Vol. 34, No. 4 (2024) | DOI: 10.37190/ord240404

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

Adapting the insurance pricing model for distribution channel expansion using the Bayesian generalized linear model

Carina Gunawan[1] Muhamad Ivan Faizal[2] Nanang Susyanto[1*]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Asuransi Tugu Pratama Indonesia Tbk., Jakarta, Indonesia*
[*]*Corresponding author, email address: nanang_susyanto@ugm.ac.id*

Abstract

The insurance market is changing due to new distribution channels, requiring insurers to update their pricing models. We propose a mathematical approach using Bayesian generalized linear models (GLM) to adjust insurance pricing. Our strategy modifies the pricing model by incorporating distribution channels while utilizing the initial model as a baseline. Bayesian GLM enable effective model updates while incorporating existing knowledge. We validated our approach using data from the general insurance sector, comparing it with the traditional approach. Results show that Bayesian GLM outperforms the traditional method in accurately estimating pricing. This superiority highlights its potential as a powerful tool for insurers to remain competitive in a rapidly changing market. Our approach makes a significant mathematical contribution to insurance pricing, allowing insurers to adapt to market conditions and enhance their competitive edge.

Keywords: *distribution channel, insurance pricing, competitiveness, predictive performance, Bayesian GLM*

1. Introduction

General insurance is one of the leading insurance products in the Indonesian insurance environment. The reliability and often low price captivate a lot of the public, especially health and vehicle insurance. In addition, a much more intuitive approach to the empirical data at hand is that the past COVID-19 pandemic has dramatically grown awareness of the importance of insurance as a financial protection instrument and further complicates the insurance and risk landscape of general insurance.

Seeing the growth of demand and the general insurance market, a lot of companies would naturally see this as an opportunity to expand their existing business models. This could be done in many ways, one of which is opening and expanding the number of distribution channels. On the other side of this advancement, actuaries would also have to adjust to the massive inflation of risk and policies. These prove a new demand for a flexible model that can accommodate the ever-changing and growing landscape of

Received 5 November 2023, 26 September 2024, accepted 26 September 2024, published online 19 December 2024
ISSN 2391-6060 (Online)/

The costs of publishing this issue have been co-financed by the Department of Operations Research and Business Intelligence at the Faculty of Management, Wrocław University of Science and Technology, Wrocław, Poland

the Indonesian insurance industry. One famously known method is the generalized linear model (GLM), which is an expansion of the traditional regression. GLM liberates users from the constraint of the traditional regression, which is its limitations to model strictly normally distributed data [15]. This particular flexibility is very advantageous to actuaries, as insurance data such as claims and the number of claims are rarely (in other words, never) normally distributed. More often, it is more suitable to use Poisson distribution or Gamma distribution to describe claim patterns [13].

Again, further advancements are made to the existing GLM model. This time, incorporated into the GLM is the Bayesian principle, specifically on the regression coefficients. As with traditional Bayesian, these coefficients are assumed to be random variables [14], therefore prior distributions are assumed for each regression coefficient. These assumptions are made considering there is little to no reliable data when we are initially making models for product expansions. As such, we may use the limited data available, as the prior distribution of the regression coefficient of the Bayesian GLM (BGLM). These data could be obtained by borrowing data from another area, or existing products that may be similar in terms of risk and claim patterns (i.e., sharing common characteristics such as demographic, geographic, climate, etc.).

As its namesake, the basic method that makes up the foundation for the Bayesian GLM modeling is the generalized linear model (GLM). There is an extensive array of literature on the subject such as [12, 15], and [6], as they have extensive information on the basics of general insurance pricing (such as aggregate models, rating factors, frequency and severity modeling, etc.) and the application of GLM on the subject. As for Bayesian analysis, readers could refer to [4] and [5]. In addition, some articles that show some applications and advancements in Bayesian modeling are [16] that cover mainly Bayesian analysis applications in pharmaceutical research and a lot of discussions involved in the paper that show various other applications of the Bayesian inference technique, [8] presents Bayesian's ability to resolve null values, and [3] which illustrates Bayesian's implementation on various case studies, ranging from medicine, economics, and engineering.

The leading advantage of using Bayesian methods lies in its ability to combine information from multiple sources. This ability accounts for the model's flexibility, by incorporating multiple levels of randomness and combining information from multiple sources. In addition, by estimating a posterior conditionally based on real or newly incorporated data, Bayesian gives more objective results in its model. It also takes into greater account a statistical problem's uncertainty, hence making it a great tool for presenting a certain unknown landscape to clients [4]. Conversely, Bayesian is highly dependent on its prior distribution, even more so in cases where there is little to no data available. While prior distributions are a great tool to incorporate personal judgment that may aid in model enhancement, an inaccurate prior distribution may also result in a misleading model due to its subjective nature, especially where there is no data to override it with. An overly flexible prior may also result in overfitting if there is the data involved is insufficient. Bayesian analysis is also computationally expensive, as it relies heavily on simulations such as Markov chain Monte Carlo (MCMC) to compensate for its minimally required observations [9].

Particularly, in the actuarial science landscape, Bayesian statistics has been an intensive and promising topic. It started from Bühlmann and Straub's innovation on Bayesian credibility theory in 1967, and ever since then it has grown into a reliable method for almost every aspect of insurance modeling. In this paper, we are about to take a look at Bayesian statistics' contribution to the pricing aspect of insurance.

In the insurance pricing industry, Bayesian GLM has been shown as a useful method in automobile insurance risk modeling, where it is deemed an excellent model for incorporating the personal judgment of actuaries and available yet limited data, hence leading to more reasonable vehicle insurance rates [2]. Furthermore, Bayesian GLM also shows potential in modeling in conditions where there is little to no data available, by borrowing data from another class of business that is similar to the risk of interest [20]. This ability is particularly useful in the practice of pricing new and unique insurance products, where it cannot be guaranteed that an actuary will have reliable data on hand to work with. The Bayesian GLM, as described, was implemented in [18], where spatial variable selection was employed to assess the impact of weather on insurance claims. In a similar vein, [19] also utilizes Bayesian modeling for insurance risk assessment. However, there has been a lack of methods addressing the adoption pricing model, particularly in the context of the expansion of distribution channels using real insurance data, specifically within the Indonesian market.

This paper aims to propose a method of adjusting the insurance pricing model to incorporate the expansion of distribution channels by using BGLM. Here we use the data of vehicle insurance of a general insurance company in Indonesia, from 2016 to 2021, to present a more realistic take on the model with more real and representative data, albeit messier. We are using the data of 2016–2020 as the prior distribution, early 2021 as the data that will update the prior, and late 2021 as the data for testing. We also emphasize that there is a strikingly high increase in the number of claims received by the company in the year 2021. The reason behind this mighty increase is that the company made a significant expansion of its vehicle insurance. As there was an increase in the policyholder from the expansion, the spike in claims naturally came following. The effect of this spike will be captured by the proposed method in adjusting the insurance pricing model.

In the following section, we will further explore the application of the Bayesian GLM technique for adjusting pricing models. In Chapter 2, the Bayesian GLM model that we are utilizing will be explained. It includes model and prior distribution specifications, updating the prior distribution based on newly observed data, and incorporating new information into an existing pricing model using Bayesian GLM. Chapter 3 will state the data used and the implementation of the model on the data we have used. Some discussion of the results and suggestions for further research on the topic is provided in Chapter 4 which also delivers the conclusion of the study.

2. Methodology

2.1. Overview of Bayesian GLM methodology

The basis of the Bayesian GLM is, as its name, to incorporate Bayesian theories into the GLM method. Usually, in traditional GLM, we would determine the coefficients by using the frequentist method, more commonly known as the maximum likelihood estimator (MLE). However, this method comes with a big stumbling block, that is its reliability on good data. Estimations that are based solely on past data can become problematic when there is insufficient data available. This is particularly true when expanding the distribution channels, as the past data may no longer accurately represent a newly transformed landscape.

To address the aforementioned problem, Bayesian principles come into play. A trademark of Bayesian statistics is its subjective approach to a previously exact one. Bayesian challenges the notion that model parameters are fixed, and as probability is the best tool in the box to tackle uncertainty and find unknown values, so can it be used to determine unknown model parameters [11]. The Bayesian method does so by determining a prior distribution for the unknown values, this represents our subjective judgment on the parameter of interest. As more data is gathered, we can update the model we are using, until theoretically, we finally reach the real value of each parameter, which is represented in the posterior distribution. Mathematically, the posterior of the parameter of interest can be described as such

$$\pi(\theta X) = \frac{\pi(X\theta)\pi(\theta)}{\int_{\theta} \pi(X\theta)\pi(\theta)d\theta} = \frac{\pi(X\theta)\pi(\theta)}{\pi(x)} \tag{1}$$

where given a data set of $X = \{x_1, x_2, \ldots, x_n\}$. Here, $\pi(X\theta)$ stands for the likelihood function of θ, $\pi(\theta)$ is the prior distribution, and the denominator is the marginal distribution of x. Another much simpler form of the posterior distribution is given by

$$\pi(\theta X) \propto \pi(X|\theta)\pi(\theta) \tag{2}$$

Theoretically, as we have more data, the likelihood value will get larger and eventually override the uncertainty presented by the prior distribution $\pi(\theta)$, therefore we get closer to the real value.

As with traditional GLM modeling for insurance pricing, we will construct two separate models for each claim frequency and claim severity. Claim frequency pertains to the count of claims submitted by policyholders to the insurance company, while claim severity quantifies the actual monetary losses incurred within a specific period, typically one year. Each model, aside from differences in model specifications, will for the most part receive the same treatment of prior and model specification, posterior estimation, model checking, and posterior analysis. However, it is advised for analysts to do some unique adjustments if needed for each model, such as parameter range adjustments or changing the distribution to better suit the data, depending on the result we get from the program.

It is common for Bayesian statistics to use simulation techniques due to the high dimensional integration and analytic calculation that is required to estimate the posterior. Here we use Markov Chain Monte Carlo (MCMC) simulation to estimate the posterior. More specifically we are using the No-U-Turn Sampling (NUTS) algorithm by [7], which is an update to the Hamiltonian Monte Carlo (HMC) algorithm previously done by [1], where it basically determines its own required leapfrog steps by using a slice variable and declaring a stop criterion.

2.2. Prior distribution specification

A critically important step in Bayesian analysis is the specification of prior distribution, as this distribution represents the aspect of uncertainty, due to the lack of informative data. Two ways are typically used to discern which prior distribution to use. The first one is the population interpretation, where we assume that the prior distribution represents a population of possible parameter values. A more subjective approach is the state of knowledge interpretation, in which we include our uncertainty about the parameter of interest and assume that the parameter value could be realized at random from our prior distribution. The prior distribution need not be concentrated and close to the real value which is unknown, this is due to the information from the real data which is expressed in the likelihood distribution will usually outweigh any reasonable prior distribution [4].

Some distributions are commonly acquainted and used in the insurance and risk modeling industry. For frequency modeling, the most commonly used distribution is the Poisson distribution, but aside from that, in conditions where there are over-dispersion properties present, negative binomial and zero-inflated Poisson may be a reliable option. In terms of severity modeling, usually, the gamma distribution is mostly utilized due to its ability to accommodate heavy-tailed characteristics.

2.3. Updating prior and incorporating new information to an existing model with Bayesian GLM

We will be employing the so-called three-pricing period principle to incorporate and update the prior distribution based on the newly obtained data information. The three pricing period principle is in essence splitting the pricing process into three separate parts:

1. First period, where we have no data available representing the new situation and therefore we use the old data (prior data) to fit the model.
2. Second period, we update the pricing model with newly available data (this process is repeated for several periods until it is deemed to be enough experience).
3. When enough experience is obtained (in other words, we have sufficient data to make a good model), traditional GLM may be used to fit the model.

It is however, important to notice that in the process of fitting each model pay careful attention to the model fit of each BGLM and traditional GLM, as well as the number of observations used in each data, as it may occur after a certain amount of information is used, the GLM model is deemed as a superior than of its Bayesian counterpart.

3. Results

3.1. Description of data used

The data used in this study is the loss register for vehicle insurance from a general insurance company based in Indonesia from the year 2016–2021. The data set includes the underwriting year, loss year, claim amount, number of claims, and vehicle category (5 levels). The number of claims of the dataset we are using consists of 96.57% of zero claims and 6.43% of 1 claim. Meanwhile, for the severity, the smallest claim is at IDR 4,860, the maximum claim is IDR 556,843,605, and the mean claim IDR 7,908,519.

In addition to the statistical information stated above, business and practical information, the company has recently expanded its distribution channels in 2019. This expansion accounts for massive inflation of the number of policies and claims received in 2020–2021, to accommodate the change in risk landscape as well as claims, we will use Bayesian GLM to adjust and incorporate the new information obtained post-expansion. As a standard procedure for Bayesian techniques, it is required to declare a data set for our prior information.

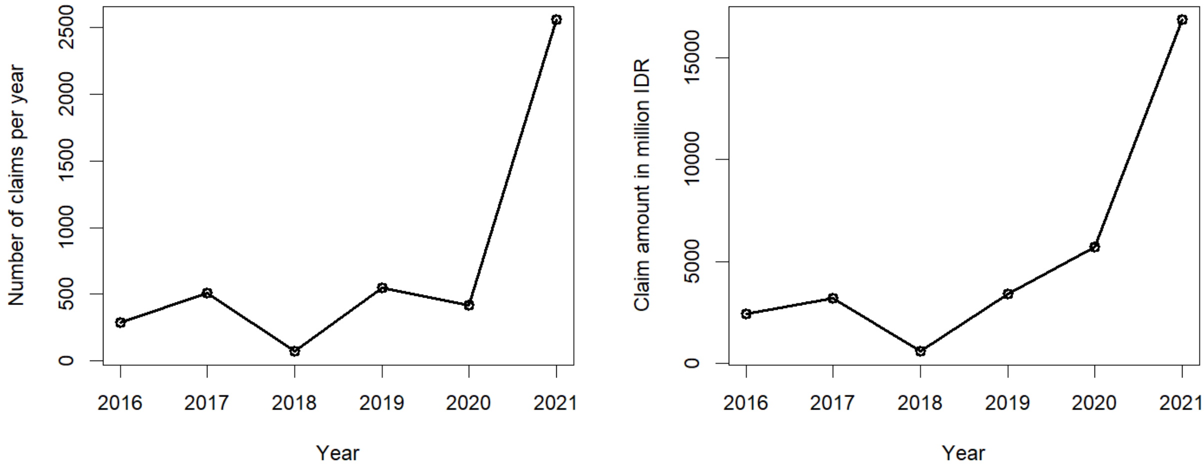

Figure 1. Claim frequency (left) and severity (right) per loss year

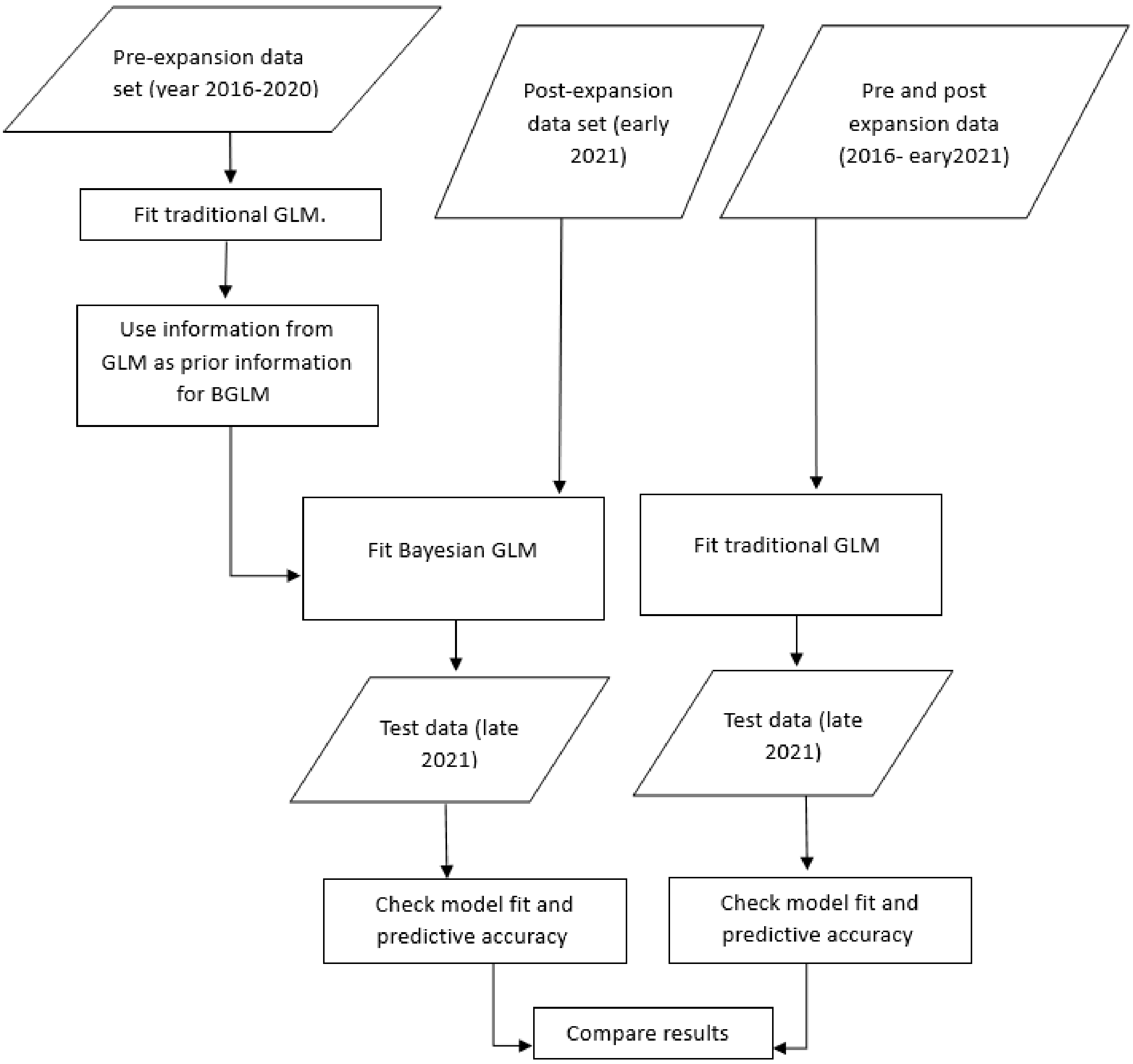

Figure 2. Workflow of incorporating new information with Bayesian GLM

Based on the information we have, as seen in Figure 1, there is a drastic change in the claims incurred in the year 2021. When comparing our approach with the traditional one, we follow the same procedure, as illustrated in Figure 2. Indeed, in 2020, the company expanded its distribution channels, and it took a year to reach the customers. The substantial increase in this expansion was primarily due to the company's significant expansion of its vehicle insurance offerings. With an increase in policyholders resulting from this expansion, there was a subsequent surge in claims. The proposed method aims to account for the impact of this spike in claims when adjusting the insurance pricing model. The prolonged duration of this process could have been influenced by the COVID-19 pandemic. Therefore, we have chosen to utilize the data from 2016–2020 as prior information.

3.2. Prior distribution specification for insurance pricing

3.2.1. Frequency model

For this paper, we use the logistic regression. This uncommon approach is made mainly due to the limitation of the data that we have. The data does not include exposure, normally represented by policy duration which is critical for the usual Poisson regression. Another supporting circumstance is that the number of claims filed is always either 1 or 0, hence the classification method chosen. We model

$$Y_i \sim \text{Bernoulli}(\pi_i) \tag{3}$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \tag{4}$$

where $Y_i\theta_i$ stands for the claims filed by the ith policyholder, provided we know θ_i, which is the rate of claims. Meanwhile, the vector $(x_{1i}, \ldots, x_{pi})$ stands for the rating factors of policyholder i. Here, $p+1$ is the number of parameters needed.

In the model, we assume the coefficients to have a prior distribution that follows the normal distribution

$$\beta_j|\mu_j, \sigma_j^2 \sim \text{normal}(\mu_j, \sigma_j^2), \quad \text{for} \quad j = 0, 1, \ldots, p \tag{5}$$

and for the priors for each parameter, the mean μ_j will take the information we have from the traditional GLM with prior distribution data, and the variance parameter σ_j^2 will follow the uniform distribution $\text{uniform}(0, 10)$.

3.2.2. Severity model

As stated before, we make a separate model for claim severity, with its own predicted variable and rating factors. The rating factors are denoted as the frequency as $x_i = (x_{1i}, \ldots, x_{1i})'$, however, the rating factors used here can differ from the frequency model if need be. Here we take an unorthodox approach, that is we use the log-normal regression. This decision is primarily based on the data at hand, which exhibits a better fit with the lognormal distribution

$$Z_i \sim \text{lognormal}(\mu_i, \sigma_i) \quad \text{for} \quad i = 0, 1, \ldots, p \tag{6}$$

$$\mu_i = \gamma_0 + \gamma_1 x_{1i} + \cdots + \gamma_q x_{qi} \tag{7}$$

and

$$\sigma_i^2 \sim \text{uniform}(0, 10) \tag{8}$$

As with the frequency model, we are interested in determining the coefficient $\gamma_0, \gamma_1, \ldots, \gamma_q$ with both frequentist and Bayesian approaches. For the Bayesian approach, we choose the normal distribution as the weakly informative prior for the aforementioned regression coefficients

$$\gamma_j|\mu_j, \sigma_j^2 \sim \text{normal}(\mu_j, \sigma_j^2) \quad \text{for} \quad j = 0, 1, \ldots, q \tag{9}$$

where equivalently as the frequency method, the mean μ_j will follow the information we have from the coefficient of the traditional GLM model with prior data, and the variance parameter follows the uniform distribution as such that $\sigma_j^2 \sim \text{uniform}(0, 10)$.

3.3. Implementation of Bayesian GLM in adjusting insurance pricing models

Both the frequency and severity models are implemented using Python with the package PyMC by [17]. The advantage that this package has compared to others is its ease and flexibility. The package is equipped with several built-in packages, such as Arviz by [10], that support its users even further in model construction and analysis.

Regarding the rating factors for our GLM models, both traditional and Bayesian, we refer to the Financial Services Authority Circular Letter No. 6/SEOJK.05/2017. According to this circular letter, there are five categories of vehicles for non-bus and non-truck types, which are determined by the Sum Insured as outlined in Table 1. It is important to note that our data do not include policies for buses and trucks. We have selected vehicle category 1 as the base variable.

Table 1. Sum insured for each vehicle category [10^6 IDR]

Category	1	2	3	4	5
Sum insured	0–125	125–200	200–400	400–800	> 800

3.3.1. Frequency model

As a part of the three-period pricing principle, we initially construct a traditional GLM model. This choice is primarily based on the availability of sufficient prior data, which allows us to employ the frequentist method. The results obtained from the traditional GLM method are summarized in Table 2.

Table 2. Traditional GLM for frequency with prior data (2016–2020)

	Estimate	Standard error	95% CI	
Intercept	–1.9021	0.0758	–2.0287	–1.7793
Vehicle category 2	0.8166	0.0902	0.6695	0.9662
Vehicle category 3	0.5634	0.0836	0.4274	0.7023
Vehicle category 4	0.9936	0.0851	0.8549	1.1351
Vehicle category 5	0.9560	0.1043	0.7850	1.1281

As our objection, here is to use the prior data as our judgment basis for the Bayesian GLM, we use the parameter estimates found from the traditional GLM as the value of parameter mean μ_j, which is the mean parameter of the prior distribution of the BGLM coefficients. Meanwhile, as we do not have any information on the variance parameter σ_j^2, we leave it as it is, which follows the uniform distribution. By using MCMC from the package PyMC, we obtained the result given in Table 3. The parameter estimates are obtained from 24,000 MCMC iterations; 4 chains of 5,000 draws and 1,000 tuning iterations each.

Table 3. Bayesian GLM for frequency with post-expansion data (first half of 2021)

	Estimate	Standard error	95% CI	
Intercept	–5.7988	0.0011	–5.9273	–5.6763
Vehicle category 2	2.6959	0.0013	2.5417	2.8418
Vehicle category 3	2.6780	0.0012	2.5370	2.8286
Vehicle category 4	2.7166	0.0014	2.5447	2.8956
Vehicle category 5	2.5844	0.0025	2.1958	2.9511

3.3.2. Severity model

While the method we use here for the severity model mirrors the one for frequency, modeling severity requires some adjustments to the data we are using. We only model claims that have incurred, this means that we only use claims that are greater than zero. We construct a traditional GLM using the data from pre-expansion, that is year 2016–2020, and the result is illustrated in Table 4

Table 4. Traditional GLM for severity with prior data (2016–2020)

	Estimate	Standard error	95% CI	
Intercept	15.0553	0.07627	14.9297	15.1808
Vehicle category 2	–0.0153	0.0891	–0.1619	0.1314
Vehicle category 3	0.0284	0.0848	–0.1112	0.1679
Vehicle category 4	0.1167	0.0924	–0.0354	0.2688
Vehicle category 5	0.2054	0.1089	0.0261	0.3846

Next, we use the parameter estimation result for building the Bayesian GLM. This model is constructed using the information we obtained in the first year of expansion, that is the first half of 2021. To get the results, we ran 4 chains of MCMC iterations, with each chain containing 5,000 draws and 1,000 tuning draws. The estimation from the Bayesian GLM model is provided by Table 5.

Table 5. Bayesian GLM for severity with post-expansion data (first half of 2021)

	Estimate	Standard error	95% CI	
Intercept	14.8757	0.0014	14.7500	14.9912
Vehicle category 2	0.0578	0.0014	-0.0750	0.1977
Vehicle category 3	–0.0054	0.0013	–0.1388	0.1351
Vehicle category 4	0.1820	0.0017	–0.0052	0.3684
Vehicle category 5	0.6045	0.0044	0.2140	0.9790

3.4. Comparison with traditional pricing models

In this section, we compare each model with their traditional GLM counterparts. For each BGLM model that was constructed with the year 2020 data, we also made their mirroring GLM model, which we then compare their errors and predictive powers to.

For the frequency model, we obtain the GLM model as presented in Table 6. The results reveal a notable contrast in the standard errors between the two models. In Table 3, the BGLM model demonstrates significantly smaller errors compared to its GLM counterpart. Moreover, while the confidence intervals from the traditional GLM still encompass 0 for some variables, the Bayesian GLM's confidence intervals do not. These findings collectively demonstrate that the Bayesian model exhibits superior capability in describing limited data compared to the traditional approach.

Table 6. Traditional GLM for the frequency with post-expansion data (2016–early 2021)

	Estimate	Standard error	95% CI	
Intercept	–4.9810	0.0709	–5.0999	–4.8665
Vehicle category 2	2.6168	0.0835	2.4807	2.7557
Vehicle category 3	2.7804	0.0782	2.6534	2.9109
Vehicle category 4	3.4666	0.0796	3.3372	3.5993
Vehicle category 5	3.8192	0.0994	3.6563	3.9834

The same was done for the severity model which also compares our approach to the traditional one with the same post-expansion data constructed. The results we got for this traditional approach are provided by Table 7.

Table 7. Traditional GLM for severity model with post-expansion data (2016–early 2021)

	Estimate	Standard error	95% CI	
Intercept	14.9281	0.0549	14.8378	15.0184
Vehicle category 2	0.0565	0.0636	-0.0482	0.1612
Vehicle category 3	0.0750	0.0626	-0.0279	0.1781
Vehicle category 4	0.2201	0.0724	0.1011	0.3392
Vehicle category 5	0.4012	0.0918	0.2501	0.5523

We can assess the performance of each model by comparing the results in Tables 5 and 7, particularly focusing on the standard error and confidence interval. Notably, the Bayesian GLM exhibits considerably smaller standard errors for each variable, providing a clear indication that it outperforms the traditional GLM method.

From Tables 8 and 9, we can see how parameter estimates were updated from the GLM prior by traditional and Bayesian GLM.

Table 8. Parameter change for each method (frequency)

	GLM with prior	GLM with post-expansion	Bayesian GLM
Intercept	–1.9021	–4.981	–5.802
Category 2	0.81663	2.61683	2.669
Category 3	0.56336	2.78041	2.681
Category 4	0.99357	3.46662	2.720
Category 5	0.95601	3.81923	2.588

Table 9. Parameter change for each method (severity)

	GLM with prior	GLM with post-expansion	Bayesian GLM
Intercept	15.05525	14.92818	14.8757
Category 2	–0.01526	0.05649	0.0578
Category 3	0.02835	0.07508	–0.0054
Category 4	0.11672	0.22013	0.182
Category 5	0.20541	0.40124	0.6045

We also tested the prediction power of each model both for frequency and severity models. Because our data only contains claim frequencies of 0 and 1, we utilize the true positive rate (TPR) and true negative rate (TNR) to assess the accuracy of both frequency models. TPR calculates the probability that a policyholder submits a claim when they do, whereas TNR signifies the probability that a policyholder does not submit a claim when they indeed do not. Whereas, for comparing the performance of severity models, we employ mean absolute error (MAE) and mean absolute percentage error (MAPE). As we can see from Table 10, the Bayesian GLM outperforms the traditional one. As a reminder, our approach involves utilizing data from 2016 to 2020 as the prior dataset, using the first half of 2021 as the data to compute the posterior, and reserving the second half of 2021 for testing the models, which includes the calculation of MAE and MAPE.

Table 10. Prediction errors

	TPR [%]	TNR [%]	MAE	MAPE
Traditional GLM (2016–2020)	89.252	73.83	3,144,814	1.3705
Traditional GLM (2016–early 2021)	89.252	73.83	3,107,431	1.2694
Bayesian GLM (early 2021)	93.614	78.94	3,073,123	1.1465

Table 11. Pure premium with traditional GLM and Bayesian GLM

	Traditional GLM with prior	Traditional GLM	Bayesian GLM
Category 1	448,663.09	20,749.48	8,696.56
Category 2	872,212.75	261,472.48	120,576.10
Category 3	717,598.15	303,308.44	121,970.23
Category 4	992,534.58	548,495.65	126,608.30
Category 5	1,186,495.32	1,083,088.62	204,186.96

Furthermore, the pure premium estimated by the Bayesian GLM is lower than that of the traditional GLM as given by Table 11, indicating that it offers a cost advantage. Aside from the model advantage, the explanation behind the striking difference in premium prices can be credited to the massive increase in policy numbers after the expansion of distribution channels.

Therefore, by utilizing Bayesian GLM, insurers can maintain better competitiveness in the market. The improved performance and cost-effectiveness of the Bayesian GLM make it a valuable tool for accurately estimating premiums and staying competitive in the insurance industry.

4. Conclusion

In this research, we have shown how the Bayesian GLM performs as an updating model to accommodate a shift in the risk landscape, in this case from an expansion of distribution channels. The results are shown in the model we constructed present clear evidence that in terms of prediction ability, the Bayesian model surpasses its traditional GLM counterpart. The evidence is presented in the Bayesian GLM's larger accuracy score for the frequency model, where the Bayesian GLM yields a greater true positive rate (TPR) and true negative rate (TNR) compared to the traditional GLM. For the severity model, we compared their prediction ability through mean absolute error (MAE) and mean absolute predicted error (MAPE), the Bayesian GLM performs better than the traditional GLM. The Bayesian GLM also performs well to illustrate new and lower premium prices from an increase in the number of policies, which has happened as a consequence of the channel of distribution's expansion.

The GLM also offers a high level of interpretability, that can give an insight into the relationship and risks related to each variable. This is particularly helpful as a tool for statisticians and actuaries to explain their findings to business persons who might not be familiar with statistics and actuarial methods. In the future, this model may be further explored to see its performance in other aspects of insurance pricing, such as making a model pricing for a newly opened class of business, with prior data borrowed from other existing classes of business.

References

[1] Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[2] Gao, C., Li, Q., and Guo, Z. Automobile insurance pricing with Bayesian general linear model. In *Innovative Computing and Information. International Conference, ICCIC 2011, Wuhan, China, September 17-18, 2011. Proceedings, Part I* (Berlin, 2011), M. Dai, Ed., Springer, pp. 359–365.

[3] Gatsonis, C., Hodges, J. S., Kass, R. E., and Singpurwalla, N. D. *Case Studies in Bayesian Statistics*. Springer-Verlag, 1993.

[4] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B, Vehtari, A., and Rubin, D. B. *Bayesian Data Analysis*. 3rd Edition. Chapman & Hall/CRC. 2013.

[5] Gelman, A., and Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

[6] Goldburd, M., Khare, A., and Tevet, D. *Generalized Linear Models for Insurance Rating*. Casualty Actuarial Society, 2016.

[7] Hoffman, M. D., and Gelman, A. The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research 15*, 1 (2011), 1593–1623.

[8] Kruschke, J. K. Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science 6*, 3 (2011), 299–312.

[9] Kruschke, J. K. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, 2015.

[10] Kumar, R., Carroll, C., Hartikainen, A., and Martin, O. A. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software 4*, 33 (2019), 1143.

[11] Lynch, S. M. Bayesian theory, history, applications, and contemporary directions. In *International Encyclopedia of the Social & Behavioral Sciences* (Oxford, 2015), J. D. Wright, Ed., Elsevier, pp. 378–382.

[12] McCullagh, P. *Generalized Linear Models*, Routledge, New York, NY, 1989.

[13] Ntzoufras, I. *Frontmatter*. In *Bayesian Modeling Using WinBUGS* (Hoboken, HJ, 2009), I. Ntzoufras, Ed., John Wiley & Sons, pp. i–xxiii.

[14] Ntzoufras, I. Introduction to Bayesian inference. In *Bayesian Modeling Using WinBUGS* (Hoboken, HJ, 2009), I. Ntzoufras, Ed., John Wiley & Sons, pp. 1–29.

[15] Ohlsson, E., and Johansson, B. *Non-Life insurance pricing*. In *Non-Life Insurance Pricing with Generalized Linear Models* (Berlin, 2010), E. Ohlsson and B. Johansson, Eds., Springer, pp. 1–14.

[16] Racine, A., Grieve, A. P., Flühler, H., and Smith, A. F. M. Bayesian methods in practice: Experiences in the pharmaceutical industry. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 35*, 2 (1986), 93–150.

[17] Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science 2* (2016), e55.

[18] Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., and Meze-Hausken, E. A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *Journal of the Royal Statistical Society Series C: Applied Statistics 62*, 1 (2013), 85–100.

[19] Tufvesson, O., Lindström, J., and Lindström, E. Spatial statistical modeling of insurance risk: a spatial epidemiological approach to car insurance. *Scandinavian Actuarial Journal 2019*, 6 (2019), 508–522.

[20] Zhang, J., and Miljkovic, T. Ratemaking for a new territory: Enhancing GLM pricing model with a Bayesian analysis. *Casualty Actuarial Society E-Forum 2* (2018), 1–32.