A New and Flexible Regression Model for Ratemaking and Reserving

Andrei Badescu

Department of Statistical Sciences
University of Toronto
badescu@utstat.utoronto.ca

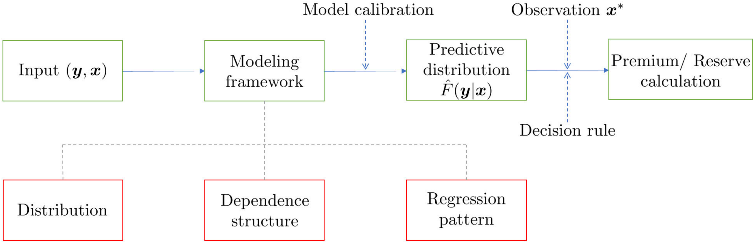
Joint work with Tsz Chai (Samson) Fung and Sheldon Lin

Annual Meeting, Casualty Actuarial Society
Nov 10-13, 2019

Agenda

Process on ratemaking and reserving/prediction

- Policy attributes (covariates) and data:
  - $\boldsymbol{x} = (x_0, x_1, \cdots, x_P)^T$: Policyholder information (covariates)
  - $\boldsymbol{y} = (y_1, y_2, \cdots, y_K)^T$: Claim frequencies or severities of multiple business lines
- A typical modelling process



Some of existing modelling approaches

Parametric models

- GLM, GAM, copula, ...
- Specified assumptions
- Easy to interpret/ compute
- but often inflexible due to linear or special structure of the models

Machine learning models

- Decision tree, neural network
- Fewer assumptions
- More flexible
- Hard to interpret (problematic for premium calculation)
- Potentially overfitting to data and hence not robust

Alternative approach: Mixture-based models

- Mixture of Erlang distribution (Lee and Lin (2010, 2012)), Multivariate Pascal distribution (Badescu et al. (2015)), ...
- Semi-parametric and interpretable: Classification of policyholders into various risk subgroups
- Flexible and data-driven: Number of subgroups can be adjusted to control model complexity

Desirable properties for a “good model”

A model should be

- Flexible and parsimonious so that it can fit (without overfitting) any type of insurance claim data (multi-modal, light/heavy tailed..) and accommodate various dependence structures. We can also minimize the issue of model selection, an often difficult task;
- Interpretable in order to identify the (often highly nonlinear) relationships between policy attributes and claims;
- Mathematical tractable so that premiums, reserves, dependency measures and other quantities of interest can be calculated with ease;
- Statistical tractable, i.e. having efficient estimation/fitting algorithms.

Our proposed model

Start with a class of Generalized Mixture of Experts (GMoE) models...

$$h^*(\boldsymbol{y};\boldsymbol{x}) := h^*(\boldsymbol{y};\boldsymbol{x},\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*,g) = \sum_{j=1}^{g} \pi_j^*(\boldsymbol{x};\boldsymbol{\alpha}^*) \prod_{k=1}^{K} f_k(y_k;\boldsymbol{\theta}_{jk}^*(\boldsymbol{x};\boldsymbol{\beta}_{jk}^*))$$

gating function expert function

- Flexible, but extremely complicated!

How to make the GMoE useful in general/P&C insurance applications?

- Logit-weighted **reduced** Mixture of Experts Models (LRMoE)

$$h(\boldsymbol{y};\boldsymbol{x}) := h(\boldsymbol{y};\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\Theta},g) = \sum_{j=1}^{g} \pi_j(\boldsymbol{x};\boldsymbol{\alpha}) \prod_{k=1}^{K} f_k(y_k;\boldsymbol{\theta}_{jk})$$

logit-linear gating

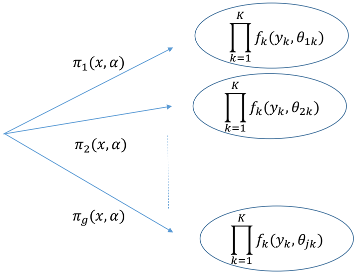
$$\pi_j(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{exp\{\boldsymbol{\alpha}_j^T\boldsymbol{x}\}}{\sum_{j'=1}^{g} exp\{\boldsymbol{\alpha}_{j'}^T\boldsymbol{x}\}}$$

no regression

Model description

- Interpretation
  - Policyholders classified in various homogeneous risk subgroups
  - Policyholder characteristics affects probabilities of assignments
  - Each risk group has its own frequency/severity distribution

A graphical interpretation



Denseness property

The full flexibility of the LRMoE is theoretically justified by its "denseness properties" = "approximates close enough"

- Denseness of LRMoE in the space of GMoE
  - Reducing GMoE to LRMoE does not reduce flexibility
  - Model parsimony
- Denseness in any regression distributions
  - LRMoE can be "fully flexible" to capture any distribution, dependence and regression patterns including nonlinear regression and covariate interactions
  - Calibrated model resembles any input data

Mathematical tractability

Marginalization properties (the LRMoE is closed under marginalization)

- Response marginalization: any marginal distribution is in the same class of distributions with easily calculated parameters
- Covariates marginalization: if some covariates are missing the response is still in the same class of distributions with with easily calculated parameters

Moment properties

- Mean, Variance, Covariance as well Kendall's tau and Spearman's rho all have analytical expressions.

Statistical estimation: an ECM algorithm

- Aim: Efficiently estimate maximum-likelihood parameters $\mathbf{\Phi} = (\boldsymbol{\alpha}, \mathbf{\Theta})$
- Observed data log-likelihood (hard to optimize directly!):

$$l(\mathbf{\Phi}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{g} \pi_j(\boldsymbol{x}_i; \boldsymbol{\alpha}) \prod_{k=1}^{K} f(y_{ik}; \boldsymbol{\theta}_{jk}) \right]$$

Expectation conditional maximization (ECM) algorithm:

- E-step: Compute the expectation of complete data log-likelihood
- M-step: Maximize the likelihood requires high-dimensional optimization of non-concave functions!

How to choose an appropriate expert function?

Recall the denseness mild regularity condition:

- the expert function can be arbitrarily close to any degenerate distributions.

Frequency distributions

- Poisson ✗
- Negative Binomial ✗
- Conway-Maxwell-Poisson (CMP) ✓
- Renewal Count model ✓
- Discretized Severity model ✓

Severity distributions

- Gamma ✓
- Weibull ✓
- Log-normal ✓
- Inverse Burr ✓
- Exponential ✗
- Pareto ✗

The choice of experts for frequency distributions

Erlang Count Model

$$f(y;\boldsymbol{\theta}) := P(N_1 = y; m, \beta) = e^{-\beta} \sum_{b=0}^{m-1} \frac{\beta^{my+b}}{(my+b)!}, \qquad y = 0, 1, 2, \ldots$$

- Denseness condition
- Closed-form pmf
- Interpretation as a renewal process

The resulting model is called Erlang Count LRMoE (EC-LRMoE)

The choice of experts for severity distributions

Transformed Gamma Model

$$\tilde{Y}_i = \frac{(1 + Y_i)^\gamma - 1}{\gamma}, \qquad \gamma > 0$$

- Box Cox transforming the response variable
- Transformed variable $\tilde{Y}_i$ follows LRMoE with gamma expert function
- Captures light tail ($\gamma > 1$), heavy tail ($0 < \gamma < 1$) and extreme tail ($\gamma \to 0$)

The resulting model is called Transformed Gamma LRMoE (TG-LRMoE)

Application to real insurance count data

Data Overview

- European automobile insurance dataset
- 18,019 policyholders
- Bivariate response $\boldsymbol{Y}_i$ (2 business lines):
  - Y_{i1}: number of claims of third body liabilities
  - Y_{i2}: number of claims of car damages
- 11 covariates for $\boldsymbol{x}_i$: Policyholder age, car age, geographical location etc.

Application to real insurance count data

Covariate Information

Discrete valued covariates

Variable	Description	Mean	Stddev	Minimum	Maximum
x_{i1}	Age of policyholder	51.000	11.702	20	88
x_{i2}	Car age	6.248	3.335	0	26

Categorical covariates

Variable	Description	Levels	Proportions
x_{i3}	Car fuel	Diesel: $x_{i3} = 1$	0.383
		Gasoline: $x_{i3} = 0$	0.617
x_{i4}–x_{i5}	Policyholder's history	Renewal with claims last year: $x_{i4} = 1$	0.148
		New contract: $x_{i5} = 1$	0.235
		Renewal, no claims last year: $x_{i4}, x_{i5} = 0$	0.618
x_{i6}–x_{i9}	Geographical location	Region I: $x_{i6} = 1$	0.187
		Region II: $x_{i7} = 1$	0.146
		Region III: $x_{i8} = 1$	0.111
		Capital: $x_{i9} = 1$	0.420
		Region IV: $x_{i6}, x_{i7}, x_{i8}, x_{i9} = 0$	0.136
x_{i10}–x_{i11}	Car brand class	Class A: $x_{i10} = 1$	0.193
		Class B: $x_{i11} = 1$	0.513
		Class C: $x_{i10}, x_{i11} = 0$	0.284

Table: Summary of the covariates.

Application to real insurance count data

Fitting result (EC-LRMoE)

- 5 components detected based on AIC

	$\hat{m}$ $k=1$	$\hat{m}$ $k=2$	$\hat{\beta}$ $k=1$	$\hat{\beta}$ $k=2$	$E[Y_{ik}\|Z_{ij}=1]$ $k=1$	$E[Y_{ik}\|Z_{ij}=1]$ $k=2$	$P(Z_{ij}=1)$
$j=1$	15	3	3.299	0.732	0.000	0.038	0.743
	(12.975,15)	(3,3)	(2.550,7.123)	(0.374,0.799)	(0.000,0.008)	(0.007,0.048)	(0.689,0.756)
$j=2$	3	1	1.470	2.042	0.188	2.042	0.087
	(1,3)	(1,1)	(0.207,1.587)	(1.839,2.229)	(0.147,0.236)	(1.839,2.229)	(0.073,0.107)
$j=3$	1	1	1.585	2.487	1.585	2.487	0.003
	(1,3)	(1,15)	(0.807,7.661)	(1.899,27.036)	(0.752,2.742)	(1.449,3.128)	(0.001,0.010)
$j=4$	2	4	0.650	5.232	0.143	0.933	0.119
	(2,3)	(4,4)	(0.506,1.239)	(4.653,5.570)	(0.100,0.183)	(0.786,1.018)	(0.098,0.166)
$j=5$	6	3	5.665	1.902	0.513	0.310	0.047
	(4,8)	(3,3)	(2.621,7.648)	(1.415,2.465)	(0.191,0.578)	(0.172,0.420)	(0.037,0.076)

Table: Estimates of parameters

- Subgroup 1 is the lowest risk group
- Subgroup 3 is the highest risk group

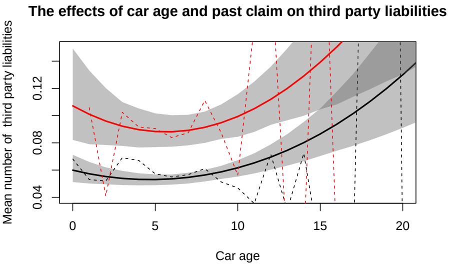
Application to real insurance count data

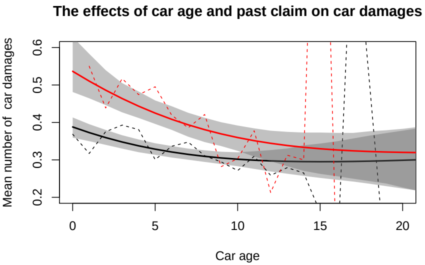
Preliminary fitting

Y_{i1}	Empirical	Fitted			Y_{i2}	Empirical	Fitted		
		NB GLM	ZINB GLM	EC-LRMoE			NB GLM	ZINB GLM	EC-LRMoE
0	16971	16975.06	16976.66	16965.19	0	14182	14177.32	14205.60	14188.88
1	991	972.64	969.88	1001.73	1	2499	2498.57	2386.71	2484.87
2	48	65.90	66.81	40.75	2	752	810.45	883.92	777.23
3	3	4.95	5.14	7.31	3	359	307.02	333.24	317.83
4	5	0.41	0.45	2.82	4	129	125.77	127.56	155.43
5+	1	0.04	0.05	1.20	5	66	54.19	49.51	64.01
					6	22	24.22	19.46	22.04
					7	7	11.15	7.74	6.52
					8+	3	10.30	5.26	2.19
χ^2		**81.31**	**70.61**	**5.66**	χ^2		**22.59**	**33.88**	**11.13**
loglik		-4224.94	-4213.99	-4208.77	loglik		-13279.18	-13204.95	-13178.68
mean	0.062	0.062	0.062	0.062	mean	0.340	0.340	0.340	0.340
% diff		0.011%	-0.001%	0.161%	% diff		-0.011%	-0.037%	-0.007%
variance	0.069	0.068	0.068	0.069	variance	0.649	0.669	0.644	0.650
% diff		-2.148%	-1.857%	-0.426%	% diff		3.077%	-0.805%	0.221%
skewness	5.084	4.522	4.544	5.096	skewness	3.265	3.672	3.305	3.261
% diff		-11.049%	-10.615%	0.244%	% diff		12.451%	1.215%	-0.132%
kurtosis	40.248	26.386	26.755	41.938	kurtosis	16.509	23.063	17.988	16.399
% diff		-34.441%	-33.526%	4.199%	% diff		39.702%	8.962%	-0.668%

Application to real insurance count data

Model visualization – Covariates influences





Application to real insurance count data

Three individual profiles

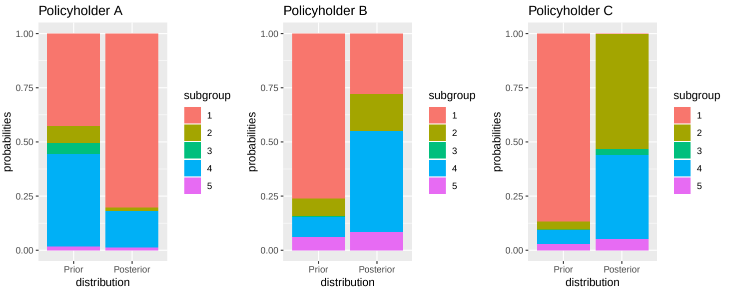
Policyholder	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
A	0	0	36	1	1	1	0	0	0	1	0	0	0
B	0	1	59	7	1	0	1	0	0	1	0	0	1
C	1	2	72	7	0	0	0	1	0	0	0	1	0

Table: Three selected policyholders to be considered for the calculations of subgroup probabilities.

Application to real insurance count data

Model visualization – Subgroup probabilities

- Consider 3 policyholders with very different characteristics
- Prior and posterior subgroup probabilities:



Application to real insurance count data

Predictive Applications

- Construct 6 hypothetical risk profiles: "Best", "Good", "Average 1", "Average 2", "Bad", "Worst"

Profile	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
Best	60	8	0	0	0	0	0	0	0	1	0
Good	40	5	0	0	0	1	0	0	0	1	0
Average 1	40	6	0	0	0	0	0	0	1	0	1
Average 2	80	10	1	0	0	0	0	1	0	0	0
Bad	30	0	1	0	1	0	1	0	0	0	1
Worst	30	1	1	1	0	0	0	0	1	0	0

Table: Six different hypothetical risk profiles to be considered.

Application to real insurance count data

Predictive Applications

Profile	Mean and variance of claims					
	$E[Y_{i1}]$	$\mathrm{Var}[Y_{i1}]$	$E[Y_{i2}]$	$\mathrm{Var}[Y_{i2}]$	$E[L_i]$	$\mathrm{Var}[L_i]$
Best	0.033	0.035	0.179	0.316	0.212	0.396
Good	0.040	0.048	0.246	0.440	0.286	0.565
Average 1	0.050	0.051	0.347	0.710	0.397	0.847
Average 2	0.053	0.054	0.331	0.710	0.383	0.850
Bad	0.098	0.147	0.482	0.873	0.580	1.237
Worst	0.121	0.152	0.677	0.909	0.798	1.227

Table: Mean and variance of the number of claims.

Application to real insurance count data

Predictive Applications

- Various premiums based on the fitted model and various decision rules

	Quantile Premium		SD Premium		SD Prem (Indep)	
Profile	75%	95%	$\gamma = 0.1$	$\gamma = 0.5$	$\gamma = 0.1$	$\gamma = 0.5$
Best	0.219	0.232	0.275	0.526	0.271	0.508
Good	0.302	0.321	0.361	0.662	0.356	0.635
Average 1	0.411	0.427	0.489	0.857	0.484	0.833
Average 2	0.407	0.452	0.476	0.844	0.471	0.820
Bad	0.595	0.680	0.691	1.136	0.681	1.085
Worst	0.835	0.921	0.908	1.352	0.901	1.313

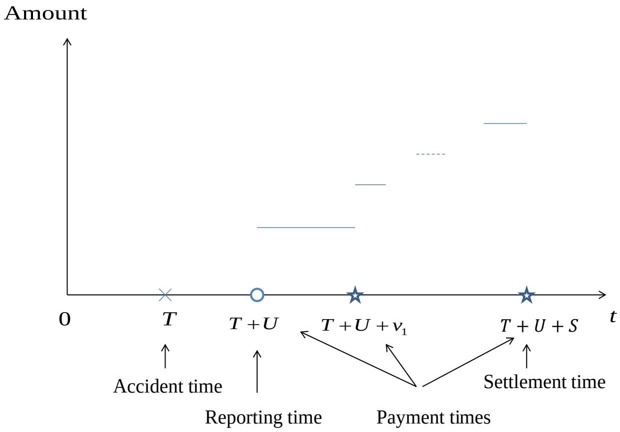
Application to IBNR prediction

A micro-level approach

- Instead of looking at aggregate level as the classical "triangle" type methods, the idea is to look at granular level where one can use more information with respect to each indivudual claim development
- The number of accidents form a homogenous Poisson process
- Let T_i, $i = 1, \dots N^a(t)$, represent the i^{st} arrival eppoch
- Associated to each arrival time T_i we have
- U_i - the reporting delay
- S_i - the settlement delay
- $\mathbf{Z}_i$ - the payment mark

Application to IBNR prediction

A sample path of the claim developmental process



Application to IBNR prediction

Data Overview

- European automobile insurance dataset
- Observation period: 1/1/2007 – 12/31/2012
- Verification period: 1/1/2013 – 01/01/2018
- 594,908 third-party liability insurance contracts
  - Contract number, start date and end date
  - Covariates information $\boldsymbol{x}_i$ available
- 28,256 reported claims
  - Contract number, loss date, reporting date, loss severity

Application to IBNR prediction

Covariate information

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: $x_{k10} = 0$

Table: Summary of the covariates for the k^{th} contract.

Application to IBNR prediction

Modeling framework

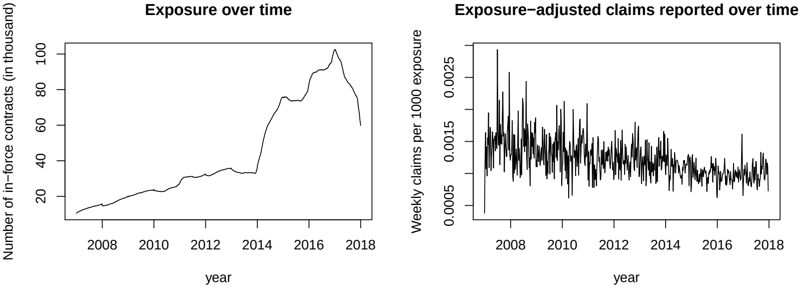


Figure: Number of in-force contracts and weekly number of claims reported (exposure adjusted) versus time.

Application to IBNR prediction

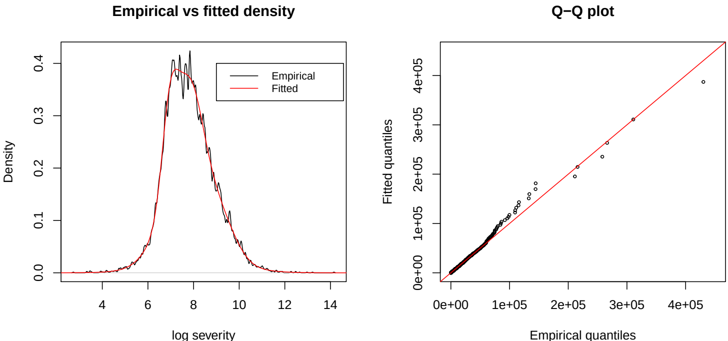
Modeling framework

- Claim arrival process follows a Poisson process with intensity $\lambda(t|\boldsymbol{x}_k) = \omega_k \exp\{\boldsymbol{x}_k^T \beta\}$ with regression coefficients β
- Modeling frequency: Poisson GLM by standard R package
- Modeling severity and reporting delay: Transformed Gamma LRMoE (TG-LRMoE)
- 11 covariates for $\boldsymbol{x}_i$: Policyholder age, car age, geographical location etc.

Application to IBNR prediction

Fitting result (Severity)

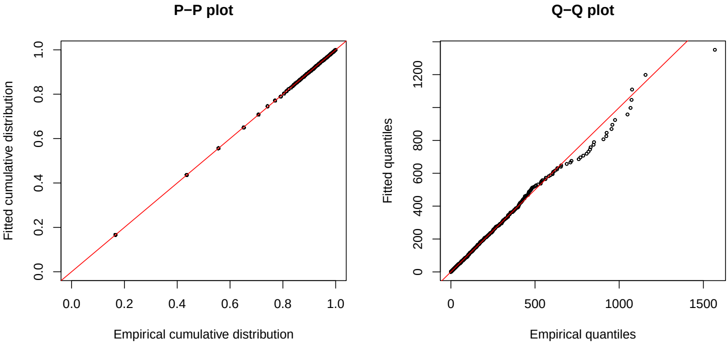
- 9 components detected based on AIC
- Heavy-tailed distribution detected ($\hat{\gamma} = 0.198 < 1$)
- Goodness-of-fit tests:



Application to IBNR prediction

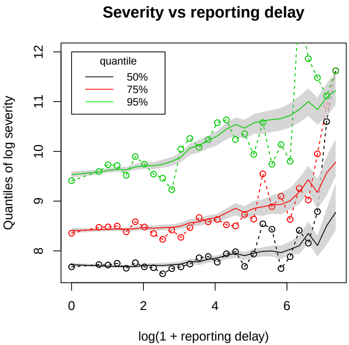
Fitting result (Reporting delay)

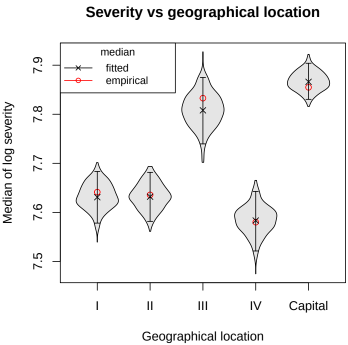
- 8 components detected based on AIC
- Goodness-of-fit tests:



Application to IBNR prediction

- Visualizing covariate influences
  - Relationship between loss severity and reporting delay/ geographical location



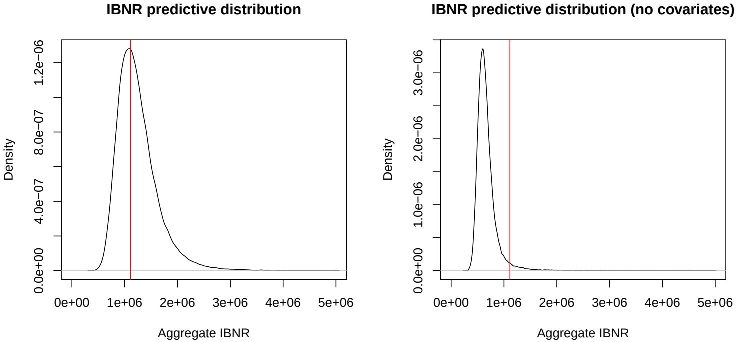


Application to IBNR prediction

- Methodology: Monte-Carlo simulation
  1. Generate the claim frequency, reporting delay and severity directly from the fitted distributions for each contract
  2. Compute the unreported amount of claims
  3. Aggregate the unreported claims across all contracts

Application to IBNR prediction

- Out-of-sample IBNR prediction performance
  - Left: IBNR prediction by the proposed modeling framework
  - Right: IBNR prediction without covariates



Application to IBNR prediction

Why is the inclusion of covariates important for adequate IBNR prediction?

	Mean value/ proportion		
	IBNR claims	All claims	Difference (%)
Policyholder age	46.113	45.994	0.3%
Car age	4.429	3.726	18.9%
Car fuel			
>Diesel	0.471	0.410	14.8%
>Gasoline	0.529	0.590	-10.3%
Geographical location			
>Region I	0.131	0.221	-40.6%
>Region II	0.160	0.227	-29.8%
>Region III	0.112	0.107	4.7%
>Region IV	0.089	0.138	-35.6%
>Capital	0.508	0.307	65.7%
Car brand class			
>Class A	0.247	0.268	-7.7%
>Class B	0.394	0.387	1.8%
>Class C	0.359	0.345	3.9%
Contract type			
>Renewal	0.644	0.572	12.7%
>New	0.356	0.428	-17.0%
Reporting delay (days)	308.583	17.839	1629.8%

Table: The mean value of each covariate for all simulated IBNR claims (simulated from the fitted model) and for all claims (based on the empirical training set).

Concluding Remarks

In this presentation, we...

- Propose the LRMoE as a multivariate regression model
  - Transparent interpretation – Risk group classification
  - Denseness property – "Full flexibility"
  - Mathematical and computational tractability – Moment and marginalization properties, efficient model calibration algorithm
- Erlang Count expert function for frequency applications
  - Verify the denseness property empirically
  - Real insurance count data – Demonstrate usefulness of EC-LRMoE in insurance ratemaking

Concluding Remarks

- Transformed Gamma expert function for severity modeling
  - Capturing a wide range of tail behavior
  - Fitting severity and reporting delay component of real insurance dataset
  - Predicting adequate IBNR reserves through out-of-sample testing

Upcoming ...

- Fitting censored and truncated data
  - Developing efficient fitting algorithm
  - Modeling reporting delay
  - Modeling losses subject to deductibles and policy limits

- An R package for insurance loss modeling implementation!!

Selected References I

Fung, T. C., Badescu, A. L. and Lin X. S.
A class of Mixture of Experts Models for General Insurance: Theoretical developments.
Insurance: Mathematics and Economics, in press, 2019.

Fung, T. C., Badescu, A. L. and Lin X. S.
A class of Mixture of Experts Models for General Insurance: Application to correlated claim frequencies.
ASTIN Bulletin: The Journal of the IAA, accepted, 2019.

Fung, T. C., Badescu, A. L. and Lin X. S.
A new class of severity regression models with an application to IBNR prediction.
Submitted, 2019.

Fung, T. C., Badescu, A. L. and Lin X. S.
Fitting censored and truncated regression data using the Mixture of Experts models.
Working Paper.

– Thank You –