BONUS-MALUS SCALES USING EXPONENTIAL LOSS FUNCTIONS

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

August 2000

Abstract

This paper focuses on techniques for constructing Bonus-Malus systems in third party liability automobile insurance. Specifically, the article presents a practical method for constructing optimal Bonus-Malus scales with reasonable penalties that can be commercially implemented. For this purpose, the symmetry between the overcharges and the undercharges reflected in the usual quadratic loss function is broken through the introduction of parametric asymmetric loss functions of exponential type. The resulting system possesses the desirable financial stability property.

Key words and phrases: Bonus-Malus system, Markov chains, exponential loss functions

1 Introduction and Motivation

Rating systems penalizing insureds responsible for one or more accidents by premium surcharges (or *maluses*), and rewarding 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. Several studies have shown that, if insurers are 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 well 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 (BMS, in short). We will adopt here the latter terminology. For a thorough presentation of the techniques relating to BMS, see Lemaire (1995).

In the EU, competition based on BMS seems to have been limited to a few member countries. In Portugal for instance, each company developed its own experience rating system. The Portugese market is characterized by many movements among different companies, partly explained by competition but also by the lack of data disclosure among insurers. Policyholders have indeed the faculty to leave a company and to declare another one that it is the first policy they subscribe. As a result, policyholders placed in the highest classes tend to leave the company.

In the near future, complete freedom about *a posteriori* ratemaking will be given to all insurers operating in EU member states, in accordance with European Directives. Therefore, the turnover of policyholders is expected to increase in the near future. It seems thus hopeless to implement in practice the very principle of credibility theory, that is, to make correspond the premium paid to the true risk in the long run. Consequently, unless the regulatory authorities organize an information system so that each insurer has access to the past record of claims at fault, companies will not be able to apply credibility techniques. It seems therefore more realistic to apply BMS in order to counteract moral hazard: the BMS becomes a simple incentive to drive carefully. The systems considered in this paper for the purpose of illustration conform to this philosophy.

All the methods used so far to determine the relative premiums are characterized by important penalties in case of claims and moderate (or even small) premium discounts awarded to claim-free policyholders. This is a by-product of the use of a quadratic loss function. In this paper, we apply an idea proposed by Ferreira (1977) and Lemaire (1979) to the Bayes scales studied by Norberg (1976), Borgan, Hoem and Norberg (1981) and Gilde and Sundt (1989): specifically, we derive the optimal relativities under an exponential loss function. The main advantage of this method is that actuaries have now the freedom to determine the severity of the system by selecting the value of a single parameter.

A few words on the notation and terminology used troughout the paper. In the remainder, we denote a point of the real n-dimensional space $\mathbb{R}^n$ by a bold letter $\mathbf{x}$; the ith component of $\mathbf{x}$ is x_i, $i = 1, 2, \ldots, n$. All the vectors are tacitly assumed to be column vectors. A matrix is denoted by a capital letter in boldface, for instance $\mathbf{M}$; $\mathbf{M}^t$ denotes the transposition of $\mathbf{M}$. The vector of ones, that is $(1, 1, \ldots, 1)^t$, will be denoted by $\mathbf{e}$. The identity matrix (with entries 1 on the main diagonal and 0 elsewhere) is denoted by $\mathbf{I}$. Finally, we denote by $\mathbb{N}$

the set $\{0, 1, 2, \dots\}$ of the non-negative integers, by $\mathbb{N}_0$ the set $\{1, 2, 3, \dots\}$ of the positive integers and by $\mathbb{R}^+$ the half-positive real line $[0, +\infty)$.

2 Portfolio model

The framework of credibility theory, with its fundamental notion of randomly distributed risk parameters, was employed in analysis of BMS by Pesonen as early as 1963. To be specific, let us consider a portfolio of n policies. The ith policy of the portfolio, $i = 1, 2, \dots, n$, is represented by a sequence $(\Theta_i, K_{i1}, K_{i2}, K_{i3}, \dots)$ where K_{ij} represents the number of claims incurred by this policyholder during the jth year the policy is in force, i.e. during the period $(j-1, j)$. At the portfolio level, the sequences $(\Theta_i, K_{i1}, K_{i2}, K_{i3}, \dots)$ are assumed to be independent and identically distributed for $i = 1, 2, \dots, n$. 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 occurence of claims; it is regarded as a random variable. Given $\Theta_i = \theta$, the random variables $K_{i1}, K_{i2}, K_{i3}, \dots$ are assumed to be independent and identically distributed. Unconditionally, these random variables are obviously dependent. At the outset, Θ_i is totally unknown. As time goes on, its value is reflected by the risk performance on the policy. This fact makes individual experience rating possible.

Let us denote as X_{ijk}, $k = 1, 2, \dots, K_{ij}$ the amounts of the K_{ij} claims reported by the ith policyholder during the jth year. The total claim amount for this risk in year j is

$$S_{ij} = \sum_{k=1}^{K_{ij}} X_{ijk},$$

with the convention that the empty sum equals 0. The severities X_{ijk}, $i = 1, 2, \dots, n$, $j, k \in \mathbb{N}_0$, are assumed to be independent and identically distributed, and independent of the claim frequencies K_{ij}, $j \in \mathbb{N}_0$. This assumption 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. This assumption seems acceptable for third party liability insurance. Indeed, the payments of the insurance company are for the third party, not for the policyholder. Therefore, the amount paid by the insurer mostly depends on the characteristics of the third party, and not on those of the insured. Henceforth, we put $\mathbb{E}X_{ijk} \equiv 1$, 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

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

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

The very basic tenets of a BMS are as follows:

(i) a premium calculation principle;

(ii) a distribution for the $[K_{ij}|\Theta_i = \theta]$'s;

(iii) a distribution F_Θ for the Θ_i's;

(iv) a loss function of which expectation has to be minimized in order to find the optimal experience premium.

Considering (i), we assume throughout this paper, as it is common in practice, that each policy is charged an amount proportional to its expected number of claims (expected value principle). Let us now turn to item (ii). In the numerical illustrations, we assume in this paper that the occurrence of the claims reported by individual policyholders, given their risk parameter, is described by a homogeneous Poisson process. The assumptions underlying the Poisson counting model are as follows:

A1: the probability of an accident during a small period of time $(t, t + \Delta t)$ is, ignoring higher-order terms, proportional to the duration Δt of this interval;

A2: the probability in A1 does not depend on the start t of the interval;

A3: the probability of two or more accidents in time interval $(t, t + \Delta t)$ is negligible;

A4: the number of accidents relating to disjoint time intervals are independent.

If we assume that the accident pattern of a policyholder conforms to A1-A4 then the number of claims generated by this individual is Poisson distributed. Even if A2 eliminates seasonal effects and A4 rules out the learning experience from an accident, the set of all Poisson assumptions should at least provide (locally in time) a good approximation to the accident generating mechanism. The annual numbers of claims $[K_{i1}|\Theta_i = \theta], [K_{i2}|\Theta_i = \theta], \ldots$ reported by policyholder i are then independent and 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 constant over time.

Items (iii) and (iv) will be addressed further in this work.

All the numerical illustrations given in this paper are based on the data displayed in Table 2.1. Specifically, Table 2.1 shows the distribution of the number of claims in the automobile third-party liability portfolio of a typical Benelux company; n_k is the number of policies with k claims reported during the year 1995, $k = 0, 1, 2, 3, 4 (= k_{\max})$. It contains $n = 112,031$ policies and has mean $\overline{x} = 0.0935$ and variance $s^2 = 0.1023$. First, let us fit the distribution of Table 2.1 by a Poisson distribution. The Maximum Likelihood method leads to the selection of the observed mean as the estimator of θ. Computing the probabilities of a Poisson distribution with mean $\overline{x}$ and multiplying them by the sample size n leads to the fitted theoretical frequencies $\widehat{n}_k$ presented in column A of Table 2.1. The fit is extremely poor: there is not enough probability mass in the right tail of the Poisson distribution. To measure the goodness-of-fit, standard χ^2-statistics is used, with the following calculation procedure:

$$\chi^2_{obs} = -2 \sum_{k=0}^{k_{\max}} n_k \ln\left(\frac{\widehat{n}_k}{n_k}\right).$$

Considering the p-value, the Poisson assumption is rejected without doubt. The characteristic standard sign-sequence "+,−,+" for the $n_k - \widehat{n}_k$'s is clearly observed, indicating that the data probably come from a mixture of Poisson rather than from an homogenous Poisson distribution. The incompatibility of the homogeneity assumption underlying the Poisson model with statistical analysis already proves that the introduction of a BMS in automobile insurance is justified. Column D of Table 2.1 displays the negative binomial fit (that is, F_Θ corresponds to a two-parameter Gamma distribution with mean a/τ and variance a/τ^2). Parameters a and τ were estimated by Maximum Likelihood (using the moment estimations as starting values). This model provides a reasonable fit to the data, as can be seen from the expected $\widehat{n}_k$'s presented in Column D of Table 2.1. Nevertheless, the p-value is only of 9%, indicating that the model must be considered with prudence. The fits displayed in columns B and C correspond to nonparametric maximum likelihood estimation of F_Θ; see Section 3.4 for details.

3 Markov Models for Practical BMS

3.1 BMS as Markov chains

In practice, a BMS consists of a finite number of classes, each with its own premium level. New policyholders have access to a specified class. After each year, the policy moves up or down according to transition rules. In case a BMS is in force, all policies can be partitioned into a finite number of classes, so that the annual premium depends only on the class. The knowledge of the present level and of the number of claims of the present year suffice to determine the next class. This ensures that the BMS may be represented by a Markov chain: the future (the class for year $t+1$) depends on the present (the class for year t and the number of accidents reported during year t) and not on the past (the complete claim history and the levels occupied during years $1, 2, \ldots, t-1$). Sometimes, fictitious classes have to be introduced in order to meet this memoryless property.

Formally, an insurance company uses a BMS when

- the policyholders are partitioned into a finite number of disjoint classes $C_0, C_1, \ldots, C_s$ so that the annual premium depends only on the class;
- the policyholders begin their driving career in a specified starting class C_{in};
- an insured's class for a given year is determined uniquely by the class of the preceding annual period and the number of claims reported during this period.

The relativity associated to class C_ℓ is r_ℓ; the meaning is that an insured in class C_ℓ pays an amount of premium equals to r_ℓ% of the base premium corresponding to class C_{in} (hence $r_{in} = 1$). By convention, the classes $C_0, C_1, \ldots, C_s$ have been numbered so that $r_0 \leq r_1 \leq \ldots \leq r_s$. The transition rules, i.e. the rules determining the transfer from one class to another when the number of claims of the preceding period is known, are described by transformations T_k such that $T_k(\ell_1) = \ell_2$ if the policy is transferred from class C_{ℓ_1} to class C_{ℓ_2} when k claims have been reported; T_k can be represented as a matrix $\{t^{(k)}_{\ell_1\ell_2}\}$, $\ell_1, \ell_2 = 0, 1, \ldots, s$, such that

$$t^{(k)}_{\ell_1\ell_2} = \begin{cases} 1 \text{ if } T_k(\ell_1) = \ell_2 \\ 0 \text{ otherwise.} \end{cases}$$

k	n_k	A	B	C	D
0	102,435	102,026	102,435	102,435	102,442
1	8,804	9,544	8,805	8,811	8,778
2	714	446	712	703	743
3	65	14	68	76	63
4	12	0	10	8	5
5	1	0	2	1	0
≥ 6	0	0	0	0	0
χ^2_{obs}		365.67	1.25	3.78	8.18
# d.f.		5	1	3	4
p-value		$< 10^{-6}$	0.26	0.29	0.09

Column A: expected frequency with homogeneous Poisson
Column B: expected frequency with 3-point NPMLE $\widehat{F}_\Theta$
$\widehat{\theta}_1 = 0.132$, $\widehat{\theta}_2 = 0.829$, $\widehat{\theta}_3 \approx 0.000$
$\widehat{\pi}_1 = 0.651$, $\widehat{\pi}_2 = 0.009$ and $\widehat{\pi}_3 = 0.340$
Column C: expected frequency with 2-point NPMLE $\widehat{F}_\Theta$
$\widehat{\theta}_1 = 0.068$, $\widehat{\theta}_2 = 0.446$, $\widehat{\pi}_1 = 0.933$ and $\widehat{\pi}_2 = 0.067$
Column D: expected frequency with Negative Binomial
$\widehat{a} = 1.0255$ and $\widehat{\tau} = 10.9672$

Table 2.1: Data set from a typical automobile third-party liability insurance portfolio observed during the year 1995.

The evolution of a policyholder with mean claim frequency θ in such a BMS can thus be represented as a homogeneous Markov chain $\mathcal{Z}(\theta) = \{Z_\nu(\theta),\ \ \nu \in \mathbb{N}_0\}$ where $Z_\nu(\theta) = \ell$ if the policyholder is in class C_ℓ during the νth period of insurance. For an excellent introduction to the use of Markov chains in connection with BMS, we refer the interested reader to Chapter 7 in Rolski, Schmidli, Schmidt and Teugels (1999).

3.2 Transient distributions

In our model, the probability $p_{\ell_1\ell_2}(\theta)$ of moving from C_{ℓ_1} to C_{ℓ_2} for a policyholder with mean frequency θ is equal to

$$p_{\ell_1\ell_2}(\theta) = \sum_{k=0}^{+\infty} \exp(-\theta)\frac{\theta^k}{k!} t_{\ell_1\ell_2}^{(k)};$$

$\mathbf{M}(\theta)$ is the corresponding one-step transition matrix, i.e. $\mathbf{M}(\theta) = \{p_{\ell_1\ell_2}(\theta)\}$, $\ell_1, \ell_2 = 0, 1, \ldots, s$. The probability $p_{\ell_1\ell_2}$ defined as

$$p_{\ell_1\ell_2} = \int_{\theta\in\mathbb{R}^+} p_{\ell_1\ell_2}(\theta) dF_\Theta(\theta),\ \ \ell_1, \ell_2 = 0, 1, \ldots, s,$$

is the corresponding probability for a randomly selected policyholder of the portfolio. Taking the νth power of $\mathbf{M}(\theta)$ yields the ν-step transition matrix whose element $(\ell_1\ell_2)$, denoted as $p^{(\nu)}_{\ell_1\ell_2}(\theta)$, is the probability of moving from class C_{ℓ_1} to class C_{ℓ_2} in ν transitions, i.e.

$$p^{(\nu)}_{\ell_1\ell_2}(\theta) = \mathbb{P}[Z_{k+\nu}(\theta) = \ell_2 | Z_k(\theta) = \ell_1],\ \ k \in \mathbb{N}_0;$$

$p^{(\nu)}_{\ell_1\ell_2}$ is defined analogously to $p_{\ell_1\ell_2}$ by averaging for all the possible values for θ.

3.3 Stationary distribution

Let us define $p^{(\nu)}_\ell(\theta) = P[Z_\nu(\theta) = \ell]$. All BMS in practical use have a "best" class, with the property that a policy in that class remains in the class after a claim-free period. In the following, we restrict attention to such non-periodic bonus rules. The transition matrix $\mathbf{M}(\theta)$ associated to such a BMS is regular, i.e. there exists some integer $\xi_0 \geq 1$ such that all entries of $\{\mathbf{M}(\theta)\}^{\xi_0}$ are strictly positive. Consequently, the Markov chain $\mathcal{Z}(\theta)$ is ergodic and thus possesses a stationary distribution $\boldsymbol{\pi}(\theta) = (\pi_0(\theta), \pi_1(\theta), \ldots, \pi_s(\theta))^t$; $\pi_\ell(\theta)$ is the stationary probability for a policyholder with mean frequency θ to be in level ℓ i.e.

$$\pi_\ell(\theta) = \lim_{\nu\to+\infty} p^{(\nu)}_\ell(\theta),\ \ \ell = 0, 1, \ldots, s.$$

Note that $\boldsymbol{\pi}(\theta)$ does not depend on the starting class. The term $\pi_\ell(\theta)$ is the limit value of the probability that the policyholder is in class C_ℓ, when the number of periods tends to $+\infty$. It is also the fraction of the time a policyholder with claim frequency θ spends in class C_ℓ, once stationarity has been reached.

Let us now recall how to compute the $\pi_\ell(\theta)$'s. The vector $\boldsymbol{\pi}(\theta)$ is the solution of the system

$$\begin{cases} \boldsymbol{\pi}^t(\theta) = \boldsymbol{\pi}^t(\theta)\mathbf{M}(\theta), \\ \boldsymbol{\pi}^t(\theta)\mathbf{e} = 1. \end{cases}$$

Let $\mathbf{E}$ be the $(s+1)\times(s+1)$ matrix all of whose entries are 1, i.e. consisting of $s+1$ column vectors $\mathbf{e}$. Then, it can be shown that

$$\boldsymbol{\pi}(\theta) = \mathbf{e}^t\big(\mathbf{I} - \mathbf{M}(\theta) + \mathbf{E}\big)^{-1},$$

which provides a direct method to get $\boldsymbol{\pi}(\theta)$.

Let Z be a random variable valued in $\{0,1,\ldots,s\}$ such that $[Z|\Theta=\theta]$ conforms to the distribution $\boldsymbol{\pi}(\theta)$ i.e.

$$\mathbb{P}[Z=\ell|\Theta=\theta] = \pi_\ell(\theta),\ \ \ell=0,1,\ldots,s;$$

$[Z|\Theta=\theta]$ can be seen as the weak limit of the $Z_\nu(\theta)$'s. The unconditional distribution of Z is

$$\mathbb{P}[Z=\ell] = \int_{\theta\in\mathbb{R}^+} \pi_\ell(\theta)dF_\Theta(\theta) \equiv \pi_\ell;$$

Z may be interpreted as the bonus class of a randomly selected policy in the BMS once stationarity has been reached.

3.4 Determination of the transient and stationary distributions from observed claim frequencies

Several algorithms have been proposed in order to compute the stationary distribution of the policyholders in a given BMS. Dufresne (1988) proposed a very elegant technique requiring independence between the annual number of accidents per policyholder $K_{i1}, K_{i2}, K_{i3}, \ldots$; unfortunately, this independence assumption rules out all the mixed Poisson distributions. Dufresne (1995) adapted the reasoning to the mixed Poisson case, but at the cost of many numerical difficulties. Here, we follow the method proposed by Walhin and Paris (1999). The idea is to resort on the NonParametric Maximum Likelihood Estimator (NPMLE, in short) of the structure function F_Θ.

In a seminal paper, Simar (1976) gave a detailed description of the NPMLE of F_Θ, as well as an algorithm for its computation. The NPMLE $\widehat{F}_\Theta$ of F_Θ is a discrete distribution putting positive probability masses $\varphi_1, \varphi_2, \ldots, \varphi_q$ on q support points $\theta_1, \theta_2, \ldots, \theta_q$. The resulting model for the claim number is a finite mixture model, allowing for easy computations. Simar (1976) obtained an upper bound for the size q of the support of the NPMLE: let κ be the number of observed distinct values, i.e.

$$\kappa = \#\{k \in \mathbb{N} \text{ such that } n_k > 0\}$$

(in most cases, $\kappa = k_{\max}+1$) then $\widehat{F}_\Theta$ exists, is unique and has a number of support points less than or equal to

$$\widehat{q} = \min\left\{\left[\frac{k_{\max}+1}{2}\right], \kappa\right\}, \tag{3.1}$$

where $[x]$ denotes the integer part of the real x.

The solution $\widehat{F}_\Theta$ puts probability masses $\widehat{\varphi}_1, \widehat{\varphi}_2, \ldots, \widehat{\varphi}_{\widehat{q}}$ at the atoms $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_{\widehat{q}}$, i.e.

$$\mathbb{P}[K_{ij}=k] = \int_{\theta\in\mathbb{R}^+} \exp(-\theta)\frac{\theta^k}{k!}d\widehat{F}_\Theta(\theta) = \sum_{\zeta=1}^{\widehat{q}} \widehat{\varphi}_\zeta \exp(-\widehat{\theta}_\zeta)\frac{\widehat{\theta}_\zeta^k}{k!},\ \ k\in\mathbb{N}.$$

This finite mixture model has a nice intuitive interpretation. To fix the ideas, assume that $\widehat{q} = 3$ and $\widehat{\theta}_1 < \widehat{\theta}_2 < \widehat{\theta}_3$. Then, the portfolio consists of $\widehat{\varphi}_1$% of good drivers with annual mean claim frequency of $\widehat{\theta}_1$, of $\widehat{\varphi}_2$% of medium-quality drivers with annual mean claim frequency $\widehat{\theta}_2$ and of $\widehat{\varphi}_3$% of bad drivers with annual mean claim frequency of $\widehat{\theta}_3$. A desirable property of $\widehat{F}_\Theta$ is that

$$\int_{\theta\in\mathbb{R}^+} \theta d\widehat{F}_\Theta(\theta) = \overline{x},$$

so that the observed claim amount is kept unchanged.

For the portfolio in Table 2.1, Simar's upper bound (3.1) is 3. The fit based on the 3-point $\widehat{F}_\Theta$ is displayed in column B of this table; column C shows the corresponding fit with a 2-point $\widehat{F}_\Theta$. Considering the p-values, these two fits can be regarded as statistically equivalent. We therefore prefer the 2-point structure function since it involves less parameters (statistical principle of parcimony). The NPMLE is thus a good-risk/bad-risk model. It is worth mentioning that the purely discrete nature of the NPMLE is sometimes undesirable. Therefore, Denuit and Lambert (2000) have proposed a smoothed version of the NPMLE, using a Gamma kernel.

Using NPMLE, the computation of the $p_{\ell_1\ell_2}$'s and of the π_{ℓ_2}'s become rather easy. The idea is to compute $\mathbf{M}(\theta_\zeta)$, the transition matrix for policyholders with claim frequency θ_ζ, $\zeta = 1, 2, \ldots, q$. The element $(\ell_1\ell_2)$ of $\big(\mathbf{M}(\theta_\zeta)\big)^\nu$ is $p^{(\nu)}_{\ell_1\ell_2}(\theta_\zeta)$, and hence

$$p^{(\nu)}_{\ell_1\ell_2} = \int_{\theta\in\mathbb{R}^+} p^{(\nu)}_{\ell_1\ell_2}(\theta) d\widehat{F}_\Theta(\theta) = \sum_{\zeta=1}^{\widehat{q}} \widehat{\varphi}_\zeta p^{(\nu)}_{\ell_1\ell_2}(\widehat{\theta}_\zeta).$$

Norming the left eigenvector of $\mathbf{M}(\widehat{\theta}_\zeta)$ or applying Dufresne's (1988) method yields the stationary distribution for the policyholders with annual mean claim frequency of $\widehat{\theta}_\zeta$, $\boldsymbol{\pi}(\widehat{\theta}_\zeta)$. It suffices then to compute the weighted average of these eigenvectors to get the stationary distribution of a randomly selected policyholder of the portfolio, i.e.

$$\pi_{\ell_2} = \sum_{\zeta=1}^{\widehat{q}} \widehat{\varphi}_\zeta \pi_{\ell_2}(\widehat{\theta}_\zeta), \;\; \ell_2 = 0, 1, \ldots, s.$$

3.5 Example of a BMS

In this paper, we consider the following type of Bonus-Malus scale for our practical illustrations. The policyholders are classified according to the number of claim-free years since their last claim. After a claim all premiums reductions are lost. Such systems are widely used in UK. They have been considered e.g. by De Pril and Goovaerts (1983) who studied the bonus-hunger phenomenon induced by such merit rating plans. See also Lemaire (1995).

Specifically, let us consider a BMS with s classes (numbered 0 to s). The starting class is s. Each claim-free year is rewarded by one bonus class. In case an accident is reported, all the discounts are lost and the policyholder is transferred to class s.

Note that the philosophy behind such a BMS is different from credibility theory. Indeed, this BMS only aims to counteract moral hazard: it is in fact more or less equivalent to a deductible which is not paid at once but smoothed over s years (the time needed to go back to the lowest class). In the authors' opinion, such a system is pragmatic in a very competitive market. The turnover of policyholders is indeed important and it seems hopeless to make correspond in the long run the premium paid to the true risk, since policyholders stay on average just a few years in the same company (and no official information system is created or it is easily eluded).

For instance, let us consider such a BMS with 6 classes (numbered 0 to 5). This choice is commercially reasonable in a country where the average claim frequency is approximately equal to 10%. Indeed, every ten years, the average driver will stay five years in the super bonus class and the other five years, the policyholder will move down to come back to the lowest level. The transition matrix $\mathbf{M}(\theta)$ associated to this BMS is given next:

$$\mathbf{M}(\theta) = \begin{pmatrix} \exp(-\theta) & 0 & 0 & 0 & 0 & 1-\exp(-\theta) \\ \exp(-\theta) & 0 & 0 & 0 & 0 & 1-\exp(-\theta) \\ 0 & \exp(-\theta) & 0 & 0 & 0 & 1-\exp(-\theta) \\ 0 & 0 & \exp(-\theta) & 0 & 0 & 1-\exp(-\theta) \\ 0 & 0 & 0 & \exp(-\theta) & 0 & 1-\exp(-\theta) \\ 0 & 0 & 0 & 0 & \exp(-\theta) & 1-\exp(-\theta) \end{pmatrix}.$$

Let us now consider a large group of policyholders entering such a system at time 0 in level 5 (we recall that all the computations are based on the 2-point NPMLE fit of Table 2.1). In Figures 3.3.1-3.3.3, the transient distributions $p_{5\ell}^{(k)}$ are depicted for $\ell = 0, 1, \ldots, 5$ and $k = 1, 3, 5$. It is easily checked that $p_{5\ell}^{(5)} = \pi_\ell$ for $\ell = 0, 1, \ldots, 5$, so that the system needs 5 years to reach stationarity (i.e. the time needed by the best policyholders starting from level 5 to arrive in class 0). Therefore, Figure 3.3.3 also gives the stationary distribution.

4 Determination of the relativities

4.1 Norberg's method with a quadratic loss

To each of the $s+1$ levels of a BMS, we would like to attach a relativity r_ℓ; the premium for class ℓ, $\ell = 0, 1, \ldots, s$, is the product of the base premium and a fraction r_ℓ%. Let us denote as P_ℓ the premium corresponding to level ℓ.

Predictive accuracy is a useful measure of the efficiency of a BMS. The idea behind this notion is as follows. A BMS is good at discriminating among the good and the bad risks if the premium they pay is close to their "true" premium. According to Norberg (1976), once the number of classes, the starting level and the transition rules have been fixed, the optimal premium P_ℓ^{quad} associated to level ℓ is determined by maximizing the asymptotic predictive accuracy. Formally, the P_ℓ^{quad}'s minimize the mean squared deviation between a policy's expected claim frequency and its premium in the year t as $t \to +\infty$: denoting as Θ the unknown claim frequency of a randomly selected policyholder of the portfolio, our aim is to choose the function P such that the expected squared difference between the "true" premium Θ and the premium P_Z paid by a policyholder in the system (after the stationary state has been reached), i.e. the goal is to minimize

$$Q_0 = \mathbb{E}(\Theta - P_Z)^2 = \int_{\theta\in\mathbb{R}^+} \sum_{j=0}^{s} (\theta - P_j)^2 \pi_j(\theta) dF_\Theta(\theta)$$

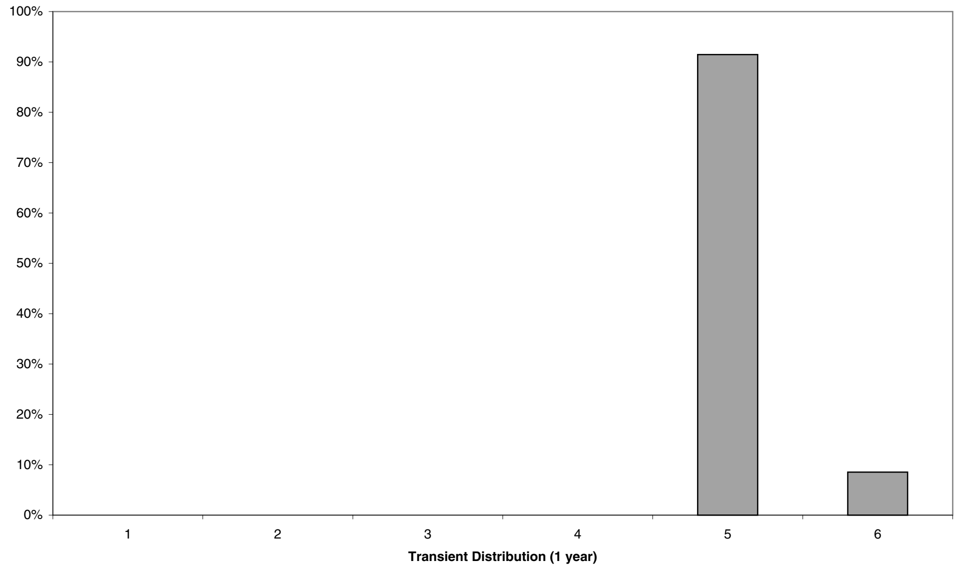

Figure 3.3.1: *1-year transient distribution*

where Z is as defined in Section 3.3. The solution P_Z^{quad} is given by $\mathbb{E}[\Theta|Z]$ so that

$$P_\ell^{quad} = \mathbb{E}[\Theta|Z = \ell] = \frac{\int_{\theta\in\mathbb{R}^+} \theta\pi_\ell(\theta)dF_\Theta(\theta)}{\pi_\ell}, \quad \ell = 0, 1, \ldots, s.$$

The optimal premium for class C_ℓ is thus equal to the conditional expected claim frequency for an infinitely old policy, given that the policy is in class C_ℓ. The corresponding relativities are given by

$$r_\ell^{quad} = 100 \times \frac{\mathbb{E}[\Theta|Z = \ell]}{\mathbb{E}\Theta}\%, \qquad \ell = 0, 1, \ldots, s.$$

It is easily seen that $\mathbb{E}r_Z^{quad} = 1$, resulting in financial equilibrium once steady state is reached. This fundamental property is highly desirable: the introduction of a BMS has no impact on the yearly premium collection. The repartition of the amounts paid by the policyholders is modified according to the reported claims but on the whole, the company gets the same amount of money. The relativities r_ℓ^{quad} for the practical example described in Section 3.5 are displayed in Table 4.1. The penalties for policyholders reporting claim are very severe. For instance, a policyholder in the superbonus class pay 77.2% of the base premium. If he files a claim, he will pay 187.1% of the base premium, that is more than twice his preceding premium.

This method entirely relies on the stationary distribution of the BMS. It can therefore be recommended only if the steady state is reached after a relatively short period, as it is the case for the BMS considered in Section 3.5. Moreover, because of the construction of

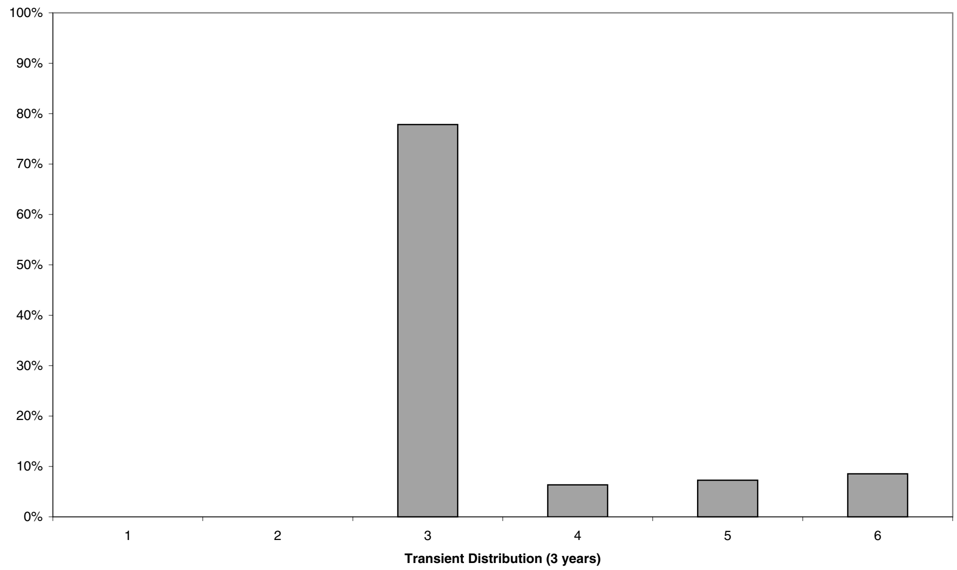

Figure 3.3.2: *3-year transient distribution*

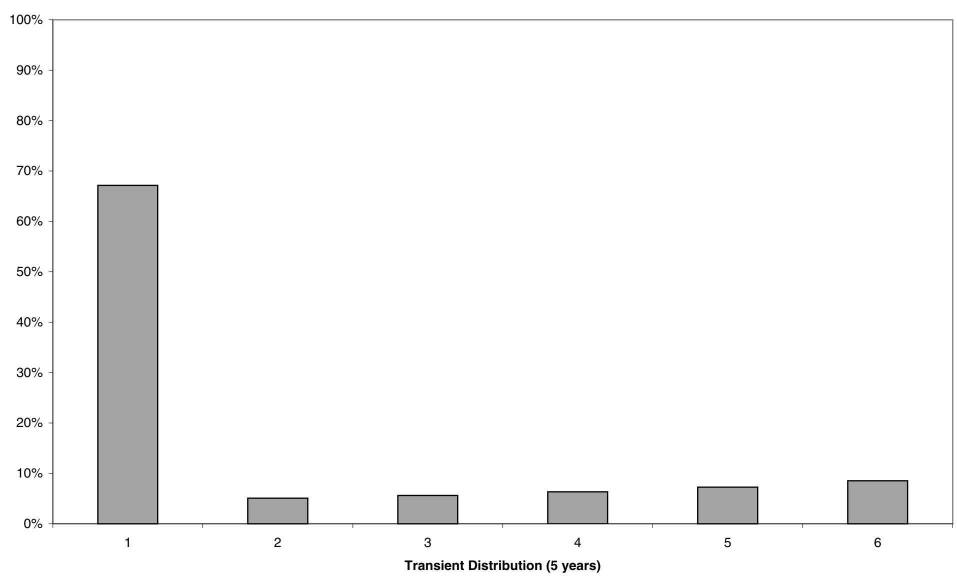

Figure 3.3.3: *5-year transient or stationary distribution*

this system, using the stationary distribution for the computation yields higher premiums than those obtained using transient distributions, with the method of Borgan, Hoem and Norberg (1981).

Level ℓ	0	1	2	3	4	5
r_ℓ^{quad}	77.2	105.2	118.3	136.0	158.8	187.1
r_ℓ^{exp}, $\eta = 75\% \Leftrightarrow c = 1.018$	80.4	103.9	115.2	130.5	150.1	176.2
r_ℓ^{exp}, $\eta = 50\% \Leftrightarrow c = 2.465$	85.1	102.8	111.9	124.3	141.1	162.8
r_ℓ^{exp}, $\eta = 25\% \Leftrightarrow c = 5.108$	88.8	101.7	108.0	116.8	128.8	144.8
r_ℓ^{lin}, $a = 0.0721$, $b = 0.0198$	77.3	98.5	119.7	140.9	162.0	183.2
r_ℓ^{e-lin}, $\eta = 75\%$, $a = 0.0750$, $b = 0.0171$	80.3	98.7	117.0	135.4	153.7	172.0
r_ℓ^{e-lin}, $\eta = 50\%$, $a = 0.0784$, $b = 0.0140$	84.0	98.9	113.9	128.9	143.8	158.8
r_ℓ^{e-lin}, $\eta = 25\%$, $a = 0.0828$, $b = 0.0099$	88.7	99.2	109.8	120.4	131.0	141.5

Table 4.1: Relativities r_ℓ^{quad}, r_ℓ^{exp}, r_ℓ^{lin} and r_ℓ^{e-lin}.

4.2 Norberg's method with exponential loss function

When the new premium amount is fixed by the insurer, two kinds of errors may arise: either the policyholder is undercharged and the insurance company looses its money or the insured is overcharged and the insurer is at risk of losing the policy. In order to penalize large mistakes to a greater extent, it is usually assumed that the loss functions is a non-negative convex function of the error. The loss is zero when no error is made and strictly positive otherwise. In most papers devoted to BMS, the loss function is taken to be quadratic. Among other choices we find also the absolute loss and the 4-degree loss; see e.g. Lemaire and Vandermeulen (1983). The problem with the two latter losses is that the resulting BMS are unbalanced; in practice, actuaries most often resort on a quadratic loss.

Our aim is to propose an asymmetric loss function with one parameter; the latter reflects the severity of the BMS. In order to reduce the maluses obtained with a quadratic loss, keeping a financially balanced system, we resort on an exponential loss function. It is worth mentioning that such loss functions have been first proposed by Ferreira (1977) and Lemaire (1979) in the classical credibility setting. Our purpose here is to apply exponential loss function to determine the optimal Bonus-Malus scale.

In order to soften the penalties, keeping the financial stability property condition, let us determine P_Z so to minimize

$$\mathbb{E} \exp\big\{ -c(\Theta - P_Z)\big\} \quad (4.1)$$
$$\text{under the financial stability constraint } \mathbb{E}P_Z = \mathbb{E}\Theta$$

where the parameter $c > 0$ determines the "severity" of the BMS. The choice of the loss (4.1) is made in order to reduce the penalties compared to those obtained under a quadratic loss. Indeed, (4.1) puts more weight on the errors resulting in an overestimation of the premium (i.e. $P_Z > \Theta$) than on those coming from an underestimation. Consequently, the maluses are reduced, as well as the bonuses since financial stability has been imposed.

Let us derive the general solution of (4.1).

Proposition 4.1. *The solution of the constrained optimization problem (4.1) is*

$$P_Z^{exp} = \mathbb{E}\Theta + \frac{1}{c}\Big\{\mathbb{E}\big[\ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\big] - \ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\Big\}.$$

Proof. First, note that

$$\exp\{cP_Z^{exp}\} = \frac{\exp\{c\mathbb{E}\Theta\}\exp\Big\{\mathbb{E}\Big[\ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\Big]\Big\}}{\mathbb{E}\big[\exp(-c\Theta)|Z\big]}.$$

Now, considering the latter formula, we have to minimize

$$\begin{aligned}\mathbb{E}\Big[\exp\big\{-c(\Theta - P_Z)\big\}\Big] &= \mathbb{E}\Big[\exp\big\{c(P_Z - P_Z^{exp})\big\}\Big] \\ &\qquad \exp\{c\mathbb{E}\Theta\}\exp\Big\{\mathbb{E}\big[\ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\big]\Big\}\end{aligned}$$

Invoking Jensen's inequality yields

$$\begin{aligned}\mathbb{E}\Big[\exp\big\{-c(\Theta - P_Z)\big\}\Big] &\geq \underbrace{\exp\Big\{c\mathbb{E}\big[P_Z - P_Z^{exp}\big]\Big\}}_{=1} \\ &\qquad \exp\{c\mathbb{E}\Theta\}\exp\Big\{\mathbb{E}\ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\Big\} \\ &= \mathbb{E}\Big[\exp\big\{-c(\Theta - P_Z^{exp})\big\}\Big],\end{aligned}$$

which ends the proof. □

The optimal relativities are then given by $r_\ell^{exp} = 100 \times P_Z^{exp}/\mathbb{E}\Theta\%$. In order to compute these quantities, it suffices to evaluate

$$\mathbb{E}[\exp(-c\Theta)|Z = j] = \frac{\int_{\theta\in\mathbb{R}^+}\exp(-c\theta)\pi_j(\theta)dF_\Theta(\theta)}{\pi_j}$$

and

$$\begin{aligned}E\Big(\ln \mathbb{E}[\exp(-c\Theta)|Z]\Big) &= \sum_{j=0}^{s}\pi_j \ln \mathbb{E}[\exp(-c\Theta)|Z = j] \\ &= \sum_{j=0}^{s}\pi_j \ln\left(\frac{\int_{\theta\in\mathbb{R}^+}\exp(-c\theta)\pi_j(\theta)dF_\Theta(\theta)}{\pi_j}\right).\end{aligned}$$

Let us briefly explain a possible criterion to fix the value of the parameter c. First, note that

$$\lim_{c\to 0} P_j^{exp} = \mathbb{E}[\Theta|Z = j] = P_j^{quad},$$

so that letting c tend to 0 yields Norberg's approach. In other words, the BMS becomes more severe as c decreases. Now, the ratio of the variances of the premiums obtained with an exponential and a quadratic loss is given by

$$\frac{\mathbb{V}\mathrm{ar}[P_Z^{exp}]}{\mathbb{V}\mathrm{ar}[P_Z^{quad}]} = \frac{1}{c^2}\frac{\mathbb{V}\mathrm{ar}\Big[\ln \mathbb{E}\left[\exp(-c\Theta)|Z\right]\Big]}{\mathbb{V}\mathrm{ar}\Big[\mathbb{E}\left[\Theta|Z\right]\Big]} = \eta\% \leq 100\%.$$

The idea is then to select the variance of the premium in the new system as a fraction of the corresponding variance under a quadratic loss (for instance $\eta = 25$, 50 or 75%). Of course, other procedures can be applied. For instance, the actuary could select the value of r_0, or of r_s, and then compute c in order to match this value.

The relativities r_ℓ^{exp} for the BMS described in Section 3.5 are displayed in Table 4.1 for the aforementioned η's. The severity of the BMS clearly increases as c decreases, the r_ℓ^{quad}'s corresponding to $c = 0$.

4.3 Gilde and Sundt method

In practice, a linear scale of the form $P_\ell = a + b\ell$, $\ell = 0, 1, \ldots, s$, could be desirable. Such scales have been studied by Gilde and Sundt (1989). In their paper, the optimal linear premium is the solution of the minimization of

$$\mathbb{E}(\Theta - P_Z)^2 = \mathbb{E}(\Theta - a - bZ)^2.$$

It is well-known that the solution of this optimization problem is given by

$$b = \frac{\mathbb{C}\mathrm{ov}[Z,\Theta]}{\mathbb{V}\mathrm{ar}[Z]} \text{ and } a = \mathbb{E}\Theta - \frac{\mathbb{C}\mathrm{ov}[Z,\Theta]}{\mathbb{V}\mathrm{ar}[Z]}\mathbb{E}Z. \tag{4.2}$$

The linear premium scale is thus of the form

$$P_j^{lin} = \mathbb{E}\Theta + \frac{\mathbb{C}\mathrm{ov}[Z,\Theta]}{\mathbb{V}\mathrm{ar}[Z]}(j - \mathbb{E}Z),$$

where

$$\begin{aligned}\mathbb{C}\mathrm{ov}[Z,\Theta] = \mathbb{C}\mathrm{ov}\Big[\mathbb{E}[Z|\Theta],\Theta\Big] &= \sum_{\ell=0}^{s} \ell\mathbb{E}[\Theta\pi_\ell(\Theta)] - \mathbb{E}Z\mathbb{E}\Theta \\ &= \sum_{\ell=0}^{s} \ell \int_{\theta\in\mathbb{R}^+} \theta\pi_\ell(\theta)dF_\Theta(\theta) - \mathbb{E}Z\mathbb{E}\Theta.\end{aligned}$$

Finally, the relativities are given by $r_\ell^{lin} = 100 \times P_\ell^{lin}/\mathbb{E}\Theta\%$.

The relativities r_ℓ^{lin} for the BMS described in Section 3.5 are displayed in Table 4.1; they are rather similar to the unconstrained r_ℓ^{quad}'s.

Let us now indicate how Gilde and Sundt's (1989) approach can be extended using exponential loss functions. The aim is now to minimize the objective function

$$\mathcal{O}(a,b) = \mathbb{E}\exp\Big\{ - c(\Theta - a - bZ)\Big\}$$

under the constraint $\mathbb{E}\Theta = a + b\mathbb{E}Z$. The latter relation implies that $a = \mathbb{E}\Theta - b\mathbb{E}Z$ so that it suffices to minimize

$$\widetilde{\mathcal{O}}(b) = \mathbb{E}\exp\Big\{ - c\Big(\Theta - \mathbb{E}\Theta - b(Z - \mathbb{E}Z)\Big)\Big\}.$$

Differentiating $\widetilde{\mathcal{O}}$ with respect to b and equating to zero yields

$$\begin{aligned} & \mathbb{E}\Big[(Z - \mathbb{E}Z)\exp\Big\{ - c\Big(\Theta - \mathbb{E}\Theta) - b(Z - \mathbb{E}Z)\Big)\Big\}\Big] = 0 \\ \Leftrightarrow \quad & \int_{\theta\in\mathbb{R}^+} \sum_{\ell=0}^{s} (\ell - \mathbb{E}Z)\exp\Big\{ - c\Big(\theta - \mathbb{E}\Theta) - b(\ell - \mathbb{E}Z)\Big)\Big\}\pi_\ell(\theta)dF_\Theta(\theta) = 0 \end{aligned}$$

which has to be solved numerically to get the value of b (and hence of a). Convenient starting values for the numerical search are provided by (4.2). This yields the relativities

$$r_\ell^{e-lin} = 100 \times \frac{a + b\ell}{\mathbb{E}\Theta}\%;$$

the values of the r_ℓ^{e-lin}'s in our example are listed in Table 4.1.

5 Conclusions

In this paper, we have demonstrated on the basis of a typical Benelux data set how to build optimal BMS for asymmetric loss functions of exponential-type. This allows the actuary to design financially balanced BMS with moderate penalties, that can be implemented in practice. In that respect, this work solves an open problem faced with the quadratic loss function, namely too high *maluses*.

Let us briefly comment some of the simplifying assumptions made throughout this paper. First, we considered a closed portfolio: no policy cancellations and no new policyholders entering the portfolio. This is of course unrealistic. In a market where competition is (partly) based on BMS, policy lapses cannot be neglected, in particular for policyholders with high maluses. In order to take this phenomenon into account, a convenient way is to introduce an additional $(s+2)$th state in the Markov chain: when a policy is transferred to level $s+1$, it means that the policyholder changes of company. It is then realistic to use transition probabilities $p_{\ell;s+1}$ increasing with ℓ.

Another crucial assumption is the constancy of the claim frequency θ for each policyholder during his whole driving carreer. This is also unrealistic since claim frequencies are generally convex functions of the age, for instance. An answer to this question is provided by the inclusion of explanatory variables, as it is done in Dionne and Vanasse (1989,1992), Pinquet (1997,1999) and Bermúdez, Denuit, Dhaene and Morillo (2000) in the classical framework of credibility theory. In a forthcoming work, we expand on the ideas given in Taylor (1997) and we build optimal Bonus-Malus scales taking into account *a priori* risk classification.

Acknowledgements

The first named author gratefully acknowledges the support of the Belgian Government under "Projet d'Actions de Recherche Concertées" (No. 98/03-217).

References

[1] Bermúdez, L., Denuit, M., Dhaene, J., and Morillo, I. (2000). Optimal Bonus-Malus systems integrating a priori risk classification. Mimeo.

[2] Borgan, Ø, Hoem, J.M., and Norberg, R. (1981). A nonasymptotic criterion for the evaluation of automobile bonus systems. *Scandinavian Actuarial Journal*, 265-178.

[3] Denuit, M., and Lambert, Ph. (2000). Smoothed NPML estimation of the risk distribution. Discussion Paper, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[4] De Pril, N., and Goovaerts, M.J. (1983). Bounds for the optimal critical claim size of a bonus system. *Insurance: Mathematics and Economics* **2**, 27-32.

[5] 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.

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

[7] Dufresne, F. (1988). Distribution stationnaire d'un système bonus-malus et probabilité de ruine. *ASTIN Bulletin* **18**, 31-46.

[8] Dufresne, F. (1988). The efficiency of the Swiss Bonus-Malus system. *Bulletin of the Swiss Association of Actuaries*, 29-41.

[9] 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.

[10] Gilde, V., and Sundt, B. (1989). On Bonus systems with credibility scales. *Scandinavian Actuarial Journal*, 13-22

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

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

[13] Lemaire, J., and Vandermeulen, E. (1983). Une propriété du principe de l'espérance mathématique. *Bulletin Trimestriel de l'Institut des Actuaires Français*, 5-14.

[14] Norberg, R. (1976). A credibility theory for automobile bonus system. *Scandinavian Actuarial Journal*, 92-107.

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

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

[17] Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1999). *Stochastic Processes for Insurance and Finance*. John Wiley & Sons, New York.

[18] Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Annals of Statistics* **4**, 1200-1209.

[19] Taylor, G. (1997). Setting a Bonus-Malus scale in the presence of other rating factors. *ASTIN Bulletin* **27**, 319-327.

[20] Walhin, J.-F., and Paris, J. (1999). Using mixed Poisson distributions in connection with Bonus-Malus systems. *ASTIN Bulletin* **29**, 81-99.