A predictive random effects model of dependent claims frequency and severity

Himchan Jeong   Emiliano A. Valdez

University of Connecticut

Data Science in Action

3 April, 2018

Table of Contents

Longitudinal Claim Data Structure

- Policyholder i is followed over time $t = 1, \ldots, T_i$ years. (Here T_i is at most 9 years.)
- Unit of analysis "*it*" – an insured driver i over time t (year)
- For each "*it*", could have several claims, $k = 0, 1, \ldots, n_{it}$
- Have available information on: number of claims n_{it}, amount of claim c_{itk}, exposure e_{it} and covariates (explanatory variables) x_{it}
  - covariates often include age, gender, vehicle type, driving history and so forth
- We will model the pair $(n_{it}, \overline{c}_{it})$ where

$$\overline{c}_{it} = \begin{cases} \dfrac{1}{n_{it}} \displaystyle\sum_{k=1}^{n_{it}} c_{itk}, & n_{it} > 0 \\ 0, & n_{it} = 0 \end{cases}$$

  is the observed average claim size and $S_{it} = \sum_{k=1}^{n_{it}} c_{itk}$ is the observed aggregate claim size.

The Frequency-Severity Two-Part Model

- For ratemaking in auto insurance, we have to predict the cost of claims $S = \sum_{k=1}^{n} C_k$.
- Traditional approach is

$$\text{Cost of Claims} = \text{Frequency} \times \text{Average Severity}$$

- The joint density of the number of claims and the average claim size can be decomposed as

$$\begin{aligned} f(N, \overline{C}|\mathbf{x}) &= f(N|\mathbf{x}) \times f(\overline{C}|N, \mathbf{x}) \\ \text{joint} &= \text{frequency} \times \text{conditional severity}. \end{aligned}$$

- This natural decomposition allows us to investigate/model each component separately and it does not preclude us from assuming N and $\overline{C}$ are independent.

Premium for Compound Loss under Independence

- If we assume that N and $C_1, C_2, \dots, C_n$ are independent, then we can calculate the premium for compound loss as

$$
\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}\left[\sum_{k=1}^{N} C_k\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{N} C_k | N\right]\right] \\
&= \mathbb{E}\left[N\mathbb{E}\left[\frac{1}{N}\sum_{k=1}^{N} C_k | N\right]\right] = \mathbb{E}\left[N\mathbb{E}\left[\overline{C}|N\right]\right] \\
&= \mathbb{E}\left[N\mathbb{E}\left[\overline{C}\right]\right] = \mathbb{E}[N]\,\mathbb{E}\left[\overline{C}\right]
\end{aligned}
$$

  In other words, we just multiply the expected values from frequency model and the average severity model.

- In general, $\mathbb{E}[S] \neq \mathbb{E}[N]\,\mathbb{E}\left[\overline{C}\right]$.

Why is the Dependence Important?

- If we have positive correlation between N and $\overline{C}$, then

$$\mathbb{E}[S] > \mathbb{E}[N]\,\mathbb{E}\left[\overline{C}\right]$$

  so the company suffers from the higher loss relative to earned premium.

- If we have negative correlation between N and $\overline{C}$, then

$$\mathbb{E}[S] < \mathbb{E}[N]\,\mathbb{E}\left[\overline{C}\right]$$

  so we come up with the loss of market share due to higher premium.

Literature Review for the Proposed Model

- Modeling dependence between the frequency and average severity
  - Shi et al. (2015) and Garrido et al. (2016) used observed frequency as a covariate for the mean of average severity, in other words, $\mathbb{E}\left[\overline{C}|N\right] = e^{\mathbf{x}\beta + n\theta}$ while they only work on cross-sectional data.
- Modeling insurance claim using longitudinal data
  - Antonio and Beirlant (2007) and Boucher et al. (2008) suggested generalized linear mixed model (GLMM) for analyzing data in actuarial science.

Literature Review for the Proposed Model

- Jeong et al. (2017) extended the work of Garrido et al. (2016) to GLMM so that it can consider the dependency between the frequency and average severity with longitudinal data.
- Here, we want to suggest a model framework that incorporates dependence between the frequency and average severity, as well as provide a closed form of likelihood in longitudinal setting by adopting conjugate random effects.

Random Effects Model

Random effects model allows for subject-specific effects which reflects the idea that there is a natural heterogeneity across subjects. For insurance applications, our subject i is usually a policyholder observed for a period of T_i periods.

Given the random effects vector R_i for each subject i, we can write down our likelihood where κ stands for the 'fixed effect', whereas σ stands for the dispersion of random effects.

$$\ell(\kappa, \sigma | y) = \log \prod_{i=1}^{M} \int \prod_{t=1}^{T_i} f(y_{it} | R_i, \kappa) f(R_i | \sigma) dR_i$$

If $y_{it} | R_i$ and R_i are not conjugate, then above likelihood has no explicit form. (For example, $y_{it} | R_i \sim$ Gamma and $R_i \sim$ Gaussian.)

Frequency Model Specifications

We calibrated frequency models with the following specifications:

- For the count of claims (frequency)
  (1) Simple Poisson GLM: $N_{it} \sim \text{Pois}(e^{\mathbf{x}_{it}\alpha})$ so that $\mathbb{E}\left[N_{it}|\mathbf{x}_{it}\right] = e^{\mathbf{x}_{it}\alpha}$.
  (2) Poisson/Gamma Random Effect Model (= Multivariate NB Model)

Poisson/Gamma Random Effects Model

- Here N_{it} is count of claims of i-th insured in t-th year and $\mathbf{x}_{it}$ is the covariate, respectably. And let $\nu_{it} = e^{\mathbf{x}_{it}\alpha}$ where α is fixed effects parameter for the mean of frequency.
- Let $N_{it}|b_i \sim \text{Pois}(\nu_{it} b_i)$, $\mathbb{E}\left[N_{it}|b_i\right] = \mathbb{V}ar\left(N_{it}|b_i\right) = \nu_{it} b_i$ where $b_i \sim \text{Gamma}(r, 1/r)$ so that $\mathbb{E}\left[b_i\right] = 1$ and $\mathbb{V}ar\left(b_i\right) = \frac{1}{r}$.
- In fact, $N_{it} \sim NB(r, \frac{\nu_{it}}{r+\nu_{it}})$ where

$$f_N(n) = \binom{n+r-1}{n} \left(\frac{r}{r+\nu_{it}}\right)^r \left(\frac{\nu_{it}}{r+\nu_{it}}\right)^n$$

Unconditional Mean and Variance

$$\mathbb{E}\left[N_{it}\right] = \mathbb{E}\left[\mathbb{E}\left[N_{it}|b_i\right]\right] = \mathbb{E}\left[b_i \nu_{it}\right] = \nu_{it},$$

$$\mathbb{V}ar\left(N_{it}\right) = \mathbb{V}ar\left(\mathbb{E}\left[N_{it}|b_i\right]\right) + \mathbb{E}\left[\mathbb{V}ar\left(N_{it}|b_i\right)\right] = \nu_{it}(1 + \frac{\nu_{it}}{r})$$

Multivariate NB Distribution

- Let $dF_b = g(b)db$, which is probability measure with respect to gamma distribution. Then according to Boucher et al.(2008), it is known that marginal pdf for the frequency as following, which can be called multivariate negative binomial (MVNB) distribution.

$$
\begin{aligned}
f_{N_i}(n_{i1}, \ldots, n_{iT_i}) &= \int \prod_t f_{N|b}(n_{it}|b)dF_b \\
&= \prod_t \left(\frac{e^{\mathbf{x}_{it}\alpha}}{\sum e^{\mathbf{x}_{it}\alpha} + r} \right)^{n_{it}} \left(\frac{r}{\sum e^{\mathbf{x}_{it}\alpha} + r} \right)^{r} \frac{\Gamma(\sum_t n_{it} + r)}{\Gamma(r) \prod_t n_{it}!}
\end{aligned}
$$

- This is different from the setting which assumes no association structure within the claims of each policyholder.

$$
N_{it} \perp N_{ij} | b_i \quad \text{BUT} \quad N_{it} \not\perp N_{ij}.
$$

Estimation of Fixed Effects with MVNB Distribution

- We can estimate $\hat{r}$ and $\hat{\alpha}$ by differentiating following likelihood and solving the likelihood equations.

$$
\begin{aligned}
\ell_N &= \sum_i \left(\log \int \prod_t f_{N|b}(n_{it}|b) dF_b \right) \\
&= \sum_i \sum_t \left[n_{it} \mathbf{x}_{it} \alpha - \log \Gamma(n_{it} + 1) \right] \\
&\quad + \sum_i \log(\Gamma(\sum_t n_{it} + r)) \\
&\quad - \sum_i [(\sum_t n_{it} + r) \log(\sum_t e^{\mathbf{x}_{it}\alpha} + r)] \\
&\quad + M \left[r \log r - \log \Gamma(r) \right]
\end{aligned}
$$

Severity Model Specifications

We calibrated severity models with the following specifications:

- For the average size of claims (severity)
  1. Simple Gamma GLM: $\overline{C}_{it}|N_{it} \sim \text{Gamma}(\frac{n_{it}}{\phi}, e^{\mathbf{x}_{it}\beta+n_{it}\theta}\frac{\phi}{n_{it}})$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it},\mathbf{x}_{it}\right] = e^{\mathbf{x}_{it}\beta+n_{it}\theta}$ and $\frac{\mathbb{V}ar\left(\overline{C}_{it}|N_{it},\mathbf{x}_{it}\right)}{\mathbb{E}\left[\overline{C}_{it}|N_{it},\mathbf{x}_{it}\right]^2} = \frac{\phi}{n_{it}}$
  2. Gamma/Normal Random Effect Model (= Gamma GLMM): $\overline{C}_{it}|N_{it}, U_i \sim \text{Gamma}(\frac{n_{it}}{\phi}, e^{\mathbf{x}_{it}\beta+n_{it}\theta+u_i}\frac{\phi}{n_{it}})$ where $u_i \sim \text{Normal}(0,\sigma_u^2)$ so that $\mathbb{E}\left[\overline{C}_{it}|N_{it},\mathbf{x}_{it},u_i\right] = e^{\mathbf{x}_{it}\beta+n_{it}\theta+u_i}$ and $\frac{\mathbb{V}ar\left(\overline{C}_{it}|N_{it},\mathbf{x}_{it},u_i\right)}{\mathbb{E}\left[\overline{C}_{it}|N_{it},\mathbf{x}_{it},u_i\right]^2} = \frac{\phi}{n_{it}}$
  3. Gamma/I-gamma Random Effect Model (= Multivariate GP Model)
  4. G-gamma/GI-gamma Random Effect Model (= Multivariate GB2 Model)

I-gamma/G-gamma/GI-gamma Distributions

- Inverse gamma (I-gamma) is a two-parametric distribution such that $f_{Y_1}(y) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\dfrac{\beta}{y}\right)$.
- Generalized gamma (G-gamma) is a three-parametric distribution such that $f_{Y_2}(y) = \dfrac{p\beta^{\alpha p}}{\Gamma(\alpha)} y^{\alpha p-1} \exp\left(-\beta^p y^p\right)$.
- Generalized inverse gamma (GI-gamma) is a three-parametric distribution such that $f_{Y_3}(y) = \dfrac{p\beta^{\alpha p}}{\Gamma(\alpha)} y^{-\alpha p-1} \exp\left(-\beta^p y^{-p}\right)$.
- Note that when p=1, then G-gamma is equivalent to Gamma distribution and GI-gamma is equivalent I-gamma distribution, respectively.

Gamma/Inverse-gamma Random Effects Model

- First let $\mu_{it} = e^{\mathbf{x}_{it}\beta + n_{it}\theta}$, and $\bar{C}_{it}$ is average claim size of i-th insured in t-th year and $\mathbf{x}_{it}$ is the covariate, respectably.
- Let $\bar{C}_{it}|n_{it}, U_i \sim \text{Gamma}(\frac{n_{it}}{\phi}, U_i\mu_{it}\frac{\phi}{n_{it}})$,
  $\mathbb{E}\left[\bar{C}_{it}|n_{it}, U_i\right] = U_i\mu_{it}, \quad \mathbb{V}ar\left(\bar{C}_{it}|n_{it}, U_i\right) = {U_i}^2\mu_{it}^2\frac{\phi}{n_{it}}$
  where $U_i \sim$ I-gamma$(k+1, k)$ so that $\mathbb{E}\left[U_i\right] = 1$ and $\mathbb{V}ar\left(U_i\right) = \frac{1}{k-1}$.

Unconditional Mean and Variance

$$\mathbb{E}\left[\bar{C}_{it}|n_{it}\right] = \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|n_{it}, U_i\right]\right] = \mathbb{E}\left[U\mu_{it}\right] = \mu_{it}$$

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

Generalized Pareto Distribution

- In fact, $\bar{C}_{it}|n_{it} \sim GP(k+1, \mu_{it} k \frac{\phi}{n_{it}}, \frac{n_{it}}{\phi})$. (We can show this using conjugacy of gamma and I-gamma distribution)
- $Y \sim GP(a, \xi, \tau)$

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

- Note that $f(\bar{c}|n,u) = \frac{1}{\Gamma(n/\phi)} \left(\frac{n}{u\mu\phi}\right)^{n/\phi} \bar{c}^{n/\phi-1} \exp\left(-\frac{n\bar{c}}{u\mu\phi}\right)$ and $g(u) = \frac{1}{\Gamma(k+1)} (\frac{k}{u})^{k+1} \exp(-\frac{k}{u}) \frac{1}{u}$ where $\mu = e^{\boldsymbol{x}\beta + n\theta}$. Hence

Deriving Generalized Pareto Distribution

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

Multivariate GP Distribution

- Let $dF_u = g(u)du$, which is probability measure with respect to I-gamma distribution. Then according to our model specification, we can get our marginal likelihood for the average severity as following, which can be called multivariate generalized pareto (MVGP) distribution.

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\overline{c}_{i1}, \ldots, \overline{c}_{iT_i}|n_i) &= \int \prod_t f_{\overline{C}|N,U}(\overline{c}_{it}|n_{it}, u)dF_u \\
&= \frac{k^{k+1} \prod_t \left(n_{it}\overline{c}_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta} / \phi \right)^{n_{it}/\phi}}{\left(k + \sum_t n_{it}\overline{c}_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta} / \phi \right)^{\sum_t n_{it}/\phi + k + 1}} \\
&\times \frac{\Gamma(\sum_t n_{it}/\phi + k + 1) \prod_t \overline{c}_{it}^{-1}}{\Gamma(k+1) \prod_t \Gamma(n_{it}/\phi)}
\end{aligned}
$$

Estimation of Fixed Effects with MVGP Distribution

- We can estimate $\hat{\phi}, \hat{\beta}$ and $\hat{\theta}$ by differentiating following likelihood and solving the likelihood equations.

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_i \left(\log \int \prod_t f_{\overline{C}|N,U}(\overline{c}_{it}|n_{it}, u) dF_u \right) \\
&= \sum_i \sum_t [-\log \Gamma(n_{it}/\phi) - \log \overline{c}_{it}] + \sum_i \log \Gamma(\sum_t n_{it}/\phi + k + 1) \\
&\quad + \sum_i \sum_t n_{it}/\phi(\log n_{it}\overline{c}_{it} - \mathbf{x}_{it}\beta - n_{it}\theta - \log \phi) \\
&\quad - \sum_i [\sum_t n_{it}/\phi + k + 1) \log(k + \sum_t n_{it}\overline{c}_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta}/\phi)] \\
&\quad + M\,[(k+1)\log k - \log \Gamma(k+1)]
\end{aligned}
$$

G-gamma / GI-gamma Random Effects Model

- Let $\mu_{it} = e^{\mathbf{x}_{it}\beta + n_{it}\theta}$ and we assume that average severity follows a G-gamma distribution so that $\bar{C}_{it}|n_{it}, U_i \sim$ G-gamma$(n_{it}/\phi, U_i\mu_{it}\frac{\Gamma(n_{it}/\phi)}{\Gamma(n_{it}/\phi+1/p)}, p)$.
- Then we can check that $\mathbb{E}\left[\bar{c}_{it}|n_{it}, U_i\right] = U_i\mu_{it}$ and $\mathbb{V}ar\left(\bar{c}_{it}|n_{it}, U_i\right) = U_i^2\mu_{it}^2(\frac{\Gamma(n_{it}/\phi+2/p)\Gamma(n_{it}/\phi)}{\Gamma(n_{it}/\phi+1/p)^2} - 1)$ where $U_i \sim$ GI-gamma$(k+1, k\frac{\Gamma(k)}{\Gamma(k+1-1/p)}, p)$ so that $\mathbb{E}\left[U_i\right] = 1$ and $\mathbb{V}ar\left(U_i\right) = \frac{\Gamma(k+1-2/p)\Gamma(k+1)}{\Gamma(k+1-1/p)^2} - 1$.
- Note that if p=1, then G-gamma and GI-gamma are equivalent to gamma and inverse gamma, respectively.
- We can get unconditional mean and variance as following.

G-gamma / GI-gamma Random Effects Model

Unconditional Mean and Variance

$$\mathbb{E}\left[\bar{C}_{it}|n_{it}\right] = \mathbb{E}\left[\mathbb{E}\left[\bar{C}_{it}|n_{it}, U_i\right]\right] = \mathbb{E}\left[U_i\mu_{it}\right] = \mu_{it}$$

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

GB2 Distribution

- In fact, $\bar{C}_{it}|n_{it} \sim GB2(k+1, \mu_{it}\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})}, \frac{n_{it}}{\phi}, p)$. (We can show this using conjugacy of G-gamma and GI-gamma distribution)
- $Y \sim GB2(a, \xi, \tau, p)$

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

- Note that $f(\bar{c}|n,u) = \frac{p}{\Gamma(v)}\left(\frac{z/\mu}{u}\right)^{pv}\bar{c}^{pv-1}\exp\left(-(\frac{\bar{c}z/\mu}{u})^p\right)$ and $g(u) = \frac{p}{\Gamma(k+1)}(\frac{w}{u})^{pk+p}\exp\left(-\frac{w^p}{u^p}\right)\frac{1}{u}$. where $\mu = e^{\mathbf{x}\beta+n\theta}$, $v = n/\phi$, $w = \Gamma(k+1)/\Gamma(k+1-1/p)$, and $z = \Gamma(v+1/p)/\Gamma(v)$.

Deriving GB2 Distribution

$$
\begin{aligned}
f(\bar{c}|n) &= \int_0^\infty f(\bar{c}|n,u)g(u)du \\
&= \int_0^\infty f(\bar{c}|n,u)\frac{p}{\Gamma(k+1)}(\frac{w}{u})^{pk+p}\exp(-\frac{w^p}{u^p})\frac{1}{u}du \\
[x := u^p] &= \frac{p^2\bar{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\int_0^\infty u^{-pk-pv-p-1}e^{(-\frac{w^p+(\bar{c}z/\mu)^p}{u^p})}du \\
[\frac{dx}{du} = pu^{p-1}] &= \frac{p^2\bar{c}^{pv-1}(\frac{z}{\mu})^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}|\frac{1}{p}\int_0^\infty x^{-k-v-2}e^{(-\frac{w^p+(\bar{c}z/\mu)^p}{x})}dx| \\
&= \frac{|p|\bar{c}^{pv-1}(z/\mu)^{pv}w^{pk+p}}{\Gamma(v)\Gamma(k+1)}\frac{\Gamma(v+k+1)}{(w^p+(\bar{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\bar{c}^{-1}(\bar{c}z/\mu)^{pv}w^{pk+p}}{(w^p+(\bar{c}z/\mu)^p)^{v+k+1}} \\
&= |p|\frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)}\frac{\bar{c}^{pv-1}(w\mu/z)^{pk+p}}{(\bar{c}^p+(w\mu/z)^p)^{v+k+1}}
\end{aligned}
$$

Deriving GB2 Distribution

If we back substitute $\mu = e^{\mathbf{x}\beta + n\theta}$, $v = \frac{n}{\phi}$, $w = \Gamma(k+1)/\Gamma(k+1-\frac{1}{p})$, and $z = \Gamma(\frac{n}{\phi}+\frac{1}{p})/\Gamma(\frac{n}{\phi})$, then we can get

$$
\begin{aligned}
f(\bar{c}|n) &= |p| \frac{\Gamma(v+k+1)}{\Gamma(v)\Gamma(k+1)} \frac{\bar{c}^{pv-1}(w\mu/z)^{p(k+1)}}{(\bar{c}^p + (w\mu/z)^p)^{v+k+1}} \\
&= |p| \frac{\Gamma(\frac{n}{\phi}+k+1)}{\Gamma(\frac{n}{\phi})\Gamma(k+1)} \frac{\bar{c}^{pn/\phi-1}\left(\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})} e^{\mathbf{x}\beta+n\theta}\right)^{pk+p}}{\left(\bar{c}^p + (\frac{\Gamma(k+1)\Gamma(\frac{n}{\phi})}{\Gamma(k+1-\frac{1}{p})\Gamma(\frac{n}{\phi}+\frac{1}{p})} e^{\mathbf{x}\beta+n\theta})^p\right)^{n/\phi+k+1}}
\end{aligned}
$$

GP and GB2 Distribution

- $Y \sim GP(a, \xi, \tau)$

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

- $Y \sim GB2(a, \xi, \tau, p)$

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

- GP is a special case of GB2 when p=1.

Multivariate GB2 Distribution

- Let $dF_u = g(u)du$, which is probability measure with respect to GI-gamma distribution. Then according to our model specification, we can get our marginal likelihood for the average severity as following, which can be called multivariate generalized beta-II (MVGB2) distribution where $w = \frac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, and $z_{it} = \frac{\Gamma(n_{it}/\phi+1/p)}{\Gamma(n_{it}/\phi)}$.

$$
\begin{aligned}
f_{\overline{C_i}|N_i}(\overline{c}_{i1}, \ldots, \overline{c}_{iT_i}|n_i) &= \int \prod_t f_{\overline{C}|N,U}(\overline{c}_{it}|n_{it}, u)dF_u \\
&= \frac{w^{pk+p} \prod\limits_t \left(\overline{c}_{it} z_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta}\right)^{pn_{it}/\phi}}{\left(w^p + \sum\limits_t (\overline{c}_{it} z_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta})^p\right)^{\sum\limits_t n_{it}/\phi + k + 1}} \\
&\quad \times \frac{\Gamma(\sum\limits_t n_{it}/\phi + k + 1) \prod\limits_t \overline{c}_{it}^{-1} p^{T_i}}{\Gamma(k+1) \prod\limits_t \Gamma(n_{it}/\phi)}
\end{aligned}
$$

Estimation of Fixed Effects with MVGB2 Distribution

- We can estimate $\hat{\phi}, \hat{\beta}, \hat{\theta}$ and $\hat{p}$ by differentiating following marginal likelihood and solving the likelihood equations where $w = \frac{\Gamma(k+1)}{\Gamma(k+1-1/p)}$, and $z_{it} = \frac{\Gamma(n_{it}/\phi+1/p)}{\Gamma(n_{it}/\phi)}$.

$$
\begin{aligned}
\ell_{\overline{C}|N} &= \sum_i \left(\log \int \prod_t f_{\overline{C}|N,U}(\overline{c}_{it}|n_{it}, u) dF_u \right) \\
&= \sum_i \left[\sum_t (-\log \Gamma(\frac{n_{it}}{\phi}) - \log(\frac{\overline{c}_{it}}{p})) + \log \Gamma(\sum_t \frac{n_{it}}{\phi} + k + 1) \right] \\
&\quad + p \sum_i \sum_t n_{it}/\phi (\log \overline{c}_{it} z_{it} - \mathbf{x}_{it}\beta - n_{it}\theta) \\
&\quad - \sum_i [\sum_t n_{it}/\phi + k + 1) \log(w^p + \sum_t (\overline{c}_{it} z_{it} e^{-\mathbf{x}_{it}\beta - n_{it}\theta})^p)] \\
&\quad + M \left[(k+1)p \log w - \log \Gamma(k+1) \right]
\end{aligned}
$$

Multivariate GP and GB2 Distribution

- Both MVGP and MVGB2 are different from the setting which assumes no association structure within the claim amounts of each policyholder.

$$\overline{C}_{it} \perp \overline{C}_{ij} | U_i, N_i \quad \text{BUT} \quad \overline{C}_{it} \not\perp \overline{C}_{ij} | N_i \ .$$

Observable policy characteristics used as covariates

Categorical variables	Description		Proportions	
VehType	Type of insured vehicle:	Car	99.27%	
		Motorbike	0.47%	
		Others	0.27%	
Gender	Insured's sex:	Male = 1	80.82%	
		Female = 0	19.18%	
Cover Code	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

- Singapore insurance data (1993–2000: Training set, 2001: Test set)
- $M = 50,215$ unique policyholder, 162,179 of aggregated total number of observations observed on training set.

Frequency Estimation Results

Table 1: Regression estimates of the frequency models

	Poisson			MVNB		
	Est	s.e	Pr(>|t|)	Est	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		99091.40			98989.24	

Frequency Validation Results - MSE and MAE

Table 2: Validation measures for the frequency models

	Poisson	MVNB
MSE	0.33161	0.33174
MAE	0.16996	0.16969

Frequency Validation Results - Gini Index

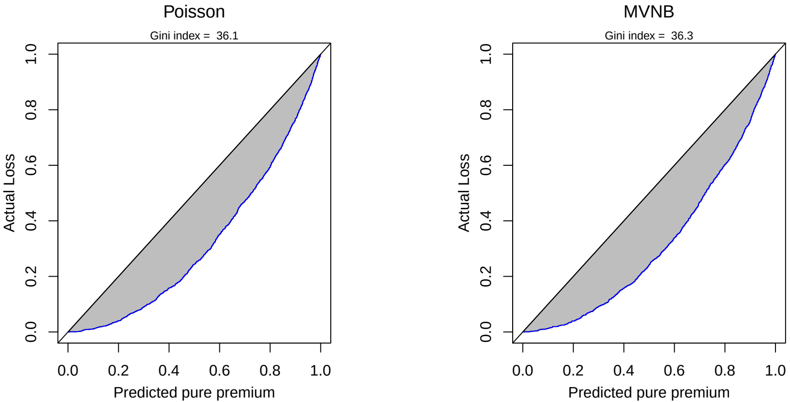

Figure 1: Gini indices for frequency models

Severity Estimation Results

Table 3: Regression estimates of the average severity models

	Gamma		Gamma GLMM		MVGP		MVGB2 (p=0.81)	
	Est	Pr(>|t|)	Est	Pr(>|t|)	Est	Pr(>|t|)	Est	Pr(>|t|)
(Intercept)	7.61	0.00	5.91	0.00	7.64	0.00	7.50	0.00
VTypeCar	-0.29	0.55	0.12	0.62	-0.18	0.36	-0.12	0.70
VTypeMBike	2.87	0.03	2.32	0.00	3.19	0.00	3.29	0.00
log(VehCapa)	0.53	0.00	0.33	0.00	0.48	0.00	0.48	0.00
VehAge	-0.03	0.00	-0.01	0.00	-0.01	0.02	-0.01	0.00
SexM	-0.01	0.91	-0.02	0.49	-0.03	0.19	-0.03	0.48
Comp	0.05	0.60	0.19	0.00	0.26	0.00	0.12	0.01
Age	-0.16	0.00	-0.05	0.03	-0.16	0.00	-0.15	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	-0.01	0.00
Count	-0.12	0.01	0.01	0.65	-0.04	0.08	-0.08	0.00
Loglikelihood		-138605		-133760		-125130		-125092
AIC		277236		267548		250289		250212
BIC		277334		267653		250395		250317

Severity Validation Results - MSE and MAE

Table 4: Validation measures for the average severity models

	Gamma	Gamma GLMM	MVGP	MVGB2
MSE	8325.652	8308.980	8308.735	8315.923
MAE	3462.561	3259.532	3334.564	3384.894

Severity Validation Results - Gini Index

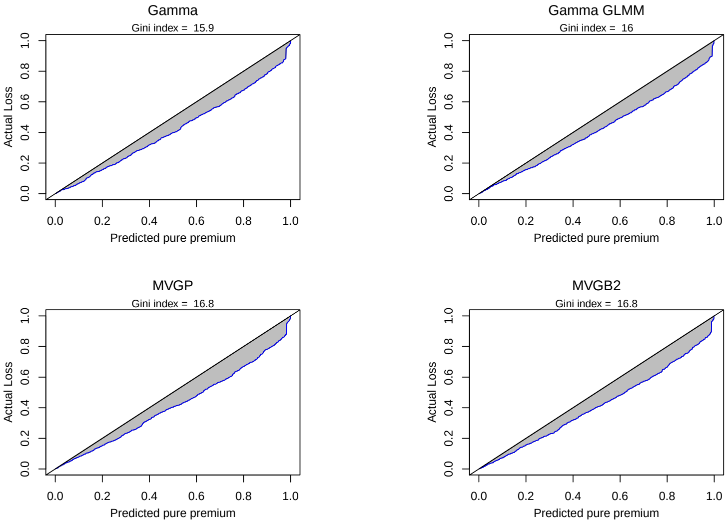

Figure 2: Gini indices for severity models

Severity Validation Results - Q-Q Plots

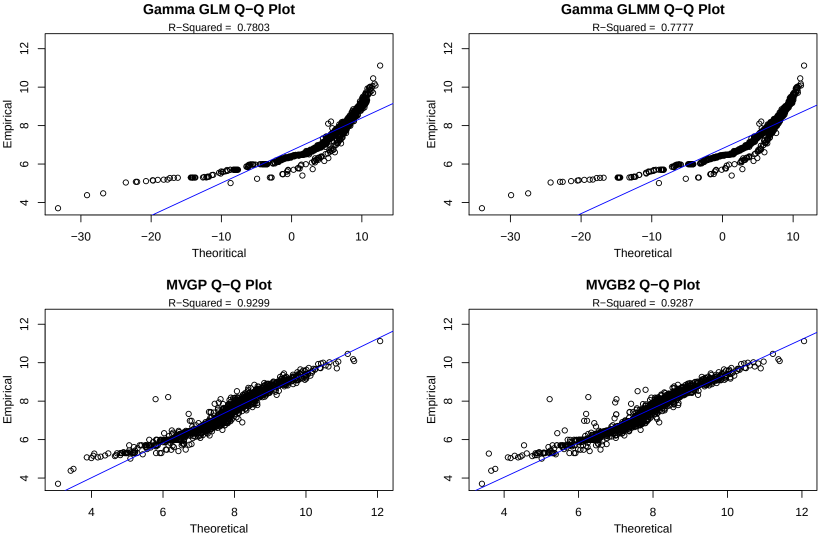

Figure 3: Q-Q Plots for different models

Concluding Remarks

- We found significant negative dependence between the frequency and average severity from our model.
- Under the conjugate random effects model framework, we obtained MVGB2 which is flexible parametric distribution as well as naturally involved with association structure within the claims of a policyholder.
- MVGB2 outperformed naive gamma GLM and gamma GLMM with respect to most of the model selection criteria, such as AIC, BIC, Gini index, and fit of Q-Q Plots.

References

Antonio, K. and Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 40(1):58 - 76.

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.

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.

Jeong, H., Ahn, J., Park, S., and Valdez, E. A. (2017). Generalized linear mixed models for dependent compound risk models. *working paper*.

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