Optimal design of a bonus-malus system: linear relativities revisited

Chong It Tan*

Research School of Finance, Actuarial Studies & Statistics, Australian National University, Australia

Abstract

In this paper, we revisit the determination of optimal relativities under the linear form of relativities that is more viable in designing a commercial bonus-malus system. We derive the analytical formulae for the optimal linear relativities subject to a financial balanced inequality constraint. We also numerically investigate the impact of different a priori risk classification towards the effectiveness of transition rules. Our results show that the a priori risk segmentation is not a sensitive factor for the effectiveness of transition rules. Furthermore, relative to the general relativities, we find that the restriction of linear relativities only produces a small amount of deterioration towards the numerical value of the optimised objective function.

Keywords

1. Introduction

Traditionally, the motor insurance rate-making process consists of two separate steps. In the a priori rating (the first step), insurers make use of certain observable risk classification variables to divide a portfolio of motor vehicle drivers into a number of homogeneous tariff classes. However, these a priori variables are not able to fully capture the risk characteristics of the insured drivers, so the a posteriori rating (the second step) – under the framework of credibility premium or bonus-malus system (BMS) – is needed to tackle the residual heterogeneity. These mechanisms are based on the claims experience information because it is reasonably believed that the unobservable risk characteristics would be partially revealed through the drivers' claims history. In particular, the design of a BMS can be regarded as the commercial version of the credibility premium (see e.g. Dionne & Vanasse, 1989) framework.

Each BMS is represented by three building blocks: the number of BMS levels that the BMS is operating in with a pre-specified starting level, the transition rules which govern the transition of policyholders between BMS levels over time, and the set of optimal relativities that are multiplied with the base premium to obtain the premium amounts payable. Given the specified number of BMS levels and the chosen transition rules, Norberg (1976) first determined the optimal relativity associated with each BMS level through the maximisation of asymptotic predictive accuracy (also known as the

*Correspondence to: Chong It Tan, Research School of Finance, Actuarial Studies & Statistics, Australian National University, Canberra, ACT 0200, Australia. Tel: +612 6125 5458; Fax: +612 6125 0087; E-mail: chongit.tan@anu.edu.au

Norberg's criterion). However, this original approach does not incorporate the heterogeneity between different tariff classes, which is equivalent to the absence of a priori rating. To address this problem, Taylor (1997) developed a simulation procedure (see also Lemaire *et al.*, 2015), whereas Pitrebois *et al.* (2003) obtained the analytical expression for the set of optimal relativities.

In practice, however, the optimal relativities are largely determined commercially to ease the implementation of BMS from the drivers' perspective. For instance, in Asian countries such as Singapore and Hong Kong, all the BMS levels have relativities that are below 100%. In other words, drivers are rewarded a no-claim discount in a claim-free year and are not subject to any relativity above 100% even if they incur claims. In such systems, the BMS would not be financially balanced because the bonuses rewarded are not offset by the maluses imposed, making it not optimal and undesirable for insurers.

Due to the lack of flexibility in the implementation of BMS, Pitrebois *et al.* (2003) also argued that a single set of optimal relativities applied to all the policyholders, regardless of their a priori risk characteristics, would induce unfairness towards the drivers who are perceived to be more risky relative to the drivers that are less risky on the a priori basis. To alleviate this inadequacy scenario, Tan *et al.* (2015) proposed a generalisation of the Norberg's criterion and analytically derived the optimal relativities under a financial equilibrium constraint first considered by Coene & Doray (1996).

However, the obtained relativities from these previous studies may not be desirable for commercial reasons. Specifically, the numerical values of optimal relativities may be of irregular forms due to the possibility of an abrupt rise or drop for successive BMS levels. For instance, the Japanese BMS is a practical example of having non-linear relativities, where the varying differences between adjacent BMS levels may not be appealing for motor vehicle drivers. In light of this issue, Pitrebois *et al.* (2004) considered a linear form of optimal relativities first proposed by Gilde & Sundt (1989), in which a constant amount is imposed as the difference of relativities for any two adjacent BMS levels.

In this paper, we extend the framework of Tan *et al.* (2015) to accommodate a constrained linear form of optimal relativities. Moreover, we also discuss how to incorporate further commercial constraints into the design of a BMS. It follows that the analytical formulae of the optimal linear relativities can be readily derived by using Lagrangian method with Kuhn–Tucker conditions to allow for any inequality constraints.

On the other hand, building upon the measure of the interaction between a priori and a posteriori ratemakings developed by Pitrebois *et al.* (2003), Tan *et al.* (2015) suggested to quantify the different specifications of transition rules using a measure called effectiveness of transition rules. They also argued that a set of level-varying transition rules that are more flexible[1] may be more effective than the corresponding commonly adopted simple transition rules. This effectiveness metric is dependent on the a priori rating outcomes in addition to the choices of transition rules. In this paper, we numerically analyse the impact of a priori risk classification onto the resulting effectiveness of transition rules. We find that the a priori risk segmentation is not a sensitive factor in evaluating the transition rules.

After the specifications of a priori model and transition rules are chosen, we proceed with the determination of optimal relativities. In our numerical illustrations, we consider both the general and

---

[1] Relative to simple (level-fixed) transition rules, level-varying transition rules can be used to correct the likelihoods of drivers to continue occupying certain clusters of BMS levels.

linear forms of optimal relativities in order to examine the deterioration of the optimisation when the linear restriction is incorporated. The results indicate that the restricted form of optimal linear relativities only produces a small amount of deterioration in the optimised objective function as compared with the optimal general relativities.

The remainder of this paper is structured as follows. We describe an overview of the motor insurance rate-making framework in Section 2. In Section 3 we study the determination of optimal linear relativities that takes practical constraints into consideration. The selection criterion of transition rules in designing an optimal BMS and its dependence on the a priori rating are discussed in Section 4. Numerical illustrations are presented in Section 5. Section 6 concludes the paper.

2. Motor Insurance Rate-Making Framework

In this section, we provide a methodological overview of the motor insurance rate-making process. As our aim in this paper mainly focusses on the optimal design of a BMS itself, we shall only briefly discuss the modelling techniques of claim frequencies and residual heterogeneity that are required in the later sections for the determination of optimal relativities and the evaluation of transition rules. We refer interested readers to Lemaire (1995) and Denuit *et al.* (2007) for more extensive details.

2.1. Modelling of a priori claim frequencies

Suppose we have a portfolio with n policies, where d_i is the length of exposure period and Y_i the number of claims reported by driver i during the period. The Poisson regression is widely used in the a priori classification of Y_i under the generalised linear models framework (see McCullagh & Nelder, 1989). In particular, we can express the predicted a priori claim frequency for driver i as

$$\lambda_i = d_i \exp\left(\hat{\beta}_0 + \sum_{j=1}^{q} \hat{\beta}_j x_{ij}\right)$$

where $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_q$ are the estimated regression coefficients and $(x_{i1}, x_{i2}, \ldots, x_{iq})$ represents the vector of observable risk classification variables for driver i. The purpose of this rate-making step is to classify the portfolio of drivers into a number of risk classes. However, it is well known that driving behaviours are influenced by other characteristics that are not observable by the insurer. As a result, the residual heterogeneity within the risk classes motivates the implementation of a posteriori rating via the BMS mechanism as the second step in the rate-making process.

2.2. Modelling of residual heterogeneity

Specifically, the residual heterogeneity is modelled by incorporating a random effect Θ_i into the conditional distribution of Y_i. Conditional on $\Theta_i = \theta$, Y_i is Poisson distributed with mean $\lambda_i \theta$, that is

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

We further suppose that all the Θ_i's are independent and follow a Gamma distribution with parameters (a, a). The probability density function for Θ_i can be written as

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

54

The use of a Poisson-Gamma mixture produces a negative binomial distribution for the number of claims reported Y_i. With these specifications, we have $\mathbb{E}[\Theta_i] = 1$ and $\mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i | \Theta_i]] = \lambda_i$. Note that other specifications of mixed Poisson models can also be used (see e.g. Yip & Yau, 2005).

Now suppose that a motor vehicle driver is picked at random from the portfolio that has been subdivided into h risk classes. The actual expected annual claim frequency for this driver can be written as $\Lambda\Theta$, where Λ is the unknown a priori expected annual claim frequency and Θ captures the random residual heterogeneity. We further denote n_g as the number of drivers in the g-th risk class whose expected claim frequency is λ_g, so we have the proportion of drivers in the g-th risk class as $w_g = \Pr[\Lambda = \lambda_g] = \frac{n_g}{n}$.

As the current level occupied and the number of claims reported in the current period is sufficient to determine the next level occupied, the mechanism of BMS can be modelled as a Markov chain. Let us denote the transition probability of moving from level ℓ_1 to level ℓ_2 for a driver with expected claim frequency $\lambda\theta$ as $p_{\ell_1\ell_2}(\lambda\theta)$, then the one-step transition matrix is $\boldsymbol{P}(\lambda\theta) = \{p_{\ell_1\ell_2}(\lambda\theta)\}$ in a BMS with j levels. The row vector of the stationary distribution $\boldsymbol{\pi}(\lambda\theta) = (\pi_1(\lambda\theta), \pi_2(\lambda\theta), \ldots, \pi_j(\lambda\theta))$ can be easily determined by solving the following two conditions:

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

where $\pi_\ell(\lambda\theta)$ is the stationary probability for a driver with expected claim frequency $\lambda\theta$ to be in level ℓ and $\boldsymbol{1}$ is the column vector of 1's. With these setup, we obtain the following expression for the probability of drivers staying in level ℓ

$$\Pr[L = \ell] = \sum_{g=1}^{h} \Pr[L = \ell | \Lambda = \lambda_g] \Pr[\Lambda = \lambda_g]$$

$$= \sum_{g=1}^{h} \Pr[\Lambda = \lambda_g] \int_{0}^{+\infty} \Pr[L = \ell | \Lambda = \lambda_g, \Theta = \theta] f(\theta)\, d\theta$$

$$= \sum_{g=1}^{h} w_g \int_{0}^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta) d\theta$$

3. Determining Optimal Relativities of a BMS

In practice, the number of BMS levels is usually set in advance and the transition trajectory of drivers are governed by the chosen transition rules. Given these two inputs, the optimal relativities are determined and applied to all drivers independent of their a priori risk characteristics. To partially alleviate the unfairness towards the a priori more risky drivers, Tan *et al.* (2015) proposed the minimisation of the objective function

$$\mathbb{E}\left[(\Lambda\Theta - \Lambda r_L)^2\right] \tag{1}$$

under a financial equilibrium constraint $\mathbb{E}[r_L] = 1$. They also derived an analytical formula for optimal relativity r_L that minimises the expected squared difference between the true premium and the actual premium in absolute terms, which is a generalisation of the asymptotic Norberg's criterion (see Pitrebois *et al.*, 2003) by incorporating the a priori base premiums directly into the objective function.

However, by allowing a general form for optimal relativity r_L in the optimisation, the obtained numerical values may be subject to irregular patterns from one BMS level to another, as pointed out by Pitrebois *et al.* (2004). A more viable solution for commercial BMS in practice is to impose a linear restriction onto the values of optimal relativities, as proposed by Gilde & Sundt (1989). In more detail, the linear relativities can be written as $r_L^{\text{linear}} = \alpha + \beta L$, where it features a difference of a constant premium amount β for any two adjacent BMS levels.

In this paper, we analyse the determination of optimal relativities for a BMS when the linear restriction is imposed. As before, we take into account the financial balanced condition recommended by Coene & Doray (1996) and Baione *et al.* (2002). Hence we obtain the following constrained optimisation with an inequality constraint:

$$\min \mathbb{E}\left[(\Lambda\Theta - \Lambda\alpha - \lambda\beta L)^2\right] \text{ subject to } \alpha + \beta\mathbb{E}[L] \geq 1 \tag{2}$$

We use the Lagrangian method to solve the optimisation problem in equation (2). In this case, the Lagrangian is defined as

$$\mathcal{L}(\alpha, \beta, \gamma) = \mathbb{E}\left[(\Lambda\Theta - \Lambda\alpha - \lambda\beta L)^2\right] + \gamma(1 - \alpha - \beta\mathbb{E}[L]) \tag{3}$$

We can then derive the following Kuhn–Tucker conditions:

$$-2\left(\mathbb{E}[\Lambda^2\Theta] - \alpha\mathbb{E}[\Lambda^2] - \beta\mathbb{E}[\Lambda^2 L]\right) - \gamma = 0 \tag{4a}$$

$$-2\left(\mathbb{E}[\Lambda^2\Theta L] - \alpha\mathbb{E}[\Lambda^2 L] - \beta\mathbb{E}[\Lambda^2 L^2]\right) - \gamma\mathbb{E}[L] = 0 \tag{4b}$$

$$\gamma(1 - \alpha - \beta\mathbb{E}[L]) = 0 \tag{4c}$$

$$1 - \alpha - \beta\mathbb{E}[L] \leq 0 \tag{4d}$$

$$\gamma \geq 0 \tag{4e}$$

For brevity, let us define

$$A = \mathbb{E}\left[\Lambda^2\right] = \sum_{g=1}^{b} \lambda_g^2 \sum_{\ell=1}^{j} \Pr\left[\Lambda = \lambda_g \mid L = \ell\right] \Pr[L = \ell]$$

$$= \sum_{g=1}^{b} \lambda_g^2 \sum_{\ell=1}^{j} \Pr\left[L = \ell \mid \Lambda = \lambda_g\right] \Pr\left[\Lambda = \lambda_g\right]$$

$$= \sum_{g=1}^{b} w_g \lambda_g^2 \sum_{\ell=1}^{j} \int_0^{+\infty} \Pr\left[L = \ell \mid \Lambda = \lambda_g, \Theta = \theta\right] f(\theta) d\theta$$

$$= \sum_{\ell=1}^{j} \sum_{g=1}^{b} w_g \lambda_g^2 \int_0^{+\infty} \pi_\ell(\lambda_g\theta) f(\theta) d\theta \tag{5}$$

Similarly, we define the following notations:

$$B = \mathbb{E}[\Lambda^2\Theta] = \sum_{\ell=1}^{j} \sum_{g=1}^{b} w_g \lambda_g^2 \int_0^{+\infty} \theta\pi_\ell(\lambda_g\theta) f(\theta) d\theta \tag{6}$$

56

$$C = \mathbb{E}[\Lambda^2 L] = \sum_{\ell=1}^{j} \ell \sum_{g=1}^{h} w_g \lambda_g^2 \int_0^{+\infty} \pi_\ell(\lambda_g \theta) f(\theta) d\theta \tag{7}$$

$$D = \mathbb{E}[\Lambda^2 L^2] = \sum_{\ell=1}^{j} \ell^2 \sum_{g=1}^{h} w_g \lambda_g^2 \int_0^{+\infty} \pi_\ell(\lambda_g \theta) f(\theta) d\theta \tag{8}$$

$$E = \mathbb{E}[\Lambda^2 \Theta L] = \sum_{\ell=1}^{j} \ell \sum_{g=1}^{h} w_g \lambda_g^2 \int_0^{+\infty} \theta \pi_\ell(\lambda_g \theta) f(\theta) d\theta \tag{9}$$

$$F = \mathbb{E}[L] = \sum_{\ell=1}^{j} \ell \sum_{g=1}^{h} w_g \int_0^{+\infty} \pi_\ell(\lambda_g \theta) f(\theta) d\theta \tag{10}$$

The solution set (α, β, γ) is analysed as follows:

Case A1: $\gamma = 0$ and if condition (4d) holds with the following solution set

$$\alpha = \frac{B - \frac{CE}{D}}{A - \frac{C^2}{D}} \tag{11a}$$

$$\beta = \frac{E - \frac{CB}{A}}{D - \frac{C^2}{A}} \tag{11b}$$

Otherwise, we have to select

Case A2: if $\gamma > 0$ based on the following expression

$$\gamma = \frac{\left(A - \frac{C^2}{D}\right)\left(D - \frac{C^2}{A}\right) - \left(A - \frac{C^2}{D}\right)\left(EF - \frac{BCF}{A}\right) - \left(D - \frac{C^2}{A}\right)\left(B - \frac{EC}{D}\right)}{\left(A - \frac{C^2}{D}\right)\left(\frac{F^2}{2} - \frac{CF}{2A}\right) + \left(D - \frac{C^2}{A}\right)\left(\frac{1}{2} - \frac{CF}{2D}\right)} \tag{12a}$$

$$\alpha = \frac{B - \frac{C\left(E + \frac{\gamma}{2}F\right)}{D} + \frac{\gamma}{2}}{A - \frac{C^2}{D}} \tag{12b}$$

$$\beta = \frac{E - \frac{C\left(B + \frac{\gamma}{2}\right)}{A} + \frac{\gamma}{2}F}{D - \frac{C^2}{A}} \tag{12c}$$

If we relax the financial balanced inequality constraint, we have $\gamma^{\text{unconstrained}} = 0$ and the solution set as shown in expressions (11a) and (11b) except that the condition (4d) may not be satisfied. In other words, the resulting unconstrained premium relativity may have an average of less than 100%, which may cause the insurer to be insolvent in the long run. Otherwise, the fulfilment of condition (4d) implies that it is not a binding constraint, such that solution set for the constrained optimisation is identical to that of the unconstrained optimisation.

Further constraints for a commercial BMS can be easily incorporated into the minimisation of (2). For instance, the insurer may want to set $\alpha = 0$ such that the drivers in the first BMS level are subject to a premium relativity of β instead of $\alpha + \beta$. In this case, the Kuhn–Tucker condition (4a) is not required and the conditions (4b)–(4e) can be simplified by substituting $\alpha = 0$, where we obtain the following two cases:

Case B1: $\gamma = 0$ and if condition (4d) holds with the following solution set

$$\alpha = 0 \tag{13a}$$

$$\beta = \frac{E}{D} \tag{13b}$$

If Case B1 does not hold, we need to consider

Case B2: if $\gamma > 0$ based on the following expression

$$\gamma = \frac{2(D-EF)}{F^2} \tag{14a}$$

$$\alpha = 0 \tag{14b}$$

$$\beta = \frac{E + \frac{\gamma}{2}F}{D} = \frac{1}{F} \tag{14c}$$

Alternatively, the insurer may wish to establish a minimum premium relativity of $c_{\min}$, so we need to impose an additional inequality constraint of $\alpha + \beta \geq c_{\min}$ for the BMS level $\ell = 1$. The insurer may also want to impose a maximum relativity of $c_{\max}$ that introduces an extra constraint of $\alpha + \beta j \leq c_{\max}$ for the BMS level $\ell = j$. The resulting set of optimal linear relativities can be derived accordingly to satisfy these commercial considerations. However, it should be noted that the minimised value of the objective function would increase with each additional constraint.

4. Choosing Effective Transition Rules for a BMS

In this section, we describe the measures of the interaction between the two rate-making steps and discuss how the varying extent of the a priori risk segmentation may affect the effectiveness of transition rules.

First, after the a priori rating step is carried out, the mean and variance of Λ can be expressed as

$$\mathbb{E}[\Lambda] = \sum_{g=1}^{b} w_g \lambda_g$$

$$\mathbb{V}[\Lambda] = \mathbb{E}[\Lambda^2] - [\mathbb{E}[\Lambda]]^2 = \sum_{g=1}^{b} w_g \lambda_g^2 - \left(\sum_{g=1}^{b} \omega_g \lambda_g \right)^2$$

Note that the value of $\mathbb{E}[\Lambda]$ for a portfolio of drivers is fixed, regardless of the choices of variables used to obtain the predicted claim frequencies λ_g's. This is because of the following necessary condition in the estimation procedure:

$$\sum_i \lambda_i = \sum_i \frac{y_i}{d_i} \rightarrow \frac{\sum_{g=1}^{b} n_g \lambda_g}{n} = \frac{\sum_i \frac{y_i}{d_i}}{n}$$

so we have $\sum_{g=1}^{b} w_g \lambda_g$ equals to the average number of claims reported per unit exposure $\frac{\sum_i \frac{y_i}{d_i}}{n}$ in all circumstances. On the other hand, the value of $\mathbb{V}[\Lambda]$ would change according to the use of different a priori variables.

58

Pitrebois *et al*. (2003) proposed to measure the interaction extent of the two rating mechanisms by

$$\mathbb{E}[\Lambda \,|\, L = \ell] = \frac{\sum_{g=1}^{b} w_g \int_0^{+\infty} \lambda_g \pi_\ell(\lambda_g \theta) f(\theta) d\theta}{\sum_{g=1}^{b} w_g \int_0^{+\infty} \pi_\ell(\lambda_g \theta) f(\theta) d\theta} \tag{15}$$

Apart from conditional on the predicted claim frequencies λ_g's, this measure is dependent on the stationary distribution of drivers in level ℓ – which in turn is reliant on the transition rules – therefore its value reflects how the two rate-making steps interact with one another. However, there are j values of $\mathbb{E}[\Lambda \,|\, L = \ell]$ for $\ell = 1, 2, \ldots, j$, making them difficult to be used as a standard measure in the design of a BMS.

To address this problem, Tan *et al*. (2015) suggested the following metric called effectiveness of transition rules

$$\tau_{\text{rule}} = \frac{\mathbb{E}[\mathbb{V}[\Lambda \,|\, L]]}{\mathbb{V}[\Lambda]} = 1 - \frac{\mathbb{V}[\mathbb{E}[\Lambda \,|\, L]]}{\mathbb{V}[\Lambda]}$$

$$= 1 - \frac{\sum_{\ell=1}^{j} (\mathbb{E}[\Lambda \,|\, L = \ell] - \mathbb{E}[\Lambda])^2 \Pr[L = \ell]}{\mathbb{E}[\Lambda]^2 - [\mathbb{E}[\Lambda]]^2} \tag{16}$$

to compare the different sets of transition rules. Other things being equal, a larger value of τ_{rule} is preferred in designing an optimal BMS.

Nonetheless, although the measure of τ_{rule} utilises its denominator to normalise the impact of a priori rating in order to quantify the effectiveness of transition rules, we note that its numerical value would change based on the varying extent of risk classification. In particular, the value of $\mathbb{E}[\Lambda^2]$ (and hence $\mathbb{V}[\Lambda]$) is expected to increase due to a finer risk segmentation if we include more observable variables in the Poisson regression. However, it is not clear analytically how this increment would be attributed to either of the two components $\mathbb{E}[\mathbb{V}[\Lambda \,|\, L]]$ and $\mathbb{V}[\mathbb{E}[\Lambda \,|\, L]]$, respectively. Consequently, we may have to select different candidates of transition rules subject to the choices of risk segmentation variables. In this paper, we examine the sensitivity of the τ_{rule} with respect to the use of a priori rating variables using numerical illustrations in section 5.

5. Numerical Illustrations

For our illustrations, we use the motor vehicle claims data as documented in De Jong & Heller (2008). The data set contains the information of 67,856 1-year motor insurance policies of a portfolio in the period of 2004/2005. The distribution of the number of claims reported is shown in Table 1, whereby its mean and variance are 0.0728 and 0.0774, respectively.

We make use of the following risk classification variables: driver's age ("Age", six categories), driver's area of residence ("Area", six categories) and vehicle's age ("Vehicle", four categories). For categorical variable s with ν_s categories, we create $\nu_s - 1$ dummy variables. To investigate the impact of different a priori variables towards the effectiveness of transition rules, we consider all the possible subsets of the above three categorical variables except the case of not using any variables. In more detail, the seven subsets of variables are as follows: single categorical variable (three cases), combination of two categorical variables (three cases) and all three categorical variables (one case). For instance, the model with only driver's age contains $\nu_{\text{Age}} = 5$ dummy variables, whereas the

Table 1. Distribution of the number of claims reported in the portfolio.

Number of claims	Observed number of policies
0	63,232
1	4,333
2	271
3	18
4	2
≥ 5	0

model with driver's area of residence and vehicle's age consists of $v_{\text{Area}} + v_{\text{Vehicle}} = 5 + 3 = 8$ dummy variables. The resulting number of risk classes is $\prod_{s^*} v_{s^*}$, where s^* denotes the categorical variable included in the model. For example, the model that considers all three categorical variables has $6 \times 6 \times 4 = 144$ risk classes. We can then evaluate the mean and variance of Λ from the predicted a priori claim frequencies.

For each of these models, we estimate the parameter a of the Gamma distribution for Θ accordingly via the following expression:

$$\alpha = \frac{\sum_{i=1}^{n} \left(d_i \exp\left(\hat{\beta}_0 + \sum_{j=1}^{q} \hat{\beta}_j x_{ij} \right) \right)^2}{\sum_{i=1}^{n} \left[\left(y_i - d_i \exp\left(\hat{\beta}_0 + \sum_{j=1}^{q} \hat{\beta}_j x_{ij} \right) \right)^2 - y_i \right]}$$

Similar to the BMS analysed by Taylor (1997), we consider a BMS that comprises of $j = 9$ levels with two sets of level-fixed rules of $-1/+2$ and $-1/+4$. We also examine the concept of level-varying rules as proposed by Tan *et al.* (2015). In more detail, let $t_{\ell,k}$ be the level transition imposed on the drivers staying in level ℓ and making k claims in the current year, the proposed level-varying rules[2] can be represented as follows:

$$t_{\ell,0} = \begin{cases} 0, & \text{for } l = 1 \\ -1, & \text{for } 2 \leq \ell \leq \left\lceil \frac{j}{2} \right\rceil + 1 \\ -2, & \text{for } \ell > \left\lceil \frac{j}{2} \right\rceil + 1 \end{cases}$$

$$t_{\ell,k} = \begin{cases} \min\left(j - \ell, \max\left(k, \left\lceil \frac{j-\ell}{p} \times k \right\rceil \right) \right), & \text{for } k \geq 1, \ell < j \\ 0, & \text{for } k \geq 1, \ell = j \end{cases} \tag{17}$$

where $[x]$ is the ceiling function of x and the pre-specified parameter p is the smallest number of claims required for the drivers to move from level ℓ to level j. The specifications of the transition rules are illustrated in Table 2, where the level-fixed rules of $-1/+2$ are compared against the level-varying rules based on expression (17) with $p = 4$. In particular, the basis of comparison is the smallest number of claims required for all drivers to move to the highest BMS level j. For instance, the rules of $-1/+2$ require $p = 4$ claims for the drivers in level 1 to move to level 9. Similarly, the level-fixed rules of $-1/+4$ are compared against the level-varying rules with $p = 2$. Note that the

---

[2] The proposed functional form is not unique. Any appropriate form of varying transition rules can be considered as long as it satisfies the conditions as detailed in Tan *et al.* (2015).

Table 2. Specifications of the four sets of transition rules under consideration.

	Next level occupied if						Next level occupied if				
	0	1	2	3	≥4		0	1	2	3	≥4
Current level	claims are reported					Current level	claims are reported				
	(a) Level-fixed −1/+2						(b) Level-varying with $p = 4$				
9	8	9	9	9	9	9	7	9	9	9	9
8	7	9	9	9	9	8	6	9	9	9	9
7	6	9	9	9	9	7	5	8	9	9	9
6	5	8	9	9	9	6	5	7	8	9	9
5	4	7	9	9	9	5	4	6	7	8	9
4	3	6	8	9	9	4	3	6	7	8	9
3	2	5	7	9	9	3	2	5	6	8	9
2	1	4	6	8	9	2	1	4	6	8	9
1	1	3	5	7	9	1	1	3	5	7	9

	Next level occupied if				Next level occupied if		
	0	1	≥2		0	1	≥2
Current level	claims are reported			Current level	claims are reported		
	(c) Level-fixed −1/+4				(d) Level-varying with $p = 2$		
9	8	9	9	9	7	9	9
8	7	9	9	8	6	9	9
7	6	9	9	7	5	8	9
6	5	9	9	6	5	8	9
5	4	9	9	5	4	7	9
4	3	8	9	4	3	7	9
3	2	7	9	3	2	6	9
2	1	6	9	2	1	6	9
1	1	5	9	1	1	5	9

Table 3. Effectiveness of transition rules for different combinations of a priori model and transition rules.

		τ_{rule}			
	$\mathbb{V}[\Lambda]$	−1/+2	Level-varying with $p = 4$	−1/+4	Level-varying with $p = 2$
Vehicle	0.00052	98.75%	99.05%	98.99%	99.20%
Age	0.00291	93.81%	94.76%	96.28%	96.45%
Area	0.00079	98.11%	98.56%	98.52%	98.80%
Vehicle + Age	0.00349	92.88%	93.97%	95.47%	95.77%
Vehicle + Area	0.00132	96.94%	97.64%	97.58%	98.03%
Age + Area	0.00374	92.51%	93.65%	95.14%	95.48%
All	0.00433	91.61%	92.87%	94.38%	94.82%

functional form of the level-varying transition rules in expression (17) may not be straightforward for motor vehicle drivers, but the representation in Table 2 should be much easier for their references. Hence, for a practical BMS, insurers should specify the specifications of transition rules using a table representation.

Table 4. χ^2 likelihood ratio test statistic for nested models.

Model1 (M1)	Model2 (M2)	$\Delta = 2(LL(M1)-LL(M2))$	d.f. = d.f.$_{M1}$−d.f.$_{M2}$	$\Pr(\chi^2_{d.f.} > \Delta)$
Vehicle + Age	Vehicle	805.5217	5	<0.0001
Vehicle + Age	Age	172.7151	3	<0.0001
Vehicle + Area	Vehicle	255.7550	5	<0.0001
Vehicle + Area	Area	169.1066	3	<0.0001
Age + Area	Age	250.3631	5	<0.0001
Age + Area	Area	796.5213	5	<0.0001
All	Vehicle + Age	251.9487	5	<0.0001
All	Vehicle + Area	801.7154	5	<0.0001
All	Age + Area	174.3007	3	<0.0001

Note: *LL*, value of log-likelihood; d.f., degrees of freedom.

With these details, we calculate the effectiveness of transition rules for each combination of a priori model and transition rules. The results are tabulated in Table 3. As expected, we observe that including additional variables into the model lead to an increase in the variability of Λ. In other words, using a richer set of a priori variables cause a higher variance in the predicted claim frequency relative to the reduced model. The values of τ_{rule} are all highly satisfactory above 90% independent of the underlying a priori risk classification model.

For each set of transition rules, the value of τ_{rule} reduces with a larger variance of Λ, which indicates that the transition rules become less effective with a finer risk classification model. Moreover, the level-varying transition rules are more effective than the corresponding simple transition rules for each risk segmentation model, suggesting that the level-varying transition rules perform better, regardless of how the insurer specifies its risk classification model. In addition, it can be seen that the most effective transition rules are the level-varying rules with $p = 2$, whereas the least effective rules are the level-fixed rules with $-1/+2$ consistently for all the models under consideration, implying that the varying extent of a priori risk segmentation is not a sensitive factor towards the ranking of the effectiveness of transition rules. More importantly, these results provide further empirical justifications for introducing the idea of level-varying transition rules into the design of BMS in practice.

In deciding the most suitable combination of a priori model and transition rules, we should take into account both the significance of a priori model as well as the effectiveness of transition rules instead of merely picking the combination that produces the highest value of τ_{rule}. On the basis of the χ^2 likelihood ratio test statistic based for nested models in Table 4, we find that the best model is the model that includes all three categorical variables. Hence, in our subsequent analysis we choose to design the BMS based on this model and the level-varying transition rules with $p = 2$, where its estimated parameter a of the Gamma distribution for Θ is 2.7859.

In our next step, we determine the optimal linear relativities for the BMS based on the analytical formulae presented in section 3. The results are shown in Table 5, where the corresponding value of $\mathbb{E}[r_L]$ is also displayed. For comparison purpose, the optimal general relativities r_ℓ (i.e. without the constraint of linear form) and the optimised value of the objective function (1) are also provided to examine the extent of the deterioration in $\mathbb{E}[(\Lambda\Theta-\Lambda r_L)^2]$ when the linear restriction is imposed onto the BMS design.

Note that $r_\ell^{\text{constrained}}$ and $r_\ell^{\text{unconstrained}}$ correspond to the optimal relativities with and without the financial balanced inequality constraint, respectively. We see that the values of r_ℓ are higher than $r_\ell^{\text{unconstrained}}$ for

Table 5. Optimal relativities for the bonus-malus system.

| Level ℓ | $r_\ell^{\text{linear}} = \alpha + \beta\ell$ | | $r_\ell^{\text{linear}} = \beta\ell$ | | r_ℓ^{general} | |
	$r_\ell^{\text{constrained}}$	$r_\ell^{\text{unconstrained}}$	$r_\ell^{\text{constrained}}$	$r_\ell^{\text{unconstrained}}$	$r_\ell^{\text{constrained}}$	$r_\ell^{\text{unconstrained}}$
9	163.04%	160.16%	278.94%	198.72%	169.01%	166.18%
8	152.12%	149.00%	247.95%	176.64%	148.05%	144.79%
7	141.20%	137.84%	216.95%	154.56%	140.12%	136.68%
6	130.29%	126.69%	185.96%	132.48%	127.38%	123.65%
5	119.37%	115.53%	154.97%	110.40%	115.44%	111.44%
4	108.45%	104.38%	123.97%	88.32%	108.73%	104.58%
3	97.53%	93.22%	92.98%	66.24%	102.99%	98.71%
2	86.61%	82.06%	61.99%	44.16%	97.98%	93.59%
1	75.69%	70.91%	30.99%	22.08%	74.12%	69.21%
$\mathbb{E}[r_L]$	100%	95.75%	100%	71.24%	100%	95.66%
$\mathbb{E}[(\Lambda\Theta - \Lambda r_L)^2]$	0.01391	0.01382	0.02780	0.01993	0.01382	0.01373

each BMS level ℓ under each form of optimal relativities. This finding can be explained by the values of $\mathbb{E}[r_L]$ that are all below 100% for unconstrained optimal relativities. Put it differently, higher optimal relativities are needed such that the binding constraint of the financial balanced condition can be fulfilled.

Nonetheless, the constrained optimal relativities for both the values of $r_\ell^{\text{linear}} = \alpha + \beta\ell$ and r_ℓ^{general} are very close to that of their unconstrained versions, as further supported by the small difference between their values of $\mathbb{E}[(\Lambda\Theta - \Lambda r_L)^2]$. More crucially, the optimised value of the objective function (1) for the optimal linear relativities is only about 0.7% higher relative to that of the optimal general relativities. These findings suggest that imposing a linear restriction to the numerical relativities do not cause a significant deterioration in the optimisation and hence optimal linear relativities should be preferred in designing a commercial BMS. However, the constrained linear relativities of the form $r_\ell^{\text{linear}} = \beta\ell$ produce much larger differences in the optimised $\mathbb{E}[(\Lambda\Theta - \Lambda r_L)^2]$ as compared with its unconstrained values as well as to the optimal general relativities. In summary, the numerical results in this section suggest that the insurer should consider using optimal linear relativities that are more viable in the practical implementation of BMS with a set of level-varying transition rules after a proper a priori risk classification model has been selected.

6. Concluding Remarks

In this paper, we study the determination of optimal relativities for a BMS when the set of relativities is a linear form of the BMS level. The analytical formulae for the optimal linear relativities are derived under the financial balanced inequality constraint. Moreover, it follows that further practical constraints can be easily taken into account in the optimisation procedure.

Our results indicate that imposing a set of linear relativities do not cause a significant change in the minimised objective function. This finding provides supporting justification for insurers in designing an optimal BMS that is both sufficiently optimal on a theoretical basis and commercially viable from a practical perspective. We also find that the varying extent of the a priori risk segmentation is not a sensitive factor towards the resulting effectiveness of transition rules. Furthermore, the numerical results also suggest that the level-varying transition rules are more effective than the level-fixed

rules, regardless of the a priori risk classification model, hence insurers should definitely consider incorporating this idea into the design of an optimal BMS.

Throughout our analysis, we assume that the number of BMS levels is fixed in advance by the insurer. It is possible to relax this assumption by adding the number of BMS levels j as a parameter into the optimisation of (1). In future research, it is warranted to numerically examine the marginal improvement of incorporating j as an extra parameter into the optimal design of a BMS.

Acknowledgements

References

Baione, F., Levantesi, S. & Menzieti, M. (2002). The development of an optimal bonus-malus system in a competitive market. *ASTIN Bulletin*, **32**(1), 159–170.

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

De Jong, P. & Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press, Cambridge.

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.

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, MA.

Lemaire, J., Park, S. & Wang, K. (2015). The impact of covariates on a bonus-malus system: an application of Taylor's model. *European Actuarial Journal*, **5**(1), 1–10.

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.

Pitrebois, S., Denuit, M. & Walhin, J.-F. (2003). 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.

Tan, C.I., Li, J., Li, J.S.-H. & Balasooriya, U. (2015). Optimal relativities and transition rules of a bonus-malus system. *Insurance: Mathematics and Economics*, **61**, 255–263.

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

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.