Optimal Relativities and Transition Rules of a Bonus-Malus System

Chong It Tan[a,*], Jackie Li[b], Johnny Siu-Hang Li[c], Uditha Balasooriya[a]

[a] *Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798.*

[b] *Department of Mathematics and Statistics, Curtin University, Perth, Australia.*

[c] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1.*

Abstract

When a bonus-malus system with a single set of optimal relativities and a simple transition rules is implemented, two inadequacy scenarios are induced because all policyholders are subject to the same *a posteriori* premium relativities (level transitions) independent of their *a priori* characteristics (current levels occupied). In this paper we propose a new objective function in the determination of optimal relativities that directly incorporates the *a priori* expected claim frequencies to partially address one of the inadequacy scenarios. We derive the analytical solution for the optimal relativities under a financial equilibrium constraint. Furthermore, we introduce a metric called effectiveness of transition rules to compare the different specifications of transition rules. We also argue that varying transition rules which are more flexible in addressing the other inadequacy scenario may be more effective than their corresponding simple rules.

Keywords: bonus-malus system, relativities, transition rules

*Corresponding author. Tel: +65 9793 8939.
Email addresses: `citan1@e.ntu.edu.sg` (Chong It Tan), `kakijackie.li@curtin.edu.au` (Jackie Li), `shli@uwaterloo.ca` (Johnny Siu-Hang Li), `auditha@ntu.edu.sg` (Uditha Balasooriya)

Preprint submitted to Insurance: Mathematics and Economics *February 8, 2015*

1. Introduction

A bonus-malus system (BMS) is a form of *a posteriori* rating used in motor insurance ratemaking to complement the *a priori* rating.[1] It consists of the following three building blocks:

- BMS levels: The number of BMS levels is specified in advance by the insurer. Every new driver enters into the BMS at a pre-specified starting level.
- transition rules: After each year, the driver moves up or down these levels according to a set of transition rules, depending on the claims experience in that year.
- relativities: Each level is associated with a premium adjustment coefficient called relativity, which is to be multiplied with the base premium to obtain the premium amount.

Generally, insurers consider the frequency of claims reported (see, e.g., Lemaire, 1995; Tremblay, 1992; Walhin and Paris, 1999) as the relevant claims experience information, but a few studies (see, e.g., Frangos and Vrontos, 2001; Mahmoudvand and Hassani, 2009; Pinquet, 1997; Tzougas et al., 2014) took into account both the frequency and severity of claims.

Since the occupancy of BMS level is a function of the claims experience, a driver who has a lower *a priori* expected claim frequency is expected to gravitate towards the lower BMS levels. Conversely, a driver with a higher *a priori* expected claim frequency tends to occupy the higher BMS levels. Using such an argument, Taylor (1997) pointed out that the drivers occupying the BMS level with a lower average *a priori* expected claim frequency tend to pay a lower *a priori* premium. Hence, he further commented that the justifiable BMS relativities need to consider the differentiation of underlying claim frequencies by experience, but only to the extent that this differentiation has not been recognized within the *a priori* premiums.

The above remarks have three important implications. Firstly, they tell us that the main purpose of a BMS is to deal with the heterogeneity within each risk class but not the heterogeneity between different risk classes. Secondly, the information of *a priori* expected claim frequency should be incorporated into the determination of premium relativities. Finally, the integration of both *a priori* and *a posteriori* ratemakings should produce a BMS in which the average *a priori* expected claim frequency appears to be as indistinguishable as possible across different BMS levels once the steady state has been reached. Put it differently, the average *a priori* expected claim frequencies of BMS levels should exhibit as little variation as possible to ensure that the implementation of BMS does not further reinforce the *a priori* risk classification. This last implication can be used to alleviate the clustering of policyholders in the BMS levels according to their *a priori* risk characteristics, but surprisingly it is often overlooked in the design of BMS.

In general, the set of transition rules is easy to understand except for a few special bonus rules such as in the former Belgian BMS (see Pitrebois et al., 2003a). Specifically, the same number of level transitions is applied to all drivers regardless of their current levels occupied

[1] For this reason, *a posteriori* rating is often called *a posteriori* correction. It is also called experience rating because it relies on the claims experience information.

and *a priori* characteristics, as long as the drivers make the same number of claims in the current year. For instance, a common set of transition rules is $-1/+2$ (i.e. one level bonus for no claims; two levels malus for each claim). Given the number of BMS levels and the specified transition rules, the optimal relativity for each BMS level was derived by Norberg (1976) through the maximization of asymptotic predictive accuracy, more formally known as Norberg's criterion.

When the general BMS above – with a single set of optimal relativities and a set of simple transition rules – is implemented, it leads to two inadequacy scenarios. The first inadequacy scenario was pointed out by Pitrebois et al. (2003b) who argued that unfairness is induced because of the same *a posteriori* correction in terms of relativity being applied to the policyholders independent of their *a priori* characteristics, hence overpenalizing the drivers who are *a priori* bad risks.[2] Similarly, in this paper we argue that the second inadequacy scenario arises because of the same *a posteriori* correction in terms of level transition being applied to the policyholders independent of their current levels occupied, hence overpenalizing the drivers staying in the higher BMS levels.

To tackle the first inadequacy scenario, we propose a new objective function in the determination of optimal relativities. In more detail, our objective function, as a generalization of the Norberg's criterion, aims to minimize the expected squared difference between the true premium and the actual premium (in absolute terms). By minimizing the new objective function, the first inadequacy scenario could be alleviated because on average the actual premiums payable by policyholders are of the smallest difference from their corresponding true premiums. A financial equilibrium constraint is added to the optimization such that the set of optimal relativities has an average of 100%. This constraint was first considered by Coene and Doray (1996) who devised a financially-balanced BMS. The solution can then be derived analytically by using the Lagrangian method. Moreover, we argue that the last implication should be given more consideration in designing an optimal BMS and address this issue by introducing a new metric called effectiveness of transition rules.

On top of that, we introduce the concept of varying transition rules, which can alleviate the second inadequacy scenario and may improve effectiveness. As an illustrative example, we propose to specify the transition rules as a functional form of the current level occupied instead of a rigid setting. The numerical results indicate that the proposed varying transition rules can lead to a higher level of effectiveness.

The remainder of this paper is organized as follows. We review the motor insurance ratemaking methodology by describing the roles of *a priori* and *a posteriori* ratings in Section 2. The three implications and the two inadequacy scenarios as noted above are discussed in Section 3. In Section 4, we propose the new objective function for the determination of optimal relativities and introduce the effectiveness of transition rules as a metric in designing a BMS. The rationales of using a set of varying transition rules are presented, followed by the specification of the proposed rules. Numerical illustrations are provided in Section 5.

[2] Specifically, they argued that the bad drivers that are required to pay a larger *a priori* premium should be rewarded more and penalized less than the good drivers.

Section 6 concludes the paper.

2. Motor Insurance Ratemaking

The motor insurance ratemaking process consists of two steps called *a priori* ratemaking and *a posteriori* ratemaking. Specifically, the *a priori* ratemaking classifies the policyholders into different tariff risk classes using the observable risk classification variables under the framework of generalized linear models (see McCullagh and Nelder, 1989), such that the policyholders in each tariff class pay the same *a priori* premium. The residual heterogeneity that still remains is tackled by an *a posteriori* ratemaking mechanism, which uses past claims experience information to determine the premium corrections. The two common *a posteriori* ratemaking mechanisms are credibility premium and BMS. In this paper we focus on the design of BMS, which can be perceived as the commercial version of the credibility premium. We refer readers to the seminal paper by Dionne and Vanasse (1989) for discussions on the credibility premium.

2.1. A priori ratemaking

Let us consider a portfolio with n policies. Denote d_i as the risk exposures of policyholder i with number of claims reported Y_i and a vector of *a priori* variables $\boldsymbol{X}_i^T = (X_{i1}, X_{i2}, \dots, X_{iq})$, for $i = 1, 2, \dots, n$. The Poisson regression technique is commonly used to model the number of claims reported Y_i, but other count distributions can also be used. Mathematically, the Poisson regression model for Y_i can then be written as

$$Y_i \sim \text{Poisson}\left(d_i \exp(\beta_0 + \sum_{m=1}^{q} \beta_m x_{im})\right), i = 1, 2, \dots, n, \tag{1}$$

where β_m's are the regression coefficients. If we denote $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_q$ as the estimated regression coefficients, we can express the predicted *a priori* expected claim frequency for policyholder i as

$$\lambda_i = d_i \exp(\hat{\beta}_0 + \sum_{m=1}^{q} \hat{\beta}_m x_{im}), i = 1, 2, \dots, n.$$

Policyholders with the same predicted *a priori* expected claim frequency are then grouped into the same tariff class.

In practice, the insurer uses an extensive list of *a priori* variables to obtain more accurate λ_i's. However, the residual heterogeneity would remain in the risk classes because there are many unobservable variables that have not been considered. This residual heterogeneity is modeled in the *a posteriori* ratemaking. The modeling of residual heterogeneity has been studied extensively (see, e.g., Lemaire, 1995; Denuit et al., 2007). For example, by assuming a conditional distribution for the number of claims reported Y_i, the residual heterogeneity is modeled by adding a random heterogeneity term into the conditional distribution.

2.2. A posteriori ratemaking

2.2.1. Structure of BMS

As mentioned, the structure of BMS consists of BMS levels, transition rules and relativities. The descriptions for BMS levels and transition rules are detailed as follows:

- There are s BMS levels, numbered from 1 to s.
- A specified starting level is assigned to a new driver.
- Each claim-free year is rewarded by a bonus transition (i.e. the driver moves down the BMS level), whereas each reported claim is penalized by a malus transition (i.e. the driver goes up the BMS level).
- Since the number of claims in the current year and the current level occupied suffice to determine the level occupied in the next year, the BMS can be represented by a Markov Chain: the future level in year $t+1$ depends on the present information (the current level occupied and the number of claims in year t) but not on the past history.
- Generally, the number of transition levels imposed is easy to understand because it is applied to all policyholders independent of their *a priori* expected claim frequencies and current levels occupied.
- Nevertheless, any forms of transition rules can be used in determining the occupancy of BMS level. For instance, the old Belgian (see Pitrebois et al., 2003a) and the old Portuguese (see Centeno and Silva, 2002) systems are both dependent on the number of consecutive claim-free years. In this case, the states of BMS should be redefined in such a way that the Markovian property continues to hold.[3]

Given the number of BMS levels and the transition rules, the optimal relativity for each level ℓ, denoted as r_ℓ can be determined through the optimization of certain objective function. For example, under the assumption of conditional Poisson distribution in modeling the number of claims, the policyholders who occupy level ℓ are required to pay an amount of premium equal to r_ℓ% of the *a priori* premium, but this multiplicative form may not hold true for other count distributions.

2.2.2. Modeling of residual heterogeneity

The modeling details of residual heterogeneity in BMS are presented in this subsection. Let Θ_i be the random effect that captures the residual heterogeneity of policyholder i. Given $\Theta_i = \theta$, Y_i's are assumed to be independent and follow a Poisson distribution with mean $\lambda_i\theta$, so we have

$$\Pr\left[Y_i = k|\Theta_i = \theta\right] = \exp(-\lambda_i\theta)\frac{(\lambda_i\theta)^k}{k!}, k = 0, 1, 2, \ldots. \quad (2)$$

[3]Fictitious levels have to be introduced in order to establish the Markovian property. For instance, Lemaire (1995) was the first to propose splitting certain levels into sublevels to accommodate the special bonus rule.

In addition, we assume that all the Θ_i's are independent and follow a Gamma distribution with the following density function:

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

where the use of Poisson-Gamma mixture leads to a Negative Binomial distribution for the number of claims reported Y_i. With these specifications, we obtain $\mathbb{E}[\Theta_i] = 1$ and $\mathbb{E}[Y_i] = \mathbb{E}\big[\mathbb{E}[Y_i|\Theta_i]\big] = \lambda_i$. Note that other specifications of mixed Poisson models can also be used. We refer readers to Yip and Yau (2005) for further details.

The previous descriptions correspond to the policyholder i. In designing a BMS, the insurer is considering a portfolio of policyholders rather than an individual. Now suppose that a policyholder is picked at random from the portfolio. By incorporating the random effect Θ, we can express the actual expected claim frequency for this policyholder as $\Lambda\Theta$, where Λ is defined as the unknown *a priori* expected claim frequency. Note that since the random effect Θ captures the residual heterogeneity, the random variables Λ and Θ may reasonably assumed to be mutually independent. We also assume that the portfolio of policyholders are subdivided into h risk classes, where w_g is denoted as the proportion of policyholders belonging to the g-th risk class whose expected claim frequency is λ_g[4], i.e. $\Pr[\Lambda = \lambda_g] = w_g$.

Let us further denote the transition probability of moving from level ℓ_1 to level ℓ_2 for a policyholder with expected claim frequency $\lambda\theta$ as $p_{\ell_1\ell_2}(\lambda\theta)$, for $\ell_1, \ell_2 = 1, 2, \ldots, s$. The one-step transition matrix can be written as $\boldsymbol{P}(\lambda\theta) = \{p_{\ell_1\ell_2}(\lambda\theta)\}$. We denote the stationary distribution by $\boldsymbol{\pi}(\lambda\theta) = \big(\pi_1(\lambda\theta), \pi_2(\lambda\theta), \ldots, \pi_s(\lambda\theta)\big)^T$, where $\pi_\ell(\lambda\theta)$ is the stationary probability for a policyholder with expected claim frequency $\lambda\theta$ to be in level ℓ. The stationary distribution $\boldsymbol{\pi}(\lambda\theta)$ can be obtained by solving the following two conditions:

$$\begin{cases} \boldsymbol{\pi}(\lambda\theta)^T = \boldsymbol{\pi}(\lambda\theta)^T \boldsymbol{P}(\lambda\theta) \\ \boldsymbol{\pi}(\lambda\theta)^T \mathbf{1} = 1, \end{cases}$$

where $\mathbf{1}$ is the column vector of 1's. These modeling details are needed in Section 4 when we discuss the determination of optimal relativities and the design of transition rules.

The stationary distribution is required because in this paper we choose to determine the set of optimal relativities on the basis of the asymptotic Norberg's criterion. This approach is different from the transient criterion proposed by Borgan et al. (1981), who argued that the age of policy should be taken into consideration especially when a relatively longer period is needed for a BMS to reach its steady state.

3. Objectives of An Optimal BMS

In this section we discuss the objectives of an optimal BMS by examining the three implications from the remarks made by Taylor (1997) and describing the problems of existing

[4]Note that the subscript g in λ_g refers to the g-th risk class. It is different from the subscript i in λ_i as defined earlier, where i refers to the policyholder i.

BMS which lead to two inadequacy scenarios. In particular, we aim to pinpoint the important issues that should be put into consideration in designing a BMS.

3.1. The three implications

Due to the clustering of policyholders in the BMS levels according to their *a priori* risks, Taylor (1997) pointed out that a BMS – as an *a posteriori* correction mechanism – should rely on the claims experience information in its design, but only restricted to the part of the information that does not reflect the observable *a priori* characteristics. In our opinion, these remarks carry the following three implications:

- implication 1: The main purpose of a BMS is to deal with the heterogeneity within each risk class but not the heterogeneity between different risk classes.
  This implication is due to the fact that heterogeneity remains no matter how extensive the use of *a priori* classification variables is, because it is clearly not feasible in practice to consider all the relevant information that may affect the claims behavior of policyholders. It is also important to note that a BMS should not attempt to tackle the heterogeneity between different risk classes, which implies that neither different transition rules nor different set of optimal relativities should be imposed on the policyholders in different risk classes.

- implication 2: The information of *a priori* expected claim frequency should be incorporated into the determination of premium relativity.
  Following from implication 1, the heterogeneity between different risk classes should be treated as given and be embedded into the computation of optimal relativities. This approach ensures the role of BMS as an *a posteriori* correction mechanism to complement the *a priori* ratemaking. Taylor (1997) considered this implication by performing simulations, followed by Pitrebois et al. (2003b) who derived an analytical formula for the optimal relativity.

- implication 3: The integration of both *a priori* and *a posteriori* ratings should produce a BMS in which the average *a priori* expected claim frequency appears to be as indistinguishable as possible across different BMS levels once the steady state has been reached.
  This implication ensures that the implementation of BMS does not further reinforce the *a priori* risk classification. It can be used to alleviate the clustering of policyholders in the BMS levels according to their *a priori* risk characteristics, which is exactly the main argument used by Taylor (1997) in his remarks. However, this implication has not been analyzed in previous studies.

There are two explanations for the clustering phenomenon of policyholders in certain BMS levels according to their *a priori* risk classifications. Firstly, it is due to the probabilistic inferences of different predicted λ_g's for the policyholders who belong to the different tariff classes. Other things being equal, the policyholders with a higher (lower) λ_g are expected to occupy a higher (lower) BMS level. It is therefore essential for a BMS to incorporate the information of *a priori* ratemaking into the design of BMS, particularly in determining the set of optimal relativities.

Secondly, it may be attributed to the fact that there is no differentiation of level transitions according to the *a priori* characteristics and the current level occupied. While it is indeed not necessary to enforce different transition rules according to the *a priori* characteristics on the basis of implication 1, this explanation provides us an objective that the transition rules should be specified in such a way that implication 3 can be achieved. In addition, this objective motivates us to construct a measure of interest as a basis to compare different transition rules of BMS.

3.2. The two inadequacy scenarios

In the practical implementation of BMS, two inadequacy scenarios arise because of the single set of optimal relativities and the simple but inflexible transition rules, respectively. These two scenarios are discussed below:

- inadequacy scenario 1: The unfairness is induced because of the identical optimal relativities applied to all the policyholders regardless of their *a priori* expected claim frequencies, as pointed out by Pitrebois et al. (2003b). As a result, given the same relativity of a level, the drivers with a higher λ_g and so a higher base premium are penalized more.
- inadequacy scenario 2: The unfairness is induced because of the identical level transitions applied to all the policyholders independent of their current levels occupied, using the similar argument as above. Since the drivers staying in higher BMS levels tend to be of a higher *a priori* risk (due to clustering) and so have a higher base premium, they are penalized more if they are subject to the same transition rules.

Note that if we are using credibility premium, these two inadequacy scenarios would not arise. However, as a commercial version of credibility premium, a BMS is restricted in its implementation. For instance, the set of optimal relativities is fixed for all policyholders, while the set of transition rules has to be introduced as part of the experience rating mechanism.[5]

With regards to the first inadequacy scenario, Pitrebois et al. (2003b) proposed to design the optimal relativities for urban and rural drivers separately. Under this proposal, the unfairness could be reduced by imposing a set of lower optimal relativities to urban drivers who are required to pay a higher base premium. This proposed solution is equivalent to the approach of separating a single portfolio of heterogeneous drivers into a number of more homogeneous portfolios and subsequently computing the corresponding set of optimal relativities for each portfolio. In practice, however, insurers prefer to design a single set of optimal relativities for the entire portfolio of drivers due to commercial reasons. Moreover, the introduction of multiple sets of optimal relativities means that the heterogeneity between different risk classes are dealt with separately by the insurer, contradicting implication 1.

[5]In computing credibility premium, there are no BMS levels, transition rules and optimal relativities. The premium is derived on an individual basis as a function of past claims experience and *a priori* expected claim frequency.

4. Designing An Optimal BMS

The issues discussed in the previous section are crucial towards the two aspects of BMS design, namely with respect to the determination of optimal relativities and the specification of transition rules, which are the main subjects in this section. Specifically, we take into account the first two implications and the first inadequacy scenario in proposing a new objective function to determine the optimal relativities. On the other hand, we propose a metric called effectiveness of transition rules by considering the third implication and the second inadequacy scenario.

4.1. Determination of optimal relativities

Pitrebois et al. (2003b) derived an analytical formula for optimal relativity by minimizing the expected squared difference between the 'true' relative premium Θ and the relative premium r_L applicable to the policyholders after the steady state has been reached. Mathematically, it can be written as

$$\min \mathbb{E}\big[(\Theta - r_L)^2\big]. \tag{3}$$

This approach takes into account the *a priori* expected claim frequency (implication 2) in the computation but the objective function itself does not consider the obtained base premium λ_g's. By ignoring the amount of base premium in the objective function, the resulting optimal relativity may not minimize the expected squared difference between the 'true' premium $\Lambda\Theta$ and the actual premium Λr_L paid by the policyholders staying in the BMS level L.

In practice, the insurer is concerned about the collected premiums in absolute terms rather than relative terms, which suggests that the amount of base premium should be incorporated into the objective function. In fact, Norberg (1976) considered the expected squared difference between the true premium and the actual premium (in absolute rather than relative terms), in which the policyholders are not separated into different risk classes and the minimization function can be further simplified as follows:

$$\min \mathbb{E}\big[(\bar{\lambda}\Theta - \bar{\lambda} r_L)^2\big] \equiv \min \mathbb{E}\big[(\Theta - r_L)^2\big],$$

where $\bar{\lambda}$ is the constant expected claim frequency for all policyholders in the absence of *a priori* risk classification. Hence, we propose the minimization of the following objective function in the determination of optimal relativities:

$$\begin{aligned}
\mathbb{E}\big[(\Lambda\Theta - \Lambda r_L)^2\big] &= \sum_{\ell=1}^{s} \mathbb{E}\big[(\Lambda\Theta - \Lambda r_L)^2 | L=\ell\big] \Pr\big[L=\ell\big] \\
&= \sum_{\ell=1}^{s} \mathbb{E}\big[\mathbb{E}\left[(\Lambda\Theta - \Lambda r_L)^2 | L=\ell, \Lambda\right] \big| L=\ell\big] \Pr\big[L=\ell\big] \\
&= \sum_{\ell=1}^{s}\sum_{g=1}^{h} \mathbb{E}\left[(\Lambda\Theta - \Lambda r_L)^2 | L=\ell, \Lambda=\lambda_g\right] \Pr\big[\Lambda=\lambda_g | L=\ell\big] \Pr\big[L=\ell\big] \\
&= \sum_{\ell=1}^{s}\sum_{g=1}^{h} \int_0^{+\infty} (\lambda_g\theta - \lambda_g r_\ell)^2 \pi_\ell(\lambda_g\theta) w_g f(\theta)\, d\theta \\
&= \sum_{g=1}^{h} w_g \int_0^{+\infty} \sum_{\ell=1}^{s} (\lambda_g\theta - \lambda_g r_\ell)^2 \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta. \qquad (4)
\end{aligned}$$

Specifically, the proposed objective function (4) aims to minimize the expected squared difference between the true premium and the actual premium, so its solution has already taken into account the differences in *a priori* premiums, which is a major issue that leads to the first inadequacy scenario. Although the problem of overpenalization of higher *a priori* risk drivers still exists given the set of optimal relativities[6], the solution obtained can partially address the unfairness because the *a priori* base premiums are embedded directly into the objective function (4).

In addition, it is important that the optimal relativity has an average of 100%, so that the bonuses and maluses exactly offset each other and result in a financial equilibrium condition. As such, the financial equilibrium constraint is added to the minimization of (4) such that it becomes the following constrained optimization:[7]

$$\min \mathbb{E}\big[(\Lambda\Theta - \Lambda r_L)^2\big], \quad \text{subject to } \mathbb{E}\left[r_L\right] = 1. \qquad (5)$$

In this case, we can make use of the Lagrangian method to solve the optimization problem in equation (5). Specifically, we define the Lagrangian as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{r}, \alpha) &= \mathbb{E}\big[(\Lambda\Theta - \Lambda r_L)^2\big] + \alpha(\mathbb{E}\left[r_L\right] - 1) \\
&= \sum_{\ell=1}^{s} \mathbb{E}\big[(\Lambda\Theta - \Lambda r_L)^2 | L=\ell\big] \Pr\big[L=\ell\big] + \alpha\Big(\sum_{\ell=1}^{s} r_\ell \Pr\big[L=\ell\big] - 1\Big), \qquad (6)
\end{aligned}$$

[6] Given the set of optimal relativities, the drivers with a higher *a priori* risk are still penalized more than the drivers with a lower *a priori* risk.

[7] In the derivation by Pitrebois et al. (2003b), the solution for optimal relativity is $r_\ell = \mathbb{E}\left[\Theta | L=\ell\right]$, so it is guaranteed that $\mathbb{E}\big[r_L\big] = \mathbb{E}\big[\mathbb{E}\left[\Theta|L\right]\big] = \mathbb{E}\big[\Theta\big] = 1$.

where $\boldsymbol{r} = (r_1, r_2, \ldots, r_\ell)^T$. We can derive the following first order conditions:

$$\Pr\big[L=\ell\big]\big(2\mathbb{E}\big[\Lambda^2\Theta - \Lambda^2 r_L | L=\ell\big] - \alpha\big) = 0, \quad \text{for } \ell = 1, 2, \ldots, s,$$
$$1 - \sum_{\ell=1}^{s} r_\ell \Pr\big[L=\ell\big] = 0.$$

The solution set for α and $r_\ell, \ell = 1, 2, \ldots, s$ is given as:

$$\alpha = \frac{\left(\sum_{\ell=1}^{s} \frac{\mathbb{E}\left[\Lambda^2\Theta|L=\ell\right] \Pr\left[L=\ell\right]}{\mathbb{E}\left[\Lambda^2|L=\ell\right]}\right) - 1}{\sum_{\ell=1}^{s} \frac{\Pr\left[L=\ell\right]}{2\mathbb{E}\left[\Lambda^2|L=\ell\right]}}, \tag{7}$$

$$r_\ell = \frac{\mathbb{E}\big[\Lambda^2\Theta|L=\ell\big]}{\mathbb{E}\big[\Lambda^2|L=\ell\big]} - \frac{\alpha}{2\mathbb{E}\big[\Lambda^2|L=\ell\big]}, \tag{8}$$

where
$$\Pr\big[L=\ell\big] = \sum_{g=1}^{h} w_g \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta, \tag{9}$$

$$\mathbb{E}\big[\Lambda^2\Theta|L=\ell\big] = \frac{\sum_{g=1}^{h} w_g \int_0^{+\infty} \lambda_g^2\theta\pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}{\sum_{g=1}^{h} w_g \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}, \tag{10}$$

$$\mathbb{E}\big[\Lambda^2|L=\ell\big] = \frac{\sum_{g=1}^{h} w_g \int_0^{+\infty} \lambda_g^2\pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}{\sum_{g=1}^{h} w_g \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}. \tag{11}$$

Note that if we relax the financial equilibrium constraint, we have $\alpha^{\text{unconstrained}} = 0$ and

$$r_\ell^{\text{unconstrained}} = \frac{\mathbb{E}\big[\Lambda^2\Theta|L=\ell\big]}{\mathbb{E}\big[\Lambda^2|L=\ell\big]}.$$

However, the premium relativity from the unconstrained optimization may have an average of less than 100%, which is undesirable for the insurer in the long run. In addition, if the insurer does not enforce any *a priori* ratemaking (i.e. all the λ_g's are assumed to be $\bar{\lambda}$), we have equations (7) and (8) respectively reduced to

$$\begin{aligned}\alpha^{\text{without } a\ priori} &= \frac{\left(\sum_{\ell=1}^{s} \frac{\bar{\lambda}^2\mathbb{E}\left[\Theta|L=\ell\right] \Pr\left[L=\ell\right]}{\bar{\lambda}^2\mathbb{E}\left[1|L=\ell\right]}\right) - 1}{\sum_{\ell=1}^{s} \frac{\Pr\left[L=\ell\right]}{2\bar{\lambda}^2\mathbb{E}\left[1|L=\ell\right]}} \\ &= \frac{\mathbb{E}\big[\Theta\big] - 1}{\frac{1}{2\bar{\lambda}^2}} = 0\end{aligned}$$

and

$$
\begin{aligned}
r_\ell^{\text{without } a \text{ } priori} &= \frac{\bar{\lambda}^2 \mathbb{E}\big[\Theta | L = \ell\big]}{\bar{\lambda}^2 \mathbb{E}\big[1 | L = \ell\big]} \\
&= \mathbb{E}\big[\Theta | L = \ell\big] \\
&= \frac{\int_0^{+\infty} \theta \pi_\ell(\bar{\lambda}\theta) f(\theta)\, d\theta}{\int_0^{+\infty} \pi_\ell(\bar{\lambda}\theta) f(\theta)\, d\theta},
\end{aligned}
$$

which have been derived by Norberg (1976) and Pitrebois et al. (2003b).

4.2. Design of transition rules

In this subsection we present a new measure called effectiveness of transition rules to quantify implication 3. It can be used to compare different transition rules in the implementation of BMS. Moreover, there is no existing standard measure for transition rules in the BMS literature, so this new measure can be used as an important metric in deciding the proper specification of transition rules.

As before, let us assume that the policyholders are partitioned into h risk classes in the *a priori* classification, where $w_g = \Pr\left[\Lambda = \lambda_g\right]$ is denoted as the proportion of policyholders in the g-th risk class. In this case, we have

$$
\begin{aligned}
\mathbb{E}\big[\Lambda\big] &= \sum_{g=1}^{h} \lambda_g w_g, \\
\mathbb{V}\big[\Lambda\big] &= \sum_{g=1}^{h} (\lambda_g - \mathbb{E}\big[\Lambda\big])^2 w_g.
\end{aligned}
$$

Pitrebois et al. (2003b) proposed the following expression to quantify the interaction of *a priori* and *a posteriori* ratemakings

$$
\mathbb{E}\big[\Lambda | L = \ell\big] = \frac{\sum_{g=1}^{h} w_g \int_0^{+\infty} \lambda_g \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}{\sum_{g=1}^{h} w_g \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta}. \tag{12}
$$

In more detail, this measure of interaction can be interpreted as the average *a priori* expected claim frequency in the BMS level $L = \ell$. From implication 3, we understand that a good BMS should exhibit small variations of $\mathbb{E}\left[\Lambda | L = \ell\right]$ across different BMS levels. Hence, we can decompose $\mathbb{V}\left[\Lambda\right]$ as a summation of two components:

$$
\begin{aligned}
\mathbb{V}\big[\Lambda\big] &= \mathbb{E}\big[\mathbb{V}\big[\Lambda | L\big]\big] + \mathbb{V}\big[\mathbb{E}\big[\Lambda | L\big]\big] \\
&= \sum_{\ell=1}^{s} \Big(\sum_{g=1}^{h} (\lambda_g - \mathbb{E}\big[\Lambda | L = \ell\big])^2 w_g \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta)\, d\theta \Big) \\
&\quad + \sum_{\ell=1}^{s} \Big(\mathbb{E}\big[\Lambda | L = \ell\big] - \mathbb{E}\big[\Lambda\big] \Big)^2 \Pr\big[L = \ell\big],
\end{aligned} \tag{13}
$$

where the first component $\mathbb{E}\big[\mathbb{V}[\Lambda|L]\big]$ is the expected value of the conditional variances and the second component $\mathbb{V}\big[\mathbb{E}[\Lambda|L]\big]$ is the variance of the conditional means.

Given the value of $\mathbb{V}[\Lambda]$, implication 3 means that the variations should be mainly driven by the first component, which is equivalent to say that the policyholders staying in each BMS level should have large variations in the *a priori* expected claim frequency. In other words, a BMS should provide sufficient flexibility such that the policyholders from any *a priori* risk classes would occupy any BMS levels in the steady state. A larger value of the first component then implies a smaller value of the second component, which indicates that the BMS does not further reinforce the *a priori* risk classification.

It can be seen that the value of the first component is dependent on the number of risk classes h, the values of *a priori* expected claim frequency λ_g's, the number of BMS levels s and also the specification of transition rules, whereas the value of $\mathbb{V}[\Lambda]$ depends only on the first two factors. Since the number of BMS levels is usually fixed in advance, we propose to use

$$\tau_{\text{rule}} = \frac{\mathbb{E}\big[\mathbb{V}[\Lambda|L]\big]}{\mathbb{V}\big[\Lambda\big]} \tag{14}$$

as a measure of the effectiveness of transition rules. Other things being equal, a larger value of τ_{rule} is preferred because of the arguments of implication 3. The insurer may consider using this metric to compare different candidates of transition rules. Alternatively, the insurer could specify a minimum threshold value for τ_{rule} such that the set of transition rules to be used in the BMS should satisfy the criterion.

4.3. Varying transition rules

With regards to the second inadequacy scenario, a set of varying transition rules may help reduce the overpenalization on the drivers staying in the higher BMS levels. The logic behind is to introduce a set of varying transition rules based on the current levels occupied to correct the tendency of those policyholders occupying the higher BMS levels (who tend to have a higher λ_g) to continue staying in the higher BMS levels. As a result, varying transition rules may lead to a fulfilment of implication 3. In more detail, it can be achieved by imposing a lower malus transition and a higher bonus transition on those policyholders occupying the higher BMS levels instead of a fixed number of level transitions for all policyholders. Conversely, the varying transition rules could also impose a higher malus transition and a lower bonus transition on the policyholders occupying the lower BMS levels (who tend to have a lower λ_g).

Let us denote $t_{\ell,k}$ as the effective level transition[8] imposed on the policyholders occupying level ℓ in the BMS and making k claims in the current year, for $\ell = 1, 2, \ldots, s$ and $k \geq 0$. The varying transition rules should satisfy the following structure:

[8]Since the BMS consists of a finite number of levels, the actual level transition for the policyholders staying in the top or bottom levels may not be the same as the level transition specified by the transition rules. Since it is the effective level transition that will be considered in the transition matrix, we choose to work with the effective level transition.

- malus transition: a non-negative effective malus transition[9], and a smaller effective malus transition for the drivers occupying the higher BMS levels:

$$\begin{aligned} t_{\ell,k} &\geq 0, \quad \text{for } k \geq 1, \\ t_{\ell_2,k} &\leq t_{\ell_1,k}, \quad \text{for } k \geq 1, \ell_2 > \ell_1. \end{aligned} \tag{a}$$

- bonus transition: a non-positive effective bonus transition, and a larger effective bonus transition (in magnitude) for the drivers occupying the higher BMS levels,

$$\begin{aligned} & t_{\ell,0} \leq 0, \\ & |t_{\ell_2,0}| \geq |t_{\ell_1,0}|, \text{for } \ell_2 > \ell_1. \end{aligned} \tag{b}$$

As an example, we propose the following set of transition rules

$$\begin{aligned} t_{\ell,0} &= \begin{cases} 0, & \text{for } \ell = 1, \\ -1, & \text{for } 2 \leq \ell \leq \lceil \frac{s}{2} \rceil + 1, \\ -2, & \text{for } \ell > \lceil \frac{s}{2} \rceil + 1, \end{cases} \\ t_{\ell,k} &= \begin{cases} \min \left[s - \ell, \max \left[k, \lceil \frac{s-\ell}{p} \times k \rceil \right] \right], & \text{for } k \geq 1, \ell < s, \\ 0, & \text{for } k \geq 1, \ell = s, \end{cases} \end{aligned} \tag{15}$$

where $\lceil x \rceil$ is the ceiling function of x, which is used because the level transition has to be an integer. If the policyholders are currently staying in level 1, no effective bonus transition can be given. We propose using at most two levels of bonus transition due to practical consideration. The cut-off level of $\lceil s/2 \rceil + 1$ is chosen to ensure that more levels are subject to one level bonus transition, so that the transition rules will not be too soft.[10] For the effective malus transition, the insurer is given the freedom to specify p, which is the smallest number of claims required for the drivers to move from level ℓ to level s. The maximum function in $t_{\ell,k}$ ensures that there is at least one level of malus transition for each claim reported, while the minimum function in $t_{\ell,k}$ limits the number of malus transitions to $s - \ell$. Moreover, there is no effective malus transition for the policyholders who are staying in the highest level s. With these specifications, the varying level transitions imposed in the BMS are both sufficiently flexible and commercial.

Note that the proposed functional form is not unique, so any appropriate functional forms of transition rules that satisfy conditions (a) and (b) can be used. More importantly, the insurer should ensure that the BMS does not further reinforce the *a priori* risk classification via, for example, checking the value of τ_{rule} in choosing the transition rules.

[9]The strict inequality holds true for certain BMS levels but not all BMS levels, due to the fact that the BMS consists of a finite number of levels. In the same vein, all the inequalities presented here are non-strict inequalities.

[10]The numbers of BMS levels that are subject to 0, 1 and 2 levels of bonus transitions are 1, $\lceil \frac{s}{2} \rceil$ and $s - 1 - \lceil \frac{s}{2} \rceil$, respectively, where $\lceil \frac{s}{2} \rceil > s - 1 - \lceil \frac{s}{2} \rceil$ for all s.

5. Numerical Illustrations

We use the motor claims data as studied by Pitrebois et al. (2004), where the policyholders are partitioned into 24 risk classes through the use of risk classification variables. The predicted $\lambda_g, g = 1, 2, \ldots, 24$ and the corresponding w_g's are listed in their work. The estimated parameter a of the Gamma distribution for Θ is 1.2401. With these details, we obtain $\mathbb{E}[\Lambda] = 0.1462$ and $\mathbb{V}[\Lambda] = 0.002625$.

We consider a BMS that consists of $s = 9$ levels, similar to the BMS as discussed by Taylor (1997). Moreover, we analyze four sets of transition rules, two of which are simple while the other two of which are varying. In particular, we compare each set of simple rules with a corresponding set of varying rules. For instance, the simple rules of $-1/+2$ is compared against the varying rules according to equation (15) with $p = 4$.[11] Similarly, the simple rules of $-1/+3$ is compared against the varying rules with $p = 3$.

The effective level transitions for these rules are shown in Tables 1 and 2. For each pair of comparable rules, it is evident that the varying rules are more flexible than the simple rules, as shown by the smaller effective malus transitions and the larger effective bonus transitions for the drivers occupying the higher BMS levels. Also, it can be seen that the rules of $-1/+2$ and $p = 4$ are softer than the rules of $-1/+3$ and $p = 3$.

For each set of transition rules, the values of $\Pr[L = \ell]$ and $\mathbb{E}[\Lambda|L = \ell]$ are tabulated in Table 3. In comparison with the corresponding simple rules, we observe that the proportion of drivers staying in the lower BMS levels (in this case, $\ell \leq 5$) under the varying rules are higher, while the reverse is true for the higher BMS levels. In addition, it can be seen that the drivers tend to occupy the three lowest BMS levels under the softer rules of $-1/+2$ and $p = 4$ as compared to the more severe rules of $-1/+3$ and $p = 3$.

Furthermore, the value of $\mathbb{E}[\Lambda|L = \ell]$ increases with the BMS level ℓ for all four sets of transition rules, providing supporting evidence for the remarks made by Taylor (1997). Particularly, the values of $\mathbb{E}[\Lambda|L = \ell]$ under the rules of $-1/+3$ are the smallest for every BMS level. This is largely due to its severe malus transitions, whereby it would result in a larger proportion of drivers staying in the higher BMS levels, hence the average *a priori* expected claim frequency would be smaller. Conversely, the policyholders occupying the lower BMS levels under such transition rules are expected to have a lower average *a priori* expected claim frequency, otherwise they would have been staying in the higher BMS levels due to the severe malus transitions.

More importantly, we should jointly consider the values of $\Pr[L = \ell]$ and $\mathbb{E}[\Lambda|L = \ell]$ to derive meaningful interpretations regarding the complementary role of *a posteriori* rating. Therefore, we examine the effectiveness of transition rules in alleviating the clustering problem of policyholders by considering the values of τ_{rule} shown in Table 4. We observe that both the varying transition rules are more effective than the corresponding simple rules. This can

[11] The basis of comparison is the smallest number of claims required to move to the highest level s. For example, here the rules of $-1/+2$ require $p = 4$ claims for the drivers occupying level 1 to move to level 9.

Table 1
Transition rules for the simple rules of $-1/+2$ (varying rules with $p = 4$).

Starting level	Level occupied if 0 claims are reported	1	2	3	≥ 4
9	8 (7)	9 (9)	9 (9)	9 (9)	9 (9)
8	7 (6)	9 (9)	9 (9)	9 (9)	9 (9)
7	6 (5)	9 (8)	9 (9)	9 (9)	9 (9)
6	5 (5)	8 (7)	9 (8)	9 (9)	9 (9)
5	4 (4)	7 (6)	9 (7)	9 (8)	9 (9)
4	3 (3)	6 (6)	8 (7)	9 (8)	9 (9)
3	2 (2)	5 (5)	7 (6)	9 (8)	9 (9)
2	1 (1)	4 (4)	6 (6)	8 (8)	9 (9)
1	1 (1)	3 (3)	5 (5)	7 (7)	9 (9)

Table 2
Transition rules for the simple rules of $-1/+3$ (varying rules with $p = 3$).

Starting level	Level occupied if 0 claims are reported	1	2	≥ 3
9	8 (7)	9 (9)	9 (9)	9 (9)
8	7 (6)	9 (9)	9 (9)	9 (9)
7	6 (5)	9 (8)	9 (9)	9 (9)
6	5 (5)	9 (7)	9 (8)	9 (9)
5	4 (4)	8 (7)	9 (8)	9 (9)
4	3 (3)	7 (6)	9 (8)	9 (9)
3	2 (2)	6 (5)	9 (7)	9 (9)
2	1 (1)	5 (5)	8 (7)	9 (9)
1	1 (1)	4 (4)	7 (7)	9 (9)

Table 3
Interaction between a priori and a posteriori ratemakings.

	$-1/+2$		Varying $p = 4$		$-1/+3$		Varying $p = 3$	
level ℓ	$\Pr[L = \ell]$	$\mathbb{E}[\Lambda \mid L = \ell]$	$\Pr[L = \ell]$	$\mathbb{E}[\Lambda \mid L = \ell]$	$\Pr[L = \ell]$	$\mathbb{E}[\Lambda \mid L = \ell]$	$\Pr[L = \ell]$	$\mathbb{E}[\Lambda \mid L = \ell]$
9	2.87%	19.26%	0.70%	21.35%	5.06%	18.02%	1.19%	20.10%
8	2.35%	18.27%	0.54%	19.23%	4.09%	17.21%	1.12%	18.07%
7	2.31%	17.37%	1.26%	18.73%	4.17%	16.36%	2.83%	17.38%
6	2.31%	16.78%	2.34%	17.34%	3.94%	15.90%	2.14%	16.59%
5	3.20%	16.01%	4.24%	16.70%	3.72%	15.58%	5.51%	16.28%
4	3.37%	15.68%	4.17%	16.17%	7.29%	14.83%	9.18%	15.25%
3	8.09%	14.90%	8.97%	15.16%	6.20%	14.70%	7.57%	15.04%
2	6.84%	14.77%	7.44%	14.97%	5.33%	14.59%	6.34%	14.87%
1	68.66%	13.97%	70.34%	14.03%	60.21%	13.87%	64.13%	13.97%

Table 4
Effectiveness of transition rules.

Transition rules	τ_{rule}
$-1/+2$	93.87%
varying $p = 4$	94.74%
$-1/+3$	94.56%
varying $p = 3$	95.18%

be explained by the fact that the varying level transitions imposed onto the drivers are able to produce a higher variation of λ_g's in each BMS level and thus resulting in a higher value of $\mathbb{E}\big[\mathbb{V}\left[\Lambda|L\right]\big]$. On the contrary, the simple rules may lack sufficient flexibility to achieve implication 3.

In addition, it can be seen that the more severe rules of $-1/+3$ ($p = 3$) are more effective than the rules of $-1/+2$ ($p = 4$). This is probably because most of the drivers have a relatively lower expected *a priori* claim frequency. By imposing a more severe transition rules, the variability of λ_g's in each BMS level would be higher, otherwise these drivers would tend to stay in the lower BMS levels under the simple transition rules.

In practice, more sets of transition rules should be compared especially when the number of BMS levels s is much higher, and those rules that have a τ_{rule} value larger than the pre-specified threshold can be used in the BMS. Specifically, if the BMS consists of a higher number of levels, it allows us to consider more choices of transition rules and the effectiveness of the chosen rules would be more crucial in forming a good motor ratemaking system. However, it should be noted that the value of τ_{rule} would be different if the *a priori* risk classification is carried out in a different way.[12]

Next, we perform the calculation of optimal relativities for the BMS under each set of transition rules. The results are displayed in Table 5, where the resulting value of $\mathbb{E}\left[r_L\right]$ is also shown for each set of optimal relativities. Note that r_ℓ and $r_\ell^{\text{unconstrained}}$ correspond to the determination of optimal relativities with and without the financial equilibrium constraint, respectively. For each set of transition rules, the values of r_ℓ are higher than $r_\ell^{\text{unconstrained}}$. This observation is supported by the finding that the values of $\mathbb{E}\left[r_L\right]$ for unconstrained optimal relativities are all below 100%. In other words, when the financial equilibrium constraint is imposed, higher optimal relativities are required such that $\mathbb{E}\left[r_L\right] = 100\%$ is obtained.

Moreover, the set of optimal relativities under the transition rules of $-1/+3$ are considerably lower for every level ℓ, mainly because the policyholders are more likely to occupy the higher BMS levels when the claims are more heavily penalized and hence a lower set of optimal relativities are required to achieve financial equilibrium condition. In contrast, the varying transition rules with $p = 4$ have the highest optimal relativities for all BMS levels except the lowest level $\ell = 1$. One possible explanation is that the probability of drivers occupying the higher BMS levels are the smallest under this set of transition rules, thus

[12] For instance, if the policyholders are partitioned into more or fewer risk classes, the value of τ_{rule} would change accordingly and may not be above 90% as shown in our illustrations.

Table 5
Optimal relativities for the BMS.

	−1/+2		Varying $p = 4$		−1/+3		Varying $p = 3$	
level ℓ	r_ℓ	$r_\ell^{\text{unconstrained}}$	r_ℓ	$r_\ell^{\text{unconstrained}}$	r_ℓ	$r_\ell^{\text{unconstrained}}$	r_ℓ	$r_\ell^{\text{unconstrained}}$
9	266.57%	261.07%	420.41%	419.11%	233.67%	227.73%	293.04%	288.85%
8	235.37%	229.27%	317.11%	315.62%	205.54%	199.03%	230.71%	225.56%
7	207.95%	201.22%	307.54%	305.95%	176.56%	169.34%	211.92%	206.36%
6	188.51%	181.29%	234.82%	233.11%	159.35%	151.70%	183.85%	177.74%
5	162.42%	154.47%	215.07%	213.20%	146.58%	138.59%	173.05%	166.69%
4	150.04%	141.74%	186.85%	184.91%	115.41%	106.55%	133.57%	126.30%
3	118.70%	109.49%	127.08%	125.08%	109.31%	100.29%	124.15%	116.67%
2	112.85%	103.46%	118.27%	116.22%	103.93%	94.75%	116.18%	108.51%
1	72.92%	62.35%	69.43%	67.10%	66.34%	56.11%	70.86%	62.08%
$\mathbb{E}[r_L]$	100%	90.20%	100%	97.79%	100%	90.68%	100%	91.90%

yielding a larger optimal relativities for these BMS levels.

6. Concluding Remarks

In this paper, we first argue that there are two inadequacy scenarios in a bonus-malus system that consists of a single set of optimal relativities and a simple transition rules. To partially address the first inadequacy scenario, we propose a new objective function in the determination of optimal relativities that directly incorporates the *a priori* claim frequencies and derive its analytical solution under a financial equilibrium constraint. We also introduce a metric called effectiveness of transition rules to measure the performance of transition rules in alleviating the clustering phenomenon of policyholders according to their *a priori* characteristics. In addition, we argue that varying transition rules may be more effective than their corresponding simple rules due to the flexibility of varying rules in tackling the second inadequacy scenario.

As mentioned, the value of $\mathbb{V}[\Lambda]$ is dependent on the number of risk classes h and the values of *a priori* expected claim frequency λ_g's. Suppose the insurer categorizes the policyholders into a number of very different risk classes given a set of more useful *a priori* variables such that the variability of Λ is much larger. In this case, the appropriate transition rules with an effectiveness larger than the pre-specified threshold may be different from the case of having a smaller $\mathbb{V}[\Lambda]$.[13] It is warranted in future research to investigate how the varying extent of *a priori* risk classification would affect the corresponding choice of sufficiently effective transition rules.

In this paper, the optimal relativities are obtained through a Bayesian approach. However, as pointed out by Pitrebois et al. (2004), the numerical values may exhibit irregular patterns in terms of an abrupt rise or drop from one level to another, which are undesirable

[13]For instance, the use of a smaller set or less useful information content carried by the classification variables may lead to a smaller value of $\mathbb{V}[\Lambda]$.

for commercial reasons. An alternative for Bayesian relativities is to derive the optimal relativities in a linear form. This approach was proposed by Gilde and Sundt (1989) who expressed the relativities in the form of $r_\ell^{\text{linear}} = \alpha + \beta\ell, \ell = 1, 2, \ldots, s$. Under this linear form, the relativities for two adjacent levels are separated by a constant amount β, making it more suitable for practical implementation of BMS. In future research, it would be desirable to derive the optimal linear relativities that minimizes the objective function (4). Since the linear form is a special case of the Bayesian approach, we can investigate the difference between the Bayesian relativities and the linear relativities, as well as examine the deterioration of the optimization when the linear restriction is incorporated.

Acknowledgements

The first author acknowledges financial support from Aon Benfield Asia Pte. Ltd.

References

Borgan, Ø., Hoem, J.M. & Norberg, R. (1981). A nonasymptotic criterion for the evaluation of automobile bonus systems. *Scandinavian Actuarial Journal*, **1981** (3), 165-178.

Centeno, M.L. & Andrade e Silva, J.M. (2002). Optimal bonus scales under path-dependent bonus rules. *Scandinavian Actuarial Journal*, **2002** (2), 129-136.

Coene, G. & Doray, L.G. (1996). A financially balanced bonus-malus system. *ASTIN Bulletin*, **26** (1), 107-116.

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

Dionne, G. & Vanasse, C. (1989). A generalization of actuarial automobile insurance rating models: the Negative Binomial distribution with a regression component. *ASTIN Bulletin*, **19** (2), 199-212.

Frangos, N. & Vrontos, S. (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin*, **31** (1), 1-22.

Gilde, V. & Sundt, B. (1989). On bonus systems with credibility scales. *Scandinavian Actuarial Journal*, **1989** (1), 13-22.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, Boston.

Mahmoudvand, R. & Hassani, H. (2009). Generalized bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin*, **39** (1), 307-315.

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

Norberg, R. (1976). A credibility theory for automobile bonus system. *Scandinavian Actuarial Journal*, **1976** (2), 92-107.

Pinquet, J. (1997). Allowance for costs of claims in bonus-malus systems. *ASTIN Bulletin*, **27** (1), 33-57.

Pitrebois, S., Denuit, M. & Walhin, J.-F. (2003a). Fitting the Belgian bonus-malus system. *Belgian Actuarial Bulletin*, **3** (1), 58-62.

Pitrebois, S., Denuit, M. & Walhin, J.-F. (2003b). Setting a bonus-malus scale in the presence of other rating factors: Taylor's work revisited. *ASTIN Bulletin*, **33** (2), 419-436.

Pitrebois, S., Denuit, M. & Walhin, J.-F. (2004). Bonus-malus scales in segmented tariffs: Gilde & Sundt's work revisited. *Australian Actuarial Journal*, **10** (1), 107-125.

Taylor, G. (1997). Setting a bonus-malus scale in the presence of other rating factors. *ASTIN Bulletin*, **27** (2), 319-327.

Tremblay, L. (1992). Using the Poisson Inverse Gaussian in bonus-malus systems. *ASTIN Bulletin*, **22** (1), 97-106.

Tzougas, G., Vrontos, S. & Frangos, N. (2014). Optimal bonus-malus systems using finite mixture models. *ASTIN Bulletin*, **44** (2), 417-444.

Walhin, J.-F. & Paris, J. (1999). Using Mixed Poisson distributions in connection with bonus-malus systems. *ASTIN Bulletin*, **29** (1), 81-99.

Yip, K.C.H. & Yau, K.K.W. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, **36** (2), 153-163.