Available online at http://iorajournal.org/indx.php/orics /index

Operations Research: International Conference Series

e-ISSN: 2722-0974
p-ISSN: 2723-1739

Vol. 5, No. 2, pp. 38-45, 2024

Aggregate Loss Models to Calculate Risk Measures

Septi Rahmawati[1*], Andhita Zahira Adib[2], Volodymyr Rusyn[3]

[1,2]Department of Mathematics University of Padjadjaran, Sumedang, Indonesia
[3]Yuri Fedkovych Chernivtsi National University, Kotsyubynsky str., 58012 Chernivtsi, Ukraine

**Corresponding author email: septi20001@mail.unpad.ac.id*

Abstract

The concept of aggregate loss models pertains to a stochastic variable representing the total sum of all losses encountered within a set of insurance policies. In the non-life insurance sector, it is employed to assess the potential losses that an insurance company may face when claims made by policyholders exceed the allocated claim reserves. The purpose of studying aggregate loss models is to ascertain risk measures such as standard deviation of premium principles, value at risk (VaR), and conditional tail expectation (CTE). These steps aid insurance companies in the management and quantification of risks associated with aggregate losses. The standard deviation of premium principles is calculated analytically by substituting expected values and variances, while VaR is estimated using the Monte Carlo method to determine quantile values and confidence intervals. CTE is evaluated by computing the average losses that surpass the VaR threshold. These distributions and parameters require the Pareto distribution, which characterizes claim sizes, and the Poisson or Negative Binomial distribution, which factors in the number of claims. It is crucial to carefully consider the selection of the appropriate distribution, as it plays a significant role in determining the accuracy and reliability of the model. Furthermore, other influencing factors, such as loading factors and confidence intervals, should also be taken into account. These factors have the potential to significantly impact the quantification of risk arising from the model.

Keywords: Aggregate loss model, standard deviation premium principles, Value at Risk (VaR), Conditional Tail Expectation (CTE), Monte Carlo Method.

1. Introduction

Insurance companies face significant potential losses if claims filed by policyholders exceed the claim reserves estimated by the insurance company. This potential is interpreted as a risk that must be managed by the insurance company to avoid losses. Risk can be assumed as a random variable with claims having a distribution, so risk calculations are usually related to probability models, one of which is the aggregate loss model. The aggregate loss model is a random variable that represents the total of all losses that occur in an insurance policy block. The aggregate loss model can be modeled using a collective risk approach, where the number of claims is a discrete random variable, and the size of the claims is a continuous random variable. The number of claims used in this research is a random variable with Poisson and Negative Binomial distributions, while the size of the claims is modeled with random variables having Gamma, Pareto, and Exponential distributions. The aggregate loss model can be measured analytically using the standard risk measure, the standard deviation premium principle. Furthermore, numerical methods such as Monte Carlo simulations are employed to determine the Value at Risk and Conditional Tail Expectation risk measures.

Insurance companies operate in an environment filled with uncertainty, especially when facing the risk of claims that can significantly impact their operational sustainability. This risk becomes increasingly complex when companies must manage claims from policyholders that may exceed the budgeted reserve estimates. Therefore, risk management is key to maintaining financial stability and the continuity of insurance companies.

Risk in the insurance context can be considered a random variable, with claims as critical elements forming the risk distribution. A deep understanding of the nature of this risk distribution enables insurance companies to take proactive steps in planning and managing their risks. One approach used to measure claims risk is through aggregate loss models. The aggregate loss model is a mathematical representation of the total losses that may occur in a specific period or within an insurance policy portfolio. In this study, focus is given to two types of random variables: the number of claims represented by the Poisson and Negative Binomial distributions, and the size of claims represented by the Gamma, Pareto, and Exponential distributions.

To measure risk analytically, standard risk measures such as standard deviation and the premium principle are used. Additionally, a numerical approach is applied through Monte Carlo methods to calculate the Value at Risk (VaR) and Conditional Tail Expectation (CTE) risk measures. The combination of analytical analysis and numerical approaches provides a more comprehensive understanding of the potential losses that insurance companies may face. Thus, this research aims not only to investigate the inherent risks in insurance claims but also to provide a solid foundation for insurance companies to develop effective risk management strategies. With a deep understanding of claims risk, insurance companies can be better prepared to face emerging challenges and ensure the sustainability of their operations in a dynamic and uncertain market.

Aggregate loss models are commonly used in the insurance and financial industries to manage risks. Poisson-Tweedie distribution family for modeling loss frequency, which provides more flexibility and reduces the chance of model misspecification. The proposed model is applied to the Transportation Security Administration (TSA) claims data to demonstrate the modeling capacity of the Poisson-Tweedie distribution. Chen, Wang, & Kelly (2021) studied this and the result highlight the effectiveness and applicability of the Poisson-Tweedie distribution in modeling loss frequency and its potential for improving risk management and decision-making in the insurance industry

Risk measure of Value-at-Risk (VaR) has shown its performance and benefit in many applications, it is in fact not a coherent risk measure. Josaphat & Syuhada (2021) introduces a new risk measure called Dependent CoVaR (DCoVaR) that provides better forecast than existing measures like Modified CoVaR (MCoVaR) and Copula CoVaR (CCoVaR) for a target loss dependent on another random loss. Empirical studies on financial returns data show that DCoVaR, when used with Gumbel Copula, accurately describes the dependence structure of returns and outperforms other measures in comprehending the connection between bivariate losses and optimizing investment positions.

Septiany, Setiawaty, & Purnaba (2020) on their study focuses on using the Monte Carlo method to model the aggregate loss distribution in the context of health insurance claims. The researchers selected the Z12M-NBGE distribution for claim frequency and the lognormal distribution for claim severity based on goodness of fit tests. These distributions were then used to form the aggregate loss distribution using the Monte Carlo method. The simulation results were obtained for the measurement of Value at Risk (VaR) and Shortfall Expectations (ES). The Monte Carlo method was found to be simple to implement and capable of handling various risks with dependency.

2. Literature Review

2.1 Aggregate Loss Model

The model for aggregate loss, depicting the distribution of the total loss within a specific timeframe, serves as a foundation for operational choices, determination of insurance premiums, enhancement of reinsurance strategies, and the effective management of solvency and liquidity risks. Regulatory bodies, tasked with ensuring the financial stability of insurance firms, mandate that these entities maintain sufficient capital to safeguard against unforeseen or severe losses. Risk metrics based on percentiles, like value at risk (VaR) or expected shortfall (ES), originate from aggregate loss models and enable the computation of scenarios representing the most unfavorable situations (Chen, Wang, & Kelly, 2021).

According to Li, et.al. [4] the aggrate loss models is based on the interaction between risks described through their correlation coefficients. Let $X_1, X_{2,} \ldots, X_n$ be random variables representing n risks that are not independent, so there exist $\rho(X_1, X_2), \ldots, \rho(X_{n-1}, X_n)$. Then, the formed aggregate risk is

$$S = X_1 + X_2 + \cdots + X_N \tag{1}$$

2.2 Monte Carlo Method

The Monte Carlo method is a numerical approach to solving loss model problems that cannot be solved analytically. This method involves randomly sampling observations according to the required distribution. In finding aggregate losses, this method uses a simple algorithm with a loading factor proportional to the standard deviation. The greater the standard deviation of the loss distribution, the higher the risk, so this method is important for measuring risk appropriately. Thus, the Monte Carlo method is an effective tool in calculating aggregate losses and measuring the risks associated with them.

The Monte Carlo method generates simulated values for a probabilistic variable by employing a random number generator with a uniform distribution in the $[0, 1]$ interval. Additionally, it utilizes the cumulative distribution function linked to these stochastic variables. It's crucial to understand that the use of simulation techniques does not imply a decision optimization process. Instead, solving problems through simulation involves the utilization of interactive algorithms and following well-defined steps to reach objectives. The input data usually consists of random variables generated by a random number generator (Platon & Constantinescu, 2014).

2.3 Standard Deviation Premium Principles

Standard deviation premium principles are a type of premium principle used in optimal reinsurance models. In the study by Yichun (2011) the insurer seeks to minimize the value at risk (VaR) or the conditional value at risk (CVaR) of their total risk exposure. Then, adapt the equation of the standard deviation premium principle is defined as:

$$SD(X) = E(X) + g\sqrt{Var\ (X)} \quad (2)$$

2.4 Value at Risk (VaR)

Value at Risk (VaR) is one of widely-used risk measure and defined as a maximum loss that can be tolerated at level of confidence (Rohmawati & Syuhada, 2015). Their study said, risk measure prediction with VaR will be accurate, when the coverage probability equal to a given level of confidence α.

Carlo (2022) discussed that inancial losses arise from statistical analyses and the models and parameters employed in their computation. As a result, there are multiple approaches to calculate Value at Risk (VaR), with three prominent methods outlined:

a) Monte Carlo Simulation Method: This involves estimating VaR by generating numerous potential outcomes derived from the initial input data.
b) Historical Simulation Method: VaR is calculated by utilizing historical price data for each financial asset.
c) Analytical/Parametric Method - Delta-Gamma: This method estimates VaR using projected profitability data.

Definition 1 (Klugman, et.al., 2004)

If the random loss variable is denoted by X, then the value at risk of X at the $100\delta\delta\%$ level, denoted by VaR_δ or π_δ, is the 100δ percentile (or quantile) of the distribution X that satisfies

$$Var_\delta = P(X > \pi_\delta) = 1 - \delta$$

2.5 Conditional Tail Expectation (CTE)

Conditional Tail Expectation (CTE) is a measure used to calculate premiums and quantify the global risk of an insurer. It takes into account both underestimation and overestimation losses by using asymmetric loss functions. The premium is calculated as the quantity that minimizes an objective function related to the conditional tail expectation of the loss. This approach ensures that the premium is a coherent risk measure and helps practitioners assess the risk associated with the total claims amount. The methodology of using CTE is applied to model the risks of composite models and is compared with other risk measures such as Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) (Enrique, et.al., 2023).

Definition 2 (Klugman, et.al., 2004)

If X is denoted as the random loss variable, the conditional tail expectation of X at the $100\delta\%$ confidence level, denoted $CTE_\delta[X]$, is an estimate of losses exceeding the 100δ percentile (or quantile) of the distribution X.

$$CTE_\delta[X] = E[X|X > Var_\delta]$$

3. Materials and Methods

3.1. Materials

The data used in this study is simulation data generated with the help of the Python application. The data size for the simulation calculations is 100,000, then the distributions and parameters used in the calculations for various combinations of Poisson and Negative Binomial (NB) distributions with an average large claim amount budgeted by the insurance company of 50,000,000 for each distribution. Subsequently, common loading factors used by insurance companies, 1 and 2, as well as confidence levels of 95% and 99%, were applied.

3.2. Method

Analytical and numerical methods are used to measure risk, including the use of the standard deviation premium principle and the Monte Carlo method to determine risk measures such as Value at Risk and Conditional Tail Expectation.

The steps taken in this research are as follows:

a) Describes the properties of aggregate loss models, including the Probability Distribution Function (PDF), Cumulative Distribution Function (CDF), expectations, and variance.
b) Replacing the expectation and variance formulas of the aggregate loss model obtained in step (1) to calculate the risk measure with the *principle standard deviation method analytically*, by combining random variables such as the number of distributed claim $Poisson\ (\mu)$ and $Binomoal\ negatif\ (r,p)$ and the distributed $Gamma\ (\alpha,\beta), Peto\ (\tau,\omega),\ Exponential\ (\lambda)$.
c) Modifying the Monte Carlo method algorithm to calculate the *Value at Risk* and *Conditional Tail Expectation risk* measures from the aggregate loss model, with random variables in the form of the number of distributed $Poisson\ (\mu)$ and $Binomoal\ negatif\ (r,p)$ and the distributed $Gamma\ (\alpha,\beta), Peto\ (\tau,\omega),\ Exponential\ (\lambda)$.
d) Simulating the calculation of risk measures *Standard Deviation Premium Principle*, *Value at Risk*, and *Conditional Tail Expectation* for aggregate loss models with known parameters for each.
e) Explain the interpretation of the calculation simulation results in step (4).
f) Summarize the findings found.

4. Results and Discussion

4.1 Characteristics of Aggregate Loss Models

Aggregate loss represents the total losses that occur within a block of insurance policies. These losses consist of the number of claims, which will be modeled as a discrete random variable N, and the size of the claims, modeled as a continuous random variable, X. The assumption is made that the magnitudes of losses are identically and independently distributed (i.i.d) between each other. The aggregate loss model can be expressed as follows:

$$S = X_1 + X_2 + \cdots + X_N$$

its cumulative distribution function is defined as:

$$F_S(s) = \sum_{n=0}^{\infty} f_N(n) F_X^{*i}(x) \tag{3}$$

where $f_N(n)$ is the PDF of the random variable N, and $F_X^{*i}(x)$ is the convolution function for the total claim size $X_1 + X_2 + \cdots + X_N$, expressed as:

$$F_X^{*i}(x) = \int_0^x F_X^{*(i-1)}(s - x_{i-1}) f_{x_{i-1}}(x_{i-1}) dx_{i-1}, i = 2,3,\ldots,n$$

then, its expectation and variance are as follows:

$$E[S] = E\big[NE[X]\big] \tag{4}$$

$$Var[S] = E[N][Var(X)] + [E(X)]^2 Var[N] \tag{5}$$

4.2 Standard Deviation Premium Principle for Aggregate Loss Models

This risk measure utilizes the expectation and variance of the aggregate loss model by substituting Equations (4) and (5) into Equation (2), allowing for an analytical solution depending on the parameters of its distribution as follows:

$$SD(S) = E[N]E[X] + g\sqrt{E[N][Var(X)] + [E(X)]^2 Var[N]} \tag{6}$$

Table 1: Risk measurement of the standard deviation premium principle for several distribution combinations.

Model	N	X	$SD(S)$
1.	$N \sim Pois(\mu)$	$X \sim Gam(\alpha,\beta)$	$\beta(\mu\alpha + g\sqrt{\mu\alpha(1+\alpha)})$
2.		$X \sim Par(\tau,\omega)$	$\frac{\omega}{\tau-1}\left(\mu + g\sqrt{2\mu\left(\frac{\tau-1}{\tau-2}\right)}\right)$
3.		$X \sim Exp(\lambda)$	$\lambda(\mu + g\sqrt{2\mu})$
4.	$N \sim BN(r,p)$	$X \sim Gam(\alpha,\beta)$	$\frac{\beta}{p}(r\alpha + g\sqrt{r\alpha(p(1-\alpha)+\alpha)})$

5.	$X \sim Par(\tau, \omega)$	$\frac{\omega}{p(\tau-1)}\left(r + g\sqrt{\frac{r(p(\tau-1)+1}{(\tau-2)}}\right)$
6.	$X \sim Exp(\lambda)$	$\frac{\lambda}{p}(r + g\sqrt{r})$

4.3 Simulation Calculation of Aggregate Loss Model

The simulation involves determining the standard deviation premium principle, estimating VaR using a modified Monte Carlo algorithm for quantile value and confidence interval estimation, and predicting CTE. The simulation uses a data size of 100,000, and the distributions and parameters for various distribution combinations are detailed in Table 4.2. The average claim amount budgeted by the insurance company is 50,000,000 for each distribution. Additionally, loading factors of 1 and 2, along with confidence levels of 95% and 99%, are applied.

Table 2:

Model	N	X	g	δ
1.	$N \sim Pois(\mu)$	$X \sim Gam(\alpha, \beta)$	1 & 2	95% 99%
2.		$X \sim Par(\tau, \omega)$	1 & 2	95% 99%
3.		$X \sim Exp(\lambda)$	1 & 2	95% 99%
4.	$N \sim BN(r, p)$	$X \sim Gam(\alpha, \beta)$	1 & 2	95% 99%
5.		$X \sim Par(\tau, \omega)$	1 & 2	95% 99%
6.		$X \sim Exp(\lambda)$	1 & 2	95% 99%

4.4 Results of Simulation Calculation for Standard Deviation Premium Principle

The standard deviation premium principle risk measure is calculated by substituting each parameter from the specified distribution in Table 1 using the formula obtained in Table 1. Consequently, the standard deviation premium principle values for each aggregate loss model are obtained as follows.

Table 3: Results of Calculation Simulation for Standard Deviation Premium Principle

Model	g	$SD(S)$
1	1	12.260754100745366
	2	17.31859613098863
2	1	14.80118129091347
	2	20.88940267714721
3	1	7.110528807362238
	2	9.99599279027379
4	1	24.213891514483464
	2	34.396180837912105
5	1	27.921248560613154
	2	39.301701405512624
6	1	9.913321086513186
	2	13.999219387496998

The standard deviation value of the loss distribution is a metric that measures how far loss values are spread from their mean. A higher standard deviation indicates a higher level of variability in the loss distribution and greater risk or uncertainty in loss projections. Conversely, a lower standard deviation indicates a higher level of consistency or predictability in the loss distribution and lower risk or more stable loss projections.

The value of 12.260754100745366 indicates the extent of variation or variability in the aggregate losses generated by the simulation. The higher the SD value, the greater the expected fluctuation in losses. If SD is small, there is a tendency to approach the mean value, depicting high predictability in the results. On the other hand, a large SD is likely to experience significant gaps.

4.5 Simulation Results for Value at Risk (VaR)

The VaR risk measure is calculated by first generating random numbers from the distribution combinations in Table 3 using the Monte Carlo Method to obtain sample values for the aggregate loss model. Then, based on the VaR determination algorithm, the quantile value is estimated using the smoothed empirical estimate. After obtaining the estimated quantile value, the confidence interval for the actual quantile is also estimated, resulting in the VaR calculation for each aggregate loss model as follows.

Table 4: Results of Calculation Simulation for Value at Risk (VaR)

Model	δ	γ	$\hat{\pi}_\delta$	$s_a \leq \pi_\delta \leq s_b$
1	95%	95%	34.84772129268849	$29.81889567 \leq \pi_\delta \leq 35.37861684$
	99%	99%	53.38010078304078	$51.4459752 \leq \pi_\delta \leq 62.95636524]$
2	95%	95%	41.98941736721508	$39.35935221 \leq \pi_\delta \leq 50.401842883120544$
	99%	99%	88.5777810245933	$87.7373182858008 \leq \pi_\delta \leq 97.95955832210473$
3	95%	95%	19.500433399896	$11.97723257 \leq \pi_\delta \leq 24.53332945548397$
	99%	99%	43.27026862122348	$41.3294231920019 \leq \pi_\delta \leq 48.69032041961238$
4	95%	95%	68.651329659987	$65.74143231401906 \leq \pi_\delta \leq 82.94606537225745$
	99%	99%	145.7069534748949	$116.95452946 \leq \pi_\delta \leq 160.36572674860875$
5	95%	95%	79.25470603740997	$79.38077383306853 \leq \pi_\delta \leq 94.57440170340368$
	99%	99%	166.09805193212324	$124.07568940017002 \leq \pi_\delta \leq 184.27972841436363$
6	95%	95%	27.480290956126968	$25.378977659117357 \leq \pi_\delta \leq 34.09919074505118$
	99%	99%	60.21395990013581	$58.2938278720175 \leq \pi_\delta \leq 67.65240917040532$

This implies that with a confidence level of 95%, aggregate losses will not exceed 34.84. In other words, there is a 5% chance that losses will exceed this value. This interval provides an estimate of the range where the actual VaR may lie. With a confidence level of 95%, VaR is estimated to be between 29.82 and 35.38. This means that this interval encompasses values that may be generated by the simulation with a 95% confidence level. The smaller the VaR interval, the higher the confidence in the given VaR estimate.

4.6 Simulation Results for Conditional Tail Expectation (CTE)

The CTE risk measure is calculated by extending the calculation from VaR, with each aggregate loss sample for the distribution combinations in Table (4.1) obtained in the previous calculations. The average value of losses exceeding the estimated quantile value is estimated using Equation (7), resulting in the CTE value estimates as follows.

Table 5: Results of Calculation Simulation for Conditional Tail Expectation (CTE)

Model	δ	$\widehat{\ }_{\delta}$
1	95%	44.95115064382386
	99%	65.94605807611651
2	95%	54.07474341198557
	99%	79.25098364164518
3	95%	26.511452937704913
	99%	38.29929885882577
4	95%	88.33106426513393
	99%	130.09351943323776
5	95%	100.7900714517955

	99%	149.84450741350932
6	95%	36.879796202120566
	99%	53.16748534429482

CTE provides information about the magnitude of anticipated losses in the extreme part of the distribution, beyond the VaR value. The value of 44.95115064382386 indicates the average expectation of losses under certain confidence levels in extreme conditions. This means that the larger the CTE, the higher the expectation of losses.

5. Conclussion

Determining risk measures for the standard deviation premium principle, value at risk (VaR), and conditional tail expectation (CTE) in this study can be summarized as follows.

a) Analytically establishing the risk measure for the standard deviation premium principle can be accomplished as follows.

$$SD[S] = E[N]E[X] + g\sqrt{E[N][Var(X)] + [E(X)]^2 Var[N]}$$

b) The algorithm for computing VaR for the combined distribution of the aggregate loss model with the number of claims, namely $Poisson(\mu)$ and $NB(r,p)$, and the size of claims distributed as $Gamma(\alpha,\beta)$, $Pareto(\tau,\omega)$, $Exponential\ (\lambda)$ is as follows.
   (a) Defining k as the measure of data to be generated.
   (b) Generating the number of claims distributed $Poisson(\mu)$ and $NB(r,p)$.
   (c) Generating the size of claims distributed $Gamma(\alpha,\beta)$, $Pareto(\tau,\omega)$, $Exponential\ (\lambda)$.
   (d) Calculating $S = \sum_{i=1}^{n_v} X_{iv}$to obtain $s_1, s_2, \dots, s_k$.
   (e) Sorting $s_1, s_2, \dots, s_k$ from smallest to largest.
   (f) Calculating the estimated quantile value $\hat{\pi}_\delta = (1-h)s_j + hs_{j+1}$.
   (g) To estimate its confidence interval, first calculate $c = \Phi^{-1}\frac{1+\gamma}{2}k\sqrt{k\delta(1-\delta)}$.
   (h) Rounding c to the nearest integer.
   (i) Determining the confidence interval $[s_a \leq \pi_\delta \leq s_b]$ with $a = k\delta - c$ and $b = k\delta + c$ based on the data obtained in step (5).

c) For the CTE risk measure, continue the calculation from VaR, where sample loss data has been obtained using the Monte Carlo method from each combination distribution of the aggregate loss model with the number of claims, namely $Poisson(\mu)$ and $NB(r,p)$ and the size of claims distributed as $Gamma(\alpha,\beta)$, $Pareto(\tau,\omega)$, $Exponential\ (\lambda)$. Then, the CTE value can be estimated with $\widehat{\Omega}_\delta$ as follows,

$$\widehat{\Omega}_\delta = \frac{1}{k\delta(1-\delta)} \sum_{j=k\delta+1}^{k} \tilde{s}_j$$

where $\tilde{s}_j$ represents the ordered sample that exceeds $k\delta$. If the data size is large or tends to infinity, the estimated value of $\widehat{\Omega}_\delta$ will converge to the true value.

d) The method used in this study only employs a simple method, namely Monte Carlo, as the data used are simulation data. It is hoped that in future research, other methods such as the Fast Fourier Transform method or the Panjer Recursion method using the CDF determined in this study can be used. The expectation is to use real data to understand how the actual risk measure values are based on insurance company claim data.

References

Carlo, L. (2022). Value at Risk (VaR). *Financial And Economic Indicators*, 95-106.

Chen, S., Wang, Z., & Kelly, M. (2021). Aggregate loss model with Poisson-Tweedie frequency. *Big Data and Information Analytics*, 56-73.

Enrique, C.-O., Emilio, G.-D., Francisco, J., & Vázquez-Polo. (2023). Conditional Tail Expectation and Premium Calculation under Asymmetric Loss. *Axioms*, 496-496.

Josaphat, B. P., & Syuhada, K. (2021). Dependent conditional value-at-risk for aggregate risk models. *Heliyon*.

Klugman, S., Panjer, H., & Willmout, G. (2004). Loss Models From Data to Decision. *Society of Actuaries*.

Platon, V., & Constantinescu, A. (2014). Monte Carlo Method in risk analysis for investment. *Procedia Economics and Finance*, 393-400.

Rohmawati, A. A., & Syuhada, K. I. (2015). Value-at-Risk and Expected Shortfall Relationship. *International Journal of Applied Mathematics and Statistics*, 200-205.

Septiany, R., Setiawaty, B., & Purnaba, I. G. (2020). The Use of Monte Carlo Method to Model the Aggregate Loss Distribution. *Al-Jabar: Jurnal Pendidikan Matematika*, 179-190.

Yichun, C. (2011). Optimal Reinsurance under Variance Related Premium Principles. *Social Science Research Network*, 1-23.