Probability in the Engineering and Informational Sciences (2022), **36**:4 963–987
doi:10.1017/S0269964821000188

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Designing a Bonus-Malus system reflecting the claim size under the dependent frequency–severity model

Rosy Oh[1], Joseph H.T. Kim[2] and Jae Youn Ahn[3]

[1] Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea. E-mail: rosy.oh5@gmail.com
[2] Department of Applied Statistics, Department of Statistics and Data Science, College of Business and Economics, Yonsei University, Seoul, Korea. E-mail: jhtkim@yonsei.ac.kr
[3] Department of Statistics, Ewha Womans University, Seoul, Korea. E-mail: jaeyahn@ewha.ac.kr.

Keywords: Bivariate random effects model, Bonus-Malus system, Dependence, Generalized linear model, Rate making

JEL Classification: C300

Abstract

In the auto insurance industry, a Bonus-Malus System (BMS) is commonly used as a posteriori risk classification mechanism to set the premium for the next contract period based on a policyholder's claim history. Even though the recent literature reports evidence of a significant dependence between frequency and severity, the current BMS practice is to use a frequency-based transition rule while ignoring severity information. Although Oh *et al.* [(2020). Bonus-Malus premiums under the dependent frequency-severity modeling. *Scandinavian Actuarial Journal* 2020(3): 172–195] claimed that the frequency-driven BMS transition rule can accommodate the dependence between frequency and severity, their proposal is only a partial solution, as the transition rule still completely ignores the claim severity and is unable to penalize large claims. In this study, we propose to use the BMS with a transition rule based on both frequency and size of claim, based on the bivariate random effect model, which conveniently allows dependence between frequency and severity. We analytically derive the optimal relativities under the proposed BMS framework and show that the proposed BMS outperforms the existing frequency-driven BMS. Later, numerical experiments are also provided using both hypothetical and actual datasets in order to assess the effect of various dependencies on the BMS risk classification and confirm our theoretical findings.

1. Introduction

In the auto insurance industry, the ratemaking process begins by classifying a given policyholder into different risk classes at the outset of the contract. This initial or a priori ratemaking process is then followed by the a posteriori ratemaking system in subsequent policy years based on the claim history of the policyholder. For practitioners, this a posteriori system for the premium prediction essentially constitutes a Markov chain, a standard stochastic process, and is known as the Bonus-Malus System (BMS). The BMS system typically consists of three elements: the Bonus-Malus (BM) levels, transition rules to navigate the BM levels, and the relativity attached to each BM level. In the classical BMS, the transition rule is governed by frequency; it ignores the severity information, which implies that the future loss of a policyholder can be appropriately modeled by predicting the frequency only (see, e.g., [5,13,22]). In fact, this frequency-driven BMS transition rule is the standard practice in many jurisdictions. It carries an implicit assumption that the frequency and severity are independent, such that the premium can be simply computed as a product of the mean frequency and mean severity. Theoretically, the structure of the traditional BMS is directly related to the classical theory of the

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

collective risk model, which assumes independence between frequency and severity for mathematical tractability and convenience [12].

However, a series of recent empirical studies have shown that the dependence between frequency and severity in auto insurance is statistically significant [8,9]. This phenomenon invalidates the practice of using frequency-driven BMS and highlights the need to extend the classical collective risk model by allowing some dependence structure between frequency and severity. The existing studies on dependent frequency–severity models and the associated insurance premiums include copula-based models [4,8], two-step frequency–severity models [7,9,18,21], and bivariate random effect-based models [1,11,14,20].

The random effect model is especially popular in insurance ratemaking because of the mathematical tractability in its prediction. The bivariate random effect model consists of two random effect components. The first random effect induces the dependence among frequencies and the second induces the dependence among individual severities. These two random effects are then jointly modeled to induce the dependence between frequency and severities at the distribution level. Statistical methods for prediction using the random effect model are well developed in the statistical and insurance literature, such as in Pinquet [20], Boudreault *et al.* [2], Hernández-Bastida *et al.* [11], Baumgartner *et al.* [1], Lu [14], and Cheung *et al.* [3].

While statistical modeling and analyses for accommodating the various dependence structures in the collective risk model have been actively conducted, incorporating such a dependence structure in the BMS is largely overlooked in the literature. This lacuna could be attributed to the fact that developing and constructing such a BMS structure is not straightforward once the frequency–severity relationship becomes dependent in complicated ways. An exception is the recent work of Oh *et al.* [17], where the authors show that the classical BMS can actually accommodate the frequency–severity dependence if we were to properly adjust the optimal BM relativity. However, the transition rule in this BMS model is still governed by the frequency while ignoring the severity history; in this sense, the BMS proposed by Oh *et al.* [17] is only a partial solution to the task of creating a more complete BMS that depends on both frequency and severity. We note that Frangos and Vrontos [6], Tzougas *et al.* [23], Gómez-Déniz [10], Lu [14], and Cheung *et al.* [3] also consider the a posteriori ratemaking system with the past claim history of both frequency and severity, but the proposed premiums in these studies are written as mathematical formulas. Hence, the premiums are not easily convertible to a proper BMS with levels, transition rule, and relativity.

To this extent, we study a BMS that depends on the history of both frequency and severity, with the bivariate random effect model as the underlying statistical model. Specifically, we generalize the approach of Oh *et al.* [17] to propose a new BMS transition rule where the past claim's size as well as the number of claims affect the next year's BM level. Next, using the analogy between the Bayesian premium and the BMS premium, we analytically derive a set of optimal relativities that minimize the expected squared error in the presence of a frequency–severity dependence under this transition rule. Furthermore, we show that, with a proper choice of the severity size threshold, the proposed BMS structure is guaranteed to outperform other frequency-driven BMSs for a statistical criterion. Given that the current BMS practice neglects the severity history and is unable to penalize large-sized claims, the proposed method can provide a new, theoretically solid means of developing a BMS that is more faithful to the true premium or loss prediction.

The remaining paper is organized as follows. In Section 2, we review the existing BMS along with the bivariate random effect model as the underlying statistical model. Section 3 proposes a new BMS structure whose transition rule incorporates the history of both frequency and severity; we, thus, derive the optimal relativities analytically. In Section 4, we show that the proposed BMS structure is guaranteed to outperform other frequency-driven BMS, with an optimally chosen severity size threshold; we then numerically illustrate this using a model inspired from actual dataset. A real insurance dataset is analyzed in Section 5 to confirm the usefulness and superiority of the proposed BMS. Finally, Section 6 concludes the paper.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

2. Existing BMS models: a review

In a typical BM, the insurer classifies a given policyholder at the outset of the contract based on a priori risk characteristics that are observable. These characteristics would include, for example, gender, age, and vehicle type. As time passes, the policyholder's claim experiences are recorded, and the insurer will periodically reassess the risk of the policyholder and assign her to different risk classes. This so-called a posteriori risk classification based on the claim history involves the Bayesian premium; however, in practice, it is simplified and expressed as the BM relativity along with the transition rule.

Statistical modeling of a BMS is based on the collective risk model, where the parameters of the frequency and severity distributions vary over policyholders. These parameters are first estimated from the a priori risk characteristics at the moment of contract, and subsequently readjusted using the claim history over time, known as the a posteriori risk classification process.

In this section, we review two existing BMSs. The first one is a classical system that aims for the predicted frequency, ignoring the severity component. The second BMS is a more recent one and targets the predicted aggregate severity based on the joint modeling of frequency and severity. However, the a posteriori risk classifications in these BMSs are determined solely based on the frequency history under both BMSs.

To fix the notation, we denote $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$, and $\mathbb{R}^+$ for the set of natural numbers, non-negative integers, real numbers, and positive real numbers, respectively. For a given policyholder, define N_t to indicate the frequency in the tth policy year, that is, the time interval between time $t-1$ and t. Use

$$\boldsymbol{Y}_t = \begin{cases} (Y_{t,1}, \ldots, Y_{t,N_t}), & N_t > 0, \\ \text{Not defined}, & N_t = 0, \end{cases}$$

to denote the associated individual severities. We further define the aggregate and average severity as

$$S_t = \begin{cases} \sum_{j=1}^{N_t} Y_{t,j}, & N_t > 0, \\ 0, & N_t = 0, \end{cases} \quad \text{and} \quad M_t = \begin{cases} \frac{\sum_{j=1}^{N_t} Y_{t,j}}{N_t}, & N_t > 0, \\ 0, & N_t = 0, \end{cases}$$

respectively, so that we have the following relationship between S_t and M_t:

$$S_t = N_t M_t.$$

The realizations of N_t, $\boldsymbol{Y}_t$, S_t, and M_t are denoted by n_t, $\boldsymbol{y}_t$, s_t, and m_t, respectively. Furthermore, we define the claim history of the given policyholder at the end of year T as

$$\mathcal{F}_T^{[\text{full}]} = \{(n_t, \boldsymbol{y}_t) \mid t = 1, \ldots, T\} \quad \text{and} \quad \mathcal{F}_T^{[\text{freq}]} = \{n_t \mid t = 1, \ldots, T\},$$

where the latter contains the frequency information only.

As is common in the BMS literature, we differentiate the a priori and a posteriori risk classification. The former is based on the risk characteristics of policyholders (or explanatory variables), which are observed at the moment of contract, and the latter is based on a residual effect, which cannot be explained by the a priori risk characteristics. For our purposes, we like to differentiate these two for the frequency and severity components, separately. Thus, we write

$$\boldsymbol{X}^{[1]} = (X_1^{[1]}, \ldots, X_{d_1}^{[1]}) \quad \text{and} \quad \Theta^{[1]}$$

to denote the a priori risk characteristics and the residual effect for the frequency and

$$\boldsymbol{X}^{[2]} = (X_1^{[2]}, \ldots, X_{d_2}^{[2]}) \quad \text{and} \quad \Theta^{[2]}$$

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

to denote the same quantities for the severity for the given policyholder. Superscripts [1] and [2] will be extensively used throughout the paper. As done in the literature, the a priori risk characteristics—$\boldsymbol{X}^{[1]}$ and $\boldsymbol{X}^{[2]}$—are constant in time and may overlap; they share the same explanatory variables. Later, we will sometimes use an additional subscript i for all the quantities introduced above to represent the ith policyholder, as needed. Regarding the a priori risk classification, the insurer is assumed to predetermine $\mathcal{K}$ risk classes based on the policyholders' risk characteristics. We will let $(\boldsymbol{x}_\kappa^{[1]}, \boldsymbol{x}_\kappa^{[2]})$ define the κth risk class and w_κ be the weight of the risk class:

$$w_\kappa = \mathbb{P}(\boldsymbol{X}^{[1]} = \boldsymbol{x}_\kappa^{[1]}, \boldsymbol{X}^{[2]} = \boldsymbol{x}_\kappa^{[2]}), \quad \kappa = 1, \ldots, \mathcal{K}.$$

and sometimes, it is convenient to define the frequency-only case

$$w_\kappa^{[1]} = \mathbb{P}(\boldsymbol{X}^{[1]} = \boldsymbol{x}_\kappa^{[1]}), \quad \kappa = 1, \ldots, \mathcal{K}.$$

For the statistical modeling of severity, we consider the *(reproductive) exponential dispersion family* (EDF) in McCullagh and Nelder [15]. The EDF is denoted by $\mathrm{ED}(\mu, \psi)$, where μ and ψ represent the mean and dispersion parameter, respectively. It has the probability density/mass function of the form

$$p(y \mid \theta, \psi) = \exp[(y\theta - b(\theta))/\psi]c(y, \psi),$$

where $b(\cdot)$ and $c(\cdot)$ are predetermined functions, and θ is the canonical parameter. With such parametrization, the mean and variance are given as

$$\mu = \mathbb{E}[Y] = b'(\theta),$$

and

$$\mathrm{Var}(Y) = b''(\theta)\psi \equiv V(\mu)\psi,$$

for some variance function $V(\cdot)$. The inverse of $b'(\cdot)$ is known as the link function, often denoted by $\eta(\cdot)$. One important property of the EDF used in this paper is the reproductive property, that is, for $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{ED}(\mu, \psi)$, the sample average is distributed as

$$\frac{1}{n}\sum_{j=1}^{n} Y_j \sim \mathrm{ED}\left(\mu, \frac{\psi}{n}\right).$$

2.1. Frequency-based BMS

In the classical BMS, both the a priori and a posteriori classification are fully determined by the frequency while ignoring the severity. We present a random effect model for this BMS, calling it "the frequency model" under the general frame work of *generalized linear models* (GLMs) [16]. For a given policyholder, the frequency model consists of several components, described as follows.

The frequency, N_t, of a given policyholder in the tth policy year is specified using a count regression model conditioning on $\boldsymbol{X}^{[1]} = \boldsymbol{x}_\kappa^{[1]}$ and $\Theta^{[1]} = \theta^{[1]}$, that is,

$$N_t \mid (\boldsymbol{x}_\kappa^{[1]}, \theta^{[1]}) \overset{\text{i.i.d.}}{\sim} F_1(\cdot; \lambda_\kappa^{[1]}\theta^{[1]}, \psi^{[1]}),$$

for some discrete distribution F_1 with the mean parameter $\lambda_\kappa^{[1]}\theta^{[1]}$ and the other parameter $\psi^{[1]}$. Following the convention, the a priori rate is modeled as $\lambda_\kappa^{[1]} = (\eta^{[1]})^{-1}(\boldsymbol{x}_\kappa^{[1]}\boldsymbol{\beta}^{[1]})$, where $\boldsymbol{\beta}^{[1]}$ is the regression coefficient vector to be estimated from the data. Hence, having information on $\boldsymbol{x}_\kappa^{[1]}$ is equivalent to knowing $\lambda_\kappa^{[1]}$, provided that the model has been fully specified. For the brevity of exposition, we will use $\lambda_\kappa^{[1]}$ instead of $\boldsymbol{x}_\kappa^{[1]}$ when the a priori characteristics $\boldsymbol{x}_\kappa^{[1]}$ appear as a conditional

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

quantity; for example, $\mathbb{E}[N_{T+1} \mid \lambda_\kappa^{[1]}, \theta^{[1]}]$ will be used instead of $\mathbb{E}[N_{T+1} \mid \boldsymbol{x}_\kappa^{[1]}, \theta^{[1]}]$. Note that $\lambda_\kappa^{[1]}$ and $\theta^{[1]}$ vary over policyholders, whereas $\psi^{[1]}$ remains constant, common for all policyholders. As such, we will suppress this common parameter for quantities that do not vary over individual policyholders. In a special case where $F_1 \in \text{EDF}$, $\psi^{[1]}$ becomes the common dispersion parameter.

By definition, the observations N_t are conditionally independent given the a priori risk characteristics $\boldsymbol{x}_\kappa^{[1]}$ and residual (or random) effect $\theta^{[1]}$, but are dependent unconditionally. We assume that $\Theta^{[1]}$ is distributed as

$$\Theta^{[1]} \overset{\text{i.i.d.}}{\sim} G_1,$$

where G_1 is a distribution function. Individual policyholders have different realizations of $\Theta^{[1]}$, which capture the residual heterogeneity; it is assumed that $\mathbb{E}[\Theta^{[1]}] = 1$ for convenience, following the convention.

The ideal premium for the next year in this frequency model is the hypothetical mean, also known as individual premium in the credibility literature, written as

$$\mathbb{E}[N_{T+1} \mid \lambda_\kappa^{[1]}, \theta^{[1]}].$$

This, however, can be numerically obtained only when $(\Lambda^{[1]}, \Theta^{[1]})$ are known. While $\Lambda^{[1]}$ is known at the outset of the contract and remains fixed over time, the value of $\Theta^{[1]}$ is never known, as it is a random variable. The solution is then to revise the distribution parameters of $\Theta^{[1]}$ based on the past claim frequency history $\mathcal{F}_T^{[\text{freq}]}$ and compute the predictive mean $\mathbb{E}[N_{T+1} \mid \lambda_\kappa^{[1]}, \mathcal{F}_T^{[\text{freq}]}]$. This is called the Bayesian premium.

Although the Bayesian premium is deemed to be the best premium possible, most ratemaking systems in the auto insurance industry maintain a BMS, which can be considered a discrete approximation of the Bayesian premium, as explained below. The classical BMS constitutes three elements: the BMS levels, transition rule, and relativities. Without loss of generality, let us assume that the insurer has $z+1$ BM levels varying from 0 to z, with $-1/+h$ system as a transition rule. Thus, in this system, a claim-free year is rewarded by moving one BM level downward, while each reported claim, regardless of the claim size, is penalized by climbing h levels per claim at the end of the period. Theoretically, the choice of the number of BM levels and the transition rule are arbitrary, though insurers may have constraints subject to local regulations in practice. The transition rule affects how the drivers are populated over $z+1$ different BM levels, which, in turn, determines the optimal premium of a driver in each level for the following year.

The last element of the classical BMS is the relativity. In the classical frequency-driven BMS, the relativity associated with BM level ℓ, denoted by $r^{[\text{freq}]}(\ell)$, is used to predict the frequency in the next year through the multiplicative form

$$\lambda_\kappa^{[1]} r^{[\text{freq}]}(\ell), \quad \ell = 0, \ldots, z, \tag{1}$$

where the a priori rate $\lambda_\kappa^{[1]} = (\eta^{[1]})^{-1}(\boldsymbol{x}_\kappa^{[1]} \boldsymbol{\beta}^{[1]})$ reflects the risk class to which the policyholder belongs. Expression (1) is obtained from the following minimization by Tan *et al.* [22]:

$$(\tilde{r}^{[\text{freq}]}(0), \ldots, \tilde{r}^{[\text{freq}]}(z)) = \arg\min_{(r(0),\ldots,r(z)) \in \mathbb{R}^{z+1}} \mathbb{E}\left[\left(\mathbb{E}[N_{T+1} \mid \Lambda^{[1]}, \Theta^{[1]}] - \Lambda^{[1]} r(L) \right)^2 \right],$$

where L denotes the BM level occupied by a randomly selected policyholder in the steady state[1] under $-1/+h$ system.

[1] As the BM level in the next period is completely determined by the current BM level and the number of reported claims in the current period, the BMS can be represented by a Markov chain which naturally leads to the derivation of the steady state.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

2.2. BMS based on both frequency and severity

Though widely used in practice, the frequency-based BMS completely ignores the claim severity information. A more statistically solid approach would be based on the aggregate severity, or the total loss of the collective risk model, which reflects both claim frequency and severity. In traditional collective risk models, the frequency and severity are assumed independent. Although a series of recent studies shows that this unwarranted assumption can be relaxed in the collective risk model, and, thus, further in the Bayesian premium and BMS. In fact, the dependence between the frequency and severity in auto insurance is shown to be statistically significant in recent empirical studies such as Frees *et al.* [8] and Garrido *et al.* [9].

While various dependent collective risk models exist as mentioned in the "Introduction" section, including the two-step frequency–severity model and the copula-based model, we find that the bivariate random effect model is most convenient for the development of a better BMS because of its conditioning structure in the model specification. In this section, we focus on the model by Oh *et al.* [17], as it offers the most comprehensive bivariate random effect model: It contains other existing bivariate random effect models in the literature—Pinquet [20], Boudreault *et al.* [2], Hernández-Bastida *et al.* [11], Baumgartner *et al.* [1], Lu [14], and Cheung *et al.* [3]—as special cases.

Paralleling the frequency model in Section 2.1, the dependent collective risk model of Oh *et al.* [17] is specified as follows. We call this "the frequency–severity model," as it requires both frequency and severity models. For the given a priori rate $(\Lambda^{[1]}, \Lambda^{[2]}) = (\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]})$ and residual effect $(\Theta^{[1]}, \Theta^{[2]}) = (\theta^{[1]}, \theta^{[2]})$, the observations

$$N_t \quad \text{and} \quad Y_{t,j} \quad \text{for } t, j \in \mathbb{N}$$

are independent. In the frequency–severity model, the distributional assumption for N_t is identical to the frequency model in Section 2.1. The individual severity $Y_{t,j}$ is specified using a regression model conditioning on the a priori risk characteristics $\boldsymbol{X}^{[2]} = \boldsymbol{x}_\kappa^{[2]}$, or equivalently $\Lambda^{[2]} = \lambda_\kappa^{[2]}$ as previously mentioned, and the residual effect $\Theta^{[2]}$, such that

$$Y_{t,j} \mid (\lambda_\kappa^{[2]}, \theta^{[2]}) \overset{\text{i.i.d.}}{\sim} \text{EDF}(\lambda_\kappa^{[2]}\theta^{[2]}, \psi^{[2]}),$$

where the EDF has mean parameter $\lambda_\kappa^{[2]}\theta^{[2]}$ with $\lambda_\kappa^{[2]} = (\eta^{[2]})^{-1}(\boldsymbol{x}_\kappa^{[2]}\boldsymbol{\beta}^{[2]})$ and dispersion parameter $\psi^{[2]}$. Finally, the residual effect characteristic $(\Theta^{[1]}, \Theta^{[2]})$, which is assumed to be independent of the a priori rate $(\Lambda^{[1]}, \Lambda^{[2]})$, has a joint distribution H. That is,

$$(\Theta^{[1]}, \Theta^{[2]}) \overset{\text{i.i.d.}}{\sim} H = C(G_1, G_2), \tag{2}$$

where G_1 and G_2 denote the marginal distribution functions for $\Theta^{[1]}$ and $\Theta^{[2]}$, respectively, and C is a bivariate copula that defines the dependence structure in H. We use g_1, g_2, and h to denote the density functions of G_1, G_2, and H, respectively. As before, it is further assumed

$$\mathbb{E}[\Theta^{[1]}] = \mathbb{E}[\Theta^{[2]}] = 1$$

for convenience.

There are several conditional independence relations under this frequency–severity model specification. For example, given $(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]})$, and $(\theta^{[1]}, \theta^{[2]})$, the claim history $(N_t, \boldsymbol{Y}_t)$ of a given policyholder is independent across time, and independence between frequency and individual severities and independence among individual severities are further implied. However, marginally or unconditionally, various dependence relations are observed:

- Dependence among frequencies induced by residual effect $\Theta^{[1]}$;
- Dependence among severities induced by residual effect $\Theta^{[2]}$;
- Dependence between frequency and severity induced by residual effects $(\Theta^{[1]}, \Theta^{[2]})$ via the copula C.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Recalling that the individual severity $Y_{t,j}$ is assumed EDF distributed, we note that the conditional average severity is again EDF distributed. That is,

$$M_t \,|\, (\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, N_t) \overset{\text{i.i.d.}}{\sim} \text{ED}(\lambda_\kappa^{[2]}\theta^{[2]}, \psi^{[2]}/N_t), \quad N_t > 0,$$

and

$$\mathbb{P}(M_t = 0 \,|\, \lambda_\kappa^{[2]}, \theta^{[2]}, N_t) = 1, \quad \text{if } N_t = 0,$$

where the mean parameter of the EDF, $\lambda_\kappa^{[2]}\theta^{[2]}$, is modeled through $\lambda_\kappa^{[2]} = (\eta^{[2]})^{-1}(\boldsymbol{x}_\kappa^{[2]}\boldsymbol{\beta}^{[2]})$. Thus, the distribution of M_t is the same as that of $Y_{t,j}$, except for the dispersion parameter. This allows us to efficiently model the aggregate severity S_t as a product form $N_t M_t$, whose mean serves as the premium under the frequency–severity model. There are two noteworthy sub-modes of the frequency–severity model. First, it reduces to the frequency model in Section 2.1 by removing the severity part. Second, if the copula C embedded in H is an independent copula, then the frequency and severity can be independently modeled, leading to the standard collective risk model.

We found that the optimal frequency-premium takes a form of $\lambda_\kappa^{[1]}\tilde{r}^{[\text{freq}]}(\ell)$ in the frequency model from (1). Now, suppose we predict the aggregate severity as the premium using the history of the frequency only. A simple, but widely, used approach in the literature [5] is to multiply the average severity of all policyholders in the same risk class to the frequency such that the BMS premium is predicted with

$$\lambda_\kappa^{[1]}\tilde{r}^{[\text{freq}]}(\ell) \times \lambda_\kappa^{[2]}. \tag{3}$$

This multiplicative form is less than ideal in at least two aspects. First, by incorporating the average severity of the entire portfolio, it reflects the history of the frequency of a given policyholder, but completely ignores the past severity records of individual policyholders. Second, the multiplicative form can be justified only when the independence of the frequency and severity is warranted, which is not valid in general. In fact, the BMS premium (3) is motivated from the following Bayesian premium with the implied independence between frequency and severity:

$$\mathbb{E}[S_{T+1} \,|\, \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}] = \lambda_\kappa^{[1]}\mathbb{E}[\Theta^{[1]} \,|\, \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}] \times \lambda_\kappa^{[2]}. \tag{4}$$

From this, the relativity $\tilde{r}^{[\text{freq}]}(\ell)$ in (3) can be interpreted as the BMS version of the posterior estimation

$$\mathbb{E}[\Theta^{[1]} \,|\, \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}]$$

in (4). Therefore, while convenient and simple, the BMS premium (3) may result in a considerable bias when dependence between frequency and severity is strong. To adjust such bias, Oh *et al.* [17] suggested the following BMS premium from a policyholder at BM level ℓ in a $-1/+h$ system under the frequency–severity model:

$$\lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\tilde{r}^{[\text{dep}]}(\ell), \quad \ell = 0, \ldots, z, \tag{5}$$

where optimal relativity $\tilde{r}^{[\text{dep}]}(\ell)$ for $\ell = 0, \ldots, z$ is the solution of the optimization,

$$\begin{aligned}&(\tilde{r}^{[\text{dep}]}(0), \ldots, \tilde{r}^{[\text{dep}]}(z)) \\ &\quad = \arg\min_{(r(0),\ldots,r(z))\in\mathbb{R}^{z+1}} \mathbb{E}\left[(\mathbb{E}[S_{T+1} \,|\, \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]}r(L))^2\right]\end{aligned} \tag{6}$$

which is an extended setting of that in Tan *et al.* [22]. Here, L denotes the BM level occupied by a randomly selected policyholder in the steady state under the $-1/+h$ system. We refer to Oh *et al.* [17] for the detailed derivation of (6).

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Similar to (3), the premium (5) can be viewed as the BMS version of the Bayesian premium from

$$\mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}] = \lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}] \tag{7}$$

and, consequently, the relativity $\tilde{r}^{[\text{dep}]}(\ell)$ in (5) can be interpreted as the BMS version of the posterior estimation

$$\mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{freq}]}]. \tag{8}$$

Notably, the BMS premium of Oh *et al.* [17] operates under the $-1/+h$ system, where the BM level transition is driven by the claim frequency only and is unable to penalize high severity claims. Also, as seen from (8), the severity history is not used inside the condition. In reality, the severity information is indirectly incorporated only through the dependence between $\Theta^{[1]}$ and $\Theta^{[2]}$. Thus, the BMS premium of Oh *et al.* [17] approximates the Bayesian premium, but in a limited sense, because it incorporates the history of claim frequency and not severity. Later, we propose to improve the BMS premium prediction with a direct incorporation of the history of severity as well.

2.3. Motivation for transition rules driven by the severity history

As discussed in the previous subsection, the BMS premium of Oh *et al.* [17] in (5) can be viewed as the BMS version of the Bayesian premium in (7). However, the premium in (7) is not optimal in that it uses only part of the claim history. The most complete Bayesian premium can be obtained by incorporating the full history of the given policyholder, including both frequency and severity:

$$\mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}] = \lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}]. \tag{9}$$

The inclusion of the past claim history of severity in addition to the past claim history of frequency is important in terms of the efficiency of the prediction as the following theorem shows.

Theorem 2.1. *For the frequency–severity model in Section 2.2, we obtain the following inequality of mean square errors (MSEs):*

$$\mathbb{E}[(S_{T+1} - \mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \mathcal{F}_T^{[full]}])^2] \le \mathbb{E}[(S_{T+1} - \mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \mathcal{F}_T^{[freq]}])^2].$$

Proof. See Appendix A. □

In plain terms, Theorem 1 states that, for the frequency–severity model, adding the past claim history of severity on top of the frequency history always increases the prediction power by providing a smaller MSE. This is true even when the frequency and severity are assumed to be independent. Relatedly, we find conditions where adding the severity history does not improve the prediction performance and whose solution is presented below. The proof is straightforward and is omitted.

Corollary 2.2. *If* $\mathbb{P}(\Theta^{[2]} = 1) = 1$*, the equality in Theorem 1 holds. That is,*

$$\mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[full]}] = \mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[freq]}].$$

The condition $\mathbb{P}(\Theta^{[2]} = 1) = 1$ is equivalent to assuming only dependence among frequencies is present, removing all other dependencies. Hence, in light of the analogy between the Bayesian premium and BMS premium, the main implication of Theorem 1 is that the inclusion the past claim history of severity in the BMS transition rule can result in a more efficient estimation of the premium.

In the next section, we show how to modify the classical BMS by incorporating the transition rule for severity and, thus, derive the related optimal relativity.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

3. Transition rule driven by the frequency–severity history

The optimal relativity proposed by Oh *et al.* [17], as in (6), is "optimal" in the sense that the BMS transition rule is driven solely by the frequency. It also accommodates the severity component, but only through the dependence structure between frequency and severity and not from the past severity records. To achieve a full optimality, using the historical severity information as well as the frequency in the transition rule is essential, as discussed in Section 2.3.

Considering the severity history in a posteriori risk classification can be traced to Picard [19]. Since then, various a posteriori risk classification methods incorporating both frequency and severity history have been proposed under various collective risk models [6,10,11,14,20]. For example, Picard [19] and Gómez-Déniz [10] consider a posteriori risk classification for frequency by distinguishing small and large claims with some threshold. On the other hand, Frangos and Vrontos [6], Hernández-Bastida *et al.* [11], and Lu [14] derived the Bayesian premium for aggregate severity. However, these studies mainly concern distributional quantities and, thus, are not directly translated to the BMS with a proper level structure and the associated transition rules.

Though the inclusion of the severity history is important in a posteriori risk classification, this element has been largely overlooked in the literature for several reasons. First, historically, the main focus of the BMS in the literature has been the prediction of frequency following the industry tradition. In fact, a typical commercial BMS involves only the claim frequency in most jurisdictions. Second, the frequency-only BMS is simpler to model and analyze. Third, it is implicitly assumed that the severity is independent of the frequency. Hence, the separate analysis of severity has been considered unimportant or redundant. An exception is the recent work of Oh *et al.* [17], where the authors propose a new BMS to predict the aggregate severity under some dependence between frequency and severity. However, their BMS still uses the frequency history only while ignoring the severity history. To this extent, this section proposes a more general BMS and the associated transition rule that distinguishes the degree of claim severity, on top of the classical frequency-based BMS transition rule. We, thus, derive the optimal relativities.

3.1. New transition rule: the $-1/+h_1/+h_2$ system

To distinguishes the degree of claim severity, we classify each claim into two types based on the claim size following Picard [19] and Gómez-Déniz [10]. The severity with $y \le \varphi$ is called a Type I claim and the severity with $y > \varphi$ is called a Type II claim for some predetermined threshold φ. For a given policyholder, define $L_{t-1} \in \{0, \dots, z\}$ to be the (random) BM level occupied during time $(t-1, t)$ and assume that the new policyholder starts at level l_0 at time 0. A claim-free year in this BMS is rewarded by coming down one BM level; each Type I claim will be penalized by climbing up by h_1 levels per claim and each type II claim by h_2 levels per claim. We denote this BMS system as the $-1/+h_1/+h_2$ system with threshold φ. Naturally, we assume $h_2 \ge h_1$; if $h = h_1 = h_2$, this reduces to the classical $-1/+h$ BMS system [5]. We denote the number of Type I claims during time $(t-1, t)$ as $N_{t-1}^{\{\mathrm{I}\}}$ and the number of Type II claims as $N_{t-1}^{\{\mathrm{II}\}}$.

Now, consider a driver in level $L_{t-1} \in \{0, \dots, z\}$ during time $(t-1, t)$. The new level in the next time period L_t can be determined based on the claim history $(N_{t-1}^{\{\mathrm{I}\}}, N_{t-1}^{\{\mathrm{II}\}})$ via

$$L_t = \begin{cases} \max\{L_{t-1} - 1, 0\}, & N_{t-1}^{\{\mathrm{I}\}} = N_{t-1}^{\{\mathrm{II}\}} = 0; \\ \min\{L_{t-1} + h_1 N_{t-1}^{\{\mathrm{I}\}} + h_2 N_{t-1}^{\{\mathrm{II}\}}, z\}, & \text{otherwise.} \end{cases} \tag{10}$$

Because of the recursive nature, L_t is a function of

$$N_0^{\{\mathrm{I}\}}, \dots, N_{t-1}^{\{\mathrm{I}\}} \quad \text{and} \quad N_0^{\{\mathrm{II}\}}, \dots, N_{t-1}^{\{\mathrm{II}\}}$$

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

for a given $L_0 = \ell_0$. Also, as L_t does not depend on $L_0, \ldots, L_{t-2}$, provided that L_{t-1} is given, the process L_t for $t = 0, 1, \ldots,$ is Markovian. The transition probabilities of this Markov chain from $L_{t-1} = \ell$ to $L_t = \ell^*$, given the observed and residual effect characteristics $\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}$, is defined as

$$p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) = \mathbb{P}(L_t = \ell^* \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, L_{t-1} = \ell),$$

and can be calculated as follows.

Theorem 3.1. *Consider the frequency–severity model in Section 2.2 and a policyholder with*

$$(\Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}) = (\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]})$$

under the $-1/+h_1/+h_2$ BMS system with threshold φ. Then, for $\ell, \ell^* \in \{0, 1, \ldots, z\}$ and $\ell^* \neq z$, the transition probabilities of moving from level ℓ to ℓ^* are

$$p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) = \begin{cases} \displaystyle\sum_{(k_1,k_2)\in\mathcal{J}(\ell,\ell^*)} \left[q_1(k_1+k_2) \binom{k_1+k_2}{k_1} (q_2(\varphi))^{k_2} (1-q_2(\varphi))^{k_1} \right], & \ell < \ell^* \text{ or } \ell^* = \ell = z; \\ q_1(0), & \max\{\ell - 1, 0\} = \ell^*; \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

where k_1 and k_2 are the number of Type I and II claims, respectively, and

$$\mathcal{J}(\ell, \ell^*) = \{(k_1, k_2) \in \mathbb{N}_0 \times \mathbb{N}_0 \mid \ell^* = \max\{\ell + k_1 h_1 + k_2 h_2, z\}\} \tag{12}$$

with $q_1(k)$ and $q_2(\varphi)$ being probabilities defined as

$$q_1(k) = \mathbb{P}(N_t = k \mid \lambda_\kappa^{[1]}, \theta^{[1]}) \tag{13}$$

and

$$q_2(\varphi) = \mathbb{P}(Y_{t,j} > \varphi \mid \lambda_\kappa^{[2]}, \theta^{[2]}). \tag{14}$$

Proof. See Appendix A. □

While the above result is precise, the implementation of the transition probability is cumbersome, mainly because of the complicated definition of the index function in (12). An alternative expression of this quantity, which can be easily implemented using a computer, is provided in the following lemma.

Lemma 3.2. *The transition probability in (11) in Theorem 2 can be written as*

$$p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) = \begin{cases} \displaystyle\sum_{k_2\in\mathcal{D}} \left[q_1(k_2+\zeta(k_2)) \binom{\zeta(k_2)+k_2}{\zeta(k_2)} (q_2(\varphi))^{k_2} (1-q_2(\varphi))^{\zeta(k_2)} \right], & \ell < \ell^*; \\ q_1(0), & \max\{\ell - 1, 0\} = \ell^*; \\ 0, & \text{otherwise}; \end{cases} \tag{15}$$

for $\ell^* \neq z$, where q_1 and q_2 are defined in (13) and (14), and

$$\zeta(k_2) = \frac{(\ell_2 - \ell_1) - k_2 h_2}{h_1} \quad \text{and} \quad \mathcal{D} = \left\{ k_2 \in \mathbb{N}_0 \,\middle|\, \frac{(\ell_2 - \ell_1) - k_2 h_2}{h_1} \in \mathbb{N}_0 \right\}.$$

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

For $\ell^* = z$, we have

$$p_{\ell,z}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) = 1 - \sum_{\ell^*=0}^{z-1} p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}).$$

Proof. It is satisfactory to show (15). For $\ell^* \neq z$, we obtain

$$\ell^* = \max\{\ell + k_1 h_1 + k_2 h_2, z\}$$

if and only if $\ell^* = \ell + k_1 h_1 + k_2 h_2$. Hence, we obtain $(k_1, k_2) \in \mathcal{J}(\ell, \ell^*)$ if and only if $k_2 \in \mathcal{D}$, which concludes (15) from (11) for $\ell^* \neq z$. □

3.2. Long-term behavior

From the transition probability $p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]})$ in Theorem 2, we can construct the one-step transition matrix for the policyholder

$$\boldsymbol{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]})$$

of size $(z+1) \times (z+1)$. The associated conditional stationary distribution for the policyholder to be in level ℓ is then given by

$$\boldsymbol{\pi}(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) = (\pi_0(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}), \ldots, \pi_z(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}))^{\mathrm{T}} \tag{16}$$

whose $(l+1)$th element is the stationary probability for the policyholder to remain in BM level ℓ. An explicit expression for the stationary probability from the transition probability is presented in Lemma 2 in Appendix B.

Finally, with this proposed transition rule, the BM level L occupied by the randomly selected policyholder in the steady state, that is, the unconditional stationary distribution, is given as

$$\mathbb{P}(L = \ell) = \sum_{\kappa \in \mathcal{K}} w_\kappa \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) h(\theta^{[1]}, \theta^{[2]}) \mathrm{d}\theta^{[1]} \mathrm{d}\theta^{[2]}, \quad \text{for } \ell = 0, \ldots, z. \tag{17}$$

3.3. Optimal relativities under the $-1/+h_1/+h_2$ system

In this subsection, we develop the optimal relativities for the frequency–severity model in Section 2.2 by incorporating the history of severity as well as frequency under the new $-1/+h_1/+h_2$ rule. For a policyholder in the κth risk class at the outset of the contract, the BMS premium takes the following form

$$\lambda_\kappa^{[1]} \lambda_\kappa^{[2]} r^{[\mathrm{new}]}(\ell), \quad \ell = 0, \ldots, z,$$

based on the analogy from the Bayesian premium in (9). The set of optimal relativities $\tilde{r}^{[\mathrm{new}]}(\ell)$ in this setting is the solution of the following optimization problem for the aggregate severity:

$$\begin{aligned}
&(\tilde{r}^{[\mathrm{new}]}(0), \ldots, \tilde{r}^{[\mathrm{new}]}(z)) \\
&\quad = \arg\min_{(r(0),\ldots,r(z)) \in \mathbb{R}^{z+1}} \mathbb{E}\left[(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]} r(L))^2 \right] \\
&\quad = \arg\min_{(r(0),\ldots,r(z)) \in \mathbb{R}^{z+1}} \sum_{\ell=0}^{z} \mathbb{E}\left[(\Lambda^{[1]}\Lambda^{[2]}\Theta^{[1]}\Theta^{[2]} - \Lambda^{[1]}\Lambda^{[2]} r(L))^2 | L = l \right] \mathbb{P}(L = \ell).
\end{aligned} \tag{18}$$

We note that the optimization problem in (18) is identical to that in (6), except that the stationary distribution of L in (18) is now determined by both frequency and severity history as shown in the

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

previous subsection. This is in contrast to (6), where the distribution of L is determined by the parameters of frequency history only. Hence, the analytical solution of (18) can be obtained in a similar manner as (6).

Theorem 3.3. *Consider the frequency–severity model in Section 2.2 under the $-1/+h_1/+h_2$ BMS system. The optimal relativity $\tilde{r}^{[new]}(\ell)$ that solves (18) is given by*

$$\tilde{r}^{[new]}(\ell) = \frac{\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2\Theta^{[1]}\Theta^{[2]} \mid L = l]}{\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2 \mid L = l]} \quad \text{for } \ell = 0, \ldots, z, \tag{19}$$

where

$$\begin{aligned}&\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2\Theta^{[1]}\Theta^{[2]} \mid L = l] \\ &= \frac{\sum_{\kappa\in\mathcal{K}} w_\kappa (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \iint \theta^{[1]}\theta^{[2]}\pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]})h(\theta^{[1]}, \theta^{[2]})d\theta^{[1]}d\theta^{[2]}}{\sum_{\kappa\in\mathcal{K}} w_\kappa \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]})h(\theta^{[1]}, \theta^{[2]})d\theta^{[1]}d\theta^{[2]}}\end{aligned} \tag{20}$$

and

$$\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2 \mid L = l] = \frac{\sum_{\kappa\in\mathcal{K}} w_\kappa (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]})h(\theta^{[1]}, \theta^{[2]})\, d\theta^{[1]}\, d\theta^{[2]}}{\sum_{\kappa\in\mathcal{K}} w_\kappa \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]})h(\theta^{[1]}, \theta^{[2]})\, d\theta^{[1]}\, d\theta^{[2]}}. \tag{21}$$

Proof. See Appendix B. □

While the optimal relativity (19) has a similar expression as in (6), it differs in that the distribution of L in (19) is determined by both the frequency and severity components, while the distribution of L in (6) is determined by the frequency component only. As a result, $\tilde{r}^{[\text{new}]}(\ell)$ in (19) is different from $\tilde{r}^{[\text{dep}]}(\ell)$ in (6). We remark that the Bayesian version of the BM premium in this section can be represented as follows:

$$\mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}] = \lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}]. \tag{22}$$

Hence, BM relativity $\tilde{r}^{[\text{new}]}(\ell)$ can be interpreted as the BMS version of the posterior estimation

$$\mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}]$$

in (22).

4. Comparison of the BMS

For the comparison of the quality of the various BMS systems, Oh *et al.* [17] proposed the hypothetical mean square error (HMSE), a measure defined as the MSE between the aggregate severity and the BMS premium in the stationary state. For example, for the frequency–severity model in Section 2.2, we obtain

$$\begin{aligned}\text{HMSE}(\boldsymbol{r}, -1/+h) &= \mathbb{E}\left[(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]}r(L))^2\right] \\ &= \sum_{\kappa\in\mathcal{K}} w_\kappa \int\int \sum_{l=1}^{z} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\theta^{[1]}\theta^{[2]} - \lambda_\kappa^{[1]}\lambda_\kappa^{[2]}r_l)^2\pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]})h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}\end{aligned}$$

for a given relativity set $\boldsymbol{r} = (r(0), \ldots, r(z))$. Here L is governed by the transition rule $-1/+h$, and its distribution is the stationary distribution of the BM level of a randomly chosen policyholder. If we use

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

the proposed $-1/+h_1/+h_2$ system (with threshold φ), the HMSE is defined as

$$\begin{aligned}&\mathrm{HMSE}(\boldsymbol{r}, -1/+h_1/+h_2)\\ &\quad= \mathbb{E}\left[(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]} r_L)^2\right]\\ &\quad= \sum_{\kappa\in\mathcal{K}} w_\kappa \int\int \sum_{l=1}^{z} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]}\theta^{[1]}\theta^{[2]} - \lambda_\kappa^{[1]}\lambda_\kappa^{[2]} r_l)^2 \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]} t) h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}.\end{aligned}$$

Again, L is governed by the transition rule $-1/+h_1/+h_2$ system with threshold φ, and its distribution is the stationary distribution of the BM level of a randomly chosen policyholder in this system. While the following result can be immediately obtained from the definitions of two BMS premiums, it is one of the main results in this paper which justifies the use of the size of severity in the transition rule.

Theorem 4.1. *The premium in BMS with the transition rule by the past claim history of both frequency and severity can provide a more efficient premium in the BMS with the transition rule by the past claim history of frequency only in the following sense:*

$$\min_{\boldsymbol{r}\in\mathbb{R}^{z+1}, h_1, h_2\in\mathbb{N}_0, \varphi\in\mathbb{R}} \mathit{HMSE}(\boldsymbol{r}, -1/+h_1/+h_2) \le \min_{\boldsymbol{r}\in\mathbb{R}^{z+1}, h\in\mathbb{N}_0} \mathit{HMSE}(\boldsymbol{r}, -1/+h).$$

Proof. The proof is trivial from the observation that the transition rule $-1/+h_1/+h_2$ is the same as the transition rule $-1/+h$ if $h_1 = h_2 = h$. Consequently, we obtain

$$\mathrm{HMSE}(\boldsymbol{r}, -1/+h/+h) = \mathrm{HMSE}(\boldsymbol{r}, -1/+h)$$

which concludes the proof. □

The remaining section provides a numerical study to show the performance of the proposed BMS compared with the BMS in the recent work of Oh *et al.* [17] in Section 2.2. Specifically, we consider the frequency–severity model with the following specification: We assume 10 different BM levels (i.e., $z = 9$), but for simplicity, consider a single a priori risk class ($\mathcal{K} = 1$).

We model the frequency part as

$$N_t|(\theta^{[1]}, \lambda_\kappa^{[1]}) \sim \mathrm{Poisson}(\lambda_\kappa^{[1]}\theta^{[1]}),$$

and the individual severity part as

$$Y_{t,j}|(\theta^{[2]}, \lambda_\kappa^{[2]}) \sim \mathrm{Gamma}(\lambda_\kappa^{[2]}\theta^{[2]}, 1/\psi^{[2]}),$$

where $\lambda_\kappa^{[2]}\theta^{[2]}$ is the mean and $1/\psi^{[2]}$ is the shape parameter. The bivariate random effect $(\Theta^{[1]}, \Theta^{[2]})$ follows the joint distribution $H = C(G_1, G_2)$, where C is the Gaussian copula with correlation coefficient ρ, and the marginal distributions, G_1 and G_2, are given by

$$\begin{cases}\Theta^{[1]} \sim \text{Log Normal}(-(\sigma^{[1]})^2/2, (\sigma^{[1]})^2);\\ \Theta^{[2]} \sim \text{Log Normal}(-(\sigma^{[2]})^2/2, (\sigma^{[2]})^2),\end{cases}$$

and whose parameters are chosen to preserve the unity of the mean. In the following subsections, we investigate the impact of different factors using various selected parameter values.

4.1. Impact of different dependence between frequency and severity

We first study the impact of different degrees of dependence between frequency and severity on the HMSE. In particular, we consider three different copula dependence parameters: $\rho = -0.8,\ -0.4$, and

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

0.4. Hence, the larger the value of ρ, the stronger the dependence becomes. We assess the impact of these changes on two BMSs: the $-1/+1$ system of [17] and our proposed $-1/+1/+2$ system. As our BMS depends on the severity threshold value φ, we consider four different thresholds, namely the 75th, 90th, 99th, and 99.9th quantile of severity, for φ. The remaining parameters are fixed at $\lambda^{[1]} = 0.5$, $\lambda^{[2]} = \exp(8.8)$, $(\sigma^{[1]})^2 = 0.99$, $(\sigma^{[2]})^2 = 0.29$, and $\psi^{[2]} = 1/0.67$, which are motivated from the real data analysis in Section 6. Under this specification, the threshold values are given by φ = 8,200, 16,800, 48,100, and 94,300.

The result is summarized in Table B.1, which shows the optimal relativity, stationary distribution, and HMSE for each $\rho = -0.8,\ -0.4$, and 0.4. Several comments are in place from this table:

- For a given ρ value, as φ grows larger, the optimal relativities, stationary distribution, and HMSE become identical for both the $-1/+1$ and $-1/+1/+2$ BMSs. This is from a general property that, by definition, the $-1/+1/+2$ system with infinitely large threshold is the same as the $-1/+1$ system.
- On the other hand, a smaller φ means heavier penalty and more drivers end up in a higher BM level. This is confirmed by the stationary probability at the highest level $\mathbb{P}(L = 9)$ growing larger as φ decreases for all ρ values. For $\rho = -0.8$ and -0.4, the threshold φ at the 90th quantile shows the best performance in terms of the HMSE between four thresholds, whereas, in $\rho = 0.4$ case, the 99th quantile is the best threshold. In general, the HMSE of the $-1/+1/+2$ system varies by the threshold and is not always smaller than that of the $-1/+1$ system. However, the best HMSE of the $-1/+1/+2$ system beats that of the $-1/+1$ system with an appropriate choice of the threshold φ, following the argument of Theorem 4. The optimal threshold can be found with numerical iterations, though its general solution is not straightforward.
- As the dependence between frequency and individual severities becomes stronger, the aggregate severity becomes larger and the premium increase will be more substantial. We observe this from comparing the optimal relativities across different ρ values. In particular, as ρ becomes larger, the optimal relativities become larger for most BM levels with the exception in the lower BM levels, especially at level 0, in both the $-1/+1$ and $-1/+1/+2$ systems.

4.2. Impact of different mean frequency and transition rule

To study the impact of different frequency parameters and the transition rule, we consider several combinations on the mean frequency, whose result is shown in Table B.2, while other parameters are fixed at $\lambda^{[2]} = \exp(8.8)$, $\psi^{[2]} = 1/0.67$, $(\sigma^{[1]})^2 = 0.99$, and $(\sigma^{[2]})^2 = 0.29$. For the dependence, we set $\rho = -0.45$, the choice made from an actual dataset in the next section.

- In comparing (a) and (b) in Table B.2, we note that both keep the same frequency parameter $\lambda^{[1]}$, but Table B.2(b) adopts a more stringent transition rule by having the $-1/+1/+3$ system. Consequently, the stationary probabilities $\mathbb{P}(L = \ell)$ in Table B.2(b) tend to be larger in the upper BM levels and smaller in the lower BM levels compared with Table B.2(a), as they should.
- Furthermore, the best threshold among other alternatives in terms of the HMSE is φ = 16,800 in Table B.2(a), but φ = 48,100 in Table B.2(b). This implies that the optimal threshold depends on the transition rule as well as other parameters, as previously mentioned.
- Table B.2(a) and (c) share the same transition rule but have different mean frequency $\lambda^{[1]}$ values. By having a larger frequency mean, Table B.2(c) leads to a higher proportion of drivers in higher BM levels compared with Table B.2(a) under both the $-1/+1$ and $-1/+1/+2$ systems. However, the relativities are generally smaller in Table B.2(c) under both the BMS rules because of the negative ρ value. That is, in this specific setting, the average claim size becomes smaller as there are more claims, and this, in turn, leads to a smaller aggregate loss on average. Similar to the above point, we also see that the optimal threshold in terms of the HMSE changes as $\lambda^{[1]}$ does; for $\lambda^{[1]} = 0.5$, the optimal threshold is φ = 16,800, corresponding to the 90th quantile, whereas φ = 94,300, corresponding to the 99.9th quantile, when $\lambda^{[1]} = 2$.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

4.3. Impact of different dependence among severities

Here, we study the impact of the dependence among individual severities by changing $\sigma^{[2]}$. We assume the independence between frequency and individual severities by setting $\rho = 0$ for this study. The other parameters are fixed at $\lambda^{[1]} = 0.5$, $\lambda^{[2]} = \exp(8.8)$, $\psi^{[2]} = 1/0.67$, and $(\sigma^{[1]})^2 = 0.99$.

- Focusing on the $-1/+1$ system, the optimal relativities remain the same over different $\sigma^{[2]}$ values. This is because the $-1/+1$ system is driven by the frequency only and its relativities from the independence between frequency and severity; it responds to the mean $E(\Theta^{[2]})$, but this mean is fixed at 1, although $\sigma^{[2]}$ changes.
- The independence between frequency and severity similarly affects the $-1/+1/+2$ system, where the stationary distribution stays the same over different $\sigma^{[2]}$ values, which is justified from (19). However, the optimal relativities do change. In particular, the relativities become larger in higher BM levels as $\sigma^{[2]}$ increases, for a given threshold φ. This occurs because a larger $\sigma^{[2]}$ value yields larger claims.
- As before, the optimal threshold varies as the $\sigma^{[2]}$ value changes. For $(\sigma^{[2]})^2 = 0.29$ and 1, the optimal threshold in terms of the HMSE is found at the 90th quantile, whereas the optimal threshold is suggested to be much larger for $(\sigma^{[2]})^2 = 0.01$.

5. Data analysis

5.1. Data description and estimation results

To examine the effect of the dependence on BM relativities, we now consider an actual insurance dataset concerning a collision coverage of new and old vehicles from the Wisconsin Local Government Property Insurance Fund (LGPIF) [8]. A detailed information on the project can be found at the LGPIF project website. The collision coverage provides cover for the impact of a vehicle with an object, impact of a vehicle with an attached vehicle, and overturn of a vehicle. The observations include policyholders who have either a new collision coverage, an old collision coverage, or both. In our analysis, longitudinal data over the policy years from 2006 to 2010 with 497 governmental entities are used. There are two categorical variables identical to the dataset studied in Oh *et al.* [17]: the entity type with six levels and the coverage with three levels, as shown in Table B.4. Following this, we model the frequency part as

$$N_t \mid (\lambda_\kappa^{[1]}, \theta^{[1]}) \sim \text{Poisson}(\lambda_\kappa^{[1]}\theta^{[1]}) \quad \text{with } \lambda_\kappa^{[1]} = \exp(\boldsymbol{x}_\kappa^{[1]}\boldsymbol{\beta}^{[1]}),$$

and the individual severity part as

$$Y_{t,j} \mid (\lambda_\kappa^{[2]}, \theta^{[2]}) \sim \text{Gamma}(\lambda_\kappa^{[2]}\theta^{[2]}, 1/\psi^{[2]}) \quad \text{with } \lambda_\kappa^{[2]} = \exp(\boldsymbol{x}_\kappa^{[2]}\boldsymbol{\beta}^{[2]}),$$

where $\lambda_\kappa^{[2]}\theta^{[2]}$ is the mean and $1/\psi^{[2]}$ is the shape parameter. In this data analysis, we use same a priori risk characteristics for both the frequency and severity parts, that is, $\boldsymbol{x}_\kappa^{[1]} = \boldsymbol{x}_\kappa^{[2]}$, and $\boldsymbol{\beta}^{[1]}$ and $\boldsymbol{\beta}^{[2]}$ are the corresponding regression coefficient vectors. For the bivariate random effect $(\Theta^{[1]}, \Theta^{[2]})$, we assume bivariate Gaussian copula C in (2) with the correlation coefficient ρ, and the Lognormal distributions with different parameters for the marginal distributions specified as

$$\Theta^{[1]} \sim \text{Lognormal}(-(\sigma^{[1]})^2/2, (\sigma^{[1]})^2) \quad \text{and} \quad \Theta^{[2]} \sim \text{Lognormal}(-(\sigma^{[2]})^2/2, (\sigma^{[2]})^2).$$

The estimation results of this model are summarized in Table B.5, and we refer to Oh *et al.* [17] for further detailed estimation procedures. Note that $*$ sign indicates that the parameters whose 95% credible interval does not contain zero.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

5.2. Optimal BM relativities in the modified BMS

With the parameter estimates in Table B.5, the optimal BM relativities in (19) and the stationary distribution in (17) are calculated under four different BMSs. As per the threshold φ, we consider the estimated 75th, 90th, 99th, and 99.9th quantile of severity distribution, and the values of the HMSE are also provided to compare the performance. The results are summarized in Table B.6.

In Table B.6(a) and (b), the BMSs with $h_1 = h_2$ are identical to those in Table B.2 of Oh *et al.* [17]. Paying attention to the $-1/+1/+2$ and $-1/+2/+3$ systems, we note that:

- As φ decreases, the proportion of policyholders increases in higher BM levels and decreases in lower BM levels, as seen from the stationary distributions. This occurs because more drivers end up in high BM levels under smaller threshold φ. This tendency is naturally more pronounced for the $-1/+2/+3$ system compared with the $-1/+1/+2$ system, as the penalty under the former system is heavier.
- Interestingly, the HMSEs of the optimal relativities under the $-1/+1/+2$ system are generally smaller than those under the $-1/+2/+3$ system for each given threshold. Though difficult to explain this in plain terms, we believe that the stationary distributions under the $-1/+1/+2$ system is more evenly spread than that under the $-1/+2/+3$ system, which, in turn, marginally stabilizes the BMS and produces smaller relativities. For the relationship between the stationary distribution and the optimal relativity, see Eqs. (19)–(21).
- In terms of the HMSE, we see that the optimal threshold φ must be quite large under both the $-1/+1/+2$ and $-1/+2/+3$ systems in this particular dataset.

Overall, based on the HMSE criterion, we would recommend the $-1/+1/+2$ transition rule with a highest threshold φ, the 99.9th quantile, as a new BMS of choice; different datasets may lead to different optimal thresholds. Though the improvement in the HMSE may seem marginal compared with the frequency-driven BMS case in this dataset, the actual amount of dollar saved can be substantial for insurers with large auto portfolios.

6. Conclusion

In auto insurance ratemaking, the premium based on a BMS can be understood as a discretely approximated analogy of the Bayesian premium. Traditionally, insurers have been using the frequency-driven BMS, which completely ignores the claim severity information because of mathematical convenience and industry conventions. More sophisticated BMSs have also been proposed based on both frequency and severity of claim, but these adopt the implicit assumption of independence between frequency and severity, which is often invalid in reality. A recent study by Oh *et al.* [17] extends these previous works to show how one could include both the frequency and severity components in a BMS that targets the aggregate severity when frequency and severity are dependent. However, its transition rule is still determined by the history of claim frequency only. Thus, the resulting BMS premium approximates the Bayesian premium, in a limited sense, and is unable to incorporate the history of claim severity.

In the current study, we extend the approach of Oh *et al.* [17] and propose a new BMS. The proposed BMS also targets the aggregate severity, where frequency and severity are dependent, but is more faithful to the purpose of the optimal relativity, that is, estimating of the true premium by reflecting the history of both frequency and severity. Under the proposed BMS, the new transition rule is $-1/-h_1/+h_2$ based on both frequency and severity, with some threshold on the size of severity. With this transition rule, the Bayesian premium can fully incorporate the history of both frequency and severity, and so can the BMS premium from the analogy of the two premiums.

As theoretical contributions, we analytically derive the set of optimal relativities under this new BMS. Furthermore, with the proper choice of the severity threshold and transition rule, the proposed BMS is guaranteed to outperform the alternative frequency-driven BMS in terms of the HMSE. In this study, we assume that the random effects are static in time. We may generalize the proposed model to incorporate a more general stochastic process by allowing the random effects to vary in time, and

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

then, the similar method as in this paper can define the corresponding set of optimal relativities. The calculation of such optimal relativities can be a possibly interesting topic in the future.

Acknowledgments. Rosy Oh was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. 2020R1I1A1A01067376). Joseph H.T. Kim is grateful for the support of the National Research Foundation of Korea (NRF 2019R1F1A1058095). Jae Youn Ahn was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (2020R1F1A1A01061202).

References

[1] Baumgartner, C., Gruber, L.F., & Czado, C. (2015). Bayesian total loss estimation using shared random effects. *Insurance: Mathematics and Economics* 62: 194–201.
[2] Boudreault, M., Cossette, H., Landriault, D., & Marceau, E. (2006). On a risk model with dependence between interclaim arrivals and claim sizes. *Scandinavian Actuarial Journal* 2006(5): 265–285.
[3] Cheung, E.C., Ni, W., Oh, R., & Woo, J.-K. (2021). Bayesian credibility under a bivariate prior on the frequency and the severity of claims. Working Paper.
[4] Czado, C., Kastenmeier, R., Brechmann, E.C., & Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal* 2012(4): 278–305.
[5] Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J.-F (2007). *Actuarial modelling of claim counts: Risk classification, credibility and Bonus-Malus systems*. Chichester: John Wiley & Sons.
[6] Frangos, N.E. & Vrontos, S.D. (2001). Design of optimal Bonus-Malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin: The Journal of the IAA* 31(1): 1–22.
[7] Frees, E.W., Derrig, R.A., & Meyers, G. (2014). *Predictive modeling applications in actuarial science*, vol. 1. New York: Cambridge University Press.
[8] Frees, E.W., Lee, G., & Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks* 4(1): 4.
[9] Garrido, J., Genest, C., & Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics* 70: 205–215.
[10] Gómez-Déniz, E. (2016). Bivariate credibility Bonus–Malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics* 70: 117–124.
[11] Hernández-Bastida, A., Fernández-Sánchez, M.P., & Gómez-Déniz, E. (2009). The net Bayes premium with dependence between the risk profiles. *Insurance: Mathematics and Economics* 45(2): 247–254.
[12] Klugman, S.A., Panjer, H.H., & Willmot, G.E. (2012). *Loss models: From data to decisions*, vol. 715. Hoboken: John Wiley & Sons.
[13] Lemaire, J. (2012). *Bonus-Malus systems in automobile insurance*, vol. 19. New York: Springer Science & Business Media.
[14] Lu, Y. (2019). Flexible (panel) regression models for bivariate count–continuous data with an insurance application. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182(4): 1503–1521.
[15] McCullagh, P. & Nelder, J.A (1989). *Generalized linear models*, 2nd ed. London: Chapman and Hall.
[16] Nelder, J.A. & Wedderburn, R.W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* 135(3): 370–384.
[17] Oh, R., Shi, P., & Ahn, J.Y. (2020). Bonus-Malus premiums under the dependent frequency-severity modeling. *Scandinavian Actuarial Journal* 2020(3): 172–195.
[18] Park, S.C., Kim, J.H., & Ahn, J.Y. (2018). Does hunger for bonuses drive the dependence between claim frequency and severity?. *Insurance: Mathematics and Economics* 83: 32–46.
[19] Picard, P. (1976). Generalisation de l'etude sur la survenance des sinistres en assurance automobile. *Bulletin Trimestriel de l'Institut des Actuaires Français* 296: 204–268.
[20] Pinquet, J. (1997). Allowance for cost of claims in Bonus-Malus systems. *ASTIN Bulletin* 27(1): 33–57.
[21] Shi, P., Feng, X., & Ivantsova, A. (2015). Dependent frequency–severity modeling of insurance claims. *Insurance: Mathematics and Economics* 64: 417–428.
[22] 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.
[23] Tzougas, G., Vrontos, S., & Frangos, N. (2014). Optimal Bonus-Malus systems using finite mixture models. *ASTIN Bulletin: The Journal of the IAA* 44(2): 417–444.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Appendix A. Proofs

Proof. First, observe that

$$\mathbb{E}[S_{T+1} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}] = \arg\min_\eta \mathbb{E}[(\Theta^{[1]}\Theta^{[2]}\lambda_\kappa^{[1]}\lambda_\kappa^{[2]} - \eta(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T))^2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \mathcal{F}_T^{[\text{full}]}],$$

where the right-hand side is optimized among any function η. Then, the proof follows as such:

$$\begin{aligned}
&\mathbb{E}[(S_{T+1} - \mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \mathcal{F}_T^{[\text{full}]}])^2] \\
&\quad = \mathbb{E}[\mathbb{E}[(S_{T+1} - \mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \mathcal{F}_T^{[\text{full}]}])^2 \mid \mathcal{F}_T^{[\text{full}]}], \Lambda^{[1]}, \Lambda^{[2]}]
\end{aligned}$$

□

Proof. Before we provide the proof, we found it beneficial to define

$$T_{\ell,\ell^*}(\cdot,\cdot) : \{0, 1, \ldots, z\} \times \{0, 1, \ldots, z\} \mapsto \{0, 1\},$$

which indicates whether the BM level has changed from ℓ in the current year to ℓ^* in the next year if there were k_1 number of Type I claims and k_2 number of Type II claims in the current year. Mathematically, it is defined as, for $(k_1, k_2) = (0, 0)$,

$$T_{\ell,\ell^*}(k_1, k_2) = \begin{cases} 1, & \text{if } \max\{\ell - 1, 0\} = \ell^*; \\ 0, & \text{otherwise;} \end{cases} \tag{A.1}$$

and, for $(k_1, k_2) \neq (0, 0)$,

$$T_{\ell,\ell^*}(k_1, k_2) = \begin{cases} 1, & \text{if } \min\{\ell + h_1 k_1 + h_2 k_2, z\} = \ell^*; \\ 0, & \text{otherwise.} \end{cases} \tag{A.2}$$

By definition of transition rule in (10), it is known that the BM level L_{t+1} is completely explained by L_t and $(N_t^{\{\text{I}\}}, N_t^{\{\text{II}\}})$. Now, the transition probability can be calculated as

$$\begin{aligned}
&p_{\ell,\ell^*}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) \\
&\quad = \mathbb{P}(L_{t+1} = \ell^* \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, L_t = \ell) \\
&\quad = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \mathbb{P}(L_{t+1} = \ell^* \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2, L_t = \ell) \\
&\qquad \times \mathbb{P}(N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, L_t = \ell) \\
&\quad = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \mathbb{P}(L_{t+1} = \ell^* \mid N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2, L_t = \ell) \\
&\qquad \times \mathbb{P}(N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, L_t = \ell) \\
&\quad = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} T_{\ell,\ell^*}(k_1, k_2)\mathbb{P}(N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2 \mid L_t = \ell, \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) \\
&\quad = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} T_{\ell,\ell^*}(k_1, k_2)\mathbb{P}(N_t^{\{\text{I}\}} = k_1, N_t^{\{\text{II}\}} = k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}),
\end{aligned} \tag{A.3}$$

where the last equality comes from the independence assumption among frequency, among individual severities, and between frequency and individual severity for given pair of risk characteristics, $(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]})$.

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Now, we consider the case $\ell < \ell^*$ or $\ell = \ell^* = z$. Then, for $(k_1, k_2) \in \mathbb{N}_0 \times \mathbb{N}_0$, we obtain

$$T_{\ell,\ell^*}(k_1, k_2) = \begin{cases} 1, & \ell^* = \max\{\ell + k_1 h1 + k_2 h_2, z\}; \\ 0, & \text{elsewhere}, \end{cases} \tag{A.4}$$

and

$$\begin{aligned} &\mathbb{P}(N_t^{\{\mathrm{I}\}} = k_1, N_t^{\{\mathrm{II}\}} = k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) \\ &= \mathbb{P}(N_t^{\{\mathrm{I}\}} = k_1, N_t^{\{\mathrm{II}\}} = k_2, N_t = k_1 + k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) \\ &= \mathbb{P}(N_t^{\{\mathrm{I}\}} = k_1, N_t^{\{\mathrm{II}\}} = k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}, N_t = k_1 + k_2) \mathbb{P}(N_t = k_1 + k_2 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) \\ &= \mathbb{P}(N_t = k_1 + k_2 \mid \lambda_\kappa^{[1]}, \theta^{[1]}) \binom{k_1 + k_2}{k_1} (\mathbb{P}(Y_{t,j} > \varphi \mid \lambda_\kappa^{[2]}, \theta^{[2]}))^{k_2} (1 - \mathbb{P}(Y_{t,j} > \varphi \mid \lambda_\kappa^{[2]}, \theta^{[2]}))^{k_1}. \end{aligned} \tag{A.5}$$

Then, combining (A.3), (A.4), and (A.5), we have the first line in (11).
Now, consider the case $\ell^* = \max\{\ell - 1, 0\}$. Then, we have

$$T_{\ell,\ell^*}(k_1, k_2) = \begin{cases} 1, & k_1 = k_2 = 0, \\ 0, & \text{elsewhere}. \end{cases} \tag{A.6}$$

Hence, combining (A.3) and (A.6), conclude the second line in (11).
In case the pair (ℓ, ℓ^*) satisfies none of the following

$$\ell < \ell^*, \quad \ell = \ell^* = z, \quad \text{or} \quad \ell^* = \max\{\ell - 1, 0\},$$

we obtain

$$T_{\ell,\ell^*}(k_1, k_2) = 0$$

which concludes the proof. □

Proof. Under the frequency–severity model, we obtain

$$\begin{aligned} &\mathbb{E}\left[(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]} r^{[\mathrm{new}]}(L))^2\right] \\ &\quad = \sum_{\ell=0}^{z} \mathbb{E}\left[(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \Lambda^{[1]}\Lambda^{[2]} r^{[\mathrm{new}]}(L))^2 \mid L = l\right] \mathbb{P}(L = \ell) \end{aligned}$$

which further implies that the optimal solution $\tilde{r}^{[\mathrm{new}]}(\ell)$ in (18) satisfies

$$0 = \mathbb{E}\left[-2\Lambda^{[1]}\Lambda^{[2]}(\mathbb{E}[S_{T+1} \mid \Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]}] - \tilde{r}^{[\mathrm{new}]}(L)\Lambda^{[1]}\Lambda^{[2]}) \mid L = l\right] \tag{A.7}$$

for $\ell = 0, \ldots, z$. Then, (A.7) is equivalent with (19).

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

First, we calculate $\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2\Theta^{[1]}\Theta^{[2]} \mid L = l]$ in (19) as

$$\begin{aligned}
&\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2\Theta^{[1]}\Theta^{[2]} \mid L = l] \\
&\quad = \sum_{\kappa\in\mathcal{K}} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \mathbb{E}[\Theta^{[1]}\Theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, L = l]\mathbb{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]} \mid L = \ell) \\
&\quad = \sum_{\kappa\in\mathcal{K}} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \iint \theta^{[1]}\theta^{[2]} f(\theta^{[1]}, \theta^{[2]} \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, L = \ell)\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}\, \mathbb{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]} \mid L = \ell) \\
&\quad = \frac{\sum_{\kappa\in\mathcal{K}} w_\kappa (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \iint \theta^{[1]}\theta^{[2]} \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}}{\sum_{\kappa\in\mathcal{K}} w_\kappa \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}},
\end{aligned}$$

where the last equation comes from the independence between the a priori and a posteriori. Similarly, we can calculate $\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2 \mid L = l]$ in (19) as

$$\begin{aligned}
\mathbb{E}[(\Lambda^{[1]}\Lambda^{[2]})^2 \mid L = l] &= \sum_{\kappa\in\mathcal{K}} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \mathbb{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]} \mid L = \ell) \\
&= \sum_{\kappa\in\mathcal{K}} (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \frac{\mathbb{P}(L = \ell \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]})\mathbb{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]})}{\mathbb{P}(L = \ell)} \\
&= \frac{\sum_{\kappa\in\mathcal{K}} w_\kappa (\lambda_\kappa^{[1]}\lambda_\kappa^{[2]})^2 \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}}{\sum_{\kappa\in\mathcal{K}} w_\kappa \iint \pi_\ell(\lambda_\kappa^{[1]}\theta^{[1]}, \lambda_\kappa^{[2]}\theta^{[2]}) h(\theta^{[1]}, \theta^{[2]})\, \mathrm{d}\theta^{[1]}\, \mathrm{d}\theta^{[2]}},
\end{aligned}$$

where the last equation comes from the independence between the a priori and a posteriori. □

Appendix B. Miscellaneous

Lemma A.1. *Assume*

$$\mathbb{P}(N_t > 0 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) > 0 \tag{B.1}$$

and

$$\mathbb{P}(N_t = 0 \mid \lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) > 0. \tag{B.2}$$

Then, for the given observed and residual effect characteristic $\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}$, we obtain

$$\boldsymbol{\pi}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) = \boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{P}(\lambda_\kappa^{[1]}, \lambda_\kappa^{[2]}, \theta^{[1]}, \theta^{[2]}) + \boldsymbol{E})^{-1},$$

where $\boldsymbol{E}$ is the $(z+1)\times(z+1)$ matrix all of whose entries are 1 and $\boldsymbol{e}$ is $(z+1)$ column vector all of whose entries are 1.

Proof. Note that (B.1) and (B.2) implies that the matrix

$$\boldsymbol{P}(\Lambda^{[1]}, \Lambda^{[2]}, \Theta^{[1]}, \Theta^{[2]})$$

is regular. Hence, the remaining proof follows from Property 4.2 in Denuit *et al.* [5], for example, among many other references. □

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Table B.1. *Optimal BM relativities and stationary distribution of L for various threshold φ and dependence parameter ρ under the two BMSs: the $-1/+1$ and $-1/+1/+2$ systems.*

	$-1/+1$ system		$-1/+1/+2$ system							
φ	–		8,200		16,800		48,100		94,300	
Level ℓ	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$
(a) For $\rho=-0.8$										
9	1.328	0.135	1.292	0.148	1.320	0.139	1.328	0.135	1.328	0.135
8	1.052	0.055	1.018	0.064	1.046	0.057	1.052	0.055	1.052	0.055
7	0.936	0.034	0.898	0.041	0.928	0.036	0.936	0.034	0.936	0.034
6	0.858	0.026	0.815	0.033	0.849	0.028	0.859	0.026	0.858	0.026
5	0.795	0.024	0.746	0.031	0.782	0.027	0.796	0.024	0.796	0.024
4	0.737	0.026	0.681	0.034	0.719	0.030	0.737	0.027	0.737	0.026
3	0.676	0.034	0.619	0.043	0.654	0.039	0.675	0.035	0.676	0.034
2	0.607	0.055	0.545	0.067	0.579	0.061	0.604	0.056	0.606	0.055
1	0.522	0.114	0.493	0.096	0.510	0.107	0.521	0.113	0.522	0.114
0	0.414	0.496	0.395	0.445	0.406	0.476	0.413	0.495	0.414	0.496
HMSE	1.297		1.293		1.282		1.295		1.297	
(b) For $\rho=-0.4$										
9	2.047	0.135	1.968	0.151	2.026	0.141	2.047	0.136	2.047	0.135
8	1.460	0.055	1.398	0.065	1.449	0.059	1.462	0.055	1.461	0.055
7	1.237	0.034	1.173	0.041	1.225	0.037	1.240	0.034	1.238	0.034
6	1.096	0.026	1.025	0.033	1.080	0.029	1.099	0.026	1.097	0.026
5	0.987	0.024	0.907	0.031	0.964	0.027	0.989	0.024	0.987	0.024
4	0.889	0.026	0.801	0.033	0.859	0.029	0.890	0.027	0.889	0.026
3	0.791	0.034	0.701	0.041	0.754	0.038	0.790	0.035	0.791	0.035
2	0.683	0.055	0.595	0.064	0.643	0.060	0.680	0.056	0.683	0.055
1	0.559	0.114	0.501	0.095	0.527	0.106	0.554	0.113	0.559	0.114
0	0.411	0.496	0.376	0.445	0.392	0.475	0.408	0.494	0.411	0.496
HMSE	6.022		5.995		5.930		5.996		6.018	
(c) For $\rho=0.4$										
9	4.489	0.135	4.129	0.156	4.338	0.145	4.478	0.136	4.489	0.135
8	2.512	0.055	2.285	0.067	2.423	0.060	2.508	0.055	2.512	0.055
7	1.918	0.034	1.715	0.042	1.837	0.037	1.913	0.034	1.918	0.034
6	1.581	0.026	1.387	0.033	1.501	0.029	1.575	0.026	1.580	0.026
5	1.341	0.024	1.154	0.030	1.260	0.027	1.334	0.024	1.340	0.024
4	1.142	0.026	0.964	0.032	1.062	0.029	1.134	0.027	1.141	0.026
3	0.958	0.034	0.797	0.039	0.881	0.037	0.948	0.035	0.957	0.034
2	0.772	0.055	0.636	0.060	0.704	0.057	0.762	0.055	0.771	0.055
1	0.578	0.114	0.487	0.094	0.524	0.105	0.567	0.113	0.576	0.114
0	0.371	0.496	0.322	0.448	0.344	0.476	0.366	0.494	0.370	0.496
HMSE	53.792		55.196		54.138		53.725		53.772	

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Table B.2. *Optimal BM relativities and stationary distribution of L for various threshold φ and $\lambda^{[1]}$ under the three BMSs: the −1/+1, −1/+1/+2, and −1/+1/+3 systems.*

(a) −1/+1 system and −1/+1/+2 system for $\lambda^{[1]} = 0.5$											
		−1/+1 system		−1/ + 1/ + 2 system							
	φ	–		8,200		16,800		48,100		94,300	
Level ℓ		$r(\ell)$	$\mathbb{P}(L = \ell)$	$r(\ell)$	$\mathbb{P}(L = \ell)$	$r(\ell)$	$\mathbb{P}(L = \ell)$	$r(\ell)$	$\mathbb{P}(L = \ell)$	$r(\ell)$	$\mathbb{P}(L = \ell)$
9	3.070	0.135	2.893	0.154	3.009	0.143	3.069	0.136	3.071	0.135	
8		1.951	0.055	1.827	0.066	1.915	0.059	1.954	0.055	1.952	0.055
7		1.572	0.034	1.451	0.042	1.535	0.037	1.574	0.034	1.573	0.034
6		1.344	0.026	1.220	0.033	1.303	0.029	1.345	0.026	1.345	0.026
5		1.174	0.024	1.046	0.030	1.127	0.027	1.174	0.024	1.175	0.024
4		1.029	0.026	0.898	0.033	0.976	0.029	1.026	0.027	1.029	0.026
3		0.888	0.034	0.761	0.040	0.831	0.037	0.883	0.035	0.888	0.035
2		0.741	0.055	0.625	0.062	0.684	0.058	0.733	0.056	0.740	0.055
1		0.579	0.114	0.499	0.094	0.531	0.105	0.569	0.113	0.577	0.114
0		0.396	0.496	0.352	0.447	0.371	0.475	0.392	0.494	0.396	0.496
HMSE		19.120		19.345		19.025		19.049		19.105	
(b) −1/+2 system and −1/+2/+3 system for $\lambda^{[1]} = 0.5$											
9	2.438	0.202	2.367	0.213	2.412	0.206	2.437	0.202	2.438	0.202	
8		1.493	0.093	1.446	0.100	1.478	0.096	1.493	0.093	1.493	0.093
7		1.161	0.059	1.115	0.065	1.145	0.062	1.160	0.060	1.161	0.059
6		0.949	0.047	0.911	0.050	0.935	0.048	0.948	0.047	0.949	0.047
5		0.815	0.040	0.766	0.045	0.796	0.042	0.814	0.040	0.815	0.040
4		0.676	0.045	0.643	0.044	0.659	0.044	0.673	0.045	0.675	0.045
3		0.606	0.042	0.558	0.048	0.587	0.045	0.605	0.043	0.606	0.042
2		0.476	0.074	0.445	0.066	0.457	0.071	0.472	0.074	0.475	0.074
1		0.446	0.059	0.417	0.053	0.429	0.056	0.443	0.059	0.446	0.059
0		0.312	0.338	0.293	0.316	0.301	0.329	0.310	0.337	0.311	0.338
HMSE		21.288		21.509		21.325		21.275		21.285	
(c) −1/+1 system and −1/+1/+2 system for $\lambda^{[1]} = 2$											
9		1.515	0.554	1.479	0.574	1.501	0.563	1.514	0.555	1.515	0.554
8		0.640	0.111	0.616	0.119	0.631	0.115	0.640	0.112	0.640	0.111
7		0.476	0.050	0.450	0.054	0.466	0.052	0.475	0.050	0.476	0.050
6		0.393	0.031	0.364	0.033	0.380	0.032	0.392	0.031	0.393	0.031
5		0.339	0.024	0.307	0.025	0.324	0.024	0.338	0.024	0.339	0.024
4		0.298	0.021	0.265	0.022	0.281	0.022	0.296	0.022	0.298	0.021
3		0.263	0.023	0.228	0.022	0.245	0.023	0.260	0.023	0.262	0.023
2		0.229	0.029	0.196	0.026	0.210	0.028	0.226	0.029	0.229	0.029
1		0.194	0.043	0.165	0.032	0.176	0.038	0.190	0.042	0.194	0.043
0		0.154	0.114	0.133	0.093	0.141	0.105	0.152	0.113	0.154	0.114
HMSE		394.566		398.329		394.815		394.581		394.553	

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Table B.3. *Optimal BM relativities and stationary distribution of L for various threshold φ and $\sigma^{[2]}$ under the two BMSs: the $-1/+1$ and $-1/+1/+2$ systems.*

		(a) $-1/+1$ system and $-1/+1/+2$ system for $(\sigma^{[2]})^2 = 0.01$									
		$-1/+1$ system		$-1/+1/+2$ system							
	φ	–		9,000		16,800		38,000		60,400	
Level ℓ		$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$
9	3.070	0.135	2.852	0.154	2.976	0.143	3.060	0.136	3.069	0.135	
8		1.951	0.055	1.786	0.066	1.879	0.059	1.944	0.055	1.951	0.055
7		1.572	0.034	1.416	0.042	1.503	0.037	1.564	0.034	1.571	0.034
6		1.344	0.026	1.193	0.033	1.276	0.029	1.337	0.026	1.343	0.026
5		1.174	0.024	1.027	0.030	1.107	0.027	1.167	0.024	1.174	0.024
4		1.029	0.026	0.886	0.033	0.963	0.029	1.021	0.027	1.028	0.026
3		0.888	0.034	0.761	0.041	0.827	0.038	0.881	0.035	0.888	0.035
2		0.741	0.055	0.628	0.062	0.688	0.058	0.735	0.056	0.740	0.055
1		0.579	0.114	0.531	0.094	0.558	0.105	0.576	0.113	0.578	0.114
0		0.396	0.496	0.367	0.445	0.384	0.474	0.395	0.494	0.396	0.496
HMSE		9.480		10.158		9.760		9.507		9.482	
		(b) $-1/+1$ system and $-1/+1/+2$ system for $(\sigma^{[2]})^2 = 0.29$									
9	3.070	0.135	2.893	0.154	3.009	0.143	3.069	0.136	3.071	0.135	
8		1.951	0.055	1.827	0.066	1.915	0.059	1.954	0.055	1.952	0.055
7		1.572	0.034	1.451	0.042	1.535	0.037	1.574	0.034	1.573	0.034
6		1.344	0.026	1.220	0.033	1.303	0.029	1.345	0.026	1.345	0.026
5		1.174	0.024	1.046	0.030	1.127	0.027	1.174	0.024	1.175	0.024
4		1.029	0.026	0.898	0.033	0.976	0.029	1.026	0.027	1.029	0.026
3		0.888	0.034	0.761	0.040	0.831	0.037	0.883	0.035	0.888	0.035
2		0.741	0.055	0.625	0.062	0.684	0.058	0.733	0.056	0.740	0.055
1		0.579	0.114	0.499	0.094	0.531	0.105	0.569	0.113	0.577	0.114
0		0.396	0.496	0.352	0.447	0.371	0.475	0.392	0.494	0.396	0.496
HMSE		19.120		19.345		19.025		19.049		19.105	
		(c) $-1/+1$ system and $-1/+1/+2$ system for $(\sigma^{[2]})^2 = 1$									
9	3.070	0.135	2.968	0.154	3.078	0.143	3.092	0.136	3.076	0.135	
8		1.951	0.055	1.901	0.066	1.987	0.059	1.980	0.055	1.958	0.055
7		1.572	0.034	1.512	0.041	1.598	0.037	1.598	0.034	1.578	0.034
6		1.344	0.026	1.266	0.032	1.352	0.029	1.365	0.026	1.350	0.026
5		1.174	0.024	1.073	0.030	1.161	0.027	1.189	0.024	1.179	0.024
4		1.029	0.026	0.914	0.032	0.995	0.029	1.035	0.027	1.031	0.026
3		0.888	0.034	0.750	0.040	0.827	0.037	0.885	0.035	0.889	0.034
2		0.741	0.055	0.616	0.062	0.675	0.058	0.729	0.056	0.739	0.055
1		0.579	0.114	0.439	0.095	0.480	0.106	0.551	0.113	0.573	0.114
0		0.396	0.496	0.322	0.448	0.346	0.476	0.383	0.494	0.394	0.496
HMSE		59.706		59.053		58.814		59.365		59.631	

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Table B.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 B.5. *(Data analysis) Estimation results under the frequency–severity model.*

			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	
Coverage 2	1.247	0.212	0.829	1.667	*
Coverage 3	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	
Coverage 2	0.180	0.244	−0.308	0.646	
Coverage 3	−0.027	0.261	−0.533	0.493	
$1/\psi^{[2]}$	0.670	0.041	0.592	0.752	*
Copula part					
$(\sigma^{[1]})^2$	0.992	0.142	0.746	1.292	*
$(\sigma^{[2]})^2$	0.293	0.067	0.176	0.433	*
ρ	−0.447	0.130	−0.690	−0.190	*

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press

Table B.6. *(Data analysis) Optimal BM relativities and stationary distribution of L for various threshold φ and under the different BMSs.*

		(a) $-1/+1/+1$ system and $-1/+1/+2$ system									
		$-1/+1/+1$ system		$-1/+1/+2$ system							
	φ	–		10,100		21,400		64,500		132,100	
Level ℓ		$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$	$r(\ell)$	$\mathbb{P}(L=\ell)$
9		0.966	0.146	0.952	0.153	0.959	0.148	0.965	0.146	0.966	0.146
8		0.454	0.034	0.441	0.038	0.448	0.035	0.453	0.034	0.454	0.034
7		0.369	0.018	0.354	0.021	0.361	0.019	0.368	0.018	0.369	0.018
6		0.327	0.013	0.312	0.016	0.318	0.014	0.326	0.013	0.327	0.013
5		0.302	0.012	0.286	0.014	0.292	0.013	0.300	0.012	0.302	0.012
4		0.284	0.013	0.268	0.016	0.273	0.014	0.282	0.013	0.284	0.013
3		0.271	0.017	0.256	0.021	0.260	0.018	0.268	0.017	0.270	0.017
2		0.259	0.029	0.248	0.037	0.250	0.032	0.256	0.029	0.259	0.029
1		0.250	0.075	0.243	0.066	0.243	0.071	0.247	0.074	0.250	0.075
0		0.243	0.645	0.239	0.619	0.239	0.635	0.241	0.645	0.242	0.645
HMSE		91.242		92.128		91.634		91.269		91.241	
		(b) $-1/+2/+2$ system and $-1/+2/+3$ system									
9	0.932	0.178	0.927	0.183	0.929	0.180	0.931	0.178	0.932	0.178	
8		0.428	0.051	0.424	0.054	0.426	0.053	0.428	0.052	0.428	0.051
7		0.343	0.030	0.339	0.033	0.341	0.031	0.343	0.030	0.343	0.030
6		0.302	0.023	0.298	0.025	0.300	0.024	0.301	0.023	0.302	0.023
5		0.278	0.020	0.274	0.023	0.276	0.021	0.278	0.020	0.278	0.020
4		0.262	0.024	0.260	0.025	0.260	0.025	0.261	0.024	0.262	0.024
3		0.253	0.024	0.251	0.031	0.251	0.027	0.253	0.025	0.253	0.024
2		0.245	0.055	0.244	0.051	0.244	0.054	0.245	0.055	0.245	0.055
1		0.243	0.046	0.242	0.043	0.242	0.045	0.242	0.046	0.243	0.046
0		0.239	0.547	0.238	0.531	0.238	0.541	0.238	0.546	0.239	0.547
HMSE		93.483		93.786		93.625		93.498		93.484	

Cite this article: Oh R, Kim JHT and Ahn JY (2022). Designing a Bonus-Malus system reflecting the claim size under the dependent frequency–severity model. *Probability in the Engineering and Informational Sciences* **36**, 963–987. https://doi.org/10.1017/S0269964821000188

https://doi.org/10.1017/S0269964821000188 Published online by Cambridge University Press