Marketing and Bonus-Malus Systems

SANDRA PITREBOIS[‡], MICHEL DENUIT[†*]
& JEAN-FRANÇOIS WALHIN[*‡]

[‡]Secura Belgian Re
Avenue des Nerviens 9-31 boîte 6
B-1040 Bruxelles, Belgium

[†]Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium

[*]Institut des Sciences Actuarielles
Université Catholique de Louvain
Grand-Rue, 54
B-1348 Louvain-la-Neuve, Belgium

April 4, 2003

Abstract

Based on the example of the Belgian Bonus-Malus scale, we will show in this article how to calculate the relative premiums associated with each level. These premiums will depend on the *a priori* ratemaking carried out preliminary. On the basis of a few examples we will also demonstrate how a Bonus-Malus system can be used as a marketing tool.

1 Introduction.

One of the main tasks of the actuary is to design a tariff structure that will fairly distribute the burden of claims among policyholders. To this end he often has to divide all policies into homogeneous classes, with all policyholders belonging to the same class paying the same premium. The classification variables introduced to divide risks into different classes are called *a priori* variables (as their values can be determined before the policyholder starts to drive). In motor third-party liability (MTPL, in short) insurance, they include age, gender and occupation of the policyholders, type and use of their car, place where they live and sometimes even number of cars in the household or marital status. It is convenient to achieve an *a priori* classification by resorting to generalized linear models (e.g. Poisson regression).

However, there are important factors that 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, risk classes are still quite heterogeneous despite the use of many *a priori* variables. But it is reasonable to believe that these hidden factors are revealed by the number of claims reported by the policyholders over the successive insurance periods. Hence the premium amount is adjusted each year on the basis of the individual claims experience in order to restore fairness among policyholders.

Rating systems penalizing insureds responsible for one or more accidents by premium surcharges (or *mali*), and rewarding claimsfree policyholders by awarding them discounts (or *boni*) are now in force in many developed countries. This *a posteriori* ratemaking is a very efficient way of classifying policyholders according to their risk. Besides encouraging policyholders to drive carefully (i.e. counteracting moral hazard), such systems aim to better assess individual risks. They are called no-claim discount systems, experience rating, merit rating, or Bonus-Malus systems (BMS, in short). In this paper we will adopt the latter terminology. For a thorough presentation of the techniques relating to BMS, see Lemaire (1995).

When a BMS is in force, the premium amount paid by the policyholder depends on the rating factors of the current period but also on the claims history. In practice, a BMS consists of a finite number of levels, each with its own relative premium. New policyholders have access to a specified class. After each year, the policy moves up or down according to transition rules and to the number of claims at fault. These transition rules are fixed and depend only on the present level and on the number of claims of the present year. The premium charged to the policyholder is obtained by applying the relative premium corresponding to his current level in the BMS to a base premium depending on his observable characteristics incorporated into the price list.

Until recently, the Belgian MTPL Bonus-Malus system was defined by the legislator, and all the insurers were to conform to it (they thus had freedom to fix the level of the basis premium, but were all to penalize the claims in the same way). This Belgian characteristic was an infringment of the European directives as regards insurance, which recommend the suppression of all tariff restrictions (only an *a posteriori* control remains). The royal decree of January 16, 2002 thus envisages the gradual end of the official Bonus-Malus scale. As from 2004, Belgian companies will have complete freedom of using their own Bonus-Malus systems. The years 2002 and 2003 are transition periods. Companies are obliged to use the old scale but may set up the relative premiums they want. A company that wishes to abandon the bonus-malus system may do so by imposing the same relativities to each step of the scale. In order to guarantee public information (which allows the new insurer to form an idea of the recent claims history of a policyholder who wishes to change insurer), which was guaranteed by the position occupied in the uniform scale, the legislator provided that in the event of cancellation of the policy, the insurer will have to provide the insured who leaves with a certificate describing his claims history of the past 5 years.

The aim of this article is to build a ratemaking for a given portfolio of policyholders. This ratemaking occurs in two stages, initially an *a priori* risk classification is carried out and next, on the basis of the example of the Belgian scale, the *a posteriori* ratemaking can be built. The latter must depend on the degree of the *a priori* ratemaking and the more precise the *a priori* ratemaking is, the weaker the amplitude of the *a posteriori* corrections are.

Moreover, the liberalization of the *a posteriori* ratemaking will allow the insurer to use the latter as a marketing tool. Nothing will indeed prevent him from proposing different Bonus-Malus scales to different categories of policyholders in order to attract one category rather than another.

The paper is further organized as follows. Section 2 describes the used data portfolio. Section 3 displays the results of the *a priori* ratemaking carried out within the Poisson regression framework. Section 4 shows how to build a Bonus-Malus scale depending on the *a priori* risk classification carried out preliminary. Moreover, on the basis of a few examples we will also demonstrate how a Bonus-Malus system can be used as a marketing tool. Section 5 shows how to adapt formulas when there are some constraints on relativities due to a path-dependent bonus rule. Section 6 will allow us to conclude.

2 Data presentation.

The data used to illustrate this paper relate to a Belgian MTPL portfolio observed during the year 1997. The data set comprises 14 505 policies. For each of those we know some characteristics of the policyholder and the number of claims he filed. The overall mean claims frequency is 14.6%.

The retained covariates used to build the *a priori* tariff are the following : the gender of the driver (male-female), the age of the driver (four classes : $18-24$ years, $25-30$ years, $31-60$ years and >60 years), the use of the vehicle (leisure and commuting only or also professional), the kind of district where the policyholder lives (urban ou rural) and whether the policyholder spreads the payment of the premium or not. We decided to choose only a few risk factors in order to simplify the presentation of the results.

Figures 2.1 to 2.5 display histograms giving, for each of the chosen risks factors, the frequency of each level of the classification variable and, for each level, the average claims frequency (in %).

These histograms require some comments. We can observe that women, who represent only 35% of the policyholders, have a lower claims frequency (13.6% against 15.2% for men). Young people show a markedly higher claims frequency. Claims frequencies seem to decrease according to age, going from 21.3% to 15.5%, then to 12.3% and finally to 10.8%. Regarding the use of the vehicle, we do not observe a big difference between private and professional

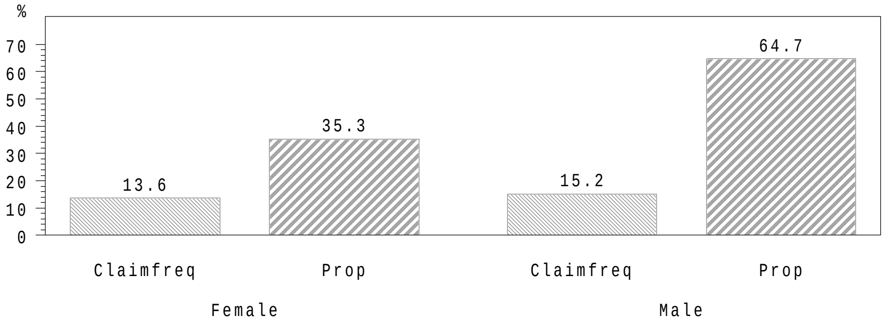

Figure 2.1: Histogram and claims frequency according to the gender of the driver.

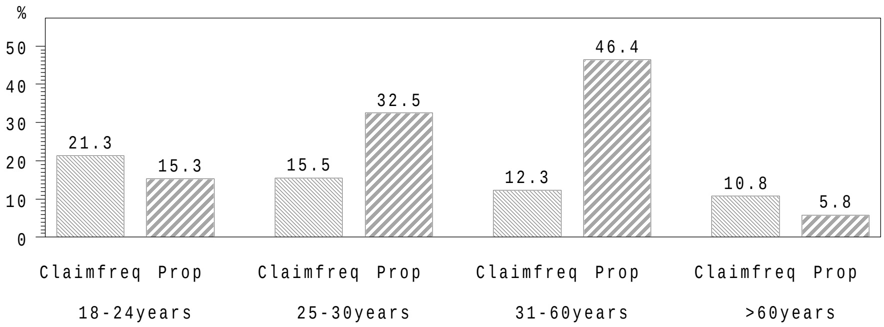

Figure 2.2: Histogram and claims frequency according to the age of the driver.

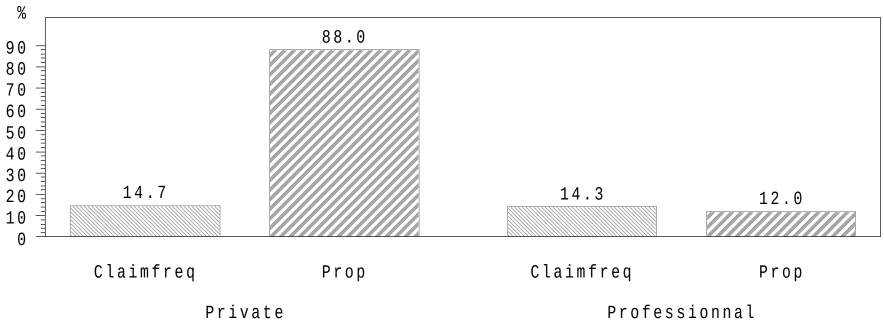

Figure 2.3: Histogram and claims frequency according to the use of the vehicle.

use. People living in a urban district report more claims than those living in a rural district. Finally, it seems that people who spread the payment of the premium show a particularly high claims frequency.

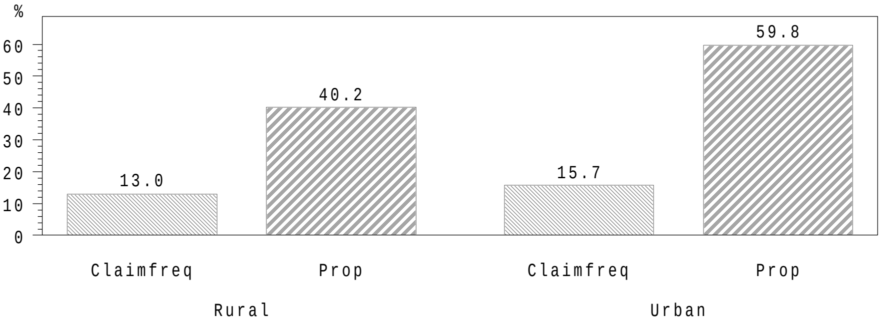

Figure 2.4: Histogram and claims frequency according to the kind of district.

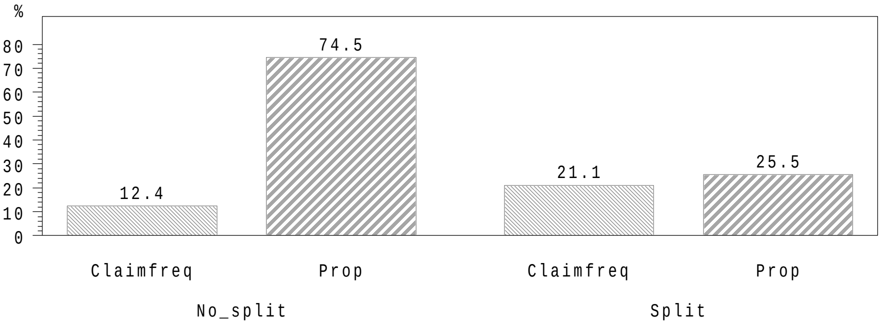

Figure 2.5: Histogram and claims frequency according to whether the premium payment is split or not.

3 *A priori* ratemaking.

The analysis of the number of claims will be performed within a Poisson regression framework. Let N_i represent the number of claims incurred by policyholder i during the observed period, $i = 1, 2, \ldots, n$. Let d_i represent the length of the observation period for policyholder i. We assume that we have some other variables $\boldsymbol{x}_i$ at disposal, known at the beginning of the period, and that these covariates are useful to explain the claims frequency of the policyholder i. Each time the value of a covariate is modified, a new period begins, d_i can therefore be different from 1.

We assume that, given $\boldsymbol{x}_i$, N_i follow a Poisson distribution, i.e.

$$N_i =_d \text{Poisson}\,\left(d_i \exp(\boldsymbol{\beta}^t \boldsymbol{x}_i)\right),\ \ i = 1, 2, \ldots, n. \tag{3.1}$$

From now, we define as $\eta_i = \boldsymbol{\beta}^t \boldsymbol{x}_i$ the linear predictor, also called score because we can sort the policyholders from those with a small risk factor to those with a high risk factor, according to increasing values of η_i. The annual expected claims frequency for policyholder i is $\lambda_i = d_i \exp(\eta_i)$. We obtain the pure premium by multiplying λ_i by the average cost of a claim.

The GENMOD procedure of SAS allows us to get the maximum likelihood estimators for the vector $\boldsymbol{\beta}$. We will here work with a combination of the variables gender and age. Indeed, if we try to build a model that considers these two variables separately, we find that the variable gender is not significant and that we should eliminate it. On the other hand, the interaction gender-age is actually significant as we will see and provides a better model (according to the loglikelihood) than a model that merely takes into account the effect of age and not of gender. The latter will be clearly shown at the end of the section. Table 3.1 displays the results of the Poisson regression.

Variable	Level	Coeff β	Std Error	Wald 95% Conf Limit		Chi-Sq	Pr>ChiSq
Intercept		-2.2131	0.0582	-2.3271	-2.0991	1447.40	$< .0001$
Gender*Age	Female $18-24$ years	0.3072	0.1117	0.0883	0.5261	7.57	0.0059
Gender*Age	Female $25-30$ years	0.1620	0.0876	-0.0098	0.3337	3.42	0.0646
Gender*Age	Female $31-60$ years	0.0651	0.0802	-0.0920	0.2222	0.66	0.4166
Gender*Age	Female >60 years	-0.0010	0.2350	-0.4616	0.4596	0.00	0.9967
Gender*Age	Male $18-24$ years	0.6429	0.0797	0.4867	0.7990	65.10	$< .0001$
Gender*Age	Male $25-30$ years	0.2875	0.0713	0.1477	0.4273	16.24	$< .0001$
Gender*Age	Male >60 years	-0.0623	0.1425	-0.3416	0.2170	0.19	0.6621
Gender*Age	Male $31-60$ years	0	0	0	0	.	.
Kind of district	Rural	-0.1828	0.0503	-0.2814	-0.0842	13.19	0.0003
Kind of district	Urban	0	0	0	0	.	.
Split of payment	Yes	0.4615	0.0515	0.3607	0.5624	80.43	$< .0001$
Split of payment	No	0	0	0	0	.	.
Use of vehicle	Professional use	0.2213	0.0784	0.0677	0.3749	7.98	0.0047
Use of vehicle	Leisure and commuting	0	0	0	0	.	.

Table 3.1: Results of the Poisson regression for a model taking into account the 5 variables.

The column "Coeff β" gives the estimation of the parameter β_j corresponding to each level of each covariate used in the model.
Columns "Chi-Sq" and "Pr>ChiSq", which is the corresponding p-value, allow us to test whether the corresponding coefficient β_j is significantly different from 0. We reject the null hypothesis $\beta_j = 0$ when the p-value is smaller than 5%.
The value of the loglikelihood for this model taking into account the 5 covariates with all their levels is $-5\,339.0$. Table 3.2 displays the results of the Type 3 analysis. This analysis allows us to examine the contribution of each covariate compared to a model not taking account of it.

Source	DF	ChiSquare	Pr > ChiSq
Gender*Age	7	75.16	$< .0001$
Kind of district	1	13.41	0.0003
Split of payment	1	77.27	$< .0001$
Use of vehicle	1	7.62	0.0058

Table 3.2: Results of the Type 3 analysis for the model taking into account the 5 variables.

When examining Tables 3.1 and 3.2 we may observe that even if all the covariates are statistically significant, some levels of the interaction gender-age may be omitted without deteriorating the quality of the model. We thus will combine some of these levels. For that purpose, we can use the CONTRAST (or ESTIMATE) option of the GENMOD procedure of SAS. This option allows us to test whether the difference between two levels of one variable is significant or not. This analysis leads us to combine women over 60 with men between 31 and 60 (p-value of 99%) and then similarly with men over 60 (p-value of 66%). We then add women between 31 and 60 to this group made up of men over 31 and of women over 60 (p-value of 35%). We can still combine into only one level women from 18 up to 24 and those from 25 up to 30 (p-value of 23%) and finally the behaviour of women from 18 up to 30 is not significantly different from that of men from 25 up to 30 (p-value of 30%). We then obtain the model described in Table 3.3.

Variable	Level	Coeff β	Std Error	Wald 95% Conf Limit		Chi-Sq	Pr>ChiSq
Intercept		-2.1975	0.0466	-2.2888	-2.1062	2225.51	$< .0001$
Gender*Age	Female $18-30$ + Male $25-30$	0.2351	0.0538	0.1297	0.3405	19.10	$< .0001$
Gender*Age	Male $18-24$	0.6235	0.0719	0.4826	0.7644	75.24	$< .0001$
Gender*Age	Female > 30 + Male > 30	0	0	0	0	.	.
District	Rural	-0.1809	0.0503	-0.2795	-0.0823	12.94	0.0003
District	Urban	0	0	0	0	.	.
Split payment	Yes	0.4677	0.0513	0.3671	0.5683	83.06	$< .0001$
Split payment	No	0	0	0	0	.	.
Use of vehicle	Professional	0.2150	0.0776	0.0630	0.3670	7.69	0.0056
Use of vehicle	Leisure and commuting	0	0	0	0	.	.

Table 3.3: Results of the Poisson regression for the final model.

The loglikelihood is equal to $-5\,340.7$ and Table 3.4 provides the results of the Type 3 analysis. All the retained covariates and all the levels of the interaction gender-age are now statistically significant.

Source	DF	ChiSquare	Pr > ChiSq
Gender*Age	2	71.66	$< .0001$
District	1	13.15	0.0003
Split of payment	1	79.68	$< .0001$
Use of vehicle	1	7.34	0.0067

Table 3.4: Results of the Type 3 analysis for the final model.

As previously announced, we still have to say a few words about the comparison between the retained model, which takes into account the interaction gender-age, and another model taking into account these two variables separately. For that purpose, we will use the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) criteria :

$$AIC = l - dim \text{ and } BIC = l - \frac{1}{2}\ln(n)dim,$$

where l is the loglikelihood, dim is the number of parameters to estimate and n is the sample size. Both of these criteria penalize the loglikelihood in case of overparametrization. Moreover, the BIC criterion takes account of the sample size.

In our example, the retained model with interaction gender-age leads to a loglikelihood equal to $-5\,340.7$ for 6 estimated parameters and a sample size equal to $14\,505$. The values of the AIC and BIC criteria are respectively −5 346.7 and −5 369.5. If we start with a model integrating the variables age and gender separately, the latter must be eliminated because it is not significant and two age groups of the four must be combined. We then obtain a model with 6 parameters providing a loglikelihood of −5 345.0. The values of the AIC and BIC criteria are respectively −5 351.0 and −5 373.8. Let us notice that here, as we compare models with the same number of parameters, criteria AIC and BIC have any added value compared to a simple comparison of loglikelihood and we can easily conclude that the model with interaction gender-age is better.

4 Building a Bonus-Malus scale.

Due to the fact that many important factors (such as aggressiveness behind the wheel or swiftness of reflexes) cannot be taken into account in the *a priori* segmentation performed in the previous section, risk classes remain heterogeneous. This residual heterogeneity can be represented by a random effect Θ_i superposed to the annual claims frequency. Specifically, given $\Theta_i = \theta$, the annual numbers of claims N_i are assumed to be independent and to conform to a Poisson distribution with mean $\lambda_i\theta$, i.e.

$$\Pr[N_i = k|\Theta_i = \theta] = \exp(-\lambda_i\theta)\frac{(\lambda_i\theta)^k}{k!}, \;\; k \in \mathbb{N}.$$

Moreover, all the Θ_i's are assumed to be independent and to follow a standard Gamma distribution with probability density function

$$u(\theta) = \frac{1}{\Gamma(a)} a^a \theta^{a-1} \exp(-a\theta), \;\; \theta \in \mathbb{R}^+.$$

Since $\mathbb{E}[\Theta_i] = 1$ we have that $\mathbb{E}[N_i] = \lambda_i$; λ_i is the expected number of claims for a policyholder for which no information about past claims is available.

The premium can then be adjusted over time with the help of credibility techniques. But the results of these techniques cannot be enforced in practice for MTPL, essentially because of commercial reasons and legal constraints. Instead, companies resort to BM scales, that may be considered as simplified versions of credibility theory. Such scales possess a number of levels, $s + 1$ say, numbered from 0 to s. The policyholders move from one level to another according to the number of claims they report. The relativity associated to level ℓ is r_ℓ; the meaning is that an insured occupying that level pays an amount of premium equals to r_ℓ% of the *a priori* premium determined on the basis of his observable characteristics.

A specified level is assigned to a new driver (often according to the use of the vehicle). Each claims free year is rewarded by a bonus point (i.e. the driver goes down one level). Claims are penalized by malus points (i.e. the driver goes up a certain number of levels each time he files a claim). After a sufficient number of claims free years, the driver enters level 0 where he enjoys the maximal bonus.

In commercial BMS, the knowledge of the present level and of the number of claims of the present year suffice to determine the next level. 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 claims 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. Indeed, in some BMS, policyholders occupying the highest levels are sent to the starting level after a few claim-free years.

In this paper we will concentrate on the old Belgian Bonus-Malus system. This system consists of a scale of 23 levels (numbered from 0 to 22). The level 0 provides the maximal bonus and the relativities are increasing up to the level 22 where the malus is maximal. Business users enter the system in level 14 whereas commuters and pleasure users enter in level 11.

The transition rules are the following : each year a one-level bonus is given and each claim is penalized by 5 levels up.

Actually, the Belgian BMS is not Markovian due to the fact that policyholders occupying high levels are sent to level 14 after 4 claim-free years. Fortunately, it is possible to introduce fictitious classes in order to meet the memoryless property. Lemaire (1995) proposed to split the classes 16 to 21 into subclasses, depending on the number of consecutive years without accident. The total number of levels is then 35 instead of 23. In the rest of this section we will ignore this special bonus rule. This will be treated into details in the section 5.

Let us define some useful quantities : let $p_{\ell_1\ell_2}(\vartheta)$ be the probability of moving from level ℓ_1 to level ℓ_2 for a policyholder with mean frequency ϑ. Further, $p^{(\nu)}_{\ell_1\ell_2}(\vartheta)$ is the probability of moving from level ℓ_1 to level ℓ_2 in ν transitions. If the stationary distribution exists, and this is the case for the Markov chain associated to the Belgian system, this distribution writes

$$\pi_{\ell_2}(\vartheta) = \lim_{\nu\to+\infty} p^{(\nu)}_{\ell_1\ell_2}(\vartheta).$$

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

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

where $\boldsymbol{e}$ is a column vector of 1's and $\boldsymbol{M}(\vartheta)$ is the one-step transition matrix, i.e. $\boldsymbol{M}(\vartheta) = \{p_{\ell_1\ell_2}(\vartheta)\}$, $\ell_1, \ell_2 = 0, 1, \ldots, s$. Let $\boldsymbol{E}$ be the $(s+1)\times(s+1)$ matrix all of whose entries are 1, i.e. consisting of $s+1$ column vectors $\boldsymbol{e}$. Then, it can be shown that

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

which provides a direct method to get $\boldsymbol{\pi}(\vartheta)$. For a derivation of the latter result, see e.g. Rolski et al. (1999).

Let us now introduce the random variable L_ϑ valued in $\{0, 1, \ldots, s\}$ such that L_ϑ conforms to the distribution $\boldsymbol{\pi}(\vartheta)$ i.e.

$$\Pr[L_\vartheta = \ell] = \pi_\ell(\vartheta), \;\; \ell = 0, 1, \ldots, s.$$

The variable L_ϑ thus represents the level occupied by a policyholder with annual claims frequency ϑ once the stationary state has been reached.

Let us now pick at random a policyholder from the portfolio. Let us denote as Λ his (unknown) *a priori* claims frequency and as Θ the residual effect of the risk factors not included in the ratemaking. The actual (unknown) annual expected claims frequency of this policyholder is then $\Lambda\Theta$. Since the random effect Θ represents residual effects of hidden covariates, the random variables Λ and Θ may reasonably be assumed to be mutually independent. Let w_k be the weight of the kth risk class whose annual claims frequency is λ_k. Clearly, $\Pr[\Lambda = \lambda_k] = w_k$. The distribution of L writes

$$\Pr[L = \ell] = \sum_k w_k \int_{\theta \geq 0} \pi_\ell(\lambda_k \theta) u(\theta) d\theta; \tag{4.1}$$

$\Pr[L = \ell]$ represents the proportion of the policyholders in level ℓ.

The interaction between *a priori* ratemaking and BMS may be checked by examining

$$\mathbb{E}[\Lambda | L = \ell] \;=\; \frac{\sum_k \lambda_k w_k \int_{\theta \geq 0} \pi_\ell(\lambda_k \theta) u(\theta) d\theta}{\sum_k w_k \int_{\theta \geq 0} \pi_\ell(\lambda_k \theta) u(\theta) d\theta}. \tag{4.2}$$

Indeed, if $\mathbb{E}[\Lambda | L = \ell]$ is increasing in level ℓ, it means that policyholders with low *a priori* expected claim frequencies will tend to gravitate in the lowest levels of the scale, and conversely for individuals with high *a priori* expected claim frequencies. So, those policyholders who have been granted premium discounts at policy issuance (on the basis of their observable characteristics) will be also rewarded *a posteriori* (because they occupy the lowest levels of the bonus-malus scale). Conversely, the policyholders who have been penalized at policy issuance (because of their observable characteristics) will cluster in the highest bonus-malus levels and will consequently be penalized again. The level occupied by the policyholders in the Bonus-Malus scale can thus be partly explained by their observable characteristics included in the price list. It is thus fair to isolate that part of the information contained in the level occupied by the policyholder that does not reflect observable characteristics. *A posteriori* corrections should be only driven by this part of the Bonus-Malus information. This phenomenon was put in evidence by TAYLOR (1997) and we here propose an analytic analogue to the simulation procedure he described.

According to Norberg's criterion, the relativities associated to each level will be obtained by minimizing the expected squared difference between the "true" relative premium Θ and the relative premium r_L applicable to a policyholder in the system (after the stationary state has been reached). The goal is thus to minimize $\mathbb{E}\Big[(\Theta - r_L)^2\Big]$.

The solution is given by

$$r_\ell \;=\; \frac{\sum_k w_k \int_{\theta \geq 0} \theta \pi_\ell(\lambda_k \theta) u(\theta) d\theta}{\sum_k w_k \int_{\theta \geq 0} \pi_\ell(\lambda_k \theta) u(\theta) d\theta}. \tag{4.3}$$

The interested reader will find more details about the derivation of the formula (4.3) in PITREBOIS ET AL. (2003a).

To end with, let us mention that if the insurance company does not enforce any *a priori* ratemaking system, all the λ_k's are equal to $\overline{\lambda}$ and (4.3) reduces to the formula

$$r_\ell = \frac{\int_{\theta \geq 0} \theta \pi_\ell(\overline{\lambda}\theta) u(\theta) d\theta}{\int_{\theta \geq 0} \pi_\ell(\overline{\lambda}\theta) u(\theta) d\theta} \tag{4.4}$$

that has been derived in NORBERG (1976). In practice the $u(\theta)$ from formulas (4.3) and (4.4) are different, because the estimation of parameter a is not the same in the model with *a priori* ratemaking as it is in the model without *a priori* ratemaking.

4.1 Numerical results for the Belgian Bonus-Malus system.

The numerical results for the Belgian Bonus-Malus system and for the portfolio described in section 2 are displayed in Table 4.1. Specifically, the values in the third column are computed with the help of (4.4) with $\widehat{a} = 0.8888$ and $\widehat{\lambda} = 0.1474$. Those values were obtained by fitting a Negative Binomial distribution to the portfolio's observed claims frequencies. The fifth column is based on (4.3) with $\widehat{a} = 1.2401$ and the $\widehat{\lambda}_k$'s obtained from the *a priori* risk classification described in section 3. The last column is computed with the help of (4.2). Once the stationary state has been reached, more or less half of the policies occupy level 0 and enjoy the maximum discount. If we build the Bonus-Malus scale without taking into account the *a priori* ratemaking, the relativities vary between 37.5% for level 0 and 306.0% for level 22. On the other hand, if we adapt the relativities to the *a priori* risk classification, these relativities vary between 48.8% and 256.0%; the severity of the *a posteriori* corrections is thus weaker in this case.

Level ℓ	$\Pr[L=\ell]$ without *a priori* ratemaking	Relativity $r_\ell = \mathbb{E}[\Theta \mid L=\ell]$ without *a priori* ratemaking	$\Pr[L=\ell]$ with *a priori* ratemaking	Relativity $r_\ell = \mathbb{E}[\Theta \mid L=\ell]$ with *a priori* ratemaking	Average *a priori* expected claims frequency in level ℓ $\mathbb{E}[\Lambda \mid L=\ell]$ with *a priori* ratemaking
22	5.6%	306.0%	5.0%	256.0%	18.3%
21	4.0%	273.9%	3.7%	235.9%	17.6%
20	3.0%	248.7%	2.8%	220.0%	17.1%
19	2.4%	228.2%	2.3%	206.8%	16.7%
18	2.0%	211.2%	1.9%	195.5%	16.4%
17	1.7%	196.5%	1.7%	185.4%	16.2%
16	1.5%	183.8%	1.5%	176.3%	15.9%
15	1.4%	172.6%	1.4%	168.1%	15.7%
14	1.3%	162.5%	1.3%	160.4%	15.6%
13	1.2%	152.8%	1.3%	152.7%	15.4%
12	1.2%	143.9%	1.3%	145.4%	15.3%
11	1.2%	135.9%	1.3%	138.8%	15.1%
10	1.2%	128.9%	1.3%	132.9%	15.0%
9	1.3%	119.8%	1.5%	124.9%	14.9%
8	1.5%	111.1%	1.7%	117.0%	14.7%
7	1.6%	104.8%	1.8%	111.4%	14.6%
6	1.7%	99.8%	1.9%	106.9%	14.5%
5	1.7%	95.6%	1.9%	103.1%	14.5%
4	4.0%	75.2%	4.5%	84.0%	14.2%
3	3.6%	72.7%	4.0%	81.6%	14.1%
2	3.3%	70.3%	3.6%	79.3%	14.1%
1	2.9%	68.0%	3.3%	77.1%	14.1%
0	50.4%	37.5%	49.1%	48.8%	13.7%

Table 4.1: Numerical characteristics for the Belgian system.

4.2 Interaction between *a priori* ratemaking and Bonus-Malus scale.

The *a posteriori* ratemaking we are building here depends on the accuracy of the *a priori* ratemaking carried out preliminary. We can check this by a simple example. If we use a less accurate *a priori* risk classification, we can expect to obtain a Bonus-Malus scale with a severity in between a scale without risk classification and another based on the risk classification described in section 3. This is what we can observe in Table 4.2 where we have built a scale taking account of an *a priori* ratemaking based only on the interaction gender-age and the kind of district where the policyholder lives. We decided thus not to use the criteria split of premium payment and use of vehicle. The relativities corresponding to this less accurate risk classification vary between 45.4% and 270.1% whereas for the first risk classification we built, they varied between 48.8% and 256.0%.

Level ℓ	$\Pr[L=\ell]$	Relativity $r_\ell = \mathbb{E}[\Theta\|L=\ell]$ with *a priori* ratemaking	Average *a priori* expected claims frequency in level ℓ $\mathbb{E}[\Lambda\|L=\ell]$ with *a priori* ratemaking
22	5.1%	270.1%	16.6%
21	3.7%	247.3%	16.3%
20	2.9%	229.0%	16.0%
19	2.3%	213.8%	15.8%
18	2.0%	200.8%	15.6%
17	1.7%	189.3%	15.5%
16	1.5%	179.5%	15.3%
15	1.4%	169.8%	15.2%
14	1.3%	161.3%	15.1%
13	1.3%	152.9%	15.0%
12	1.3%	145.0%	15.0%
11	1.3%	137.9%	14.9%
10	1.3%	131.6%	14.8%
9	1.5%	123.1%	14.7%
8	1.7%	115.0%	14.7%
7	1.8%	109.0%	14.6%
6	1.9%	104.4%	14.6%
5	1.9%	100.5%	14.5%
4	4.4%	81.0%	14.4%
3	3.9%	78.6%	14.3%
2	3.5%	76.3%	14.3%
1	3.2%	74.1%	14.3%
0	49.1%	45.4%	14.1%

Table 4.2: Numerical characteristics for the Belgian system based on a partial risk classification.

4.3 Different Bonus-Malus scales depending on observable characteristics.

When a Bonus-Malus system is in force, the same *a posteriori* corrections apply to all policyholders, whatever their *a priori* claims frequency. This of course induces unfairness in the portfolio. In order to reduce the unfairness of the tariff, we could propose several BM scales, according to the *a priori* characteristics.

For example, here we consider that the company differentiates policyholders according to the type of district they live in (urban or rural). People living in urban areas have higher *a priori* claim frequencies. They should therefore be more rewarded in case they do not file any claim and less penalized when they report accidents compared to people living in rural zones. This is indeed what we observe when comparing the relative premiums obtained in Tables 4.3 et 4.4 : the maximal discount is 46.2% for urban policyholders, compared to 52.0% for rural ones. Similarly, the highest penalty is 250.4% for urban policyholders against 268.2% for rural ones.

Urban			
Level ℓ	$\Pr[L=\ell]$	Relativity $r_\ell = \mathbb{E}[\Theta \mid L=\ell]$ with *a priori* ratemaking	Average *a priori* expected claims frequency in level ℓ $\mathbb{E}[\Lambda \mid L=\ell]$ with *a priori* ratemaking
22	5.8%	250.4%	19.2%
21	4.1%	229.4%	18.5%
20	3.2%	213.1%	18.0%
19	2.5%	199.6%	17.7%
18	2.1%	188.1%	17.3%
17	1.8%	177.9%	17.1%
16	1.6%	168.8%	16.9%
15	1.5%	160.6%	16.7%
14	1.4%	153.0%	16.5%
13	1.4%	145.5%	16.4%
12	1.3%	138.3%	16.2%
11	1.3%	131.9%	16.1%
10	1.4%	126.2%	16.0%
9	1.5%	118.4%	15.8%
8	1.7%	110.9%	15.7%
7	1.8%	105.5%	15.6%
6	1.9%	101.2%	15.5%
5	1.9%	97.5%	15.5%
4	4.4%	79.6%	15.2%
3	3.9%	77.3%	15.2%
2	3.5%	75.1%	15.1%
1	3.2%	73.0%	15.1%
0	46.7%	46.3%	14.7%

Table 4.3: Numerical characteristics of the Belgian system for urban policyholders.

Rural			
Level ℓ	$\Pr[L = \ell]$	Relativity $r_\ell = \mathbb{E}[\Theta \mid L = \ell]$ with *a priori* ratemaking	Average *a priori* expected claims frequency in level ℓ $\mathbb{E}[\Lambda \mid L = \ell]$ with *a priori* ratemaking
22	3.9%	268.2%	16.3%
21	3.0%	249.2%	15.7%
20	2.3%	233.8%	15.3%
19	1.9%	220.9%	15.0%
18	1.7%	209.5%	14.7%
17	1.5%	199.2%	14.5%
16	1.3%	189.9%	14.3%
15	1.3%	181.3%	14.1%
14	1.2%	173.2%	13.9%
13	1.2%	165.1%	13.8%
12	1.2%	157.3%	13.7%
11	1.2%	150.2%	13.5%
10	1.3%	143.9%	13.4%
9	1.4%	135.0%	13.3%
8	1.6%	126.5%	13.2%
7	1.8%	120.4%	13.1%
6	1.9%	115.5%	13.0%
5	1.9%	111.4%	12.9%
4	4.6%	90.3%	12.7%
3	4.1%	87.8%	12.6%
2	3.7%	85.3%	12.6%
1	3.4%	83.0%	12.6%
0	52.7%	52.0%	12.2%

Table 4.4: Numerical characteristics of the Belgian system for rural policyholders.

4.4 A more severe alternative to the current Belgian system.

Too many policyholders occupy the lowest levels of the Belgian Bonus-Malus scale. Consequently the system is not financially balanced. Let us consider another scale with still 23 levels but with higher penalty for a filed claim. We assume that each claim reported by the policyholder is now penalized by 7 levels instead of 5 in the actual scale. We still compare the results of a scale not based on an *a priori* ratemaking and another one taking into account the *a priori* risk classification described in section 3. Table 4.5 displays the results. Within this new scale, the proportion of policyholders occupying level 0 at stationary state is smaller, 39.3% instead of 49.1% for the scale +5. The relative premiums computed without taking into account the *a priori* ratemaking vary between 30.5% and 266.5% instead of respectively 37.5% and 306.0% for the scale +5. In case of a scale based on the *a priori* risk classification carried out in section 3, the relativities vary between 40.9% and 226.4% instead of 48.8% and 256.0 for the scale +5. For each level the relativity is smaller in scale +7 than in scale +5.

Level ℓ	$\Pr[L=\ell]$ without *a priori* ratemaking	Relativity $r_\ell = \mathbb{E}[\Theta \mid L=\ell]$ without *a priori* ratemaking	$\Pr[L=\ell]$ with *a priori* ratemaking	Relativity $r_\ell = \mathbb{E}[\Theta \mid L=\ell]$ with *a priori* ratemaking	Average *a priori* expected claims frequency in level ℓ $\mathbb{E}[\Lambda \mid L=\ell]$ with *a priori* ratemaking
22	7.7%	266.5%	7.2%	226.4%	17.5%
21	5.6%	236.3%	5.4%	206.7%	16.9%
20	4.3%	212.5%	4.2%	191.1%	16.4%
19	3.4%	192.9%	3.4%	177.9%	16.1%
18	2.9%	176.4%	2.9%	166.4%	15.8%
17	2.4%	162.4%	2.5%	156.3%	15.5%
16	2.1%	150.3%	2.3%	147.3%	15.3%
15	1.9%	140.0%	2.0%	139.4%	15.2%
14	1.8%	131.0%	1.9%	132.3%	15.0%
13	1.7%	121.3%	1.8%	124.4%	14.9%
12	1.7%	111.9%	1.8%	116.4%	14.7%
11	1.7%	104.5%	1.8%	109.9%	14.6%
10	1.6%	98.3%	1.8%	104.6%	14.5%
9	1.6%	93.2%	1.8%	100.1%	14.4%
8	1.6%	88.8%	1.7%	96.1%	14.4%
7	1.5%	84.9%	1.7%	92.6%	14.3%
6	3.1%	66.3%	3.4%	75.1%	14.0%
5	2.8%	64.1%	3.1%	73.0%	14.0%
4	2.5%	62.0%	2.8%	71.0%	14.0%
3	2.3%	60.0%	2.6%	69.1%	13.9%
2	2.1%	58.1%	2.3%	67.3%	13.9%
1	2.0%	56.4%	2.1%	65.6%	13.9%
0	41.7%	30.5%	39.3%	40.9%	13.5%

Table 4.5: Numerical characteristics of the modified Belgian system (-1/+7).

5 Special bonus rule.

As announced in section 4 we still have to derive formulas taking account of the special bonus rule of the Belgian Bonus-Malus system. This rule implies that policyholders occupying high levels are sent to level 14 after 4 claim-free years. Due to this rule the Belgian BMS is not Markovian. Fortunately, it is possible to introduce fictitious classes in order to meet the memoryless property. Lemaire (1995) proposed to split the classes 16 to 21 into subclasses, depending on the number of consecutive years without accident. This authorizes to take account of the special bonus rule. Let n_j be the number of subclasses to be associated to bonus class j. A class $j.i$ is to be understood as level j and i consecutive years without accidents. The transition rules are completely defined in Table 5.1 and the different values for n_j are given in Table 5.2. We take some liberty with the notations by using the value 0 for the subscript i and by not using a subscript when $n_j = 1$.

All the notations of the section 4 hold but with index $j.i$ instead of ℓ. According to Norberg's criterion, the relativities, $r_{j.i}$ will be obtained by minimizing the squared difference between the true relative premium Θ and the relative premium r_L applicable to the policyholder when stationary state has been reached.

The current situation is more complicated because some states have to be constrained to have the same relativity. Indeed the artificial states $j.i$ have the property that

$$r_j = r_{j.1} = \cdots = r_{j.n_j} \quad , \quad j = 0, \ldots, s.$$

This point has been addressed by Centeno et al. (2002) in a framework without *a priori* segmentation. We extend it here in the general framework.

Class	Class after k accidents					
k	0	1	2	3	4	5
22	21.1	22	22	22	22	22
21.0	20.1	22	22	22	22	22
21.1	20.2	22	22	22	22	22
20.0	19.1	22	22	22	22	22
20.1	19.2	22	22	22	22	22
20.2	19.3	22	22	22	22	22
19.0	18.1	22	22	22	22	22
19.1	18.2	22	22	22	22	22
19.2	18.3	22	22	22	22	22
19.3	14	22	22	22	22	22
18.0	17	22	22	22	22	22
18.1	17.2	22	22	22	22	22
18.2	17.3	22	22	22	22	22
18.3	14	22	22	22	22	22
17	16	21.0	22	22	22	22
17.2	16.3	21.0	22	22	22	22
17.3	14	21.0	22	22	22	22
16	15	20.0	22	22	22	22
16.3	14	20.0	22	22	22	22
15	14	19.0	22	22	22	22
14	13	18.0	22	22	22	22
13	12	17	22	22	22	22
12	11	16	21.0	22	22	22
11	10	15	20.0	22	22	22
10	9	14	19.0	22	22	22
9	8	13	18.0	22	22	22
8	7	12	17	22	22	22
7	6	11	16	21.0	22	22
6	5	10	15	20.0	22	22
5	4	9	14	19.0	22	22
4	3	8	13	18.0	22	22
3	2	7	12	17	22	22
2	1	6	11	16	21.0	22
1	0	5	10	15	20.0	22
0	0	4	9	14	19.0	22

Table 5.1: Transition rules of the Belgian BMS

j	n_j
22	1
21	2
20	3
19	4
18	4
17	3
16	2
15	1
14	1
13	1
12	1
11	1
10	1
9	1
8	1
7	1
6	1
5	1
4	1
3	1
2	1
1	1
0	1

Table 5.2: Subclasses of the Belgian BMS

The Norberg's criterion becomes a minimization under constraints :

$$\min \mathbb{E}\left[(\Theta - r_L)^2\right]$$

such that

$$r_j = r_{j.1} = r_{j.2} = r_{j.n_j} \quad , \quad j = 0, 1, \ldots, s.$$

The solution is given by

$$r_j = \frac{\sum\limits_k w_k \int\limits_0^\infty \sum_{i=1}^{n_j} \theta \pi_{j.i}(\lambda_k \theta) u(\theta) d\theta}{\sum\limits_k w_k \int\limits_0^\infty \sum_{i=1}^{n_j} \pi_{j.i}(\lambda_k \theta) u(\theta) d\theta}. \tag{5.1}$$

The interested reader will find more details about the derivation of the formula (5.1) in PITREBOIS ET AL. (2003b).

This formula extends Norberg (1976), Centeno et al. (2002) and the result of section 4. Indeed

1. Setting $n_j = 1 \quad \forall j$ gives the formula of section 4.
2. Setting $\lambda_k = \lambda \quad \forall k$ gives the formula of Centeno et al. (2002).
3. Combining the previous two particular cases gives the formula of Norberg (1976).

Note that it is easily seen that

$$r_j = \sum_{i=1}^{n_j} \frac{\mathbb{P}[L = j.i]}{\sum_{i=1}^{n_j} \mathbb{P}[L = j.i]} r_{j.i},$$

where the $r_{j.i}$'s represent the non constrained solution of $\min \mathbb{E}\left[(\Theta - r_L)^2\right]$.

We also immediately verify that

$$\sum_{j=0}^{s} r_j \mathbb{P}[L = j] = 1$$

which ensures that the BMS is financially balanced at stationary state.

Tables 5.3, 5.4 and 5.5 display the results for the Belgian Bonus-Malus system with the special bonus rule and for the *a priori* ratemaking described in section 3.

$\Pr[L = j]$		
Level j	Without Special Bonus Rule	With Special Bonus Rule
22	5.0%	4.0%
21	3.7%	2.8%
20	2.8%	2.1%
19	2.3%	1.7%
18	1.9%	0.9%
17	1.7%	1.0%
16	1.5%	1.0%
15	1.4%	1.0%
14	1.3%	2.2%
13	1.3%	2.0%
12	1.3%	1.8%
11	1.3%	1.8%
10	1.3%	1.7%
9	1.5%	1.8%
8	1.7%	2.0%
7	1.8%	2.1%
6	1.9%	2.1%
5	1.9%	2.1%
4	4.5%	4.7%
3	4.0%	4.2%
2	3.6%	3.8%
1	3.3%	3.4%
0	49.1%	49.7%

Table 5.3: Distribution of L with *a priori* ratemaking

Relativities		
Level j	Without Special Bonus Rule	With Special Bonus Rule
22	256.0%	267.5%
21	235.9%	246.1%
20	220.0%	228.8%
19	206.8%	214.4%
18	195.5%	193.6%
17	185.4%	188.8%
16	176.3%	182.1%
15	168.1%	175.0%
14	160.4%	186.7%
13	152.7%	175.2%
12	145.4%	164.7%
11	138.8%	155.4%
10	132.9%	147.2%
9	124.9%	137.0%
8	117.0%	127.2%
7	111.4%	120.0%
6	106.9%	114.3%
5	103.1%	109.6%
4	84.0%	88.2%
3	81.6%	85.3%
2	79.3%	82.7%
1	77.1%	80.1%
0	48.8%	49.9%

Table 5.4: Relativities $r_j = \mathbb{E}[\Theta|L = j]$ with *a priori* ratemaking

$\mathbb{E}[\Lambda \mid L = j]$		
Level j	Without Special Bonus Rule	With Special Bonus Rule
22	18.3%	18.7%
21	17.6%	17.9%
20	17.1%	17.4%
19	16.7%	17.0%
18	16.4%	16.4%
17	16.2%	16.3%
16	15.9%	16.1%
15	15.7%	15.9%
14	15.6%	16.2%
13	15.4%	15.9%
12	15.3%	15.7%
11	15.1%	15.5%
10	15.0%	15.3%
9	14.9%	15.1%
8	14.7%	14.9%
7	14.6%	14.8%
6	14.5%	14.7%
5	14.5%	14.6%
4	14.2%	14.2%
3	14.1%	14.2%
2	14.1%	14.1%
1	14.1%	14.1%
0	13.7%	13.7%

Table 5.5: Average *a priori* claims frequency in level ℓ, $\mathbb{E}[\Lambda \mid L = j]$ with *a priori* ratemaking

We observe that the average *a priori* expected claims frequency in level j is always higher with the special bonus rule than without that rule. The effect is more pronounced in the highest levels of the scale and less pronounced in the lowest levels of the scale. This fact is obvious from the definition of the special bonus rule. The policyholders attaining the highest classes of the scale benefit from the special bonus rule. Those staying in these highest classes show therefore a higher expected frequency. Even below level 14 the effect remains true because the policyholders have benefitted of it before attaining the lowest levels. Obvioulsy the effect is less and less pronounced at the bottom of the scale.

Some insurance companies use the bonus-malus scale as an acceptation tool. They e.g. systematically refuse drivers with a BM level > 14. Our calculations show that this is stupid because drivers at level 15 are less dangerous than drivers at level 14.

We observe that without the special bonus rule, the relativities are always increasing from level 0 to level 22. Although there is no certainty about that fact, it is logical and necessary from a commercial point of view. The same pattern is observed for $\mathbb{E}[\Lambda \mid L = j]$.

When looking at the results for the Bonus-Malus system with special bonus rule, we observe that the relativities at level 13, 14 and 15 are nor ordered anymore. This fact may be explained as follows : there are many drivers at level 14 that have benefitted of the special bonus rule, i.e. they have made lot's of claims but are sent back to level 14, which is not very much representative of their claims frequency. Level 13 which is attained from level 14 after a claim free year is also polluted by this fact.

It is clear that such a situation is not acceptable from a commerial point of view. We may constrain the scale to be linear in the spirit of Gilde and Sundt (1989). However we propose a local adjustement to the scale in order to keep $\mathbb{E}[(\Theta - r_L)^2]$ as small as possible.

Let us constrain the scale to be linear between levels 13 and 16. We are looking for updated value for r_j' , $j = 13, \ldots 16$. They are such that $r_j' = r_{j-1}' + a$, $j = 14, 15, 16$ where

$a = \frac{r'_{16} - r'_{13}}{3}$. We also want to keep the financial equilibrium of the system. Therefore we constrain a local equilibrium :

$$\sum_{j=13}^{16} r_j \pi_j = \sum_{j=13}^{16} r'_j \pi_j.$$

Choosing $r'_{13} = 177.00\%$, we obtain $r'_{16} = 185.69\%$.

Now let us compare the value of the expected error $Q = \mathbb{E}\left[(\Theta - r_L)^2\right]$ with the original model, Q_1, and with the constrained model, Q_2 :

$$\begin{aligned} Q_1 &= 0.36857 \\ Q_2 &= 0.36876 \end{aligned}$$

This shows that the error induced by the commercial constraint is really small. So we may adapt the scale without resorting to a full linear scale constraint.

We can perform numerically the local minimization without imposing a linear scale between levels 13 and 16. We use the following constraints :

$$\begin{aligned} r'_{13} &<= r'_{14}, \\ r'_{14} &<= r'_{15}, \\ r'_{15} &<= r'_{16}, \\ \sum_{j=13}^{16} r_j \pi_j &= \sum_{j=13}^{16} r'_j \pi_j, \\ r'_j &>= 165\% \quad j = 13, ..., 16, \\ r'_j &<= 188\% \quad j = 13, ..., 16. \end{aligned}$$

And we obtain $r'_{13} = 175.2\%$ and $r'_{14} = r'_{15} = r'_{16} = 182.7\%$. The value of Q is now 0.36867.

6 Conclusion.

In this article, we have shown how to build a simple Bonus-Malus system taking account of a path-dependent bonus rule or not and based on an *a priori* ratemaking and we saw how the latter influenced the scale obtained. Thus, a company can decide on the degree of *a priori* risk classification it retains by choosing a certain number and a certain type of risk variables and this will enable the company to propose to its policyholders a more or less severe *a posteriori* ratemaking. We also saw that the company can propose different scales to different policyholders and this in order to counter the inherent injustice of a uniform scale. But this idea could also be used to draw aside a class of risks regarded as too risky by proposing a very severe scale or on the other hand to attract another class of risks by proposing a favourable scale. To conclude, the liberalization of the Bonus-Malus system allows the insurance companies to use it as a real marketing tool.

References

[1] Centeno, M. and Silva, J.M.A. (2002) Optimal Bonus Scales Under Path-Dependent Bonus Rules, Scandinavian Actuarial Journal, 24, 615-627.

[2] Denuit, M., Pitrebois, S. and Walhin, J.-F. (2001). Méthodes de construction de systèmes Bonus-Malus en RC Auto. *Actu-L*, **1**, 7-38.

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

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

[5] Pitrebois, S., Denuit, M. and Walhin, J.F. (2003a). Setting a Bonus-Malus scale in the presence of other rating factors : Taylor's Work Revisited. Submitted.

[6] Pitrebois, S., Denuit, M. and Walhin, J.F. (2003b). Fitting the Belgian Bonus-Malus system. Submitted.

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

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