Predictive compound risk models with dependence

Himchan Jeong, University of Connecticut[*]

Emiliano A. Valdez, University of Connecticut[†]

Abstract

The two-part regression model, which is subdivided into claims frequency and severity, is well-known in the actuarial literature for predicting insurance pure premium based on aggregate claims. While independence between frequency and severity is conventionally assumed, there is an increase of interest in developing models to capture the possible dependence such as Garrido et al. (2016). This paper, which extends our work in Jeong et al. (2018a), explores the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time, within a two-part framework relaxing the assumption of independence between claims frequency and severity. As a result, we introduce a generalized formula for credibility premium of compound sum with dependence, which extends and integrates previous research on both credibility premium of compound sum and dependent two-part compound risk model. In the generalized forumla of credibility premium of compound sum, $D_N(\gamma)$ is introduced as an informative measure of association between frequency and severity not only the sign of it but also the magnitude, which can be easiliy interpreted and implemented in property and casualty actuarial ratemaking practice.

Keywords: Dependent frequency and severity, generalized linear model (GLM), hierarchical model, generalized Pareto (GPareto), generalized beta of the second kind (GB2), credibility premium of compound sum with dependence

1 Introduction

In the actuarial practice of predicting insurance pure premium based on aggregate claims, two-part model has been widely used. It decomposes the aggregate claims into two separate parts: one for claims frequency and the other for claims severity. Due to the non-normalness of claim frequency and severity, there have been a lot of attempts to incorporate various types of distribution, and it is noteworthy that most of them are closely related to use of generalized linear models (GLMs), which was introduced by Nelder and Wedderburn (1972). By the idea of GLMs, we need not adhere to ordinary linear regression models which implicitly assumes normally distributed random error. Indeed, it extends the possible distributions for the regression to the members of the exponential family, which includes Gaussian, Poisson, logistic, gamma, and so on. Therefore, two-part GLM has been used for modeling loss frequency and the average of loss severity as a benchmark.

[*]himchan.jeong@uconn.edu

[†]Corresponding author, emiliano.valdez@uconn.edu

Even though two-part GLM is popular because of the flexibility, there are some issues which need to be discussed. First, typically two-part GLM assumes independence between claims frequency and severity so as to ease modeling and computation. However, we may not preclude the possible dependence between the frequency and severity component. Inspired by this idea, there have been an increasing interest in developing models to capture the possible dependence such as Shi et al. (2015) and Lee et al. (2019).

Furthermore, it is important to consider the repeated measurement of claims because it can provide us the information on the unobserved heterogeneity of policyholders. For example, suppose a situation that there are two identical groups of policyholders in terms of observed covariates. If one group turns out to have very high claim for consecutive 3 years whereas the other has no claim for 3 years, then we may suspect that the observed covariates might not be enough for capturing the risk of each policyholder so that we need to incorporate the concept of random effects on the claim occurence, which is analogous to the unobserved heterogeneity in the risk of policyholders. For detailed examples of capturing unobserved heterogeneity via random effects, see Frees and Kim (2006) or Kim et al. (2017).

In this regard, this article explores the benefits of using random effects for predicting insurance claims observed longitudinally, or over a period of time, within a two-part framework without adhering to the distributionsin exponential family and independence assumption between frequency and severity components. It extends our work in Jeong et al. (2018a), which specified distributional models within the family of generalized linear models, and used Gaussian random effects. For the construction of the dependent random effects model in this paper, we propose the use of the families of distributions with conjugate random effects, which enables us to obtain explicit moments, marginal likelihood, and predictive distributions, as proposed in Lee and Nelder (1996) and Molenberghs et al. (2010). For the frequency part, we use Poisson model with gamma random effects. In our search for such models for claims severity, we utilize the family of GB2 distributions derived with an underlying distribution based on G-gamma and a random effect based on generalized inverse gamma (GI-gamma).

Note that under our proposed dependent compound risk random effects model, we may derive the credibility premium of the compound sum which exploits the longitudinal property as well as adjustments due to the possible dependence between the frequency and severity components. Credibility premium based on random effects has been explored in some actuarial literature. For example, Frees et al. (1999) provided a general framework which integrates well-known credibility models based on the use of linear mixed model. After that, Frangos and Vrontos (2001) suggested the use of conjugate random effects in both frequency and severity component so that the credibility premium of compound sum can be expressed as a product of credibility premiums of frequency and severity, which implicitly assumes the independence between the frequency and severity components. Shevchenko and Wuthrich (2006) also considered similar approach to Frangos and Vrontos (2001) in order to model operation risks using Bayesian posterior premium. However, since we cannot preclude the possible dependence between the frequency and severity, there have been some trial to incorporate the dependence in the calculation of credibility premium of compound sum, such as Hernández-Bastida et al. (2009) and Gómez-Déniz (2016). Our proposed dependent compound risk random effects model enables us to extend both Frangos and Vrontos (2001) and Garrido et al. (2016) and includes them as special cases so that the credibility premium of compound sum can be expressed as a product of not only credibility premiums of frequency and severity, but also $D_N(\gamma)$, which accounts for the possible dependence between frequency and severity components in a flexible way. For example, if there is strong and positive association between frequency and severity components, then $D_N(\gamma)$ is far greater than 1, which means credibility premium for compound sum should be much higher than mere product of credibility premiums of frequency and severity. On the other hand, if there is weak and negative association between frequency and severity components, then $D_N(\gamma)$ is slightly less than 1. Finally, if there is no dependence between frequency and severity components, then $D_N(\gamma)$ is exactly 1, which means credibility premium for compound sum is the same as product of credibility premiums of frequency and severity, as in Frangos and Vrontos (2001).

For the calibration of proposed models, we used a longitudinal claim dataset from a Singapore automobile insurance company, which includes both claim observations and policyholder characteristics for multi-year. This dataset has been used for various actuarial research, including Antonio and Valdez (2012) and Jeong et al. (2018b).

Organization of this paper is as follows. In Section 2, we introduce general concept of the dependent compound risk random effects model and our model specifications for the frequency and average severity components so that our paper can be self-contained and provide the notation used throughout the paper. In Section 3, we provide our main theoretical result on the derivation of credibility premium of compound sum without assuming independence between frequency and severity component. In Section 4, we describe our data with some preliminary investigation. In Section 5, we analyze the estimation results and introduce some model validation methods we used in here as well. In the end, in Section 6 we conclude this paper with some remarks. The appendices are attached for showing the detailed calculation of the the credibility mean of the compound sum of claim, and both marginal and predictive densities of the average severity for GP and GB2 distribution.

2 The dependent compound risk random effects model

Garrido et al. (2016) incorporated the dependence between loss frequency and loss severity in their model as follows. According to their framework, we can consider an insurer's portfolio as a cross-sectional data (in other words, for a fixed time period which is usually a single year). Here N refers to the number of claims and the size of claims are denoted by $C_1, C_2, \ldots, C_N$. Then, total loss can be expressed as follows:

$$S = C_1 + C_2 + \cdots + C_N.$$

Conventionaly, when $N = 0$, the total loss $S = 0$ as well. Only in case of $N > 0$, we define the average of loss severity as $\overline{C} = S/N$, which leads to the expression of the aggregate loss as $S = N\overline{C}$. Moreover, let us denote $\mathbf{x} = (x_1, \ldots, x_p)$ as a set of p explanatory variables, then we can introduce the dependence between the loss frequency and the average of loss severity as follows.

If we set link functions g_N and g_C for frequency and severity in GLMs, respectively, then the conditional expectation of the loss frequency and the average of loss severity is given by

$$\nu = \mathbb{E}\left[N|\mathbf{x}\right] = g_N^{-1}(\mathbf{x}\alpha) \qquad \text{and} \qquad \mu_\gamma = \mathbb{E}\left[\overline{C}|\mathbf{x}, N\right] = g_C^{-1}(\mathbf{x}\beta + \gamma N). \tag{1}$$

Although Garrido et al. (2016) set a foundation of two-part dependent GLM, their work is limited to cross-sectional data. Indeed, it is certain that a typical property and casualty (P&C) insurance policies portfolio is in a longitudinal format. In other words, it contains $(N_{it}, C_{itj}, \mathbf{x}_{it}, e_{it})'$ as observations of independent policyholders for calendar year $t = 1, \ldots, T_i$ and for policyholder $i = 1, \ldots, M$. Now we can fix a number T so that $T_i \leq T$ and it does not preclude us from allowing unbalanced data. As usual, $\mathbf{x}_{it}$ refers to the covariates vector which describes characteristics of each policyholder and $e_{it} \in (0, 1]$ refers to the length of exposure of the policyholder within calendar year t because in some cases a policyholder may not have a full exposure for given calender year.

From now on, N_{it} stands for the number of claims and C_{itj} denotes the observed claim size where j is additionally required to distinguish the multiple claims that may happen per calendar year so that $j = 1, \ldots, N_{it}$. For each calendar year t, we specify the claim severity distribution by defining $\overline{C}_{it}$ as follows provided $N_{it} > 0$.

$$\overline{C}_{it} = \frac{1}{N_{it}} \sum_{j=1}^{N_{it}} C_{itj}. \tag{2}$$

Thus, the joint density of our dependent compound risk random effects model is given as

$$f(n_{it}, \overline{c}_{it}|\theta_i^N, \theta_i^C) = f_N(n_{it}|\theta_i^N) \times f_{\overline{C}|N}(\overline{c}_{it}|\theta_i^C, n_{it}). \tag{3}$$

f_N and $f_{\overline{C}|N}$ refer to the density functions for frequency and average severity, respectively. Likewise, θ_i^N and θ_i^C refer to the random effects for the frequency and average severity of policyholder i with corresponding prior densities π_N and π_C, respectively. The construction in (3) is similar to the basic two-part model of frequency and severity, while our specification allows dependence between frequency and severity as well as the presence of random effects per each policyholder. It is understood that when in case of $N = 0$, it is conventional to simplify the joint distribution as $f(0, 0) = f_N(0)$, which represents the zero claims probability. Furthermore, we consider dependence between the frequency and average severity by using the number of claims N as a linear predictor in the mean function for the average severity in our specification of this dependent compound risk random effects model as defined in (1).

Now we can define the compound sum as

$$S_{it} = \sum_{j=1}^{N_{it}} C_{itj} = N_{it}\overline{C}_{it} \tag{4}$$

in order to refer to the aggregate claims for policyholder i in calendar year t. Let us denote $\mathbf{n}_T = (n_1, n_2, \ldots, n_T)$

and $\bar{\mathbf{c}}_T = (\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_T)$. Note that due to the presence of random effects both in frequency and severity, predictive distribution of frequency component can be expressed as follows after suppressing the subscript i:

$$
\begin{aligned}
f(n_{T+1}|\mathbf{n}_T) &= \frac{f(n_1, \ldots, n_T, n_{T+1})}{f(n_1, \ldots, n_T)} = \frac{\int \left(\prod_{t=1}^{T+1} f(n_t|\theta)\right) \pi_N(\theta) d\theta}{f(n_1, \ldots, n_T)} \\
&= \int f(n_{T+1}|\theta) \frac{\pi_N(\theta) \prod_{t=1}^{T} f(n_t|\theta)}{f(n_1, \ldots, n_T)} d\theta = \int f(n_{T+1}|\theta) \pi_N(\theta|\mathbf{n}_T) d\theta.
\end{aligned}
$$

Likewise, predictive distribution of the average severity component can be expressed as follows after suppressing the subscript i and conditioning argument on n:

$$
\begin{aligned}
f(\bar{c}_{T+1}|\bar{\mathbf{c}}_T) &= \frac{f(\bar{c}_1, \ldots, \bar{c}_T, \bar{c}_{T+1})}{f(\bar{c}_1, \ldots, \bar{c}_T)} = \frac{\int \left(\prod_{t=1}^{T+1} f(\bar{c}_t|\theta)\right) \pi_C(\theta) d\theta}{f(\bar{c}_1, \ldots, \bar{c}_T)} \\
&= \int f(\bar{c}_{T+1}|\theta) \frac{\pi_C(\theta) \prod_{t=1}^{T} f(\bar{c}_t|\theta)}{f(\bar{c}_1, \ldots, \bar{c}_T)} d\theta = \int f(\bar{c}_{T+1}|\theta) \pi_C(\theta|\bar{\mathbf{c}}_T) d\theta.
\end{aligned}
$$

Therefore, based on the predictive distributions of both frequency and the average severity components, predictive mean of S_{T+1} can be expressed as follows:

$$
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \bar{\mathbf{c}}_T\right].
$$

If we assume that $\mathbb{E}\left[\overline{C}_t|N_t\right] = \hat{\mu}_t\, e^{\gamma N_t}$ such that $\hat{\mu}_t$ is independent to N_t, then we can obtain the following predictive mean of S_{T+1}:

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \\
&= \hat{\mu}_{T+1} \times \mathbb{E}\left[N_{T+1} e^{\gamma N_{T+1}}|\mathbf{n}_T\right].
\end{aligned} \tag{5}
$$

2.1 Frequency component model specification

For the frequency N_{it}, we have the following two candidate models where we denote $\nu_{it} = e^{\mathbf{x}_{it}\alpha}$ with α as a $p \times 1$ parameter vector for the covariates associated with frequency for which they may be different from that of the average severity.

(1) Simple Poisson GLM: $N_{it} \sim \mathcal{P}(\nu_{it})$

(2) Poisson/Gamma Random Effect Model (= Multivariate NB Model)

Now let us explain the Poisson/gamma random effect model, which is given as

$$
N_{it}|\theta_i^N \sim\ \mathcal{P}(\nu_{it}\theta_i^N) \text{ and } \theta_i^N \sim\ \mathcal{G}(r, 1/r) \tag{6}
$$

so that $\mathbb{E}\left[N_{it}|\theta_i^N\right] = \mathbb{V}ar\left(N_{it}|\theta_i^N\right) = \nu_{it}\theta_i^N$. From above specification, we can see that $N_{it} \sim \mathcal{NB}\left(r, \dfrac{\nu_{it}}{r+\nu_{it}}\right)$ and the following relations hold:

$$
\begin{aligned}
&f_N(n) = \binom{n+r-1}{n}\left(\frac{r}{r+\nu_{it}}\right)^r\left(\frac{\nu_{it}}{r+\nu_{it}}\right)^n, \\
&\mathbb{E}\left[N_{it}\right] = \mathbb{E}\left[\mathbb{E}\left[N_{it}|\theta_i^N\right]\right] = \mathbb{E}\left[\theta_i^N\nu_{it}\right] = \nu_{it}, \quad \mathbb{V}ar\left(N_{it}\right) = \mathbb{V}ar\left(\mathbb{E}\left[N_{it}|\theta_i^N\right]\right) + \mathbb{E}\left[\mathbb{V}ar\left(N_{it}|\theta_i^N\right)\right] = \nu_{it}\left(1+\frac{\nu_{it}}{r}\right)
\end{aligned}
\tag{7}
$$

Although here we have $N_{it} \sim \mathcal{NB}\left(r, \dfrac{\nu_{it}}{r+\nu_{it}}\right)$, our model specification is different from usual negative binomial GLM for frequency component because N_{it} is not marginally independent. Let $\pi_N(\theta)$ be the probability density with respect to gamma distribution. Then according to Boucher et al. (2008), it is known that joint density for the multi-year claim count is given as follows, which is called multivariate negative binomial (MVNB) distribution.

$$
f_{N_i}(n_{i1}, \ldots, n_{iT_i}) = \int \prod_{t=1}^{T_i} f_{N|\theta^N}(n_{it}|\theta)\pi_N(\theta)d\theta = \prod_{t=1}^{T_i}\left(\frac{e^{\mathbf{x}_{it}\alpha}}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha}+r}\right)^{n_{it}}\left(\frac{r}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha}+r}\right)^r \frac{\Gamma\left(\sum\limits_{t=1}^{T_i} n_{it}+r\right)}{\Gamma(r)\prod\limits_{t=1}^{T_i} n_{it}!}
$$

Note that we have to account for the exposure e_{it} and this can be done by incorporating e_{it} as an offset to the mean parameter. With derived joint density of MVNB distribution, we can estimate $\hat{\alpha}$ by maximizing the following likelihood function:

$$
\begin{aligned}
\ell_N &= \sum_{i=1}^{M}\left(\log\int\prod_{t=1}^{T_i} f_{N|\theta^N}(n_{it}|\theta)\pi_N(\theta)d\theta\right) \\
&= \sum_{i=1}^{M}\sum_{t=1}^{T_i}\left[n_{it}\mathbf{x}_{it}\alpha - \log\Gamma(n_{it}+1)\right] + \sum_{i=1}^{M}\log\Gamma\left(\sum_{t=1}^{T_i} n_{it}+r\right) - \\
&\quad \sum_{i=1}^{M}\left[\left(\sum_{t=1}^{T_i} n_{it}+r\right)\log\left(\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}\alpha}+r\right)\right] + M\left[r\log r - \log\Gamma(r)\right]
\end{aligned}
\tag{8}
$$

Finally, it is not difficult to show that predictive distribution of N_{T+1} given $\mathbf{n}_T$ still follows negative binomial distribution as follows:

Lemma 1. *Suppose $N_1, N_2, \ldots, N_t$ follows MVNB distribution as defined in (6). Then*

$$
N_{T+1}|\mathbf{n}_T \sim \mathcal{NB}\left(r_T,\, \frac{\nu_{T+1}}{\tilde{r}_T+\nu_{T+1}}\right),
$$

where $r_T = r + \sum_{t=1}^{T} n_t$ and $\tilde{r}_T = r + \sum_{t=1}^{T} \nu_t$.

Proof. From (7), we can see that if $N_t|\theta^N \sim \mathcal{P}(\nu_t\theta^N)$ and $\theta^N \sim \mathcal{G}(r, 1/r)$, then $N_t \sim \mathcal{NB}\left(r, \dfrac{\nu_t}{r+\nu_t}\right)$. Furthermore, it is easy to show that $\theta^N|\mathbf{n}_t \sim \ \mathcal{G}(r_T, 1/\tilde{r}_T)$ because

$$
\begin{aligned}
\pi_N(\theta|\mathbf{n}_T) &\propto \pi_N(\theta)\prod_{t=1}^{T} f(n_t|\theta) \propto \theta^{r-1}\exp(-r\theta)\prod_{t=1}^{T}\theta^{n_t}\exp(-\nu_t\theta) \\
&\propto \theta^{\sum_{t=1}^{T} n_t + r - 1}\exp\left(-\left[r+\sum_{t=1}^{T}\nu_t\right]\theta\right) = \theta^{r_T-1}\exp\left(-\tilde{r}_T\theta\right).
\end{aligned}
$$

Therefore, $N_{T+1}|\mathbf{n}_T \sim \mathcal{NB}\left(r_T,\ \dfrac{\nu_{T+1}}{\tilde{r}_T+\nu_{T+1}}\right)$ and $\mathbb{E}\left[N_{T+1}|\mathbf{n}_T\right] = \dfrac{r_T}{\tilde{r}_T}\nu_{T+1} = \dfrac{r+\sum_{t=1}^{T} n_t}{r+\sum_{t=1}^{T}\nu_t}\nu_{T+1}$. □

2.2 Severity component model specification

For the average severity $\overline{C}_{it}|N_{it}$, we have four candidate models where we denote $\mu_{it} = e^{\mathbf{x}_{it}\beta + N_{it}\gamma}$ with β as a $p \times 1$ parameter vector for the covariates associated with average severity and γ is parameter which is used to measure the dependency between frequency and the average severity components.

(1) Simple Gamma GLM: $\overline{C}_{it}|N_{it} \sim \mathcal{G}\left(\dfrac{N_{it}}{\phi}, \mu_{it}\dfrac{\phi}{N_{it}}\right)$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it}\right] = \mu_{it}$ and $\dfrac{\mathrm{V}ar\left(\overline{C}_{it}|N_{it}\right)}{\mathbb{E}\left[\overline{C}_{it}|N_{it}\right]^2} = \dfrac{\phi}{N_{it}}$

(2) Gamma/Normal Random Effect Model (= Gamma GLMM):
$\overline{C}_{it}|N_{it}, \theta_i^C \sim \mathcal{G}\left(\dfrac{N_{it}}{\phi}, \theta_i^C\mu_{it}\dfrac{\phi}{N_{it}}\right)$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it}, \theta_i^C\right] = \theta_i^C\mu_{it}$ and $\dfrac{\mathrm{V}ar\left(\overline{C}_{it}|N_{it}, \theta_i^C\right)}{\mathbb{E}\left[\overline{C}_{it}|N_{it}, \theta_i^C\right]^2} = \dfrac{\phi}{N_{it}}$, where $\log\theta_i^C \sim \mathcal{N}\left(-\dfrac{\sigma^2}{2}, \sigma^2\right)$ so that $\mathbb{E}\left[\theta_i^C\right] = 1$.

(3) Gamma/Inv-gamma Random Effect Model (= MVGP Model)

(4) G-gamma/GI-gamma Random Effect Model (= MVGB2 Model)

Since our proposed severity models are MVGP model and MVGB2 model, let us examine them more carefully. First, gamma/inv-gamma random effect model is given as follows:

$$
\overline{C}_{it}|N_{it}, \theta_i^C \sim \ \mathcal{G}\left(\frac{N_{it}}{\phi}, \theta_i^C\mu_{it}\frac{\phi}{N_{it}}\right) \text{ and } \theta_i^C \sim \mathcal{IG}(k+1, k)
$$

so that $\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right] = \theta_i^C\mu_{it}, \mathrm{V}ar\left(\bar{C}_{it}|N_{it}, \theta_i^C\right) = (\theta_i^C)^2\mu_{it}^2\dfrac{\phi}{N_{it}}$. From above specification, we can see that $\bar{C}_{it}|N_{it}$ follows a generalized patero (GP) distribution, in other words, $\bar{C}_{it}|N_{it} \sim \mathcal{GP}\left(k+1, \mu_{it}k\dfrac{\phi}{N_{it}}, \dfrac{N_{it}}{\phi}\right)$ and the following relations hold:

$$
\begin{aligned}
\mathbb{E}\left[\bar{C}_{it}|N_{it}\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right]\right] = \mathbb{E}\left[\theta_i^C\mu_{it}\right] = \mu_{it}, \\
\mathrm{V}ar\left(\bar{C}_{it}|N_{it}\right) &= \mathrm{V}ar\left(\mathbb{E}\left[\bar{C}_{it}|N_{it}, \theta_i^C\right]\right) + \mathbb{E}\left[\mathrm{V}ar\left(\bar{C}_{it}|N_{it}, \theta_i^C\right)\right] = \frac{\mu_{it}^2}{k-1}\left(1+\frac{k\phi}{N_{it}}\right).
\end{aligned}
$$

Note that generalized Pareto distribution is widely used in ratemaking application by itself, as mentioned in Klugman et al. (2012).

Let $\pi_C(\theta)$ be the probability density with respect to inverse gamma distribution. Then we can get the marginal likelihood for the average severity as follows, which can be called multivariate generalized Pareto (MVGP) distribution and gives a natural association structure within the claims of policyholder i:

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\overline{c}_{i1},\ldots,\overline{c}_{iT_i}|n_i) &= \int \prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta \\
&= \frac{k^{k+1}\displaystyle\prod_{t=1}^{T_i}\left(n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}/\phi\right)^{n_{it}/\phi}}{\left(k+\displaystyle\sum_{t=1}^{T_i} n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}/\phi\right)^{\sum_{t=1}^{T_i} n_{it}/\phi+k+1}} \times \frac{\Gamma\left(\displaystyle\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right)\displaystyle\prod_{t=1}^{T_i}\overline{c}_{it}^{-1}}{\Gamma(k+1)\displaystyle\prod_{t=1}^{T_i}\Gamma(n_{it}/\phi)}.
\end{aligned}
$$

Note that detail of deriviation is provided in Appendix B. Now using the joint density function for $\overline{c}_{i1},\ldots,\overline{c}_{iT_i}|\mathbf{n}_{T_i}$, we can estimate $\hat{\phi}, \hat{\beta}$ and $\hat{\gamma}$ by maximizing the following likelihood function:

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_{i=1}^{M}\left(\log\int\prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\overline{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta\right) \\
&= \sum_{i=1}^{M}\sum_{t=1}^{T_i}\left[-\log\Gamma(n_{it}/\phi)-\log\overline{c}_{it}\right]+\sum_{i=1}^{M}\log\Gamma\left(\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right) \\
&\quad +\sum_{i=1}^{M}\sum_{t=1}^{T_i} n_{it}/\phi(\log n_{it}\overline{c}_{it}-\mathbf{x}_{it}\beta-n_{it}\gamma-\log\phi) \\
&\quad -\sum_{i=1}^{M}\left[\left(\sum_{t=1}^{T_i} n_{it}/\phi+k+1\right)\log\left(k+\sum_{t=1}^{T_i} n_{it}\overline{c}_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}/\phi\right)\right]+M\left[(k+1)\log k-\log\Gamma(k+1)\right].
\end{aligned} \tag{9}
$$

Likewise, G-gamma/GI-gamma random effect model is given as follows by denoting $w=\dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$:

$$
\overline{C}_{it}|N_{it},\theta_i^C \sim \mathcal{GG}\left(\frac{N_{it}}{\phi},\theta_i^C\mu_{it}\frac{\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)},p\right) \quad\text{and}\quad \theta_i^C \sim \mathcal{GIG}(k+1,w,p)
$$

so that $\mathbb{E}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]=\theta_i^C\mu_{it}$ and $\mathrm{V}ar\left(\bar{C}_{it}|N_{it},\theta_i^C\right)=(\theta_i^C)^2\mu_{it}^2\left(\dfrac{\Gamma(N_{it}/\phi+2/p)\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)^2}-1\right)$. From above specification, we can see that $\bar{C}_{it}|N_{it}$ follows a second-kind of generalized beta (GB2) distribution, in other words, $\bar{C}_{it}|N_{it}\sim\mathcal{GB}2\left(k+1,\, w\mu_{it}\dfrac{\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)},\,\dfrac{N_{it}}{\phi},\,p\right)$ and the following relations hold:

$$
\begin{aligned}
\mathbb{E}\left[\bar{C}_{it}|N_{it}\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]\right]=\mathbb{E}\left[\theta_i^C\mu_{it}\right]=\mu_{it}, \\
\mathrm{V}ar\left(\bar{C}_{it}|N_{it}\right) &= \mathrm{V}ar\left(\mathbb{E}\left[\bar{C}_{it}|N_{it},\theta_i^C\right]\right)+\mathbb{E}\left[\mathrm{V}ar\left(\bar{C}_{it}|N_{it},\theta_i^C\right)\right]=\mathrm{V}ar\left(\theta_i^C\mu_{it}\right)+\mathbb{E}\left[(\theta_i^C)^2\mu_{it}^2\left(\frac{\Gamma(N_{it}/\phi+2/p)\Gamma(N_{it}/\phi)}{\Gamma(N_{it}/\phi+1/p)^2}-1\right)\right] \\
&= \mu_{it}^2\left[\frac{\Gamma(k+1-\frac{2}{p})\Gamma(k+1)}{\Gamma(k+1-\frac{1}{p})^2}\frac{\Gamma(\frac{N_{it}}{\phi}+\frac{2}{p})\Gamma(\frac{N_{it}}{\phi})}{\Gamma(\frac{N_{it}}{\phi}+\frac{1}{p})^2}-1\right].
\end{aligned}
$$

Let $\pi_C(\theta)$ be the probability density with respect to generalized inverse gamma distribution. Then one can get the following marginal likelihood for the average severity, which can be called multivariate generalized beta-II (MVGB2) distribution introduced by Yang et al. (2011), a flexible parametric distribution with four parameters describing scale and various shapes, and gives a natural association structure within the claims of policyholder i as well.

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\bar{c}_{i1},\ldots,\bar{c}_{iT_i}|n_i) &= \int \prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\bar{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta \\
&= \frac{w^{pk+p}\prod_{t=1}^{T_i}\left(\bar{c}_{it}z_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma}\right)^{pn_{it}/\phi}}{\left(w^p+\sum_{t=1}^{T_i}(\bar{c}_{it}z_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma})^p\right)^{\sum_{t=1}^{T_i}n_{it}/\phi+k+1}} \times \frac{\Gamma\left(\sum_{t=1}^{T_i}n_{it}/\phi+k+1\right)\prod_{t=1}^{T_i}\bar{c}_{it}^{-1}p^{T_i}}{\Gamma(k+1)\prod_{t=1}^{T_i}\Gamma(n_{it}/\phi)}
\end{aligned}
$$

Note that detail of deriviation is provided in Appendix B. One can see that GB2 distribution has various advantages besides the flexibility, such as the presence of explicit moment, cumulative distribution function and joint likelihood, better fit on the tail part, and relationship with other well-known distribution. For example, when the power parameter p equals to 1 in GB2 distribution, then it is reduced to aforementioned generalized Pareto (GP) distribution.

Now using the joint density function for $\bar{c}_{i1},\ldots,\bar{c}_{iT_i}|n_i$, we can estimate $\hat{\phi},\hat{\beta},\hat{\gamma}$ and $\hat{p}$ by maximizing the following likelihood function where $w=\dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, and $z_{it}=\dfrac{\Gamma(n_{it}/\phi+1/p)}{\Gamma(n_{it}/\phi)}$:

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_{i=1}^{M}\left(\log\int\prod_{t=1}^{T_i} f_{\overline{C}|N,\theta^C}(\bar{c}_{it}|n_{it},\theta)\pi_C(\theta)d\theta\right) \\
&= \sum_{i=1}^{M}\left[\sum_{t=1}^{T_i}\left(-\log\Gamma\left(\frac{n_{it}}{\phi}\right)-\log\left(\frac{\bar{c}_{it}}{p}\right)\right)+\log\Gamma\left(\sum_{t=1}^{T_i}\frac{n_{it}}{\phi}+k+1\right)\right]+p\sum_{i=1}^{M}\sum_{t=1}^{T_i}n_{it}/\phi(\log\bar{c}_{it}z_{it}-\mathbf{x}_{it}\beta-n_{it}\gamma) \\
&\quad -\sum_{i=1}^{M}\left[\left(\sum_{t=1}^{T_i}n_{it}/\phi+k+1\right)\log\left(w^p+\sum_{t=1}^{T_i}(\bar{c}_{it}z_{it}e^{-\mathbf{x}_{it}\beta-n_{it}\gamma})^p\right)\right]+M\left[(k+1)p\log w-\log\Gamma(k+1)\right].
\end{aligned} \tag{10}
$$

3 Credibility premium of compound sum with dependence

Once we assume the dependence between the frequency and the average severity, then it might not be appropriate to naively use the product of estimated values of frequency and average severity as the estimated value of compound sum. Furthermore, it is desirable to consider the dependence among the compound sums of the same policyholder, due to unobserved heterogeneity. Our proposed dependent two-part random effects model enables us to obtain the credibility premium of S_{T+1} given information on $\mathbf{n}_T$, $\bar{\mathbf{c}}_T$ as follows:

Theorem 1. *Suppose $(N_1,N_2,\ldots,N_t)$ follows MVNB distribution as defined in (6). Moreover, let us assume that*

$\mathbb{E}\left[\overline{C}_t|n_t\right] = \theta^C e^{\mathbf{x}_t\beta}\, e^{n_t\gamma} = \theta^C \tilde{\mu}_t\, e^{n_t\gamma}$. *Then, the credibility premium of S_{T+1} is given as follows:*

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1} \\
&= \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right] \times D_N(\gamma).
\end{aligned} \tag{11}
$$

If $\gamma = 0$, then (11) is reduced as follows:

$$
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} = \mathbb{E}\left[\mu_{T+1}\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \times \mathbb{E}\left[\nu_{T+1}\theta^N|\mathbf{n}_T\right],
$$

which is suggested by Frangos and Vrontos (2001).

Proof. The proof is provided in Appendix A. □

Note that if $r \to \infty$, then MVNB distribution defined in (6) converges to naive independent Poisson model and the formula for credibility premium of S_{T+1} in (11) is reduced as follows:

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]\, \tilde{\mu}_{T+1} \times \lim_{r\to\infty}\left\{\frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}\right\} \\
&= \mathbb{E}\left[\tilde{\mu}_{T+1}\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \times \nu_{T+1} \times \exp\left(\nu_{T+1}(e^{\gamma}-1)+\gamma\right).
\end{aligned}
$$

According to Theorem 1, it can be observed that the credibility premium of S_{T+1} can be expressed as the product of credibility premium of N_{T+1}, credibility premium of $\overline{C}_{T+1}$, and adjustment factor $D_N(\gamma)$ which accounts for the dependence between frequency and the average severity. Therefore, we may obtain expressions of $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]$ in all four models for the average severity component as follows:

Corollary 1. *Suppose $(N_1, N_2, \ldots, N_t)$ follows MVNB distribution as defined in (6). If the average severity component follows Gamma GLM, then the credibility premium of S_{T+1} is given as follows:*

$$
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}.
$$

If the average severity component follows Gamma GLMM, then we have

$$
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \simeq \frac{\sum_{j=1}^J \hat{\theta}_j\, f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)}{\sum_{j=1}^J f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)}\tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1},
$$

where $\hat{\theta}_j$'s are generated from $\mathcal{LN}\left(-\sigma^2/2,\, \sigma^2\right)$.

If the average severity component follows MVGP Model, then we have

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \frac{k\phi + \sum_{t=1}^{T} S_t/\mu_t}{k\phi + \sum_{t=1}^{T} n_t}\tilde{\mu}_{T+1} \times \frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t}\nu_{T+1} \times e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T-1}.$$

Finally, if the average severity component follows MVGB2 Model, then we have

$$\begin{aligned}\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = &\left(\sum_{t=1}^{T}\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi + 1/p)}{\phi\Gamma(n_t/\phi + 1)}\right)^{p} + w^{p}\right)^{1/p}\frac{\Gamma(k_T + 1 - 1/p)}{\Gamma(k_T + 1)}\tilde{\mu}_{T+1}\times \\ &\frac{r + \sum_{t=1}^{T} n_t}{r + \sum_{t=1}^{T} \nu_t}\nu_{T+1} \times e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T-1}.\end{aligned}$$

Proof. The proof is provided in Appendix A. □

Note that the credibility premium formula for S_{T+1} in (11) is very flexible and proper extension of the formulas for mean of compound sum in previous literature such as Frangos and Vrontos (2001) and Garrido et al. (2016). For example, if a policyholder is a newcomer so that $T = 0$, in other words, there is no claim history of the policyholder, then (11) is reduced to

$$\mathbb{E}\left[S_{0+1}|\mathbf{n}_0, \bar{\mathbf{c}}_0\right] = \mathbb{E}\left[S_1\right] = \mathbb{E}\left[\theta^{C}\tilde{\mu}_1\right] \times \mathbb{E}\left[N_{T+1}e^{N_1\gamma}\right] = \tilde{\mu}_1 \times \nu_1 \times e^{\gamma}\left[1 - \left(\frac{\nu_1}{r}\right)(e^{\gamma} - 1)\right]^{-r-1},$$

which is also reduced to $\tilde{\mu}_1\nu_1 \exp\left(\nu_1(e^{\gamma} - 1) + \gamma\right)$ as $r \to \infty$. Finally, if a policyholder has been observed for T years but had no claim, then $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[S_{T+1}|\mathbf{n}_T = 0\right]$ so that

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \mathbb{E}\left[S_{T+1}|\mathbf{n}_T = 0\right] = \tilde{\mu}_{T+1}\frac{r}{r + \sum_{t=1}^{T} \nu_t}\nu_{T+1} \times e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T-1},$$

which means the policyholder gets sure discount on his/her policy due to the favorable claim history.

The following theorem tells us that $D_N(\gamma)$ is an informative metric to measure the dependence between frequency and severity components.

Theorem 2. *For $D_N(\gamma)$ in (11), the following are true:*

(i) $D_N(\gamma)$ is well-defined if and only if $\gamma \leq \log(1 + \tilde{r}_T/\nu_{T+1})$.

(ii) $D_N(\gamma)$ is a strictly increasing function of γ.

(iii)

$$D_N(\gamma) = e^{\gamma}\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma} - 1)\right]^{-r_T-1}\begin{cases} = 1 & \text{if} \quad \gamma = 0 \\ > 1 & \text{if} \quad \gamma > 0 \\ < 1 & \text{if} \quad \gamma < 0 \end{cases}. \tag{12}$$

Proof. (i) One can see that $D_N(\gamma)$ is well-defined if and only if $1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1) \geq 0$, which is equivalent to $\gamma \leq \log(1 + \tilde{r}_T/\nu_{T+1})$.

(ii) First, $g(\gamma) = 1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)$ is a strictly decreasing function of γ. Then one can see that $\left[1 - \left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T-1} = g(\gamma)^{-r_T-1}$ is increasing function of γ since $r_T > 0$. Finally, it is clear that $D_N(\gamma)$ is a product of two strictly increasing functions of γ, $g(\gamma)^{-r_T-1}$ and e^γ.

(iii) It follows from (ii) and $D_N(0) = 1$.

□

From (iii) of Theorem 2, if there is a positive (negative) dependence between frequency and severity components, then the credibility premium of compound sum would be greater (less) than mere products of credibility premium for each components so that one needs to compensate the difference by multiplying an adjustment factor for dependence which is greater (less) than 1 by the products of premium. Further, $D_N(\gamma)$ can measure not only the sign of dependence but also the magnitude of dependence as we can see in (ii) of Theorem 2. Finally, although the upper bound for feasible γ is $\log(1 + \tilde{r}_T/\nu_{T+1})$, it is not problematic in practice because usually ν_{T+1} is around 0.1 and $\tilde{r}_T \geq 2.3$ in our model specification. Therefore, as long as $\gamma \leq 3.178054 = \ln(1 + 2.3/0.1)$, $D_N(\gamma)$ is well-defined, which is satisfied with all observations in the following empirical analysis.

4 Data description

To calibrate the proposed models in this article, we used a dataset of a Singapore automobile insurance company, which contains both policy characteristic and claims experience of the policyholders. The dataset consists of observations of ten years, 1993–2002. The dataset is provided by General Insurance Association of Singapore, which represents every general insurance companies operated in Singapore during that period. The same dataset or sampled data has been used in some actuarial research articles, including but not limited to Frees and Valdez (2008) and Shi and Valdez (2012). In the dataset, the observations from the first eight years, 1993-2000 were used as a training set for estimating the associated parameters in each model, whereas the observations from the last two years, 2001–2002 were used for out-of-sample validation. In Table 1, the summary statistics only summarize the training set. There are $M = 50,215$ unique policyholders who are tracked T_i years. It is natural that the maximum value of T_i is 8.

The observed policy characteristics were used as covariates in both components, frequency and the average severity. The summarized results of the covariates information and descriptions are provided in Table 1. In total, nine variables were used as covariates, which can be either categorical or continuous, which contain both driver and vehicle information. Gender and issue age are related to driver information, and the others are related to vehicle information. Although the table is self-contained, we provide a few remarks to explain the dataset more clearly. First, the proportion of female drivers in Singapore is quite lower than that in United States, as shown in Table 1. Second, there are few observation

Table 1: Observable policy characteristics used as covariates in the training set

Categorical variables	Description		Proportions	
VehType	Type of insured vehicle:	Car	99.27%	
		MotorBike	0.47%	
		Others	0.26%	
Gender	Insured's sex:	Male = 1	80.82%	
		Female = 0	19.18%	
CoverCode	Type of insurance cover:	Comprehensive = 1	78.65%	
		Others = 0	21.35%	
Continuous variables		Minimum	Mean	Maximum
VehCapa	Insured vehicle's capacity in cc	10.00	1587.44	9996.00
VehAge	Age of vehicle in years	-1.00	6.71	48.00
Age	The policyholder's issue age	18.00	44.46	99.00
NCD	No Claim Discount in %	0.00	35.67	50.00

on insured motorbike but they are not ignorable in terms of risk characteristic. Finally, VehAge is defined as the difference of issue year from the model year of the insured car. Therefore, it is not unusual to have -1 as an observed value of VehAge because it is possible to purchase a car in 2018, which has 2019 as its model year.

Table 2: Measures for assessing correlation between the frequency and the (log) average severity

	Pearson	Kendall	Spearman
Estimate	0.04052	0.04227	0.05201
p-value	0.00000	0.00000	0.00000

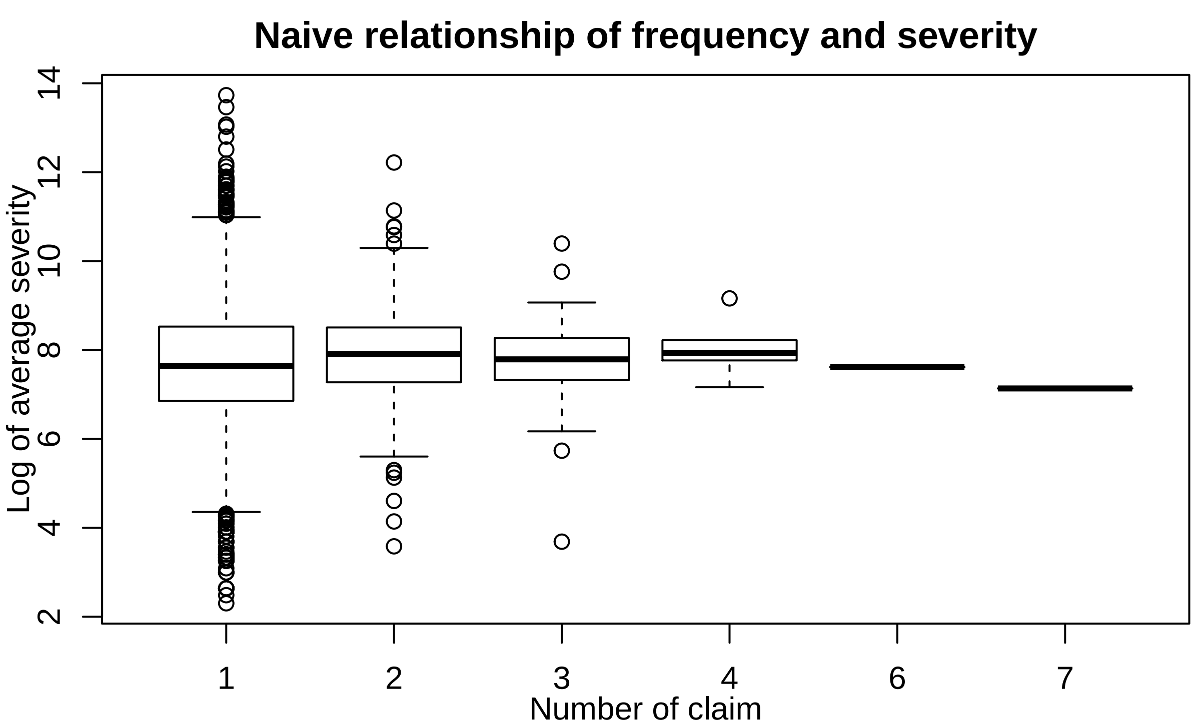

Figure 1: Naive relationship of frequency and severity

As a preliminary observation, we may consider the idea of correlation measure. Table 2 provides the values of estimated correlation measures, which include Pearson coefficient, Kendall's tau, and Spearman's rho. In all correlation measure, we can see there is strong positive correlation between the frequency and (logarithm of) the average severity, which is also shown in Figure 1. Intuitively, this result is reasonable because we can think the common risk characteristic of each policyholder affects both the frequency and the average severity. However, this might be only a preliminary investigation since we did not control any effects of heterogeneous characteristics of each policyholder with covariates.

It is well known that Poisson with gamma random effects leads to a negative binomial which usually fits better than a simple Poisson model, especially with the presence of overdispersion. In our training data, the overall average number of claims is 0.0941367 whereas the variance is 0.1024172 which indicate a possibility of overdispersion. This idea can be validated with a goodness-of-fit test for the Poisson in comparison to the negative binomial, which is shown in Table 3.

Table 3: Goodness-of-fit test for the frequency component

Count	Observed	Poisson	Negative Binomial
0	148198	147608.6	148193.5
1	12788	13895.4	12809.9
2	1109	654	1077.6
3	77	20.5	89.8
4	5	0.5	7.5
5	2	0	0.6
χ^2		104550.3	104079.2

Figure 2 provides log quantile-quantile (log-QQ) plots of fitting the gamma and generalized gamma distribution for the training set without consideration on the observed covariates. Although G-gamma distribution shows slightly better fit on the observed average severities, we need further analysis on the dataset considering both the effects of covariates and longitudinal property of the dataset.

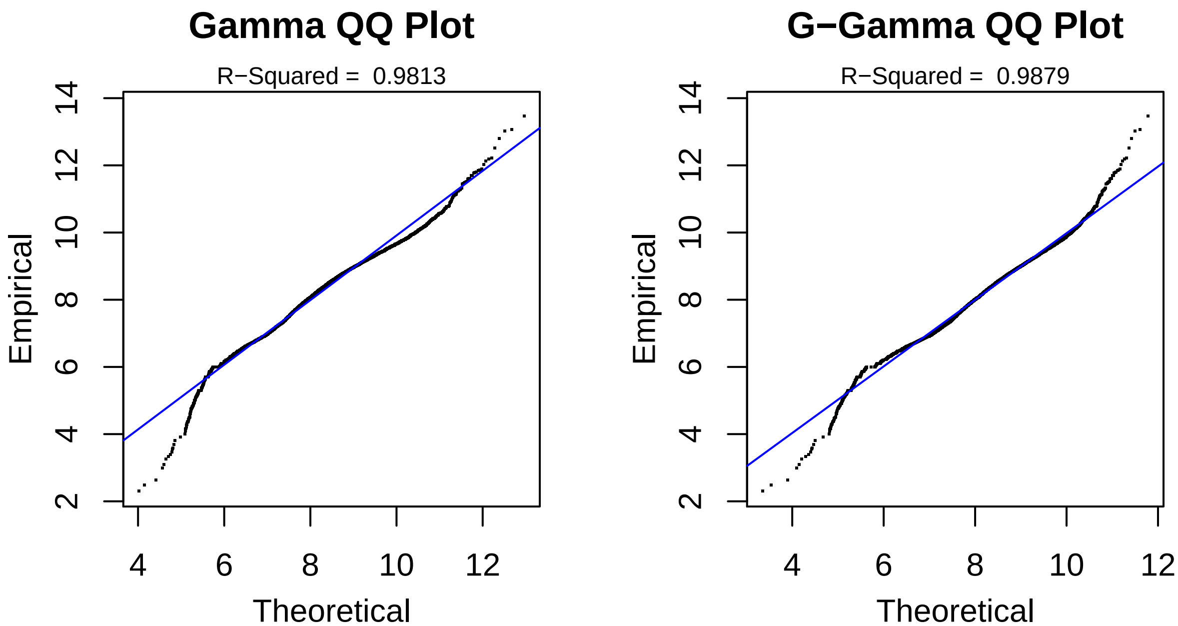

Figure 2: log-QQ plots of fitting gamma and G-gamma to average severity on training set

5 Results of estimation and model validation

Table 4 below provides the details of the estimation results from each of the frequency models described in the subsection 2.1. In the estimation procedure, $\theta^N \sim \mathcal{G}(2.3, 1/2.3)$ has been used as a prior information of θ_N so that the 95% credible interval of θ^N can include $(0.54, 2.00)$, which has been proposed as the range of Belgian bonus-malus system based on the claim frequency in Lemaire (1998). According to this table, the estimates from each model are more or less similar but MVNB model outperforms the naive Poisson model in terms of model selection criteria such as AIC and BIC.

Table 4: Regression estimates of the frequency models

	Poisson			MVNB		
	Estimate	s.e.	Pr(>|t|)	Estimate	s.e.	Pr(>|t|)
(Intercept)	-4.33	0.40	0.00	-4.93	0.52	0.00
VTypeCar	0.19	0.19	0.33	1.44	0.37	0.00
VTypeMBike	-1.41	0.49	0.00	-1.83	0.92	0.05
log(VehCapa)	0.33	0.03	0.00	0.20	0.04	0.00
VehAge	-0.02	0.00	0.00	-0.02	0.00	0.00
SexM	0.11	0.02	0.00	0.09	0.02	0.00
Comp	0.81	0.04	0.00	0.74	0.04	0.00
Age	-0.03	0.02	0.12	-0.01	0.01	0.37
Age ^{2}	0.00	0.00	0.36	0.00	0.00	0.37
Age ^{3}	0.00	0.00	0.77	0.00	0.00	0.34
NCD	-0.01	0.00	0.00	-0.01	0.00	0.00
Loglikelihood		-49565.37			-49494.62	
AIC		99152.75			99013.24	
BIC		99274.70			98989.24	

For the validation of the calibrated frequency models, we compare the actual claim count in the test set with the predicted claim count based on the corresponding observed covariates. As measures of prediction performance, here root-mean-square error (RMSE) and mean absolute error (MAE) are considered. RMSE and MAE measure the discrepancy between the actual loss and predicted loss in terms of L_2 norm and L_1 norm, respectively. Even though there are no big differences in all validation measures as shown in Table 5, consideration of frequency random effects model can be still worthy for incentivizing policyholder to have less accident so that they can get discount on their premium in the following years.

Table 5: Validation measures for the frequency models

	Poisson	MVNB
RMSE	0.28188	0.28183
MAE	0.13749	0.13563

Table 6 below provides the details of the estimation results from each of the average severity models described in the subsection 2.2. In both calibration of MVGP and MVGB2 model, $k = 11$ was used as a hyperparameter so that when $\theta^C \sim \mathcal{IG}(k+1, k)$, 95% credible interval of θ^C can be around $(0.49, 1.62)$, which is much shorter than the 95% credible interval of θ^N. It can be validated because claim severity is not used for bonus-malus system in most of countries except for south Korea, which supports the assertion that there is less variability on the severity component random effects. According to Table 6, both MVGP and MVGB2 model outperform the Gamma GLM and Gamma GLMM in terms of model selection criteria. Note that except for Gamma GLMM, the signs of the coefficients of claim count as a covariate for the average severity are significantly negative, which implies strong negative correlation between frequency and the average severity component, as shown in Frees et al. (2011).

Table 6: Regression estimates of the average severity models

	Gamma GLM		Gamma GLMM		MVGP		MVGB2	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
(Intercept)	7.61	0.00	6.43	0.00	9.49	0.00	9.74	0.00
VTypeCar	-0.29	0.55	0.12	0.62	-0.42	0.00	-0.20	0.07
VTypeMBike	2.87	0.03	2.32	0.00	5.04	0.02	5.94	0.00
logVehCapa	0.53	0.00	0.33	0.00	0.28	0.00	0.22	0.00
VehAge	-0.03	0.00	-0.01	0.00	-0.02	0.00	-0.01	0.00
SexM	-0.01	0.91	-0.02	0.49	-0.05	0.11	-0.09	0.00
Comp	0.05	0.60	0.19	0.00	0.06	0.00	0.23	0.00
Age	-0.16	0.00	-0.05	0.03	-0.16	0.00	-0.17	0.00
Age ^{2}	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Age ^{3}	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
NCD	-0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00
Count	-0.12	0.01	0.01	0.65	-0.09	0.00	-0.07	0.00
Loglikelihood		-138605.00		-133760.00		-125641.00		-125055.00
AIC		277236.83		267548.24		251309.61		250138.04
BIC		277334.50		267653.42		251407.28		250243.22

As shown in (12), negative (positive) value of $\hat{\gamma}$ implies negative (postive) correlation between the frequency and severity. As a result, we can observe that adjustment factors for Gamma GLM, MVGP, MVGB2 models are all below 1, which accounts for negative dependence between the frequency and severity. Summary for the values of $D_N(\gamma)$, the

Table 7: Summary of adjustment factors for the average severity models

	Gamma GLM	Gamma GLMM	MVGP	MVGB2
Min	0.76	1.01	0.81	0.85
Max	0.88	1.02	0.91	0.93
Mean	0.87	1.01	0.90	0.92
Range	0.13	0.01	0.10	0.08
Std. Dev	0.01	0.00	0.01	0.01

dependence adjustment factor, is provided in Table 7.

For the validation of the calibrated average severity models, we compare the actual compound sum of the claims in the test set with the predicted compound sum of the claims, which are calculated using the formula for the credibility premium of compound sum in (11). From Table 8 we can see that MVGP model, as a special case of MVGB2 model, outperforms both Gamma GLM and Gamma GLMM in terms of RMSE and MAE.

Table 8: Validation measures for the average severity models

	Gamma GLM	Gamma GLMM	MVGP	MVGB2
RMSE	1967.3387	1967.1304	1964.729	1975.3112
MAE	475.5807	455.8223	433.621	451.7382

Finally, we may compare the empirical distribution of actual compound sum of the claims with the predictive distributions under each average severity model. According to Figure 3 and Table 9, it shows us that both Gamma GLM and Gamma GLMM fail to describe the overall distribution of S for whole portfolio. On the other hands, both MVGP and MVGB2 model, especially MVGP model is closest to the histogram of actual loss. Therefore, we can claim that MVGP distribution outperforms the naive Gamma models in terms of describing overall risks distribution of the portfolio. Note that the claim data were collected at the end of calender year 2003 so that some of claim in calender year 2001 and 2002 might not be fully reported or developed, which leads to left-skewed shape of distribution of observed S.

Table 9: Summary statistics for predicted log of compound sum

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
Actual	3.7028	6.4770	7.4138	7.4377	8.3657	12.4315	1.1626
Gamma GLM	-67.0280	1.5667	5.7338	3.9695	8.2302	13.0487	6.1526
Gamma GLMM	-71.6026	1.5438	5.6869	3.9252	8.1909	12.9809	6.1528
MVGP	-4.4411	7.0704	7.9425	7.7616	8.6558	14.2374	1.2824
MVGB2	1.3340	7.5908	8.4265	8.3769	9.2155	16.1610	1.2314

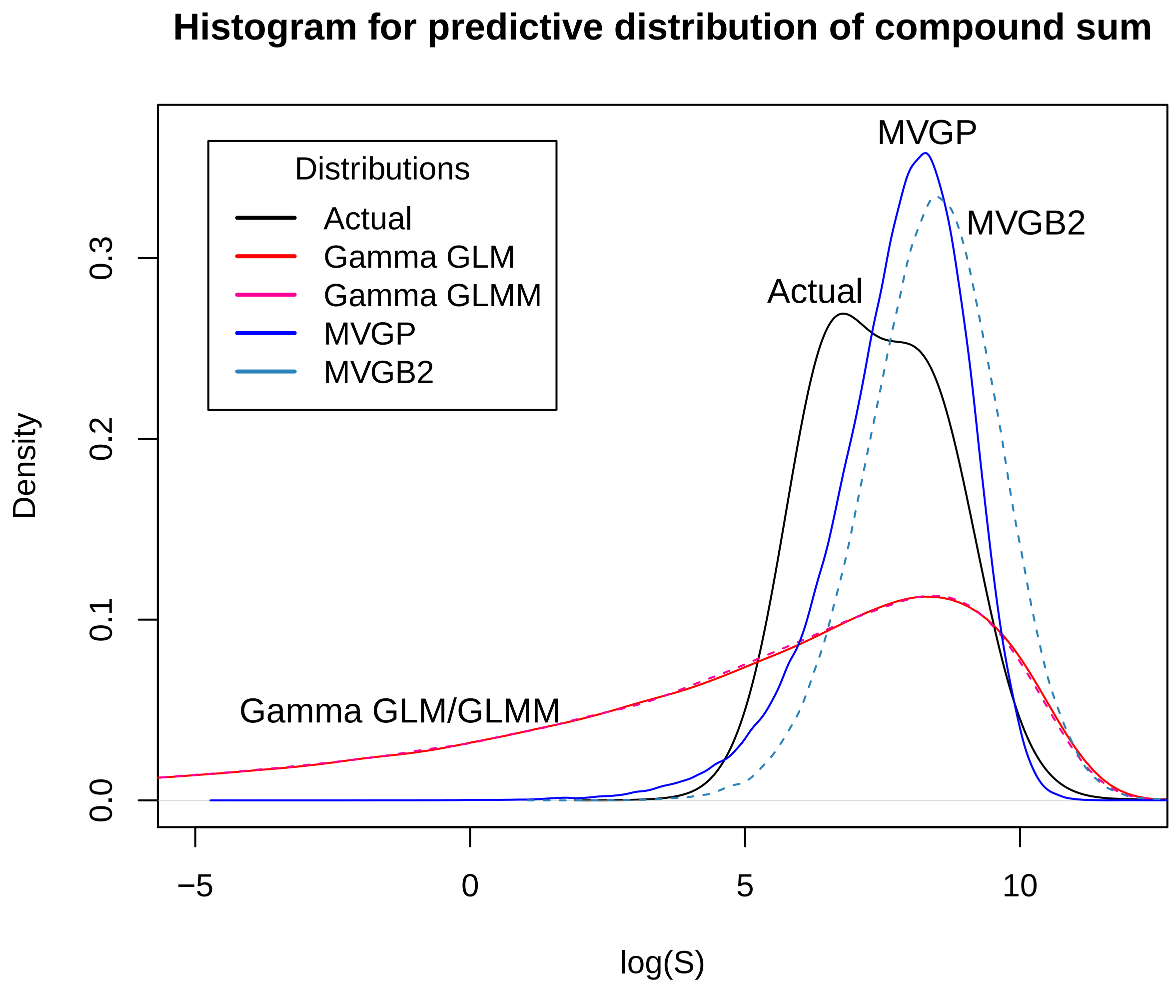

Figure 3: Predictive distribution of compound sum under severity models

6 Concluding remarks

In actuarial practice, both the longitudinal property of dataset and possible dependence between the claim frequency and the average severity might have impact on the construction of ratemaking model. In this article, we explored the possiblity of incorporating those two aspects in ratemaking model via the use of dependent compound random effects model, based on the conjugate famliy of distributions in both frequency and severity. The results show us that proposed MVGP and MVGB2 distribution outperform naive Gamma GLM and Gamma GLMM in terms of better prediction and analytical tractability for severity component. Furthermore, we could obtain a formula for credibility premium of the compound sum, incorporating the possible dependence between the frequency and severity conponents.

Appendix A. Proof of Theorem 1 and Corollary 1

In this appendix, we provide the details of the derivation for the expression of the predictive mean of the aggregate claim as defined by $S_{T+1} = N_{T+1}\overline{C}_{T+1}$ according to our random effects model specification. For simplicity, here we drop the subscript i and conditioning argument on $\mathbf{x}$ so that

$$\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \overline{\mathbf{c}}_T\right].$$

In all cases of the average severity model, conditional mean of $\overline{C}_t$ is given as

$$\mathbb{E}\left[\overline{C}_t|N_t\right] = \theta^C e^{\mathbf{x}_t\beta}\, e^{N_t\gamma} = \theta^C \tilde{\mu}_t\, e^{N_t\gamma}.$$

Therefore, predictive mean of S_{T+1} can be expressed as follows under our severity model specifications:

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[N_{T+1}\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] = \mathbb{E}\left[N_{T+1}\mathbb{E}\left[\overline{C}_{T+1}|N_{T+1}\right]|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \\
&= \mathbb{E}\left[N_{T+1}\theta^C\tilde{\mu}_{T+1}\, e^{N_{T+1}\gamma}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \\
&= \mathbb{E}\left[\theta^C\tilde{\mu}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[N_{T+1}e^{N_{T+1}\gamma}|\mathbf{n}_T\right] \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \mathbb{E}\left[M'_{N_{T+1}|\mathbf{n}_T}(\gamma)\right] \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \frac{r_T}{\tilde{r}_T}\nu_{T+1}e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1} \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1}\; \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}
\end{aligned}$$

where we have used the following results which can be immediately deduced: $M_{N_{T+1}|\mathbf{n}_T}(z) = [1-(\nu_{T+1}/\tilde{r}_T)(e^z-1)]^{-r_T}$ and $\mathbb{E}\left[N_{T+1}e^{\gamma N_{T+1}}|\mathbf{n}_T\right] = M'_{N_{T+1}|\mathbf{n}_T}(\gamma)$.

Note that if $\gamma = 0$, then predictive mean of S_{T+1} is reduced to the product of predictive means of N_{T+1} and $\overline{C}_{T+1}$ as follows:

$$\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\; \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \\
&= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \overline{\mathbf{c}}_T\right]\; \mu_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} = \mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_T, \overline{\mathbf{c}}_T\right] \times \mathbb{E}\left[N_{T+1}|\mathbf{n}_T\right],
\end{aligned}$$

because $\mu_{T+1} = \exp(\mathbf{x}_{T+1}\beta + N_{T+1}\gamma) = \exp(\mathbf{x}_{T+1}\beta + N_{T+1}\cdot 0) = \exp(\mathbf{x}_{T+1}\beta) = \tilde{\mu}_{T+1}$.

Since $\theta^C = 1$ under Gamma GLM, we have that

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \; \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \; \times e^\gamma \left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T-1} \\
&= \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \; \times e^\gamma \left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T-1}.
\end{aligned}
$$

Under Gamma GLMM, $\mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]$ has no closed form but needs to be numerically evaluated, which may cause additional burden on computation since we need to calculate $\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right]$ for each policyholder. For numerical integration, the following identity can be used:

$$
\mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \int \theta \, \pi_C(\theta|\mathbf{n}_T, \bar{\mathbf{c}}_T) d\theta = \frac{\int \theta \, f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \theta)\pi_C(\theta) d\theta}{\int f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \theta)\pi_C(\theta) d\theta} \simeq \frac{\sum_{j=1}^J \hat{\theta}_j \, f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)}{\sum_{j=1}^J f(\bar{\mathbf{c}}_T|\mathbf{n}_T, \hat{\theta}_j)},
$$

where $\hat{\theta}_j$'s are generated from $\mathcal{LN}\left(-\frac{\sigma^2}{2}, \sigma^2\right)$.

Under MVGP, we know that $\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T \sim \mathcal{IG}(k_T+1, w_T)$ where $w_T = k + \sum_{t=1}^T S_t/\phi\mu_t$ and $k_T = k + \sum_{t=1}^T n_t/\phi$ because

$$
\begin{aligned}
\pi_C(\theta|\mathbf{n}_T, \bar{\mathbf{c}}_T) &\propto \pi_C(\theta) \prod_{t=1}^T f(\bar{c}_t|n_t) \propto \theta^{-k-2} \exp\left(-\frac{k}{\theta}\right) \prod_{t=1}^T \theta^{\frac{n_t}{\phi}} \exp\left(-\frac{S_t/\phi\mu_t}{\theta}\right) \\
&\propto \theta^{-(\sum_{t=1}^T n_t/\phi + k+1)-1} \exp\left(-\frac{1}{\theta}(k + \sum_{t=1}^T S_t/\phi\mu_t)\right).
\end{aligned}
$$

Hence, $\mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] = \frac{w_T}{k_T} = \frac{k+\sum_{t=1}^T S_t/\phi\mu_t}{k+\sum_{t=1}^T n_t/\phi} = \frac{k\phi+\sum_{t=1}^T S_t/\mu_t}{k\phi+\sum_{t=1}^T n_t}$. Therefore, we have

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T\right] \; \tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \; \times e^\gamma \left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T-1} \\
&= \frac{k\phi+\sum_{t=1}^T S_t/\mu_t}{k\phi+\sum_{t=1}^T n_t}\tilde{\mu}_{T+1} \times \frac{r+\sum_{t=1}^T n_t}{r+\sum_{t=1}^T \nu_t}\nu_{T+1} \; \times e^\gamma \left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^\gamma - 1)\right]^{-r_T-1}.
\end{aligned}
$$

Finally, under MVGB2, we know that $\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T \sim \mathcal{GIG}(k_T+1, w^*_{T,p}, p)$ where $z_t = \frac{\Gamma(n_t/\phi + 1/p)}{\Gamma(n_t/\phi)}$, $w = \frac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, $w^*_{T,p} = [w^p + \sum_{t=1}^T (z_t \bar{c}_t \mu_t^{-1})^p]^{1/p}$, $k_T = k + \sum_{t=1}^T n_t/\phi$ because

$$
\begin{aligned}
\pi_C(\theta|\mathbf{n}_T, \bar{\mathbf{c}}_T) &\propto \pi_C(\theta) \prod_{t=1}^T f(\bar{c}_t|n_t) \propto \theta^{-p(k+1)-1} \exp\left(-\frac{w^p}{\theta^p}\right) \prod_{t=1}^T \theta^{-pn_t/\phi} \exp\left(-\left(\frac{\bar{c}_t z_t/\mu_t}{\theta}\right)^p\right) \\
&\propto \theta^{-p(\sum_{t=1}^T n_t/\phi + k+1)-1} \exp\left(-\frac{1}{\theta^p}\left[w^p + \sum_{t=1}^T (\bar{c}_t z_t/\mu_t)^p\right]\right),
\end{aligned}
$$

which leads to the following posterior mean of θ^C:

$$
\begin{aligned}
\mathbb{E}\left[\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] &= w^*_{T,p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)} = \left(w^p+\sum_{t=1}^{T}(z_t\overline{c}_t\mu_t^{-1})^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\\
&= \left(w^p+\sum_{t=1}^{T}\left(\frac{\overline{c}_t}{\mu_t}\frac{\Gamma(n_t/\phi+1/p)}{\Gamma(n_t/\phi)}\right)^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\\
&= \left(w^p+\sum_{t=1}^{T}\left(\frac{\overline{c}_t n_t}{\mu_t\phi}\frac{\Gamma(n_t/\phi+1/p)}{\Gamma(n_t/\phi+1)}\right)^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\\
&= \left(\sum_{t=1}^{T}\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi+1/p)}{\phi\Gamma(n_t/\phi+1)}\right)^p+w^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\mathbb{E}\left[S_{T+1}|\mathbf{n}_T,\overline{\mathbf{c}}_T\right] &= \mathbb{E}\left[\theta^C|\mathbf{n}_T,\overline{\mathbf{c}}_T\right]\ \tilde{\mu}_{T+1}\times\frac{r+\sum_{t=1}^{T}n_t}{r+\sum_{t=1}^{T}\nu_t}\nu_{T+1}\ \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}\\
&= \left(\sum_{t=1}^{T}\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi+1/p)}{\phi\Gamma(n_t/\phi+1)}\right)^p+w^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\tilde{\mu}_{T+1}\times\\
&\quad \frac{r+\sum_{t=1}^{T}n_t}{r+\sum_{t=1}^{T}\nu_t}\nu_{T+1}\ \times e^{\gamma}\left[1-\left(\frac{\nu_{T+1}}{\tilde{r}_T}\right)(e^{\gamma}-1)\right]^{-r_T-1}.
\end{aligned}
$$

Appendix B. The derivation of GP and GB2 distributions

Here we suppress all the subscripts and let $\mu=e^{\boldsymbol{x}\beta+n\gamma}$. We can see that when $\overline{C}|n,\theta^C\sim\mathcal{G}\left(\frac{n}{\phi},\theta^C\mu\frac{\phi}{n}\right)$, the density of $\overline{C}|n,\theta^C$ is given as

$$
f_{\overline{C}|N,\theta^C}(\overline{c}|n,\theta)=\frac{1}{\Gamma(n/\phi)}\left(\frac{n}{\theta\mu\phi}\right)^{n/\phi}\overline{c}^{\,n/\phi-1}\exp\left(-\frac{n\overline{c}}{\theta\mu\phi}\right).
$$

For the random effect, when $\theta^C\sim\mathcal{IG}(k+1,k)$, the density of θ^C is given as

$$
\pi_C(\theta)=\frac{1}{\Gamma(k+1)}\left(\frac{k}{\theta}\right)^{k+1}\exp\left(-\frac{k}{\theta}\right)\frac{1}{\theta}.
$$

By integrating out the random effect, we can show that $\overline{C}|n\sim\mathcal{GP}\left(k+1,\mu k\frac{\phi}{n},\frac{n}{\phi}\right)$ as follows:

$$
\begin{aligned}
f(\bar{c}|n) &= \int_0^\infty f(\bar{c}|n,\theta)g(\theta)d\theta \\
&= \int_0^\infty f(\bar{c}|n,\theta)\frac{1}{\Gamma(k+1)}\left(\frac{k}{\theta}\right)^{k+1}\exp\left(-\frac{k}{\theta}\right)\frac{1}{\theta}d\theta \\
&= \frac{\bar{c}^{n/\phi-1}(n/\mu\phi)^{n/\phi}k^{k+1}}{\Gamma(n/\phi)\Gamma(k+1)}\int_0^\infty \theta^{-k-n/\phi-2}\exp\left(-\frac{k+n\bar{c}/\mu\phi}{\theta}\right)d\theta \\
&= \frac{\bar{c}^{n/\phi-1}(n/\mu\phi)^{n/\phi}k^{k+1}}{\Gamma(n/\phi)\Gamma(k+1)}\frac{\Gamma(n/\phi+k+1)}{(k+n\bar{c}/\mu\phi)^{n/\phi+k+1}} \\
&= \frac{\Gamma(n/\phi+k+1)}{\Gamma(n/\phi)\Gamma(k+1)}\frac{\bar{c}^{-1}(n\bar{c}/\mu\phi)^{n/\phi}k^{k+1}}{(k+n\bar{c}/\mu\phi)^{n/\phi+k+1}} \\
&= \frac{\Gamma(n/\phi+k+1)}{\Gamma(n/\phi)\Gamma(k+1)}\frac{\bar{c}^{n/\phi-1}(k\mu\phi/n)^{k+1}}{(\bar{c}+k\mu\phi/n)^{n/\phi+k+1}}
\end{aligned}
$$

Note that if $Y \sim \mathcal{GP}(a,\xi,\tau)$, then the density function is given as follows:

$$
f(y|a,\xi,\tau) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)}\frac{\xi^a y^{\tau-1}}{(y+\xi)^{a+\tau}}.
$$

We know that $\theta^C|\mathbf{n}_T,\bar{\mathbf{c}}_T \sim \mathcal{IG}(k_T+1,w_T)$ where $w_T = k+\sum_{t=1}^T S_t/\phi\mu_t$ and $k_T = k+\sum_{t=1}^T n_t/\phi$ from Appendix A. Thus, it can be shown that $\overline{C}_{T+1}|\mathbf{n}_{T+1},\bar{\mathbf{c}}_T \sim \mathcal{GP}\left(k_T+1,\mu_{T+1}w_T\frac{\phi}{n_{T+1}},\frac{n_{T+1}}{\phi}\right)$ combined with the derivation of marginal GP density and

$$
\mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_{T+1},\bar{\mathbf{c}}_T\right] = \frac{w_T}{k_T}\mu_{T+1} = \frac{k\phi+\sum_{t=1}^T S_t/\mu_t}{k\phi+\sum_{t=1}^T n_t}\mu_{T+1}.
$$

Likewise, let us assume that $\overline{C}|n,\theta^C \sim \mathcal{GG}(n/\phi,\theta^C\mu/z,p)$ with the following density

$$
f(\bar{c}|n,\theta) = \frac{p}{\Gamma(n/\phi)}\left(\frac{z}{\theta\mu}\right)^{pv}\bar{c}^{pv-1}\exp\left(-\left(\frac{\bar{c}z}{\theta\mu}\right)^p\right)
$$

where $\mu = e^{\boldsymbol{x}\beta+n\gamma}$, $v = \dfrac{n}{\phi}$, and $z = \dfrac{\Gamma(n/\phi+1/p)}{\Gamma(n/\phi)}$.

For the random effect, when $\theta^C \sim \mathcal{GIG}(k+1,w,p)$, the density of θ^C is given as

$$
\pi_C(\theta) = \frac{p}{\Gamma(k+1)}\left(\frac{w}{\theta}\right)^{pk+p}\exp\left(-\frac{w^p}{\theta^p}\right)\frac{1}{\theta}
$$

where $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$.

By integrating out the random effect, we can show that $\overline{C}|n \sim \mathcal{GB}2\left(k+1, \mu\dfrac{w}{z}, \dfrac{n}{\phi}, p\right)$ as follows:

$$
\begin{aligned}
f(\bar{c}|n) &= \int_0^\infty f(\bar{c}|n,\theta)\pi_C(\theta)d\theta \\
&= \int_0^\infty f(\bar{c}|n,\theta)\frac{p}{\Gamma(k+1)}\left(\frac{w}{\theta}\right)^{pk+p}\exp\left(-\frac{w^p}{\theta^p}\right)\frac{1}{\theta}d\theta \\
[x := \theta^p] &= \frac{p^2\bar{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\int_0^\infty \theta^{-pk-pv-p-1}e^{(-\frac{w^p+(\bar{c}z/\mu)^p}{\theta^p})}d\theta \\
[\frac{dx}{d\theta} = p\theta^{p-1}] &= \frac{p^2\bar{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}|\frac{1}{p}\int_0^\infty x^{-k-v-2}e^{(-\frac{w^p+(\bar{c}z/\mu)^p}{x})}dx| \\
&= \frac{|p|\bar{c}^{pv-1}(z/\mu)^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\frac{\Gamma(v+k+1)}{(w^p+(\bar{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\bar{c}^{-1}(\bar{c}z/\mu)^{pv}w^{pk+p}}{(w^p+(\bar{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\bar{c}^{pv-1}(w\mu/z)^{pk+p}}{(\bar{c}^p+(w\mu/z)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(\frac{n}{\phi}+k+1)}{\Gamma(\frac{n}{\phi})\Gamma(k+1)}\frac{\bar{c}^{pn/\phi-1}\left(\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})}\mu\right)^{pk+p}}{\left(\bar{c}^p+(\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})}\mu)^p\right)^{n/\phi+k+1}}
\end{aligned}
$$

Note that if $Y \sim \mathcal{GB}2(a, \xi, \tau, p)$, then the density function is given as follows:

$$
f(y|a,\xi,\tau,p) = \frac{\Gamma(a+\tau)}{\Gamma(a)\Gamma(\tau)}|p|\frac{\xi^{ap}y^{\tau p-1}}{(y^p+\xi^p)^{a+\tau}}.
$$

We know that $\theta^C|\mathbf{n}_T, \bar{\mathbf{c}}_T \sim \mathcal{GIG}(k_T+1, w^*_{T,p}, p)$ where $z_t = \dfrac{\Gamma(n_t/\phi+1/p)}{\Gamma(n_t/\phi)}$, $w = \dfrac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, $w^*_{T,p} = [w^p + \sum_{t=1}^T (z_t\bar{c}_t\mu_t^{-1})^p]^{1/p}$, $k_T = k + \sum_{t=1}^T n_t/\phi$ from Appendix A.

Thus, it can be shown that $\overline{C}_{T+1}|\mathbf{n}_{T+1}, \bar{\mathbf{c}}_T \sim \mathcal{GB}2\left(k_T+1, \mu_{T+1}\dfrac{w^*_{T,p}}{z_{T+1}}, \dfrac{n_{T+1}}{\phi}, p\right)$ combined with the derivation of marginal GB2 density and

$$
\mathbb{E}\left[\overline{C}_{T+1}|\mathbf{n}_{T+1}, \bar{\mathbf{c}}_T\right] = w^*_{T,p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\mu_{T+1} = \left(\sum_{t=1}^T\left(\frac{S_t}{\mu_t}\frac{\Gamma(n_t/\phi+1/p)}{\phi\Gamma(n_t/\phi+1)}\right)^p + w^p\right)^{1/p}\frac{\Gamma(k_T+1-1/p)}{\Gamma(k_T+1)}\mu_{T+1}
$$

as shown in Jeong and Valdez (2019).

References

Antonio, K. and Valdez, E. A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. *Advances in Statistical Analysis*, 96(2):187–224.

Boucher, J.-P., Denuit, M., and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Frangos, N. E. and 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.

Frees, E. W., Gao, J., and Rosenberg, M. (2011). Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal*, 15(3):377 – 392.

Frees, E. W. and Kim, J.-S. (2006). Multilevel model prediction. *Psychometrika*, 71(1):79–104.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457 – 1469.

Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229–247.

Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205–215.

Gómez-Déniz, E. (2016). Bivariate credibility bonus–malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics*, 70:117–124.

Hernández-Bastida, A., Fernández-Sánchez, M., and Gómez-Déniz, E. (2009). The net bayes premium with dependence between the risk profiles. *Insurance: Mathematics and Economics*, 45(2):247–254.

Jeong, H., Ahn, J., Park, S., and Valdez, E. A. (2018a). Generalized linear mixed models for dependent compound risk models. *Variance*. accepted for publication.

Jeong, H., Gan, G., and Valdez, E. (2018b). Association rules for understanding policyholder lapses. *Risks*, 6(3):69.

Jeong, H. and Valdez, E. A. (2019). Bayesian shrinkage estimators with GB2 copulas. *Available at SSRN 3373377*.

Kim, G.-S., Paik, M. C., and Kim, H. (2017). Causal inference with observational data under cluster-specific non-ignorable assignment mechanism. *Computational Statistics & Data Analysis*, 113:88–99.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*, volume 715. John Wiley & Sons.

Lee, W., Park, S. C., and Ahn, J. Y. (2019). Investigating dependence between frequency and severity via simple generalized linear models. *Journal of the Korean Statistical Society*, 48(1):13–28.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(4):619–678.

Lemaire, J. (1998). Bonus-malus systems: the european and asian approach to merit-rating. *North American Actuarial Journal*, 2(1):26–38.

Molenberghs, G., Verbeke, G., Demétrio, C., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, 25(3):325–347.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Shevchenko, P. V. and Wuthrich, M. V. (2006). The structural modelling of operational risk via bayesian inference: Combining loss data with expert opinions. *The Journal of Operational Risk*, 1(3):3–26.

Shi, P., Feng, X., and Ivantsova, A. (2015). Dependent frequency-severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64:417 – 428.

Shi, P. and Valdez, E. A. (2012). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2012:1 – 21.

Yang, X., Frees, E. W., and Zhang, Z. (2011). A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathematics and Economics*, 49(2):265–284.