Determination Of Insurance Premiums Using The Optimal Bonus-Malus System With The Bayesian Method

Fatichatu Zahro[1]
[1]*Mathematics Study Program, Faculty of Science and Technology*
UIN Sunan Kalijaga Yogyakarta
Sri Istiyarti Uswatun Chasanah[2*]
[2]*Mathematics Study Program, Faculty of Science and Technology*
UIN Sunan Kalijaga Yogyakarta
Syarif Abdullah[3]
[3]*Statistics Study Program, Faculty of Technology*
Universitas Sultan Ageng Tirtayasa
*Corresponding author: *sri.chasanah@uin-suka.ac.id*

Abstract

This study develops an optimal bonus–malus premium model within a Bayesian decision-theoretic framework. Claim frequency is modeled using a Poisson distribution, while the number of claims exceeding a predefined critical threshold is modeled conditionally using a Binomial distribution. The Poisson intensity parameter is assigned an Exponential prior distribution, while the Binomial probability parameter follows a Beta prior distribution. The Exponential and Beta distributions are applied to model parameters rather than to observed data, ensuring probabilistic consistency. Since the Exponential distribution is a special case of the Gamma distribution, the Bayesian updating process remains mathematically coherent. To illustrate the model, a dataset of 1,000 simulated motor vehicle policyholders is generated in R under specified distributional assumptions. The results indicate that premiums increase with the number of claims exceeding the critical value and decrease with longer claim-free duration. The proposed framework provides a coherent and flexible approach for premium determination in bonus–malus systems. However, the findings are based on simulated data and specific modeling assumptions, which may limit direct empirical generalization.

Keywords: Bayesian Method; Bonus–Malus System; Poisson Distribution; Exponential Prior; Beta Prior.

INTRODUCTION

The rapid growth of the human population in Indonesia has been accompanied by a substantial increase in the number of motor vehicles. The dominance of private vehicles over public transportation has led to higher exposure to traffic-related risks, particularly road accidents that may result in financial losses for drivers and third parties. To mitigate these financial consequences, motor vehicle insurance plays a crucial role in providing financial risk protection (Adisti & Mutaqin, 2021).

Insurance functions as a financial protection mechanism in which the insurer compensates losses in exchange for premium payments. To ensure fairness and efficiency in premium determination, insurers implement various premium-adjustment schemes, one of which is the bonus–malus system. According to (Noviyanti et al., 2017) , the bonus–malus system adjusts premiums based on the policyholder's past claim history. Policyholders who do not report claims receive premium reductions (bonus), whereas those who submit claims are subject to premium increases (malus). This mechanism promotes safer behavior by providing financial incentives for maintaining claim-free periods.

In practice, claim experience differs not only in frequency but also in severity. Large claims typically contribute disproportionately to total loss exposure compared to small claims. Therefore, distinguishing claims based on a critical threshold becomes important in improving premium fairness and accurately reflecting risk heterogeneity among policyholders.

Several studies have applied Bayesian approaches to bonus–malus systems, ie(Hutasoit, 2019) (Moumeesri & Pongsart, 2022) (Deniz, 2016)(Lemaire, 1995). Bayesian methods incorporate prior information about risk parameters and update these beliefs using observed data. For example, (Hernawati et al., 2017) and (Deniz, 2016) proposed a Bayesian premium model in which claim frequency is modeled using a Poisson distribution, and claims are classified according to a critical value. Although this framework provides a structured probabilistic foundation, most existing models primarily emphasize claim frequency adjustment and do not explicitly examine the combined role of claim severity classification and policy duration in premium differentiation.

Motivated by these limitations, this study develops a Bayesian premium model (Hogg et al., 2019) for determining optimal premiums in a motor vehicle bonus–malus system. The model jointly considers claim frequency and the number of claims exceeding a predefined critical threshold, while also incorporating policy duration into the premium updating mechanism. The contribution of this study lies in integrating critical-value-based severity classification and policy duration within a coherent Bayesian premium optimization framework.

RESEARCH METHODS

Simulated Data

To illustrate the proposed Bayesian bonus–malus premium model, a simulated dataset consisting of $1,000$ motor vehicle policyholders is generated using R software. Simulation is employed to evaluate the structural properties of the premium model under controlled assumptions

Claim Frequency
Let X denote the number of claims submitted by a policyholder during one insurance period. Claim frequency is assumed to follow a Poisson distribution with intensity parameter $\lambda > 0$:

$$X|\lambda \sim Poisson(\lambda) \tag{1}$$

(Walpole & Myers, 1995) In the simulation, the parameter is set to:

$$\lambda = 0.4 \tag{2}$$

This represents a relatively low-frequency motor vehicle insurance portfolio.

Claim Size
Let $Y_1, Y_2, ..., Y_i$ denote individual claim sizes.
Each claim size is generated independently from a Weibull distribution:

$$Y_i \sim Weibull(\alpha_\gamma, \beta_\gamma) \tag{3}$$

Where

$\alpha_\gamma > 0$ is the shape parameter

$\beta_\gamma > 0$ is the scale parameter

In the simulation:

$$\alpha_\gamma = 3, \beta_\gamma = 50 \tag{4}$$

The Weibull distribution is used solely for data generation and does not directly enter the Bayesian updating structure.

Bayesian Bonus–Malus Model

Likelihood Specification

To incorporate claim size information into the premium determination process, a predefined critical value ψ is introduced.

For a policyholder with $X = x$ claims, define:

$$Z = \sum_{i=1}^{x} I(Y_i > \Psi) \tag{5}$$

Where I (•) denotes the indicator function. Thus:

- X = total number of claims

- Z = number of claims exceeding the critical value ψ

Conditional on $X = x$, the number of large claims follows a Binomial distribution:

$$Z|X = x, p \sim Binomial(x, p) \tag{6}$$

where:

$$p = P(Y > \psi) \tag{7}$$

Given observed data (x, z), the joint likelihood function of parameters λ and p is:

$$L(\lambda, p|x, z) = \frac{\lambda^x e^{-\lambda}}{x!} xzp^z (1 - p)^{x-z} \tag{8}$$

for $x = 0, 1, 2, ...$, and $z = 0, 1, 2, ..., x$ (Tse, 2009) (Hines & Montgomery, 1990)

Prior Distributions

Consistent with the model specification in the abstract and following (Hernawati et al., 2017), prior distributions are assigned to the model parameters.

Prior for Claim Intensity

The Poisson intensity parameter λ is assigned an Exponential prior with hyperparameter $\alpha > 0$:

$$\lambda \sim Exponential(\alpha), \tag{9}$$

with density:

$$\pi(\lambda) = \alpha e^{-\lambda \alpha}; \lambda > 0 \tag{10}$$

The Exponential distribution is a special case of the Gamma distribution with shape parameter 1, ensuring conjugacy with the Poisson likelihood.

Prior for Large-Claim Probability

The Binomial probability parameter p is assigned a Beta prior with hyperparameters $a > 0$ and $b > 0$:

$$p \sim Beta(a, b), \tag{11}$$

with density:

$$\pi(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, 0 < p < 1 \tag{12}$$

The parameters λ and p are assumed independent a priori. (Papoulis & Pillai, 2002)

Posterior Distributions

By Bayes theorm:

$$\pi(\lambda, p \mid x, z) \propto L(\lambda, p \mid x, z)\, \pi(\lambda)\, \pi(p) \tag{13}$$

Posterior Distribution of λ
Combining the Poisson likelihood and the Exponential prior yields:

$$\pi(\lambda \mid x) \propto \lambda^x e^{-(\alpha+1)\lambda} \tag{14}$$

Thus,

$$\lambda \mid x \sim Gamma(x+1, \alpha+1) \tag{15}$$

where the second parameter denotes the rate parameter.

The posterior mean is:

$$E[\lambda \mid x] = \frac{x+1}{\alpha+1} \tag{16}$$

Posterior Distribution of p
Combining the Binomial likelihood and the Beta prior yields:

$$p \mid x, z \sim Beta(a+z, b+x-z) \tag{17}$$

The posterior mean is:

$$E[p \mid x, z] = \frac{a+z}{a+b+x} \tag{18}$$

Conjugacy ensures analytical tractability of the posterior updating process.

Optimal Bonus–Malus Premium

Let:
x = observed total claims,
z = observed number of large claims
t = policy duration,
w_L = weight assigned to large claims,
w_S = weight assigned to ordinary claims,
With:

$$w_L > w_S > 0 \tag{19}$$

Define the risk function:

$$g(x, z) = w_L z + w_S(x - z). \tag{20}$$

This can be rewritten as

$$g(x, z) = x[w_S + (w_L - w_S)\frac{z}{x}]. \tag{21}$$

Bonus Function
To incorporate the effect of the claim-free duration, the following bonus factor is defined:

$$B(t) = \frac{1}{1+\delta t}, \delta > 0 \tag{22}$$

The function $B(t)$ is strictly decreasing in t, ensuring that a longer claim-free duration results in a premium reduction.

Under a squared-error loss function, the Bayes estimator minimizing posterior expected loss is the posterior mean:

$$p * (x, z, t) = \frac{x+1}{\alpha+1}\left[w_S + (w_L - w_S)\frac{a+z}{a+b+x}\right]\frac{1}{1+\delta t} \tag{23}$$

This expression shows that the premium depends jointly on the claim frequency, proportion of large claims, and policy duration (Lemaire, 1995).

RESULT AND DISCUSSION

Descriptive Statistics of Simulated Portfolio

Table 1: Summary Statistics of Claim Frequency and Claim Size

	Min	Median	Max	Mean
Claim Frequency	0	0	4	0.39
Claim Size	0	0	84.19	17.16

The claim size distribution exhibits substantial variability, with a relatively high standard deviation compared to its mean. This right-skewed structure supports the introduction of a critical threshold ψ to distinguish large claims from ordinary claims in the premium calculation process.

Distribution of Claim Frequency and Claim Size

Figure 1: Distribution of Claim Frequency

Figure 1 indicates that most policyholders report zero claims, followed by a smaller proportion that reports one claim. Higher claim frequencies rarely occur, which is consistent with the low-intensity Poisson assumption.
    This structure validates the suitability of the Poisson distribution for modeling the claim frequency in motor vehicle insurance.

Distribution of Claim Frequency and Claim Size

Figure 2: Distribution of Claim Size

Figure 2 shows a right-skewed claim-size distribution. Large claims occur with a lower frequency but contribute disproportionately to potential loss exposure. This characteristic motivates the use of a critical value of ψ to differentiate premium adjustments.

Joint Distribution of Total and Large Claims

Table 2: Joint Distribution of Total Claims (X) and Large Claims (Z) ($\psi = 35.21$)

$x \backslash z$	0	1	2	3	4	Total
0	680	0	0	0	0	680
1	78	175	0	0	0	253
2	5	27	25	0	0	57
3	0	1	4	4	0	9
4	0	0	0	1	0	1
Total	763	203	29	5	0	1000

Table 2 presents the joint distribution of total claims X and number of claims exceeding the critical value Z. The results show that:

- The majority of policyholders have Z = 0,

- The number of large claims increases as total claims increase,

- However, large claims remain relatively infrequent compared to the total claims.

This demonstrates that the selected critical value provides sufficient variation in Z while preserving the realistic portfolio characteristics.

A higher critical value further reduces the number of claims classified as large however the overall premium adjustment mechanism remains structurally consistent.

Posterior Estimation

From Section 2, the posterior distributions are:

$$\lambda \mid x \sim Gamma(x + 1, a + 1) \tag{24}$$

$$p \mid x, z \sim Beta(a + z, b + x - z) \tag{25}$$

The corresponding posterior means are:

$$E[\lambda \mid x] = \frac{x + 1}{a + 1}, E[p \mid x, z] = \frac{a + z}{a + b + x}, \tag{26}$$

The posterior of λ increases with the total claims, whereas the posterior of p increases with the number of large claims.

The prior hyperparameters introduce a shrinkage effect, moderating extreme adjustments and enhancing premium stability.

Premium Behavior

Substituting the posterior means into the risk function yields the explicit premium:

$$P^*(x, z, t) = \frac{x + 1}{\alpha + 1} \left[w_S + (w_L - w_S) \frac{a + z}{a + b + x} \right] \frac{1}{1 + \delta t} \tag{27}$$

The monotonicity of each premium component can be directly verified from the analytical expression. Premiums increase with total claims and the number of large claims, while decreasing with longer policy duration. These findings confirm that the proposed model successfully integrates frequency, severity classification, and duration effects within a coherent Bayesian premium adjustment framework. Table 3 illustrates the premium structure for various combinations of total claims, large claims, and policy duration.

The results reveal three key pattems:

- Premiums increase as total claims increase.

Table 3: Premium Amount at the Critical Value ($\psi = 35.21$)

$f(x,z)\backslash t$	0	1	2	3	4	5
(0,0)	1	0.71	0.55	0.45	0.38	0.33
(1,0)		1.39	1.08	0.88	0.75	0.65
(1,1)		1.45	1.13	0.92	0.78	0.68
(2,0)		2.03	1.58	1.29	1.09	0.94
(2,1)		2.12	1.65	1.35	1.14	0.99
(2,2)		2.21	1.71	1.40	1.19	1.03
(3,0)		2.64	2.05	1.68	1.42	1.23
(3,1)		2.75	2.14	1.75	1.48	1.28
(3,2)		2.87	2.23	1.82	1.54	1.34
(3,3)		2.98	2.31	1.89	1.60	1.39
(4,0)		3.24	2.52	2.06	1.74	1.51
(4,1)		3.37	2.62	2.14	1.81	1.57
(4,2)		3.50	2.72	2.22	1.88	1.63
(4,3)		3.63	2.82	2.31	1.95	1.69
(4,4)		3.76	2.92	2.39	2.02	1.75

- For a fixed number of total claims x, higher values of z result in substantially higher premiums.

- Premiums decrease with longer policy duration t, reflecting the bonus component of the system

These findings confirm that the proposed model successfully integrates claim frequency and claim size classification into a coherent Bayesian premium adjustment framework.

CONCLUSION

This study develops an optimal premium determination model within a Bayesian bonus–malus framework that jointly incorporates claim frequency, severity classification, and policy duration. Claim frequency is modeled using a Poisson distribution with an Exponential prior, while the probability of claims exceeding a predefined critical threshold is modeled using a Binomial distribution with a Beta prior. The conjugate structure ensures analytical tractability and yields closed-form posterior distributions.

The resulting premium formula

$$P^*(x, z, t) = \frac{x+1}{\alpha+1}\left[w_S + (w_L - w_S)\frac{a+z}{a+b+x}\right]\frac{1}{1+\delta t} \tag{28}$$

demonstrates that premium adjustments are driven by three structural components: total claim frequency, the proportion of large claims, and claim-free duration. The model ensures that premiums increase monotonically with total claims and the number of large claims, while decreasing with longer policy duration. This behavior is fully consistent with the economic principles underlying bonus–malus systems.

From a theoretical perspective, the study extends existing Bayesian bonus–malus models by explicitly integrating severity classification through a critical-value mechanism within a coherent decision-theoretic framework. The closed-form solution provides transparency in parameter interpretation and facilitates sensitivity analysis with respect to prior hyperparameters and severity weights.

From a practical standpoint, the proposed model offers insurers a flexible premium adjustment mechanism that simultaneously accounts for frequency risk, severity differentiation, and policyholder loyalty. The Bayesian structure also introduces a natural shrinkage effect, enhancing premium stability and reducing extreme fluctuations.

Nevertheless, the study relies on simulated data and specific distributional assumptions. Future research may extend the framework to real insurance datasets, explore alternative prior structures such as hierarchical Bayesian models, or incorporate dynamic updating across multiple policy periods.

Overall, the proposed framework provides a mathematically coherent and operationally feasible approach for optimal premium determination in motor vehicle bonus–malus systems.

References

Adisti, R. I. & Mutaqin, A. K. (2021). Perhitungan premi murni pada sistem bonus malus untuk frekuensi klaim berdistribusi binomial negatif dan besar klaim berdistribusi weibull pada data asuransi kendaraan bermotor di indonesia. *Gaussian*, 10:170–179.

Deniz, E. G. (2016). Bivariate credibility bonus-malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics*, DOI: `10.1016/j.insmatheco.2016.06.009`.

Hernawati, L., Purnaba, I. G., & Mangku, I. W. (2017). Comparing bonus malus premiums of two types claim which is assessed using bayesian method. *Applied Mathematical Sciences*, 11:1687–1694.

Hines, W. W. & Montgomery, D. C. (1990). *Probabilitas dan Statistik Dalam Ilmu Rekayasa dan Manajemen*. UI-Press, Jakarta.

Hogg, R. V., Craig, A. T., & McKean, J. W. (2019). *Introduction to Mathematical Statistics*. Pearson Education International, United State of Amerika.

Hutasoit, E. P. (2019). Penentuan premi bonus malus pada asuransi kendaraan menggunakan metode bayes.

Lemaire, J. (1995). *Bonus Malus Systems in Automobile Insurance*. Springer Science, New York.

Moumeesri, A. & Pongsart, T. (2022). Bonus-malus premium based on claim frequency and the size of claims. *Risks*, DOI: `10.3390/risks10090181`.

Noviyanti, L., Soleh, A. Z., & Handoko, B. (2017). Premi asuransi dengan sistem bonus malus optimal. *BIAStatistics*, 11:52–63.

Papoulis, A. & Pillai, S. U. (2002). *Probability*. McGraw-Hill, New York.

Tse, Y.-K. (2009). *Nonlife Actuarial Models*. Cambridge University Press, New York.

Walpole, R. E. & Myers, R. M. (1995). *Ilmu Pengetahuan dan Statistika untuk Insinyur dan Ilmuan*. ITB, Bandung.