UDC 004.852
DOI: 10.20535/SRIT.2308-8893.2025.4.04

ANALYSIS OF ACTUARIAL RISK WITH GENERALIZED LINEAR MODELS

R.S. PANIBRATOV, P.I. BIDYUK

Abstract. The problem of applying generalized linear models to the analysis of actuarial risks in the context of premium charges to clients was considered. The Monte-Carlo method for Markov chains was applied. Two situations were considered for the computational experiment. For the first one, insurance indicators and the target variable were randomly assigned due to the problem of public data access. To create three datasets, charges were generated from normal, gamma, and Pareto distributions with dynamic variance, and noise was added to stimulate a non-stationary process. In the second situation, actual actuarial data from the Singapore Actuarial Society was used. Generalized Linear Models with normal distribution and logarithmic link function, an exponential distribution and logarithmic link function, and Laplace distribution with identity link function were constructed. Based on the model-fitting quality metrics, conclusions were drawn about their structure.

Keywords: actuarial risk, generalized linear models, simulation modeling, exponential family of distributions, Bayesian data analysis, Monte Carlo method for Markov chains.

INTRODUCTION

Since insurance protects people and organizations financially against a variety of risks, it is seen as a fundamental component of the economy. Because it assists in managing and reducing the risks involved in providing insurance to both consumers and businesses: actuarial science is essential to the insurance sector. A thorough understanding of mathematics, statistics, finance, and economics is necessary to work as an actuary. Actuaries apply their knowledge to assist insurance companies in estimating the cost of possible risks and estimating the probability of future events.

In order to reduce the risks and minimize the financial impact of unpredictable events, the insurance sector is essential. The frequency or timing of these occurrences, however, cannot be predicted. Actuarial risk, or the likelihood of an event happening and the possible financial impact it may have, is a key component that insurance companies utilize to prevent themselves from financial catastrophe. Because actuarial risk is a complicated process that calls for certain knowledge and skills, actuaries are important to the insurance sector. Actuarial risk is fundamentally about estimating the probability of an unfavorable event happening and the possible financial consequences it may have. Actuaries analyze data and forecast the probability of an event by using complex mathematical models. They then use this data to estimate the event's financial effect and compute the premium needed to cover the risk. The business of insurance companies is risk management. Actuaries are essential in assisting insurance firms in figuring out how much risk they may accept while maintaining their financial stability. They accomplish this via examining historical data and applying statistical techniques to forecast the probability that comparable occurrences will take place in the future.

The insurance business uses the Generalized Linear Model (GLM), a statistical technique, to calculate insurance policy prices. In order to analyze and forecast the anticipated cost of claims based on different risk indicators related to the insured entities, generalized linear models are used. Compared to simpler linear models, these models offer a more complex and precise pricing mechanism by allowing actuaries and analysts to include various data types and variable relationships, such as the linear or exponential relationship between risk factors and claim costs.

Linear models are a specific instance of the many models that comprise up GLM. The assumptions of normality, constant variance, and additive effect of that are restricted in linear models are eliminated. Rather, it is assumed that the response variable belongs to the exponential distribution family.

The exponential distributions family consists of the next structure [1]:

$$f(y_i;\theta_i;\varphi)=\exp\left\{\frac{y_i\theta_i-b(\theta_i)}{a_i(\varphi)}+c(y_i,\varphi)\right\},$$

where $a_i(\varphi)$, $b(\theta_i)$ and $c(y_i,\varphi)$ are prior defined functions; θ_i is parameter, associated with mean; φ is parameter, associated with variance.

Additionally, the variance is allowed to change simultaneously with the distribution mean. Lastly, on a transformed scale, it is believed that the variables' effects on the response variable are additive [2].

For GLM, the following assumptions are made:

1. **Stochastic component**: every component of Y comes from the single exponential family distribution and is independent.

2. **Systematic component**: the linear predictor η is formed from p explanatory variables:

$$\eta=X\beta,$$

where X is design matrix; β is vector of estimation parameters.

3. **Link function**: relationship between stochastic and systematic component is defined by the link function, which is monotonic and differentiable:

$$E[Y]=\mu=g^{-1}(\eta).$$

Problem Statement. The purpose of the study is to apply GLM for analysis of actuarial risks using different distributions and specified link functions and previously applying Bayesian data analysis.

IMPORTANCE OF GLM

Because they offer a versatile framework for modeling the link between the response variable (say, such as the frequency or cost of a claim) and one or more predictor factors (such as age, vehicle type or geographic area), GLMs are also utilized in insurance pricing.

The authors of [3] emphasized that when doing statistical studies with GLMs, non-robustness against outliers is an important consideration. Additionally, they demonstrated that there aren't many reliable options, particularly when performing Bayesian statistical analysis. Focusing on gamma GLM, a widely used tool in actuarial science, they put forth a robust and efficient modeling-based method that can be applied to both frequentists and Bayesian studies. The suggested model can be easily estimated, at least on small-to-moderate-sized data sets, and is simple to analyze and comprehend.

The authors of [4] presented a brand-new deep learning technique called Deeply-learned Generalized Linear Model with Missing Data (DLGLM), which can make predictions and estimate coefficients even when there is missing not at random (MNAR) data. The creation of the data matrix and the connections between the response variable and the mask of missing values are modeled by DLGLM using deep learning neural network architecture. They were able to generalize the conventional GLM this way, taking into consideration both ignorable and non-ignorable types of missing values in the data, as well as intricate nonlinear relationships between the features. Through simulations and actual data analyses, the authors also showed that DLGLM outperforms alternative impute-then-regress techniques, such as mean and mouse imputation, in terms of coefficient estimation and prediction when MNAR missing values are present.

The problem of GLM transfer learning was studied in [5]. Bounds for estimate error and the prediction error measure with fast and slow rates under various scenarios are derived by the authors, who also suggested GLM transfer learning methods. To create confidence intervals for each coefficient component with theoretical assurances, they took into account the two-step transfer learning approach. At last, they used a real-data research and simulations to show how effective their algorithms were.

In the context of claim counts modeling, the authors of [6] suggested a method for identifying the next-best interaction to be added to an arbitrary but fixed benchmark GLM. They started by training a combined actuarial neural network (CANN) model, which is essentially a neural network that improves the benchmark GLM. Second, they sorted interactions by their strength and quantified the strength of interactions between each pair of characteristics using a quick model-specific technique called Neural Interaction Detection. Third, they compared a few small GLMs that matched the top-ranked interactions to determine the next-best interaction. This technique offers two benefits. First of all, it is completely automatable method of adding the next-best interaction that is absent from the benchmark GLM. Second, according to Friedman's H-statistic, the authors' methodology is quicker than alternative strategies. As a result, enormous data sets containing millions of observations and dozens of attributes are particularly well-suited for the proposed technique. Consequently, it can significantly reduce the time that price actuaries spend looking for interactions to enhance their GLMs, which is often time-consuming and visual process.

It was demonstrated in [7] that GLM is the best choice for estimation of operational risk. This approach demonstrated excellent risk estimating quality with minimum errors.

Alternative methods of estimating parameters of GLM were analyzed in [8].

MONTE-CARLO METHOD FOR MARKOV CHAIN

Finding the posterior distribution is the primary objective of Bayesian data analysis:

$$P(\theta \mid X) = \frac{P(X \mid \theta) P(\theta)}{P(X)},$$

where X is state space vector; θ is a parameter of distribution; $P(X \mid \theta)$ is the likelihood; $P(\theta)$ is the prior; $P(X)$ is a normalizing constant, also known as the evidence or marginal likelihood.

The denominator can be expressed as follows:

$$P(X) = \int P(X \mid \theta^*) P(\theta^*) d\theta^*.$$

The challenge of assessing the integral in the denominator is the computing problem. Markov Chain Monte Carlo (MCMC) is the most significant of the Monte Carlo techniques that may be employed.

MCMC is the method that uses a Markov chain mechanism to generate samples $x^{(i)}$ while exploring the state space, X. The purpose of this technique is to increase the amount of time the chain spends in the most crucial areas [9]. It is specifically designed to make the samples $x^{(i)}$ resemble samples generated from the desired distribution, $p(x)$.

Monte Carlo is the method for approximating a desired quantity by sampling from a probability distribution. It estimates a deterministic quantity of interest using randomization. The Monte Carlo approach is used to approximate such numbers by averaging over samples. For example, if there is an expectation or expectations to estimate, s, they may be extremely complicated integrals or perhaps impossible to estimate:

$$s = \int p(x) f(x) dx = E_p[f(x)],$$

$$\widetilde{s}_n = \frac{1}{n} \sum_{i=1}^{n} f(x^i),$$

where $f(x)$ is the probability density function.

The standard error might be decreased and a reasonably good estimate could be obtained by calculating the average across a large number of samples. One drawback of this approach is that it makes the assumption that sampling from a probability distribution is simple, which isn't always feasible. In many cases, sampling from the distribution is not even feasible. In these situations, we efficiently sample from an intractable probability distribution by using Markov chains.

With a modification, MCMC techniques function similarly to normal Monte Carlo methods, but the produced drawings $x_1, ..., x_n$ are serially correlated rather than independent. Specifically, they are the realizations of a Markov Chain consisting of N random variables, $X_1,...,X_n$.

If and only if, for all positive integers k and, n, these future observations X_{i+n} are conditionally independent of the previous values X_{i-k} given the present value p X_i, then a random sequence $\{X_i\}$ is Markov chain:

$$P(X_{i+n} = x \mid X_i, X_{i-1},..., X_{i-k}) = P(X_{i+n} = x \mid X_i).$$

This condition that sometimes is referred to as Markov property, indicates that the process is memoryless: the probability distribution of the chain's future values is only dependent on its present value con X_i, independent of how the value was arrived at (e. g. the chain's previous transition).

Although MCMC comes in a variety of flavors, the Metropolis–Hastings random walk algorithm is the easiest to implement. Standard uniform distribution, proposal distribution $p(x)$ and the target distribution must be used for applying Metropolis–Hastings algorithm.

The following steps how this algorithm works when given an initial prediction for θ that has a positive probability of being drawn.

1. Select a new suggested value θ_p that equals

$$\theta_p = \theta + \Delta\theta,$$

where $\Delta\theta$ has specific distribution for transition (for example, Normal).

2. Calculate the ratio

$$\rho = \frac{g(\theta_p \mid X)}{g(\theta \mid X)},$$

where g is the posterior probability.

3. To preserve the precise balance of the stationary distribution in the event that the proposal distribution is not symmetrical, the acceptance probability must be weighted and then calculated:

$$\rho = \frac{g(\theta_p \mid X) p(\theta \mid \theta_p)}{g(\theta \mid X) p(\theta_p \mid \theta)}.$$

Given that ratios are being taken, any distribution proportional to g will likewise be canceled by denominator, therefor it may be utilized as follows:

$$\rho = \frac{p(X \mid \theta_p) p(\theta_p)}{p(X \mid \theta) p(\theta)}.$$

4. If $\rho \geq 1$, then $\theta = \theta_p$.

If $\rho < 1$, then $\theta = \theta_p$ with probability ρ, else $\theta = \theta$, where the uniform distribution is used.

5. Repeat earlier steps.

Authors in [10] showed that MCMC approaches appear to be quite helpful in a wide range of applications. However, because MCMC methods are imprecise, deviations from the correct findings may occur due to their unpredictability. Because no guaranty can be provided, MCMC should only be utilized in extreme cases and only when there are no other options. As the parameters change over time, performance may also be maximized by dynamically modifying the parameters, especially the covariance matrix, without changing the distribution. Furthermore, for low correlations in higher dimensions, other modifications to Metropolis–Hastings are needed.

The authors of [11] presented a Poisson–Rayleigh model, which is also known as the PR-distribution, with two parameters. They were able to get a number of distinct features. The parameters of the PR distribution have been estimated using Bayesian methods, maximum likelihood, and maximum product spacing. For Bayesian estimation, the estimators were approximated using point and interval estimation using the MCMC approach, which is based on a symmetric loss function. A Bayesian estimator based on gamma priors has been proposed.

New diagnostics for evaluating MCMC algorithms efficiency, reliability, and flexibility using control and attainment maps were presented in [12]. The time needed for hyper-parameter adjustment may be shortened by the results of these new diagnostics. The diagnostics themselves can be carried out on computationally reasonable test problems with known posteriors, as demonstrated there, but they need a non-trivial computational experiment. The results of these diagnostics may be used to determine the optimal algorithm and matching hyper-parameter setup for calibrating a real-world issue that is more computationally demanding and shares traits with the test problems. The convergence of that particular search procedure may then be evaluated by applying the current MCMC diagnostics to the single calibration run of the real-world issue.

In order to increase effectiveness of posterior exploration using MCMC techniques, a Kalman-inspired proposal distribution was presented in [13]. Similar to the analysis stage in the Kalman filter, this novel proposal distribution creates candidate states by taking use the cross covariances of model parameters, measurements, and model outputs. The asymmetric nature of the Kalman-inspired proposal distribution limits its application to a brief burn-in time, following which the chains are evolved using a combination of parallel direction and snooker candidate states. The sampled chains will converge to the precise target distribution thanks to diminishing adaptability. The new proposal distribution may be easily included into any suitable MCMC technique and is not restricted to any particular MCMC methodology.

The authors of [14] investigated Metropolis–Hastings Markov chain convergence rates. The validity of appropriate central limit theorems for Markov chains can be ensured by qualitative convergence rates. The impact of growing dimensions, data size, and other variables on these algorithms' efficiency can be better understood by looking at explicit convergence rates. However, a significant amount of work is still needed in this field since explicit quantitative convergence rates are difficult to establish and remain elusive in many situations of relevance. These subjects are crucial for comprehending Metropolis–Hastings behavior in contemporary issues where there may be a lot of data, a lot of dimensions, or both.

NUMERICAL EXPERIMENT WITH ARTIFICIAL DATA

Due to the case, that actuarial data is not always available, it was decided to simulate first actuarial insurance data artificially following the next structure. Three datasets for experiment were created. For imitating data of policyholders the next features were used:

1. **Age:** numerical variable, which shows age of client and ranges between 19 and 64.

2. **Sex:** categorical variable, which identifies sex of client and has states ‘M’ for male and ‘F’ for female.

3. **BMI:** numerical variable, which shows body mass index of client. Uniform distribution was used for generation.

4. **Region:** categorical variable, which shows place of client’s residence and has state ‘A’, ‘B’, ‘C’ and ‘D’.

5. **Medical History:** categorical variable, which identifies history of previous illnesses of clients and has state ‘Diabetes’, ‘High blood pressure’ or ‘None’.

6. **Exercise:** categorical variable, which shows if client does exercise. It has states ‘Always’, ‘Rarely’ or ‘Never’.

7. **Worker Status:** categorical variable, which shows working status of client and has states ‘Employed’, ‘Student’ and ‘Unemployed’.

8. **Charges:** numerical variable which shows total charges by the insurance company. This is target variable.

For the last feature next 3 distributions were use:

- Normal;
- Gamma;
- Pareto.

For making charges as non-stationary process, algorithm of mixture distribution was applied, which consist of the next steps:

1. Generate random variable p, which has uniform distribution $p \sim U(0,1)$.

2. If $p \in \left[\sum_{i=1}^{k-1} p_i, \sum_{i=1}^{k} p_{i,}\right)$, then generate variable with chosen distribution with fixed parameter of centre and randomly generated scale parameter.

3. Repeat until size of the dataset will be reached.

After generating target variable, the noise, which has zero mean and variable standard deviation was added.

Three GLMs were built for forecasting were implemented with specified link functions:

1. GLM with normal distribution and logarithmic link function.

2. GLM with exponential distribution and logarithmic link function.

3. GLM with Laplace distribution and identity link function.

After implementing GLMs by using MCMC method next metrics of models quality were used:

- Logarithm of maximized value of a likelihood function.
- Akaike information criterion (AIC):

$$AIC = 2 * k - 2 * \ln(\widetilde{L}),$$

where $\widetilde{L}$ is maximized value of likelihood function; k is the number of estimated parameters.

- Bayesian information criterion (BIC):

$$BIC = k * \ln(n) - 2 * \ln(\widetilde{L}),$$

where $\widetilde{L}$ is maximized value of likelihood function; k is the number of estimated parameters; n is the number of data points.

The metric results of GLM parameters estimation for three distinct datasets are shown in Tables 1–3.

Table 1. Results of GLM construction using simulated actuarial insurance data, where charges have normal distribution

Metric	GLM Normal	GLM Exponential	GLM Laplace
Log-Likelihood	1.248	2.552	1.943
AIC	15.503	10.896	14.1134.11
BIC	56.464	47.305	55.0735.0

Table 2. Results of GLM construction for simulated actuarial insurance data, where claim payments have gamma distribution

Metric	GLM Normal	GLM Exponential	GLM Laplace
Log-Likelihood	1.16	3.3053.	2.232
AIC	15.68	9.39	13.535
BIC	56.64	35.798	54.495

Table 3. Results of GLM construction for simulated actuarial insurance data, where claim payments have Pareto distribution

Metric	GLM Normal	GLM Exponential	GLM Laplace
Log-Likelihood	0.261	1.678	0.641
AIC	17.479	12.644	16.718
BIC	58.439	49.053	57.677

From the results of fitting GLMs it can be seen, that GLM with exponential distribution and log link function demonstrated the best results for all datasets. On the other side, GLM with Laplace distribution and identity link function also showed acceptable results for dataset with normal distributions of charges.

Results of forecasting for best GLM models using different datasets are shown on Figs. 1–4.

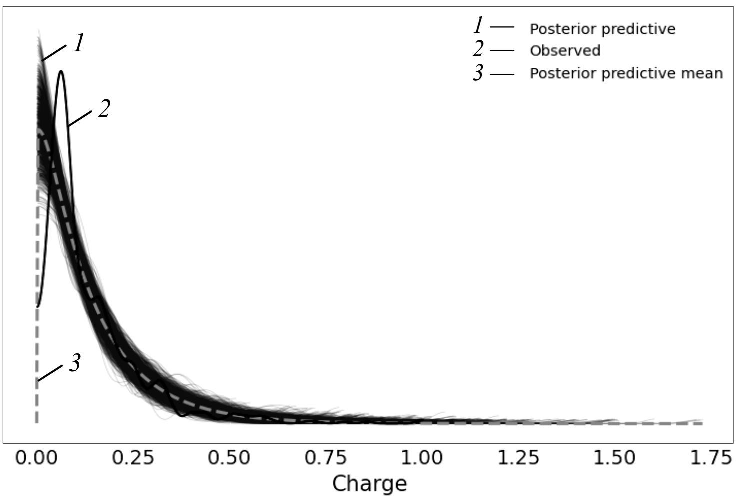

Fig. 1. Result of forecasting GLM with exponential distribution and log link function for charges, which have normal distribution

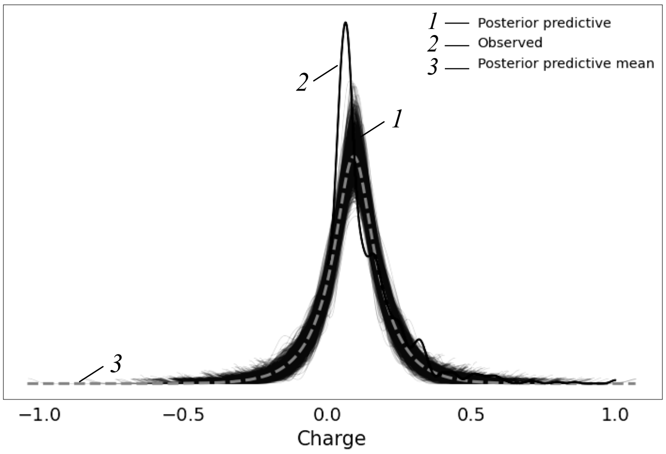

Fig. 2. Result of forecasting GLM with Laplace distribution and identity link function for charges, which have normal distribution

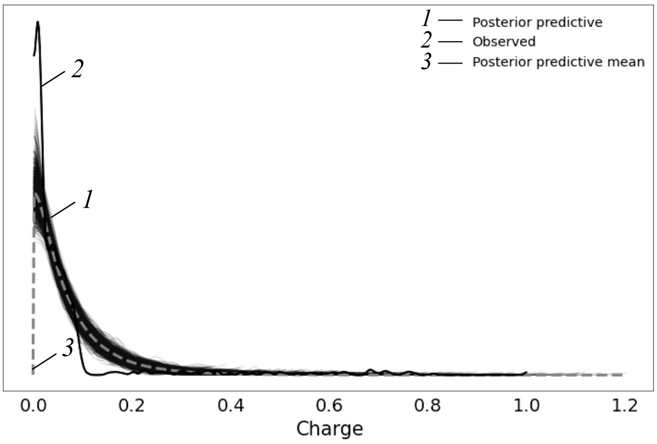

Fig. 3. Result of forecasting GLM with exponential distribution and log link function for charges, which have gamma distribution

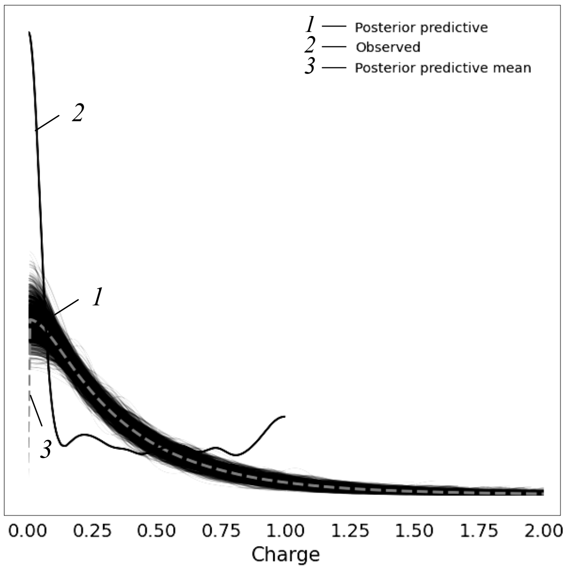

Fig. 4. Result of forecasting GLM with exponential distribution and log link function for charges, which have Pareto distribution

Tables 4–7 show numerical summaries of posterior parameter estimates for the best GLMs with different datasets, which include mean value, standard deviation and highest density region (3% and 97%).

Table 4. Numerical characteristics of posterior parameter estimates for exponential GLM and charges with normal distribution

Parameter	Mean	Std	HDI-3%	HDI-97%
Intercept	-1.986	0.147	-2.263	-1.722
Age	0.047	0.124	-0.192	0.283
Sex	-0.035	0.076	-0.181	0.097
BMI	0.071	0.131	-0.149	0.338
Region	-0.003	0.033	-0.066	0.056
MedHistory	0.001	0.047	-0.079	0.092
Exercise	-0.045	0.048	-0.131	0.038
WorkerStatus	0.000	0.046	-0.089	0.088

Table 5. Numerical characteristics of posterior parameter estimates for Laplace GLM and charges with normal distribution

Parameter	Mean	Std	HDI-3%	HDI-97%
b	0.082	0.003	0.077	0.088
Intercept	0.087	0.012	0.062	0.109
Age	0.019	0.012	-0.003	0.043
Sex	-0.012	0.007	-0.025	0.001
BMI	0.015	0.012	-0.007	0.038
Region	-0.000	0.003	-0.006	0.005
MedHistory	-0.007	0.004	-0.015	0.000
Exercise	-0.000	0.004	-0.009	0.007
WorkerStatus	0.003	0.004	-0.005	0.010

Table 6. Numerical characteristics of posterior parameter estimates for exponential GLM and charges with gamma distribution

Parameter	Mean	Std	HDI-3%	HDI-97%
Intercept	-2.975	0.157	-3.266	-2.659
Age	-0.091	0.136	-0.335	0.17
Sex	-0.062	0.077	-0.203	0.083
BMI	0.253	0.139	-0.014	0.498
Region	0.150	0.036	0.078	0.212
MedHistory	-0.001	0.049	-0.087	0.1
Exercise	0.094	0.047	0.008	0.185
WorkerStatus	-0.06	0.046	-0.142	0.027

Table 7. Numerical characteristics of posterior parameter estimates for exponential GLM for claim payments with Pareto distribution

Parameter	Mean	Std	HDI-3%	HDI-97%
Intercept	-1.034	0.152	-1.326	-0.774
Age	0.013	0.13	-0.218	0.271
Sex	0.069	0.075	-0.061	0.212
BMI	-0.174	0.135	-0.42	0.093
Region	0.067	0.037	-0.004	0.136
MedHistory	-0.019	0.047	-0.114	0.059
Exercise	-0.087	0.047	-0.17	0.002
WorkerStatus	0.038	0.046	-0.044	0.121

NUMERICAL EXPERIMENT WITH ACTUAL DATA

For this scenario the actual actuarial data of insurance company were applied for fitting GLM. Dataset was taken from Singapore Actuarial Society. All of the worker compensation insurance policies in this dataset have experienced an accident. The next features were used:

1. **Age**.
2. **Sex**.
3. **MaritalStatus:** categorical variable, which identifies marital status of clients.
4. **DependentChildren:** numerical variable, which shows number of dependent children.
5. **DependentOthers:** numerical variable, which shows number of dependent, excluding children.
6. **WeeklyWages:** numerical variable, which shows total weekly wage.
7. **PartFullTime:** categorical variable, which shows working mode.
8. **HoursWorkedPerWeek:** numerical variable, which shows total hours worked per week.
9. **DaysWorkedPerWeek:** numerical variable, which shows number of days worked per week.
10. **UltimateIncurredClaimCost:** numerical variable which shows total claims payments by the insurance company. This is target variable.

Results of fitting GLM from previous experiment are shown in Table 8.

Table 8. Results of GLM construction using actual insurance actuarial data

Metric	GLM Normal	GLM Exponential	GLM Laplace
Log-Likelihood	2.155	6.231	3.593
AIC	17.691	7.538	14.815
BIC	67.753	53.048	64.877

It can be observed that the exponential GLM with logarithmic link demonstrated best results among others for real dataset.

Results of forecasting for best GLM model for real dataset are shown in Fig. 5.

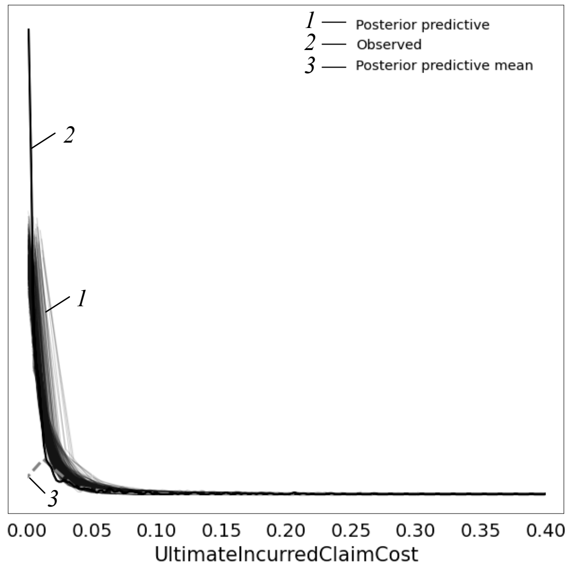

Fig. 5. Result of forecasting GLM with exponential distribution and log link function for actual actuarial insurance data

Table 9 show numerical summaries of posterior parameter estimates for best GLM.

Table 9. Numerical characteristics of posterior parameter estimates for exponential GLM and claim payments from real dataset

Parameter	Mean	Std	HDI-3%	HDI-97%
Intercept	-4.051	0.246	-4.485	-3.594
Age	0.934	0.191	0.577	1.266
Sex	-0.326	0.1	-0.515	-0.142
MaritalStatus	-0.346	0.064	-0.467	-0.229
DependentChildren	3.440	0.502	2.521	4.403
DependentOthers	-0.232	0.505	-1.064	0.776
WeeklyWages	4.722	0.459	3.882	5.582
PartFullTime	-0.217	0.199	-0.579	0.146
HourWorkedPerWeek	0.986	0.497	0.1	1.964
DaysWorkedPerWeek	-1.799	0.572	-2.864	-0.759

CONCLUSIONS

The application of GLM to the analysis of actuarial risks in the context of client claim payments is taken into consideration. For estimation parameters of models the MCMC method was implemented. The insurance indicators and the target variable were created artificially since actuarial insurance data is frequently not made public: age, sex, BMI, region, medical history, exercise, worker status and charges. The last one was generated by applying algorithm of mixture distribution, using normal, gamma and Pareto distribution with adding Gaussian noise, which had zero mean and variable standard deviation to create non-stationary process. Also real actuarial insurance data from Singapore Actuarial Society were used for experiments. Three GLM were implemented for experiments: normal with logarithmic link function, exponential with logarithmic link function and Laplace distribution with identity link function. Based on the experiment findings, it can be said that exponential GLM generally produced the best results for both artificial and real data. For the case of the normal distribution, Laplace GLM also produced positive results for artificial data.

In future studies it is planned to automatize the process of insurance data analysis using artificial intelligence and simulation techniques. As far as most of financial processes belong to the class of non-linear and non-stationary the methodology will be proposed for constructing such models. It is also planned to apply the methods of generating alternative managerial decision using Bayesian approach to data and expert estimates analysis.

REFERENCES

1. P. McCullagh, J. Nelder, *Generalized Linear Models*; 2nd edition. Chapman & Hall, 1989, 532 p.
2. D. Anderson et al., *A Practitioner's Guide to Generalized Linear Models – a foundation for theory, interpretation and application*; 3rd edition. Towers Watson, 2007, 122 p.
3. P. Gagnon, Y. Wang, "Robust heavy-tailed versions of generalized linear models with applications in actuarial science," *Computational Statistics & Data Analysis*, vol. 194, pp. 1–16, 2024. doi: 10.1016/j.csda.2024.107920
4. D.K. Lim et al., "Deeply Learned Generalized Linear Models with Missing Data," *Journal of Computational and Graphical Statistics*, vol. 33, no. 2, pp. 638–650, 2024. doi: 10.1080/10618600.2023.2276122

5. Y. Tian, Y. Feng, "Transfer learning under high-dimensional generalized linear models," *Journal of the American Statistical Association*, vol. 118, no. 544, pp. 2684–2697, 2023. doi: 10.1080/01621459.2022.2071278
6. Y. Havrylenko, J. Heger, "Detection of interacting variables for generalized linear models via neural networks," *European Actuarial Journal*, vol. 14, no. 551–580, 2024. doi: 10.1007/s13385-023-00362-4
7. R. Panibratov, P. Bidyuk, "Estimation of the parameters of generalized linear models in the analysis of actuarial risks," *System Research and Information Technologies*, no. 2, pp. 139–148, 2023. doi: 10.20535/SRIT.2308-8893.2023.2.10
8. L. Levenchuk, P. Bidyuk, O. Tymoshchuk, "Operational risk estimation using system analysis methodology," *System Research and Information Technologies*, no. 1, pp. 42–61, 2024. doi: 10.20535/SRIT.2308-8893.2024.1.04
9. C. Andrieu et al., "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, pp. 5–43, 2003. doi: 10.1023/A:1020281327116
10. C. Karras et al., "An overview of mcmc methods: From theory to applications," *Proceedings of international conference on artificial intelligence applications and innovations, IFIP, 2022, Crete, Greece, 17–20 June 2022*, pp. 319–332. Springer International Publishing. doi: 10.1007/978-3-031-08341-9_26
11. N. Alsadat et al, "Bayesian and non-Bayesian analysis with MCMC algorithm of stress-strength for a new two parameters lifetime model with applications," *AIP Advances*, vol. 13, no. 9, pp. 1–20, 2023. doi: 10.1063/5.0167295
12. H. Kavianihamedani, J.D. Quinn, J.D. Smith, "New Diagnostic Assessment of MCMC Algorithm Effectiveness, Efficiency, Reliability, and Controllability," *IEEE Access*, vol. 12, pp. 42385–42400, 2024. doi: 10.1109/ACCESS.2024.3378752
13. J. Zhang et al, "Improving simulation efficiency of MCMC for inverse modeling of hydrologic systems with a Kalman inspired proposal distribution," *Water Resources Research*, vol. 56, no. 3, pp. 1–24, 2020. doi: 10.1029/2019WR025474
14. A. Brown, G.L. Jones, "Convergence rates of Metropolis–Hastings algorithms," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 16, no. 5, pp. 1–15, 2024. doi: 10.1002/wics.70002

Received 09.01.2025

INFORMATION ON THE ARTICLE

Roman S. Panibratov, ORCID: 0000-0002-8604-4420, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: roman.panibratov@gmail.com

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pbidyuke_00@ukr.net

АНАЛІЗ АКТУАРНИХ РИЗИКІВ ЗА ДОПОМОГОЮ УЗАГАЛЬНЕНИХ ЛІНІЙНИХ МОДЕЛЕЙ / Р.С. Панібратов, П.І. Бідюк

Анотація. Розглянуто задачу побудови узагальнених лінійних моделей для аналізу актуарних ризиків із ситуацією виплат премій клієнтам. Для цього застосовано метод Монте-Карло для Марківських ланцюгів. Для дослідження розглянуто дві ситуації. У першій ситуації страхові показники та цільова змінна налаштовувалися випадковим чином через проблему вільного доступу до даних. Для створення трьох наборів даних виплати генерувалися за допомогою нормального, гамма та розподілу Парето зі змінною дисперсією та додаванням шуму для імітації нестаціонарного процесу. У другій ситуації використано реальні актуарні дані, взяті з Singapore Actuarial Society. Побудовано узагальнені лінійні моделі з нормальним розподілом із логарифмічною функцією зв'язку, експоненційним розподілом із логарифмічною функцією зв'язку і розподіл Лапласа з тотожною функцією зв'язку. За метриками якості побудови моделей зроблено висновки щодо їх структури.

Ключові слова: актуарний ризик, узагальнені лінійні моделі, імітаційне моделювання, експоненційна множина розподілів, Байєсівський аналіз даних, метод Монте-Карло для Марківських ланцюгів.