Insurance: Mathematics and Economics 83 (2018) 32–46

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Does hunger for bonuses drive the dependence between claim frequency and severity?

Sojung C. Park [a,1], Joseph H.T. Kim [b,1], Jae Youn Ahn [c,*]

[a] College of Business Administration, Seoul National University, Gwanak-gu, Seoul 151-916, Republic of Korea
[b] Department of Applied Statistics, College of Business and Economics, Yonsei University, Seoul, 120-749, Republic of Korea
[c] Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong,Seodaemun-Gu, Seoul 120-750, Republic of Korea

ARTICLE INFO

Article history:
Received March 2018
Received in revised form July 2018
Accepted 5 September 2018
Available online 14 September 2018

JEL classification:
C6

Keywords:
Dependence
Generalized linear model
Bonus Hunger
Bonus-malus system
Optimal retention

ABSTRACT

Auto ratemaking models have traditionally assumed independence between claim frequency and severity. With the development of insurance claim models that can accommodate dependence between claim frequency and severity, a series of recent studies has revealed that the aforementioned dependence between frequency and severity exists for auto insurance claims, demonstrating the validity of such models. However, the underlying process that creates this dependence has received little attention in the literature. Thus, we show that a rational decision-making process of drivers known as bonus hunger can systemically induce dependence between the claim frequency and severity even when the ground-up loss frequency and severity are, in fact, independent. Our model, based on the random effect model coupled with the standard bonus–malus system, successfully explains the seemingly contradictory results from the existing literature of weak positive dependence, between the claim frequency and severity for liability claims, and moderately negative dependence for collision claims. Our findings show that the seemingly contradicting dependence structures reported in the literature may be neither accidental nor sample specific. Furthermore, the bonus-hunger process also implies that the level of the claim frequency-severity dependence varies across bonus–malus classes, suggesting that a uniform dependency structure may not be appropriate for auto ratemaking modeling.

1. Introduction

Modeling insurance claims is a critical task in the auto insurance industry. Traditional claim models, such as two-part models (Frees et al., 2014), the Tweedie compound Poisson models (Jørgensen and Paes De Souza, 1994), and their various extensions, mostly impose independence between the claim frequency and severity, and independence among the claim severities. Such independence assumptions are convenient but are not always justifiable. Recently, a series of studies attempted to relax such assumptions and incorporate dependence into ratemaking models. These studies can be largely classified into three categories based on the dependence structure formulation: a frequency-severity model based on a generalized linear model (GLM) framework in which the claim frequency is a covariate in the severity regression model (e.g., Frees et al. 2014, Shi et al. 2015, Garrido et al. 2016, and Jeong et al. 2017), a copula model in which the frequency and severity components are linked via some parametric copula function (e.g., Czado et al. 2012 and Frees et al. 2016), and a (shared) random effect model (e.g., Dimakos and Di Rattalma 2002 and Baumgartner et al. 2015).

Although these new developments lead to a better understanding of the inherent risk and provide more appropriate pricing and risk management tools for insurers, the studies are potentially limited in some aspects. First, although the empirical analyses in these studies support the evidence of dependence between claim frequency and severity, they mostly assume a constant dependence parameter. As auto insurance premiums depend on the claim experiences of previous years and a driver can use his or her discretion to report losses, assuming a fixed constant parameter to model the dependence may not be flexible enough. Second, the dependence reported in the literature is somewhat inconsistent. For example, Frees et al. (2014), Shi et al. (2015), and Garrido et al. (2016) find negative dependence between the claim frequency and severity in the case of collision coverage but no dependence or weak positive dependence in the case of liability coverage. The reason for these apparently contradictory results has not been investigated. Finally, there has been no discussion of the underlying process that drives the dependence between claim frequency and severity. This point is actually related to the first two issues; if we properly understand the hidden driver of this dependence, we may

* Corresponding author.
E-mail addresses: sojungpark@snu.ac.kr (S.C. Park), jhtkim@yonsei.ac.kr (J.H.T. Kim), jaeyahn@ewha.ac.kr (J.Y. Ahn).
[1] The First Authors.

https://doi.org/10.1016/j.insmatheco.2018.09.002
0167-6687/

be able to explain the inconsistent behavior of the dependence and build the pricing model in a more realistic and appropriate manner.

To tackle these issues, we extend the existing claim frequency-severity dependence models by incorporating a behavioral element of the auto driver based on the bonus–malus system (BMS), a common mechanism for premium setting in auto insurance. In particular, we show that a particular driver behavior called bonus hunger can induce dependence between the frequency and severity of auto claims. To our knowledge, this is the first attempt to integrate the underlying process of the claim frequency-severity dependence in the pricing model. The BMS is a standard *a posteriori* ratemaking device to adjust premium levels based on claim experiences. Since the premium generally increases as a driver reports more claims, drivers tend to avoid reporting small losses to maintain a low premium. This behavior is known as bonus hunger. Bonus hunger is known to be a prevalent phenomenon in insurance contracts within any bonus–malus rating system (Lemaire, 2012). It is also known that bonus–malus rating systems induce discrepancies between accident records and claim records (Lemaire, 1977; Walhin and Paris, 2000), which means that ground-up losses and claim amounts generally differ in practice. In this study, we argue that bonus-hunger behavior not only distorts the distributions of claim frequency and severity (Lemaire, 1977; Walhin and Paris, 2000) but can also be a factor in driving the dependence between the claim frequency and severity even when the original accident (ground-up loss) frequency and severity are independent under a given BMS.

Our argument is laid out as follows. We first set up a typical frequency-severity model and build a BMS alongside this model as a mechanism to control the number of claims made by each driver. More specifically, we use the Lemaire algorithm (Lemaire, 2012) to compute and construct an optimal retention table, which serves as the threshold for whether a loss resulting from an accident is reported or not. If a loss exceeds the optimal retention level, the driver reports the claim, but the driver does not bother to file claims for small losses below the threshold. Our claim is that for higher bonus–malus (BM) levels, the optimal retention is a decreasing function of the number of claims already reported, whereas it is an increasing function for lower BM levels. In other words, at higher BM levels, policyholders only report expensive losses for the first few accidents but start to report cheaper accidents for subsequent accidents, whereas this pattern reverses at lower BM levels. This behavioral pattern in turn suggests positive or no dependence between the reported claim frequency and severity at lower BM levels and negative dependence at higher BM levels. An important implication of this pattern is that, because of bonus hunger, the dependence between the frequency and severity may vary by BM level rather than being constant for all policyholders. A simplified mathematical analysis and a simulation study are shown to support our claim. Furthermore, we propose a frequency-severity model that can accommodate varying dependences across BM levels, and we quantify the changes in dependence across different BM levels based on this model.

As a real-world application, we apply our model to Korean auto insurance claim data. In this empirical analysis, we consider two different types of coverage separately: first-party physical damage coverage and liability coverage. In Korea, first-party physical damage covers the driver's car damage, which can be thought of as a combination of the collision and comprehensive coverages in US auto insurance contracts. For first-party physical damage coverage, we show that the dependence parameter is on average negative and does gradually change from positive to negative as the BM level increases, which empirically confirms that our theoretical mechanism of bonus hunger affects the dependence between frequency and severity. In contrast, the analysis of liability claims shows that this dependence is on average positive and is largely invariant to changes in BM levels. This seemingly contradicting result, again, can be attributed to the driver's bonus-hunger behavior because, although the driver can exercise his or her discretion in filing a claim for first-party physical damage coverage, it is much more difficult for the driver to avoid claiming an accident or to reduce the size of the loss because liability losses involve a third party. Interestingly, these findings can be related to previous studies on dependence in auto insurance claims. Specifically, by employing a constant dependence parameter on the frequency-severity model, Frees et al. (2014) and Shi et al. (2015) find weak positive dependence between the claim frequency and severity using auto liability insurance data from the state of Massachusetts in 2006 (Ferreira Jr and Minikel, 2012). In contrast, Garrido et al. (2016) report negative dependence based on an analysis of collision claims in a Canadian automobile insurance dataset, again using a constant dependence parameter in the frequency-severity model.

In summary, we argue that, provided that the insured party can exercise his or her discretion to decide to report or not to report an accident, opportunistic claim behavior can induce a stochastic dependence pattern between claim frequency and severity, with the degree of dependence varying across BM levels. Furthermore, the dependence structure qualitatively varies depending on whether the insured has control over whether or not to report an accident. The contributions of the present study are, thus, threefold. First, to our knowledge, this is the first study that attempts to explain the underlying cause that drives the dependence between claim frequency and severity based on bonus hunger, a behavioral decision-making process of drivers. Second, this is the first study to suggest the necessity of using heterogeneous dependence structures across different groups of drivers and different types of coverage. Although we use the BMS class to group drivers and differentiate dependence structures, it is possible that other classifications may produce different heterogeneous groups as well. This heterogeneity also suggests that a strong dependence may exist in certain sub-groups even when the dependence may seem insignificant when considering a pooled group.

The remainder of this article is organized as follows. In Section 2, we set up a frequency-severity modeling framework that allows dependence between claim frequency and severity. In Section 3, we describe BM systems and bonus-hunger behavior along with the Lemaire algorithm as a tool to determine the optimal retention level. A synthetic BMS is constructed, and the simulated dataset is analyzed in Section 4 to validate and measure the frequency-severity dependence. Section 5 provides in-depth analyses based on a real insurance claim dataset, focusing on validating our claim on the frequency-severity dependence and comparing our findings with those of previous studies. We conclude the study in Section 6.

2. The frequency-severity model

In this section, we construct a standard frequency-severity model found in the literature. Consider an auto insurance portfolio consisting of k drivers. Assuming no policy modifications such as deductibles or limits, the standard model of the ith driver's total loss during a given time period is expressed as a compound risk model

$$S_i = \sum_{j=1}^{N_i} Y_{i,j}, \qquad i = 1, \dots, k, \tag{1}$$

where N_i represents the random claim counts (or frequencies) and $Y_{i,j}$ is the claim size (or severity) of the jth claim with $j = 1, \dots, N_i$. We set $S_i = 0$ when $N_i = 0$ by convention so that the loss history of the driver is written as

$$\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,N_i})^T, \quad N_i \geq 1. \tag{2}$$

For later use, we further denote the (random) average claim severity for the ith driver as

$$M_i = \frac{S_i}{N_i}, \quad N_i \geq 1, \tag{3}$$

where $\mathbf{Y}_i$ and M_i are not defined when $N_i = 0$.

The aggregate loss variable S_i has been an important actuarial topic in general insurance, and its various distributional properties are now well known (Klugman et al., 2008). When relevant covariates are available at the individual level and are used in the ratemaking process, relatively simple and mathematically tractable models have been preferred in the literature, such as members of the GLM class; see, for example, Denuit et al. (2007), De Jong et al. (2008), Frees (2010), and Lemaire (2012). Following similar lines, in this study, we consider the negative binomial and Poisson distributions for the frequency.[2] For the severity, we choose gamma and inverse Gaussian distributions, both of which belong to the exponential dispersion family (Garrido et al., 2016). These models are popular members of the GLM class found in the literature and easily accommodate relevant covariates as needed.

2.1. Model assumptions for frequency

We now specify parametric model assumptions, which are used throughout the analysis, under the random effects model. These models are found in previous studies, such as Denuit et al. (2007), De Jong et al. (2008), Lemaire (2012), Frees et al. (2014), Shi et al. (2015), Garrido et al. (2016), and Jeong et al. (2017). Suppose that a vector $\mathbf{x}_i = (x_{1,i}, \ldots, x_{p,i})$ of covariates is available for the ith driver to explain the frequency and severity variables. As each element of the vector can be relevant to frequency, severity, or both, we may create two vectors, $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$, so that the former is associated with the claim frequency and the latter with the severity, where the first elements are equal to one by convention. Some elements of these two vectors may overlap if the corresponding covariates are relevant to both variables. Given this structure, we characterize the frequency model of the ith driver using a Poisson regression model:

$$N_i \sim \mathcal{P}(\lambda_i \Theta) \qquad \text{with} \quad \log(\lambda_i) = \boldsymbol{\alpha}^T \mathbf{x}_{1,i}, \tag{4}$$

where λ_i is the a priori expected frequency of the policyholder and $\Theta = \theta_i$ is the residual effect of hidden covariates that are not included in a priori ratemaking. The log link ensures that the Poisson intensity remains positive. We further assume that the prior distribution of Θ follows a gamma distribution. For our purpose, we reparameterize the gamma distribution, where the density of $\mathcal{G}(\xi, \upsilon)$ is defined as

$$f(y|\xi, \upsilon) = \frac{1}{\Gamma(\upsilon)} \left(\frac{\upsilon}{\xi}\right)^{\upsilon} y^{\upsilon-1} \exp\left(-\frac{y\upsilon}{\xi}\right), \quad y > 0 \tag{5}$$

with mean ξ and the variance ξ^2/υ. Here, $1/\upsilon$ is called the dispersion parameter. Under this reparameterization, the prior variable Θ is assumed to follow $\mathcal{G}(1, \upsilon)$ with mean $1/\upsilon$ and variance $1/\upsilon^2$. Note that N_i unconditionally follows a negative binomial distribution $NB(\lambda, \upsilon)$ with a probability mass function given by

$$f(n|\lambda, \upsilon) = \frac{\Gamma(\upsilon + n)}{\Gamma(\upsilon)\Gamma(n+1)} \left(\frac{\lambda}{\upsilon + \lambda}\right)^n \left(\frac{\upsilon}{\upsilon + \lambda}\right)^{\upsilon}$$

with mean λ and variance $\lambda + \lambda^2/\upsilon$.

2.2. Model assumptions for severity

For severity, we consider two popular distributions in the exponential dispersion family (Garrido et al., 2016): the gamma distribution in (5) and the inverse Gaussian distribution (IG in short), whose density is given by

$$f(y|\mu, \rho) = \frac{\rho}{\sqrt{2\pi y^3}} \exp\left[-\frac{\rho(y-\mu)^2}{2\mu^2 y}\right] \tag{6}$$

with mean μ and variance μ^3/ρ. When applied to the regression setup with covariates $\boldsymbol{x}_{2,i}$, therefore, we have two candidate regression models:
The Gamma regression model defined as

$$Y_{i,j} \sim \mathcal{G}(\mu_i, \upsilon) \qquad \text{with} \quad \log(\mu_i) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_{2,i} \tag{7}$$

with mean μ_i and variance μ_i^2/υ, and the inverse Gaussian regression model defined as

$$Y_{i,j} \sim \mathcal{IG}(\mu_i, \rho) \qquad \text{with} \quad \log(\mu_i) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_{2,i} \tag{8}$$

with mean μ_i and variance μ_i^3/ρ. Here, we adopt a typical assumption of the severity model, where $Y_{i,1}, \ldots, Y_{i,N_i}$ are independent and identically distributed, and independent of N_i. In order to account for possible dependence between N_i and $Y_{i,j}$, we also consider an extended version of these two models in which the dependence is modeled through conditioning. That is, we add an extra covariate in (7) and (8) so that the claim size can respond to the number of claims. In this case, conditional on $N_i = n_i > 0$, the severity models become

$$\begin{aligned} &M_i \big| n_i \sim \mathcal{G}(\mu_i, n_i\upsilon) \quad \text{or} \quad M_i \big| n_i \sim \mathcal{IG}(\mu_i, n_i\rho) \\ &\text{with} \quad \log(\mu_i) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_{2,i} + \beta^* \zeta(n_i), \end{aligned} \tag{9}$$

where $\zeta(n_i)$ is a function of the observed frequency and β^* is its associated coefficient. For simplicity, we will use the identity function for ζ. Frees et al. (2014), Shi et al. (2015), and Garrido et al. (2016) use such specification to detect and model the dependence between frequency and severity.

As mentioned in the introduction, recent studies report that the dependence between the claim frequency and severity is weakly positive for liability insurance coverage and negative for collision coverage. Although we will attempt to explain this seemingly contradicting phenomenon in more detail later, it suffices to say at this point that it is important and necessary for a frequency-severity model to be able to capture different dependence levels. In this regard, we note that the sign and magnitude of β^* in (9) can modulate the dependence between claim frequency and severity.

3. Review: bonus–malus system

A BMS or merit rating system is a premium adjustment mechanism widely used in the a posteriori ratemaking process to set the premium for the next time period based on the previous claim history of a driver. Apart from peripheral differences, all BMSs have the following common features. First, a BMS assigns a driver to one of predetermined levels (or classes) depending on the driver's past claim history. Thus, each level represents a group of drivers with homogeneous risk profiles, and the drivers in the same class pay identical premiums. For example, a BMS with $s + 1$ levels would have levels labeled $0, 1, \ldots, s$, where level 0 represents the best driver group and level s stands for the worst group. Second, the premium is directly linked to the level in the BMS so that good drivers in lower levels pay smaller premiums and bad drivers in higher levels are charged larger premiums. The premium differential is determined based on level-specific multipliers r_l associated with each level $l \in \{0, \ldots, s\}$; a driver in level l has to pay r_l% of the

[2] Since we use a two-part model to derive the dependence between the frequency and severity, the choice of the frequency distribution is not critical to the dependence structure.

base premium to be covered by the insurer. Clearly, r_l, known as the premium relativity, becomes larger as the level increases. Third, as the claim experience evolves over time, a driver's level migrates from one level to another at each policy renewal. The insurer grants bonuses (premium discounts) to drivers who reported no claims in the previous year by lowering the level and maluses (premium increases) to drivers with many claims by elevating the level. Because of this structural design, a BMS can induce policyholders to refrain from filing small claims, as the premium increase may exceed the indemnity payment from the insurer. This phenomenon is known as *bonus hunger* (Philipson, 1960). In this section, we set up a standard vanilla BMS, determine the optimal claim strategy for policyholders, and describe the bonus-hunger behavior.

In a typical BMS with $s+1$ classes, a claim-free year is rewarded by dropping one level, and each claim is penalized by climbing h levels per claim, commonly known as the $-1/+h$ system. This system can be mathematically described by a quantity related to a transition of a driver from one level to another. Suppose that a driver is currently in level $l \in \{0, \dots, s\}$. If the number of accidents reported is z, the new level at the next policy renewal based on this year's claim history, denoted as $T_z(l)$, becomes

$$T_z(l) = \begin{cases} \max\{l-1, 0\}, & z = 0; \\ \min\{l+zh, s\}, & z > 0. \end{cases} \tag{10}$$

Now, regarding the premium relativity r_l, we formulate a Bayesian-type relativity model following the work of Norberg (1976). In this model, Λ is the (random) a priori expected claim frequency of a driver independent of Θ which is the residual effect that is not included in the a priori ratemaking. The true, unknown expected claim frequency of the driver is then $\Lambda\Theta$. It is assumed that Λ is a discrete random variable with possible realizations λ_k and probability

$$\mathbb{P}(\Lambda = \lambda_k) = w_k,$$

where w_k is the weight of the kth risk class, whose annual claim frequency is λ_k. If we denote L as the level occupied by the driver once the steady state has been reached, the standard theory says that r_l is shown to be the conditional mean

$$r_l = \mathbb{E}\left[\Theta \middle| L = l\right] \tag{11}$$

under the criterion of the expected squared difference between the true relative premium Θ and the relative premium r_l; see Norberg (1976) and Denuit et al. (2007) for more details. Variations of r_l under different circumstances can be found in Pitrebois et al. (2005), Tan et al. (2015), Tan (2016), for example.

3.1. Bonus hunger

Throughout the analysis, we assume that a driver can calculate an optimal retention level and make a rational decision in the sense that she reports a claim to the insurer only when the loss amount exceeds some optimal retention level. Thus, it is implicitly assumed that the insured has complete freedom regarding whether or not to report an accident, which may not be possible for some coverages involving a third party. The optimal retention level is determined by comparing two alternative present values (PVs): (1) the loss PV coming from the sum of the concurrent premium increases in the subsequent years assuming that the loss is claimed to the insurer, and (2) the loss PV borne by the driver assuming that the accident is not filed. The optimal retention level, then, is the loss amount that is indifferent between the two alternatives; see Lemaire (1977), Denuit et al. (2007), and Lemaire (2012) for further details.

Drawing on the notation of Denuit et al. (2007), we now briefly review how bonus hunger can be formulated. We assume that the original ground-up loss frequency and severity are independent in our discussion. Our goal is to determine the optimal retention level of a driver in level l with expected accident frequency λ, denoted by $\eta(l, \lambda)$. We start with the probability that the driver does not report her loss with the retention level $\eta(l, \lambda)$:

$$p_l(\lambda) := \int_0^{\eta(l,\lambda)} f(y; \mu, \upsilon)\mathrm{d}y, \tag{12}$$

where $f(y; \mu, \upsilon)$ is the severity density following a given distribution. The expected severity of the non-reported accident is then calculated as

$$\mu_l(\lambda) = \frac{1}{p_l(\lambda)} \int_0^{\eta(l,\lambda)} y f(y; \mu, \upsilon)\mathrm{d}y. \tag{13}$$

Next, the probability that this driver files z claims to the insurer during the year is given by

$$q_l(z|\lambda) = \sum_{y=z}^{\infty} \mathbb{P}(N = y) \binom{y}{z} (1 - p_l(\lambda))^z (p_l(\lambda))^{y-z}, \quad z = 0, 1, 2, \dots,$$

the mean of which then gives the expected frequency of reported accidents

$$\lambda_l^* = \sum_{z=0}^{\infty} z\, q_l(z|\lambda).$$

If we assume that the occurrence of accidents is uniformly distributed during the policy year, the PV of the average annual total cost (TC) of the driver at the onset of the policy year ($t = 0$) is given by

$$TC_0(\eta(l, \lambda), \lambda) = \begin{cases} b(l) + \upsilon^{1/2}\mu_l(\lambda)\left(\lambda - \lambda_l^*\right), & \eta(l, \lambda) > 0; \\ b(l), & \eta(l, \lambda) = 0, \end{cases} \tag{14}$$

where υ is a discount factor and $b(l)$ is the premium paid by a policyholder at BM level l at the beginning of the policy year, given by r_l% of the base premium. Note that when $\eta(l, \lambda) = 0$, the TC reduces to $b(l)$ as $\mu_l(\lambda) = 0$, implying that when a driver in level l reports every loss regardless of its size, the cost for the driver is simply the premium in that level l. Here, $\mu_l(\lambda)\left(\lambda - \lambda_l^*\right)$ is interpreted as the average annual cost caused by non-reported accidents. If we further include losses arising from level migrations in the next year due to reported claims, the PV of all concurrent payments of a driver at level l with expected accident frequency λ is recursively defined as:

$$V_l(\lambda) = TC_0(\eta(l, \lambda), \lambda) + \upsilon \sum_{z=0}^{\infty} q_l(z|\lambda) V_{T_z(l)}(\lambda). \tag{15}$$

Here, both (14) and (15) are PVs of future payments at the beginning of the year $t = 0$. We note that the PV, evaluated at time $t \in (0, 1]$, of the average total cost in the remaining of the first year $[t, 1]$ is obtained as

$$TC_t(\eta(l, \lambda), \lambda) = \begin{cases} \upsilon^{(1-t)/2}\mu_l(\lambda)\left(\lambda - \lambda_l^*\right)(1-t), & \eta(l, \lambda) > 0; \\ 0, & \eta(l, \lambda) = 0. \end{cases}$$

Now, in order to derive the optimal retention level, suppose that a driver in level l just caused an accident of loss size x at time $t \in [0, 1]$ and that g accidents were already reported to the insurer during time $[0, t]$. If the driver reports the new loss, her expected PV of payments in the subsequent years at time t is

$$TC_t(\eta(l, \lambda), \lambda) + \upsilon^{1-t} \sum_{z=0}^{\infty} q_l(z|\lambda(1-t)) V_{T_{g+z+1}(l)}(\lambda). \tag{16}$$

Note that the claim frequency inside q_l has been adjusted proportionally to reflect the remaining time until the next policy renewal.

If she decides not to report, the expected PV of the payments in the subsequent years at t plus the loss size x becomes

$$TC_t\left(\eta\left(l,\lambda\right),\lambda\right)+x+v^{1-t}\sum_{z=0}^{\infty}q_l(z\left|\lambda(1-t)\right)V_{T_{g+z}(l)}(\lambda). \tag{17}$$

Since the retention level $\eta\left(l,\lambda\right)$ is the claim severity x^* at which the driver is indifferent between the two scenarios in (16) and (17), we equate these two expressions to yield

$$x^*=\eta\left(l,\lambda\right)=v^{1-t}\sum_{z=0}^{\infty}q_l(z\left|\lambda(1-t)\right)\left[V_{T_{g+z+1}(l)}(\lambda)-V_{T_{g+z}(l)}(\lambda)\right]. \tag{18}$$

Whereas $\eta\left(l,\lambda\right)$ is a function of t and g, we fix time t following the standard practice so that we can fairly compare the impact of the number of already reported accidents, denoted by g. In the literature, the impact of different values of t on the optimal retention is known to be minor compared to those of other variables (Denuit et al., 2007), and different choices are used in other studies. Among possible $t\in[0,1]$, we choose $t=1$, which effectively implies that the policyholder believes that the latest accident would be the last one in the remaining year. This assumption is reasonable for most drivers. In addition, this choice of t substantially simplifies (18) to yield the following intuitive expression for the retention level

$$\eta\left(l,\lambda\right)=V_{T_{g+1}(l)}(\lambda)-V_{T_g(l)}(\lambda). \tag{19}$$

On the other hand Denuit et al. (2007) choose $t=0$ to reduce the amount of computation. We note that the assumption $t=1$ is used only for the calculation of the optimal retention, and our argument in the remaining analysis is valid with the optimal retention $\eta(l,\lambda)$ in (18) under any $t\in[0,1]$ as long as it is fixed upfront.

3.2. Calculation of optimal retention using the Lemaire algorithm

As there is no closed-form solution for $\eta\left(l,\lambda\right)$ in (19) due to its highly non-linear recursive form, we adopt the Lemaire algorithm (Lemaire, 1977; Denuit et al., 2007; Lemaire, 2012), which gives the solution to (19) using iterations. For a policyholder in class l with fixed relativity r_l, the algorithm computes the optimal retention level $\eta\left(l,\lambda\right)$ using the following steps:

i. For initial values of retention $\eta\left(l,\lambda\right)$, set

$$\eta^{[0]}(l,\lambda)=0,\quad \text{for}\quad l=0,\ldots,s.$$

Since $\eta\left(l,\lambda\right)=0$ implies that, from (14),

$$TC_0\left(\eta\left(l,\lambda\right),\lambda\right)=b(l),$$

we have, according to (15),

$$V_l(\lambda)=b(l)+v\sum_{z=0}^{\infty}q_l(z\left|\lambda\right)V_{T_z(l)}(\lambda),\quad \text{for}\quad l=0,1,\ldots,s. \tag{20}$$

We can solve this equation system using the rule of $T_z(l)$ in (10) and obtain $V_0(\lambda),\ldots,V_s(\lambda)$. This result in turn gives $\eta\left(l,\lambda\right)$ from (19) for $l=0,\ldots,s$, which is set to be the first iterated values $\eta^{[1]}(l,\lambda)$.

ii. For a given $\eta\left(l,\lambda\right)=\eta^{[j]}(l,\lambda), j\geq 1$, we calculate

$$TC_0\left(\eta\left(l,\lambda\right),\lambda\right) \tag{21}$$

using (14). Then, (15) results in an equation system whose solution gives $V_l(\lambda)$ for $l=0,1,\ldots,s$, and (19) provides the next iterated value $\eta^{[j+1]}(l,\lambda)$.

iii. Repeat (ii) from above to produce the sequence of $''^{[j]}(l,\lambda)$, which converges to the optimal retention $\eta\left(l,\lambda\right)$.

Since $\eta\left(l,\lambda\right)$ in (19) also depends on g, the number of accidents, we will sometimes use $\eta_g\left(l,\lambda\right)$ to emphasize this aspect.

Table 1
Information on covariates.

Proportion		Factor 1	
		$x_{1,i}=0$	$x_{1,i}=1$
Factor 2	$x_{2,i}=0$	0.02	0.04
	$x_{2,i}=1$	0.23	0.71

4. A synthetic BMS and its analysis

In this section, we explain how bonus hunger can induce dependence between the reported claim frequency and severity, even when the original ground-up loss frequency and severity are independent. To this end, we construct a synthetic BMS in which the drivers retain small losses if the loss size is less than the optimal retention depending on their current BM levels; this system is idealistic in that all drivers make sequential claim reporting decisions based on the optimal retention level computed by the Lemaire algorithm in the previous section. Under this BM system, we produce optimal retention tables where the optimal retention levels are shown as a function of the current BM level l and the number of already reported accidents g. The produced tables not only demonstrate how the optimal retention induces the dependence between the reported claim frequency and severity but, more importantly, also prove that the induced dependence does vary by BM level rather than remaining constant. Later in this section, we analyze the simulated dataset with a regression approach to support our claim.

4.1. Setup for ground-up loss frequency and severity

The parameters of the frequency and severity models in this subsection are motivated by the real dataset in Section 5 for an easier comparison. However, we choose simple, arbitrary values for covariates, as they can vastly vary in theory over different datasets. Referring to Section 2, we model the original ground-up loss frequency for the ith driver with a Poisson model, $N_i\sim\mathcal{P}(\lambda_i\Theta)$, where $\log(\lambda_i)$ is described by a regression with two covariate factors $x_{1,i},\ x_{2,i}$ so that

$$\lambda_i=\exp\left(\alpha_0+\alpha_1x_{1,i}+\alpha_2x_{2,i}\right), \tag{22}$$

where $x_{1,i},\ x_{2,i}\in\{0,1\}$ are binary covariates with the proportion of each pair specified in Table 1 and the coefficients are set as $(\alpha_0,\alpha_1,\alpha_2)=(-0.5,-0.15,-0.16)$. The residual effect is set to be $\Theta\ \sim\ \mathcal{G}(1,0.5)$. We also assume that the individual severity $Y_{i,j}\overset{\text{i.i.d.}}{\sim}\mathcal{G}(\xi,v)$, where $\xi\ =\ \exp(14)/1000$ and $v\ =\ 1/2.5$. It is noted that we assume the independence of the frequency and severity of the ground-up loss. Thus, the mean of the aggregated ground-up loss of the ith driver is calculated as the product of the mean frequency and mean severity, $\lambda_i\xi$. From this specification, we simulate losses for n drivers, each of which has a history of ground-up losses of

$$N_i \text{ and } \mathbf{Y}_i=(Y_{i,1},\ldots,Y_{i,N_i})^T \text{ for } i=1,\ldots,n.$$

The key element of this construction is that N_i and $\mathbf{Y}_i$ are independent. Note that not all of these generated losses are reported, as each driver makes claim decisions based on the optimal retention level.

4.2. Optimal retention table

In addition to the frequency-severity model specification, we also construct a BM system where each driver is populated. In terms of the premium relativity r_l, we consider two distinct BM relativity types, as shown in Table 2, to assess their impacts on

Table 2
Two versions of BM relativities, r_l and r_l^{ed}, in units of 0.01.

Level (l)	0	1	2	3	4	5	6	7	8	9	10
r_l	40	42	45	48	51	54	58	61	65	69	74
r_l^{ed}	40	48	55	63	70	78	86	93	101	109	116
Level (l)	11	12	13	14	15	16	17	18	19	20	21
r_l	83	90	95	100	108	115	123	135	150	170	200
r_l^{ed}	124	131	139	147	154	162	170	177	185	192	200

Table 3
Optimal retentions using the relativity in Table 2 in units of 10,000 won.

	$\eta_g(l, \lambda_1)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	13	25	36	45	54	61	69	76	84	95	106
g=1	28	34	39	44	48	53	59	68	76	80	82
g=2	39	44	49	54	60	70	79	84	88	98	114
g=3	54	57	60	71	83	94	109	155	172	142	122
g=4	66	82	104	151	168	140	121	107	99	96	92
l	11	12	13	14	15	16	17	18	19	20	21
g=0	115	120	126	133	142	152	166	176	187	126	65
g=1	88	94	102	115	146	152	129	63	0	0	0
g=2	155	168	141	121	110	55	0	0	0	0	0
g=3	111	102	99	51	0	0	0	0	0	0	0
g=4	93	49	0	0	0	0	0	0	0	0	0

the optimal retention level. Both relativity types have 22 levels (i.e., $s = 21$), but their scales are different. The first set in the table, r_l, reflects the actual BM relativities used by a major Korean insurer. Note that the difference $r_{l+1} - r_l$ increases as the BM level l increases. The other set in the table, r_l^{ed}, is the equal distance version of r_l. The base premium for the ith driver is defined as $b(l) = cr_l$, where $c = \lambda_i \xi$ is used as opposed to other choices for convenience. Note that our argument in this section is valid with other choices of c.

To generate level transitions in this BMS for a driver, we assume a $-1/+2$ system with an annual effective interest rate of 10%, or $v = 1/1.1$. Based on the Lemaire algorithm, the optimal retention levels $\eta(l, \lambda)$ are computed for $l = 0, \ldots, 21$ for r_l and r_l^{ed} as given in Table 2. For example, the optimal retentions calculated based on r_l in Table 2 for a driver with $\lambda = \lambda_1$ are presented in Table 3.

The full optimal retentions are provided in Tables A.10 and A.11, respectively, from which the following important observations are made:

i. $\eta_g(l, \lambda)$ is an increasing function of g at lower BM levels (shown as the gray area). In other words, at lower BM levels, a driver's retention level increases as she reports more claims.

ii. $\eta_g(l, \lambda)$ is a decreasing function of g at higher BM levels (shown as the white area). This result means that, at higher BM levels, the retention level goes down as a driver reports more claims.

As our rational policyholders only report losses greater than the optimal retention level, higher value of $\eta_g(l, \lambda)$ will result in larger claimed loss amount. These observations therefore imply that at lower BM levels, the claim size will get larger as more claims are reported. For example, using the numbers in Table 3, a driver in level 0 with $g = 1$ will report the second loss as long as it is greater than 28, but another driver in the same level with $g = 4$ would report the fifth loss only when it is greater than 66. Hence, this progressively increasing retention scheme essentially suggests a positive dependence between the reported claim frequency and the (average) reported claim severity for lower BM levels. In contrast, at higher BM levels, the optimal retention $\eta_g(l, \lambda)$ decreases as more claims get reported. Since more small losses are getting reported at these higher level as number of claims increases, this will lead to a negative dependence between claim frequency and the average severity.

These two patterns gradually move toward each other as a driver migrates among the lower and higher BM levels. We formally articulate our findings in the following claim, which will be supported through the subsequent subsections.

Claim 1. *Under the BMS in* Section 3*, due to the bonus-hunger phenomenon, the dependence between claim frequency and claim severity gradually shifts from positive to negative as BM levels increase.*

In the following subsection, we offer an analytical expression of the optimal retention under a simplified BMS structure. Then, returning to the original BMS, we provide a heuristic intuition for Claim 1.

4.3. Analytic result for the infinite BM level case

Even though the optimal retention level is in general determined by complicated numerical iterations using the Lemaire algorithm in Section 3.2, it is possible to derive a closed-form version for a simple BMS structure. For this analytic derivation, we make the following assumptions:

A1. The BMS has a $-1/+2$ system for all drivers.

A2. The set of possible BM levels is given by $\mathcal{C}_\infty = \{\ldots, -2, -1, 0, 1, 2, \ldots,\}$. Thus, the BM scale has infinitely many levels with no upper or lower limit.

A3. The premium relativity scales are linear or equally distant, that is, $b(l+1) - b(l) = c$ for all $l \in \mathcal{C}_\infty$ with $b(0) = b_0 \in \mathbb{R}$.

A4. Drivers retain small losses based on their optimal retention levels.

We first note that, under the unrestricted class structure $\mathcal{C}_\infty$, the optimal retention $\eta_g(l, \lambda)$ must be a constant for any l, based on the following rationale. Due to the equal-distant BM scales, a driver would obtain the same amount of benefit from forgoing reporting the loss and from shifting down one level regardless of her current level l. Likewise, as the driver has the fixed loss frequency and severity model, her expected PV of future ground-up losses always remains the same. This property means that the driver will take a decision on the retention level based on the same cost and benefit regardless of her current level l. Therefore, the optimal retention level must be the same across all of the levels under $\mathcal{C}_\infty$, that is, $\eta_g(l_2, \lambda) = \eta_g(l_1, \lambda)$ for any $l_1, l_2 \in \mathcal{C}_\infty$ and any g. Its specific form is established below.

Proposition 1. *Under assumptions A1–A4, for* $0 < v < 1$, *the optimal retention level given by*

$$\eta_g(l, \lambda) = \begin{cases} \dfrac{3c}{1-v}, & g = 0 \\ \dfrac{2c}{1-v}, & g \geq 1 \end{cases} \tag{23}$$

is a solution to Eq. (19).

Proof. We prove this result by plugging (23) into the right side of (19) and showing that the result is the same as the left side, $\eta_g(l,\lambda)$. We start by iterating (15) for future years. By denoting the number of accidents reported in the next two subsequent years by z_1 and z_2, respectively, we have the iterated version for (15) as follows.

$$
\begin{aligned}
V_l(\lambda) &= TC_0(\eta(l,\lambda),\lambda) + v\sum_{z_1=0}^{\infty} q_l(z_1|\lambda) \\
&\quad\times\left[TC_0\left(\eta\left(T_{z_1}(l),\lambda\right),\lambda\right) + v\sum_{z_2=0}^{\infty} q_{T_{z_1}(l)}(z_2|\lambda)V_{T_{z_2}\circ T_{z_1}(l)}(\lambda)\right] \\
&= TC_0(\eta(l,\lambda),\lambda) + v\sum_{z_1=0}^{\infty} q_l(z_1|\lambda)TC_0\left(\eta\left(T_{z_1}(l),\lambda\right)\right) \\
&\quad + v^2\sum_{z_1=0}^{\infty}\sum_{z_2=0}^{\infty} q_l(z_1|\lambda)q_{T_{z_1}(l)}(z_2|\lambda)V_{T_{z_2}\circ T_{z_1}(l)}(\lambda) \\
&= TC_0(\eta(l,\lambda),\lambda) + v\sum_{z_1=0}^{\infty} q_l(z_1|\lambda)TC_0\left(\eta\left(T_{z_1}(l),\lambda\right)\right) \\
&\quad + v^2\sum_{z_1=0}^{\infty}\sum_{z_2=0}^{\infty} q_l(z_1|\lambda)q_{T_{z_1}(l)}(z_2|\lambda)TC_0 \\
&\quad \times\left(\eta\left(T_{z_2}\circ T_{z_1}(l),\lambda\right),\lambda\right) + \cdots.
\end{aligned}
\tag{24}
$$

Note that, from

$$TC_0\left(\eta\left(l,\lambda\right),\lambda\right) \le b(l) + \lambda\mu$$

and (24), one can show that $V_l(\lambda) < \infty$. Similarly, we have $-\infty < V_l(\lambda)$, which in turn implies that

$$|V_l(\lambda)| < \infty.$$

Then, applying (14) and (24) to the right-hand side of (19), we have

$$
\begin{aligned}
V_{T_{g+1}(l)}(\lambda) - V_{T_g(l)}(\lambda) &= b(T_{g+1}(l)) - b(T_g(l)) \\
&\quad + v\sum_{z_1=0}^{\infty} q_l(z_1|\lambda)\left[b(T_{z_1}\circ T_{g+1}(l)) - b(T_{z_1}\circ T_g(l))\right] \\
&\quad + v^2\sum_{z_1=0}^{\infty}\sum_{z_2=0}^{\infty} q_l(z_1|\lambda)q_l(z_2|\lambda)\left[b(T_{z_2}\circ T_{z_1}\circ T_{g+1}(l))\right. \\
&\quad \left. - b(T_{z_2}\circ T_{z_1}\circ T_g(l))\right] + \cdots.
\end{aligned}
\tag{25}
$$

Here, the equality holds because $v^{1/2}\mu_{l+1}(\lambda)\left(\lambda - \lambda^*_{l+1}\right) = v^{1/2}\mu_l(\lambda)\left(\lambda - \lambda^*_l\right)$, implied by the fact that the optimal retention is independent of l, as shown by (23). We also note that, under $\mathcal{C}_\infty$,

$$b(\,T_{z_k}\circ\cdots\circ T_{z_1}\circ T_{g+1}(l)\,) - b(\,T_{z_k}\circ\cdots\circ T_{z_1}\circ T_g(l)\,) > 0, \quad k = 1,2,\ldots$$

for all g. To further simplify (25), we consider two separate cases. First, when $g \ge 1$, the level difference between $T_{g+1}(l)$ and $T_g(l)$ is always two, and, thus, $b(T_{g+1}(l)) - b(T_g(l)) = 2c$. Similarly, we have that $b(T_{z_1}\circ T_{g+1}(l)) - b(T_{z_1}\circ T_g(l)) = 2c$ and that $b(T_{z_2}\circ T_{z_1}\circ T_{g+1}(l)) - b(T_{z_2}\circ T_{z_1}\circ T_g(l)) = 2c$; the same holds for all remaining future years. Hence, (25) becomes

$$
\begin{aligned}
V_{T_{g+1}(l)}(\lambda) - V_{T_g(l)}(\lambda) &= 2c + v\sum_{z_1=0}^{\infty} q(z_1|\lambda)2c \\
&\quad + v^2\sum_{z_1=0}^{\infty}\sum_{z_2=0}^{\infty} q(z_1|\lambda)q(z_2|\lambda)2c + \cdots \\
&= 2c + 2cv + 2cv^2 + 2cv^3 + \cdots \\
&= \frac{2c}{1-v}, \qquad g \ge 1.
\end{aligned}
\tag{26}
$$

On the other hand, when $g = 0$, the level difference between $T_1(l)$ and $T_0(l)$ becomes three, and, thus, $b(T_1(l)) - b(T_0(l)) = 3c$. Following the same logic, we simplify (25) as

$$
\begin{aligned}
V_{T_{g+1}(l)}(\lambda) - V_{T_g(l)}(\lambda) &= 3c + 3cv + 3cv^2 + 3cv^3 + \cdots \\
&= \frac{3c}{1-v}, \qquad g = 0,
\end{aligned}
\tag{27}
$$

which is the same across levels. Since $V_{T_{g+1}(l)}(\lambda) - V_{T_g(l)}(\lambda)$ in (26) and (27) coincides with $\eta_g(l,\lambda)$ presented in (23), we conclude that $\eta_g(l,\lambda)$ in (23) is a solution to (19). □

Proposition 1 implies that, under the assumptions A1–A4, the quantities

$$p_l(\lambda), \quad \mu_l(\lambda), \quad q_l(z|\lambda), \quad \text{and} \quad \lambda^*_l \tag{28}$$

defined in Section 3.1 are all independent of the BM levels.

Remark 1. For the simplicity of the argument, this study disregards the deductible, a common insurance instrument to mitigate the moral hazard and decrease the premium. Knowing that a deductible induces the truncated distribution of the ground-up loss variable, Proposition 1 can be applied into the BM system as long as the deductible is kept constant across the BM levels. In general, however, the deductible amount opted by a policyholder can depend on many factors such as adverse selection, advantageous selection and selection due to moral hazard. Thus, it is possible in practice for a policyholder in different BM levels or retention thresholds to select different deductibles in the middle of the policy year. However, as the interaction of these factors and deductibles are highly complex and difficult to quantify, incorporating varying deductibles is beyond the scope of our present study. In the same spirit, later we limit our empirical analysis with real data to the policies with a constant deductible throughout policy year in order to examine pure bonus-hunger effect on the dependency.

4.4. Heuristic investigation of the optimal retention for the finite BM level case

We now turn to the standard class structure $\mathcal{C} = \{0,\ldots,s\}$, under which the analysis becomes more complicated. Although the argument stated below is therefore only heuristic, it provides insight into the fundamental mechanism behind Claim 1. First, consider a driver currently at a high BM level l_{high} that is very close to s. Compared to a driver with a lower BM level, it is more likely that this driver's future level transition ends up at the maximum level s. That is, since $l_{\text{high}} \approx s$, it is more probable, relative to drivers with lower BM levels, that

$$T_{z_k}\circ\cdots\circ T_{z_1}\circ T_g(l_{\text{high}}) = s, \qquad k = 1,2,\ldots,$$

so that

$$b(\,T_{z_k}\circ\cdots\circ T_{z_1}\circ T_g(l_{\text{high}})\,) = b(s), \qquad k = 1,2,\ldots,$$

and, consequently,

$$
\begin{aligned}
&b(\,T_{z_k}\circ\cdots\circ T_{z_1}\circ T_{g+1}(l_{\text{high}})\,) - b(\,T_{z_k}\circ\cdots\circ T_{z_1}\circ T_g(l_{\text{high}})\,) \\
&= b(s) - b(s) = 0, \quad k = 1,2,\ldots.
\end{aligned}
$$

Therefore, more terms on the right side of Eq. (25) tend to be zeroed out. If g increases for this driver, the driver's tendency to fall into level s in the future also grows, because the next year's level $T_g(l_{\text{high}})$ will be closer to s. Hence, for a driver with l_{high}, increasing g leads to even more zero terms in (25), making the retention $\eta_g(l_{\text{high}},\lambda)$ a non-increasing function of g assuming that the quantities in (28) do not depend on l.

Now, consider a driver currently at a low BM level l_{low} close to 0. Compared to a driver at a higher BM level, this driver's future level transition is more likely to end up at the minimum level 0. That is, since $l_{\text{low}} \approx 0$, it is more probable, compared to drivers in higher BM levels, that

$$T_{z_k} \circ \cdots \circ T_{z_1} \circ T_g(l_{\text{low}}) = 0, \qquad k = 1, 2, \ldots$$

so that

$$b(\, T_{z_k} \circ \cdots \circ T_{z_1} \circ T_g(l_{\text{low}})\,) = b(0), \qquad k = 1, 2, \ldots$$

and

$$b(\, T_{z_k} \circ \cdots \circ T_{z_1} \circ T_{g+1}(l_{\text{low}})\,) - b(\, T_{z_k} \circ \cdots \circ T_{z_1} \circ T_g(l_{\text{low}})\,)$$
$$= b(0) - b(0) = 0, \quad k = 1, 2, \ldots .$$

Therefore, again, more terms on the right side of Eq. (25) tend to become zero. However, in this case, if g increases for this driver, the driver's tendency to fall into level 0 in the future decreases because the next year's level $T_g(l_{\text{low}})$ will move farther away from 0. Hence, for a driver with l_{low}, increasing g leads to fewer zero terms in (25), making $\eta_g(l_{\text{low}}, \lambda)$ an increasing function of g. Hence, we expect that as the driver's current BM level moves from 0 to s, the optimal retention gradually shifts from an increasing function of g to a decreasing one, again, assuming that the quantities in (28) does not depend on l.

Under class structure $\mathcal{C}$, however, the quantities in (28) do depend on l, and $TC_0(\cdot, \lambda)$ term in (24) does not vanish when we subtract $V_{T_g(l)}(\lambda)$ from $V_{T_{g+1}(l)}(\lambda)$. Furthermore, the equi-distant premium relativity in the assumption A3 is rarely satisfied in practice. Therefore, investigating the behavior of $\eta_g(l, \lambda)$ intuitively with respect to g in actual BMS is much more complicated than the argument presented above. In order to approach this issue, we may rely on numerically produced values, such as those in Tables A.10 and A.11, to understand the relationship between $\eta_g(l, \lambda)$ and g to some degree. However, these tables do not directly manifest how BM levels affect the dependence between the claim frequency and severity or how strong the dependence is for different BM levels. In the next subsection, via a simulation study, we quantify the changes in the dependence between reported frequencies and severities across different BM levels. Before closing this subsection, we make the following observation, which might not be directly linked to the main topic of the paper but has its own interest as a possible independent research topic, suggested by the reviewer.

Remark 2. Using the similar heuristical logic used this subsection, we expect that given fixed g, $\eta_g(l, \lambda)$ is an increasing function of l for the lower BM levels close to 0 and a decreasing function of l for the upper BM levels close to s. This can be also numerically checked from Tables A.10 and A.11. An intuitive explanation is that the current year's BM level l is the result of the last year's g value, and therefore, the behavioral pattern of $\eta_g(l, \lambda)$ over l resembles that over g, as observed in Section 4.2.

4.5. Analysis of simulated loss: Emulating the dependence of the reported frequency and severity

So far in this section, we have specified the ground-up loss model and the BM system. We also produced the tables for optimal retention levels and explained their patterns over the BM scales and g. Furthermore, Section 4.4 heuristically shows that the dependence between the reported claim frequency and severity gradually changes from positive to negative as the BM level increases from 0 to s. In this subsection, we adopt the frequency-severity model in Section 4.1 to validate and quantify the claim frequency-severity dependence across different BM levels using simulated claims. This simulation study can provide an alternative way to identify and measure the dependence, and can be used as a benchmark in interpreting the result if applied to actual datasets.

Table 4
Illustration of the bonus-hunger phenomenon in units of 10,000 won.

i	y_i	g prior to ith accident	$\eta_g\,(3, \lambda_3)$	report?
1	180	0	129	Yes
2	13	1	49	No
3	317	1	49	Yes
4	34	2	0	Yes
5	9	3	0	Yes

We start by filtering the ground-up losses generated in Section 4.1 into the reported losses for each driver according to the optimal retention rule. As before, the ground-up losses are denoted by N_i and $\boldsymbol{Y}_i$. Now, let us denote the frequency of the *reported* claims by N_i^* with $N_i^* \le N_i$ and the corresponding severity vector as

$$\boldsymbol{Y}_i^* = \left(Y_{i,j_1}, \ldots, Y_{i,j_{N_i^*}}\right)^T. \tag{29}$$

The average reported claim severity is then

$$M_i^* := \frac{\sum_{z=1}^{N_i^*} Y_{i,j_z}}{N_i^*}, \quad N_i^* > 0,$$

where M_i^* is not defined when $N_i^* = 0$, which is parallel to the notation in Section 2. As the retention depends on the number of claims already reported, g, it is informative to illustrate how a driver makes a series of reporting decisions through a simple example.

Example 1. Assume that the settings in Sections 4.1 and 4.2 hold. Consider a driver at the current BM level $l = 18$ with covariates $x_1 = x_2 = 1$, so that the driver's mean loss frequency is λ_4. From Table A.10, the driver's optimal retention levels are given as

$$(\eta_0\,(18, \lambda_4)\,, \eta_1\,(18, \lambda_4)\,, \eta_2\,(18, \lambda_4)\,, \eta_3\,(18, \lambda_4)\,,$$
$$\eta_4\,(18, \lambda_4)) = (129, 49, 0, 0, 0).$$

Now, suppose that she has five accidents in the year with the following severities in a chronologically ordered sequence:

$$(y_{i,1}, \ldots, y_{i,5}) = (180, 13, 317, 34, 9).$$

At the first loss $y_{i,1} = 180$, the number of already reported accidents g is zero. She then compares 180 against her optimal retention $\eta_0\,(18, \lambda_4) = 129$. Since $180 > \eta_0\,(18, \lambda_4)$, she reports the accident to the insurer, and g becomes one. When the second accident occurs, she compares $y_{i,2} = 13$ with $\eta_g\,(18, \lambda_4) = 49$. Since $y_{i,2} < \eta_g\,(18, \lambda_4)$, she will not report this loss, and g remains one. Next, the third loss $y_{i,3} = 317$ should be compared with $\eta_1\,(18, \lambda_4) = 49$. Since $y_{i,3} > \eta_1\,(18, \lambda_4)$, she reports this claim, and g becomes two. Similarly, the fourth accident $y_{i,4} = 34$ is reported because $y_{i,4} > \eta_2\,(18, \lambda_4) = 0$, and g becomes three. The last loss is also reported, as $y_{i,5} = 9 > 0 = \eta_3\,(18, \lambda_4)$. After completing this process, her claim history will be

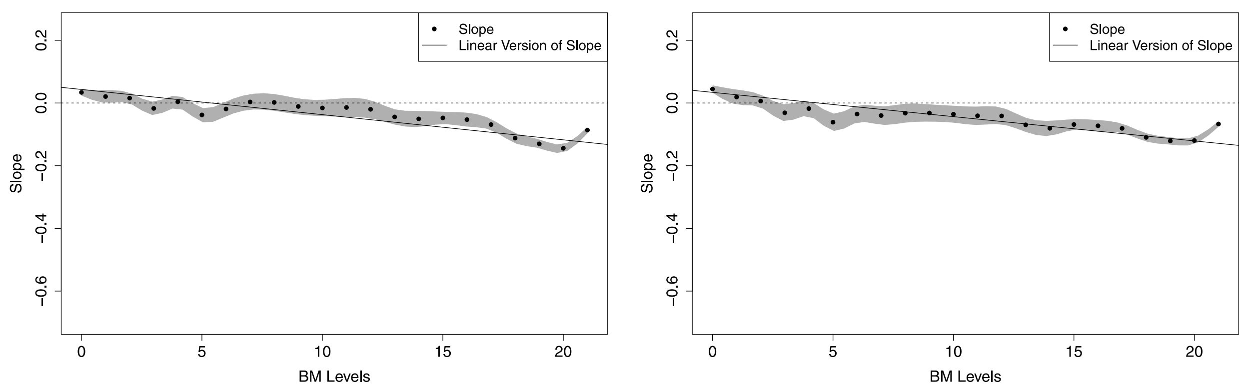
$$n_i^* = 4 \quad \text{and} \quad \boldsymbol{y}_i^{*\text{T}} = (180, 317, 34, 9),$$

and the average severity will be $m_i^* = (180 + 317 + 34 + 9)/4 = 135$. All of the losses in this example are in units of 10,000 won, and the process is summarized in Table 4. □

To assess the effect of the distribution of premium relativities on the optimal retention, we use a BMS with the premium relativities r_l and r_l^{ed} given in Table 2. We simulate 1,000,000 drivers' losses according to Section 4.1 and let individual drivers go through the sequential reporting decisions based on their optimal retention levels $\eta_g\,(l_i, \lambda_i)$, as explained in the previous example. Drivers are assumed to be distributed across each BM level as in Table 5, which

Table 5
Steady state distribution of policyholders.

l	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}\left(L=l \mid x_{1,i}=0, x_{2,i}=0\right)$	0.025	0.010	0.014	0.014	0.017	0.018	0.021	0.023	0.026	0.029	0.033
$\mathbb{P}\left(L=l \mid x_{1,i}=0, x_{2,i}=1\right)$	0.069	0.024	0.033	0.029	0.034	0.032	0.036	0.036	0.038	0.039	0.041
$\mathbb{P}\left(L=l \mid x_{1,i}=1, x_{2,i}=0\right)$	0.066	0.023	0.032	0.028	0.032	0.032	0.035	0.035	0.038	0.039	0.041
$\mathbb{P}\left(L=l \mid x_{1,i}=1, x_{2,i}=1\right)$	0.145	0.045	0.059	0.046	0.052	0.046	0.047	0.044	0.044	0.042	0.042
l	11	12	13	14	15	16	17	18	19	20	21
$\mathbb{P}\left(L=l \mid x_{1,i}=0, x_{2,i}=0\right)$	0.037	0.041	0.046	0.052	0.058	0.065	0.073	0.082	0.093	0.104	0.117
$\mathbb{P}\left(L=l \mid x_{1,i}=0, x_{2,i}=1\right)$	0.043	0.045	0.047	0.049	0.051	0.053	0.055	0.058	0.060	0.063	0.065
$\mathbb{P}\left(L=l \mid x_{1,i}=1, x_{2,i}=0\right)$	0.043	0.045	0.047	0.049	0.052	0.054	0.057	0.059	0.062	0.065	0.068
$\mathbb{P}\left(L=l \mid x_{1,i}=1, x_{2,i}=1\right)$	0.040	0.039	0.038	0.037	0.036	0.035	0.034	0.033	0.032	0.032	0.031



(a) Using r_l in Table 2 (b) Using r_l^{ed} in Table 2

Fig. 1. Coefficients of n^* at each BM level with 95% confidence intervals. Here, we have increased the scope of the y-axis on purpose so that the linear version of the slope can be easily compared with that of Fig. 2.

is the limiting stationary distribution calculated from the transition probability in the BMS (Denuit et al., 2007). The resulting dataset then contains x_{1i}, $x_{2,i}$, n_i^*, $\boldsymbol{y}_i^{*\mathrm{T}}$, and the average claim severity m_i^* for each driver $i = 1, \ldots, 1,000,000$.

With this claim dataset, we may quantify the frequency-severity dependence at each BM level. We fit the reported claims using the following gamma regression model.

$$\log(\mu_i) = \sum_{j=0}^{21} \beta_{0j} I(l_j = l_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \sum_{j=0}^{21} \beta_{3j} I(l_j = l_i) n_i^*, \quad (30)$$

where different BM levels are defined as factors. The coefficient β_{3j} represents the degree of dependence between the reported claim severity and frequency in each BM level l_i. Specifically, for the given BM level $l_i = l$, the equation in (30) becomes

$$\log(\mu_i) = \beta_{0l} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{3l} n_i^*.$$

Fig. 1 shows the estimation results for β_{3j} from the reported claims, where we denote as $\widehat{\beta_{3j}}$, of β_{3j} (shown as dots $\bullet$) and its 95% confidence interval (shaded) for each BM level under two different relativities r_l and r_l^{ed}. Clearly, the estimator $\widehat{\beta_{3j}}$ is hardly zero in both figures, indicating that the frequency and severity in the reported claims are correlated. Second, there is a solid, consistent downward trend for the estimator $\widehat{\beta_{3j}}$ as the BM level j increases. Specifically, as we move toward lower BMS levels, we observe slightly positive values of $\widehat{\beta_{3j}}$, implying a mildly positive dependence between the claim frequency and severity. As the BM level increases, however, $\widehat{\beta_{3j}}$ decreases and becomes negative, implying a strong negative dependence.

The superimposed straight lines in Fig. 1 are produced by a supplementary linear regression formulation in which the BM levels are treated as being numerical. This approach enables us to quantify the trend of dependence between frequency and severity across different BM levels. With this formulation, we can investigate the linear trend, if it exists, and measure the average change in the severity amount due to a scale shift in the BM level. The specified regression model is

$$\log(\mu_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 n_i^* + \gamma_4 l_i + \gamma_5 n_i^* \times l_i, \quad (31)$$

where $\gamma_3 + \gamma_5 l_i$, the coefficient of n_i^*, denotes the degree of dependence between the reported claim severity and frequency at BM level l_i, and γ_5 represents the change in the dependence due to the BM level change. The lines in Fig. 1 are the estimates of $\gamma_3 + \gamma_5 l$. The slope of γ_5 are estimated to be -0.0080 under r_l and -0.0077 under r_l^{ed}, with p-values less than 10^{-15} for both. Our simulation study therefore confirms Claim 1 in that the pattern in Fig. 1 clearly shows that the dependence between the *claim* frequency and severity tends to be positive for lower BM levels and negative for higher BM levels.

5. Real data application

In this section, we investigate the dependence between claim frequency and severity across BM levels, using Korean auto insurance claim data. The analyses in the previous sections suggest that the dependence should vary across different BM levels and the dependence should gradually become more negative as the BM level increases toward to the strongest malus class. This result should be the case as long as policyholders make a strategic claim reporting decision called bonus-hunger in a bounded malus structure.

Table 6
Distribution of numbers of accidents and insured parties.

(a) Number of accidents

Frequency	0	1	2	3	4	5	6	7	8	9
Collision	393,581	86,700	16,595	2,552	454	78	27	8	2	3
Liability	436,678	57,512	5,177	532	79	15	4	3	0	0

(b) Number of insured parties by factor

		Sex					
		0 – Male			1 – Female		
		Age group					
Sports	Car size	1 – All ages	2 – Above 26	3 – Above 43	1 – All ages	2 – Above 26	3 – Above 43
0 – General	1 – Small	3,893	32,716	15,015	7,228	73,247	28,097
	2 – Mid-sized	1,573	17,701	12,074	4,415	72,292	42,712
	3 – Large	568	9,596	8,815	1,648	31,498	30,369
	4 – Multi-Purp	592	10,110	7,082	2,191	49,992	34,413
1 – Sports	1 – Small	2	39	7	4	49	10
	2 – Mid-sized	37	188	48	64	791	78
	3 – Large	15	135	53	45	478	120
	4 – Multi-Purp	0	0	0	0	0	0

5.1. Auto insurance in Korea

Personal auto insurance policies in Korea are similar to the fault-based auto insurance system in most US states. Basic bodily injury liability and physical damage liability coverages are mandatory. Policyholders may choose extra bodily injury liability coverage and may increase the limit for the physical damage liability coverage with extra cost. As most policyholders choose extra bodily injury liability coverage and extra physical damage liability coverage, almost all claims are not bounded by a policyholder's choice of limits (Park and Han, 2017). Several other types of coverage, such as first-party physical damage, personal injury, and uninsured motorist coverage, are also available. About 60–70 percent of contracts carry first-party physical damage coverage. Insurers use both *a priori* and *a posteriori* ratings. For the *a posteriori* rating, Korea has a national BMS, and the BMS levels of all policyholders are shared by all auto insurance companies through the Korean Insurance Development Institute (KIDI). Although the BMS is standardized and the BMS relativities are suggested by KIDI, insurers may have their own relativities. There is one BMS per policy applied to all selected coverages.

5.2. Data

The data was provided by a major insurer in Korea and contains the records of about 500,000 auto insurance contracts during 2010. All contracts include first-party physical damage and liability coverage. Our data includes rating factors, such as the gender and age of the policyholder, the type of car insured, the value of the vehicle, and the BMS class (0 — 21 BM levels). In addition, the total number and severity of claims per coverage type are available. The following are the specific explanations of these variables:

- Gender of the main driver: male (1) or female (0)
- Allowed age range of the co-drivers: all ages (1), ages above 26 (2), or ages above 43 (3). Category 3 is considered to be the least risky.
- Vehicle type: sports (1) or general (0)
- Vehicle size: small (1), mid-sized (2), large (3), or multi-purpose (4)
- Value of vehicle: expressed as the percentage of the average car value in the market
- Claim frequency: separate numbers of claims for collision and liability coverage.
- Claim severity: the aggregate amount of collision and liability claims for each contract.

As our data only has the sum of annual claims, (n_i^*, m_i^*), and not the individual ground-up loss amounts, the optimal retention for each claim is not directly obtainable.[3] Note that although we have records for the number of claims for each coverage type separately, we do not have the total claim frequency that triggers BMS transitions. A BMS transition occurs whenever a claim is made, but the claim may be attributed to first-party physical damage, liability, or both. If one accident gives rise to claims for both coverages, it is treated as a single claim in determining the next BM level. To avoid double counting, we use the maximum of the first-party physical damage frequency and the liability frequency as a proxy for the claim frequency related to the BMS transition. Such data analyses may somewhat underestimate the total frequency if a policy has both first-party physical damage-only and liability-only claims separately.[4]

In Tables 6 and 7, we present the summary information of the dataset. Table 6 shows the number of claims reported by each coverage type and the number of insured parties by each factor; Table 7 provides some details about these claims by factor. Out of 500,000 drivers, 127,096 have claimed accidents, among which there are 63,322 non-zero liability claims and 106,419 non-zero first-party physical damage claims.

5.3. Preliminary analysis with a constant dependence parameter

As a preliminary analysis, we run a severity model with the claim frequency included as a covariate without considering varying dependence by BM level, following previous studies

[3] In cases in which the individual loss amount is available, the estimation of the optimal retention is possible under some mild assumptions (Walhin and Paris, 2000).

[4] However, such cases must be very rare. According to internal data available to us from KIDI, 48.7 percent of first-party physical damage claims were claims from first-party-only accidents, such as theft, fire, unknown causes, and so on. In our data, if we assume that all liability claims always are reported with first-party physical damage claims, 46.2 percent of first-party claims can be classified as first-party-only claims. The discrepancy between the two is about 2.5 percent. The chance that one policy claims both liability-only and first-party-only claims separately in one year is much lower than even this value. Therefore, although this estimation of the total number of claims underestimates the actual number of claims, the difference is minor.

Table 7
Information on the number of accidents and average severity by factor.

		Sex					
		0 – Male			1 – Female		
		Age group					
Sports	Car size	1 – All ages	2 – Above 26	3 – Above 43	1 – All ages	2 – Above 26	3 – Above 43
(a) Number of first-party physical damage/liability accidents by factor							
0 – General	1 – Small	1,151	7,189	3,283	2,060	14,768	4,740
		(785)	(4,491)	(1,974)	(1,351)	(8,866)	(2,979)
	2 – Mid-sized	525	4,272	2,772	1,254	15,812	8,351
		(342)	(2,469)	(1,688)	(844)	(9,019)	(4,881)
	3 – Large	183	2,492	2,179	461	7,487	6,890
		(112)	(1,274)	(1,091)	(279)	(3,601)	(3,352)
	4 – Multi-Purp	157	2,127	1,502	507	9,556	6,206
		(123)	(1,527)	(1,077)	(396)	(6,300)	(4,254)
1 – Sports	1 – Small	1	13	0	0	12	0
		(1)	(5)	(0)	(0)	(2)	(0)
	2 – Mid-sized	15	39	12	23	203	15
		(10)	(23)	(7)	(16)	(101)	(5)
	3 – Large	2	37	8	15	86	14
		(1)	(21)	(5)	(12)	(36)	(2)
	4 – Multi-Purp	0	0	0	0	0	0
		(0)	(0)	(0)	(0)	(0)	(0)
(b) Average severity of first-party physical damage/liability accidents by factor							
0 – General	1 – Small	1,121,739	920,694	899,317	1,187,316	930,078	836,142
		(1,103,755)	(893,652)	(920,386)	(1,097,813)	(940,680)	(893,268)
	2 – Mid-sized	1,535,184	1,251,998	1,112,995	1,459,561	1,184,697	1,022,244
		(1,173,445)	(1,059,536)	(970,022)	(1,246,630)	(1,055,312)	(1,019,214)
	3 – Large	2,050,200	1,974,018	1,653,136	2,097,726	1,717,536	1,512,176
		(1,000,299)	(1,080,479)	(1,046,758)	(1,247,227)	(1,193,201)	(1,024,122)
	4 – Multi-Purp	1,175,796	1,143,676	1,161,230	1,280,482	1,083,026	1,054,669
		(1,078,529)	(1,053,789)	(1,239,510)	(1,210,786)	(1,098,809)	(1,115,603)
1 – Sports	1 – Small	588,600	2,716,153	NA	NA	3,833,828	NA
		(360,000)	(918,966)	(NA)	(NA)	(2,456,770)	(NA)
	2 – Mid-sized	2,736,371	1,511,837	1,864,517	2,157,330	1,688,383	1,462,112
		(1,514,173)	(1,012,480)	(717,944)	(1,376,109)	(1,464,385)	(1,835,768)
	3 – Large	4,435,500	5,869,510	4,186,295	12,496,146	8,730,600	1,425,321
		(1,233,500)	(1,026,843)	(1,397,606)	(3,285,659)	(1,736,556)	(256,580)
	4 – Multi-Purp	NA	NA	NA	NA	NA	NA
		(NA)	(NA)	(NA)	(NA)	(NA)	(NA)

(Frees et al., 2014; Shi et al., 2015; Garrido et al., 2016). Using the same severity model formulation as in (9), we apply the gamma and inverse Gaussian regressions to both coverage types separately, with the mean specified as

$$\begin{aligned}\log(\mu_{i,\text{type}}) = \beta_0 &+ \beta_1 I(\text{sex}_i = 1) + \beta_2 I(\text{type}_i = 1) + \beta_3 \text{value}_i \\ &+ \beta_4 I(\text{size}_i = 2) + \beta_5 I(\text{size}_i = 3) + \beta_6 I(\text{size}_i = 4) \\ &+ \beta_7 I(\text{age}_i = 2) + \beta_8 I(\text{age}_i = 3) + \beta_9 n^*_{i,\text{type}},\end{aligned} \tag{32}$$

where "type" is either the first-party physical damage or the liability. The results are shown in Table 8, where the estimated coefficients and their t-values are presented along with Akaike's information criterion (AIC) and the scaled deviance (McCullagh and Nelder, 1989) for each model. From the gamma regression result, a negative dependence is observed between the claim frequency and severity on the collision side based on the estimate of $\beta_9 = -0.0180$, and a positive dependence is found for liability coverage, as seen from the estimate $\beta_9 = 0.0274$. Both estimates are of substantial statistical significance with t-values of -3.85 and 2.79, respectively. The inverse Gaussian regression reports similar results in the table. These findings are in line with the results in Frees et al. (2014), Shi et al. (2015), and Garrido et al. (2016), who observe negative dependence for collision coverage and positive dependence for liability coverage. Overall, the inverse Gaussian model performs better than the gamma model for both collision and liability in terms of AIC; the *p*-values of the deviance tests for all models considered are close to 0, showing adequate fits under both models.

5.4. Model with varying dependence parameters by BM levels

We now model the first-party physical damage claims and the liability claims separately, with the BM level information included for individual drivers. We again apply the gamma and the inverse Gaussian regression models specified as

$$\begin{aligned}\log(\mu_i) = \sum_{j=0}^{21} \beta_{0j} I(l_j = l_i) &+ \beta_1 I(\text{sex}_i = 1) + \beta_2 I(\text{type}_i = 1) \\ &+ \beta_3 \text{value}_i + \beta_4 I(\text{size}_i = 2) + \beta_5 I(\text{size}_i = 3) \\ &+ \beta_6 I(\text{size}_i = 4) + \beta_7 I(\text{age}_i = 2) + \beta_8 I(\text{age}_i = 3) \\ &+ \sum_{j=0}^{21} \beta_{9j} I(l_j = l_i) \tilde{n}^*_i .\end{aligned} \tag{33}$$

Here, $\tilde{n}^*_i$ (i.e., $\max(n^*_{i,col}, n^*_{i,liab})$) is used as a proxy for the total claim frequency, since we intend to observe the severity behavior to the BMS transition, and the BMS transition is determined by the total number of claims.

The result of this model for first-party physical damage is described in Table 9(a) and is also plotted in Fig. 2(a). In this figure, the slopes $\beta_{9j}, j = 0, \ldots, 21$ of BM level j are depicted as dots • against the BM level *j*. Two numbers that determine the size of the confidence interval are attached to each marked value; the value below the dot is the number of drivers with at least one claim in the corresponding BM level, and the number above the dot stands for the number of drivers with at least two claims in that BM level. We

Table 8
GLM regression results for severity excluding BM level information.

	First-party physical damage				Liability			
	Gamma		Inv. Gaussian		Gamma		Inv. Gaussian	
	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value
Car value	0.1678	46.05	0.1703	36.23	0.0777	9.87	0.0826	9.32
Sex = 1	−0.0105	−1.32	−0.0053	−0.70	0.0463	3.06	0.0448	3.02
Car type = 1	0.4008	8.14	0.3567	5.68	0.1682	1.60	0.1671	1.39
Car size = 2	0.1556	17.30	0.1564	18.77	0.0904	5.26	0.0895	5.32
Car size = 3	0.3323	29.10	0.3251	27.95	0.0838	3.65	0.0844	3.68
Car size = 4	0.1605	15.52	0.1690	17.69	0.1728	9.15	0.1735	9.19
Age = 2	−0.1451	−9.92	−0.1423	−9.71	−0.0880	−3.31	−0.0913	−3.35
Age = 3	−0.1861	−12.12	−0.1801	−11.77	−0.0916	−3.27	−0.0926	−3.23
$\tilde{n}_i^*$	−0.0180	−3.85	−0.0183	−4.14	0.0274	2.79	0.0260	2.60
Dispersion	1.5606		0.0000016		3.090		0.0000033	
Scaled deviance	57038		8440		20283		27839	
AIC	3819176		3801930		2060096		2040176	

Table 9
GLM regression for severity using BM level information.

	Gamma				Inverse Gaussian			
	Model: Eq. (33)		Model: Eq. (34)		Model: Eq. (33)		Model: Eq. (34)	
	Estimate	t-value	Estimate	t-value	Estimate	t-value	Estimate	t-value
(a) First-party physical damage coverage								
Car value	0.1584	43.04	0.1572	43.09	0.1571	33.03	0.1556	33.20
Sex = 1	0.0180	2.24	0.0191	2.39	0.0216	2.84	0.0224	2.94
Car type = 1	0.3868	7.91	0.3862	7.90	0.3418	5.48	0.3408	5.45
Car size = 2	0.1706	19.00	0.1713	19.10	0.1710	20.52	0.1718	20.65
Car size = 3	0.3563	31.00	0.3604	31.53	0.3484	29.69	0.3521	30.19
Car size = 4	0.1851	17.80	0.1871	18.04	0.1907	19.85	0.1921	20.06
Age = 2	−0.1548	−10.63	−0.1527	−10.50	−0.1526	−10.45	−0.1506	−10.33
Age = 3	−0.1708	−11.13	−0.1676	−10.97	−0.1677	−10.97	−0.1652	−10.85
γ_{11}	NA	NA	−0.0036	−3.17	NA	NA	−0.0039	−3.64
Dispersion	1.5368		1.5394		0.0000015		0.0000016	
Scaled deviance	57446		57395		84723		84670	
AIC	3818065		3818107		3801282		3801269	
(b) Liability coverage								
Car value	0.0682	8.54	0.0694	8.76	0.0706	7.92	0.0722	8.14
Sex = 1	0.0684	4.46	0.0677	4.41	0.0658	4.38	0.0654	4.33
Car type = 1	0.1565	1.50	0.1581	1.51	0.1538	1.29	0.1561	1.30
Car size = 2	0.1008	5.86	0.0998	5.79	0.0999	5.95	0.0980	5.81
Car size = 3	0.1095	4.72	0.1058	4.58	0.1121	4.84	0.1078	4.66
Car size = 4	0.1940	10.20	0.1919	10.09	0.1958	10.30	0.1931	10.14
Age = 2	−0.0972	−3.66	−0.0974	−3.66	−0.1016	−3.73	−0.1012	−3.70
Age = 3	−0.0829	−2.95	−0.0831	−2.96	−0.0868	−3.03	−0.0853	−2.97
γ_{11}	NA	NA	0.0008	0.31	NA	NA	0.0007	0.29
Dispersion	3.0571		3.0833		0.0000033		0.0000033	
Scaled deviance	20405		20267		28038		27714	
AIC	2059807		2059866		2040026		2040032	

find that the most claims are positioned in BM levels below 16, and, hence, the corresponding confidence intervals for the coefficients of $\tilde{n}_i^*$ at each BM level are quite narrow. Although the estimates in the three highest BM levels are relatively unstable due to scarce number of observations, reflected in the wider confidence intervals, the slope values across the BM levels clearly show an overall downward trend. We note that the overall shape and position of this figure is comparable to the simulated data result depicted in Fig. 1. We supplement this result by calibrating the same data with BM levels defined as numbers, that is, for both the gamma and inverse Gaussian models,

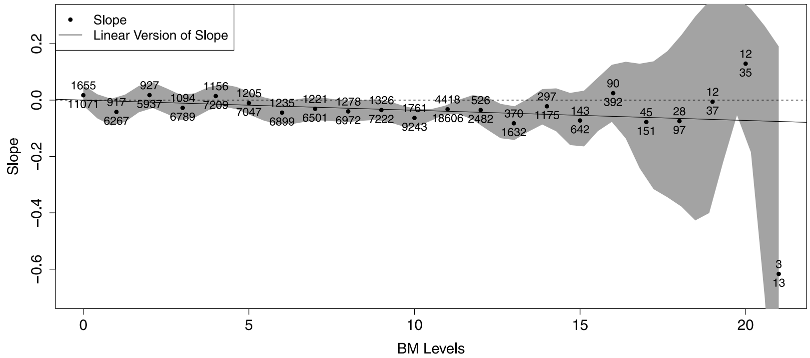
$$\begin{aligned}\log(\mu_i) = \gamma_0 + \gamma_1 I(\text{sex}_i = 1) + \gamma_2 I(\text{type}_i = 1) + \gamma_3 \text{value}_i \\ + \gamma_4 I(\text{size}_i = 2) + \gamma_5 I(\text{size}_i = 3) \\ + \gamma_6 I(\text{size}_i = 4) + \gamma_7 I(\text{age}_i = 2) + \gamma_8 I(\text{age}_i = 3) \\ + \gamma_9 l_i + \gamma_{10}\tilde{n}_i^* + \gamma_{11}\tilde{n}_i^* \times l_i.\end{aligned} \tag{34}$$

This regression setup has a slope of $\gamma_{10} + \gamma_{11} l_i$ for the claim frequency variable, which is shown as the superimposed straight line in Fig. 2(a). The resulting regression finds that all covariates other than sex are statistically highly significant. In particular, the dependence between the claim severity and frequency is positive at BM level zero, but becomes negative from level one and onward. The tendency toward negative dependence becomes stronger as the BM level increases, as anticipated. Under the gamma regression model, the slope estimate for $\tilde{n}_i^*$ at BM level l_i is

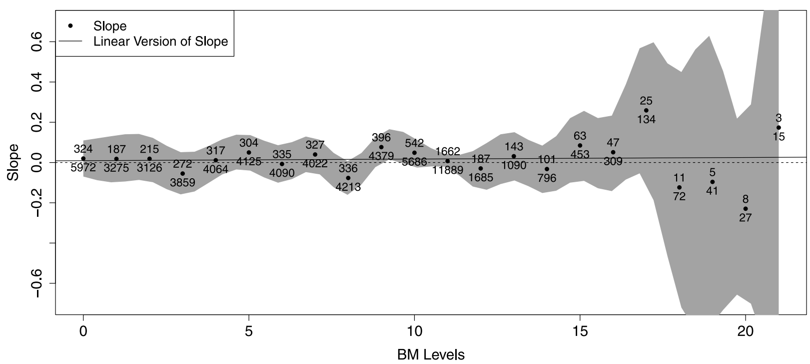
$$5.297 \cdot 10^{-6} - 3.612 \cdot 10^{-3} \cdot l_i, \qquad l_i \in \{0, 1, 2, \dots, 21\}.$$

The t-value of $\gamma_{11} = -3.6 \cdot 10^{-3}$ is -3.17, and the p-value of the corresponding chi-squared test is 0.001434, confirming a clear downward trend throughout the BM levels. The conclusion is similar for the inverse Gaussian regression in the same table, where the t-value of $\gamma_{11} = -3.9 \cdot 10^{-3}$ is -3.64 and the p-value of the corresponding chi-squared test is 0.000250. The analyses provide strong evidence of the bonus-hunger phenomenon for first-party physical damage insurance coverage.

We now turn to liability coverage. We employ the identical specification as in (33) using both gamma and inverse Gaussian regression models. The results of the fitted models are shown in



(a) First-party physical damage coverage



(b) Liability coverage

Fig. 2. Coefficients of n^* at each BM level in real data along with 95% confidence intervals.

Table 9(b), and the coefficients on $\tilde{n}_i^*$ across the BM levels are depicted in Fig. 2(b). Unlike in the case of first-party physical damage insurance, the estimated model reveals that the slope estimate for $\tilde{n}^*$ of the liability coverage, $\beta_{9j}, j = 0, \ldots, 21$, does not have a particular trend. This result is also confirmed by the t-values of the γ_{11} estimates under the alternative model (34), which are 0.31 in the gamma model and 0.29 in the inverse Gaussian model. Furthermore, the p-values of the corresponding chi-squared test are 0.7585 and 0.7788, respectively. Fig. 2(b) also illustrates the absence of a clear dependence trend. Although the analyses of the two coverage types show a stark contrast in terms of the frequency-severity dependence, this result is not necessarily contradictory if liability coverage is less likely to be affected by drivers' behavioral preferences based on bonus hunger. This assumption is quite reasonable because liability claims involve third parties and, thus, opportunistic claim behavior is less likely to occur. Although it may be possible for a driver to report a liability claim by negotiating with the third party and paying cash, such events are deemed rare for two reasons. First, the amount of the loss is much less manipulable. Second, the third party involved in the accident has zero or minimal incentive for not reporting the claim.

6. Concluding remarks

The classical auto insurance literature assumes independence between the claim frequency and severity. While this assumption yields simpler pricing models, its validity is not always justifiable. A series of recent studies suggests several approaches to account for the possible dependence between the claim frequency and severity, and empirically finds evidence for dependence in auto insurance claims. Such dependence leads to research on the BMS where claim severity, as well as the claim frequency, is used as a rating factor. For example, a pioneer study by Pinquet (1997) suggests a generalized BMS which incorporates claim severity in a BMS when there exists a dependence between severity and frequency. More recently, Gómez-Déniz (2016) proposed the so-called credibility BMS based on Bayesian approach where the claim severity is used as a rating factor by dividing claims into two types: those with severity less than some predetermined threshold and those which exceed this threshold. Our model is similar to Gómez-Déniz (2016) in the sense that we allow policyholders to report losses above certain retention level. However, our model is different from Gómez-Déniz (2016) because in our model the threshold of reporting, or the retention level, is not constant and this varying threshold essentially creates a dependence between claim frequency and severity. Also, in these studies where the claim severity is incorporated as a rating factor it is assumed that the dependence between the claim frequency and severity is constant, which is inconsistent with the empirical dependence.

To this extent, we attempt, in this study, to explain an underlying cause that drives the dependence between claim frequency and severity based on bonus hunger, a behavioral decision-making process of drivers. In particular, we build a GLM-based severity model coupled with a BMS to show that bonus hunger can induce the claim frequency-severity dependence, even when the original loss frequency and severity are independent. Through our proposed model, we claim that bonus hunger leads to dependence structures that can vary across BM levels. As a real-world application, we apply our model to actual insurance claim data and confirm our claim. We find that bonus hunger can shed light on understanding the seemingly contradictory dependence results for collision and liability insurance that appear in the existing literature.

Our results have several implications. First, the reported discrepancy of dependencies found in the literature could be a natural

Table A.10
Optimal retentions using the relativity in Table 2 in units of 10,000 won.

	$\eta_g(l,\lambda_1)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	13	25	36	45	54	61	69	76	84	95	106
g=1	28	34	39	44	48	53	59	68	76	80	82
g=2	39	44	49	54	60	70	79	84	88	98	114
g=3	54	57	60	71	83	94	109	155	172	142	122
g=4	66	82	104	151	168	140	121	107	99	96	92
l	11	12	13	14	15	16	17	18	19	20	21
g=0	115	120	126	133	142	152	166	176	187	126	65
g=1	88	94	102	115	146	152	129	63	0	0	0
g=2	155	168	141	121	110	55	0	0	0	0	0
g=3	111	102	99	51	0	0	0	0	0	0	0
g=4	93	49	0	0	0	0	0	0	0	0	0

	$\eta_g(l,\lambda_2)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	10	20	29	37	44	50	56	62	69	78	88
g=1	23	27	32	36	39	44	49	56	63	66	69
g=2	32	36	40	44	49	57	65	69	72	80	93
g=3	43	46	50	58	67	76	87	122	139	119	106
g=4	52	65	82	117	133	116	104	95	90	88	85
l	11	12	13	14	15	16	17	18	19	20	21
g=0	96	100	105	111	119	128	140	150	161	109	57
g=1	73	79	85	96	122	128	113	55	0	0	0
g=2	124	138	121	107	99	50	0	0	0	0	0
g=3	99	92	90	46	0	0	0	0	0	0	0
g=4	85	45	0	0	0	0	0	0	0	0	0

	$\eta_g(l,\lambda_3)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	11	21	30	37	44	51	57	63	70	79	89
g=1	23	27	32	36	40	44	49	56	64	67	69
g=2	32	36	40	45	50	58	66	70	73	81	94
g=3	44	47	50	58	68	77	88	124	141	121	107
g=4	53	66	83	119	135	117	105	96	90	88	85
l	11	12	13	14	15	16	17	18	19	20	21
g=0	97	101	106	112	120	129	142	152	162	110	57
g=1	74	80	86	97	123	130	114	56	0	0	0
g=2	126	139	122	108	100	50	0	0	0	0	0
g=3	100	93	91	47	0	0	0	0	0	0	0
g=4	86	45	0	0	0	0	0	0	0	0	0

	$\eta_g(l,\lambda_4)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	8	16	24	30	36	41	46	51	57	65	73
g=1	18	22	26	29	32	36	40	47	53	56	58
g=2	26	29	33	36	41	47	54	58	60	66	76
g=3	35	38	41	48	56	62	70	97	112	100	92
g=4	43	52	65	92	105	96	89	83	80	80	78
l	11	12	13	14	15	16	17	18	19	20	21
g=0	80	84	89	94	101	109	120	129	139	95	50
g=1	62	67	72	81	102	109	99	49	0	0	0
g=2	101	113	103	94	89	45	0	0	0	0	0
g=3	88	84	82	42	0	0	0	0	0	0	0
g=4	78	41	0	0	0	0	0	0	0	0	0

consequence of the bonus-hunger behavior of drivers, rather than a result of sample-specific errors. Second, depending on the BMS level, bonus hunger can induce both positive and negative dependences between claim frequency and severity, suggesting that a constant dependence structure may not be appropriate for the severity model. In this study, we find that the dependence between claim frequency and severity may exist for some groups but could be negated by other groups having the opposite dependence. We also anticipate that future work on more sophisticated models allowing a varying dependence parameter, may further improve the loss prediction.

Acknowledgments

Joseph H.T. Kim was supported by the Basic Science Research Program of the National Research Foundation (NRF) grant funded by the Korean Government (NRF-2015R1A1A1A05027336). Sojung C. Park appreciates the support from the Institute of Management Research and the Institute of Finance and Banking at Seoul National University. Jae Youn Ahn was supported by the Basic Science Research Program of the National Research Foundation (NRF) grant funded by the Korean Government (NRF-2017R1D1A1B03032318).

Table A.11
Optimal retentions using the equal distance relativity in units of 10,000 won.

	$\eta_g^{ed}(l,\lambda_1)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	29	55	76	93	107	119	128	136	143	148	153
g=1	58	68	76	83	89	94	98	101	104	106	108
g=2	75	82	88	93	98	102	106	111	117	124	131
g=3	90	96	103	112	121	131	142	162	157	124	99
g=4	118	130	144	169	168	136	112	94	79	67	55
l	11	12	13	14	15	16	17	18	19	20	21
g=0	157	160	162	163	163	160	154	140	121	72	32
g=1	109	110	111	111	114	100	72	31	0	0	0
g=2	144	135	104	79	59	26	0	0	0	0	0
g=3	80	64	50	22	0	0	0	0	0	0	0
g=4	45	21	0	0	0	0	0	0	0	0	0

	$\eta_g^{ed}(l,\lambda_2)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	24	46	63	77	89	99	107	114	120	125	129
g=1	48	56	63	69	74	79	82	85	88	90	92
g=2	63	69	74	78	82	86	89	93	97	103	109
g=3	75	80	85	91	99	107	117	135	135	111	91
g=4	95	105	117	140	143	120	102	88	75	65	54
l	11	12	13	14	15	16	17	18	19	20	21
g=0	133	135	137	139	139	137	133	122	108	65	30
g=1	93	94	95	96	99	88	66	28	0	0	0
g=2	122	117	93	73	56	25	0	0	0	0	0
g=3	75	61	48	22	0	0	0	0	0	0	0
g=4	44	20	0	0	0	0	0	0	0	0	0

	$\eta_g^{ed}(l,\lambda_3)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	25	46	64	78	90	100	108	115	121	126	131
g=1	48	57	64	70	75	79	83	86	89	91	93
g=2	63	70	75	79	83	87	90	94	99	104	111
g=3	76	80	86	93	100	108	118	137	136	111	92
g=4	96	106	119	142	145	121	103	88	76	65	54
l	11	12	13	14	15	16	17	18	19	20	21
g=0	134	137	139	140	140	138	134	123	108	66	30
g=1	94	95	96	97	100	89	66	29	0	0	0
g=2	123	118	94	73	56	25	0	0	0	0	0
g=3	76	61	49	22	0	0	0	0	0	0	0
g=4	44	20	0	0	0	0	0	0	0	0	0

	$\eta_g^{ed}(l,\lambda_4)$										
l	0	1	2	3	4	5	6	7	8	9	10
g=0	20	38	53	65	75	83	91	97	102	106	110
g=1	40	47	53	58	63	67	70	72	75	77	78
g=2	53	58	62	66	70	73	75	78	82	86	92
g=3	63	67	71	76	81	88	96	113	114	97	83
g=4	77	85	96	116	121	105	92	81	71	61	52
l	11	12	13	14	15	16	17	18	19	20	21
g=0	113	116	118	119	119	118	115	107	96	59	27
g=1	80	80	82	83	86	78	60	26	0	0	0
g=2	103	101	83	67	52	23	0	0	0	0	0
g=3	70	58	46	21	0	0	0	0	0	0	0
g=4	42	19	0	0	0	0	0	0	0	0	0

Appendix A

See Tables A.10 and A.11.

References

Baumgartner, C., Gruber, L.F., Czado, C., 2015. Bayesian total loss estimation using shared random effects. Insurance Math. Econom. 62, 194–201.
Czado, C., Kastenmeier, R., Brechmann, E.C., Min, A., 2012. A mixed copula model for insurance claims and claim sizes. Scand. Actuar. J. (4), 278–305.
De Jong, P., Heller, G.Z., et al., 2008. Generalized Linear Models for Insurance Data, volumn 10. Cambridge University Press Cambridge.
Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J.-F., 2007. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. John Wiley & Sons.
Dimakos, X.K., Di Rattalma, A.F., 2002. Bayesian premium rating with latent structure. Scand. Actuar. J. 2002 (3), 162–184.
Ferreira Jr., J., Minikel, E., 2012. Measuring per mile risk for pay-as-you-drive automobile insurance. Transp. Res. Rec. J. Transp. Res. Board (2297), 97–103.
Frees, E.W., 2010. Regression Modeling with Actuarial and Financial Applications. Cambridge University Press, New York.
Frees, E.W., Derrig, R.A., Meyers, G., 2014. Predictive Modeling Applications in Actuarial Science, volume 1. Cambridge University Press.
Frees, E.W., Lee, G., Yang, L., 2016. Multivariate frequency-severity regression models in insurance. Risks 4 (1), 4.

Garrido, J., Genest, C., Schulz, J., 2016. Generalized linear models for dependent frequency and severity of insurance claims. Insurance Math. Econom. 70, 205–215.

Gómez-Déniz, E., 2016. Bivariate credibility bonus–malus premiums distinguishing between two types of claims. Insurance Math. Econom. 70, 117–124.

Jeong, H., Valdez, E.A., Ahn, J.Y., Park, S., 2017. Generalized linear mixed models for dependent compound risk models. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3045360.

Jørgensen, B., Paes De Souza, M.C., 1994. Fitting tweedie's compound Poisson model to insurance claims data. Scand. Actuar. J. 1994 (1), 69–93.

Klugman, S., Panjer, H., Willmot, G., 2008. Loss Models, third ed. John Wiley, New York.

Lemaire, J., 1977. La soif du bonus. Astin Bull. 9 (1–2), 181–190.

Lemaire, J., 2012. Bonus-malus Systems in Automobile Insurance, volume 19. Springer science & business media.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. In: CRC Monographs on Statistics & Applied Probability, Springer Verlag, New York.

Norberg, R., 1976. A credibility theory for automobile bonus systems. Scand. Actuar. J. 1976 (2), 92–107.

Park, S.C., Han, S., 2017. Externalities from driving luxury cars. Risk Manag. Insur. Rev. 20 (3), 391–427.

Philipson, C., 1960. The swedish systems of bonus. Astin Bull. 1 (03), 134–141.

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

Pitrebois, S., Walhin, J.-F., Denuit, M., 2005. Bonus-malus systems with varying deductibles. ASTIN Bulletin J. IAA 35 (1), 261–274.

Shi, P., Feng, X., Ivantsova, A., 2015. Dependent frequency–severity modeling of insurance claims. Insurance Math. Econom. 64, 417–428.

Tan, C.I., 2016. Varying transition rules in bonus–malus systems: From rules specification to determination of optimal relativities. Insurance Math. Econom. 68, 134–140.

Tan, C.I., Li, J., Li, J.S.-H., Balasooriya, U., 2015. Optimal relativities and transition rules of a bonus–malus system. Insurance Math. Econom. 61, 255–263.

Walhin, J.F., Paris, J., 2000. The true claim amount and frequency distributions within a bonus-malus system. Astin Bull. 30 (2), 391–403.