



## PEMODELAN PREMI ASURANSI BERDASARKAN DATA SEVERITY MENGUNAKAN MODEL LOGNORMAL

### *Insurance Premium Modeling Based on Severity Data using A Lognormal Model*

Putri Isnaini Cahyaning Baiti\*<sup>1</sup>, Annisa Hevita G.K.S<sup>2</sup>, Karina Sylfia Dewi<sup>3</sup>,  
Nanda Azzanina<sup>4</sup>

<sup>1,2,3,4</sup>Institut Teknologi Sumatera

\*Email: putri.baiti@at.itera.ac.id

#### **Abstract**

*The insurance industry in Indonesia requires reliable quantitative approaches to accurately determine premium rates and manage claim risks effectively. This study aims to model pure insurance premiums based on claim severity data using the lognormal regression approach. The data used consist of historical individual claim amounts (severity) obtained from a general insurance company in Indonesia, covering the period from 2009 to 2015. Initial data exploration revealed that the distribution of claim values is positively skewed, indicating the suitability of lognormal modeling. Three models were evaluated: Generalized Linear Model (GLM) with Gamma distribution, GLM with Inverse Gaussian distribution, and linear regression with lognormal transformation. Model selection was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The results show that the lognormal model had the lowest AIC and BIC values, indicating superior performance compared to the other models. The selected model was then used to forecast pure premiums for the next 12 months, followed by the calculation of commercial premiums with a 30% loading factor. The prediction results show a consistent and proportional upward trend in premiums, demonstrating the model's effectiveness in capturing historical claim patterns and supporting data-driven premium setting.*

**Keywords:** AIC and BIC; Claim data; Insurance Premium; Lognormal Distribution; Severity Modeling

#### **Abstrak**

*Industri asuransi di Indonesia memerlukan pendekatan kuantitatif yang andal untuk menentukan besaran premi agar dapat mengelola risiko klaim secara optimal. Penelitian ini bertujuan untuk memodelkan premi murni asuransi berdasarkan data nilai klaim (severity) menggunakan pendekatan regresi lognormal. Data yang digunakan merupakan data historis klaim nominal dari salah satu perusahaan asuransi kerugian di Indonesia selama periode 2009 hingga 2015. Proses analisis dimulai dengan eksplorasi sebaran data yang menunjukkan pola distribusi positif dan berpola skew ke kanan, kemudian dilanjutkan dengan pemodelan regresi menggunakan tiga pendekatan: GLM Gamma, GLM Inverse Gaussian, dan regresi lognormal. Evaluasi performa model dilakukan berdasarkan kriteria Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC). Hasil analisis menunjukkan bahwa model lognormal memiliki nilai AIC dan BIC paling rendah, sehingga dipilih sebagai model terbaik. Model ini kemudian digunakan untuk memprediksi premi murni selama 12 bulan ke depan, dengan penyesuaian premi komersial sebesar 30% loading. Hasil prediksi menunjukkan pola pertumbuhan premi yang stabil dan proporsional, menjadikan model ini layak digunakan sebagai dasar perhitungan tarif premi berbasis*

*data historis klaim.*

**Kata Kunci:** AIC dan BIC; Data Klaim; Distribusi Lognormal; Model Severity; Premi Asuransi

## PENDAHULUAN

Industri asuransi di Indonesia memiliki potensi besar seiring dengan meningkatnya jumlah penduduk dan beragamnya risiko yang dihadapi masyarakat, seperti kehilangan atau kerusakan aset, kecelakaan, dan bencana. Untuk menghadapi risiko-risiko tersebut, banyak individu dan badan usaha memilih menggunakan asuransi sebagai sarana perlindungan. Asuransi merupakan suatu sistem pengalihan risiko, di mana risiko kerugian finansial ditransfer dari satu pihak (tertanggung) ke pihak lain (penanggung). Dalam transaksi ini, penanggung menjamin bahwa pihak tertanggung akan memperoleh penggantian atas kerugian yang dialaminya akibat suatu peristiwa yang belum diketahui waktu terjadinya, dengan kewajiban bagi tertanggung untuk membayar sejumlah uang yang disebut premi (Booth *et al.*, 2020).

Bagi perusahaan asuransi kerugian, seperti yang menyediakan perlindungan atas kendaraan bermotor, pengelolaan risiko klaim menjadi aspek penting. Semakin banyak nasabah dan polis yang dimiliki, semakin besar pula potensi kerugian yang harus ditanggung perusahaan (Fress *et al.*, 2016). Oleh karena itu, diperlukan perhitungan premi yang tepat agar dapat mengimbangi ekspektasi beban klaim yang mungkin muncul di masa mendatang. Salah satu pendekatan yang digunakan dalam menentukan premi adalah dengan memanfaatkan data historis klaim untuk membentuk *loss distribution*, yaitu distribusi statistik yang menggambarkan karakteristik besaran kerugian (*severity*) (Mustafa, 2025; Laudage *et al.*, 2019).

Dalam beberapa kasus, data klaim yang tersedia hanya mencakup nilai klaim (*severity*) tanpa disertai informasi frekuensi. Oleh karena itu, diperlukan pendekatan pemodelan yang sesuai dengan karakteristik data *severity*, yang biasanya bersifat positif dan berpola *skewness* kanan. Salah satu distribusi yang umum digunakan dalam hal ini adalah distribusi lognormal (Poudyal, 2021; Shi *et al.*, 2023). Studi ini memfokuskan pada estimasi premi murni berbasis *severity* menggunakan pendekatan regresi lognormal, yang telah terbukti efektif dalam berbagai studi aktuarial untuk mendukung pengambilan keputusan secara statistik dan aktuarial (Zhang *et al.*, 2023).

## METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif yang bertujuan untuk memodelkan premi asuransi berdasarkan data *severity* klaim menggunakan model distribusi lognormal. Studi ini dilakukan dengan menggunakan data sekunder berupa nilai klaim individual (*severity*) yang diperoleh dari salah satu perusahaan asuransi kerugian di Indonesia. Data yang digunakan mencakup nilai nominal klaim (dalam satuan rupiah) atas klaim-klaim yang diajukan selama periode tertentu, tanpa memuat informasi mengenai frekuensi klaim.

Dilakukan pemilihan model terbaik menggunakan indikator AIC dan BIC untuk melihat model terbaik. Estimasi parameter model dilakukan dengan metode *maximum likelihood estimation* (MLE). Validasi model dilakukan

melalui analisis residual dan pengujian signifikansi koefisien regresi. Selanjutnya premi murni dihitung berdasarkan ekspektasi nilai klaim dari model lognormal.

a. Indikator AIC dan BIC

Bentuk asli Kriteria Informasi Akaike (*Akaike Information Criterion / AIC*) (Akaike, 1998) adalah memilih sebuah model  $m \in \mathcal{M}$  untuk masalah estimasi densitas dengan meminimalkan:

$$AIC_m \triangleq -\frac{2}{n} \sum_{i=1}^n \log \hat{p}_{\theta_m}(y_i) + \frac{2d_m}{n} \quad (1)$$

$\theta_m$  menyatakan estimasi maksimum likelihood pada model  $m$ . Rumus di atas dapat diperluas untuk regresi dan deret waktu dengan mengganti  $\hat{p}_{\theta_m}$  menggunakan distribusi bersyarat  $y|x$  atau  $y_t|y_{t-1}, \dots, y_1, x_t$  yang diestimasi dari model. Dalam pengaturan yang lebih umum pada data nonparametrik dimana ukuran himpunan model  $\mathcal{M}$  meningkat seiring dengan  $n \rightarrow \infty$  AIC bersifat asimtotik efisien. AIC juga bersifat meminimalkan risiko maksimum terhadap skenario terburuk dari metode yang tersedia (Zhang *et al.*, 2023).

Dalam masalah estimasi densitas, Kriteria Informasi Bayesian (*Bayesian Information Criterion / BIC*) (Schwartz, 1978) memilih model dengan meminimalkan

$$BIC_m \triangleq -2 \sum_{i=1}^n \log \hat{p}_{\theta_m}(y_i) + d_m \log n \quad (2)$$

$\theta_m$  menyatakan estimasi maksimum likelihood pada model  $m$ . Sama dengan AIC, BIC juga dapat diterapkan untuk regresi linier dan deret waktu dengan asumsi distribusi terhadap galat  $\varepsilon$ . BIC memiliki interpretasi Bayesian yang menarik, yaitu kriteria pemilihan model dengan probabilitas posterior tertinggi

$$p(\mathcal{M}_m|D_n) = \frac{p \int_{\mathbb{R}^{d_m}} \exp\left(\sum_{i=1}^n \log(p_{\theta_m}(y_i))\right) p_m(\theta_m) d\theta_m}{\exp\left(\sum_{i=1}^n \log(p_{\hat{\theta}_m}(y_i))\right) p_m(\hat{\theta}_m) (2\pi)^{\frac{d_m}{2}} \det(B)^{-\frac{1}{2}}} \quad (3)$$

$p_m(\theta_m)$  menyatakan densitas prior dari parameter  $\theta_m$  pada model  $\mathcal{M}_m$ .  $B \triangleq \sum_{i=1}^n \nabla_{\theta}^2 \log p_{\hat{\theta}_m}(y_i)$  dan  $\det(\cdot)$  mendefinisikan determinand ari matriks. BIC konsisten jika digunakan pada model parametrik (Zhang *et al.*, 2023). AIC dan BIC mengevaluasi kehilangan informasi saat menggunakan model tertentu, artinya semakin kecil nilainya semakin sedikit informasi yang hilang atau dapat dikatakan semakin baik modelnya.

b. Model Lognormal

Sebagai bagian dari keluarga distribusi *log-location-scale*, distribusi lognormal memiliki berbagai aplikasi dalam ilmu aktuaria, bisnis, dan ekonomi (Pyoudal, 2021). Distribusi ini dapat mendekati jenis data kerugian (*severity*) aktuaria yang bersifat homogen (Hewitt *et al.*, 1979; Punzo *et al.*, 2018). Selain itu telah dibuktikan juga jika untuk data kerugian aktuaria yang bersifat

heterogen, distribusi lognormal mampu menggambarkan karakteristik data, baik pada bagian awal (*head*), akhir (*tail*), maupun keduanya dalam berbagai model komposit (Punzo *et al.*, 2018; Michael *et al.*, 2020)

Misalkan  $y$  adalah random variabel berdistribusi lognormal ( $LN(\theta, \sigma^2)$ ) dengan *probability density function* sebagai berikut

$$f(y|\theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)y}} e^{\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right)}, y > 0, \theta > 0, \sigma > 0 \quad (4)$$

Dimana  $\mu$  parameter lokasi dan  $\sigma$  parameter skala dari distribusi lognormal. Mean dan variansi (4) dihitung menggunakan metode *maximum likelihood estimator* (MLE) sebagai berikut

$$L(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{(2\pi\sigma^2)y_i}} e^{\left(-\frac{(\ln(y_i)-\mu)^2}{2\sigma^2}\right)}$$

*Log-likelihood* nya

$$l(\mu, \sigma) = -n \ln \sigma - n \ln \sqrt{2\pi} - \sum_{i=1}^n \ln y_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln y_i - \mu)^2 \quad (5)$$

Dengan menggunakan turunan parsial dari *log-likelihood* akan diperoleh estimasi parameter sebagai berikut

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln y_i, \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n (\ln y_i - \hat{\mu})^2 \quad (6)$$

(Habib *et al.*, 2025).

### c. Pemodelan Premi

Salah satu model yang paling umum digunakan untuk memodelkan *severity* klaim adalah model lognormal. Sebuah variabel dikatakan mengikuti distribusi lognormal apabila logaritma natural dari variabel tersebut berdistribusi normal. Model ini sangat bermanfaat dalam pemodelan nilai klaim, karena besaran klaim pada umumnya bersifat positif dan memiliki distribusi yang condong ke kanan (*skewed*). Nilai harapan dari distribusi lognormal disesuaikan dengan komponen variansi, yang merupakan faktor penting dalam penetapan tarif premi asuransi (Frees, 2010).

Regresi lognormal adalah model regresi yang digunakan ketika variabel respons  $Y$  adalah positif dan berdistribusi miring ke kanan (*right-skewed*), dan logaritma dari  $Y$  diasumsikan berdistribusi normal. Model *severity* lognormal menggunakan regresi

$$\log(Y_i) = x_i\beta + \varepsilon_i \quad (7)$$

Premi dapat diprediksi dengan menggunakan

$$E(Y_i) = \exp\left(\mu_i + \frac{1}{2}\sigma^2\right) \quad (8)$$

Dengan asumsi:

$Y_i > 0$  adalah besarnya klaim

$\varepsilon_i \sim N(0, \sigma^2)$  (normal standard error pada log scale)

Maka  $Y_i \sim LN(\mu_i, \sigma^2)$  dimana  $\mu_i = x_i\beta$

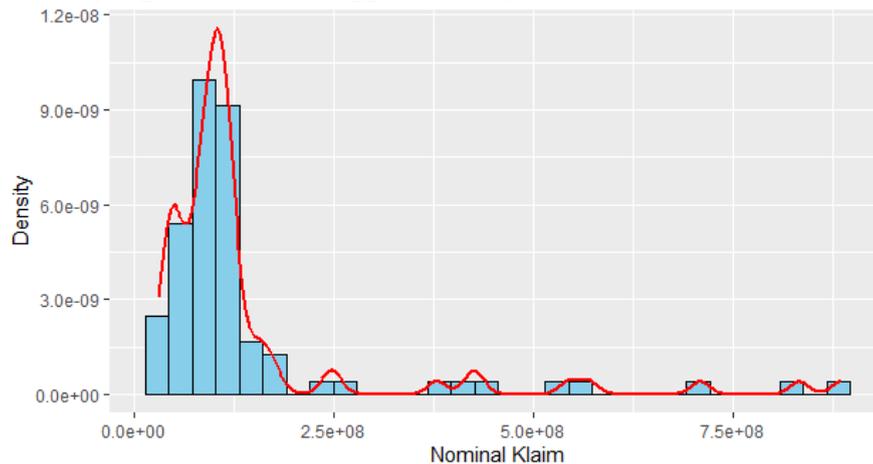
## HASIL DAN PEMBAHASAN

### a. Eksplorasi Data

Pada penelitian ini, akan digunakan data historis klaim asuransi dari PT Asuransi Jasa Indonesia (Jasindo) periode tahun 2009 sampai tahun 2015. Data terdiri atas variabel waktu (bulanan) dan nominal klaim (satuan rupiah). Untuk keperluan analisis, variabel waktu dikonversi ke dalam indeks waktu agar dapat digunakan sebagai prediktor linier.

### b. Pemilihan Model Terbaik

Distribusi dari data nominal klaim dilihat melalui kurva densitas. Berdasarkan visualisasi awal, sebaran data bersifat asimetris dengan condong ke kanan (*right-skewed*). Hal ini dapat dilihat dari nilai *skewness* positif (*skewness* = 3.033283). Distribusi ini sering ditemukan dalam data *severity* klaim asuransi, yang bernilai positif dan memiliki ekor kanan yang panjang. Berikut bentuk sebaran data yang diperoleh menggunakan *software* Rstudio.



**Gambar 1. Sebaran Data Nominal Klaim**

Untuk mengidentifikasi model terbaik dalam memodelkan data nominal klaim, dilakukan pendekatan dengan tiga jenis model. Yaitu metode GLM dengan distribusi Gamma, model GLM dengan distribusi Inverse Gaussian, dan model regresi linier dengan transformasi lognormal (model lognormal). Kriteria evaluasi yang digunakan adalah AIC dan BIC. Metode AIC (*Akaike Information Criterion*) dan BIC (*Bayesian Information Criterion*) merupakan metode populer dalam statistika untuk membandingkan kualitas relatif dari beberapa model. Keduanya mempertimbangkan kebaikan model terhadap data dan kompleksitas model. Berikut hasil yang diperoleh

**Tabel 1. Pemilihan Model dengan AIC dan BIC**

Model	AIC	BIC
Gamma	3160.5549	3167.7750
Inverse Gaussian	3136.4969	3136.4969
Lognormal	121.5924	128.8125

Dari Tabel 1, diperoleh model Lognormal memiliki nilai AIC dan BIC

paling rendah di antara dua model lainnya. Hal ini berarti informasi yang diberikan oleh model lognormal akan lebih baik dari pada model lainnya. Model lognormal dipilih untuk digunakan dalam peramalan premi murni karena kemampuannya dalam menangani distribusi *skew* dan menghasilkan estimasi rata-rata klaim yang stabil.

c. Perhitungan Premi Murni

Berdasarkan analisis menggunakan *software Rstudio* dan Model Lognormal (7), diperoleh hasil modelnya adalah

$$\log(klaim) = 17.65114 + 0.02036x_i + \varepsilon_i$$

Model tersebut kemudian digunakan untuk memprediksi premi murni selama 12 bulan ke depan. Prediksi dilakukan dengan mengekstrapolasi indeks waktu ke periode berikutnya dan menghitung estimasi median dan mean dari distribusi lognormal. Estimasi mean dilakukan menggunakan (8) dimana  $\mu$  adalah hasil prediksi log dari model regresi dan  $\sigma^2$  adalah variansi residual dari model. Berikut merupakan hasil prediksi premi murni selama 12 bulan kedepan.

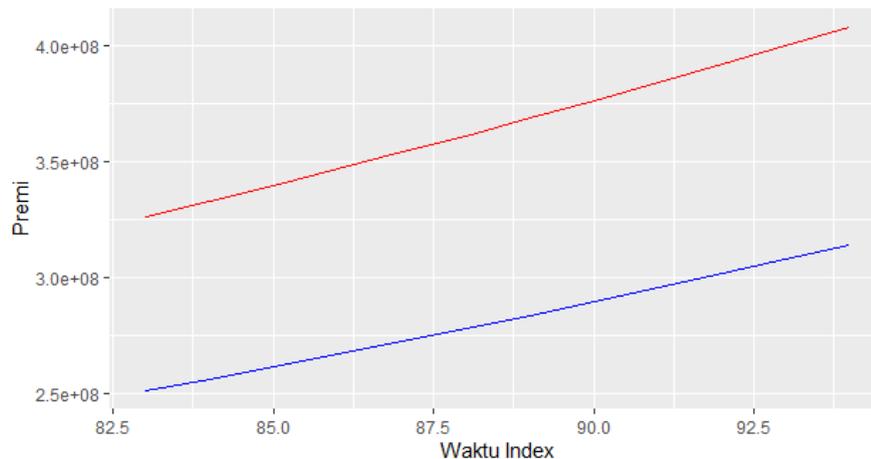
**Tabel 2. Prediksi Premi Murni**

Waktu	Premi Murni	Premi Komersil
1	251009408,3	326312230,8
2	256172281,3	333023965,6
3	261441346,5	339873750,4
4	266818788,2	346864424,7
5	272306835,5	353998886,2
6	277907763,4	361280092,5
7	283623893,7	368711061,8
8	289457595,9	376294874,7
9	295411288,3	384034674,7
10	301487438,8	391933670,5
11	307688566,3	399995136,2
12	314017241,4	408222413,8

Premi komersial dihitung dengan menambahkan faktor *loading* sebesar 30% terhadap premi murni. Faktor *loading* ini memperhitungkan biaya-biaya lainnya yang harus dikeluarkan oleh Perusahaan dalam mengelola produk asuransinya.

d. Visualisasi dan Interpretasi

Agar lebih mudah melihat pertumbuhan premi murni dan komersial dari waktu ke waktu, dilakukan visualisasi tren hasil prediksi. Garis premi murni (biru) dan komersil (merah) digambarkan untuk menunjukkan perbedaan akibat koreksi lognormal dan penambahan loading.



**Gambar 2. Prediksi Premi Asuransi Model Lognormal**

Gambar 2 menunjukkan tren kenaikan baik pada prediksi premi murni maupun premi komersial. Premi komersial yang dihitung dengan menambahkan margin atau faktor *loading* konsisten berada di atas garis premi murni dan memperlihatkan bentuk pertumbuhan yang paralel. Pola ini menunjukkan bahwa model lognormal yang digunakan tidak hanya mampu menangkap pola historis klaim, tapi juga memberikan estimasi premi ke depan yang proporsional dan sesuai ekspektasi bisnis.

## KESIMPULAN

Penelitian ini menunjukkan bahwa model regresi lognormal merupakan pendekatan yang efektif dalam memodelkan nominal klaim asuransi, khususnya ketika data yang tersedia hanya mencakup informasi mengenai *severity* klaim tanpa frekuensi. Berdasarkan evaluasi menggunakan kriteria AIC dan BIC, model lognormal terbukti memberikan informasi lebih baik dibandingkan dengan model GLM Gamma dan Inverse Gaussian. Model ini mampu menangkap karakteristik distribusi data yang bersifat positif dan *right-skewed*, serta menghasilkan estimasi premi murni yang akurat dan stabil. Prediksi premi murni selama 12 bulan ke depan menunjukkan tren peningkatan yang realistis dan dapat diandalkan sebagai dasar perhitungan premi komersial setelah penambahan *loading*. Hasil ini memberikan bukti empiris bahwa regresi lognormal dapat digunakan sebagai alat bantu aktuaria yang valid dalam perencanaan tarif premi berbasis data historis. Oleh karena itu, penerapan model ini sangat relevan bagi perusahaan asuransi dalam mendukung pengambilan keputusan tarif yang lebih objektif dan berbasis data.

## SARAN

Berdasarkan hasil penelitian, disarankan agar perusahaan asuransi mempertimbangkan penggunaan model regresi lognormal dalam proses penentuan premi, terutama ketika data yang tersedia hanya data *severity* (nilai klaim) tanpa frekuensi. Model ini terbukti mampu memberikan estimasi premi murni yang stabil, serta mencerminkan karakteristik data klaim asuransi yang cenderung berpola *right-skewed*. Untuk implementasi operasional, perusahaan dapat mengintegrasikan pendekatan ini ke dalam sistem *pricing* premi dengan penyesuaian margin (*loading*) yang sesuai berdasarkan strategi bisnis dan profil



risiko masing-masing. Selain itu, disarankan agar perusahaan melakukan evaluasi berkala terhadap model prediksi yang digunakan, dengan memperbarui data secara berkala guna mempertahankan akurasi estimasi dalam menghadapi dinamika risiko yang terus berubah. Penelitian lanjutan juga direkomendasikan untuk memperluas model dengan memasukkan variabel tambahan, seperti karakteristik tertanggung atau jenis polis, serta membandingkan performa dengan model lain seperti distribusi Tweedie atau pendekatan Bayesian.

## DAFTAR PUSTAKA

- Booth, P., Robert, C., Steven, H., Dewi, J., Zaki, K., Robert, P., & Ben, R. (2020). *Modern Actuarial Theory and Practice (2<sup>nd</sup> Edition)*. New York: Chapman and Hall/CRC.
- Frees, E. W. (2010) *Regression Modeling with Actuarial and Financial Applications*. United Kingdom: Cambridge University Press.
- Frees, E. W., Derrig, R. A., & Meyers, G. G. (2016). *Predictive modeling applications in actuarial science, volume II: Case studies in insurance*. Canada: Cambridge University Press.
- Habib, M., Muhammad, A., & Sadiyah, M.A.A. Influence Analysis in the Lognormal Regression Model with Fitted and Quantile Residuals. *Axioms*, 14, 464.
- Hewitt, C. C., Jr., & Lefkowitz, B. (1979). Methods for fitting distributions to insurance loss data. *In Proceedings of the Casualty Actuarial Society*, 66, 139–160. Casualty Actuarial Society.
- Laudage, C. S. D., & Wasmund, J. (2019). Severity modeling of extreme insurance claim for tariffication. *Insurance: Mathematics and Economics*, 88, 77–92.
- Michael, S., Tatjana, M., & Volodymyr, M. (2020). Mixture Modeling of Data with Multiple Partial Right-Censoring Levels. *Advances in Data Analysis and Classification*, 14, 355-378.
- Mustafa, Z. (2025). Tweedie distribution: A statistical solution for unusually dispersed data. *Sciencestatistics: Journal of Statistics, Probability, and Its Application*, 3 (1), 29–37.
- Poudyal, C. (2021). Robust estimation of loss models for lognormal insurance payment severity data. *arXiv*.
- Punzo, A., Luca, B., & Antenello, M. (2018). Compound unimodal distribution for insurance losses. *Insurance: Mathematics and Economics*, 81, 95-107.
- Shi, Y., Tan, K. S., & Wüthrich, M. V. (2023). Claims modelling with three-component composite models. *Risks*, 11 (11), 196.
- Su, X., & Bai, M. (2020). Stochastic gradient boosting frequency–severity model of insurance claims. *PLOS ONE*, 15(8), e0238000.
- Zhang, J., Yuhong, Y., & Jie, D. (2023). Information criteria for model selection. *Wiley Interdisciplinary Reviews: Computational Statistics*, 15(5).

