Ratemaking in a Changing Environment, ASTIN Bulletin

Online Appendices

Appendix A: Summary statistics at the policy and claim level

Table 1 summarizes the distribution of the claim frequency and severity, which are the two continuous outcomes of interest at the policy level. The Spearman correlation coefficient shows that the two outcomes are correlated. The table also shows that the coverage and deductible distributions are right-skewed, and we deal with the skewness by taking logarithmic transformations on these variables. The high correlation coefficient for coverage suggests it will be a significant predictor in the claim frequency and severity model. In addition, the table summarizes the transaction frequency to the settlement of a reported claim and the average payment amounts; the Spearman correlation coefficient shows that they are correlated. Moreover, the correlation coefficient between deductible and the transaction frequency and severity indicates it may be used as a predictor.

Table 1: Summary statistics for outcomes at the policy level (claim frequency and severity) and claim level (transaction frequency and severity), and continuous covariates (deductibles, and coverages).

	Policy Level					
	Min.	Median	Average	Max.	Claim Frequency (ρ_S)	Claim Severity (ρ_S)
Claim Frequency	0	0	0.893	231	-	0.964
Claim Severity	0	0	9,649	12,922,218	0.964	-
Deductible	500	1,000	3,407	100,000	0.051	0.090
Coverage(000'S)	0.4	11,493	11,493	2,444,797	0.404	0.396
	Claim Level					
	Min.	Median	Average	Max.	Transaction Frequency (ρ_S)	Transaction Payment (ρ_S)
Transaction Frequency	1	1	1.221	11	-	0.301
Transaction Payment (Average per claim)	8.2	2,609	11,757	1,174,293	0.301	-
Deductible	500	1,000	6,739	100,000	0.095	0.132
Coverage(000'S)	102.3	58,046	295,660	2,444,797	-0.019	-0.079

Table 2 shows the summary statistics of the claim frequency and severity at the policy level and the transaction frequency and severity at the claim level for the categorical variables in the dataset. The table suggests a high variation in the claim frequency and average severity of the claims across the categorical variables at the policy level. At the claim level, the transaction frequency does not vary much, but there is a substantial variation with the average severity.

Table 2: Summary statistics at the policy and claim level by categorical variables.

Variable	Policy Level Average Frequency	Policy Level Average Severity	Claim Level Average Frequency	Claim Level Average Severity
Entity Type				
Village	0.349	2,977	1.196	7,776
City	1.607	12,015	1.222	9,276
County	3.330	17,453	1.169	13,089
Misc	0.172	4,110	1.127	14,328
School	0.931	25,963	1.231	15,604
Town	0.092	1,204	1.126	6,351
Region				
Northeastern	0.561	7,835	1.189	12,840
Northern	0.410	10,682	1.229	13,879
Southeastern	1.389	30,123	1.206	15,605
Southern	1.196	5,820	1.263	7,921
Western	0.483	5,210	1.216	7,989
Alarm Credit				
No Alarm Credit	0.226	2,427	1.237	7,873
Alarm Credit 5%	0.290	3,508	1.183	10,566
Alarm Credit 10%	0.275	3,016	1.165	10,250
Alarm Credit 15%	1.059	15,869	1.212	11,326
Alarm Credit (Combination)	2.212	29,923	1.233	14,476

Appendix B: Algorithm to construct the simulation data.

We employ Algorithm 1 to construct the simulation data.

Algorithm 1 Data-generating process for MPP.

Input: Number of policyholders J, parameters for claim frequency $\{\alpha, \beta_{11}, \beta_{12}\}$, parameters for reporting delay distribution $\{\gamma_{10}, \gamma_{11}.\gamma_{12}, \kappa\}$, parameters for transaction frequency $\{b, \pi_{11}, \pi_{12}\}$, parameters for transaction payment $\{\phi_{10}, \phi_{11}, \phi_{12}, \sigma\}$, probability of settlement η, and ratemaking time τ.

Output:
Complete reported data - policy-level ($PL = \left\{(S_j, N_j, x_{j1}, x_{j2}); j = 1, ..., J\right\}$),
claim-level ($CL = \left\{(P_{ji}^{ULT}, x_{j1}, x_{j2}, \delta_{ji}, M_{ji}(\tau), M_{ji}); N_j > 0, i = 1, ..., N_j\right\}$),
and transaction-level ($TL = \left\{(P_{jik}, x_{j1}, x_{j2}, M_{ji}(\tau)); M_{ji} > 0, k = 1, ..., M_{ji}\right\}$)
Datasets at ratemaking date - policy-level ($PL^R = \left\{PL; k \leq M_{ji}(\tau)\right\}$),
claim-level ($CL^R = \left\{CL; k \leq M_{ji}(\tau)\right\}$),
and transaction-level ($TL^R = \left\{TL; k \leq M_{ji}(\tau)\right\}$)

for Policyholder $j = 1, ..., J$ **do**
  Generate $x_{j1} \sim \text{Bernouli}(0.3)$, $x_{j2} \sim \text{Normal}(0, 1)$;
  Generate $N_j^T \sim \text{Poisson}(\psi_j)$; $\psi_j = \exp(\log \alpha + x'_{j1}\beta_{11} + x'_{j2}\beta_{12})$;
  Generate $U_i \sim \text{Weibull}(N_j^T, \theta_i, \kappa)$; $\theta_i = \exp(\gamma_{10} + x'_{j1}\gamma_{11} + x'_{j2}\gamma_{12})$;
  Generate $v_i \sim \text{Uniform}(N_j^T, 0, \tau)$;
  Generate $N_j \sim \text{Count}(U_i + v_i \leq \tau)$;
  if $N_j > 0$ **then**
    for $i = 1, ..., N_j$ **do**
      Generate $M_{ji} \sim \text{Poisson}(\lambda_{ji})$; $\lambda_{ji} = \exp(\log b + x'_{j1}\pi_{11} + x'_{j2}\pi_{12})$;
      if $M_{ji} > 0$ **then**
        for $k = 1, ..., M_{ji}$ **do**
          Generate $P_{jik} \sim \text{Gamma}\left(\frac{\exp(\mu_{jik})}{\sigma}, \sigma\right)$;
          where $\mu_{jik} = \exp(\phi_{10} + x'_{j1}\phi_{11} + x'_{j2}\phi_{12})$;
        end for
        return $\left\{P_{ji}^{ULT} = P_{ji1} + \ldots + P_{jiM_{ji}}\right\}$;
        Generate $\delta_{ji} \sim \text{Bernouli}(\eta)$
        for $\delta_{ji} = 0$ **do**
          $M_{ji}(\tau) \sim$ Discrete Uniform$(0, M_{ji})$
        end for
      else
        $P_{jik} = 0, P_{ji}^{ULT} = 0$, $M_{ji}(\tau) = 0$ and $\delta_{ji} = 1$
      end if
    end for
    return $\left\{S_j = P_{j1}^{ULT} + \ldots + P_{jN_j}^{ULT}\right\}$;
  else
    $S_j = 0$
  end if
end for
return
$PL = \left\{(S_j, N_j, x_{j1}, x_{j2})\right\}$; $PL^R = \left\{PL; k \leq M_{ji}(\tau)\right\}$;
$CL = \left\{(P_{ji}^{ULT}, x_{j1}, x_{j2}, \delta_{ji}, M_{ji}(\tau), M_{ji})\right\}$; $CL^R = \left\{CL; k \leq M_{ji}(\tau)\right\}$
$TL = \left\{(P_{jik}, x_{j1}, x_{j2}, M_{ji}(\tau))\right\}$; $TL^R = \left\{TL; k \leq M_{ji}(\tau)\right\}$

Appendix C: Loss Cost Prediction Using the Frequency-Severity Model

A rating formula based on the Frequency-Severity model will be achieved by the product of exponentiated estimates from the claim frequency, and severity models. The following rating formula calculates the predicted loss cost:

$$\begin{aligned}\text{Loss Cost} &= e_j \exp(\ln \hat{\alpha}_T + \boldsymbol{x}_j'\hat{\boldsymbol{\beta}}) \times \exp(\boldsymbol{x}_j'\hat{\boldsymbol{\phi}}) \\ &= \text{Exposure} \\ &\times \text{Expected number of claims} \\ &\times \text{Expected payment per claim.}\end{aligned} \tag{1}$$

Where e_j is the exposure variable and $\boldsymbol{x}_j$ are rating factors for the new contract. $\{\hat{\alpha}_T\}$ are the fitted trend parameters from the most recent policy year for the reported claim. Also, $\{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}\}$ are the fitted parameters for rating variables from the claim frequency model, and the severity model building blocks. Here, the Poisson model is used to model the claim frequency, and a gamma GLM with a logarithmic link is used to model the loss amounts from claims.

References