L-estimation of Claim Severity Models: A Flexible Weighting Approach

Chudamani Poudyal[1]

Department of Statistics & Data Science
University of Central Florida

THE 60TH ACTUARIAL RESEARCH CONFERENCE (ARC–2025)

July 29 – August 1, 2025 | Toronto, Canada

[1] In partial collaboration with:

- *Gokarna R. Aryal*, Purdue University Northwest
- *Keshav Pokhrel*, University of Michigan-Dearborn

Outline

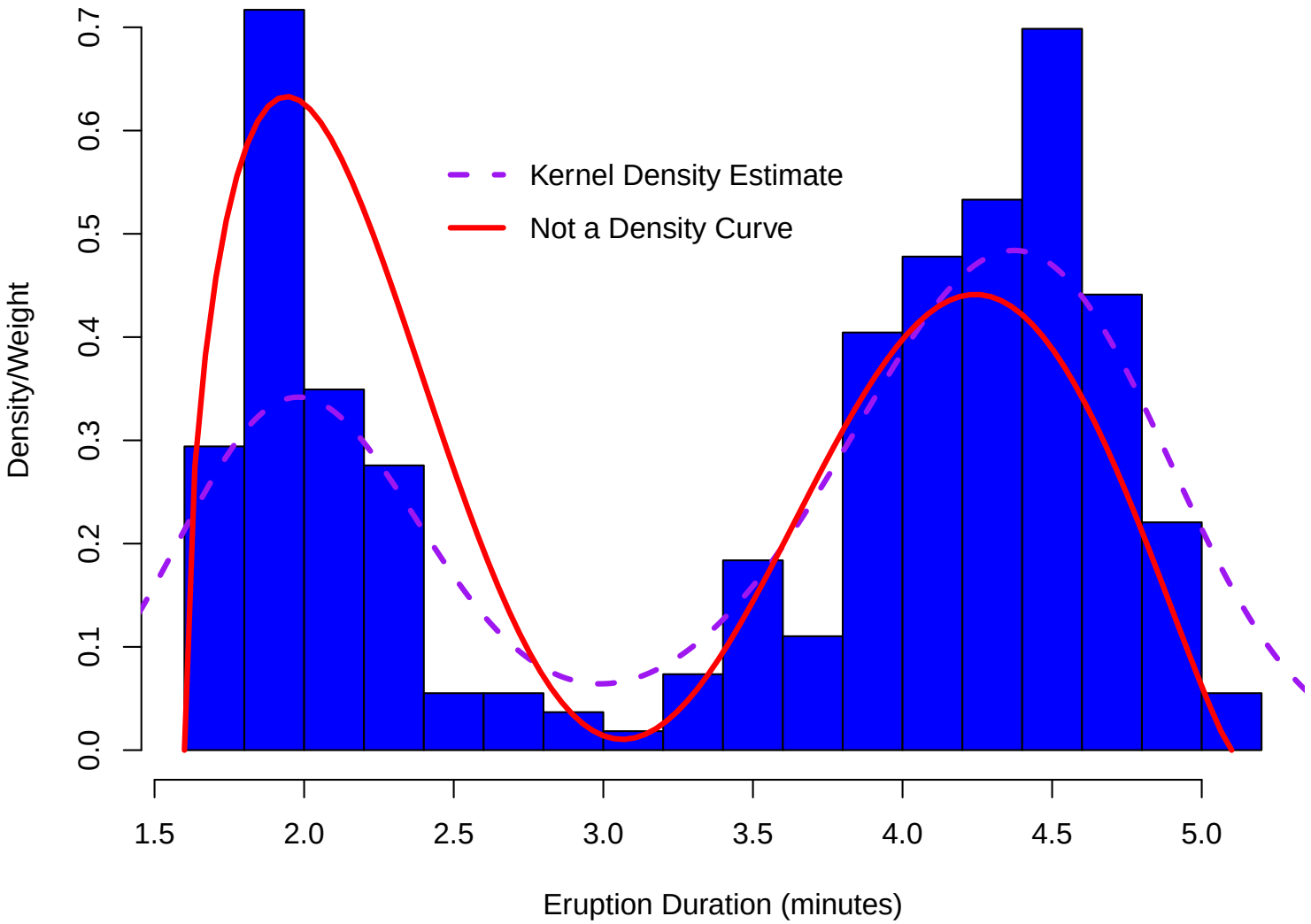
Faithful Geyser Data

Consider the *Old Faithful Geyser Data*, available on the base R-package `datasets`, representing the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

```
A data frame with 272 observations on 2 variables.

[,1]eruptions numeric Eruption time in mins
[,2]waiting numeric Waiting time to next eruption (in mins)
```

Histogram of Eruption Time



Some Notes

- The histogram clearly shows bimodal pattern.
- A single uni-modal distribution is not suitable for this dataset.
- The contemporary financial loss modeling literature has been directed toward composite loss modeling:
  - Abu Bakar et al. (2015) – *Modeling loss data using composite models*
  - Reynkens et al. (2017) – *Modeling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions*
  - Grün and Miljkovic (2019) – *Extending composite loss models using a general framework of advanced computational tools*
  - Delong et al. (2021) – *Gamma mixture density networks and their application to modeling insurance claim amounts*
- Many more ...

This Research Motivation

- Instead of using composite/slicing or mixture models, simply use a standalone model with a FLEXIBLE weighting mechanism that can capture the multimodal nature observed in the sample data (e.g., faithful data).
- Robust weighted MLE can be used (see, e.g., Fung, 2025).
- Need to search for a weight density function having support in $(0,1)$ with two goals, at least: 📢
  - Can be accommodated as desired with the multi-modal nature of the data.
  - To control the influence of possible outliers in certain parts of the data, like left tail and/or right tail, etc.
- To achieve the desired goals, we go with L-estimation strategy, Poudyal et al. (2025).

L – Statistics

- Parameter Vector: $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$
- Sample order statistics: $X_{1:n} \leq \ldots \leq X_{n:n}$
- L–statistics:

$$\widehat{\mu}_j := \frac{1}{n} \sum_{i=1}^{n} J_j \left(\frac{i}{n+1} \right) h_j(X_{i:n}), \quad 1 \leq j \leq k, \tag{1}$$

where $J_j : [0,1] \to \mathbb{R}_{\geq 0}$ represents a *weights-generating* function. Both J_j and h_j are specially chosen functions, (see, e.g., Poudyal, 2025).

- Population quantities are then given by

$$\mu_j \equiv \mu_j(\boldsymbol{\theta}) \equiv \mu_j(\theta_1, \ldots, \theta_k) = \int_0^1 J_j(u) H_j(u)\, du, \quad 1 \leq j \leq k, \tag{2}$$

where $H_j := h_j \circ F^{-1}$.

Two Asymptotic Results

Theorem 1 (Chernoff et al., 1967).

The k-variate vector $\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim \mathcal{AN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} := \left[\sigma_{ij}^2\right]_{i,j=1}^{k}$ with the entries

$$\sigma_{ij}^2 = \int_0^1 \alpha_i(u)\alpha_j(u)\,du = \int_0^1 \int_0^1 J_i(v)J_j(w)K(v,w)\,dH_i(v)\,dH_j(w), \qquad (3)$$

where $K(v,w) := K(w,v) = \min\{v,w\} - vw, \quad$ *for* $\quad 0 \le v, w \le 1.$

Theorem 2.

The L-estimator of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, has the following asymptotic distribution:

$$\widehat{\boldsymbol{\theta}} \;=\; \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right) \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}'\right), \quad \boldsymbol{D} = \left[\left.\frac{\partial g_i}{\partial \widehat{\mu}_j}\right|_{\widehat{\boldsymbol{\mu}}=\boldsymbol{\mu}}\right]_{k\times k} =: \left[d_{ij}\right]_{k\times k} \qquad (4)$$

Performance Criteria

The asymptotic performance of the L-estimators is evaluated in terms of asymptotic relative efficiency (ARE) compared to the maximum likelihood estimator (MLE).

For a scenario involving k parameters, the ARE is defined as:

$$ARE(\mathcal{C}, MLE) = \left(\frac{\det\left(\boldsymbol{\Sigma}_{\text{MLE}}\right)}{\det\left(\boldsymbol{\Sigma}_{\mathcal{C}}\right)} \right)^{1/k}, \tag{5}$$

where $\boldsymbol{\Sigma}_{\text{MLE}}$ and $\boldsymbol{\Sigma}_{\mathcal{C}}$ denote the asymptotic covariance matrices of the MLE and $\mathcal{C}$ estimators, respectively, and $\det$ represents the determinant of a square matrix.

Performance Criteria

The asymptotic performance of the L-estimators is evaluated in terms of asymptotic relative efficiency (ARE) compared to the maximum likelihood estimator (MLE).

For a scenario involving k parameters, the ARE is defined as:

$$ARE(\mathcal{C}, MLE) = \left(\frac{\det\left(\boldsymbol{\Sigma}_{\text{MLE}}\right)}{\det\left(\boldsymbol{\Sigma}_{\mathcal{C}}\right)} \right)^{1/k}, \tag{5}$$

where $\boldsymbol{\Sigma}_{\text{MLE}}$ and $\boldsymbol{\Sigma}_{\mathcal{C}}$ denote the asymptotic covariance matrices of the MLE and $\mathcal{C}$ estimators, respectively, and $\det$ represents the determinant of a square matrix.

NOTE: Optimal solution of $ARE(\mathcal{C}, MLE) = 1$ can be achieved, Serfling (1980, p. 268), for the case of estimating location and scale parameters, but the weights-generating function may not behave as expected.

An Important Inequality

Theorem 3 (Poudyal et al. 2025).

Let $f_i : (0,1) \to \mathbb{R}$, $1 \leq i \leq 2$, be two non-zero functions such that $f_1, f_2 \in L^2$-space and f_1 and f_2 are not linearly dependent. Consider the following integrals:

$$\begin{aligned}
\Omega_1 &:= \int_0^1 \int_0^1 f_1(x)\, f_1(y)\, K(x,y)\, dy\, dx, \\
\Omega_2 &:= \int_0^1 \int_0^1 f_1(x)\, f_1(y)\, f_2(y)\, K(x,y)\, dy\, dx, \\
\Omega_3 &:= \int_0^1 \int_0^1 f_1(x)\, f_1(y)\, f_2(x)\, f_2(y)\, K(x,y)\, dy\, dx,
\end{aligned}$$

where $K(\cdot,\cdot)$ is defined in Theorem 1. Then, $\Omega_2^2 < \Omega_1\, \Omega_3$.

Possible Weights-generating Densities

Any continuous probability density with support in $(0,1)$ can be used as a weights-generating function.

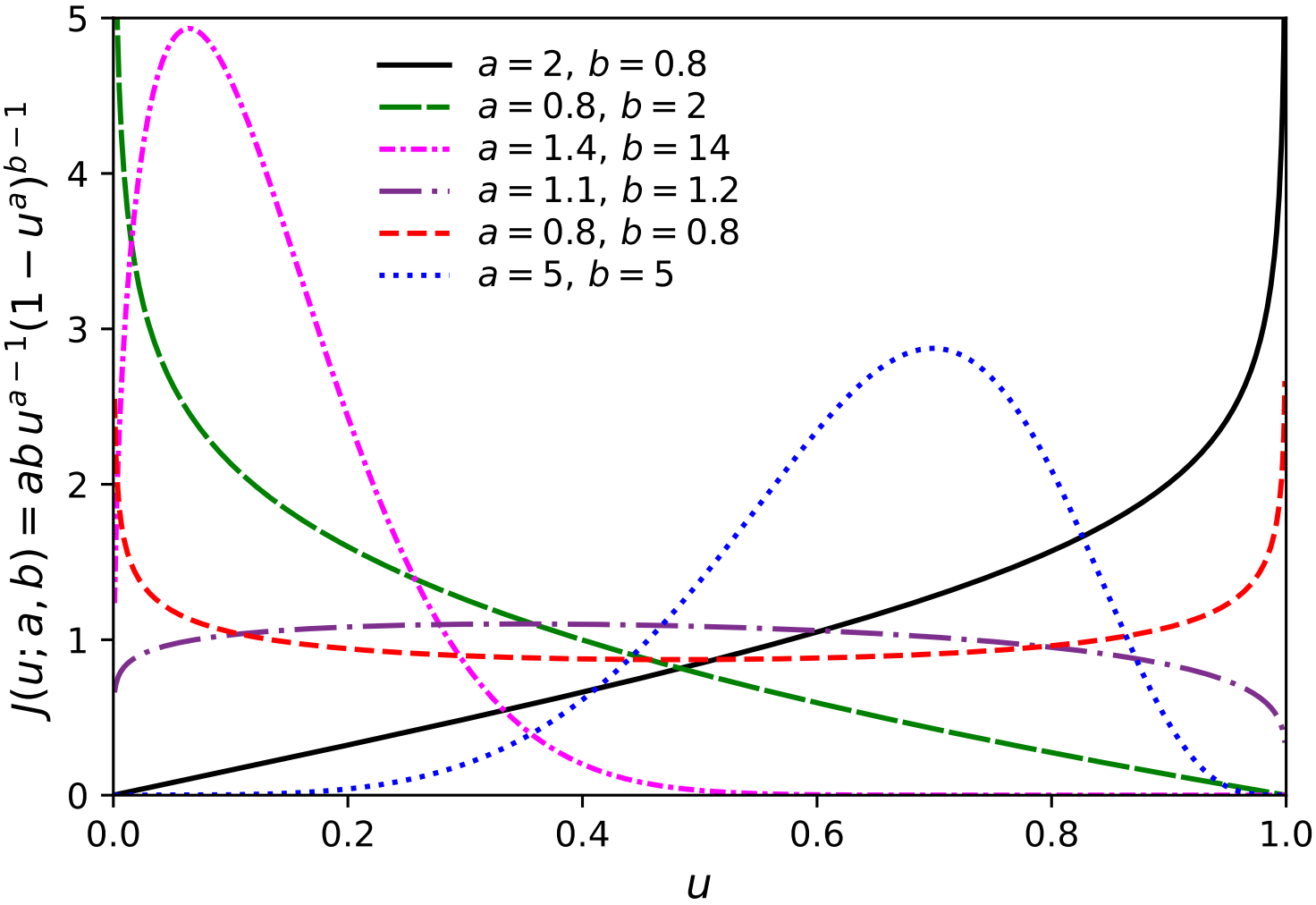
Possible Weights-generating Densities

Any continuous probability density with support in $(0,1)$ can be used as a weights-generating function.

- Uniform density – not very useful.
- Beta density – can be used, but not very flexible in order to capture the multi-modal nature of the data.
- Kumaraswamy (Kum) density – still not able to capture the multi-modal nature of the data, but more easy to implement as a weights-generating function – Hyper-parameterization!

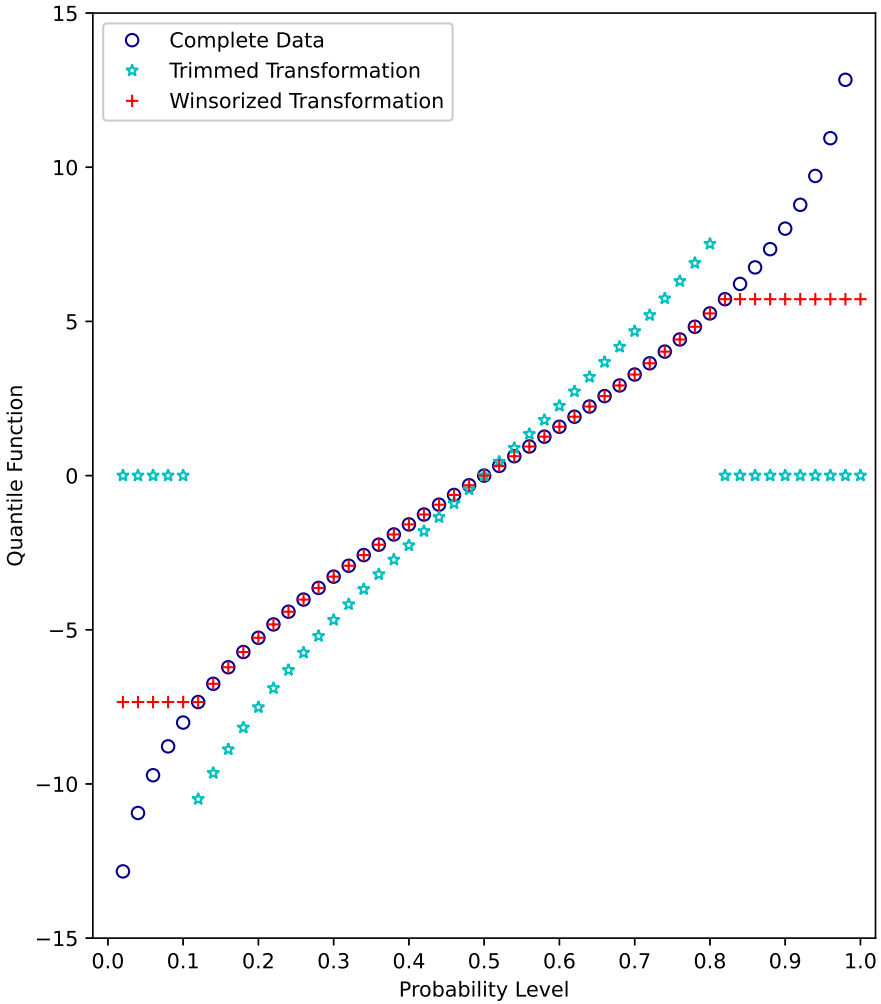
📢 We will go with this density in this presentation.

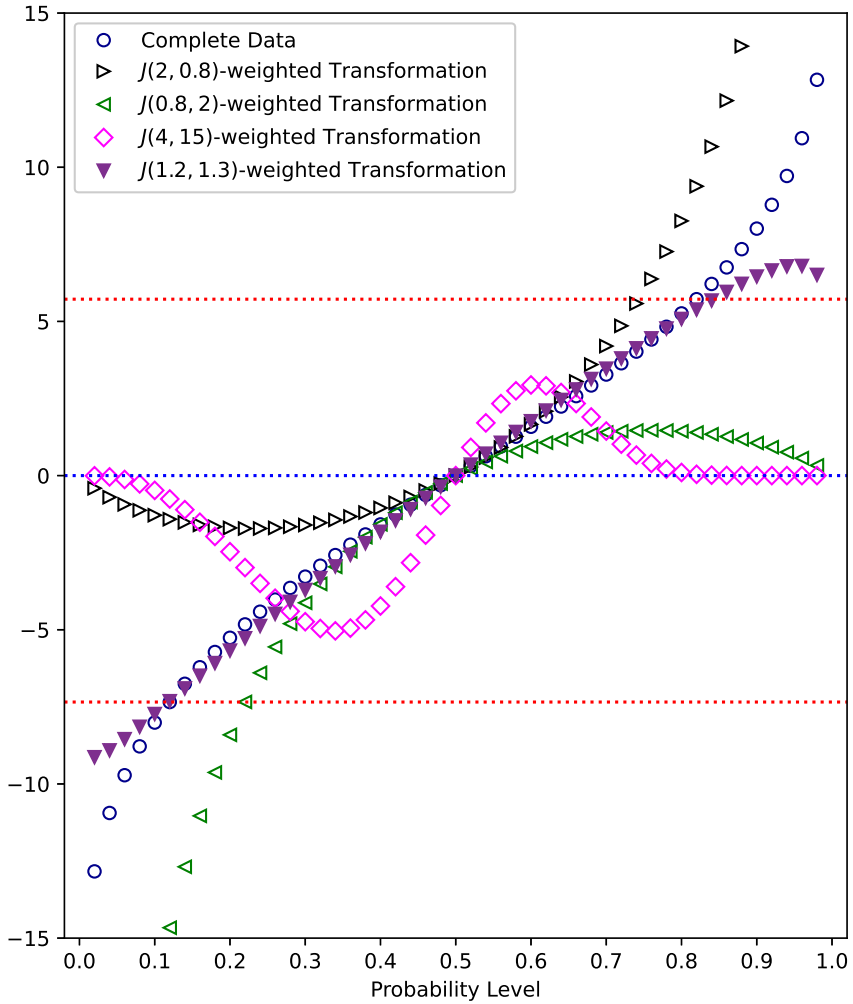
Shapes of the pdf of the Kum Distribution



▸ Go to LN ARE
▸ Go to KS Table
▸ Go to Fitted Models

Visualizing the Data Transformation





Lognormal Model

Loss severity r.v. $X \sim \text{Lognormal}(\theta, \sigma)$. Then, by using Kum density weighted L-estimators, we get

- Parameter Estimation

$$\begin{cases} \widehat{\theta}_{\text{K}} &= \widehat{\mu}_1 - c_1 \widehat{\sigma}_{\text{K}} =: g_1(\widehat{\mu}_1, \widehat{\mu}_2), \\ \widehat{\sigma}_{\text{K}} &= \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/\eta} =: g_2(\widehat{\mu}_1, \widehat{\mu}_2). \end{cases} \tag{6}$$

- ARE Calculation

$$\text{ARE}\left(\left(\widehat{\theta}_{\text{K}}, \widehat{\sigma}_{\text{K}}\right), \left(\widehat{\theta}_{\text{MLE}}, \widehat{\sigma}_{\text{MLE}}\right)\right) = \left(\frac{\left(c_2 - c_1^2\right)^2}{2\left(\Lambda_1 \Lambda_3 - \Lambda_2^2\right)}\right)^{0.5}. \tag{7}$$

- The constants c_1, c_2, η, Λ_1, Λ_2, and Λ_3 do not depend on the parameters.

Lognormal ARE Calculation

Back to Kumaraswamy Densities

a	b							
	0.3	0.8	1	1.3	2	5	15	20
0.3	0.142	0.204	0.152	0.103	0.053	0.017	0.013	0.014
0.5	0.214	0.582	0.479	0.356	0.203	0.054	0.012	0.009
0.8	0.176	0.962	0.950	0.846	0.623	0.262	0.080	0.058
1.0	0.152	0.950	1.000	0.955	0.782	0.412	0.164	0.128
1.2	0.133	0.892	0.977	0.974	0.853	0.518	0.247	0.201
2.0	0.092	0.662	0.782	0.847	0.844	0.672	0.449	0.401
4.0	0.057	0.389	0.489	0.565	0.622	0.608	0.520	0.495
5.0	0.050	0.323	0.412	0.483	0.544	0.555	0.499	0.482
7.0	0.040	0.242	0.314	0.376	0.435	0.467	0.446	0.437
10.0	0.033	0.177	0.233	0.283	0.336	0.377	0.377	0.374

ARE $\left(\left(\widehat{\theta}_{\mathrm{K}}, \widehat{\sigma}_{\mathrm{K}}\right), \left(\widehat{\theta}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}\right)\right)$ for selected values of a and b.

Lognormal ARE Calculation

Back to Kumaraswamy Densities

a	b							
	0.3	0.8	1	1.3	2	5	15	20
0.3	0.142	0.204	0.152	0.103	0.053	0.017	0.013	0.014
0.5	0.214	0.582	0.479	0.356	0.203	0.054	0.012	0.009
0.8	0.176	0.962	0.950	0.846	0.623	0.262	0.080	0.058
1.0	0.152	0.950	1.000	0.955	0.782	0.412	0.164	0.128
1.2	0.133	0.892	0.977	0.974	0.853	0.518	0.247	0.201
2.0	0.092	0.662	0.782	0.847	0.844	0.672	0.449	0.401
4.0	0.057	0.389	0.489	0.565	0.622	0.608	0.520	0.495
5.0	0.050	0.323	0.412	0.483	0.544	0.555	0.499	0.482
7.0	0.040	0.242	0.314	0.376	0.435	0.467	0.446	0.437
10.0	0.033	0.177	0.233	0.283	0.336	0.377	0.377	0.374

ARE $\left(\left(\widehat{\theta}_{\mathrm{K}}, \widehat{\sigma}_{\mathrm{K}}\right), \left(\widehat{\theta}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}\right)\right)$ for selected values of a and b.

Lognormal ARE Calculation

Back to Kumaraswamy Densities

a	b							
	0.3	0.8	1	1.3	2	5	15	20
0.3	0.142	0.204	0.152	0.103	0.053	0.017	0.013	0.014
0.5	0.214	0.582	0.479	0.356	0.203	0.054	0.012	0.009
0.8	0.176	0.962	0.950	0.846	0.623	0.262	0.080	0.058
1.0	0.152	0.950	1.000	0.955	0.782	0.412	0.164	0.128
1.2	0.133	0.892	0.977	0.974	0.853	0.518	0.247	0.201
2.0	0.092	0.662	0.782	0.847	0.844	0.672	0.449	0.401
4.0	0.057	0.389	0.489	0.565	0.622	0.608	0.520	0.495
5.0	0.050	0.323	0.412	0.483	0.544	0.555	0.499	0.482
7.0	0.040	0.242	0.314	0.376	0.435	0.467	0.446	0.437
10.0	0.033	0.177	0.233	0.283	0.336	0.377	0.377	0.374

ARE $\left(\left(\widehat{\theta}_{\mathrm{K}}, \widehat{\sigma}_{\mathrm{K}}\right), \left(\widehat{\theta}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}\right)\right)$ for selected values of a and b.

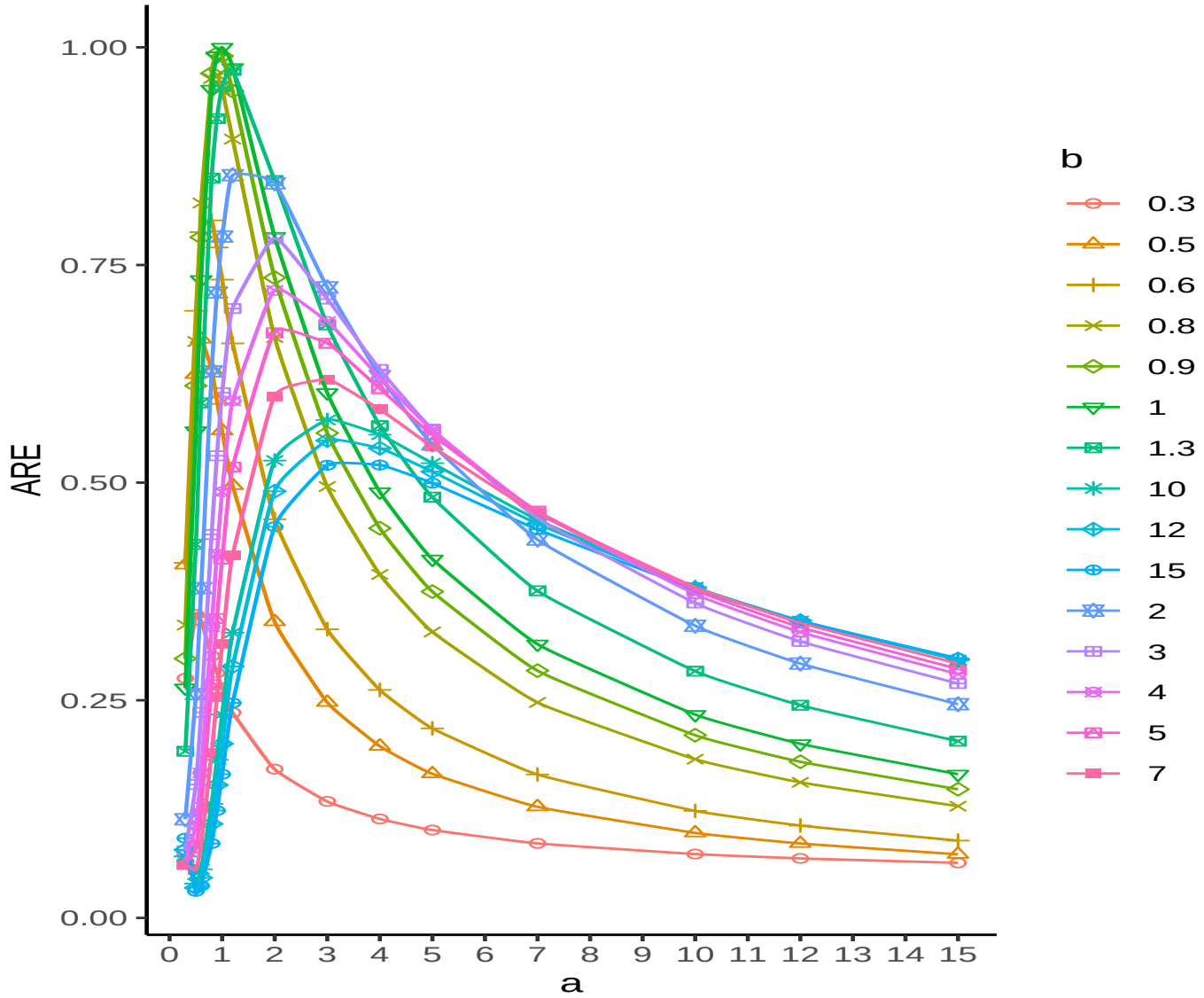
Lognormal ARE Calculation

Back to Kumaraswamy Densities

a	b							
	0.3	0.8	1	1.3	2	5	15	20
0.3	0.142	0.204	0.152	0.103	0.053	0.017	0.013	0.014
0.5	0.214	0.582	0.479	0.356	0.203	0.054	0.012	0.009
0.8	0.176	0.962	0.950	0.846	0.623	0.262	0.080	0.058
1.0	0.152	0.950	1.000	0.955	0.782	0.412	0.164	0.128
1.2	0.133	0.892	0.977	0.974	0.853	0.518	0.247	0.201
2.0	0.092	0.662	0.782	0.847	0.844	0.672	0.449	0.401
4.0	0.057	0.389	0.489	0.565	0.622	0.608	0.520	0.495
5.0	0.050	0.323	0.412	0.483	0.544	0.555	0.499	0.482
7.0	0.040	0.242	0.314	0.376	0.435	0.467	0.446	0.437
10.0	0.033	0.177	0.233	0.283	0.336	0.377	0.377	0.374

ARE $\left(\left(\widehat{\theta}_{\mathrm{K}}, \widehat{\sigma}_{\mathrm{K}}\right), \left(\widehat{\theta}_{\mathrm{MLE}}, \widehat{\sigma}_{\mathrm{MLE}}\right)\right)$ for selected values of a and b.

Lognormal ARE Graphs



Fréchet Model

Loss severity r.v. $X \sim$ Fréchet(θ, α, σ), but with $\theta = 0$. Then, by using Kum density weighted L-estimators, we get

- Parameter Estimation

$$\begin{cases} \widehat{\alpha}_{\mathsf{K}} = \sqrt{\dfrac{\tau}{\widehat{\mu}_2 - \widehat{\mu}_1^2}} =: g_1\left(\widehat{\mu}_1, \widehat{\mu}_2\right), \quad \text{where} \quad \tau \equiv \tau(J) := \kappa_2 - \kappa_1^2, \\ \widehat{\sigma}_{\mathsf{K}} = \exp\left\{\widehat{\mu}_1 + \dfrac{\kappa_1}{\widehat{\alpha}_{\mathsf{K}}}\right\} =: g_2\left(\widehat{\mu}_1, \widehat{\mu}_2\right). \end{cases}$$

- ARE Calculation

$$\text{ARE}\left(\left(\widehat{\theta}_{\mathsf{K}}, \widehat{\sigma}_{\mathsf{K}}\right), \left(\widehat{\theta}_{\mathsf{MLE}}, \widehat{\sigma}_{\mathsf{MLE}}\right)\right) = \left(\frac{6\pi^{-2}\tau^2}{\Psi_1\Psi_3 - \Psi_2^2}\right)^{0.5}, \tag{9}$$

- The constants κ_1, κ_2, τ, Ψ_1, Ψ_2, and Ψ_3 do not depend on the parameters.

Simulation Study

Comprehensive Simulation Study is available on the paper Poudyal et al. (2025) for the following three parametric distributions.

- Pareto Distribution
- Lognormal Distribution
- Fréchet Distibution

Data Introduction

- 1500 US indemnity losses.
- Sample Size: $n = 1500$.
- Found to fit Lognormal Model, see, e.g., Poudyal et al. (2024).
- Purely for illustrative/visualization purposes and to enable a clearer visualization, as it is challenging to represent all 1,500 sample observations, a random subsample of size 50 is extracted using `seed(123)`.
- Modified Data – The largest observation of 2,173,595 is inflated to be 10^7 (10 million).

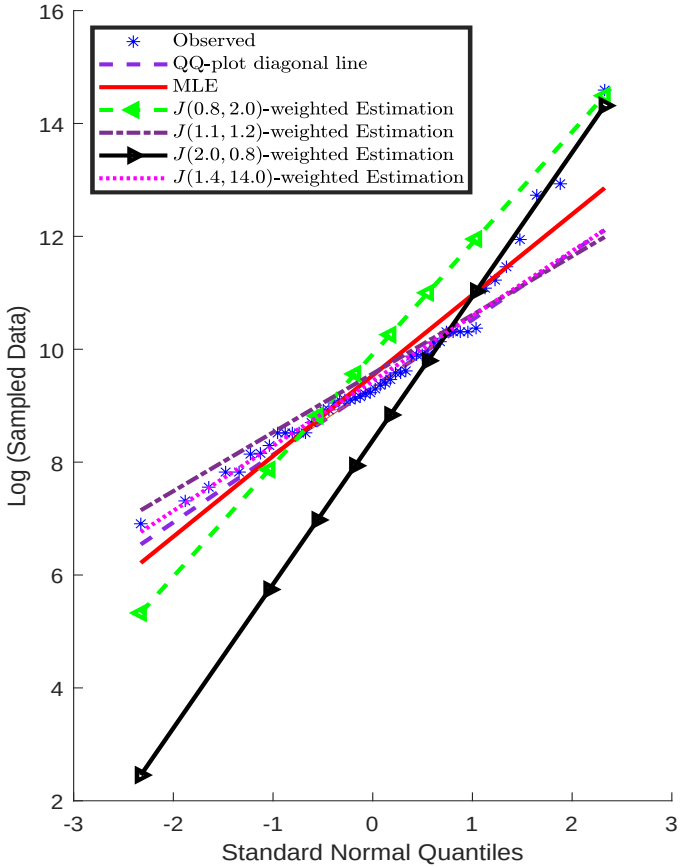
KS and CvM Test Statistics

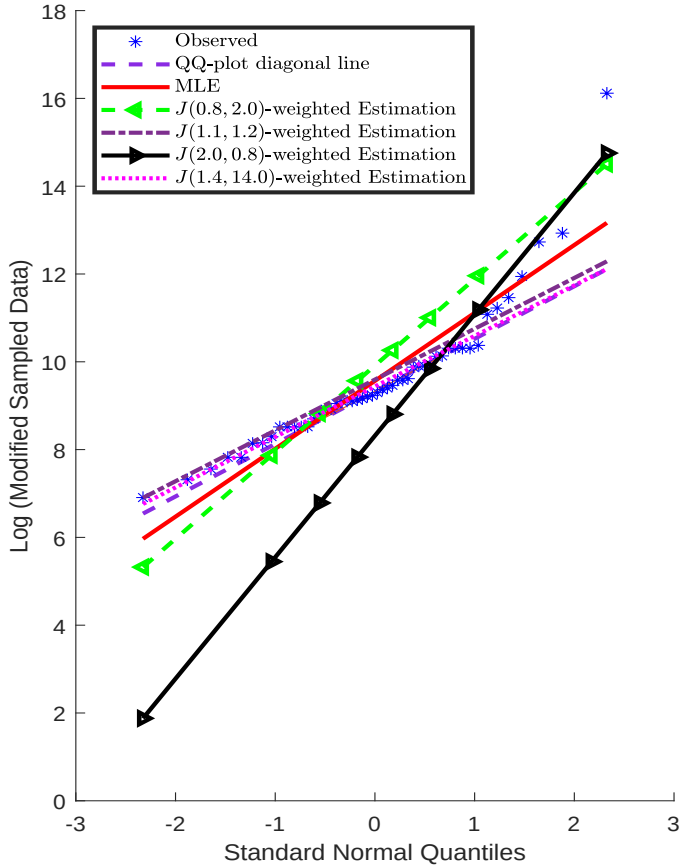
KS $\overset{\text{def}}{=}$ Kolmogorov-Smirnov and CvM $\overset{\text{def}}{=}$ Cramér-von Mises

Models	$(\widehat{\theta},\widehat{\sigma})$	KS Test		CvM	$(\widehat{\theta},\widehat{\sigma})$	KS Test		CvM
		p-value	D			p-value	D	
–	Original Data				Modified Original Data			
MLE	(9.37, 1.64)	0.24	0.03	0.11	(9.38, 1.64)	0.23	0.03	0.12
$J(1.1,1.2)$	(9.38, 1.63)	0.29	0.03	0.11	(9.38, 1.63)	0.29	0.03	0.11
–	Sampled Data				Modified Sampled Data			
MLE	(9.54, 1.43)	0.27	0.14	0.19	(9.57, 1.55)	0.13	0.16	0.25
$J(0.8,2.0)$	(9.91, 1.97)	0.00	0.27	0.81	(9.91, 1.97)	0.00	0.27	0.82
$J(1.1,1.2)$	(9.57, 0.74)	0.36	0.13	0.22	(9.60, 0.89)	0.31	0.13	0.23
$J(2.0,0.8)$	(8.39, 2.55)	0.00	0.36	2.03	(8.32, 2.77)	0.00	0.38	2.16
$J(1.4,14.0)$	(9.44, 1.15)	0.89	0.08	0.07	(9.44, 1.15)	0.89	0.08	0.07

▸ Back to Kumaraswamy Densities

Visualization of Fitted Models





Quantile-quantile plot and lognormal fits for the sampled US Indemnity loss data (left panel) and the modified version of the sampled data (right panel).

Back to Kumaraswamy Densities

Summary

- Present a flexible and robust L-estimation framework weighted by Kumaraswamy densities, addressing the limitations of the method of trimmed moments (MTM) and the method of winsorized moments (MWM).
- Explicit formulations for L-estimators were developed for key parametric models along with inferential justification on asymptotic normality and variance-covariance structures.
- A comprehensive simulation study is available on the paper.
- The implementation of the methodology on a real dataset demonstrated its superior performance in handling outliers and heavy-tailed distributions for predictive loss severity modeling.

Future Directions I

The findings of this study suggest several promising directions for future research.

- Implementation of this weighting mechanism in generalized linear models – ALGORITHMIC APPROACH.

- The ▸ Kumaraswamy Densities examined fail to capture the multimodal nature of financial loss data, such as that in ▸ Faithful Eruption Time.

  This raises an open question: how can we identify probability densities with support in $(0, 1)$ that exhibit a multimodal structure, as illustrated on the next page(s)?

  We expect that the corresponding L–estimation will offer superior performance in terms of the robustness-efficiency trade-off compared to contemporary mixture modeling approaches.

Future Directions II

For $a > 0, b > 0$ and $x \in (0,1)$, the following are the valid probability density functions:

- For $-1 \leq \lambda \leq 3$,

$$f_1(x) = abx^{a-1}(1-x^a)^{b-1}\left[1+\lambda-4\lambda(1-x^a)^b+3\lambda(1-x^a)^{2b}\right].$$

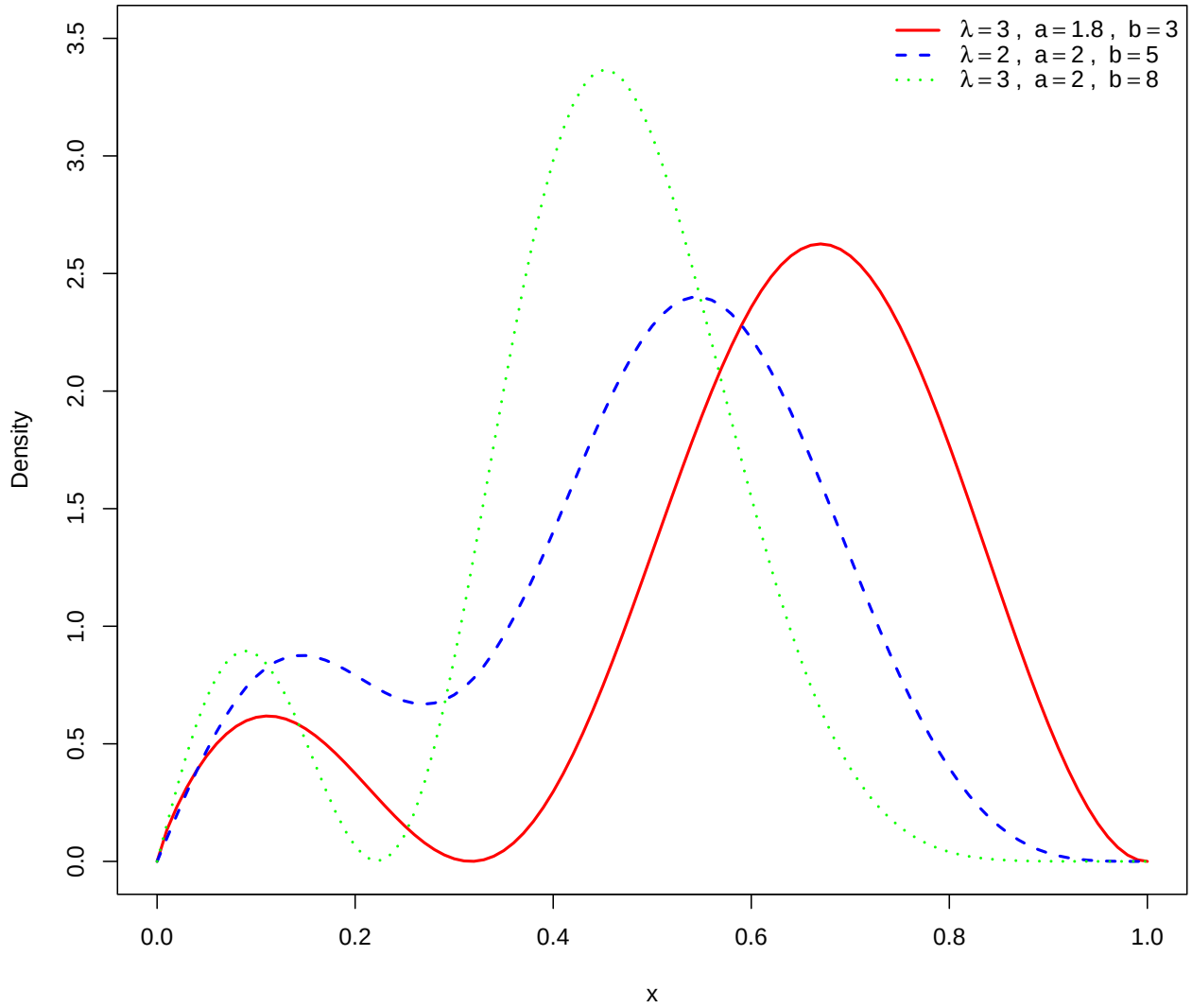
- For $0 \leq \lambda \leq 3$,

$$f_2(x) = abx^{a-1}(1-x^a)^{b-1}\left[3-\lambda+2\lambda(1-x^a)^b-6(1-x^a)^b+3(1-x^a)^{2b}\right]$$

📢 Both f_1 and f_2 can generate bimodal densities.

Future Directions III

Graphs of f_1 for different values of λ.



Mixture density to create more than two modes?

Selected References I

Abu Bakar, S. A., Hamzah, N. A., Maghsoudi, M., and Nadarajah, S. (2015). Modeling loss data using composite models. *Insurance: Mathematics & Economics*, 61:146–154.

Chernoff, H., Gastwirth, J. L., and Johns, Jr., M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Annals of Mathematical Statistics*, 38(1):52–72.

Delong, Ł., Lindholm, M., and Wüthrich, M. V. (2021). Gamma mixture density networks and their application to modelling insurance claim amounts. *Insurance: Mathematics & Economics*, 101:240–261.

Fung, T. C. (2025). Robust estimation and diagnostic of generalized linear model for insurance losses: a weighted likelihood approach. *Metrika. International Journal for Theoretical and Applied Statistics*, 88(2):149–182.

Grün, B. and Miljkovic, T. (2019). Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal*, 2019(8):642–660.

Selected References II

Poudyal, C. (2025). On the asymptotic normality of trimmed and winsorized L-statistics. *Communications in Statistics – Theory and Methods*, 54(10):3114–3133.

Poudyal, C., Aryal, G. R., and Pokhrel, K. P. (2025). L-estimation of claim aeverity models: Weighted by Kumaraswamy density. *Insurance: Mathematics & Economics*, accepted.

Poudyal, C., Zhao, Q., and Brazauskas, V. (2024). Method of winsorized moments for robust fitting of truncated and censored lognormal distributions. *North American Actuarial Journal*, 28(1):236–260.

Reynkens, T., Verbelen, R., Beirlant, J., and Antonio, K. (2017). Modelling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions. *Insurance: Mathematics & Economics*, 77:65–77.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Thanks & Questions

THANK YOU ALL!

QUESTIONS???