Ahn, Jae Youn, Eric C. K. Cheung, Rosy Oh, and Jae-Kyung Woo. 2022. "Optimal Relativities in a Modified Bonus-Malus System With Long Memory Transition Rules and Frequency-Severity Dependence." *Variance* 15 (2).

Ratemaking and Product Information

Optimal Relativities in a Modified Bonus-Malus System With Long Memory Transition Rules and Frequency-Severity Dependence

Jae Youn Ahn[1a], Eric C. K. Cheung[2b], Rosy Oh[3c], Jae-Kyung Woo[2d]

[1] Statistics, Ewha Womans University, [2] School of Risk and Actuarial Studies, UNSW Sydney, [3] Mathematics, Korea Military Academy

Keywords: Bonus-Malus System, Extended Bonus-Malus Levels, Optimal Relativity, A Priori Classification, A Posteriori Ratemaking, Frequency-Severity Dependence

Variance

Vol. 15, Issue 2, 2022

In the classical application of a bonus-malus system (BMS) to automobile insurance, the premium for the next year is adjusted according to the policyholder's claim history (particularly frequency) in the previous year. Variations on the classical BMS have been considered, such as taking more of the driver's claim experience into account to better assess an individual's risk. We note that in countries such as Belgium, Italy, Korea, and Singapore it is common practice for a BMS to adopt transition rules according to the claim history for the past several years. In this paper, we revisit a modified BMS briefly introduced by Lemaire in 1995 and Pitrebois and colleagues in 2003. Specifically, such a BMS extends the number of bonus-malus (BM) levels due to an additional component in the transition rules representing the number of consecutive claim-free years. With the extended BM levels granting a more reasonable bonus to careful drivers, this paper investigates the transition rules in a more rigorous manner, and provides the optimal BM relativities under various statistical model assumptions including the frequency random effects model and the dependent collective risk model. Numerical analysis of a real data set is provided to compare the classical BMS and our proposed BMS.

Address for Correspondence: jaeyahn@ewha.ac.kr

1. INTRODUCTION

As Lemaire (1998) notes, with respect to automobile insurance, insurers in the United States and Canada tend to employ many a priori variables (observable factors) to classify risks associated with policyholders, whereas in other parts of the world, such as Europe, Asia, Latin America, and Africa, insurers commonly employ a posteriori ratemaking. The common a posteriori ratemaking process can be accomplished via a bonus-malus system (BMS), and such a posteriori ratemaking holds the promise of resolving residual heterogeneity still remaining within risk classes. As noted in Section 1 of Lemaire (1998), although BMSs are not commonly used in North America, that situation could change, as regulatory authorities may prohibit insurers from using certain classification variables and, moreover, a posteriori ratemaking is known to be an efficient way to assess individual risks. Based on the policyholder's claims experience, such a system discounts the premium as a reward for a claim-free case (bonus) and charges an additional premium in the presence of accidents (malus). Such an a posteriori ratemaking mechanism takes on various forms across countries, but it generally consists of three major components: bonus-malus (BM) levels; transition rules; and BM relativities. Each new policyholder is classified into one of the finite number of risk classes depending on his or her observable risk characteristics, and he or she is also assigned an initial BM level. After each policy period, the BM level

a Jae Youn Ahn is an Associate Professor of the Statistics Department, Ewha Womans University. He received his Ph.D. in Statistics from the University of Iowa. His research interests are in predictive analysis in insurance.

b Eric Cheung is an Associate Professor at the School of Risk and Actuarial Studies of UNSW Sydney. His research interests include insurance risk theory, ruin theory, aggregate claims analysis, queueing theory and stochastic processes, and he has published more than 40 papers in these areas. He is an Associate of the Society of Actuaries, and he has been an Associate Editor of Insurance: Mathematics and Economics since July 2019.

c Rosy Oh is an Assistant Professor of the Department of Mathematics, Korea Military Academy. She received a Ph.D. in Statistics from Ewha Womans University. Her research area is predictive analysis.

d JK Woo is currently an Associate Professor, a Nominated Accreditation Actuary at the School of Risk and Actuarial Studies, UNSW Sydney. Her research interests are focused on risk theory, aggregate claim analysis, dependence modelling and the bonus-malus system. She is an editorial board member of ASTIN Bulletin, Probability in the Engineering and Information Sciences, and topic editor of Risks.

moves up or down according to the prespecified transition rules that penalize policyholders based on the number of reported claims and reward those without claims. In a typical BMS, claims experience in the previous policy period is used to determine the transition rule. However, for better separation of consistently and temporarily good (or bad) risks, we can consider capturing claims experience in multiple previous policy periods in the design of the transition rule, as is done in Belgium. Then, the premium is renewed at a rate that is the product of the BM relativity and the base premium. Here the BM relativity is determined by the BM level, which plays a role in the premium adjustment coefficient. In this way, the actual premium charged to policyholders in each BM level can be more closely matched to their risks reflected in the claim number record for the past few policy periods and consequently the claims costs are fairly shared. In the classical BMS, the base premium is often calculated as a product of the mean frequency and the mean severity due to the assumed independence between those two random variables (e.g., Pitrebois, Denuit, and Walhin 2003a; Denuit et al. 2007; Tan et al. 2015). However, the independence assumption is often criticized as some recent insurance reports demonstrate the existence of a significant dependence between claim frequency and severity. In the actuarial literature, various statistical models have been developed in an attempt to capture certain dependency structures between frequency and severity of claims. These include, for example, copula models (Czado et al. 2012; Frees, Lee, and Yang 2016), shared or bivariate random effects models (Hernandez-Bastida, Fernandez-Sanchez, and Gomez-Deniz 2009; Baumgartner, Gruber, and Czado 2015; Oh, Shi, and Ahn 2020), and two-part models (Shi, Feng, and Ivantsova 2015; Garrido, Genest, and Schulz 2016; Park, Kim, and Ahn 2018; Jeong et al. 2021).

The focus of the paper is twofold. First, we revisit the extension of the BM levels (Lemaire 1995; Pitrebois, Denuit, and Walhin 2003b) and define our long-memory transition rules in a more formal setting. Second, under the proposed transition rules, we consider the optimal relativities not only in a frequency-only model but also in a dependent collective risk model that allows for dependence between frequency and severity. As an example of the dependent collective model, we consider the copula-based bivariate random effects model as in Oh, Shi, and Ahn (2020), among many choices of shared or bivariate random effects models. Whereas in the classical BMS the policyholder's BM level is reduced if there is no claim in the past year, in our modified BMS the BM level moves to a lower level when the policyholder has no claim consecutively for years equivalent to the "period of penalty". A similar case is discussed in Chapter 17 of Lemaire (1985) and Section 5 of Pitrebois, Denuit, and Walhin (2003b), where the transition is made based on the number of consecutive claim-free years. However, the optimal relativities in the present paper are obtained by solving different forms of optimization problems. Under two models—namely, the frequency random effects model and the bivariate random effects model—the optimal relativities and the hypothetical mean squared error are studied. The rest of the paper is organized as follows. In Section 2, we describe two random effects models and give the optimal relativity results under the classical BMS rules. In Section 3, a period of penalty (denoted as "pen") is introduced in the transition rules such that the BM level moves downward only when there is no claim for the last consecutive $(1+pen)$ years. Under the modified transition rules, an "augmented" BM level is newly defined so that its transition process has the Markov property. In turn, the optimal relativities are provided under the two different models. Finally, in Section 4, we turn to a data set from the Wisconsin Local Government Property Insurance Fund (e.g., Frees, Lee, and Yang (2016)) to illustrate the impact of the long-memory transition rules on our BMS versus the classical BMS.

2. MODEL DESCRIPTIONS AND BASIC RESULTS

For the i-th policyholder in the t-th policy year, we let N_{it} be the number of claims and $\mathbf{Y}_{it}=(Y_{it1},\ldots,Y_{itN_{it}})$ be the vector of associated claim amounts, where Y_{itj} is the j-th claim amount. (Note that $\mathbf{Y}_{it}$ is undefined when $N_{it}=0$.) Also, the aggregate claim amount and average claim amount are defined as

$$\begin{aligned} S_{it} &:= \begin{cases} \sum_{j=1}^{N_{it}} Y_{itj}, & N_{it}>0 \\ 0, & N_{it}=0 \end{cases} \quad \text{and} \\ M_{it} &:= \begin{cases} \frac{1}{N_{it}} \sum_{j=1}^{N_{it}} Y_{itj}, & N_{it}>0, \\ 0, & N_{it}=0, \end{cases} \end{aligned} \tag{1}$$

respectively. Clearly, these two quantities are directly linked as $S_{it}=N_{it}M_{it}$.

In the insurance ratemarking process, a priori ratemaking results in risk classification of the (new) policyholder whose past claims history is insufficient for the ratemaking purposes. Using the observed risk characteristics of the policyholders (collected in the row vector $\mathbf{X}_i$ for the i-th policyholder), their risk classes are determined and the a priori premiums are also obtained. Then, the residual heterogeneity is explained by capturing the unobserved risk characteristics of policyholders (denoted as Θ_i for the ith policyholder) in an a posteriori ratemaking process, and consequently their premium amounts are adjusted based on the claims history. Here, $\mathbf{X}_i$ and Θ_i do not depend on time. In general, the frequency part may be considered enough to construct the BMS under the assumption of independence between frequency and severity. However, recent BMS research in Oh, Shi, and Ahn (2020) reveals that the dependence between frequency and severity plays a critical role in the determination of optimal relativities in the case of significant dependence. In this regard, the superscripts [1] and [2], representing the frequency and severity components, respectively, are introduced to variables and their corresponding realizations. When we are looking only at the frequency component, we drop the notation [1] for convenience. To model the frequency and the severity of claims, we employ a generalized linear model as described in de Jong and Heller (2008) along with the assumption of an exponential dispersion family for the random components (see Appendix A for a description).

2.1. RANDOM EFFECTS MODELS

For the unobserved risk characteristics in a posteriori ratemating, let us describe two different models that take into account, first, the randomness of frequency only and, second, both frequency and severity. For convenience, we assume that there are $\mathcal{K}$ risk classes predetermined at the moment of when the contracts begin. We start with the description of the model for the frequency component only.

Model 1 (frequency model with random effects). *First, the weight of the k-th risk class is defined as*

$$w_k := \mathbb{P}(\mathbf{X} = \mathbf{x}_k), \qquad k = 1, \dots, \mathcal{K}, \tag{2}$$

where $\mathbf{X}$ is the row vector of the observed risk characteristics of a randomly picked person from the population of policyholders and $\mathbf{x}_k$ is a vector of the observed risk characteristics for the k-th risk class. Note that w_k can be regarded as the fraction of policyholders with observed risk characteristics $\mathbf{x}_k$. The a priori premium for the i-th policyholder is determined as $\Lambda_i := \eta^{-1}(\mathbf{X}_i\beta)$, where $\eta(\cdot)$ is a link function, $\mathbf{X}_i$ is the row vector of observed risk characteristics of this i-th policyholder, and β is a column vector of regression coefficients in the count regression model estimated from the past data concerning the number of reported claims. Incorporating the unobserved risk component for the frequency,[1] the conditional distribution of the number of claims N_{it} for the i-th policyholder in the t-th policy year given the observed risk characteristics $\mathbf{X}_i = \mathbf{x}_i$ and unobserved risk characteristics $\Theta_i = \theta_i$ along with the distribution of Θ_i is provided in Appendix B.

The following model extends the two-part model (Frees and Valdez 1998; Garrido, Genest, and Schulz 2016) by introducing bivariate random effects to model various types of dependence, such as

- dependence among frequencies;
- dependence among severities; and
- dependence between frequency and severity.

Model 2 (collective risk model with bivariate random effects (Oh, Shi, and Ahn 2020)). *Recall that the superscripts [1] and [2] represent, respectively, the frequency and the severity components. As an extension of Model 1, this model takes into account the random component of the claim severity as well for the a priori and a posteriori ratemaking processes. In contrast to (2), we define the weight of the k-th risk class as*

$$w_k := \mathbb{P}(\mathbf{X}^{[1]} = \mathbf{x}_k^{[1]}, \mathbf{X}^{[2]} = \mathbf{x}_k^{[2]}), \qquad k = 1, \dots, \mathcal{K},$$

where $(\mathbf{X}^{[1]}, \mathbf{X}^{[2]})$ is the pair of row vectors containing the observed risk characteristics relevant to the frequency and the severity of claims for a randomly picked person from the population of policyholders, and $(\mathbf{x}_k^{[1]}, \mathbf{x}_k^{[2]})$ is the corresponding pair for the k-th risk class. The a priori premium for the i-th policyholder can be obtained from the information of the vector $(\Lambda_i^{[1]}, \Lambda_i^{[2]})$, which is in turn determined by the observed risks characteristics as

$$\Lambda_i^{[1]} := \eta_1^{-1}(\mathbf{X}_i^{[1]}\beta^{[1]}) \quad \textit{and} \quad \Lambda_i^{[2]} := \eta_2^{-1}(\mathbf{X}_i^{[2]}\beta^{[2]}),$$

where $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are link functions, $\mathbf{X}_i^{[1]}$ and $\mathbf{X}_i^{[2]}$ are the i-th policyholder's row vectors of observed risk characteristics for frequency and severity, and $\beta^{[1]}$ and $\beta^{[2]}$ are column vectors of parameters to be estimated. More importantly, the dependence between claim frequency and severity is also considered as the corresponding unobserved risk characteristics $(\Theta_i^{[1]}, \Theta_i^{[2]})$ are modeled via a bivariate copula. More specific distributional descriptions of this model are provided in Appendix C.

It is instructive to note that when Model 2 is restricted to the claim frequency only, it is conceptually equivalent to Model 1.

2.2. OPTIMAL RELATIVITIES

In the classical BMS framework, common transition rules are such that the BM level is lowered by one for a claim-free year and increased by h levels per claim, leading to what is referred to as the $-1/+h$ system. Based on such transition rules, the BM level for each policyholder evolves as a (discrete-time) Markov chain. Let us denote L as a random variable representing the BM level (from 0 to z) for a randomly picked policyholder from the population in a stationary state (independent of time). Its distribution (that is, the proportion of policyholders in each level) is determined by the frequency component as

$$\mathbb{P}(L = \ell) = \sum_{k=1}^{\mathcal{K}} w_k \int \pi_\ell(\lambda_k\theta, \psi) g(\theta) \mathrm{d}\theta, \qquad \ell = 0, \dots, z, \tag{3}$$

where $\pi_\ell(\lambda_k\theta, \psi)$ is the stationary probability that a policyholder with expected frequency $\lambda_k\theta$ is in level ℓ. (It is understood that $\lambda_k = \eta^{-1}(\mathbf{x}_k\beta)$, $g(\cdot)$ and ψ in (3) are replaced by $\lambda_k^{[1]} = \eta_1^{-1}(\mathbf{x}_k^{[1]}\beta^{[1]})$, $g_1(\cdot)$ and $\psi^{[1]}$, respectively, if one considers Model 2.) We shall provide explanations on how to obtain the stationary probabilities at the beginning of Section 3.2 for our more general model in relation to Example 2. The relativity associated with the BM level ℓ is denoted by $\zeta(\ell)$. The following lemmas review how to determine the optimal relativity $\tilde{\zeta}(\ell)$ in two optimization settings under the model assumptions in Section 2.1.

Lemma 1. (Tan et al. 2015) Consider the optimization problem

$$(\tilde{\zeta}(0), \dots, \tilde{\zeta}(z)) := \underset{(\zeta(0), \dots, \zeta(z)) \in \mathbb{R}^{z+1}}{\arg\min} \mathbb{E}[(\Lambda\Theta - \Lambda\zeta(L))^2] \tag{4}$$

under the frequency-only Model 1. Because $\mathbb{E}[N_{it}|\Lambda_i, \Theta_i] = \Lambda_i\Theta_i$ (see (B.1)) is the "correct" premium for the i-th policyholder if we did know Θ_i, one can regard $\Lambda\Theta = \eta^{-1}(\mathbf{X}\beta)\Theta$ as the "correct" premium for a policyholder randomly picked from the population having observed risk characteristics $\mathbf{X}$ and unobserved risk characteristics Θ. The optimization (4) amounts to choosing the relativities to minimize the mean squared difference between $\Lambda\Theta$ and the actual premium charged $\Lambda\zeta(L)$ when a

1 In practice, figuring out all the possible combinations of the observed risk characteristics for the model is very difficult. Hence, the introduction of the unobserved risk characteristics Θ_i is essential, and this provides theoretical and practical justifications for the a posteriori risk classification.

policyholder is in BM level L. Then, the optimal relativities can be analytically calculated as

$$\tilde{\zeta}(\ell) := \frac{\mathbb{E}[\Lambda^2\Theta|L=\ell]}{\mathbb{E}[\Lambda^2|L=\ell]} = \frac{\sum_{k=1}^{\mathcal{K}} w_k\lambda_k^2 \int \theta\pi_\ell(\lambda_k\theta,\psi)g(\theta)\mathrm{d}\theta}{\sum_{k=1}^{\mathcal{K}} w_k\lambda_k^2 \int \pi_\ell(\lambda_k\theta,\psi)g(\theta)\mathrm{d}\theta}, \quad \ell=0,\ldots,z.$$

Lemma 2. (Oh, Shi, and Ahn 2020) Consider the optimization problem

$$(\tilde{\zeta}(0),\ldots,\tilde{\zeta}(z)) := \underset{(\zeta(0),\ldots,\zeta(z))\in\mathbb{R}^{z+1}}{\arg\min}\ \mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]} - \Lambda^{[1]}\Lambda^{[2]}\zeta(L))^2] \tag{5}$$

under the frequency-severity Model 2. As $\mathbb{E}[S_{it}|\Lambda_i^{[1]},\Lambda_i^{[2]},\Theta_i^{[1]},\Theta_i^{[2]}] = \Lambda_i^{[1]}\Lambda_i^{[2]}\Theta_i^{[1]}\Theta_i^{[2]}$ (see Appendix C) is the "correct" premium for the i-th policyholder, $\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]} = \eta_1^{-1}(\mathbf{X}^{[1]}\beta^{[1]})\eta_2^{-1}(\mathbf{X}^{[2]}\beta^{[2]})\Theta^{[1]}\Theta^{[2]}$ is the "correct" premium for a policyholder randomly picked from the population having observed risk characteristics $(\mathbf{X}^{[1]},\mathbf{X}^{[2]})$ and unobserved risk characteristics $(\Theta^{[1]},\Theta^{[2]})$. The optimization (5) is concerned with choosing the relativities to minimize the mean squared difference between $\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]}$ and the actual premium charged $\Lambda^{[1]}\Lambda^{[2]}\zeta(L)$ when a policyholder is in BM level L. Then, the optimal relativities can be analytically calculated as

$$\tilde{\zeta}(\ell) := \frac{\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2\Theta^{[1]}\Theta^{[2]}|L=\ell]}{\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2|L=\ell]} = \frac{\sum_{k=1}^{\mathcal{K}} w_k(\lambda_k^{[1]}\lambda_k^{[2]})^2 \int\int \theta^{[1]}\theta^{[2]}\pi_\ell(\lambda_k^{[1]}\theta^{[1]},\psi^{[1]})h(\theta^{[1]},\theta^{[2]})\mathrm{d}\theta^{[1]}\mathrm{d}\theta^{[2]}}{\sum_{k=1}^{\mathcal{K}} w_k(\lambda_k^{[1]}\lambda_k^{[2]})^2 \int \pi_\ell(\lambda_k^{[1]}\theta^{[1]},\psi^{[1]})g_1(\theta^{[1]})\mathrm{d}\theta^{[1]}}, \quad \ell=0,\ldots,z.$$

The optimal relativities in Lemmas 1 and 2 can be applied to various transition rules and BM levels ℓ. To calculate optimal relativity in the BMS literature, the Markov property of the transition rules is typically used, and the BM level is often assumed to be stationary (Pitrebois, Denuit, and Walhin 2003a; Denuit et al. 2007; Tan et al. 2015; Oh, Shi, and Ahn 2020). The most common BMS (and also one of the simplest of the BMSs) is the $-1/+h$ system, although other variations are available, such as, for example, in Tan et al. (2015). However, insurers in many countries, including Belgium, Korea, and Singapore, adapt transition rules with a long memory. While such transition rules do not directly possess the Markov property, examples in the literature show how that can be dealt with (Lemaire 1995; Pitrebois, Denuit, and Walhin 2003a). A key technique involves augmentation of BM levels, and the following section generalizes these examples.

We remark that the optimization in (4) or (5) is not the only optimization criterion one can use. Various criteria exist to define the optimality of BM relativities—see, e.g., Denuit et al. (2007) and Tan et al. (2015). More importantly, BM relativities obtained from (4), (5), Denuit et al. (2007), and Tan et al. (2015) are known to create some systematic bias in the prediction of premium. Such systematic bias is called the double-counting problem, and we refer interested readers to Lemaire (1995), Taylor (1997), and Oh et al. (2020) for the details of the double-counting problem and its solution.

3. A MODIFIED BMS WITH AUGMENTED BM LEVELS

In this section, the number of consecutive claim-free years for a policyholder is taken into account in the transition rules. Assuming there are $z+1$ BM levels, we let $L_t \in \{0,1,\ldots,z\}$ be the BM level at time $t \in \{0,1,\ldots\}$ so that L_t is the BM level applicable for the $(t+1)$-th policy year, that is, from time t to time $t+1$. Similar to previous notation, for $t \in \{1,2,\ldots\}$ we use N_t to denote the number of claims in the t-th policy year.

3.1. MODIFIED TRANSITION RULES OF $-1/+h/pen$

While the classical BMS usually adopts the so-called $-1/+h$ system for $h \le z$ as described in Section 2.2, we shall introduce an additional component, namely "pen," to this system. Specifically, under such a $-1/+h/pen$ system, each reported claim increases the BM level by h while $1+pen_t^*$ consecutive claim-free years will result in a decrease of one BM level at time $t+1$ when moving from the $(t+1)$-th policy year to the $(t+2)$-th. Here pen_t^* is defined as

$$pen_t^* := \min\{pen, t\}, \tag{6}$$

for $t \in \{0,1,\ldots\}$. This implies that the BM level at time $t=0$ starts with no penalty, and indeed a new policyholder can have his or her BM level reduced by one every year as long as he or she maintains a no-claim record. (This is also equivalent to letting $N_0 = N_{-1} = \ldots = N_{-pen} = 0$ and saying that one requires $1+pen$ consecutive claim-free years to decrease the BM level by one.) Note that the classical BMS is retrieved from our extended model by letting $pen = 0$. We further assume that a new policyholder without any driving history belongs to the BM level ℓ_0 in the beginning. Then, L_t is mathematically represented as

$$L_t := \begin{cases} \min\{L_{t-1}+hN_t, z\}, & N_t > 0, \\ \max\{L_{t-1}-1, 0\}, & N_t = \ldots = N_{\max\{t-pen_{t-1}^*,1\}} = 0, \\ L_{t-1}, & \text{otherwise}, \end{cases} \tag{7}$$

for $t \in \{1,2,\ldots\}$, with $L_0 = \ell_0$. It is also assumed that the BM relativity $\zeta(\ell)$ is applied in the same manner as in the classical $-1/+h$ system. That is, each BM level ℓ is bestowed with BM relativity $\zeta(\ell)$. Compared with the classical $-1/+h$ system, an additional pen_t^* claim-free years are required to reduce the BM level by one, whereas the penalty of climbing h BM levels per claim remains the same. Such a system is introduced to motivate policyholders to drive carefully by applying a stricter rule for lowering the BM level (reward case) while keeping the same rule with respect to increasing the BM level (penalty case). Insurers have adopted such transition rules in some countries, such as Singapore and Korea.

While the Markov property in the transition probability appearing in certain BMSs is convenient for the analysis of BMS-related problems (e.g., Taylor 1997; Pitrebois, Denuit, and Walhin 2003a; Denuit et al. 2007), it is obvious that with the current definition of L_t, the transition rules are not Markovian as L_t depends not only on L_{t-1} but also on the numbers of claims in the past multiple periods. To resolve this issue, as explained in Lemaire (1995) and Pitrebois, Denuit, and Walhin (2003a), fictitious levels can be

included to redefine the state space of the BM levels. Subdividing some of the BM levels in order to include the information of the number of claim-free years results in an increase of the number of states versus the classical case. More precisely, for $\ell = h, \dots, z$, the BM level ℓ is augmented into

$$(\ell)_0, (\ell)_1, \dots, (\ell)_{pen},$$

where the subscript stands for the number of additional claim-free periods (compared with the classical BMS) required to get rewarded, while $\ell \in \{0, \dots, h-1\}$ is just translated into $(\ell)_0$ without augmentation.[2] Hence, the original $(z+1)$ BM levels in the $-1/+h/pen$ system have been augmented to $[h + (z-h+1) \times (pen+1)]$ BM levels. Let us denote the BM level at time t under the augmented system as L_t^*. Then, it is convenient to define the possible combinations of $(\ell)_a$ in our proposed model (that is, the state space) as

$$\mathcal{A}_{z,pen} := \{(\ell)_0 | \ell = 0, \dots, h-1\} \cup \{(\ell)_0, \dots, (\ell)_{pen} | \ell = h, \dots, z\}. \tag{8}$$

Furthermore, suppose that a policyholder who was at the BM level $L_{t-1}^* = (\ell)_a \in \mathcal{A}_{z,pen}$ at time $t-1$ has reported at time t that N_t accidents occurred during the t-th policy year. Then the new BM level L_t^* at time t is determined as

$$L_t^* := \begin{cases} (\max\{\ell - 1, 0\})_0, & \text{if} \quad N_t = 0 \text{ and } a = 0, \\ (\ell)_{a-1}, & \text{if} \quad N_t = 0 \text{ and } a \neq 0, \\ (\min\{\ell + hN_t, z\})_{pen}, & \text{if} \quad N_t > 0, \end{cases} \tag{9}$$

for $t \in \{1, 2, \dots\}$, where a new policyholder without driving history starts with $L_0^* = (\ell_0)_0$. The transitions of L_t^* are now Markovian, as L_t^* depends on $L_{t-1}^* = (\ell)_a$ and the latest claim number only. Also, we assume that the BM relativity in state $(\ell)_a \in \mathcal{A}_{z,pen}$ under the augmented system defined via (9) is the same as the one in $\ell \in \{0, 1, \dots, z\}$ for the original system defined in (7) in the sense that the relativities depend on the BM level ℓ but not the information a that is artificially introduced to make the transitions Markovian. In other words, under the proposed model one has the relativities

$$\zeta^*\left((\ell)_a\right) := \zeta(\ell), \qquad (\ell)_a \in \mathcal{A}_{z,pen}. \tag{10}$$

We call the BMS with BM levels (8), transition rules (9), and relativities (10) an augmented $-1/+h/pen$ system. The following example illustrates the transition rules with a certain penalty period.

Example 1. Let us consider the $-1/+2/2$ system with $z = 20$. That is, whereas the BM level is to be increased by two per claim, having $1 + pen_t^*$ consecutive claim-free years is to be rewarded by a decrease of one BM level. We assume that a new policyholder without driving history starts at BM level $L_0 = 10$ (so $\ell = 10$ at time 0). Given the BM level L_t at time t and the claim frequency N_{t+1} in the $(t+1)$-th year, we summarize the next year's BM level L_{t+1} according to (7) and alternatively L_{t+1}^* according to (9) in Table 1. (Note that with $pen = 2$, (6) implies that pen_t^* stays level at 2 once t reaches 2.) The first few transitions are explained as follows:

1. Transition from $t = 0$ to $t = 1$:
   At time 0, one has $pen_0^* = 0$ from (6), and therefore $1 + pen_0^* = 1$ claim-free year is required to reduce the BM level by one at time 1 (like the classical BMS), and so we set $a = 0$ at time 0. As it turns out that there is no claim reported at $t = 1$ (i.e., $N_1 = 0$), the BM level is reduced from $L_0 = 10$ to $L_1 = 9$ (and thus $\ell = 9$ at $t = 1$). Now $pen_1^* = 1$ according to (6), meaning that a total of $1 + pen_1^* = 2$ consecutive claim-free years are needed to reduce the BM level by one the next year at time 2. Since there has already been one claim-free year, only one more claim-free year is needed to earn such a reward (again like the classical BMS), and therefore one sets $a = 0$ at time 1.
2. Transition from $t = 1$ to $t = 2$:
   Since $N_2 = 0$ again, the condition of reward is satisfied and the BM level moves down to $L_2 = 8$ (and $\ell = 8$) at time 2. With $pen_2^* = 2$ at time 2, it is known that $1 + pen_t^* = 3$ consecutive claim-free years are needed to lower the BM level by one. With two consecutive claim-free years already in hand, we need only one more claim-free year to reduce the BM level in the next year, and therefore $a = 0$ at time 2.
3. Transition from $t = 2$ to $t = 3$:
   Since $N_3 = 2$ in the third year, the BM level moves up to 12 $(= 8 + (2 \times 2))$ at time 3 (and so $\ell = 12$). As $pen_3^* = 2$ at time 3, one needs $1 + pen_3^* = 3$ consecutive claim-free years to reduce the BM level by one, which means two claim-free years in addition to the one claim-free year required in the classical BMS. Consequently, we have $a = 2$ at $t = 3$.
4. Transition from $t = 3$ to $t = 4$:
   Because $N_4 = 1$ in the fourth year, the BM level increases to $L_4 = 14$ $(= 12 + (2 \times 1))$ at time 4. As there is a claim in this period, three consecutive claim-free years are still needed to reduce the BM level by one. So, we have $a = 2$ at $t = 4$.
5. Transition from $t = 4$ to $t = 5$:
   Although there is no claim in the fifth year as $N_5 = 0$, the condition of having three consecutive claim-free years in order to enjoy a reward is not satisfied, and therefore the BM level stays at the same level as last year, so that $L_5 = 14$. Then two more consecutive claim-free years are required to lower the BM level by one. This is one additional year compared with the classical $-1/+2$ system, so $a = 1$ at $t = 5$.

Example 2. Suppose that we have the $-1/+2/1$ system with $z = 7$. We consider Model 1 where a policyholder, say the i-th policyholder, has observed and unobserved risk characteristics $\mathbf{X}_i = \mathbf{x}_i$ and $\Theta_i = \theta_i$, respectively. Then, based on the augmented form of BM levels defined in (8), one finds a total of $(2 + ((7 - 2 + 1) \times 2)) = 14$ BM levels with the transition matrix $\mathbf{P}$ given by

\[\tiny\begin{align} &

2 No augmentation of the BM levels $0, \dots, h-1$ is necessary as those BM levels can be reached only as a result of reward from the previous years.

Table 1. Transition rules under a $-1/+2/2$ system

t	0	1	2	3	4	5	6	7	8	9
N_{t+1}	0	0	2	1	0	0	0	0	1	
L_t	10	9	8	12	14	14	14	13	12	14
pen_t^*	0	1	2	2	2	2	2	2	2	2
L_t^*	$(10)_0$	$(9)_0$	$(8)_0$	$(12)_2$	$(14)_2$	$(14)_1$	$(14)_0$	$(13)_0$	$(12)_0$	$(14)_2$

$$(0)_0 \quad (1)_0 \quad (2)_0 \quad (2)_1 \quad (3)_0 \quad (3)_1 \quad (4)_0 \quad (4)_1 \quad (5)_0 \quad (5)_1 \quad (6)_0 \quad (6)_1 \quad (7)_0 \quad (7)_1$$

$$\begin{cases} \pi^*(\lambda_k\theta,\psi) = \pi^*(\lambda_k\theta,\psi)\mathbf{P}, \\ \pi^*(\lambda_k\theta,\psi)\mathbf{e}_{14} = 1, \end{cases}$$

$$\begin{array}{cccccccccccccc} p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_0 \end{array}$$

(1)_0\\(2)_0\(2)_1\(3)_0\(3)_1\(4)_0\(4)_1\(5)_0\(5)_1\(6)_0\(6)_1\(7)_0\(7)_1

where

$$p_n$$

in the matrix is defined by (see (B.1))

$$p_n := \mathbb{P}(N_{it} = n|\Theta_i = \theta_i, \mathbf{X}_i = \mathbf{x}_i), \qquad n = 0, 1, \ldots.$$

For simplicity, the dependence of p_n on θ_i and $\mathbf{x}_i$ is suppressed. Similarly, if we assume Model 2 instead and the i-th policyholder has observed and unobserved risk characteristics given by $(\mathbf{X}_i^{[1]}, \mathbf{X}_i^{[2]}) = (\mathbf{x}_i^{[1]}, \mathbf{x}_i^{[2]})$ and $(\Theta_i^{[1]}, \Theta_i^{[2]}) = (\theta_i^{[1]}, \theta_i^{[2]})$, then p_n in the matrix $\mathbf{P}$ is (see (C.1))

$$p_n := \mathbb{P}(N_{it} = n|\Theta_i^{[1]} = \theta_i^{[1]}, \mathbf{X}_i^{[1]} = \mathbf{x}_i^{[1]}), \qquad n = 0, 1, \ldots.$$

3.2. OPTIMAL RELATIVITIES

In this section, under a $-1/+h/pen$ system, the optimal values of the relativities $\zeta(\ell)$ for $\ell = 0, 1, \ldots, z$ are studied. Following the classical BMS literature, we denote the stationary distribution of L_t^* by L^* for a randomly picked policyholder. Then, under the representation in Model 1, one finds in an analogous manner to (3) that

$$\begin{aligned} \mathbb{P}(L^* = (\ell)_a) &= \sum_{k=1}^{\mathcal{K}} w_k \int \pi^*_{(\ell)_a}(\lambda_k\theta,\psi) g(\theta)\mathrm{d}\theta, \\ &(\ell)_a \in \mathcal{A}_{z,pen}, \end{aligned} \tag{11}$$

where $\pi^*_{(\ell)_a}(\lambda_k\theta,\psi)$ is the stationary distribution for a policyholder to be in BM level $(\ell)_a$ given that his or her a priori claim frequency is λ_k and unobserved risk characteristics are summarized in θ. (Again, λ_k, $g(\cdot)$ and ψ are replaced by $\lambda_k^{[1]}$, $g_1(\cdot)$ and $\psi^{[1]}$, respectively, if we consider Model 2 instead.) For example, the 14-dimensional row vector of stationary probabilities corresponding to Example 2, namely

$$\pi^*(\lambda_k\theta,\psi) := (\pi^*_{(0)_0}(\lambda_k\theta,\psi), \pi^*_{(1)_0}(\lambda_k\theta,\psi), \ldots, \pi^*_{(7)_1}(\lambda_k\theta,\psi)),$$

can be obtained as the solution of

$$\begin{cases} \pi^*(\lambda_k\theta,\psi) = \pi^*(\lambda_k\theta,\psi)\mathbf{P}, \\ \pi^*(\lambda_k\theta,\psi)\mathbf{e}_{14} = 1, \end{cases}$$

where $\mathbf{P}$ is the transition matrix in Example 2 (recalculated with $\mathbf{x}_k$ and θ in place of $\mathbf{x}_i$ and θ_i, respectively), and $\mathbf{e}_{14}$ is a 14-dimensional column vector of ones.

Under the augmented system, we consider the optimal relativities as the solution of the optimization problem

$$(\tilde\zeta(0), \ldots, \tilde\zeta(z)) := \underset{(\zeta(0),\ldots,\zeta(z))\in\mathbb{R}^{z+1}}{\arg\min}\ \mathbb{E}[(\Lambda\Theta - \Lambda\zeta^*(L^*))^2] \tag{12}$$

where we wish to predict the frequency under the frequency-only Model 1. On the other hand, when we are interested in the prediction of the aggregate claim, like (5), the optimal relativities are given by the solution of the optimization problem

$$(\tilde\zeta(0), \ldots, \tilde\zeta(z)) := \underset{(\zeta(0),\ldots,\zeta(z))\in\mathbb{R}^{z+1}}{\arg\min}\ \mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]} - \Lambda^{[1]}\Lambda^{[2]}\zeta^*(L^*))^2] \tag{13}$$

under the frequency-severity Model 2. Recall that, by definition, $\zeta^* : \mathcal{A}_{z,pen} \mapsto \mathbb{R}$ in (12) or (13) is completely characterized by $\zeta(\ell)$ for $\ell = 0, \ldots, z$ (see (10)).

Proposition 1. Consider the $-1/+h/pen$ system under Model 1. The optimal relativities defined as the solution of the optimization problem (12) are given by

$$\tilde\zeta(\ell) := \frac{\mathbb{E}[\Lambda^2\Theta|L^* = (\ell)_0]}{\mathbb{E}[\Lambda^2|L^* = (\ell)_0]}, \qquad \ell = 0, \ldots, h-1, \tag{14}$$

and

$$\begin{aligned} \tilde\zeta(\ell) := &\frac{\sum_{a=0}^{pen}\mathbb{E}[\Lambda^2\Theta|L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a)}{\sum_{a=0}^{pen}\mathbb{E}[\Lambda^2|L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a)}, \\ &\ell = h, \ldots, z; a = 0, \ldots, pen, \end{aligned} \tag{15}$$

where the numerator and denominator can be calculated based on

$$\begin{aligned} &\mathbb{E}[\Lambda^2\Theta|L^* = (\ell)_a] \\ &\quad = \frac{\sum_{k=1}^{\mathcal{K}} w_k\lambda_k^2 \int \theta\pi^*_{(\ell)_a}(\lambda_k\theta,\psi)g(\theta)\mathrm{d}\theta}{\mathbb{P}(L^* = (\ell)_a)} \end{aligned} \tag{16}$$

and

$$\mathbb{E}[\Lambda^2|L^* = (\ell)_a] = \frac{\sum_{k=1}^{\mathcal{K}} w_k\lambda_k^2 \int \pi^*_{(\ell)_a}(\lambda_k\theta,\psi)g(\theta)\mathrm{d}\theta}{\mathbb{P}(L^* = (\ell)_a)}, \tag{17}$$

respectively. Note that when $pen = 0$, the optimal relativity $\tilde\zeta(\ell)$ reduces to that in Tan et al. (2015) given in Lemma 1.

Proof. Note that the objective function in (12) can be written as

$$\begin{aligned} &\mathbb{E}[(\Lambda\Theta - \Lambda\zeta^*(L^*))^2] \\ &\quad = \sum_{(\ell)_a\in\mathcal{A}_{z,pen}} \mathbb{E}[(\Lambda\Theta - \Lambda\zeta^*((\ell)_a))^2|L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a) \\ &\quad = \sum_{(\ell)_a\in\mathcal{A}_{z,pen}} \mathbb{E}[(\Lambda\Theta - \Lambda\zeta(\ell))^2|L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a), \end{aligned} \tag{18}$$

where the last line follows from (10). We observe from (8) that

- for each $\ell = 0, \ldots, h-1$, the relativity $\zeta(\ell)$ is shared by only one BM level $(\ell)_0$; and

- for each $\ell = h, \ldots, z$, the same relativity $\zeta(\ell)$ is shared by the augmented BM levels $(\ell)_0, \ldots, (\ell)_{pen}$.

Consequently, (18) becomes

$$\begin{aligned}&\mathbb{E}[(\Lambda\Theta - \Lambda\zeta^*(L^*))^2] \\ &= \sum_{\ell=0}^{h-1} \mathbb{E}[(\Lambda\Theta - \Lambda\zeta(\ell))^2 | L^* = (\ell)_0]\mathbb{P}(L^* = (\ell)_0) \\ &+ \sum_{\ell=h}^{z}\sum_{a=0}^{pen} \mathbb{E}[(\Lambda\Theta - \Lambda\zeta(\ell))^2 | L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a).\end{aligned} \tag{19}$$

For each fixed $\ell = 0, \ldots, h-1$, differentiation of (19) with respect to $\zeta(\ell)$ for optimization yields

$$\mathbb{E}[-2\Lambda(\Lambda\Theta - \Lambda\zeta(\ell)) | L^* = (\ell)_0]\mathbb{P}(L^* = (\ell)_0) = 0,$$

from which (14) follows. On the other hand, when $\ell = h, \ldots, z$, taking the derivative with respect to $\zeta(\ell)$ in (19) leads to

$$\sum_{a=0}^{pen} \mathbb{E}[-2\Lambda(\Lambda\Theta - \Lambda\zeta(\ell)) | L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a) = 0,$$

proving (15).

Here, the numerators of (14) and (15) can be calculated using

$$\begin{aligned}&\mathbb{E}[\Lambda^2\Theta | L^* = (\ell)_a] \\ &= \frac{1}{\mathbb{P}(L^* = (\ell)_a)}\mathbb{E}[\Lambda^2\Theta I(L^* = (\ell)_a)] \\ &= \frac{1}{\mathbb{P}(L^* = (\ell)_a)}\sum_{k=1}^{\mathcal{K}} \int \lambda_k^2\theta \mathbb{P}(\Theta \in \mathrm{d}\theta, \Lambda = \lambda_k, L^* = (\ell)_a) \\ &= \frac{1}{\mathbb{P}(L^* = (\ell)_a)}\sum_{k=1}^{\mathcal{K}} \int \lambda_k^2\theta \mathbb{P}(L^* = (\ell)_a | \Theta \in \mathrm{d}\theta, \Lambda = \lambda_k)\mathbb{P}(\Theta \in \mathrm{d}\theta, \Lambda = \lambda_k).\end{aligned} \tag{20}$$

Since the a priori $\mathbf{X}$ and the a posteriori Θ are independent and $\Lambda = \eta^{-1}(\mathbf{X}\beta)$, one has that $\mathbb{P}(\Theta \in \mathrm{d}\theta, \Lambda = \lambda_k) = \mathbb{P}(\Lambda = \lambda_k)\mathbb{P}(\Theta \in \mathrm{d}\theta) = w_k g(\theta)\mathrm{d}\theta$. Using this together with the fact that $\mathbb{P}(L^* = (\ell)_a | \Theta \in \mathrm{d}\theta, \Lambda = \lambda_k) = \pi^*_{(\ell)_a}(\lambda_k\theta, \psi)$ confirms that (20) reduces to (16). The expectation (17) can be obtained in an almost identical manner, and the details are omitted.

Proposition 2. Consider the $-1/+h/pen$ system under Model 2. The optimal relativities defined as the solution of the optimization problem (13) are given by

$$\tilde{\zeta}(\ell) := \frac{\mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2\Theta^{[1]}\Theta^{[2]} | L^* = (\ell)_0]}{\mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2 | L^* = (\ell)_0]}, \quad \ell = 0, \ldots, h-1,$$

and

$$\tilde{\zeta}(\ell) := \frac{\sum_{a=0}^{pen} \mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2\Theta^{[1]}\Theta^{[2]} | L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a)}{\sum_{a=0}^{pen} \mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2 | L^* = (\ell)_a]\mathbb{P}(L^* = (\ell)_a)}, \quad \ell = h, \ldots, z; a = 0, \ldots, pen,$$

where the numerator and denominator can be calculated based on

$$\begin{aligned}&\mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2\Theta^{[1]}\Theta^{[2]} | L^* = (\ell)_a] \\ &= \frac{\sum_{k=1}^{\mathcal{K}} w_k(\lambda_k^{[1]}\lambda_k^{[2]})^2 \int\int \theta^{[1]}\theta^{[2]}\pi^*_{(\ell)_a}(\lambda_k^{[1]}\theta^{[1]}, \psi^{[1]})h(\theta^{[1]}, \theta^{[2]})\mathrm{d}\theta^{[1]}\mathrm{d}\theta^{[2]}}{\mathbb{P}(L^* = (\ell)_a)}\end{aligned}$$

and

$$\mathbb{E}[\left(\Lambda^{[1]}\Lambda^{[2]}\right)^2 | L^* = (\ell)_a] = \frac{\sum_{k=1}^{\mathcal{K}} w_k(\lambda_k^{[1]}\lambda_k^{[2]})^2 \int \pi^*_{(\ell)_a}(\lambda_k^{[1]}\theta^{[1]}, \psi^{[1]})g_1(\theta^{[1]})\mathrm{d}\theta^{[1]}}{\mathbb{P}(L^* = (\ell)_a)},$$

respectively. Note that when $pen = 0$, the optimal relativity $\tilde{\zeta}(\ell)$ reduces to that in Oh, Shi, and Ahn (2020) given in Lemma 2).

Proof. The proof of Proposition 2 is similar to that of Proposition 1, and thus we omit the details.

3.3. NUMERICAL ANALYSIS

Here we provide a simple example to examine the effect of the period of penalty on the optimal BM relativity $\tilde{\zeta}(\ell)$ and the stationary probability $\mathbb{P}(L = \ell)$. First, it is convenient to introduce the relation

$$\mathbb{P}(L = \ell) = \sum_{\{a | (\ell)_a \in \mathcal{A}_{z,pen}\}} \mathbb{P}(L^* = (\ell)_a), \tag{21}$$

which follows from the definition of L^*. The performances of our modified BM transition rules under a $-1/+1/pen$ system as well as a $-1/+2/pen$ system for $pen = 0, 1, 2, 3$ are compared by computing the hypothetical mean squared error (HMSE). For Model 1 (the frequency random effects model) with the set of BM relativities $\zeta = \{\zeta(\ell)\}_{\ell=0}^{z}$, the HMSE is expressed as

$$\begin{aligned}\mathrm{HMSE}(\zeta) &:= \mathbb{E}[(\Lambda\Theta - \Lambda\zeta^*(L^*))^2] \\ &= \sum_{k=1}^{\mathcal{K}} \sum_{(\ell)_a \in \mathcal{A}_{z,pen}} w_k \int (\lambda_k\theta - \lambda_k\zeta(\ell))^2 \pi^*_{(\ell)_a}(\lambda_k\theta, \psi)g(\theta)\mathrm{d}\theta.\end{aligned} \tag{22}$$

If ζ is replaced by the vector of optimal relativities $\tilde{\zeta} = \{\tilde{\zeta}(\ell)\}_{\ell=0}^{z}$ calculated using Proposition 1, then $\mathrm{HMSE}(\tilde{\zeta})$ is the minimized value corresponding to the right-hand side of the optimization problem (12).

Similarly, for Model 2 with the set of BM relativities $\zeta = \{\zeta(\ell)\}_{\ell=0}^{z}$, following the logic in Oh, Shi, and Ahn (2020), the HMSE is expressed as

$$\begin{aligned}\mathrm{HMSE}(\zeta) &:= \mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]} - \Lambda^{[1]}\Lambda^{[2]}\zeta^*(L^*))^2] \\ &= \sum_{k=1}^{\mathcal{K}} \sum_{(\ell)_a \in \mathcal{A}_{z,pen}} w_k \int\int (\lambda_k^{[1]}\lambda_k^{[2]}\theta^{[1]}\theta^{[2]} - \lambda_k^{[1]}\lambda_k^{[2]}\zeta(\ell))^2 \pi^*_{(\ell)_a}(\lambda_k\theta, \psi^{[1]})h(\theta^{[1]}, \theta^{[2]})\mathrm{d}\theta^{[1]}\mathrm{d}\theta^{[2]}.\end{aligned}$$

Again, $\mathrm{HMSE}(\tilde{\zeta})$ is the minimum on the right-hand side of the optimization problem (13) with $\tilde{\zeta} = \{\tilde{\zeta}(\ell)\}_{\ell=0}^{z}$ calculated using Proposition 2.

Example 3. Under the frequency model with random effects described as Model 1, it is assumed that there is only one risk class, where the frequency N follows a Poisson distribution with a priori frequency $\lambda = \lambda_0$ (so that the distribution function F can be put in the form of (B.1)) and the random effect Θ in (B.2) has a lognormal distribution with mean 1 and $\sigma^2 = 0.99$, that is,

$$\begin{aligned}N | (\Lambda = \lambda, \Theta = \theta) &\sim \mathrm{Poisson}(\lambda\theta) \quad \text{with} \quad \lambda = \lambda_0, \\ \Theta &\sim \mathrm{Lognormal}(-\sigma^2/2, \sigma^2) \quad \text{with} \quad \sigma^2 = 0.99.\end{aligned}$$

With $z = 9$ and $pen = 0, 1, 2, 3$, the stationary probabilities $\mathbb{P}(L = \ell)$ and the optimal BM relativities $\tilde{\zeta}(\ell)$ for all ℓ's and the resulting HMSE (under the optimal relativities) are calculated under a $-1/+1/pen$ system and a $-1/+2/pen$ system. The results when $\lambda_0 = 0.05$ are first summarized in Tables 2(a) and (b).

In obtaining the tables, we first construct the transition probability matrix according to the transition rules of the BMS and the claim frequency distribution with the assumed parameters (see Example 2) and obtain the stationary probability $\pi^*_{(\ell)_a}$ as explained at the beginning of Section 3.2. Then, $\mathbb{P}(L = \ell)$ can be calculated from (11) and (21). Note that integration with respect to θ under the distributional assumption on the density $g(\theta)$ of Θ is needed in (11). While explicit evaluation of such a one-dimensional integral is not possible in general, numerical integration can be implemented easily in most computing software. In turn, the optimal relativities $\tilde{\zeta} = \{\tilde{\zeta}(\ell)\}_{\ell=0}^{z}$ can be obtained from Proposition 1, where numerical integration is again performed based on (16) and (17) (but the integral appearing

Table 2. (Example 3) Distribution of L and optimal relativities under the frequency-only Model 1

(a) $-1/+1/pen$ systems with various pen for $\lambda_0 = 0.05$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
9	17.096	0.001	12.635	0.002	10.288	0.004	8.801	0.008
8	13.947	0.000	9.626	0.001	7.468	0.002	6.151	0.002
7	11.978	0.000	8.413	0.001	6.597	0.002	5.477	0.002
6	10.384	0.000	7.331	0.001	5.786	0.002	4.832	0.003
5	9.408	0.001	6.282	0.001	4.985	0.003	4.187	0.004
4	7.304	0.001	5.197	0.002	4.159	0.004	3.520	0.006
3	5.552	0.002	4.035	0.005	3.284	0.008	2.815	0.012
2	3.676	0.007	2.816	0.015	2.367	0.023	2.076	0.030
1	2.003	0.044	1.676	0.073	1.483	0.095	1.348	0.111
0	0.887	0.944	0.826	0.898	0.779	0.858	0.740	0.821
HMSE	0.00287		0.00249		0.00227		0.00213	

(b) $-1/+2/pen$ systems with various pen for $\lambda_0 = 0.05$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
9	10.288	0.002	8.105	0.007	6.858	0.013	6.030	0.020
8	8.106	0.002	5.793	0.004	4.567	0.005	3.807	0.007
7	6.755	0.002	5.101	0.004	4.209	0.005	3.633	0.006
6	5.410	0.003	4.006	0.005	3.263	0.008	2.796	0.011
5	4.638	0.003	3.674	0.005	3.135	0.007	2.772	0.008
4	3.349	0.007	2.599	0.014	2.205	0.020	1.950	0.027
3	2.973	0.009	2.497	0.012	2.206	0.013	2.000	0.014
2	1.779	0.041	1.508	0.066	1.351	0.086	1.241	0.101
1	1.677	0.038	1.512	0.034	1.394	0.030	1.303	0.028
0	0.811	0.892	0.762	0.850	0.723	0.812	0.691	0.778
HMSE	0.00260		0.00244		0.00235		0.00229	

(c) $-1/+1/pen$ systems with various pen for $\lambda_0 = 1$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	2.092	0.311	1.713	0.463	1.528	0.562	1.416	0.634
13	1.147	0.089	0.791	0.060	0.618	0.045	0.512	0.035
12	0.914	0.044	0.677	0.036	0.546	0.030	0.462	0.025
11	0.794	0.027	0.602	0.025	0.494	0.022	0.422	0.020
10	0.718	0.020	0.548	0.019	0.453	0.018	0.390	0.016
9	0.662	0.016	0.505	0.016	0.419	0.015	0.362	0.014
8	0.618	0.014	0.470	0.015	0.390	0.014	0.337	0.013
7	0.579	0.014	0.439	0.014	0.363	0.014	0.314	0.013
6	0.543	0.014	0.410	0.015	0.339	0.014	0.293	0.013
5	0.508	0.016	0.382	0.016	0.315	0.016	0.273	0.014
4	0.470	0.020	0.353	0.020	0.291	0.018	0.252	0.017
3	0.429	0.027	0.322	0.026	0.266	0.024	0.230	0.021
2	0.380	0.042	0.286	0.038	0.238	0.033	0.207	0.029
1	0.321	0.078	0.246	0.065	0.206	0.053	0.180	0.044
0	0.246	0.269	0.197	0.172	0.169	0.122	0.150	0.091
HMSE	1.08320		1.23048		1.32241		1.38605	

in (17) is the same as that in (11)). HMSE can be similarly calculated via (22) once one has calculated the optimal relativities $\tilde{\zeta}$.

We first look at the $-1/+1/pen$ system in Table 2(a). For each fixed value of $pen = 0, 1, 2, 3$, most of the population is in the lowest BM level 0 (as $\mathbb{P}(L = 0)$ is at least 82%), which can be explained by the relatively low mean claim frequency of $\lambda_0 = 0.05$. By inspecting $\mathbb{P}(L = \ell)$ for $pen = 0, 1, 2, 3$, we observe that an increase in pen tends to move some of the population toward higher BM levels, leading to a diversification of the stationary BM levels. This is because a higher pen means that more claim-free years are needed to reduce the BM level by one, rendering it more difficult for drivers at a high BM level to move downward. As mentioned in Proposition 1, the optimal relativities under $pen = 0$ are the same as those in Tan et al. (2015). As pen increases, it is important to observe that the optimal relativity $\tilde{\zeta}(\ell)$ for each fixed BM level ℓ decreases. Although a lower BM relativity means that a driver occupying a given BM level pays a less expensive premium in a $-1/+1/pen$ system with a higher pen, the insurer's premium income is in turn compensated by an increased portion of drivers occupying higher BM levels as pen increases. A higher pen also results in an improvement of the HMSE in this example. Similar patterns can be found in the $-1/+2/pen$ system as shown in Table 2(b). However, for a fixed $pen = 1, 2, 3$, the optimal relativity $\tilde{\zeta}(\ell)$ in Table 2(b) is not necessarily increasing in ℓ especially for low BM levels. For example, it is observed that $\tilde{\zeta}(2)$ is slightly smaller than $\tilde{\zeta}(1)$. With $h = 2$, a possible explanation for $\tilde{\zeta}(2) < \tilde{\zeta}(1)$ is that certain policyholders occupying BM level 2 could be those who were in the lowest BM level 0 (and had one claim) in the previous year, and these are still good drivers. On the other hand, those in BM level 1 could not have come directly from BM level 0. It is interesting to note that although $\tilde{\zeta}(\ell)$ is not always increasing in ℓ in Table 2(b), reporting $n \geq 1$ claims moves a policyholder upward by $2n$ BM levels and this always leads to an increase in $\tilde{\zeta}(\ell)$. Nevertheless, potential problems arise when a policyholder moves down from BM level 2 to BM level 1 after a number of claim-free years and then suffers an increase in the optimal relativity (and hence premium). For future work, constrained optimization may be considered by requiring $\tilde{\zeta}(\ell)$ to be increasing in ℓ, but it is unlikely that this will lead to optimal relativities that admit expressions as explicit as those in Propositions 1 and 2.

When $\lambda_0 = 1$, we consider BM levels up to $z = 14$ because the assumption of a higher mean claim frequency will result in a larger proportion of policyholders occupying higher BM levels. The $-1/+1/pen$ system in Table 2(c) shows some similar features: as pen increases, $\tilde{\zeta}(\ell)$ decreases and the distribution (at equilibrium) of policyholders shifts from lower BM levels to higher ones. However, since a significant fraction of the population is already in the highest BM level 14 when $pen = 0$, increasing pen antidiversifies the stationary distribution of BM levels. We observe that HMSE is increased with higher pen.

We also notice one interesting phenomenon when comparing the cases where the claim frequencies are of different magnitude. In Tables 2(a) and (b) for which $\lambda_0 = 0.05$, the optimal relativity $\tilde{\zeta}(0)$ for BM level 0 is still at least 69% while that for the highest BM level can be as high as 17. An intuitive explanation is that given the low claim frequency $\lambda_0 = 0.05$, it is not unusual that drivers do not file claims. As a result, claim-free drivers (those occupying the lowest BM level) are not necessarily much better drivers than others, and therefore they are not rewarded with much premium discount. On the other hand, drivers occupying the highest BM level must have reported some claims in the past years, and they are more likely to be worse drivers with more risky unobserved risk characteristics (that is, higher Θ) and thus are more severely penalized. Moving to Table 2(c) for which $\lambda_0 = 1$, it becomes much more common for drivers to report claims and occupy higher BM levels, so they are not penalized too much with the highest BM relativity $\tilde{\zeta}(14)$ around 2. But it is those at lower BM levels who have claim-free years that should be regarded as significantly better drivers, and they are given a great discount with the BM relativity $\tilde{\zeta}(0)$ no larger than 24.6%.

From this example, we conclude that while the extension of the classical $-1/+h$ system to the $-1/+h/pen$ system provides room for improvement of the predictive power via lowering the HMSE, such an improvement is not always guaranteed.

Example 4. Under the collective risk model with random effects described as Model 2, we assume one risk class only, where the distribution function F_1 in (C.1) for the frequency part follows a Poisson distribution with $\lambda^{[1]} = \lambda_0^{[1]}$ and the distribution function F_2 in (C.1) for the severity part follows a gamma distribution with $\lambda^{[2]} = \lambda_0^{[2]}$ and $\psi^{[2]} = \psi_0^{[2]}$. The marginal distributions of the random effects $(\Theta^{[1]}, \Theta^{[2]})$ are specified as

$$\begin{cases} \Theta^{[1]} \sim \text{Lognormal}(-\sigma_1^2/2, \sigma_1^2) & \text{with} \quad \sigma_1^2 = 0.99, \\ \Theta^{[2]} \sim \text{Lognormal}(-\sigma_2^2/2, \sigma_2^2) & \text{with} \quad \sigma_2^2 = 0.29, \end{cases}$$

with distribution functions G_1 and G_2, respectively, and the dependence is described by a Gaussian copula C with correlation coefficient $\rho = -0.45$ (so that the joint distribution function is given by $H = C(G_1, G_2)$ in (C.4)). With $pen = 0, 1, 2, 3$, we have calculated the values of the stationary probabilities $\mathbb{P}(L = \ell)$ and the optimal BM relativities $\tilde{\zeta}(\ell)$ for all ℓ's together with the HMSE under a $-1/+1/pen$ system for the cases $z = 9$ and $\lambda_0^{[1]} = 0.05$ (Table 3(a)) as well as $z = 14$ and $\lambda_0^{[1]} = 1$ (Table 3(c)) and under a $-1/+2/pen$ system for the case $z = 9$ and $\lambda_0^{[1]} = 0.05$ (Table 3(b)). In constructing Table 3, the parameters $\lambda_0^{[2]} = \exp(8.00)$ and $\psi_0^{[2]} = 1/0.670$ are kept fixed. Note that the values of HMSE in Table 3 are much higher than those in Table 2 concerning Example 3, and that is simply because the HMSE in the present example measures the error concerning the aggregate claim (and the individual severity has large magnitude as $\lambda_0^{[2]} = \exp(8.00)$) whereas the HMSE in Example 3 concerns the error in the claim number. Moreover, some of the optimal relativities in Table 3 are quite different from those in Table 2, showing the importance of taking into account the dependence between claim frequency and claim severity for modeling purposes. Nonetheless, similar patterns to those in Table 2 are observed in Table 3. Note also that the stationary probabil-

ities in Table 3 are identical to their counterparts in Table 2 because the probabilities depend only on the distribution of the claim frequency.

4. DATA ANALYSIS

4.1. SUMMARY OF DATA ESTIMATION RESULTS

In this subsection, we summarize the data estimation results using real data. To examine the effect of dependence on ratemaking, we employ a data set concerning collision coverage for new and old vehicles from the Wisconsin Local Government Property Insurance Fund (LGPIF) (Frees, Lee, and Yang 2016)—detailed information on the project can be found at the LGPIF project website. Such collision coverage provides coverage for impact of vehicle with an object, impact of vehicle with an attached vehicle, and overturn of a vehicle. The observations include policyholders who have either new collision coverage or old collision coverage, or both. In our data analysis, longitudinal data over the policy years from 2006 to 2010 with 497 governmental entities are used. There are two categorical variables: the entity type with six levels, and the coverage with three levels as shown in Table 4 (which is the same as Table C.6 in Oh, Shi, and Ahn (2020)).

Specifically, for Model 1 in the data analysis, as in (B.1) we consider

$$N_{it}|(\Theta_i=\theta_i,\mathbf{X}_i=\mathbf{x}_i)\sim \text{Poisson}(\lambda_i\theta_i) \quad \text{with} \quad \lambda_i:=\exp(\mathbf{x}_i\beta),$$

where $\mathbf{x}_i$ is a row vector containing the i-th policyholder's information of the categorical variables in Table 4 for the frequency, and the column vector β contains the corresponding parameters. For the random effect Θ, we assume a lognormal distribution with mean 1 and variance of $\exp(\sigma^2)-1$, that is,

$$\Theta\sim \text{Lognormal}(-\sigma^2/2,\sigma^2).$$

Summary statistics of the posterior samples for the parameters in Model 1 using the Bayesian approach are presented in Table 5. The table includes the posterior median (Est), the posterior standard deviation (Std.dev), and the 95% highest posterior density Bayesian credible interval (95% CI). Note that an asterisk, $*$, indicates that the parameters are significant at the 0.05 level. In estimating the parameters in Table 5, we have run 30,000 Markov chain Monte Carlo (MCMC) iterations saving every fifth sample after a burn-in of 15,000 iterations by using the JAGS program. Standard MCMC diagnostics gave no indication of lack of convergence. For Model 2, in line with (C.1) and (C.2), we assume

$$N_{it}|(\Theta_i^{[1]}=\theta_i^{[1]},\mathbf{X}_i^{[1]}=\mathbf{x}_i^{[1]})\sim \text{Poisson}(\lambda_i^{[1]}\theta_i^{[1]})$$

for the frequency, and

$$Y_{itj}|(\Theta_i^{[2]}=\theta_i^{[2]},\mathbf{X}_i^{[2]}=\mathbf{x}_i^{[2]})\sim \text{Gamma}(\lambda_i^{[2]}\theta_i^{[2]},1/\psi^{[2]})$$

for the individual claim severity, where $\lambda_i^{[2]}\theta_i^{[2]}$ is the mean and $1/\psi^{[2]}$ is the shape parameter of the gamma distribution. With log link, we assume

$$\lambda_i^{[1]}=\exp(\mathbf{x}_i^{[1]}\beta^{[1]}) \quad \text{and} \quad \lambda_i^{[2]}=\exp(\mathbf{x}_i^{[2]}\beta^{[2]}),$$

where we take $\mathbf{x}_i^{[1]}=\mathbf{x}_i^{[2]}$, and $\beta^{[1]}$ and $\beta^{[2]}$ are the corresponding parameters.

For the bivariate random effects $(\Theta^{[1]},\Theta^{[2]})$, we assume a Gaussian copula C in (C.4) with correlation coefficient ρ, and the marginal distributions are assumed to follow lognormal distributions with different parameters specified as

$$\begin{cases}\Theta^{[1]} &\sim \text{Lognormal}(-\sigma_1^2/2,\sigma_1^2),\\ \Theta^{[2]} &\sim \text{Lognormal}(-\sigma_2^2/2,\sigma_2^2).\end{cases}$$

The estimation results from Oh, Shi, and Ahn (2020) are summarized in Table 7, and we refer interested readers to Oh, Shi, and Ahn (2020) for the details of the estimation procedure and further results. In particular, the dependence parameter of the Gaussian copula is estimated to be $\rho=-0.447$, suggesting a significant negative dependence between the random effects $\Theta^{[1]}$ and $\Theta^{[2]}$ of the frequency and the severity, respectively.

4.2. ANALYSIS OF OPTIMAL RELATIVITIES IN A MODIFIED BMS

Using the posterior median as estimates of the parameters in Models 1 and 2 when $z=14$, we calculate the optimal relativity and stationary probability for each BM level as well as the values of HMSE under the modified BM transition rules for a $-1/+1/pen$ system and a $-1/+2/pen$ system when $pen=0,1,2,3$. The results are summarized in Tables 6 and 8. We remark that unlike in Examples 3 and 4, where the stationary probabilities $\mathbb{P}(L=\ell)$ are identical across Tables 2 and 3, the stationary probabilities in Tables 6 and 8 are close but not identical. The reason is that the frequency parameters for Model 2 in Table 7 are estimated jointly with the severity parameters and therefore the results are slightly different from those for Model 1 given in Table 5, where only frequency parameters are estimated. Note also that the parameters are estimated using MCMC sampling, which has also possibly contributed to some differences.

Comparing across Tables 6 and 8, one can observe the impact of frequency-severity dependence on the optimal relativities. In particular, under negative dependence between frequency and severity, which is implied from the negative dependence between the two random effects with $\rho=-0.447$, the optimal relativities tend to decrease for higher BM levels but increase for lower BM levels. Consequently, the optimal relativities in Table 8 are less spread out than those in Table 6. It is noted that all optimal BM relativities in Table 8 are less than 1, which may be counterintuitive at first glance. However, it is noted that, roughly speaking, $\tilde{\zeta}(\ell)$ may be regarded as an estimate of $\Theta^{[1]}\Theta^{[2]}$ when $L^*=(\ell)_a$ for some a (see (10) and the optimization (13) in Model 2). Since $\Theta^{[1]}$ and $\Theta^{[2]}$ are negatively dependent (with both having mean 1), when one of them is large the other is more likely to be small, making the product unlikely to be larger than 1.

Both Tables 6 and 8 confirm that as pen increases, each optimal relativity $\tilde{\zeta}(\ell)$ (for fixed ℓ) decreases and the proportion of policyholders in higher BM levels increases. Explanations similar to those in Examples 3 and 4 are applicable. A certain degree of diversification effects on the BM levels as pen increases resembling Tables 2(a) and (b) in Example 3 and Tables 3(a) and (b) in Example 4 can be ob-

Table 3. (Example 4) Distribution of L and optimal relativities under the frequency-severity Model 2 with dependence

(a) $-1/+1/pen$ systems with various pen for $\lambda_0^{[1]} = 0.05$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
9	7.217	0.001	5.749	0.002	4.917	0.004	4.366	0.008
8	6.246	0.000	4.715	0.001	3.889	0.002	3.357	0.002
7	5.573	0.000	4.259	0.001	3.540	0.002	3.073	0.002
6	5.000	0.000	3.834	0.001	3.202	0.002	2.792	0.003
5	4.695	0.001	3.404	0.001	2.854	0.003	2.500	0.004
4	3.805	0.001	2.936	0.002	2.479	0.004	2.184	0.006
3	3.065	0.002	2.407	0.005	2.060	0.008	1.834	0.012
2	2.210	0.007	1.812	0.015	1.591	0.023	1.443	0.030
1	1.369	0.044	1.203	0.073	1.101	0.095	1.026	0.111
0	0.727	0.944	0.692	0.898	0.665	0.858	0.641	0.821
HMSE	14179.63		13189.89		12525.65		12053.38	

(b) $-1/+2/pen$ systems with various pen for $\lambda_0^{[1]} = 0.05$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
9	4.876	0.002	4.069	0.007	3.582	0.013	3.247	0.020
8	4.093	0.002	3.170	0.004	2.647	0.005	2.307	0.007
7	3.568	0.002	2.882	0.004	2.491	0.005	2.229	0.006
6	3.005	0.003	2.389	0.005	2.045	0.008	1.819	0.011
5	2.672	0.003	2.239	0.005	1.985	0.007	1.809	0.008
4	2.067	0.007	1.706	0.014	1.509	0.020	1.376	0.027
3	1.887	0.009	1.656	0.012	1.510	0.013	1.403	0.014
2	1.259	0.041	1.115	0.066	1.029	0.086	0.966	0.101
1	1.205	0.038	1.117	0.034	1.053	0.030	1.002	0.028
0	0.684	0.892	0.655	0.850	0.631	0.812	0.611	0.778
HMSE	13230.45		12675.38		12303.48		12042.78	

(c) $-1/+1/pen$ systems with various pen for $\lambda_0^{[1]} = 1$

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	1.439	0.311	1.230	0.463	1.123	0.562	1.055	0.634
13	0.939	0.089	0.711	0.060	0.590	0.045	0.513	0.035
12	0.796	0.044	0.634	0.036	0.539	0.030	0.475	0.025
11	0.717	0.027	0.581	0.025	0.500	0.022	0.444	0.020
10	0.666	0.020	0.542	0.019	0.468	0.018	0.418	0.016
9	0.627	0.016	0.510	0.016	0.442	0.015	0.395	0.014
8	0.595	0.014	0.483	0.015	0.419	0.014	0.375	0.013
7	0.567	0.014	0.459	0.014	0.397	0.014	0.356	0.013
6	0.540	0.014	0.436	0.015	0.377	0.014	0.337	0.013
5	0.513	0.016	0.413	0.016	0.357	0.016	0.319	0.014
4	0.483	0.020	0.388	0.020	0.335	0.018	0.300	0.017
3	0.450	0.027	0.361	0.026	0.313	0.024	0.280	0.021
2	0.409	0.042	0.330	0.038	0.287	0.033	0.258	0.029
1	0.359	0.078	0.293	0.065	0.256	0.053	0.232	0.044
0	0.291	0.269	0.246	0.172	0.219	0.122	0.201	0.091
HMSE	4649638		5052378		5328804		5532048	

Table 4. (Data analysis) Observable policy characteristics used as covariates

Categorical variables	Description	Proportions
Entity type	Type of local government entity	
	Miscellaneous	5.03%
	City	9.66%
	County	11.47%
	School	36.42%
	Town	16.90%
	Village	20.52%
Coverage	Collision coverage amount for old and new vehicles	
	Coverage $\in (0, 0.14]$ = 1	33.40%
	Coverage $\in (0.14, 0.74]$ = 2	33.20%
	Coverage $\in (0.74, \infty)$ = 3	33.40%

Table 5. (Data analysis) Estimation results under the frequency-only Model 1

parameter	Est	Std.dev	95% CI		
			lower	upper	
Fixed effect					
(Intercept)	-2.798	0.312	-3.433	-2.225	*
Type=City	0.601	0.337	-0.044	1.280	
Type=County	1.923	0.329	1.257	2.543	*
Type=School	0.438	0.303	-0.172	1.023	
Type=Town	-1.343	0.380	-2.087	-0.600	*
Type=Village	-0.005	0.320	-0.630	0.618	
Coverage=2	1.254	0.214	0.820	1.651	*
Coverage=3	2.156	0.231	1.688	2.587	*
Random effect					
σ^2	0.993	0.14	0.733	1.276	*

served from Tables 6 and 8. Meanwhile, the HMSE values show that a higher pen leads to a decrease in predictive power in this example. The higher HMSE in Table 8 compared with Table 6 is again attributed to the fact that Model 2 is concerned with the aggregate claim while Model 1 is about claim number, and interested readers are referred to Tables A4 and A5 in Oh, Shi, and Ahn (2020) for the magnitude of the average claim severity in Model 2.

Lastly, we can compare the stationary distribution L of the BM level and the optimal relativities $\tilde{\zeta}(\ell)$ under two different BM systems: (i) a $-1/+1/2$ system with a small increase of BM level per claim ($h = 1$) but a period of penalty ($pen = 2$); and (ii) a $-1/+2/0$ system with a larger increase of BM level per reported claim ($h = 2$) but without a period of penalty ($pen = 0$). From Tables 6(a) and (b) concerning Model 1, the presence of a period of penalty leads to a smaller proportion of policyholders in the lowest level 0 (with $\mathbb{P}(L = 0) = 0.505$ in the former model and $\mathbb{P}(L = 0) = 0.542$ in the latter) but a larger proportion in the highest level 14 (with $\mathbb{P}(L = 14) = 0.237$ in the former model and $\mathbb{P}(L = 14) = 0.173$ in the latter). For a given BM level ℓ, the optimal relativity $\tilde{\zeta}(\ell)$ in the former model is always lower than the latter, suggesting that a BMS with a period of penalty instead of a higher increase of the BM level per claim can look more attractive to the market if the relativities are available to potential customers. Such an effect is even more pronounced if one assumes $pen = 3$ instead of $pen = 2$. Hence, this numerical example clearly illustrates that we can expect different outcomes in terms of the distribution of the BM level as well as the optimal relativities depending on how more rigid the BMS is designed. Instead of (or in addition to) imposing a higher increase of the BM level per claim, our model makes it harder for a policyholder to transit to lower BM levels once he or she has reported a claim by requiring consecutive and multiple claim-free years to enjoy a bonus. We also remark that the same conclusion can be drawn from Tables 8(a) and (b) by comparing the $-1/+1/2$ (or $-1/+1/3$) system and the $-1/+2/0$ system under Model 2.

Table 6. (Data analysis) Distribution of L and optimal relativities under frequency-only Model 1

(a) $-1/+1/pen$ system with various pen								
pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	1.295	0.143	1.167	0.197	1.114	0.237	1.086	0.268
13	0.484	0.031	0.362	0.022	0.312	0.018	0.285	0.016
12	0.385	0.015	0.321	0.014	0.290	0.013	0.271	0.012
11	0.342	0.009	0.298	0.010	0.276	0.010	0.262	0.010
10	0.319	0.007	0.284	0.008	0.266	0.008	0.255	0.009
9	0.304	0.006	0.274	0.007	0.259	0.008	0.249	0.008
8	0.293	0.005	0.266	0.007	0.253	0.007	0.244	0.008
7	0.285	0.005	0.260	0.007	0.248	0.008	0.241	0.009
6	0.277	0.006	0.255	0.007	0.244	0.009	0.238	0.010
5	0.271	0.007	0.251	0.009	0.241	0.010	0.235	0.012
4	0.265	0.009	0.247	0.012	0.237	0.014	0.233	0.015
3	0.259	0.014	0.242	0.018	0.234	0.021	0.229	0.023
2	0.252	0.026	0.237	0.034	0.229	0.038	0.225	0.041
1	0.243	0.072	0.230	0.087	0.223	0.094	0.218	0.098
0	0.228	0.645	0.217	0.562	0.210	0.505	0.205	0.462
HMSE	2.42280		2.54650		2.60334		2.63588	

(b) $-1/+2/pen$ system with various pen

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	1.240	0.173	1.130	0.236	1.088	0.280	1.065	0.315
13	0.452	0.047	0.346	0.031	0.301	0.024	0.276	0.020
12	0.354	0.025	0.305	0.021	0.278	0.018	0.262	0.017
11	0.312	0.017	0.282	0.015	0.265	0.015	0.254	0.014
10	0.288	0.013	0.267	0.013	0.255	0.013	0.246	0.013
9	0.273	0.011	0.258	0.011	0.248	0.012	0.242	0.011
8	0.263	0.010	0.250	0.011	0.242	0.013	0.236	0.013
7	0.256	0.010	0.245	0.011	0.239	0.011	0.234	0.011
6	0.250	0.011	0.240	0.014	0.233	0.016	0.228	0.017
5	0.245	0.012	0.237	0.013	0.232	0.013	0.228	0.012
4	0.240	0.018	0.231	0.023	0.225	0.027	0.221	0.030
3	0.237	0.019	0.230	0.019	0.225	0.017	0.221	0.016
2	0.230	0.050	0.221	0.064	0.216	0.072	0.211	0.077
1	0.228	0.043	0.221	0.033	0.217	0.026	0.213	0.022
0	0.213	0.542	0.206	0.486	0.201	0.444	0.196	0.410
HMSE	2.48426		2.58912		2.63540		2.66122	

Table 7. (Data analysis) Estimation results under the frequency-severity Model 2 with dependence

			95% CI		
parameter	Est	Std.dev	lower	upper	
Frequency part					
Intercept	-2.767	0.318	-3.417	-2.153	*
City	0.597	0.337	-0.051	1.272	
County	1.907	0.335	1.271	2.587	*
School	0.411	0.304	-0.181	1.014	
Town	-1.351	0.384	-2.103	-0.584	*
Village	-0.012	0.323	-0.626	0.654	
Coverage2	1.247	0.212	0.829	1.667	*
Coverage3	2.139	0.230	1.713	2.615	*
Severity part					
Intercept	8.829	0.375	8.103	9.588	*
City	-0.036	0.353	-0.737	0.637	
County	0.341	0.338	-0.336	0.980	
School	-0.173	0.328	-0.805	0.484	
Town	0.497	0.440	-0.356	1.349	
Village	0.316	0.346	-0.357	0.994	
Coverage2	0.180	0.244	-0.308	0.646	
Coverage3	-0.027	0.261	-0.533	0.493	
$1/\psi^{[2]}$	0.670	0.041	0.592	0.752	*
Copula part					
σ_1^2	0.992	0.142	0.746	1.292	*
σ_2^2	0.293	0.067	0.176	0.433	*
ρ	-0.447	0.130	-0.690	-0.190	*

5. CONCLUSION

In automobile third-party liability insurance, the BMS has been broadly used as an a posteriori ratemaking mechanism to help insurers rate a policyholder's risk more accurately and thus calculate premium more fairly to reflect policyholder risk. The BMS is also designed to stimulate drivers to practice safer driving by adjusting the premium on policy renewal based on the claim history in the previous year. However, a typical BMS that immediately offers a reward to policyholders without a claim in the previous year may tend to move policyholders easily toward lower BM levels. Hence, to resolve the imbalance due to a concentration of policyholders in the lowest BM level, prevent a quick recovery (in terms of paying lower premium) for those drivers who have a good claim history for only a single period, and better distinguish between drivers who are consistently good and those who are only temporarily good, in this paper we introduce a "period of penalty" to count the number of consecutive claim-free years required to lower the BM level. Although other more rigid BMS transition rules such as a $-1/+h$ system with $h = 2$ or $h = 3$ to severely penalize claims in the previous year may put less pressure on the premium income of the insurer, they may weaken the product's competitiveness in the market and insurers may lose better customers. Through numerical illustrations with a data set, we demonstrate that compared with a $-1/+2$ system, the introduction of a penalty period to a $-1/+1$ system can result in (i) lower values of optimal BM relativities, which can potentially improve marketability of the product; and (ii) better separation of risks because the BM level 0 is less concentrated as drivers who are not consistently good are moved to higher BM levels. Once policyholders are aware of the requirement of consecutive claim-free years to enjoy lower BM relativities at lower BM levels, they are likely better motivated to drive more safely.

Moreover, in the aforementioned modified BMS with a penalty period, we take into account the randomness of (i) frequency only or (ii) both frequency and severity when modeling the unobserved risk characteristics and constructing the optimal relativities associated with the BM levels. Those relativities are then used to determine the premium actually charged to the policyholders staying in the corresponding BM levels. We hope that our extended BMS allowing for dependency between the frequency and the severity of claims via dependency of the unobserved risk characteristics provides some insights for improving the classical BMS, in particular, when the insurers have the freedom to alter the features of the BMS to reflect claim experience more accurately based on the claim history of policyholders in multiple years with product competitiveness in mind.

Table 8. (Data analysis) Distribution of L and optimal relativities under the frequency-severity Model 2 with dependence

(a) $-1/+1/pen$ system with various pen

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	0.968	0.144	0.888	0.199	0.855	0.238	0.837	0.270
13	0.459	0.031	0.356	0.022	0.309	0.018	0.282	0.016
12	0.376	0.015	0.318	0.014	0.287	0.013	0.268	0.012
11	0.338	0.010	0.295	0.010	0.272	0.010	0.259	0.010
10	0.316	0.007	0.280	0.008	0.262	0.008	0.252	0.009
9	0.301	0.006	0.270	0.007	0.255	0.008	0.247	0.008
8	0.290	0.005	0.262	0.007	0.250	0.008	0.243	0.008
7	0.281	0.005	0.256	0.007	0.245	0.008	0.240	0.009
6	0.274	0.006	0.251	0.007	0.242	0.009	0.239	0.010
5	0.267	0.007	0.247	0.009	0.240	0.011	0.237	0.012
4	0.261	0.009	0.244	0.012	0.238	0.014	0.237	0.015
3	0.256	0.014	0.241	0.018	0.237	0.021	0.237	0.023
2	0.250	0.026	0.239	0.034	0.237	0.038	0.237	0.041
1	0.245	0.072	0.238	0.088	0.238	0.095	0.238	0.098
0	0.240	0.643	0.237	0.560	0.236	0.503	0.236	0.460

HMSE	91069385	95366837	97457899	98678459

(b) $-1/+2/pen$ system with various pen

pen	0		1		2		3	
Level ℓ	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$	$\tilde{\zeta}(\ell)$	$\mathbb{P}(L=\ell)$
14	0.933	0.174	0.865	0.238	0.838	0.282	0.824	0.316
13	0.431	0.047	0.341	0.031	0.298	0.024	0.275	0.021
12	0.348	0.025	0.302	0.021	0.277	0.018	0.262	0.017
11	0.309	0.017	0.280	0.016	0.263	0.015	0.254	0.014
10	0.286	0.013	0.265	0.013	0.254	0.013	0.247	0.013
9	0.271	0.011	0.256	0.011	0.248	0.012	0.244	0.011
8	0.260	0.010	0.249	0.011	0.244	0.013	0.241	0.014
7	0.253	0.010	0.245	0.011	0.241	0.011	0.240	0.011
6	0.248	0.011	0.241	0.014	0.239	0.016	0.239	0.017
5	0.244	0.012	0.240	0.013	0.239	0.013	0.239	0.013
4	0.241	0.018	0.238	0.023	0.238	0.027	0.239	0.030
3	0.240	0.019	0.238	0.019	0.238	0.017	0.239	0.016
2	0.238	0.050	0.238	0.064	0.239	0.072	0.239	0.077
1	0.238	0.043	0.238	0.033	0.239	0.026	0.239	0.022
0	0.237	0.540	0.236	0.483	0.236	0.441	0.235	0.407
HMSE	93352237		96970265		98659783		99617247	

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for helpful comments and suggestions that improved earlier versions of the manuscript. This research was supported by a Committee on Knowledge Extension Research grant funded by the Casualty Actuarial Society (CAS). The views expressed herein are those of the authors and are not necessarily those of the CAS.

Submitted: April 20, 2019 EDT. Accepted: May 26, 2021 EDT. Published: November 21, 2022 EDT.

REFERENCES

Baumgartner, Carolin, Lutz F. Gruber, and Claudia Czado. 2015. "Bayesian Total Loss Estimation Using Shared Random Effects." *Insurance: Mathematics and Economics* 62 (May):194–201. https://doi.org/10.1016/j.insmatheco.2015.02.008.

Czado, Claudia, Rainer Kastenmeier, Eike Christian Brechmann, and Aleksey Min. 2012. "A Mixed Copula Model for Insurance Claims and Claim Sizes." *Scandinavian Actuarial Journal* 2012 (4): 278–305. https://doi.org/10.1080/03461238.2010.546147.

de Jong, Piet, and Gillian Z. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge University Press. https://doi.org/10.1017/cbo9780511755408.

Denuit, Michel, Xavier Maréchal, Sandra Pitrebois, and Jean-François Walhin. 2007. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Chichester, West Sussex, UK: Wiley. https://doi.org/10.1002/9780470517420.

Frees, Edward W., Gee Lee, and Lu Yang. 2016. "Multivariate Frequency-Severity Regression Models in Insurance." *Risks* 4 (1): 4. https://doi.org/10.3390/risks4010004.

Frees, Edward W., and Emiliano A. Valdez. 1998. "Understanding Relationships Using Copulas." *North American Actuarial Journal* 2 (1): 1–25. https://doi.org/10.1080/10920277.1998.10595667.

Garrido, J., C. Genest, and J. Schulz. 2016. "Generalized Linear Models for Dependent Frequency and Severity of Insurance Claims." *Insurance: Mathematics and Economics* 70 (September):205–15. https://doi.org/10.1016/j.insmatheco.2016.06.006.

Hernandez-Bastida, A., M.P. Fernandez-Sanchez, and E. Gomez-Deniz. 2009. "The Net Bayes Premium with Dependence between the Risk Profiles." *Insurance: Mathematics and Economics* 45 (2): 247–54.

Jeong, H., E.A. Valdez, J.Y. Ahn, and S.C. Park. 2021. "Generalized Linear Mixed Models for Dependent Compound Risk Models." *Variance* 14 (1).

Lemaire, Jean. 1985. *Automobile Insurance: Actuarial Models*. Boston: Kluwer Academic.

———. 1995. *Bonus-Malus Systems in Automobile Insurance*. Boston: Springer Netherlands. https://doi.org/10.1007/978-94-011-0631-3.

———. 1998. "Bonus-Malus Systems: The European and Asian Approach to Merit-Rating." *North American Actuarial Journal* 2 (1): 26–38. https://doi.org/10.1080/10920277.1998.10595668.

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

Oh, Rosy, Kyung Suk Lee, Sojung C. Park, and Jae Youn Ahn. 2020. "Double-Counting Problem of the Bonus-Malus System." *Insurance: Mathematics and Economics* 93:141–55.

Oh, Rosy, Peng Shi, and Jae Youn Ahn. 2020. "Bonus-Malus Premiums under the Dependent Frequency-Severity Modeling." *Scandinavian Actuarial Journal* 2020 (3): 172–95. https://doi.org/10.1080/03461238.2019.1655477.

Park, Sojung C., Joseph H.T. Kim, and Jae Youn Ahn. 2018. "Does Hunger for Bonuses Drive the Dependence between Claim Frequency and Severity?" *Insurance: Mathematics and Economics* 83:32–46.

Pitrebois, S., M. Denuit, and J.F. Walhin. 2003a. "Setting a Bonus-Malus Scale in the Presence of Other Rating Factors: Taylor's Work Revisited." *ASTIN Bulletin* 33 (2): 419–36.

———. 2003b. "Marketing and Bonus-Malus Systems." Presented at the 34th International ASTIN Colloquium, Berlin, August 24–27.

Shi, Peng, Xiaoping Feng, and Anastasia Ivantsova. 2015. "Dependent Frequency–Severity Modeling of Insurance Claims." *Insurance: Mathematics and Economics* 64 (September):417–28. https://doi.org/10.1016/j.insmatheco.2015.07.006.

Tan, Chong It, Jackie Li, Johnny Siu-Hang Li, and Uditha Balasooriya. 2015. "Optimal Relativities and Transition Rules of a Bonus–Malus System." *Insurance: Mathematics and Economics* 61 (March):255–63. https://doi.org/10.1016/j.insmatheco.2015.02.001.

Taylor, Greg. 1997. "Setting a Bonus-Malus Scale in the Presence of Other Rating Factors." *ASTIN Bulletin* 27 (2): 319–27. https://doi.org/10.2143/ast.27.2.542055.

APPENDICES

APPENDIX A. GENERALIZED LINEAR MODELS

The *exponential dispersion family* (EDF) in McCullagh and Nelder (1989) can be considered for modeling the random components of the frequency and the severity of insurance claims in generalized linear models. The EDF with mean μ and dispersion ψ, whose distribution function is denoted by $F(\cdot;\mu,\psi)$, has the probability density/mass function (in y)

$$p(y|\vartheta,\psi)=\exp[(y\vartheta-b(\vartheta))/\psi+c(y,\psi)],$$

where ϑ is the canonical parameter and $b(\cdot)$ and $c(\cdot)$ are predetermined functions. The mean of the distribution can be expressed as $\mu=b'(\vartheta)$ and the variance is $b''(\vartheta)\psi\equiv V(\mu)$, where the inverse of $b'(\cdot)$ is known as the canonical link function and $V(\cdot)$ is called the variance function.

APPENDIX B. MODEL 1

In the frequency random effects model:

1. For the i-th policyholder, the conditional distribution of the number of claims N_{it} in the t-th policy year given the observed risk characteristics $\mathbf{X}_i=\mathbf{x}_i$ and unobserved risk characteristics $\Theta_i=\theta_i$ is specified as
$$N_{it}|(\Theta_i=\theta_i,\mathbf{X}_i=\mathbf{x}_i)\overset{\text{i.i.d.}}{\sim}F(\cdot;\lambda_i\theta_i,\psi),\quad\text{(B.1)}$$
where the distribution function F has mean parameter $\lambda_i\theta_i$ with $\lambda_i=\eta^{-1}(\mathbf{x}_i\beta)$ and some parameter ψ. The "i.i.d." in (B.1) means that, conditional on $\Theta_i=\theta_i$ and $\mathbf{X}_i=\mathbf{x}_i$, the claim numbers N_{it}'s for different t's are independent and identically distributed. When F is in the class of EDF in Appendix A, ψ is the dispersion parameter.
2. The random variable Θ_i for the i-th policyholder's unobserved risk characteristics concerning the claim frequency is assumed to be independent of the observed risk characteristics $\mathbf{X}_i$. The Θ_i's are i.i.d. having distribution function G and density $g=G'$, and we write
$$\Theta_i\overset{\text{i.i.d.}}{\sim}G.\quad\text{(B.2)}$$
A generic variable with distribution G is denoted by Θ, and we assume $\mathbb{E}[\Theta]=1$ for convenience.

APPENDIX C. MODEL 2

To account for various forms of dependence such as frequency-frequency, severity-severity, and frequency-severity, we consider a copula-based random effects model as in (Oh, Shi, and Ahn 2020):

1. The frequency component N_{it} is specified using a count regression model conditional on the observed risk characteristics $\mathbf{X}_i^{[1]}$ and the unobserved risk $\Theta_i^{[1]}$ such that
$$N_{it}|(\Theta_i^{[1]}=\theta_i^{[1]},\mathbf{X}_i^{[1]}=\mathbf{x}_i^{[1]})\overset{\text{i.i.d.}}{\sim}F_1(\cdot;\lambda_i^{[1]}\theta_i^{[1]},\psi^{[1]}),\quad\text{(C.1)}$$
where the distribution function F_1 has mean parameter $\lambda_i^{[1]}\theta_i^{[1]}$ with $\lambda_i^{[1]}=\eta_1^{-1}(\mathbf{x}_i^{[1]}\beta^{[1]})$ and some parameter $\psi^{[1]}$. The "i.i.d." in (C.1) means that N_{it}'s for different t's are (conditionally) independent and identically distributed. If F_1 belongs to EDF, then $\psi^{[1]}$ corresponds to the dispersion parameter. Here we implicitly assume that the conditional distribution in (C.1) does not change even if we are given further information on the risk characteristics $\mathbf{X}_i^{[2]}$ and $\Theta_i^{[2]}$ pertaining to the severity.
2. The distribution of the i-th policyholder's j-th claim severity Y_{itj} in the t-th year conditional on the observed risk characteristics $\mathbf{X}_i^{[2]}$ and the unobserved risk characteristics $\Theta_i^{[2]}$ is specified as
$$Y_{itj}|(\Theta_i^{[2]}=\theta_i^{[2]},\mathbf{X}_i^{[2]}=\mathbf{x}_i^{[2]})\overset{\text{i.i.d.}}{\sim}F_2(\cdot;\lambda_i^{[2]}\theta_i^{[2]},\psi^{[2]}),\quad\text{(C.2)}$$
where the distribution function F_2 is in EDF with mean parameter $\lambda_i^{[2]}\theta_i^{[2]}$ (where $\lambda_i^{[2]}=\eta_2^{-1}(\mathbf{x}_i^{[2]}\beta^{[2]})$) and dispersion parameter $\psi^{[2]}$. The "i.i.d." in (C.2) means that Y_{itj}'s for different t's and different j's are (conditionally) independent and identically distributed. It is understood that the conditional distribution in (C.2) remains the same even if we are further given the risk characteristics $\mathbf{X}_i^{[1]}$ and $\Theta_i^{[1]}$ pertaining to the frequency. By the property of EDF, the distribution of the average claim amount M_{it} (see (1)) conditional on the observed risk characteristics $\mathbf{X}_i^{[2]}$, the unobserved risk characteristics $\Theta_i^{[2]}$, and the frequency N_{it} is specified via
$$M_{it}\mid(N_{it}=n_{it},\Theta_i^{[2]}=\theta_i^{[2]},\mathbf{X}_i^{[2]}=\mathbf{x}_i^{[2]})\overset{\text{i.i.d.}}{\sim}F_2(\cdot;\lambda_i^{[2]}\theta_i^{[2]},\psi^{[2]}/n_{it}),\quad n_{it}>0,\quad\text{(C.3)}$$
with updated dispersion parameter $\psi^{[2]}/n_{it}$, and
$$\mathbb{P}(M_{it}=0|N_{it}=n_{it},\Theta_i^{[2]}=\theta_i^{[2]},\mathbf{X}_i^{[2]}=\mathbf{x}_i^{[2]})=1,n_{it}=0.$$
The "i.i.d." in (C.3) means that M_{it}'s for different t's are (conditionally) independent and identically distributed.
3. The bivariate random vector of unobserved risk characteristics $(\Theta_i^{[1]},\Theta_i^{[2]})$ of the i-th policyholder, which is assumed to be independent of the observed risk characteristics $(\mathbf{X}_i^{[1]},\mathbf{X}_i^{[2]})$, has joint distribution function H specified via
$$(\Theta_i^{[1]},\Theta_i^{[2]})\overset{\text{i.i.d.}}{\sim}H=C(G_1,G_2),\quad\text{(C.4)}$$
where G_1 and G_2 denote the marginal distribution functions of $\Theta_i^{[1]}$ and $\Theta_i^{[2]}$, respectively, and C is a bivariate copula. The "i.i.d." in (C.4) means that $(\Theta_i^{[1]},\Theta_i^{[2]})$'s for different i's are independent and identically distributed. A generic pair of $(\Theta_i^{[1]},\Theta_i^{[2]})$ will be denoted by $(\Theta^{[1]},\Theta^{[2]})$. We shall use g_1, g_2, and h to denote the density versions of G_1, G_2, and H, respectively. For convenience, it is further assumed that
$$\mathbb{E}[\Theta^{[1]}]=1\quad\text{and}\quad\mathbb{E}[\Theta^{[2]}]=1.$$