Available online at http://www.iorajournal.org/index.php/ijgor/index

International Journal of Global Operations Research

e-ISSN: 2722-1016
p-ISSN: 2723-1739

Vol. 6, No. 4, pp. 299-305, 2025

Sensitivity of Premium Calculations to Distribution Assumptions: The Impact of Using Pareto vs. Pareto Distributions Exponential on Claims of Large Losses

Aisha Nurul Qolbi Hadiat[1*], Niken Pinkyvita Nur Anjani[2]

[1, 2]*Departemen Matematika, Fakultas MIPA, Universitas Padjadjaran, Bandung*

Email: aisha22003@mail.unpad.ac.id, niken22002@mail.unpad.ac.id

Abstract

Premium calculation is a fundamental aspect in the insurance industry to ensure the sustainability and profitability of the company. One of the main factors that affect the accuracy of premium calculation is the selection of the probability distribution of claims, especially on large value claims that have the potential to cause significant losses. The exponential distribution is often used due to its simplicity, however the characteristics of the light tail make it less suitable for modeling extreme claims. In contrast, the Pareto distribution with the property of the heavy tail is considered more representative in capturing the risk of large claims. This study aims to analyze the sensitivity of premium calculations to distribution assumptions by comparing the use of exponential and Pareto distributions on large loss claims data. This analysis is expected to provide a practical overview of the impact of distribution selection on premium determination and its implications in insurance risk management in Indonesia.

Keywords: insurance premiums, Pareto distribution, exponential distribution, large claims, model sensitivity, risk management

1. Introduction

The calculation of premiums in the insurance industry is an important aspect that determines the sustainability and profitability of insurance companies. The premium must accurately reflect the risks to be borne so that the company can cover the claims that occur while obtaining reasonable profits (Alfaridzi & Prabowo, 2023). One of the main factors in the calculation of premiums is the selection of the probability distribution of loss claims, especially claims with large value that have the potential for significant financial impacts.

In the context of insurance in Indonesia, Pareto and Exponential distributions are often used to model large loss claims. The Pareto distribution is known as the *heavy tail* characteristic, which allows it to capture the occurrence of large claims that are rare but have a significant impact, such as claims due to natural disasters and major accidents. In contrast, the Exponential distribution has a simpler form and is often used for claims with high frequency and relatively small loss values, assuming a *memoryless nature* that simplifies the calculation of premiums (Alfaridzi & Prabowo, 2023; Diba et al., 2017).

Choosing the right distribution is very important because it can significantly affect premium estimates. If the distribution used is not in accordance with the characteristics of the claim data, then the premium calculation can be biased, both in the form of underestimation and overestimation of risk (Putra et al., 2021). This can cause insurance companies to suffer financial losses or lose competitiveness in the market (Alfaridzi & Prabowo, 2024). Therefore, sensitivity analysis to distribution assumptions is an important step to understand how much impact distribution selection has on the results of premium calculations.

This study aims to analyze the sensitivity of premium calculations to distribution assumptions by comparing the use of Pareto and Exponential distributions on large loss claims. Through this approach, it is hoped that a clearer picture can be obtained of the differences in premium results generated by the two distributions and their practical implications in insurance risk management in Indonesia. The results of this study are expected to contribute to actuaries and risk managers in choosing a more appropriate and accurate distribution model, so as to increase the effectiveness of risk management and financial stability of insurance companies.

2. Literature Review

2.1 Calculation of Premiums in Insurance

Premiums are a major component in the insurance industry that serves as compensation for the risk transferred from the insured to the insurer. In actuarial theory, premiums are calculated based on the expected value of the claim with the addition of loading factors as profit margins and risk anticipation (Bowers et al., 1997). Errors in the estimation of claims distribution can lead to bias in the calculation of premiums that have a direct impact on the profitability and stability of the insurance company (Putra, Lesmana, & Purnaba, 2021).

2.2 Probability Distribution for Claims Data

Insurance claim modeling relies heavily on the chosen probability distribution. Exponential distributions are widely used due to their simplicity and *memoryless nature,* which makes it easy to calculate premiums on claims with high frequency but small values (Diba, Saepudin, & Rohmawati, 2017). However, this distribution tends to be inappropriate for claims with large value because they have a *light-tailed*. In contrast, the Pareto distribution is known for its *heavy-tail* characteristics that are able to capture the possibility of large claims that are rare but have significant impacts (Embrechts, Klüppelberg, & Mikosch, 1997; Cooray & Ananda, 2005).

2.3 Pareto and Exponential Comparison in Claims Modeling

Several previous studies have shown the importance of choosing a distribution that matches the characteristics of the claim data. Alfaridzi & Prabowo (2023) use exponential and gamma distributions in modeling vehicle insurance premiums, while other studies show that the use of Pareto is more appropriate in modeling large claims, especially in cases of losses due to disasters or extreme events (Resnick, 2007; Mao, Hu, & Hu, 2015). The study of Cooray & Ananda (2005) even proposed a lognormal–Pareto composite model to combine the behavior of small to large claims, demonstrating Pareto's flexibility in modeling distribution tails.

2.4 The Impact of Distribution Selection on Premiums

The selection of claim distribution directly affects the amount of premiums generated. Exponential tends to result in lower premium estimates for large claims, so there is a risk of causing *underpricing* and financial losses if extreme claims occur (Alfaridzi & Prabowo, 2024). In contrast, the Pareto distribution provides a higher premium estimate because it takes into account tail risk. This is in line with extreme risk theory which states that heavy-tailed models are more realistic in predicting large claims (Chavez-Demoulin, Embrechts, & Hofert, 2016).

2.5 State of Art

This research is rooted in actuarial studies that focus on large modeling of insurance claims. A review of the literature shows diversity in the selection of probability distributions. The selected distribution has a significant impact on premium estimates and risk management outcomes. Table 1 presents a comparison of the focuses, methodologies, and key findings of previous studies that are relevant, especially related to the type of distribution used:

Table 1: Comparison of Previous Journals in Insurance Claims Modeling

Author (Year)	Research Focus	Distribution of Claims Used	Key Findings/Contributions
Embrechts et al. (1997)	Extreme event modeling for insurance and finance	Heavy Tail *Distribution* (including Pareto)	It establishes the theoretical basis that heavy-tailed models are capable of capturing the possibility of large claims that are rare but have a significant impact.
Cooray & Ananda (2005)	Actuarial data modeling	Model composite Lognormal–Pareto	Propose a composite model to combine the behavior of small to large claims, demonstrating Pareto's flexibility in

			modeling the tail of the distribution
Resnick (2007)	Extreme Risk Theory and the Heavy Tail Model	Pareto (or other Heavy Tail Distribution)	Theoretically, it shows that heavy-tailed models are more realistic in predicting large claims.
Diba et al. (2017)	Insurance company bankruptcy opportunity modeling	Exponential	It uses Exponential distribution due to its simplicity and *memoryless nature.*
Putra et al. (2021)	Calculation of motor vehicle insurance premiums	*Generalized Linear Models* with Tweedie distributions	Leverage more complex GLM models with distributions in exponential families
Alfaridzi & Prabowo (2023)	Determination of vehicle insurance premium rates	Exponential and Gamma	Provides a premium model for vehicle insurance, but focuses on light to medium tail distribution

2.6 Research Gap

Based on the above comparison, most of the research in Indonesia (Diba et al., 2017; Putra et al., 2021; Alfaridzi & Prabowo, 2023) still tend to focus on simpler distributions such as Exponential, Gamma, or Tweedie. Light-tail models tend to result in lower premium estimates for large claims (*underpricing*). Although heavy-tailed models such as Pareto are theoretically recognized (Embrechts et al., 1997; Resnick, 2007) is more suitable for extreme risk, a study that explicitly compares and analyzes the impact of the use of light-tailed (Exponential) and heavy-tailed (Pareto) distributions on the sensitivity of premium calculations on large loss claims in Indonesia is still limited.

Therefore, this study offers a contribution by explicitly testing the sensitivity of premiums to the two different distribution assumptions on large loss claims data, so as to provide a practical picture of the risk of bias in the determination of insurance premiums in Indonesia.

3. Materials and Methods

3.1 Materials

The research data includes data on large claims (60th percentile) of accident insurance in 2023. The analysis was carried out using R software with actuar, fitdistrplus, and VaR packages, as well as Excel for the pre-processing stage.

Table 2: Distribution parameters and characteristics

Distribution	Parameter	Density Function	Formula Mean	Characteristic
Pareto	α (shape), (scale)θ	$\frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}$	$\frac{\alpha\theta}{\alpha-1}$	Heavy
Exponential	λ (rate)	$\lambda e^{-\lambda x}$	$1/\lambda$	Light

3.2 Methods

The steps of the research are as follows:

1) Data pre-processing: Data validation, inflation adjustment, and large claims selection (> P60).
2) Parameter estimation: Using the *Maximum Likelihood Estimation* (MLE) method.
3) *Goodness of Fit*: Uji Kolmogorov-Smirnov (KS), Anderson-Darling (AD), AIC, dan BIC.
4) Premium calculation: Calculates pure and total premiums.
5) Ukuran risiko: *Value at Risk* (VaR) dan *Expected Shortfall* (ES).
6) Sensitivity analysis: Calculates absolute, relative, and ratio differences between models.
7) Model validation: Using MAE, RMSE, and solvency tests.

3.2.1 Distribution Model

Pareto:

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \quad x \geq \theta > 0, \quad \alpha > 0 \tag{1}$$

Exponential:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0 \tag{2}$$

3.2.2 Parameter Estimation (MLE)

Pareto:

$$\hat{\alpha} = \frac{n}{\sum \ln\left(\frac{x_i}{\theta}\right)}, \quad \hat{\theta} = \min(x_i) \tag{3}$$

Exponential:

$$\hat{\lambda} = \frac{n}{\sum x_i} \tag{4}$$

3.2.3 Goodness of Fit

Table 3: Goodness of Fit Testing Criteria

Test Method	Formula	Admission Criteria
Kolmogorov-Smirnov	$D = \max \lvert F_n(x) - F(x) \rvert$	$p - value > 0.05$
Anderson-Darling	$A^2 = -n - \sum \frac{2i-1}{n}[lnF(X_i) + \ln(1 - F(X_{n+1-i}))]$	$p - value > 0.05$
AIC	$2k - 2\ln(L)$	Lower is better
BIC	$kln(n) - 2\ln(L)$	Lower Better

3.2.4 Award Calculation

Pareto:

$$E[X] = \frac{\alpha\theta}{\alpha-1} \tag{5}$$

Exponential:

$$E[X] = 1/\lambda \tag{6}$$

$$Premi\ Total = Premi\ Murni\ \times (1 + 20\%) \tag{7}$$

3.2.5 Risk Size

Table 4: Risk Measurement Formula at Confidence Level p

Distribution	Var	Ice
Pareto	$\theta((1-p)^{-\frac{1}{\alpha}} - 1)$	$\frac{VaR_p + \theta}{1 - 1/\alpha}$
Exponential	$-\ln(1-p)/\lambda$	$VaR_p + \frac{1}{\lambda}$

3.2.6 Sensitivity Analysis

$$\lvert Premi_{Pareto} - Premi_{Exp} \rvert \tag{8}$$

$$\frac{Premi_{Pareto} - Premi_{Exp}}{Premi_{Exp}} \times 100\% \tag{9}$$

$$\frac{Premi_{Pareto}}{Premi_{Exp}} \tag{10}$$

3.2.7 Model Validation

Validation was carried out by out-of-sample split (70/30), rolling window, and evaluation using MAE, RMSE, and solvency ratio.

4. Results and Discussion

This analysis uses insurance claim data from 34 observations with the main variable Claim Amount (in million rupiah). Based on the results of preprocessing and the determination of the threshold at the 60th percentile, there are 14 large claim data used in the distribution estimation.

4.1 Parameter Estimation

Parameter estimation is carried out using the Maximum Likelihood Estimation (MLE) method. The estimated results are shown in Table 5.

Table 5: Parameter Estimation Results

Distribution	Parameter	Value
Pareto	α	0.8139
Pareto	θ	45.313
Exponential	λ	0.000004

The values show that the Pareto distribution has a very heavy tail, which indicates the potential for a very high but rare large claim. Conversely, in exponential distributions it is very small, indicating a large average claim.$\alpha < 1\lambda$

4.2 Model Fit Test (Goodness of Fit)

Model fit tests were conducted with the Kolmogorov–Smirnov (KS), Akaike In formation Criterion (AIC), and Bayesian Information Criterion (BIC). The test results are presented in Table 6.

Table 6: Model Fit Test

Method	Pareto	Exponential
KS p-value	0.0070	0.6777
AIC	372.37	375.14
BIC	373.65	375.78

The Pareto model has lower AIC and BIC, so it is statistically better at capturing large claim distributions, although the KS test shows $p - value < 0.05$. This is natural because large claims data tend to have high deviations and a limited number of observations.

4.3 Award Calculation

Premium calculations are carried out for two distribution models with an additional *loading factor* of 20%. The premium value is shown in Table 7.

Table 7: Comparison of premiums by distribution

Types of premiums	Pareto	Exponential
Murni E[X] Prize	-198,124.82	225,563.41
Total Prize (+20%)	-237,749.78	270,676.09

The value of a negative Pareto premium is caused by $\alpha < 1$, which causes mathematical expectations to be not positively defined. This indicates a very high extreme risk, where average losses are difficult to estimate stably.4.4 Value *at Risk* and *Expected Shortfall*

Risk size analysis was conducted at a 95% confidence level. The results of the calculation of the served in Table 8.

Tabel 8: *Value at Risk* (VaR) dan *Expected Shortfall* (ES) pada $p = 0.95$

Risk Size	Pareto	Exponential
$VaR_{0.95}$	1,752,773.13	675,727.59
$ES_{0.95}$	-7,861,883.16	901,291.01

The Pareto distribution shows much greater VaR and ES values, suggesting a high sensitivity to extreme claims. Meanwhile, exponential gives more conservative results due to its lightweight tail nature.

4.5 Sensitivity Analysis

The sensitivity of premiums to changes in distribution can be seen in Table 9.

Table 9: Total Premium Sensitivity Analysis

Indicator	Value	Information
Absolute Difference	508,425.90 million rupiah	Difference Award
Relative Difference	-187.84%	Pareto is lower than Exponential
Premium Ratio	-0.88	Pareto 0.88 times Exponential premium

The difference in premium yield suggests that the distribution assumptions greatly affect the final estimate. The Pareto distribution results in a smaller (mathematically negative) premium due *to the extreme heavy tail* effect, while exponential provides a more stable premium value.

4.6 Interpretation and Implications

These results show that for large claims data with a well-tailed distribution, the Pareto model is more appropriately used in extreme risk analysis. However, for the purpose of setting a stable and realistic premium, exponential distribution can be a conservative initial approach. Combinations or composite models (e.g. normal–Pareto logs) can be considered in future studies to capture the entire range of claims more accurately.

5. Conclusion

This study shows that the selection of probability distribution has a significant influence on the results of premium calculation and risk measurement on the claim data. Based on the results of parameter estimation and model fit testing, the Pareto distribution has lower AIC and BIC values compared to the unfortunate exponents, so it is better at describing the characteristics of large claims that are extreme. However, the value of the ¡1 parameter in the Pareto model causes the mathematical expectations to be not positively defined, which indicates the existence of extreme risk with the instability of the average loss

In contrast, the Exponential model provides more stable and conservative premium yields, although it tends to ignore extreme risks due to its lightweight tail nature. Sensitivity analysis shows a significant difference between the premiums calculated using the two distributions, where the Pareto premium tends to be lower (even mathematically negative) than the exponential, indicating a high sensitivity to the distribution assumptions.

Practically, these results confirm that the selection of distribution models must be adjusted to the characteristics of claim data. For large claims data, heavy-tailed distributions such as Pareto are more representative in capturing extreme risk, while exponential distributions are more suitable for high-frequency claims with small values. Therefore, a composite model approach (such as lognormal–Pareto) or hy brid modeling is recommended in future studies to produce more realistic, stable, and accurate premium estimates in the context of insurance risk management in Indonesia.

References

Alfaridzi, S. I. R., & Prabowo, A. (2023). Determination of premium rates for vehicle insurance with claim size following exponential and gamma distributions. Premium Insurance Business Journal, 10(1), 29–41. http://ejournal.stma-trisakti.ac.id/index.php/premium/article/view/46

Alfaridzi, S. I. R., & Prabowo, A. (2023). Penentuan tarif premi pada asuransi kendaraan dengan besar klaim berdistribusi eksponensial dan gamma. Premium Insurance Business Journal, 10(1), 29–41. http://ejournal.stma-trisakti.ac.id/index.php/premium/article/view/46

Alfaridzi, S. I. R., & Prabowo, A. (2024). Calculating optimal traffic accident insurance premiums based on claim size and claim frequency. Lamda: Scientific Journal of MIPA Education and its Application. https://ejournal.baleliterasi.org/index.php/lambda/article/view/843

Alfaridzi, S. I. R., & Prabowo, A. (2024). Menghitung premi asuransi kecelakaan lalu lintas yang optimal berdasarkan besar klaim dan frekuensi klaim. Lamda: Jurnal Ilmiah Pendidikan MIPA dan Aplikasinya. https://ejournal.baleliterasi.org/index.php/lambda/article/view/843

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., & Nesbitt, C. J. (1997). Actuarial Mathematics (2nd ed.). Society of Actuaries.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., & Nesbitt, C. J. (1997). Actuarial Mathematics (2nd ed.). Society of Actuaries.

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance, 83(3), 735–776. https://doi.org/10.1111/jori.12070

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. Journal of Risk and Insurance, 83(3), 735–776. https://doi.org/10.1111/jori.12070

Cooray, K., & Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal–Pareto model. Scandinavian Actuarial Journal, 2005(5), 321–334. https://doi.org/10.1080/03461230510009719

Cooray, K., & Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal–Pareto model. Scandinavian Actuarial Journal, 2005(5), 321–334. https://doi.org/10.1080/03461230510009719

Diba, F., Saepudin, D., & Rohmawati, A. A. (2017). Modeling and simulation of the probability of bankruptcy of insurance companies with analysis of premium value and claim size assumed to be exponentially distributed. Indonesian Journal on Computing (Indo-JC), 2(2), 1–10. https://socjs.telkomuniversity.ac.id/ojs/index.php/indojc/article/view/147

Diba, F., Saepudin, D., & Rohmawati, A. A. (2017). Pemodelan dan simulasi peluang kebangkrutan perusahaan asuransi dengan analisis nilai premi dan ukuran klaim diasumsikan berdistribusi eksponensial. Indonesian Journal on Computing (Indo-JC), 2(2), 1–10. https://socjs.telkomuniversity.ac.id/ojs/index.php/indojc/article/view/147

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer.

Mao, T., Hu, T., & Hu, Z. (2015). On risk measures for heavy-tailed losses under insurance and finance. ASTIN Bulletin, 45(2), 425–452. https://doi.org/10.1017/asb.2014.30

Mao, T., Hu, T., & Hu, Z. (2015). On risk measures for heavy-tailed losses under insurance and finance. ASTIN Bulletin, 45(2), 425–452. https://doi.org/10.1017/asb.2014.30

Note: The APA format is preserved. Indonesian titles and journal names are translated for context.

Putra, T. A. J., Lesmana, D. C., & Purnaba, I. G. P. (2021). Calculation of motor vehicle insurance premiums using generalized linear models with a Tweedie distribution. Jambura Journal of Mathematics. https://ejurnal.ung.ac.id/index.php/jjom/article/view/10136

Putra, T. A. J., Lesmana, D. C., & Purnaba, I. G. P. (2021). Penghitungan premi asuransi kendaraan bermotor menggunakan generalized linear models dengan distribusi Tweedie. Jambura Journal of Mathematics. https://ejurnal.ung.ac.id/index.php/jjom/article/view/10136

Resnick, S. I. (2007). Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer.