2025-02-25

A New Class of Severity Regression Models with an Application to IBNR Prediction

Research group on AIRM Lab, POSTECH

Keywoong Bae

Information

- Title: A New Class of Severity Regression Models with an Application to IBNR Prediction
- Author: Tsz Chai Fung, Andrei L. Badescu, and X. Sheldon Lin
- Journal: North American Actuarial Journal
- Year: 2020

1. Introduction

Framework

- **Purpose:** With the use of the proposed TG-LRMoE Model, we not only obtain the excellent **goodness of fit for severity and reporting delay** but also provide **reasonable predictions for IBNR.**
- **Methodology:** Finite Mixture Model (TG-LRMoE)
- **Dataset:** A dataset supplied by European major automobile insurer

Step 1	• Introduction to TG-LRMoE model and its properties • ECM Algorithm-Based Parameter Estimation	⇨	**Theoretical analysis**
Step 2	• Dataset Introduction and its classification (In-Sample and Out-of-Sample) • Model fit on In-Sample • IBNR Prediction on Out-of-sample	⇨	**Empirical analysis**

Step 1

Introduction to TG-LRMoE model and its properties

2. The TG-LRMoE Regression Model

What is the LRMoE?

- Insurance data often exhibit heavy-tailed behavior, where the extreme losses in the tail are often the most impactful to the insurers.
- Tail behavior of insurance loss data is commonly modeled by traditional heavy-tailed distributions as Burr, log-Gamma, and generalized Pareto distributions (GPD).
- To cater to the mismatch between the body behavior and the tail behavior, a relatively **more flexible modeling approach** is the use of a composite distribution or more generally a spliced distribution.
- The logit-weighted reduced mixture of experts model (**LRMoE**) was proposed as an alternative statistical tool for frequency or severity regression.

2. The TG-LRMoE Regression Model

LRMoE Regression Model

- Its probability Density Function (PDF) is given by

$$\boldsymbol{h}_{\widetilde{Y}_i|x_i}(\widetilde{\boldsymbol{y}}_i;\boldsymbol{x}_i,\boldsymbol{\alpha},\boldsymbol{\Psi},\boldsymbol{g}) = \sum_{j=1}^{g} \boldsymbol{\pi}_j(\boldsymbol{x}_i;\boldsymbol{\alpha})\boldsymbol{f}\left(\widetilde{\boldsymbol{y}}_i;\boldsymbol{\psi}_j\right), \qquad \widetilde{\boldsymbol{y}}_i > \boldsymbol{0},$$

  - g is the number of latent classes,
  - $\Psi = (\psi_1, \dots, \psi_g)$ are the parameters of the **expert functions (f)**,
  - $\boldsymbol{\pi}_j(\boldsymbol{x}_i;\boldsymbol{\alpha}) = \frac{\exp\left\{\alpha_j^T x_i\right\}}{\sum_{j'=1}^{g}\exp\left\{\alpha_{j'}^T x_i\right\}}$ is the mixing weight for the jth class (**gating function**),
  - $\alpha = (\alpha_1, \dots, \alpha_g)$ and $\alpha_j = \left(\alpha_{j0}, \dots, \alpha_{jP}\right)^T \in R^{P+1}$ are the regression parameters for mixing weights.

2. The TG-LRMoE Regression Model

TG-LRMoE (Transformed Gamma – LRMoE)

- **TG-LRMoE** is a function where the observed data first manipulated through a Box-Cox transformation and modeled by the LRMoE with a Gamma expert function.
- First, Y_i is transformed through a **Box-Cox transformation** for a more flexible adjustment on tail behavior of the data ($\boldsymbol{\widetilde{Y}_i = \frac{(1+Y_i)^{\gamma}-1}{\gamma}}$).
- Based on this, they combine the **Gamma Expert function** with logit-weighted regression to model the latent classes.
- Through this, we can model various tail behaviors of loss distributions

2. The TG-LRMoE Regression Model

Box-Cox Transformation

- A total of n mutually independent insurance claims.

- For each claim $i \in \{1, \ldots, n\}$, we define $Y = (Y_1, \ldots, Y_n)^T$ and $x_i = \left(x_{i0}, \ldots, x_{ip}\right)^T$ as the claim severity column vector (response variable) and the information relating to claim i (covariates).

- They transform the claim severities ($\boldsymbol{Y}$) into $\widetilde{\boldsymbol{Y}} \coloneqq \left(\widetilde{\boldsymbol{Y}}_{\boldsymbol{1}}, \ldots, \widetilde{\boldsymbol{Y}}_{\boldsymbol{n}}\right)^{\boldsymbol{T}}$ using Box-Cox transformation:

$$\widetilde{\boldsymbol{Y}}_{\boldsymbol{i}} = \frac{(\boldsymbol{1} + \boldsymbol{Y}_{\boldsymbol{i}})^{\boldsymbol{\gamma}} - \boldsymbol{1}}{\boldsymbol{\gamma}}, \qquad \boldsymbol{\gamma} > \boldsymbol{0},$$

  where γ is a parameter controlling the tail heaviness of the distribution of $\widetilde{Y}_i$.

- $\widetilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ has a lighter tail than Y_i when $\gamma < 1$ and vice versa.

- It enables flexible tail modeling.

2. The TG-LRMoE Regression Model

Gamma Expert Function

- In TG-LRMoE, they choose a Gamma expert function so that $\Psi = (m, \theta), \psi_j = (m_j, \theta_j)$, and

$$\boldsymbol{f(\tilde{y}_i; \psi_j) \coloneqq f(\tilde{y}_i; m_j, \theta_j) = \frac{\tilde{y}_i^{m_j-1} e^{-\frac{\tilde{y}_i}{\theta_j}}}{\Gamma(m_j)\theta_j^{m_j}}, \qquad \tilde{y}_i, m_j, \theta_j > 0,}$$

  where $m = (m_1, \dots, m_g)$ and $\theta = (\theta_1, \dots, \theta_g)$ are, respectively, the shape and scale parameters of gamma distribution.

- Under simple probabilistic arguments, the pdf of Y_i, given x_i can also be derived:

$$\boldsymbol{h_{\tilde{Y}_i|x_i}(\tilde{y}_i; x_i, \alpha, m, \theta, g) = \sum_{j=1}^{g} \pi_i(x_i; \alpha) f(\tilde{y}_i(\gamma); m_j, \theta_j)(1 + y_i)^{\gamma-1} \quad , y_i, m_j, \theta_j, \gamma > 0,}$$

  where $\tilde{y}_i(\gamma) = ((1 + y_i)^{\gamma} - 1)/\gamma$ is a function of γ.

2. The TG-LRMoE Regression Model

TG-LRMoE (Transformed Gamma – LRMoE)

- TG-LRMoE classifies each claim into one of the g unobservable subgroups.
- Claim severity distributions vary among subgroups but are homogeneous within a group.
- Depending on the covariates x_i of the claim, each claim has different probabilities of being classified into different subgroups.
- The subgroup assignments are governed by the regression coefficients α of the gating function.
- A large positive regression coefficient α_{jp} represents a higher chance for a claim to be classified as subgroup j when x_{ip} is large.

3. Desirable Properties

Three Properties on TG-LRMoE

- **Denseness Property**
  - TG-LRMoE function has complete flexibility, meaning it can model any distribution and regression structure effectively and with great adaptability.
- **Tail flexibility**
  - By adjusting the γ parameter, the model can cover a wide range of tail behaviors, from light-tailed to heavy-tailed distributions.
- **Identifiability**
  - The model ensures identifiability, which means that it allows for statistical inference and the results can be clearly interpreted.

Step 1

ECM Algorithm-Based Parameter Estimation

4. Parameter Estimation

MLE vs ECM

- This section presents an ECM algorithm to efficiently calibrate the model parameters.

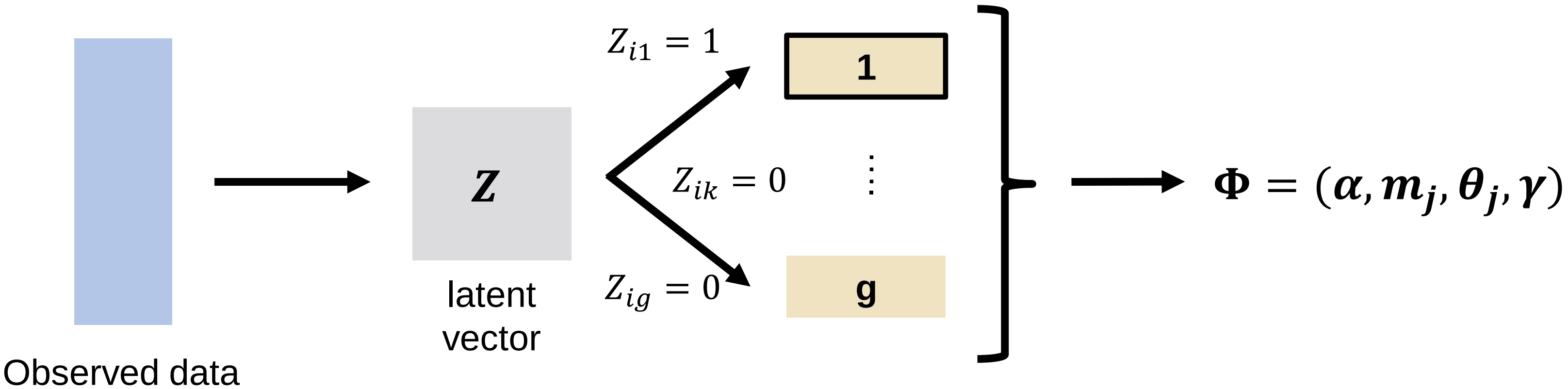
	MLE (Maximum Likelihood Estimation)	ECM (Expectation Conditional Maximization)
How estimate parameters?	By maximizing the likelihood function for given data.	Through the E-step and CM-step when latent variables are involved.
When to use?	No latent variables or for direct estimation from observed data.	With **latent variables** or when it is required.
Approach	Direct optimization of the likelihood function to estimate parameters.	By calculating the expectation in the E-step and perform conditional maximization in the CM-step.
Model stability	Can lead to overfitting.	By handling latent variables, ensure stable updates to the model and prevent overfitting.
Mainly Uses	Used in simple models without latent variables (e.g., general regression models)	Useful for mixture models or models involving latent variables.

4. Parameter Estimation

ECM Algorithm

- This paper uses TG-LRMoE combining multiple components for a flexible tail modeling.
- However, it is not directly known which component corresponds to a particular data point.
- Therefore, we use latent variables to model it.

E step | **CM step**



- Calculate the relative probability of a claim being classified to a subgroup g.

$$z_{ij}^{(l)} = E\left[Z_{ij} \middle| y, x, \Phi^{(l-1)}\right] = \frac{\pi_j\left(x_i; \alpha^{(l-1)}\right) f\left(y_i; \theta_{jk}^{(l-1)}\right)}{\sum_{j'=1}^{g} \pi_{j'}\left(x_i; \alpha^{(l-1)}\right) f\left(y_i; \theta_{j'k}^{(l-1)}\right)}$$

- Estimate the parameters $\boldsymbol{\Phi} = (\boldsymbol{\alpha}, \boldsymbol{m}, \boldsymbol{\theta}, \boldsymbol{\gamma})$, while g is fixed at each ECM run.

4. Parameter Estimation

Parameter Penalization

- Estimating parameters of the LRMoE severity distributions imposes an extra challenge that the likelihood function may be unbounded ($m_j \to \infty, \theta_j \to 0$).
- To this end, the ECM algorithm **penalizes parameters** taking extreme values through finding the AP estimates of the parameters.
- To penalize parameters taking extreme values, they adopt a **Bayesian approach** to set up a prior distribution for each parameter.
    - For the regression coefficient α, they set $\alpha_{jp} \sim N\left(0, \sigma_{jp}^2\right)$ for $j = 1, \dots, g-1$ and $p = 0, \dots, P$.
    - For the shape and size parameters, we set $m_j \sim Gamma(\nu_j^{(1)}, \lambda_j^{(1)})$ and $\theta_j \sim Gamma(\nu_j^{(2)}, \lambda_j^{(2)})$ for $j = 1, \dots, g$ to prevent spurious fitted models.
    - They choose not to penalize γ because this parameter does not result in spurious model or overfitting issue.

Step 2

Dataset Introduction and its classification

5.1 Data Overview

Data

- Dataset supplied by **European major automobile insurer**
    - It contains 594,908 third-party liability insurance contracts during the observation period from January 1, 2007 to December 31, 2017.
    - For each contract, the contract number, starting date, ending date, and various policyholder features are recorded.
    - Among all contracts, 28,256 claims are incurred and reported on or before December 31, 2017.
    - For each claim, the contract number, loss date, reporting date, settlement date, and total amount paid are recorded.
    - For unsettled claims, the insurer also provides the case reserve estimates (the expected future payments) based on detailed claim-specific information.
    - The total incurred loss of a claim is then the sum of the amount paid and the case reserve estimate.

5.1 Data Overview

Summary of the Covariates for the kth Contract

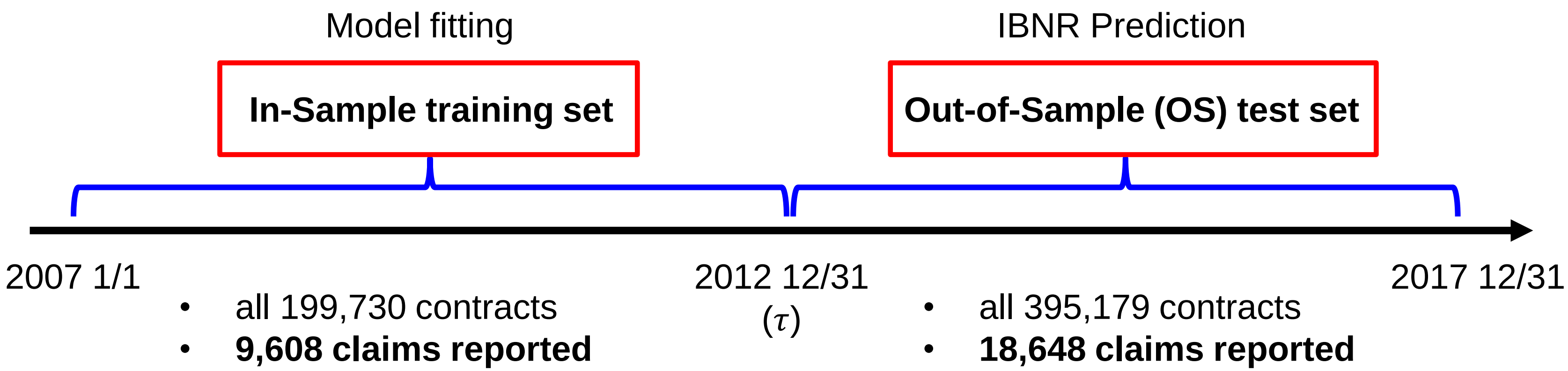
TABLE 1
Summary of the Covariates for the kth Contract

Variable	Description	Type	Levels
x_{k1}	Policyholder age	Discrete	
x_{k2}	Car age	Discrete	
x_{k3}	Car fuel	Categorical	Diesel: $x_{k3} = 1$
			Gasoline: $x_{k3} = 0$
x_{k4}–x_{k7}	Geographical location	Categorical	Region I: $x_{k4} = 1$
			Region II: $x_{k5} = 1$
			Region III: $x_{k6} = 1$
			Region IV: $x_{k7} = 1$
			Capital: $x_{k4} = x_{k5} = x_{k6} = x_{k7} = 0$
x_{k8}–x_{k9}	Car brand class	Categorical	Class A: $x_{k8} = 1$
			Class B: $x_{k9} = 1$
			Class C: $x_{k8} = x_{k9} = 0$
x_{k10}	Contract type	Categorical	Renewal contract: $x_{k10} = 1$
			New contract: $x_{k10} = 0$

5.1 Data Overview

In-Sample and Out-of-Sample

- To evaluate the predictive power through the out-of-sample test, we set a validation date τ of December 12, 2012 and divide the dataset into two part:



Step 2

Model fit on In-Sample

5.3 Model Fitting

Model Fitting

- Three components to be modeled for IBNR prediction: severity, reporting delay, and frequency.

- Direct maximization of $\mathcal{L}(= g_1 \times g_2 \times g_3)$ is difficult because of its complicated form;

    - First, the development process loglikelihood function g_1 and the reporting delay likelihood g_2 are first optimized.

    - Then, given that the parameters involved in g_2 are obtained and fixed, we maximize the frequency likelihood g_3.

Maximize log-likelihood on Severity ($g_1(Z)$) ⇨ Maximize log-likelihood on Reporting Delay ($g_2(U)$) ⇨ Maximize log-likelihood on Frequency ($g_3(N^r)$) ⇨ Maximize $\mathcal{L}$

5.3.1 Severity

GLM on loss severity

- They perform the Goodness-of-Fit (GoF) by fitting Gamma, lognormal, and Pareto GLMs.
  - As covariates, they select all of the variables ($x_{i1} - x_{i10}$) and the transformed reporting delay $x_{i11} := \log(1 + u_i)$.
- $\boldsymbol{p}$ **values** are directly obtained using Kolmogorov-Smirnov test, Chi-square test, and Anderson-Darling test.
- Lognormal GLM provides a relatively better fitting with extremely **small $\boldsymbol{p}$ values ($\mathbf{10^{-7}}$).**
- So, there is a **substantial room for improvement** of the fitting performance → TG-LRMoE

	Model selection statistics			p Values of goodness-of-fit statistics		
	Log-likelihood	AIC	BIC	Kolmogorov-Smirnov test	χ^2 test	Anderson-Darling test
Gamma GLM	–90,974.7	181,975.3	182,068.5	$<10^{-7}$	$<10^{-7}$	$<10^{-7}$
Log-normal GLM	–89,303.9	178,633.8	178,727.1	$<10^{-7}$	$<10^{-7}$	$<10^{-7}$
Pareto GLM	–89,936.3	179,898.5	179,991.7	$<10^{-7}$	$<10^{-7}$	$<10^{-7}$
TG-LRMoE (nine components)	**–88,766.4**	**177,762.8**	178,587.4	.9809	.7110	1.0000
TG-LRMoE (four components)	–88,927.1	177,944.1	**178,266.8**	.6943	.2183	.7116

Note: The bold numbers represent the largest value of the Log-likelihood and the smallest AIC and BIC values.

Fung et al. (2020)

5.3.1 Severity

GLM on loss severity (Result)

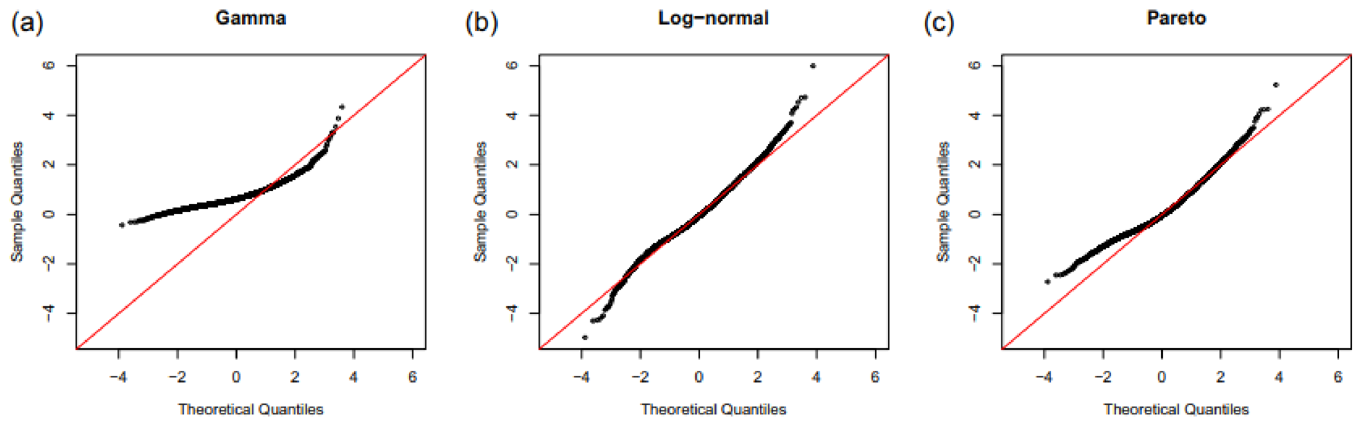


FIGURE 2. Q-Q Normal Plots for the Normalized Residuals under (a) Gamma, (b) Log-normal, and (c) Pareto GLMs.

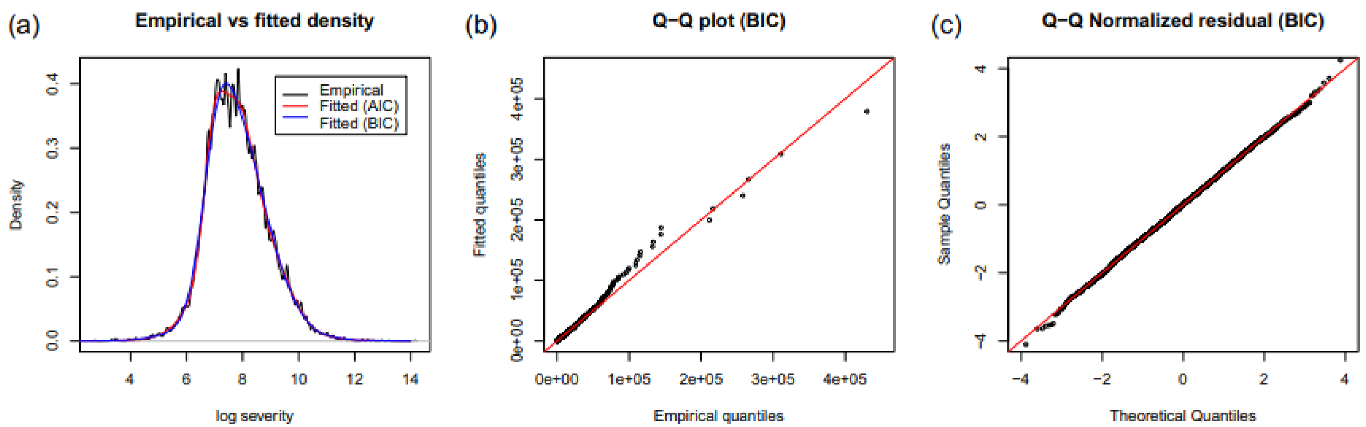


FIGURE 4. (a) Empirical and Fitted Log-Transformed Density Plot. *Note:* Empirical density is generated by kernel method with small bandwidth. (b) Q-Q Plot for Severities (BIC Model). (c) Q-Q Normal Plot for the Normalized Residuals (BIC Model).

Fung et al. (2020)

5.3.1 Severity

GLM on loss severity (Result)

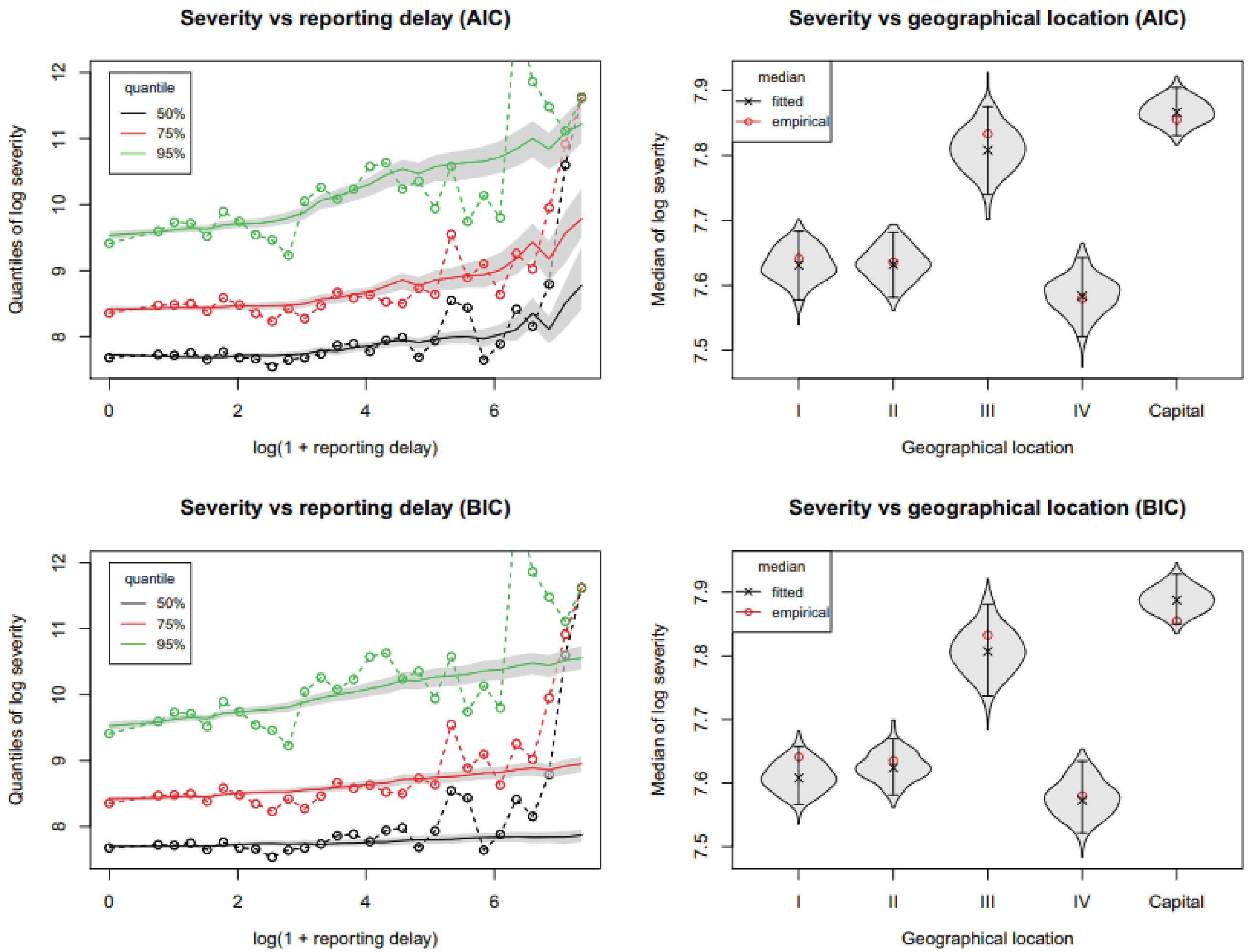


FIGURE 5. (Left) Quantiles of Log Severity versus Reporting Delay Evaluated Using the Nonparametric Approach. *Note:* Dotted curve: empirical pattern; solid curve: fitted pattern; shaded region: 95% confidence interval. (Right) Median of Log Severity versus Geographical Location via a Violin Plot with 95% Confidence Interval Shown by a Bar.

Fung et al. (2020)

5.3.1 Severity

GLM on loss severity (Result)

TABLE C.1
Fitted Severity Model Parameters and the Related Quantities (AIC)

	Component j								
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$
$\hat{\alpha}_{j0}$	3.614	0.795	–1.631	–3.069	–3.030	3.209	–0.186	1.355	0.000
$\hat{\alpha}_{j1}$	–0.100	0.002	–0.001	–0.028	–0.018	–0.035	0.011	0.001	0.000
$\hat{\alpha}_{j2}$	0.179	0.137	0.124	–0.029	0.175	–0.503	–0.072	–0.045	0.000
$\hat{\alpha}_{j3}$	0.030	–0.401	–0.334	0.735	–0.711	–0.050	–0.172	–0.095	0.000
$\hat{\alpha}_{j4}$	–1.279	–1.624	1.559	1.814	–2.203	–0.332	–0.114	–1.640	0.000
$\hat{\alpha}_{j5}$	–2.032	–0.963	1.180	1.688	–2.389	0.454	–2.111	–0.962	0.000
$\hat{\alpha}_{j6}$	0.678	–0.566	1.163	1.325	–1.203	0.691	0.036	–0.599	0.000
$\hat{\alpha}_{j7}$	–1.339	–1.156	2.245	2.555	–0.733	1.587	–1.066	–1.235	0.000
$\hat{\alpha}_{j8}$	–0.524	0.448	0.663	–0.970	–0.505	–1.750	–0.755	–0.182	0.000
$\hat{\alpha}_{j9}$	–1.021	0.248	0.456	0.241	0.034	0.251	0.010	–0.001	0.000
$\hat{\alpha}_{j10}$	–2.355	0.791	–0.267	–1.050	–0.012	–1.061	–0.869	0.302	0.000
$\hat{\alpha}_{j11}$	–0.211	–0.797	–0.032	1.058	0.984	0.135	0.424	–0.372	0.000
$\hat{m}_j$	114.253	77.849	21.701	17.304	8.074	34.003	109.477	66.706	214.113
$\hat{\theta}_j$	0.140	0.148	0.515	0.837	1.282	0.324	0.110	0.211	0.047
$\hat{\gamma}$	0.098								
$E[\tilde{Y}_i \| Z_j = 1]$	16.015	11.497	11.181	14.485	10.347	11.026	12.068	14.095	10.010
$P(Z_j = 1)$	0.024	0.214	0.161	0.059	0.019	0.152	0.096	0.147	0.129

5.3.2 Reporting Delay

Experimental Setting

- To model the reporting delay distribution, they need to consider $g_2(U)$, a likelihood function involving the right-truncated reporting delays of all 9,608 observed claims.

- However, direct maximization of $g_2(U)$ is difficult because the denominator $P_{U|x_i}(\tau - t_i^r)$ (), which is the probability of truncation, varies among the observed claims.

    - To alleviate such a computational issue, they discard some data points where the loss occur after a specified date τ_0.

    - Choosing December 31, 2011, as τ_0, the truncation probabilities $P_{U|x_i}(\tau - t_i^r)$ are very close to one for the remaining $n^r = 7{,}524$ data points because they observe that more than 99% of the claims are reported within a year.

- In addition to the data truncation issue, the observed reporting delays are interval censored (i.e., the number of days [integers] instead of continuous values are observed).

    - They assume that the observed reporting delay U_i (given the covariates $x_{i1} - x_{i10}$) follows an interval-censored version of TG-LRMoE (i.e., the uncensored delay $\tilde{U}_j$ follows the TG-LRMoE and $U_i = [\tilde{U}_i]$ where $[\cdot]$ is a floor function) for $i = 1, \dots, n^r$.

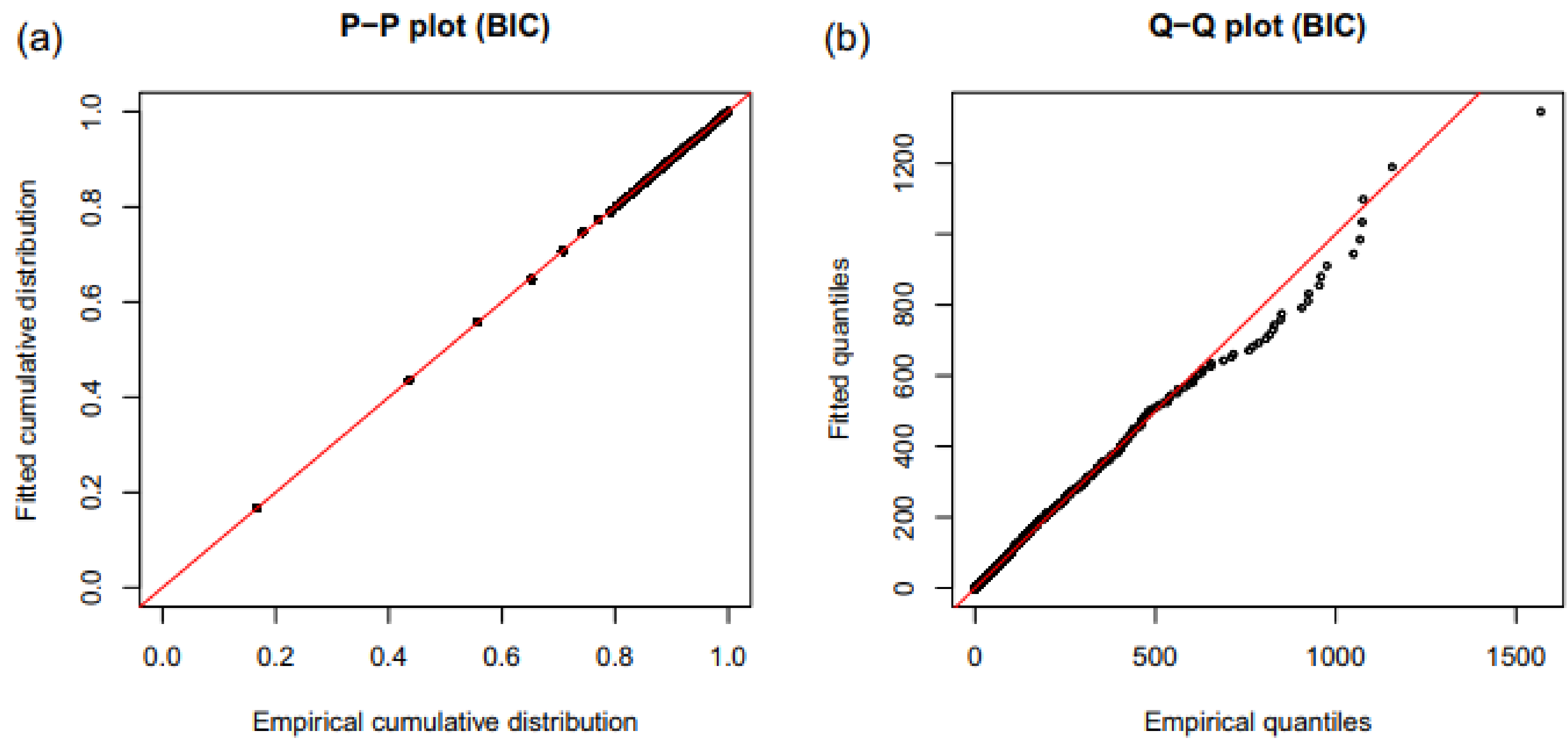
5.3.2 Reporting Delay

GLM fitting

- The AIC and BIC fitted models contain eight and four components, respectively.
- Under the AIC model as an example, a large proportion (around 70%) of claims belong to components 2 to 6, representing several types of accidents or claims that are usually quickly reported.
- In contrast, **components 1 and 7** correspond to claims with **substantial reporting delays**.
- Both plots reveal that the censored version of TG-LRMoE provides excellent fit to both the body and the tail of the reporting delay data.

5.3.2 Reporting Delay

P-P Plot and Q-Q Plot for Reporting Delay under the BIC Model



5.3.2 Reporting Delay

P-P Plot and Q-Q Plot for Reporting Delay under the BIC Model

TABLE C.3
Fitted Reporting Delay AIC Model Parameters and the Related Quantities

	Component j							
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$
$\hat{\alpha}_{j0}$	–1.227	0.763	–2.317	–1.097	–2.533	0.061	–0.364	0.000
$\hat{\alpha}_{j1}$	–0.042	–0.030	0.012	–0.003	0.003	0.002	–0.004	0.000
$\hat{\alpha}_{j2}$	0.023	–0.085	–0.076	0.064	0.176	0.005	–0.012	0.000
$\hat{\alpha}_{j3}$	0.914	0.070	0.047	–0.097	0.306	0.070	0.190	0.000
$\hat{\alpha}_{j4}$	1.840	1.871	0.535	0.266	–0.160	0.295	–0.870	0.000
$\hat{\alpha}_{j5}$	1.471	1.439	0.730	–0.058	0.608	0.430	–0.636	0.000
$\hat{\alpha}_{j6}$	–0.280	2.107	0.548	–0.441	–0.954	0.520	–0.164	0.000
$\hat{\alpha}_{j7}$	2.148	1.543	–0.681	0.644	1.240	0.964	–0.278	0.000
$\hat{\alpha}_{j8}$	–0.909	–1.461	0.462	–0.294	2.688	0.030	0.148	0.000
$\hat{\alpha}_{j9}$	–1.450	–2.364	1.071	–0.204	0.016	–0.390	–0.042	0.000
$\hat{\alpha}_{j10}$	0.046	–0.953	0.013	1.515	–1.874	0.227	0.134	0.000
$\hat{m}_j$	2.212	3.431	11.514	2.199	1.990	12.450	0.808	1.050
$\hat{\theta}_j$	28.984	1.027	0.257	1.468	1.592	0.089	226.712	11.164
$E[Y_i \mid Z_j = 1]$	64.112	3.523	2.963	3.227	3.169	1.106	183.126	11.722
$P(Z_j = 1)$	0.027	0.084	0.073	0.174	0.082	0.295	0.086	0.180

Fung et al. (2020)

5.3.3 Frequency

GLM fitting

- After fitting the severity and delay distributions, the remaining piece is to calibrate the frequency model → maximize $g_3(N^r)$.

TABLE C.5
Estimated Regression Coefficients for Poisson GLM

	Without contract date			with contract date		
	Estimate	SE	p Value	Estimate	SE	p Value
Intercept	–2.0771	0.0491	.0000	–2.0554	0.0514	.0000
Policyholder age	–0.0085	0.0009	.0000	–0.0084	0.0009	.0000
Car age	–0.0157	0.0035	.0000	–0.0146	0.0036	.0001
Car fuel						
>Diesel	0.1389	0.0213	.0000	0.1402	0.0213	.0000
>Gasoline	–			–		
Geographical location						
>Region I	–0.0324	0.0285	.2570	–0.0365	0.0287	.2038
>Region II	0.1523	0.0283	.0000	0.1498	0.0283	.0000
>Region III	–0.0996	0.0362	.0060	–0.1003	0.0362	.0056
>Region IV	–0.1346	0.0331	.0000	–0.1381	0.0332	.0000
>Capital	–			–		
Car brand class						
>Class A	–0.0461	0.0268	.0860	–0.0456	0.0268	.0890
>Class B	–0.1216	0.0239	.0000	–0.1235	0.0240	.0000
>Class C	–			–		
Contract type						
>Renewal	–0.2083	0.0212	.0000	–0.2015	0.0218	.0000
>New	–			–		
Contract date	–			–0.0092	0.0065	.1565

Note: The left panel is the frequency model adopted for IBNR prediction. The right panel includes policyholder contract date as a covariate and aims to examine the time effect on claim frequencies.

Step 2

IBNR Prediction on Out-of-sample

5.3 Model Fitting

Notations

- Suppose that the development of the lth claim of the kth contract is described as a triplet

$$\left(T_l^{(k)}, U_l^{(k)}, Z_l^{(k)}\right)$$

  where $T_k^{(k)}$ is the accident date, $U_l^{(k)}$ is the reporting delay (in days), and $Z_l^{(k)}$ is the development process after the claim is reported.

- We also define notations for the kth contract,
  - $\boldsymbol{N_k^a(t)} = \sum_{\boldsymbol{l=1}}^{\infty} \mathbf{1}_{\{\boldsymbol{T_l^{(k)}} \le \boldsymbol{t}\}}$ as the total claim count process,
  - $\boldsymbol{N_k^r(t)} = \sum_{\boldsymbol{l=1}}^{\infty} \mathbf{1}_{\{\boldsymbol{T_l^{(k)}} \le \boldsymbol{t},\, \boldsymbol{T_l^{(k)}} + \boldsymbol{U_l^{(k)}} \le \boldsymbol{\tau}\}}$ as the reported claim count process,
  - $\boldsymbol{N_k^u(t)} = \sum_{\boldsymbol{l=1}}^{\infty} \mathbf{1}_{\{\boldsymbol{T_l^{(k)}} \le \boldsymbol{t}, \boldsymbol{T_l^{(k)}} + \boldsymbol{U_l^{(k)}} > \boldsymbol{\tau}\}}$ as the IBNR claim count process.

- $N_k^a(t) = N_k^r(t) + N_k^u(t)$.

Model Prediction and Out-of-Sample test

Simulation studies

- The predictive distribution of the aggregate IBNR is obtained by simulations, requiring the following steps for each contract $k = 1, \dots, m$ (repeating this procedure 50,000 times):

1) Generate the number of claims for the contract $n_k^a(\tau)$ from the **fitted frequency GLM**.

2) For $l = 1, \dots, n_k^a(\tau)$, simulate the accident date of the lth claim $t_{kl}^{acc} \sim \mathrm{Uniform}[t_k^{start}, \min\{t_k^{end}, \tau\}]$ and the reporting delay t_{kl}^{delay} from the **fitted interval-censored version of TG-LRMoE**.

3) For $l = 1, \dots, n_k^a(\tau)$, simulate the unreported amount of each claim $\boldsymbol{y_{kl}^{IBNR}}$, which is generated from the **fitted TG-LRMoE** if $t_{kl}^{acc} + t_{kl}^{delay} > \tau$ and is equal to zero if $t_{kl}^{acc} + t_{kl}^{delay} \leq \tau$.

- The aggregate predicted amount of IBNR claims is given by

$$\boldsymbol{y_{agg}^{IBNR} = \sum_{k=1}^{m} \sum_{l=1}^{n_k^a(\tau)} y_{kl}^{IBNR}}$$

Model Prediction and Out-of-Sample test

Simulation results

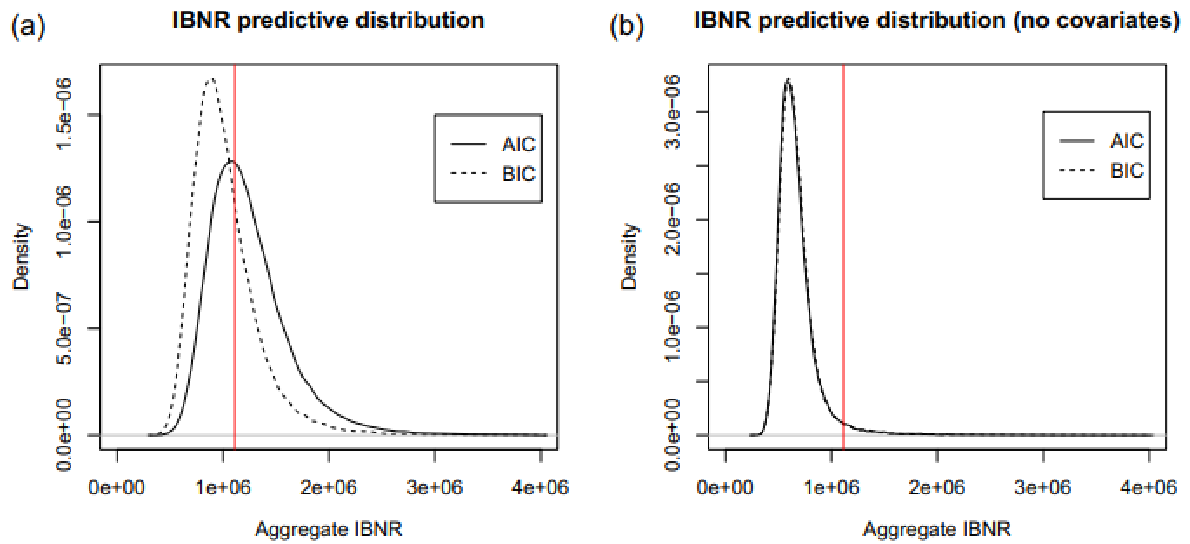


FIGURE 7. Predictive Distribution of the IBNR (a) with Covariates and (b) without Covariates. *Note:* (right panel). The vertical line is the realized aggregate IBNR based on the OS dataset.

TABLE 3
Summary Statistics of the IBNR Predictions

		CTE		VaR			
	Mean	70%	95%	95%	99.5%	Realized	p Value
With covariates (AIC)	1.250	1.718	2.418	1.966	3.023	≥1.112	.836
With covariates (BIC)	1.026	1.406	2.011	1.609	2.485	≥1.112	.597
Without covariates (AIC)	0.676	0.899	1.412	0.992	1.903	≥1.112	.063
Without covariates (BIC)	0.687	0.927	1.568	1.000	2.224	≥1.112	.066

Note: The mean, CTE, VaR, and realized value are expressed in millions.

Model Prediction and Out-of-Sample test

Simulation results

- The vertical line is the total IBNR claims from the test set.
  - Note that only information up to 2017 is available, so the true IBNR may be larger than the value displayed.
- A sharp spike in the average claim severities for very long reporting delay is captured by the AIC model but not by the BIC model.
  - Because the reporting delay of a simulated IBNR claim is usually long, the claim severity for a simulated IBNR claim may be underestimated using the BIC.
- We also highlight the importance of including covariates and using regression models for adequate IBNR predictions.
  - The realized IBNR is on the right tail of the predictive distribution, providing evidence that the IBNR may be underestimated if covariates are excluded.