as described in Table 1.1. In this case, the policyholder keeps the variations of the premiums due to the modifications in his personal characteristics $\mathbf{\Omega}$ and transfers to the company the purely random fluctuations of Y (that is, the variance of the outcomes of Y once the personal characteristics $\mathbf{X}$ and $\mathbf{Z}$ have been taken into account). As mentioned in Bowers et al. (1997), "In a competitive economy, market forces will encourage insurers to price short-term policies so that deviations from expected value will behave as independent random variables. Deviations should exhibit no pattern that might be exploited by the insured or insurer to produce consistent gains. Such consistent deviations would indicate inefficiencies in the insurance market". Consequently, the trend is towards a premium amount $\mathbb{E}[Y|\mathbf{\Omega}]$.

	Carried by the policyholder	Carried by Insurer
Risk	$\mathbb{E}[Y\|\mathbf{\Omega}]$	$Y - \mathbb{E}[Y\|\mathbf{\Omega}]$
Expectation	$\mathbb{E}[Y]$	0
Variance	$\mathbb{V}\text{ar}\Big\{\mathbb{E}[Y\|\mathbf{\Omega}]\Big\}$	$\mathbb{E}\Big\{\mathbb{V}\text{ar}[Y\|\mathbf{\Omega}]\Big\}$

Table 1.1: Risk transfer between insurance company and policyholder in case of full information.

Of course, since the elements of $\mathbf{Z}$ are unknown to the insurer, the situation described in Table 1.1 is purely theoretical. Since the company only knows $\mathbf{X}$, the reality of insurance business is rather as depicted in Table 1.2. Now, let us write

$$\begin{aligned}\mathbb{V}\mathrm{ar}[Y|\mathbf{X}] &= \mathbb{E}[Y^2|\mathbf{X}] - \Big\{\mathbb{E}[Y|\mathbf{X}]\Big\}^2 \\ &= \mathbb{E}\Big[\mathbb{E}[Y^2|\mathbf{\Omega}]\Big|\mathbf{X}\Big] - \Big\{\mathbb{E}\Big[\mathbb{E}[Y|\mathbf{\Omega}]\Big|\mathbf{X}\Big]\Big\}^2 \\ &= \mathbb{E}\Big[\mathbb{V}\mathrm{ar}[Y|\mathbf{\Omega}]\Big|\mathbf{X}\Big] + \mathbb{V}\mathrm{ar}\Big[\mathbb{E}\Big[\mathbb{E}[Y|\mathbf{\Omega}]\Big|\mathbf{X}\Big]\Big],\end{aligned}$$

so that

$$\mathbb{E}\Big\{\mathbb{V}\mathrm{ar}[Y|\mathbf{X}]\Big\} = \mathbb{E}\Big\{\mathbb{V}\mathrm{ar}[Y|\mathbf{\Omega}]\Big\} + \mathbb{E}\Big\{\mathbb{V}\mathrm{ar}\Big[\mathbb{E}\Big[\mathbb{E}[Y|\mathbf{\Omega}]\Big|\mathbf{X}\Big]\Big\}.$$

The first term in the latter sum, i.e. $\mathbb{E}\{\mathbb{V}\mathrm{ar}[Y|\mathbf{\Omega}]\}$, represents the purely random fluctuations of the risk and is supported by the insurance company in virtue of the very basic principle of insurance. On the contrary, the second term represents the variations of the expected claims due to the unknown risk characteristics $\mathbf{Z}$. This quantity should be corrected by an experience rating mechanism.

	Carried by the policyholder	Carried by Insurer
Risk	$\mathbb{E}[Y|\mathbf{X}]$	$Y - \mathbb{E}[Y|\mathbf{X}]$
Expectation	$\mathbb{E}[Y]$	0
Variance	$\mathbb{V}\mathrm{ar}\Big\{\mathbb{E}[Y|\mathbf{X}]\Big\}$	$\mathbb{E}\Big\{\mathbb{V}\mathrm{ar}[Y|\mathbf{X}]\Big\}$

Table 1.2: Risk transfer between insurance company and policyholder in case of partial information.

Now, assume the insurance company incorporates more *a priori* variables in its tariff structure, that is, $\widetilde{\mathbf{X}}$ is substituted for $\mathbf{X}$ with $\mathbf{X} \subset \widetilde{\mathbf{X}}$; then

$$\mathbb{E}\Big\{\mathbb{V}\mathrm{ar}\Big[\mathbb{E}[Y|\mathbf{\Omega}]\Big|\widetilde{\mathbf{X}}\Big]\Big\} \leq \mathbb{E}\Big\{\mathbb{V}\mathrm{ar}\Big[\mathbb{E}[Y|\mathbf{\Omega}]\Big|\mathbf{X}\Big]\Big\},$$

that is, the residual heterogeneity in the portfolio is reduced. Consequently, the variance of the insurer's result is also reduced, i.e.

$$\mathbb{E}\Big\{\mathbb{V}\mathrm{ar}[Y|\widetilde{\mathbf{X}}]\Big\} \leq \mathbb{E}\Big\{\mathbb{V}\mathrm{ar}[Y|\mathbf{X}]\Big\}.$$

The severity of the *a posteriori* corrections thus decreases as the information used by the insurer increases.

Now, the idea behind experience rating is that past claims experience reveals the hidden features $\mathbf{Z}$. Let $\mathbf{Y}^{\leftarrow}$ denotes the past claims experience available about Y. The idea is that the information contained in $(\mathbf{X}, \mathbf{Y}^{\leftarrow})$ becomes comparable to $\mathbf{\Omega}$ as time goes on. Therefore, the *a posteriori* premium is

$$\mathbb{E}[Y|\mathbf{X}, \mathbf{Y}^{\leftarrow}].$$

Experience rating is based on the following mechanism:

1. claim-free policyholders are rewarded by premium discounts called *bonuses*;

2. policyholders reporting one or more accidents at fault are penalized by premium surcharges called *maluses*.

The very aim of this paper is to examine the interaction between *a priori* ratemaking (i.e. identification of the best predictors $\mathbf{X}$ and of the risk premium $\mathbb{E}[Y|\mathbf{X}]$) and *a posteriori* ratemaking (i.e. premium corrections according to past claims history $\mathbf{Y}^{\leftarrow}$ in order to reflect the unavalaible information contained in $\mathbf{Z}$).

The paper is organized as follows. In Section 2, we briefly review the current methodology of automobile ratemaking (at least in European countries), considering risk classification and credibility as two different problems. This approach has flaws, as it will be demonstrated further. The main reason is that experience rating aims to reduce the residual heterogeneity of the portfolio, which obviously depends on the degree of *a priori* segmentation. Therefore, *a priori* and *a posteriori* ratemaking have to be integrated in a continuous risk evaluation mechanism. In Section 3, we present the results of Dionne and Vanasse (1989, 1992) and Gisler (1996), as well as an alternative approach based on an exponential loss function. Such loss functions have been considered by Ferreira (1977), Lemaire (1979), Young (1996) and Denuit and Dhaene (2001), among others. All the methods examined in the present paper are illustrated on the basis of a pedagogical example using a Spanish insurance portfolio. This example takes only two risk factors into account and allows a deep understanding of all the technical mechanisms. Adaptations of the methodology to real-life portfolio is then straightforward. Several optimization programs will be extensively used throughout this paper. Some of them are standard in actuarial sciences, others are less common. The reader will find in an Appendix a description of all these results, together with proofs for the sake of completeness.

2 Current Methodology

2.1 The model

Consider an automobile portfolio consisting of n policies. The amount of premium paid by the policyholder depends on the rating factors of the current period (think for instance of the type of the car or of the occupation of the policyholder) but also on the claim history. The insurance premium is the product of a base premium and of a bonus-malus coefficient. The base premium is a function of the current rating factors whereas the bonus-malus coefficient only depends on the history of reported claims at fault.

The usual methodology can be described as follows. First, the insurance company achieves risk classification using generalized linear models (Poisson or logistic regressions, for instance), as explained e.g. in Renshaw (1994). This yields a partition of the portfolio in disjoint risk classes RC_1, RC_2, …, RC_K. In each risk class, the policies are identical from the company point of view, whereas policies in different risk classes have distinct risk profiles. A base premium BP_i is set for each risk class RC_i; the amount BP_i is charged to a new policyholder entering in RC_i.

Of course, inside each risk class, the policies are not identical *stricto sensu*. Therefore, the premium is adjusted over time using a Bonus-Malus factor $BMF(k,t)$ depending on the number of years t the policy is in force and on the number of claims k reported during this period; this factor is computed from a mixed Poisson distribution (usually the Negative Binomial) estimated over the entire portfolio.

The base premium varies from a risk class to another and the same Bonus-Malus factor is applied to all drivers. Methodologically, this approach is of course erroneous. Indeed, the Bonus-Malus system has to correct the amounts of premium for the residual heterogeneity existing in the different risk classes. Therefore, the severity of a Bonus-Malus system depends on the risk class occupied by policyholders. Moreover, a company using many *a priori* risk factors will have to apply softer Bonus-Malus coefficients than a company resorting on just a few risk factors. Uniform Bonus-Malus systems imposed by regulatory authorities (as it is the case in Belgium or in France) creates cross-subzidiation in insurance portfolios.

In order to determine the Bonus-Malus coefficients, the ith policy of the portfolio, $i = 1,2,\ldots,n$, is represented by a sequence $(\Theta_i, K_{i1}, K_{i2}, K_{i3}, \ldots)$ where K_{ij}, $j = 1,2,\ldots$, represents the number of claims incurred by this policyholder during the jth year the policy is in force, i.e. during the period $(j-1,j)$. Let

$$K_{i\bullet}(t) = \sum_{j=1}^{t} K_{ij}$$

be the number of claims reported during the first t years of insurance.

At the portfolio level, the sequences $(\Theta_i, K_{i1}, K_{i2}, K_{i3}, \ldots)$ are assumed to be independent and identically distributed for $i = 1,2,\ldots,n$. The risk parameter Θ_i represents the risk proneness of policyholder i, i.e. unknown risk characteristics of the policyholder having a significant impact on the occurrence of claims; it is regarded as a random variable. Given $\Theta_i = \theta$, the random variables $K_{i1}, K_{i2}, K_{i3}, \ldots$ are assumed to be independent and identically distributed. Unconditionally, these random variables are dependent.

Let us denote as Z_{ijk}, $k = 1,2,\ldots,K_{ij}$ the amounts of the K_{ij} claims reported by the ith policyholder during the jth year provided policyholder i has been involved in at least one claim (i.e. $K_{ij} \geq 1$). The total claim amount for this risk in year j is

$$S_{ij} = \sum_{k=1}^{K_{ij}} Z_{ijk}.$$

The severities Z_{ijk}, $i = 1,2,\ldots,n$, $j,k \in \mathbb{N}_0$, are assumed to be independent and identically distributed, and independent of the claim frequencies K_{ij}, $j \in \mathbb{N}_0$; it is worth mentioning that this assumption has been questioned by several authors. It essentially states that the cost of an accident is for the most part beyond the control of a policyholder. The degree of care exercised by a driver mostly influences the number of accidents, but in a much lesser way the cost of these accidents. Nevertheless, this assumption seems acceptable in third-party liability insurance, e.g. because the payments of the insurance company are determined by third-party characteristics. Note that the severities $\{Z_{ijk},\ j,k \in \mathbb{N}_0\}$ are also independent of Θ_i

We put $\mathbb{E}Z_{ijk} \equiv 1$, which means that the expected claim amount is chosen as monetary unit. The pure premium for policy i in year j is then given by

$$\mathbb{E}[S_{ij}|\Theta_i = \theta] = \mathbb{E}[K_{ij}|\Theta_i = \theta] = \theta;$$

A priori (i.e. without information about claims history), an identical amount of premium $\mathbb{E}\Theta_i$ is charged to new policyholders.

The annual numbers of claims $[K_{i1}|\Theta_i = \theta], [K_{i2}|\Theta_i = \theta], \ldots$ of the policyholder i are assumed to be independent and to conform to a Poisson distribution with mean θ, i.e.

$$\mathbb{P}[K_{ij} = k|\Theta_i = \theta] = \exp(-\theta)\frac{\theta^k}{k!}, \;\; k \in \mathbb{N}, j \in \mathbb{N}_0;$$

θ is the claim frequency of this policyholder and is assumed to be constant over time. Usually, the common cumulative distribution function F_Θ of the Θ_i's, often called the structure function, belongs to the two-parameter Gamma family, i.e.

$$dF_\Theta(\theta) \equiv d\Gamma(\theta|a,\tau) = \frac{\tau^a \exp(-\tau\theta)\theta^{a-1}}{\Gamma(a)}d\theta, \;\; a, \tau > 0, \;\; \theta \in \mathbb{R}^+. \tag{2.1}$$

Henceforth, $\Gamma(.|a,\tau)$ denotes the cumulative distribution function associated to the Gamma distribution with mean a/τ and variance a/τ^2. Under (2.1), it is well-known that the number of claims for a policyholder randomly drawn from the portfolio follows a Negative Binomial distribution, i.e. for any $i = 1, 2, \ldots, n$,

$$\mathbb{P}[K_{ij} = k] = \begin{pmatrix} k+a-1 \\ k \end{pmatrix} \left(\frac{\tau}{1+\tau}\right)^a \left(\frac{1}{1+\tau}\right)^k, \;\; k \in \mathbb{N}, j \in \mathbb{N}_0;$$

the K_{ij}'s are thus identically distributed (but not independent, since they are generated by the same policyholder, and thus contingent on the same risk parameter Θ_i).

2.2 *A posteriori* premiums

Now, suppose that policy i has been observed for t years and that $k_{i\bullet}(t)$ claims have been reported during this period. Classically, the premium for year $t+1$ is defined as a function of the claims reported during the years $1, 2, \ldots, t$, $\Psi(k_{i1}, k_{i2}, \ldots, k_{it})$, say; the function Ψ is determined by minimizing $\mathbb{E}L\{\theta_i - \Psi(k_{i1}, k_{i2}, \ldots, k_{it})\}$ for some loss function L, taken to be non-negative, convex and such that $L(0) = 0$. The losses considered in this paper are the standard quadratic loss where $L(x) = x^2$ and the exponential loss with positive parameter c where $L(x) = \exp(-cx)$. From the results recalled in Appendix, we easily get the following results.

Proposition 2.1. (i) *Under a quadratic loss function, the best estimator of the pure premium Θ_i under (2.1) is given by*

$$W_{t+1}^{quad} = \frac{a}{\tau}(1-\rho_{quad}) + \frac{k_{i\bullet}(t)}{t}\rho_{quad} \;\textit{with}\; \rho_{quad} = \frac{t}{\tau+t}.$$

(ii) *Under an exponential loss function with parameter $c > 0$, the best estimator of the pure premium Θ_i under (2.1) is given by*

$$W_{t+1}^{exp} = \frac{a}{\tau}\big(1 - \rho_{exp}(c)\big) + \frac{k_{i\bullet}(t)}{t}\rho_{exp}(c) \textit{ with } \rho_{exp}(c) = \frac{t}{c}\ln\left(1 + \frac{c}{\tau + t}\right),$$

Proof. It suffices to invoke Proposition 4.1, noticing that in the Poisson-Gamma model, *a posteriori* structure functions are still Gamma with updated parameters, i.e.

$$F_\Theta(.|K_{i1} = k_1, K_{i2} = k_2, \ldots, K_{it} = k_t) = \Gamma(.|a + k_{i\bullet}(t), \tau + t).$$

□

Considering Proposition 2.1, W_{t+1}^{quad} is a convex combination of the portfolio mean a/τ and the observed average number of claims $k_{i\bullet}(t)/t$ over the period $[0, t]$. The weight ρ_{quad} given to the past claims tends to 1 as the length t of the observation period grows to $+\infty$. Similarly, W_{t+1}^{exp} is a convex combination of the theoretical mean a/τ and the average number of claims $k_{i\bullet}(t)/t$ per year over the period $[0, t]$. The weight $\rho_{exp}(c)$ given to claim history with the exponential loss function is smaller than that with a quadratic loss function; indeed,

$$\rho_{exp}(c) = \frac{t}{c}\ln\left(1 + \frac{c}{\tau + t}\right) \leq \frac{t}{\tau + t} = \rho_{quad}.$$

The credibility factor being smaller, more weight is put on the overall mean a/τ and the *a posteriori* premiums are less variable.

It is worth mentioning that in the Poisson-Gamma model, the Bayesian approach coincides with the linear credibility estimator. In other words, Proposition 2.1 can be interpreted in a semiparametric framework, as in the classical Bühlmann-Straub approach.

Remark 2.2. Let us examine the consequences of a variation of the parameter c involved. Letting c tend to 0 yields

$$\lim_{c \to 0} W_{t+1}^{exp} = W_{t+1}^{quad};$$

when c tends to 0, we thus find the *a posteriori* premium associated with the quadratic loss.

Let us examine the limiting case $c \to +\infty$: we have that

$$\lim_{c \to +\infty} \rho_{exp}(c) = 0 \text{ so that } W_{t+1}^{exp} \to a/\tau.$$

This provides an intuitive meaning of the parameter c: if c increases, then the *a posteriori* merit-rating scheme becomes less severe, and at the limit, the premium no more depends on the incurred claims. Moreover, routine calculations show that

$$\frac{d}{dc}\rho_{exp}(c) < 0,$$

so that the weight given to the observed average claim number decreases as c increases. The credibility factor $\rho_{exp}(c)$ decreases from ρ_{quad} when $c = 0$ to 0 as $c \to +\infty$.

Let $I_i(t)$ denote the index of the risk class occupied by policyholder i during year t. Now, the classical *a posteriori* premium for year $t+1$ charged to policyholder i having reported $k_{i\bullet}(t)$ claims during the first t years is given by

$$P_{t+1}^{quad}(k_{i\bullet}(t),t) = BP_{I_i(t+1)}BMF^{quad}(k_{i\bullet}(t),t) \text{ with}$$

$$BMF^{quad}(k_{i\bullet}(t),t) = \frac{W_{t+1}^{quad}}{\mathbb{E}\Theta_i} = \frac{a+k_{i\bullet}(t)}{\tau+t} \times \frac{\tau}{a} \quad (2.2)$$

under a quadratic loss. Under an exponential loss, we get

$$P_{t+1}^{exp}(k_{i\bullet}(t),t) = BP_{I_i(t+1)}BMF^{exp}(k_{i\bullet}(t),t) \text{ with}$$

$$BMF^{exp}(k_{i\bullet}(t),t) = \frac{W_{t+1}^{exp}}{\mathbb{E}\Theta_i} = 1 - \frac{t}{c}\ln\left(1+\frac{c}{\tau+t}\right) + \ln\left(1+\frac{c}{\tau+t}\right)\frac{k_{i\bullet}(t)}{c}\frac{\tau}{a}. \quad (2.3)$$

These premiums appear thus as the product between a base premium $BP_{I_i(t+1)}$ depending on the personal characteristics of policyholder i at time $t+1$ and a Bonus-Malus coefficient deduced from Proposition 2.1 (i.e. disregarding the partition of the portfolio in different risk classes). The model used to determine the Bonus-Malus coefficients indeed assumes that all the risks of the portfolio have the same *a priori* claim frequency and that the differences in the claim frequency between the risks are only due to the individual risk characteristics Θ_i. Hence, the model implicitly assumes that the tariff takes into account differences in claim frequencies only by the Bonus-Malus and that such differences are not reflected, not even to some extent, in the base premiums.

Now, let us briefly explain the reason why this approach is erroneous. The aim of the Bonus-Malus system is to adjust the amount of premium according to past claim experience, in order to reduce the residual heterogeneity in the different risk classes of the portfolio. Since Bonus-Malus coefficients of Proposition 2.1 do not take into account explanatory variables, they are function of the total heterogeneity of the portfolio, before tariff segmentation. In other words, the Bonus-Malus factors penalize once again bad risks and reward once again good risks.

2.3 Numerical illustration

In this section, we exemplify the traditional way of calculating credibility premiums. The results of this section will be compared to those of Section 3.4. Let us consider the following example involving data from a Spanish insurance company. As it can be seen from Table 2.1, policies have been categorized according to the age of the driver (three classes, namely "less than 35 years", "between 36 and 49 years" and "more than 50 years") and the power of the car (four classes, namely "less than 53 hp", "between 54 hp and 75 hp", "between 76 and 118 hp" and "more than 119 hp"). Each of the 12 cells in Table 2.1 gives the observed mean claim frequency for the risk classes RC_1, RC_2, …, RC_{12}. The complete structure of the portfolio is described in Tables 2.2 and 2.3. One can find there the observed claim distribution in each of the 12 risk classes. The symbols used are as follows: n_{ik} represents the number of policies reporting k claims in RC_i, $k=0,1,2,\dots,k_{\max}^{(i)}$, $i=1,2,\dots,12$, and

$$n_{i\bullet} = \sum_{k=0}^{k_{\max}^{(i)}} n_{ik}$$

is the number of policies in RC_i, $i = 1, 2, \ldots, 12$.

Power	Age ≤ 35	Age $35 < . \leq 49$	Age ≥ 50
≤ 53	0.1866	0.1572	0.1283
$54 \leq . \leq 75$	0.2685	0.2279	0.1986
$76 \leq . \leq 118$	0.2992	0.2526	0.2386
≥ 119	0.3217	0.2846	0.2483

Table 2.1: Observed mean claim frequencies according the classification factors Age and Power.

We assume that the total number of claims T_{ij} reported by policyholder j in RC_i during one period follows a Poisson distribution with mean λ_i. Moreover, the random variables $T_{i1}, T_{i2}, \ldots$ are assumed to be independent. Therefore, the total number of claims $T_{i\bullet} = \sum_{j=1}^{n_{i\bullet}} T_{ij}$ reported by the $n_{i\bullet}$ policyholders in RC_i conform to a Poisson distribution with mean $n_{i\bullet}\lambda_i$. The realisation of $T_{i\bullet}$ is $t_{i\bullet} = \sum_{k=1}^{k_{\max}^{(i)}} kn_{ik}$. Let us introduce the indicator variables J_2 and J_3 such that

$$J_k = \begin{cases} 1 \text{ if the policyholder is in Age category } k \\ 0 \text{ otherwise} \end{cases}$$

$k = 2, 3$. Similarly, let us define L_2, L_3 and L_4 as

$$L_k = \begin{cases} 1 \text{ if the policyholder drives a car in category } k \\ 0 \text{ otherwise} \end{cases}$$

$k = 2, 3, 4$. Each individual is represented by a vector

$$\mathbf{X}_i = (1, j_2, j_3, \ell_2, \ell_3, \ell_4);$$

the corresponding vector of parameters is

$$\boldsymbol{\eta} = (\epsilon, \gamma_2, \gamma_3, \delta_2, \delta_3, \delta_4).$$

When the counts are small, which is typically the case in automobile insurance, the normal approximation is poor and fails to account for the discreteness of the data. Normal regression should be avoided in this case. Generalized linear models provide an appropriate framework for the analysis of count data. A linear model for the logarithm of the λ_i's is often used in actuarial science (see e.g. Pinquet (1997)). This provides a regression model for count data analogous to the usual normal regression for continuous data According to standard methodology of generalized linear models, the logarithmic function is also the natural link for the Poisson distribution (see e.g. Dobson (1990)). We specify our model by a relation of the form

$$\ln \lambda_i + \ln n_{i\bullet} = \boldsymbol{\eta}^t \mathbf{X}_i = \epsilon + \sum_{k=2}^{3} \gamma_k J_k + \sum_{k=2}^{4} \delta_k L_k. \tag{2.4}$$

	≤ 35		$35 < . \leq 49$		≥ 50	
≤ 53	k	n_{1k}	k	n_{2k}	k	n_{3k}
	0	3,316	0	7,797	0	10,437
	1	548	1	1,063	1	1,159
	2	61	2	140	2	143
	3	15	3	17	3	15
	4	4	4	6	4	2
	5	1	≥ 5	0	5	1
	≥ 6				6	1
					≥ 7	0
	$n_{1\bullet} =$	3,945	$n_{2\bullet} =$	9,023	$n_{3\bullet} =$	11,758
	$t_{1\bullet} =$	736	$t_{2\bullet} =$	1,418	$t_{3\bullet} =$	1,509
	$\overline{x}_1 =$	0.1866	$\overline{x}_2 =$	0.1751	$\overline{x}_3 =$	0.1283
	$s_1^2 =$	0.227	$s_2^2 =$	0.1828	$s_3^2 =$	0.1501
$54 \leq . \leq 75$	k	n_{4k}	k	n_{5k}	k	n_{6k}
	0	9,470	0	21,031	0	22,788
	1	1,916	1	3,775	1	3,766
	2	445	2	720	2	591
	3	84	3	143	3	109
	4	21	4	36	4	24
	5	7	5	11	5	5
	6	0	6	2	6	4
	7	1	7	1	≥ 7	0
	8	3	≥ 8	0		
	≥ 9	0				
	$n_{4\bullet} =$	11,947	$n_{5\bullet} =$	25,719	$n_{6\bullet} =$	27,287
	$t_{4\bullet} =$	3,208	$t_{5\bullet} =$	5,862	$t_{6\bullet} =$	5,420
	$\overline{x}_4 =$	0.2685	$\overline{x}_5 =$	0.2279	$\overline{x}_6 =$	0.1986
	$s_4^2 =$	0.3635	$s_5^2 =$	0.2946	$s_6^2 =$	0.2451

Table 2.2: Observed claims distributions in the risk classes.

	≤ 35		$35 < . \leq 49$		≥ 50	
$76 \leq . \leq 118$	k	n_{7k}	k	n_{8k}	k	n_{9k}
	0	6,570	0	15,702	0	15,158
	1	1,423	1	3,112	1	2,848
	2	321	2	603	2	510
	3	89	3	148	3	123
	4	33	4	31	4	33
	5	6	5	11	5	11
	6	3	6	2	6	1
	7	1	≥ 7	0	7	3
	8	1		0	8	1
	≥ 9	0			≥ 9	0
	$n_{7\bullet} =$	8,447	$n_{8\bullet} =$	19,609	$n_{9\bullet} =$	18,688
	$t_{7\bullet} =$	2,527	$t_{8\bullet} =$	4,953	$t_{9\bullet} =$	4,459
	$\overline{x}_7 =$	0.2992	$\overline{x}_8 =$	0.2526	$\overline{x}_9 =$	0.2386
	$s_7^2 =$	0.4322	$s_8^2 =$	0.3288	$s_9^2 =$	0.3200
≥ 119	k	$n_{10;k}$	k	$n_{11;k}$	k	$n_{12;k}$
	0	1,125	0	4,554	0	4,680
	1	274	1	902	1	900
	2	69	2	224	2	187
	3	9	3	55	3	25
	4	7	4	15	4	12
	5	1	5	9	5	5
	6	1	6	2	6	1
	≥ 7	0	7	0	7	1
			8	1	8	1
			≥ 9	0	≥ 9	0
	$n_{10\bullet} =$	1,486	$n_{11\bullet} =$	5,762	$n_{12\bullet} =$	5,812
	$t_{10\bullet} =$	478	$t_{11\bullet} =$	1,640	$t_{12\bullet} =$	1,443
	$\overline{x}_{10} =$	0.3217	$\overline{x}_{11} =$	0.2846	$\overline{x}_{12} =$	0.2483
	$s_{10}^2 =$	0.4376	$s_{11}^2 =$	0.4214	$s_{12}^2 =$	0.3408

Table 2.3: Observed claims distributions in the risk classes.

In order to determine the maximum likelihood estimator of the paramete $\boldsymbol{\eta}$, we have to maximize

$$L(\boldsymbol{\eta}) = \prod_{i=1}^{12} \exp(-\lambda_i n_{i\bullet}) \frac{(\lambda_i n_{i\bullet})^{t_{i\bullet}}}{t_{i\bullet}!}.$$

The regularity conditions satisfied by the Poisson distribution ensure that the global maximum of the log-Likelihood function $\ln L$ is given uniquely by the solutions of $\partial \ln L / \partial \boldsymbol{\eta} = 0$. It is easy to check that the maximum likelihood estimator $\widehat{\boldsymbol{\eta}}$ of the parameter $\boldsymbol{\eta}$ are the solutions of

$$\sum_{i=1}^{12} (T_{i\bullet} - n_{i\bullet}\lambda_i) X_{ij} = 0, \;\; j = 1, 2, \ldots, 6,$$

where X_{ij} is the jth component of $\mathbf{X}_i$. As pointed out by Pinquet (1997), this can be interpreted as an orthogonality relation between the residuals and the covariates. Since the rating factors have a finite number of levels and the explanatory variables are indicators of these levels, this equation means that, for every sub-portfolio corresponding to a given level, the sum of the premiums is equal to the total number of claims. Consequently, such a system has the financial stability property. We finally get the results displayed in Table 2.4.

Parameter $\boldsymbol{\eta}$	Estimate $\widehat{\boldsymbol{\eta}}$	Standard deviation	Confidence interval 95%
ϵ	-1.7219	0.0198	$[-1.7607; -1.6831]$
γ_2	-0.1634	0.0147	$[-0.1922; -0.1345]$
γ_3	-0.2800	0.0149	$[-0.3093; -0.2508]$
δ_2	0.3987	0.0185	$[0.3625; 0.4350]$
δ_3	0.5324	0.0189	$[0.4953; 0.5694]$
δ_4	0.6150	0.0236	$[0.5688; 0.6611]$

Table 2.4: Estimation of the paramters in (2.4).

The vector $\widehat{\boldsymbol{\eta}}$ is approximately gaussian for large sample sizes, with mean $\boldsymbol{\eta}$ and variance-covariance matrix the inverse of the information matrix $\mathcal{I}$. Let us recall that the element (j, k) of $\mathcal{I}$ is

$$\mathcal{I}_{jk} = \sum_{i=1}^{12} X_{ij} X_{ik} n_{i\bullet} \lambda_i.$$

Computing the variance-covariance matrix yields

$$\widehat{\mathcal{I}}^{-1} = \begin{pmatrix} 0.000392 & -0.000144 & -0.000151 & -0.000277 & -0.000277 & -0.000265 \\ -0.000144 & 0.000217 & 0.000145 & 0.000002 & 0.000000 & -0.000014 \\ -0.000151 & 0.000145 & 0.000223 & 0.000008 & 0.000009 & -0.000006 \\ -0.000277 & 0.000002 & 0.000008 & 0.000342 & 0.000274 & 0.000273 \\ -0.000277 & 0.000000 & 0.000009 & 0.000274 & 0.000357 & 0.000273 \\ -0.000265 & -0.000014 & -0.000006 & 0.000273 & 0.000273 & 0.000555 \end{pmatrix}.$$

Considering Table 2.4 all the parameters are significantly different from 0 (since no confidence interval overlap 0), so that all the covariates are statistically significant. The expected claim numbers for each of the 12 cells are given in Table 2.5 (it is interesting to compare the fitted results to their empirical counterparts given in Table 2.1). Table 2.5 thus gives the base premiums attached to each of the 12 risk classes.

	Age		
Power	≤ 35	$35 < . \leq 49$	≥ 50
≤ 53	0.1787	0.1518	0.1351
$54 \leq . \leq 75$	0.2663	0.2262	0.2013
$76 \leq . \leq 118$	0.3044	0.2585	0.2300
≥ 119	0.3306	0.2808	0.2498

Table 2.5: Expected mean claim frequencies according the classification factors Age and Power in the model 2.4.

In order to get the Bonus-Malus factors, let us consider the claim distribution corresponding to the whole portfolio; it is given in Table 2.6. The Negative Binomial fit using Maximum Likelihood is displayed in the third column. The *a posteriori* premiums are then given by (2.2) and (2.3) with the estimated values of a and τ given by $\widehat{a} = 0.8665$ and $\widehat{\tau} = 3.9097$.

k	n_k	Neg. Bin. fit
0	122,628	122,713
1	21,686	21,656
2	4,014	4,116
3	832	801
4	224	158
5	68	31
6	17	6
7	7	1
8	7	0
≥ 9	0	0

Table 2.6: Observed claim distribution corresponding to the data in Table 2.1 together with a fit to the Negative Binomial distribution with parameters $\widehat{a} = 0.8665$ and $\widehat{\tau} = 3.9097$.

Consider for instance a 30-year-old driver whose car is in the category "≤ 53". His *a priori* expected number of accidents is 0.1787 for the first 5 periods. In period 5, he reaches 35 years old, and his expected number of accidents becomes 0.1518. In the first half of Table 2.7, one can see the bonus-malus coefficients and premiums for that individual. The second column (entitled "BP_t") represents the expected number of accidents (i.e. the base premium) for each period. The third column (entitled "BMF") represents the bonus-malus factor in case the policyholder does not cause any claims during $(0, t)$ computed on the basis of (2.2). Column 4 gives the total corresponding premium (product of elements of 2nd and 3d column). The two other blocks are analogous, for policyholders having reported 1 or

2 claims during this period. Consider now a 30-year-old policyholder driving a car in the category "$\geq$ 119". His expected claim frequency for the first five periods is 0.3306, and 0.2808 after. The second part of Table 2.7 shows the evolution of the premium amounts for this individual. The bonus-malus factors are identical in the two tables but the premiums differ substantially.

Car in category "$\leq$ 53"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.1787	0.7963	0.1423	1.7154	0.3065	2.6344	0.4708
2	0.1787	0.6616	0.1182	1.4251	0.2547	2.1887	0.3911
3	0.1787	0.5658	0.1011	1.2189	0.2178	1.8719	0.3345
4	0.1787	0.4943	0.0883	1.0648	0.1903	1.6352	0.2922
5	0.1787	0.4388	0.0784	0.9453	0.1689	1.4517	0.2594
6	0.1518	0.3945	0.0599	0.8499	0.1290	1.3052	0.1981
7	0.1518	0.3584	0.0544	0.7720	0.1172	1.1856	0.1800
8	0.1518	0.3283	0.0498	0.7072	0.1073	1.0860	0.1649
9	0.1518	0.3028	0.0460	0.6524	0.0990	1.0019	0.1521
10	0.1518	0.2811	0.0427	0.6055	0.0919	0.9299	0.1412
Car in category "$\geq$ 119"							
		0 Claim during $(0,t)$		1 Claim during $(0,t)$		2 Claims during $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.3306	0.7963	0.2633	1.7154	0.5671	2.6344	0.8709
2	0.3306	0.6616	0.2187	1.4251	0.4711	2.1887	0.7236
3	0.3306	0.5658	0.1871	1.2189	0.4030	1.8719	0.6189
4	0.3306	0.4943	0.1634	1.0648	0.3520	1.6352	0.5406
5	0.3306	0.4388	0.1451	0.9453	0.3125	1.4517	0.4799
6	0.2808	0.3945	0.1108	0.8499	0.2386	1.3052	0.3665
7	0.2808	0.3584	0.1006	0.7720	0.2168	1.1856	0.3329
8	0.2808	0.3283	0.0922	0.7072	0.1986	1.0860	0.3050
9	0.2808	0.3028	0.0850	0.6524	0.1832	1.0019	0.2813
10	0.2808	0.2811	0.0789	0.6055	0.1700	0.9299	0.2611

Table 2.7: Bonus-Malus coefficients and *a posteriori* premiums (2.2) for a 30-year-old policyholder.

Table 2.8 is the analog of Table 2.7 for an exponential loss. The Bonus-Malus factors in column 3 are computed from (2.3). The value 12.93 for the parameter c has been set in such a way that the variance of the *a posteriori* premiums paid by a policyholder during the first 10 years represents 50% of the variance if the premiums were computed under a quadratic loss; for more details, see Denuit and Dhaene (2001). It is interesting to compare the Bonus-Malus factors in Tables 2.7 and 2.8. When an exponential loss is used, the size of the *maluses* is reduced. Since the system is financially balanced, this implies that the size of the *bonuses* is also reduced.

Car in category "≤ 53"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.1787	0.9002	0.1609	1.3505	0.2413	1.8007	0.3218
2	0.1787	0.8207	0.1467	1.2253	0.2190	1.6299	0.2913
3	0.1787	0.7553	0.1350	1.1234	0.2007	1.4915	0.2665
4	0.1787	0.7003	0.1251	1.0384	0.1856	1.3765	0.2460
5	0.1787	0.6533	0.1167	0.9662	0.1727	1.2791	0.2286
6	0.1518	0.6125	0.0930	0.9039	0.1372	1.1953	0.1815
7	0.1518	0.5768	0.0876	0.8496	0.1290	1.1224	0.1704
8	0.1518	0.5452	0.0828	0.8017	0.1217	1.0583	0.1606
9	0.1518	0.5170	0.0785	0.7591	0.1152	1.0013	0.1520
10	0.1518	0.4916	0.0746	0.7210	0.1095	0.9504	0.1443
Car in category "≥ 119"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.3306	0.9002	0.2976	1.3505	0.4465	1.8007	0.5953
2	0.3306	0.8207	0.2713	1.2253	0.4051	1.6299	0.5388
3	0.3306	0.7553	0.2497	1.1234	0.3714	1.4915	0.4931
4	0.3306	0.7003	0.2315	1.0384	0.3433	1.3765	0.4551
5	0.3306	0.6533	0.2160	0.9662	0.3194	1.2791	0.4229
6	0.2808	0.6125	0.1720	0.9039	0.2538	1.1953	0.3356
7	0.2808	0.5768	0.1620	0.8496	0.2386	1.1224	0.3152
8	0.2808	0.5452	0.1531	0.8017	0.2251	1.0583	0.2972
9	0.2808	0.5170	0.1452	0.7591	0.2132	1.0013	0.2812
10	0.2808	0.4916	0.1381	0.7210	0.2025	0.9504	0.2669

Table 2.8: Bonus-Malus coefficients and *a posteriori* premiums (2.3) for a 30-year-old policyholder, with $c = 12.93$.

3 Integrated Ratemaking

3.1 Claim frequency model

In seminal papers, Dionne and Vanasse (1989, 1992) proposed a Bonus-Malus system which integrates *a priori* and *a posteriori* information on an individual basis. These authors introduced a regression component in the Poisson counting model in order to use all available information in the estimation of accident frequency.

Let us assume that the number of claims K_{it} for the ith policyholder of the portfolio during the year t conforms to a Poisson distribution with mean $\lambda_{I_i(t)}$, where $I_i(t)$ is the index of the risk class occupied by policyholder i in year t. A common problem for count data is that, even after allowing for important explanatory variables using the Poisson regression model, the fits obtained are rather poor. This indicates that, conditional upon the explanatory variables included in the final model, the variance of an observation is greater than its mean, implying that the Poisson assumption is incorrect. Most often, this is due to the fact that important explanatory variables may not have been measured and are consequently incorrectly excluded from the regression relationship.

A convenient way to take this phenomenon into account is to introduce a random effect in this model; see e.g. Pinquet (1999). We assume that K_{it} follows a Poisson distribution with mean $\lambda_{I_i(t)}\Theta_i$, where Θ_i conforms to a Gamma distribution with unit mean, i.e. with parameters (α, α). Then, K_{it} follows a Negative Binomial law, i.e.

$$\mathbb{P}[K_{it} = k | I_i(t)] = \binom{\alpha + k - 1}{k} \left(\frac{\lambda_{I_i(t)}}{\alpha + \lambda_{I_i(t)}} \right)^k \left(\frac{\alpha}{\alpha + \lambda_{I_i(t)}} \right)^\alpha, \quad k \in \mathbb{N}.$$

The meaning of Θ_i is that of an error term; Θ_i represents the impact on the mean claim frequency of all the policyholders' characteristics not taken into account *a priori*. On average, Θ_i has no impact on the claim frequency since $\mathbb{E}K_{it} = \lambda_{I_i(t)}$. Let us now derive the *a posteriori* distribution of Θ_i.

Lemma 3.1. *If the distribution function of* Θ_i *is* $\Gamma(.|\alpha, \alpha)$ *then the distribution function of* $[\Theta_i | K_{i1} = k_{i1}, K_{i2} = k_{i2}, \ldots, K_{it} = k_{it}]$ *is* $\Gamma(.|\alpha + k_{i\bullet}(t), \alpha + \lambda_{i\bullet}(t))$ *where*

$$k_{i\bullet}(t) = \sum_{j=1}^{t} k_{ij} \text{ and } \lambda_{i\bullet}(t) = \sum_{j=1}^{t} \lambda_{I_i(j)}.$$

Proof. Bayes theorem yields

$$\begin{aligned} d\mathbb{P}[\Theta_i \leq \theta | K_{i1} = k_{i1}, K_{i2} = k_{i2}, \ldots, K_{it} = k_{it}] \\ &= \frac{\mathbb{P}[K_{i1} = k_{i1}, K_{i2} = k_{i2}, \ldots, K_{it} = k_{it} | \Theta_i = \theta] d\mathbb{P}[\Theta_i \leq \theta]}{\mathbb{P}[K_{i1} = k_{i1}, K_{i2} = k_{i2}, \ldots, K_{it} = k_{it}]} \\ &= \frac{\theta^{k_{i\bullet}(t)} \exp(-\theta \lambda_{i\bullet}(t)) \alpha^\alpha \theta^{\alpha-1} \exp(-\alpha\theta) d\theta}{\alpha^\alpha \int_{\xi \in \mathbb{R}^+} \xi^{k_{i\bullet}(t)+\alpha-1} \exp(-\alpha \lambda_{i\bullet}(t) \xi) d\xi}, \end{aligned}$$

as announced. □

In order to estimate the parameter α describing the residual heterogeneity of the portfolio, we use the Maximum Likelihood method. Precisely, we maximize

$$L(\alpha) = \prod_{i=1}^{12} \prod_{k=0}^{k_{\max}^{(i)}} \left\{ \binom{\alpha + k - 1}{k} \left(\frac{\lambda_i}{\alpha + \lambda_i} \right)^k \left(\frac{\alpha}{\alpha + \lambda_i} \right)^\alpha \right\}^{n_{ik}};$$

this yields $\widehat{\alpha} = 0.8157$.

3.2 *A posteriori* premium using a quadratic loss

In the model described in the preceding section, Dionne and Vanasse (1989,1992) and Gisler (1996) have obtained the following result; it can be seen as a direct consequence of Proposition 4.1 and its proof is thus omitted.

Proposition 3.2. *Assume that the distribution function of* Θ_i *is* $\Gamma(\alpha,\alpha)$. *Under a quadratic loss, the a posteriori premium for policyholder i is given by*

$$P_{t+1}^{quad} = \lambda_{I_i(t+1)} BMF^{quad}(k_{i\bullet}(t), \lambda_{i\bullet}(t)),$$

where the Bonus-Malus coefficient is given by

$$BMF^{quad}(k_{i\bullet}(t), \lambda_{i\bullet}(t)) = \frac{\alpha + k_{i\bullet}(t)}{\alpha + \lambda_{i\bullet}(t)} = (1-\rho_{quad}) \times 1 + \rho_{quad}\frac{k_{i\bullet}(t)}{\lambda_{i\bullet}(t)}$$

with

$$\rho_{quad} = \frac{\lambda_{i\bullet}(t)}{\alpha + \lambda_{i\bullet}(t)}.$$

The *a posteriori* premium P_{t+1}^{quad} can be considered as the product of a base premium $\lambda_{i;t+1}$ and a bonus-malus coefficient $BMF^{quad}(k_{i\bullet}(t), \lambda_{i\bullet}(t))$. Note that the greater the variance of Θ_i (i.e. the smaller α) the greater ρ_{quad} (i.e. the greater the weight given to the claim history of the policyholder). Moreover, ρ_{quad} is clearly increasing in $\lambda_{i\bullet}$. If $\lambda_{i\bullet}$ is very small (which is for instance the case for policies with a high deductible) then ρ_{quad} is very small, too. The no-claim discount for such policies is thus also very small and as pointed out by Gisler (1996) the usefulness of Bonus-Malus systems can be questioned in this case. It is worth mentioning that it is very similar to the credibility factor obtained without covariates, except that now, the length of year t for policyholder i is λ_{it} instead of 1. In other words, the length of the time periods is determined by the relating base premiums.

3.3 *A posteriori* premium using exponential loss

The use of a quadratic loss function leads to very high *maluses*. Although theoretically correct, such a system is not accepted by policyholders. In order to have a model with a parameter controlling the severity of the system, let us now incorporate *a priori* variables in the exponential loss function.

Proposition 3.3. *Assume that the distribution function of* Θ_i *is* $\Gamma(.|\alpha,\alpha)$. *Under an exponential loss with parameter* $c > 0$, *the a posteriori premium for policyholder i is given by*

$$P_{t+1}^{exp} = \lambda_{I_i(t+1)} BMF^{exp}(k_{i\bullet}(t), \lambda_{i\bullet}(t))$$

where the Bonus-Malus coefficient is given by

$$BMF^{exp}(k_{i\bullet}(t), \lambda_{i\bullet}(t)) = (1-\rho_{exp}) \times 1 + \rho_{exp} \times \frac{k_{i\bullet}(t)}{\lambda_{i\bullet}(t)}$$

with

$$\rho_{exp} = \frac{\lambda_{i\bullet}(t)}{c} \ln\left(1 + \frac{c}{\alpha + \lambda_{i\bullet}(t)}\right).$$

Proof. From Lemma 3.1, we get

$$\mathbb{E}\Big[\exp(-c\Theta_i)|K_{i1}=k_{i1},K_{i2}=k_{i2},\ldots,K_{it}=k_{it}\Big]=\left(\frac{\alpha+\lambda_{i\bullet}(t)}{\alpha+\lambda_{i\bullet}(t)+c}\right)^{\alpha+k_{i\bullet}(t)},$$

whence it follows that

$$\ln\mathbb{E}\Big[\exp(-c\Theta_i)|K_{i1}=k_{i1},K_{i2}=k_{i2},\ldots,K_{it}=k_{it}\Big]=-(\alpha+k_{i\bullet}(t))\ln\left(1+\frac{c}{\alpha+\lambda_{i\bullet}(t)}\right)$$

and

$$\mathbb{E}\ln\mathbb{E}\Big[\exp(-c\Theta_i)|K_{i1}=k_{i1},K_{i2}=k_{i2},\ldots,K_{it}=k_{it}\Big]=-(\alpha+\lambda_{i\bullet}(t))\ln\left(1+\frac{c}{\alpha+\lambda_{i\bullet}(t)}\right)$$

The result then follows from Proposition 4.1. □

Let us now compare the Bonus-Malus coefficients obtained with a quadratic and exponential loss functions. Since for any $c \geq 0$,

$$\ln\left(1+\frac{c}{\alpha+\lambda_{i\bullet}}\right)\leq\frac{c}{\alpha+\lambda_{i\bullet}},$$

it is easily seen that $\rho_{exp}(c) \leq \rho_{quad}$; the weight given to past claims is thus smaller under an exponential loss.

It can be shown that $\rho_{exp}(c) \to 0$ as $c \to +\infty$. If the asymmetry factor c tends to $+\infty$ then all the risks within the same tariff class pay the same premium: there is no more experience rating. Conversely, $\rho_{exp}(c) \to \rho_{quad}$ as $c \to 0$. The results obtained by Dionne and Vanasse (1989,1992) so appear as limit cases of those obtained with an exponential loss function.

3.4 Numerical illustration

Let us first compute the premium for these two policyholders using Dionne-Vanasse's methodology. This yields the results in Table 3.1. Contrarily to Table 2.7, the Bonus-Malus factors are not the same for both individuals. The differences are explained by the presence of personal characteristics in the calculation of these factors. Comparing the bonus-malus factors of Table 2.7 and 3.1, we find that once the *a priori* variables are introduced the sizes of the *bonuses* and of the *maluses* are reduced. Technically, this means that part of the heterogeneity has been taken into account in the *a priori* differentiation of the premiums, so that the residual heterogeneity is smaller and the magnitude of the *a posteriori* corrections is reduced. It is interesting to note that even if the policyholder whose car is in category "≤ 53" always pays a premium smaller that the corresponding premium for the driver in category "≥ 119", his Bonus-Malus factors are always greater (i.e. he has less *bonuses* and more *maluses*). This comes from the fact that "good" risks are rewarded in their base premiums (through the *a priori* variables incorporated in the tariff); consequently, the size of *bonus* they require for equity is reduced. In other words, the premium discount awarded to

Car in category "≤ 53"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.1787	0.8203	0.1466	1.8259	0.3263	2.8316	0.5060
2	0.1787	0.6953	0.1243	1.5478	0.2766	2.4002	0.4289
3	0.1787	0.6034	0.1078	1.3432	0.2400	2.0829	0.3722
4	0.1787	0.5330	0.0952	1.1863	0.2120	1.8397	0.3288
5	0.1787	0.4772	0.0853	1.0623	0.1898	1.6474	0.2944
6	0.1518	0.4383	0.0665	0.9757	0.1481	1.5130	0.2297
7	0.1518	0.4053	0.0615	0.9021	0.1369	1.3989	0.2124
8	0.1518	0.3768	0.0572	0.8388	0.1273	1.3008	0.1975
9	0.1518	0.3521	0.0535	0.7838	0.1190	1.2155	0.1845
10	0.1518	0.3305	0.0502	0.7356	0.1117	1.1408	0.1732
Car in category "≥ 119"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.3306	0.7945	0.2626	1.4162	0.4682	2.0379	0.6737
2	0.3306	0.6590	0.2179	1.1747	0.3884	1.6905	0.5589
3	0.3306	0.5630	0.1861	1.0036	0.3318	1.4442	0.4775
4	0.3306	0.4914	0.1625	0.8760	0.2896	1.2606	0.4168
5	0.3306	0.4360	0.1441	0.7772	0.2569	1.1184	0.3697
6	0.2808	0.3979	0.1117	0.7092	0.1992	1.0206	0.2866
7	0.2808	0.3659	0.1027	0.6522	0.1831	0.9386	0.2635
8	0.2808	0.3387	0.0951	0.6037	0.1695	0.8687	0.2439
9	0.2808	0.3152	0.0885	0.5619	0.1578	0.8085	0.2270
10	0.2808	0.2948	0.0828	0.5255	0.1476	0.7562	0.2123

Table 3.1: Bonus-Malus coefficients and *a posteriori* premiums of Proposition 3.2 for a 30-year-old policyholder.

risks judged as "good" *a priori* has to be smaller than the *bonus* awarded to those judged as "bad" *a priori*. Conversely, the penalties in case of claims is more important.

The same remarks hold for the Bonus-Malus coefficients obtained with an exponential loss function presented in Table 3.2. The severity of the *a posteriori* corrections is weaker than with a quadratic loss function, as expected.

4 Appendix: Credibility Models with quadratic and exponential loss functions

We give here all the technical results used throughout this paper. Those involving a quadratic loss are standard. The use of an exponential loss function has been advocated in actuarial sciences by Ferreira (1977) and Lemaire (1979) in the context of Bonus-Malus systems. Such

Car in category "≤ 53"								
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$		
t	BP_t	BMF_t	Premium P_{t+1}	BMF_t	Premium P_{t+1}	BMF_t	Premium P_{t+1}
1	0.1787	0.9635	0.1722	1.1676	0.2087	1.3718	0.2451
2	0.1787	0.9313	0.1664	1.1236	0.2008	1.3159	0.2352
3	0.1787	0.9022	0.1612	1.0846	0.1938	1.2669	0.2264
4	0.1787	0.8758	0.1565	1.0495	0.1876	1.2232	0.2186
5	0.1787	0.8516	0.1522	1.0177	0.1819	1.1838	0.2115
6	0.1518	0.8324	0.1264	0.9927	0.1507	1.1531	0.1750
7	0.1518	0.8144	0.1236	0.9694	0.1472	1.1245	0.1707
8	0.1518	0.7974	0.1210	0.9476	0.1438	1.0978	0.1666
9	0.1518	0.7813	0.1186	0.9270	0.1407	1.0728	0.1628
10	0.1518	0.7660	0.1163	0.9076	0.1378	1.0492	0.1593

Car in category "≥ 119"							
		0 Claim in $(0,t)$		1 Claim in $(0,t)$		2 Claims in $(0,t)$	
t	BP_t	BMF	Premium	BMF	Premium	BMF	Premium
1	0.3306	0.9359	0.3094	1.1298	0.3735	1.3238	0.4377
2	0.3306	0.8835	0.2921	1.0597	0.3503	1.2359	0.4086
3	0.3306	0.8390	0.2774	1.0013	0.3310	1.1636	0.3847
4	0.3306	0.8003	0.2646	0.9513	0.3145	1.1023	0.3644
5	0.3306	0.7660	0.2532	0.9075	0.3000	1.0491	0.3468
6	0.2808	0.7396	0.2077	0.8743	0.2455	1.0089	0.2833
7	0.2808	0.7154	0.2009	0.8439	0.2370	0.9724	0.2731
8	0.2808	0.6931	0.1946	0.8161	0.2292	0.9391	0.2637
9	0.2808	0.6723	0.1888	0.7904	0.2219	0.9084	0.2551
10	0.2808	0.6530	0.1834	0.7665	0.2152	0.8800	0.2471

Table 3.2: Bonus-Malus coefficients and *a posteriori* premiums of Proposition 3.3 for a 30-year-old policyholder, with $c = 12.93$.

loss functions have been successively applied by Denuit and Dhaene (2001) to the design of Bonus-Malus system in a Markovian setting.

Let us consider a sequence of random variables $\{X_1, X_2, X_3, \ldots\}$ and a risk parameter Θ; in the remainder of this section, Θ is a random variable, or possibly a sequence of random variables. In the latter case, $\Theta = \{\Theta_1, \Theta_2, \ldots\}$ and it is assumed that X_i depends on Θ only through Θ_i. Given Θ, the X_i's are independent. The first two moments of the X_i's are assumed to be finite. Moreover, the conditional mean of the X_i's is given by

$$\mu_i(\Theta) = \mathbb{E}[X_i|\Theta], \;\; i = 1, 2, 3, \ldots ,$$

and $\mathbb{E}\mu_i(\Theta) = \mu_i$.

Proposition 4.1. **(i)** *The minimum of*

$$\mathbb{E}\Big(\mu_{n+1}(\Theta) - \Psi(X_1, X_2, \ldots , X_n)\Big)^2$$

on all the measurable functions $\Psi : \mathbb{R}^n \to \mathbb{R}$ *is obtained for*

$$\Psi^*(X_1, X_2, \dots, X_n) = \mathbb{E}[\mu_{n+1}(\Theta)|X_1, X_2, \dots, X_n].$$

(ii) *The minimum of*

$$\mathbb{E}\exp\Big\{-c\Big(\mu_{n+1}(\Theta) - \Psi(X_1, X_2, \dots, X_n)\Big)\Big\}$$

on all the measurable functions $\Psi : \mathbb{R}^n \to \mathbb{R}$ *satisfying the constraint* $\mathbb{E}\Psi(X_1, X_2, \dots, X_n) = \mu_{n+1}$ *is obtained for*

$$\begin{aligned}\Psi^*(X_1, X_2, \dots, X_n) &= \mu_{n+1} + \frac{1}{c}\Big\{\mathbb{E}\big[\ln\mathbb{E}\left[\exp(-c\mu_{n+1}(\Theta))|X_1, X_2, \dots, X_n\right]\big]\\ &\qquad - \ln\mathbb{E}\left[\exp(-c\mu_{n+1}(\Theta))|X_1, X_2, \dots, X_n\right]\Big\}.\end{aligned}$$

Proof. (i) is a classical result; a proof can be found in any statistical textbook. An easy way to get it consists in noting that

$$\begin{aligned}&\mathbb{E}\Big(\mu_{n+1}(\Theta) - \Psi(X_1, X_2, \dots, X_n)\Big)^2\\ =\ &\mathbb{E}\Big(\mu_{n+1}(\Theta) - \Psi^*(X_1, X_2, \dots, X_n) + \Psi^*(X_1, X_2, \dots, X_n) - \Psi(X_1, X_2, \dots, X_n)\Big)^2\\ =\ &\mathbb{E}\Big(\mu_{n+1}(\Theta) - \Psi^*(X_1, X_2, \dots, X_n)\Big)^2 + \mathbb{E}\Big(\Psi^*(X_1, X_2, \dots, X_n) - \Psi(X_1, X_2, \dots, X_n)\Big)^2,\end{aligned}$$

which is clearly minimal for $\Psi \equiv \Psi^*$.

Let us now turn to (ii). Starting from

$$\begin{aligned}&\mathbb{E}\exp\Big\{-c\Big(\mu_{n+1}(\Theta) - \Psi(X_1, X_2, \dots, X_n)\Big)\Big\}\\ =\ &\mathbb{E}\Big[\exp\left\{c\Psi(X_1, X_2, \dots, X_n)\right\}\mathbb{E}\left[\exp\{-c\mu_{n+1}(\Theta)\}|X_1, X_2, \dots, X_n\right]\Big]\\ =\ &\mathbb{E}\exp\Big\{c(\Psi(X_1, X_2, \dots, X_n) - \Psi^*(X_1, X_2, \dots, X_n))\Big\}\\ &\qquad\exp\{c\mu_{n+1}\}\exp\Big\{\mathbb{E}\ln\mathbb{E}\big[\exp\{-c\mu_{n+1}(\Theta)\}|X_1, X_2, \dots, X_n\big]\Big\}.\end{aligned}$$

Now, let us apply Jensen's inequality to get

$$\begin{aligned}&\mathbb{E}\exp\Big\{-c\Big(\mu_{n+1}(\Theta) - \Psi(X_1, X_2, \dots, X_n)\Big)\Big\}\\ \geq\ &\exp\Big\{c\mathbb{E}\left[\Psi(X_1, X_2, \dots, X_n) - \Psi^*(X_1, X_2, \dots, X_n)\right]\Big\}\\ &\qquad\exp\{c\mu_{n+1}\}\exp\Big\{\mathbb{E}\ln\mathbb{E}\big[\exp\{-c\mu_{n+1}(\Theta)\}|X_1, X_2, \dots, X_n\big]\Big\}.\end{aligned}$$

Because of the constraint on the expectation of the Ψ's, the first exponential is 1, yielding the announced result. □

Remark that in (ii) the constraint is made in order to guarantee the financial equilibrium.

Acknowledgements

The authors would like to thank the referee for his careful reading of the manuscript, as well as for several useful comments and suggestions.

References

[1] Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). *Actuarial Mathematics.* The Society of Actuaries. Itasca. Illinois.

[2] Denuit, M., and Dhaene, J. (2001). Bonus-Malus scales using exponential loss functions. *Bulletin of the German Society of Actuaries* **25**, 13-27.

[3] De Wit, G.W., and Van Eeghen, J. (1984). Rate making and society's sense of fairness. *ASTIN Bulletin* **14**, 151-163.

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

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

[6] Dobson, A.J. (1990). *An Introduction to Generalized Linear Models.* Chapman & Hall/CRC, London.

[7] Fahrmeir, L., and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer Verlag, New York.

[8] Ferreira, J. (1977). Identifying equitable insurance premiums for risk classes: an alternative to the classical approach. Lecture presented at the 23th international meeting of the Institute of Management Sciences, Athens, Greece.

[9] Gisler, A. (1996). Bonus-Malus and tariff segmentation. *Paper presented at the 27th ASTIN Colloquium, Copenhagen, September 1-5, 1996.*

[10] Lemaire, J. (1979). How to define a Bonus-Malus system with an exponential utility function. *ASTIN Bulletin* **10**, 274-282.

[11] Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance.* Kluwer Academic Publisher, Boston.

[12] Mc Cullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman & Hall, New York.

[13] Pinquet, J. (1997). Allowance for cost of claims in Bonus-Malus systems. *ASTIN Bulletin* **27**, 33-57.

[14] Pinquet, J. (1999). Experience rating through heterogeneous models. In *Handbook of Insurance*, edited by G. Dionne. Kluwer Academic Publishers.

[15] Renshaw, A.E. (1994). Modelling the claim process in the presence of covariates. *ASTIN Bulletin* **24**, 265-285.

[16] Young, V. (1996). Credibility and persistency. *ASTIN Bulletin* **26**, 53-69.