International journal of applied management and economics
Vol : 02 , N° 08 , June 2024
ISSN : 2509-0720

Actuarial Risk Modeling: Case Of A Home Insurance Portfolio.

– **Author 1 :** DIOURI Youssef,
– **Author 2 :** ZAHI Jamal,

(1) : PhD Student, Laboratoire de Modélisation Mathématique et de Calcul Economique, Faculté d'Economie et de Gestion, Hassan 1st University, Settat, Morocco.

(2) : University Professor, Laboratoire de Modélisation Mathématique et de Calcul Economique, Faculté d'Economie et de Gestion, Hassan 1st University, Settat, Morocco.

Conflict of interest: The author reports no conflict of interest.

To quote this article: DIOURI .Y & ZAHI .J (2024) « Actuarial Risk Modeling: Case Of A Home Insurance Portfolio»,

IJAME : Volume 02, N° 08 | Pp: 132 – 151.

Submission date: May 2024
Publication date: June 2024

DOI : 10.5281/zenodo.11668166

Abstract:

Insurance is a major pillar of the economy, providing protection against unforeseen events and allowing both individuals and businesses to shield against potentially devastating financial losses. This study focuses on the home insurance product for a Moroccan insurance company, which offers comprehensive coverage for the dwelling and belongings against various risks such as fire, water damage, and theft. This type of insurance guarantees significant amounts, making the insurer's risk exposure very dangerous, especially in the event of claims ratio deviation. Therefore, the primary objective of this research is to assess the exact value of the actuarial risk associated with the home insurance portfolio to preserve its financial sustainability. In this study, and in order to evaluate the overall risk for the product in question, we focus on modeling the number of annual claims and the resulting financial burden, using various statistical modeling laws to quantify the overall risk that the insurer could face and therefore anticipate its commitments to policyholders. Additionally, we analyze the results obtained through the application of Monte Carlo simulation, which is one of the most effective methods for solving complex numerical problems. Therefore, we propose a risk quantification model based on two different methods which represents a decision making tool for insurance companies.

Keywords: Monte Carlo; Statistical Laws, Liability, Portfolio, Maximum Likelihood, Mutualization

1 Introduction

The primary goal of this study is to assess the actuarial risk linked to the home insurance portfolio of a Moroccan insurance company to ensure its financial sustainability.

In fact, the main objectives of this research are first to evaluate the actuarial risk associated with the portfolio and second, to provide insurers with a decision support tool for managing actuarial risk.

As commonly known, Insurance is a major pillar of the economy. It offers protection against unforeseen events, allowing both individuals and businesses to guard against potentially devastating financial losses (8). Insurance products cover a wide range of risks, from car accidents to natural disasters, illness, and liability.

Indeed, the insurer commits to financially cover the insured for a specific period in the event of a predetermined occurrence as per the contract (9), in exchange for receiving a sum known as the "insurance premium" (2). This premium is calculated based on the expected likelihood of the occurrence of the insured event. As actual occurrences of claims tend to fluctuate around a statistical average, the insurer aims to stay close to this average by subscribing to as many contracts as possible to compensate for the risks involved. This represents the principle of mutualization (3).

The array of random events that an insurance company commits to manage (such as the death of an individual or the impact of a fire on a property) can lead to an underestimation of the likely indemnity burden. This could result in significant financial consequences and weigh heavily on the insurer's solvency (3).

Therefore, the insurer must ensure that it holds sufficient reserves to meet its financial obligations to policyholders. Solvency is not only a regulatory requirement (1) to protect consumers but also a necessity to prevent the insurer from facing bankruptcy in the event of exceptional indemnity events.

This study focuses on the home insurance product, where the insured benefits from comprehensive insurance covering the dwelling and belongings against a range of risks (fire, water damage, theft, etc.). This type of insurance guarantees significant amounts, increasing the insurer's exposure to risk (3), especially in the case of claims ratio deviation.

Consequently, the main objective of this research is to evaluate the exact value of the actuarial risk associated with the home insurance portfolio for a Moroccan insurance company to preserve its financial sustainability.

Within the framework of this research, and in order to assess the overall risk for the concerned product, we will focus on modeling the number of annual claims and the resulting financial burden (2), utilizing various statistical modeling laws (8) to quantify the overall risk the insurer could face and anticipate its commitments to policyholders.

Additionally, we will compare the results obtained through the application of Monte Carlo simulation, which is one of the most effective methods for solving complex numerical problems (1).

This paper is organized according to the following outline: after the introduction, we present the research methodology and the theorical background of the research. Afterwards, we present the results of our study. We then discuss the findings before concluding and presenting the future works.

2 RESEARCH METHODOLOGY

2.1 RISK MANAGEMENT FOR AN INSURER

Risk management in the insurance industry is a crucial aspect aimed at assessing, quantifying, and managing the financial and statistical risks associated with various uncertain future events, as precise risk assessment is essential for informed decision-making and stakeholder protection. Insurers face several potential risks, including risks related to underestimating claims frequency, as well as market fluctuations such as changes in interest rates, stock prices, and asset prices. These fluctuations can impact the value of insurer investments and profitability. Insurers must accurately assess these risks to anticipate them and develop strategies to address them.

In this context, insurance companies use mathematical and statistical models to quantify risks. These models take into account a variety of factors, such as historical claims trends, external risk factors, and correlations between risks. Once risks are assessed, actuaries work to develop strategies to manage and mitigate these risks. This may include strategies such as portfolio diversification, purchasing reinsurance, and establishing financial reserves, which are funds insurers must set aside to ensure they have sufficient funds to meet their obligations to policyholders in the event of a claim. These reserves must be calculated prudently and in compliance with regulatory standards.

In the context of this study, we will focus on modeling risk related to claims assessment. Indeed, insurers calculate the likely claims frequency for each portfolio based on the fundamental concept of risk pooling, which reduces the financial impact of individual risks by sharing them among a large number of participants, thus ensuring fair and stable financial protection for all policyholders.

Below is how risk pooling works in insurance:

1. Basic principle: Each policyholder pays a premium to the insurer in exchange for the promise of compensation in the event of a claim. These premiums are pooled to form a common fund, from which payouts are made to policyholders who suffer losses.
2. Risk diversification: By pooling the risks of many policyholders, the insurer creates natural diversification. Losses suffered by some policyholders are offset by the premiums paid by all members of the group. Thus, individual losses are cushioned by the collective.
3. Equity and solidarity: Risk pooling is based on principles of fairness and solidarity. Policyholders contribute based on the level of risk they represent, but in the event of a claim, all benefit from equivalent financial protection, regardless of the extent of their contribution.
4. Financial stability: Risk pooling helps ensure the insurer's financial stability by distributing losses more evenly. This allows the insurer to better predict its obligations and maintain solvency even in the event of significant claims.
5. Size effect: The larger the insured group, the more effective risk pooling is. Insurers often seek to expand their policyholder base to benefit from this size effect and reduce the overall risk of their portfolio.

2.2 THE AGGREGATE CLAIM AMOUNT MODEL

We define the random variable S to be the total amount of claims arising from a risk in one year.

Let the random variable N denote the number of claims from the risk in this year.

And let the random variable X_i denote the amount of the i^{th} claim. The aggregate claim amount is just the sum of individual claim amounts. Therefore we can write (2) :

$$S = \sum_{i=1}^{N} X_i$$

We make two important assumptions:

First we assume that $\{Xi\}_{i\ =1,\dots,\infty}$ is a sequence of independent and identically distributed random variables.

Second, we assume that the random variable N is independent of $\{X_i\}_{i=1}^{\infty}$.

In this case, we have (1):

E(S) = E(X) * E(N)

2.3 MODEL DISTRIBUTION

The objective of this section is to build a probabilistic model to represent, the aggregate claim amount S **(2)**.

The model would require a component that modelled the number of claims and another that modelled the amounts of those claims **(2).**

As for frequency, the two most well-known models are the Poisson distribution and the negative binomial distribution. We will seek the most suitable model for our case through validation tests.

The Poisson distribution, used in the case study of the article, has the following probability:

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

The expectation is μ = λ.

The choice of the distribution for the claim amount is crucial in modeling the risk borne by the insurer. In actuarial science, to model the random behavior of severity, continuous distributions with support in R^+ are typically used.

As for severity, the most commonly used continuous distributions are: lognormal, exponential, Pareto, gamma, and Weibull. We will demonstrate in the practical part that the lognormal distribution is the most suitable.

In this case, if X follows a lognormal distribution $LN(\mu,\sigma^2)$, then $\ln(X)$ follows a normal distribution $N(\mu,\sigma^2)$, where μ = E(X). As for the probability density function, we have **(5):**

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

- $X > 0$ is the random variable,
- μ is the mean of the logarithmic variable distribution,
- σ is the standard deviation of the logarithmic variable distribution.

2.4 PARAMETERS ESTIMATION

For parameter estimation, we will focus on the most commonly used method, namely the maximum likelihood estimation method (1).

Let $X_{1,} \dots . X_n$ be n independent and identically distributed random variables with density function f (x,θ). We define the likelihood function as follows:

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

The maximum likelihood estimator is defined as the statistic $\hat{\theta} = T(X_1 \dots X_n)$ such that:

$$L(\theta^{MV}) = \mathrm{Max}_{\theta} L(\theta)$$

Therefore, we are looking for the value of θ based on the observations $(x_1 \dots x_n)$ that maximizes the probability of observing these observations. In the case where the likelihood function is continuous and twice differentiable with respect to the parameter θ, the maximum likelihood estimator θ^{MV} is the solution of the system:

$$\begin{cases} - \quad \left(\frac{\partial L}{\partial \theta}\right)_{\theta^{MV}} = 0 \\ - \quad \left(\frac{\partial^2 L}{\partial \theta^2}\right)_{\theta^{MV}} < 0 \end{cases}$$

2.5 GOODNESS OF A FIT TEST

It's a statistical test used to assess how well a given probability distribution fits a set of observations from a data sample.

In order to choose the distribution that best fits the data, we use hypothesis tests to make a decision in favor of one distribution over another.

2.5.1 THE FREQUENCY DISTRIBUTION CASE

To choose between the Poisson or Negative Binomial distribution, we compare the empirical frequency and variance of the data because if these two moments are equal, we immediately think of equidispersion, which is the main characteristic of a Poisson distribution.
In this regard, two hypothesis tests will be analyzed: the over-dispersion test based on the Fisher dispersion index statistic and the likelihood ratio test.

- ***Over dispersion Test***

The hypothesis tested $\mathbf{H_0} : \mathbf{E(N) = V(N)}$ **Versus** $\mathbf{V(N)} > \boldsymbol{E(N)}$
In other words, the presence or absence of over dispersion.
The statistic T to estimate the dispersion index, which is equal to $\frac{\mathbf{V(N)}}{\mathbf{E(N)}}$ is given by:

$$\mathbf{T} = \frac{\frac{(\mathbf{n}-\mathbf{1})\bar{\mathbf{S}}^{\mathbf{2}}}{\mathbf{n}}}{\bar{\mathbf{X}}}$$

Where $\bar{\mathbf{X}}$ and $\bar{\mathbf{S}}^{\mathbf{2}}$ are respectively the mean and the empirical variance of the frequency distribution N.
Under the hypothesis H_0, Hoel (1943) showed that the statistic nT, known as Fisher's dispersion index, is asymptotically distributed as a chi-squared distribution with n−1 degrees of freedom, where n is the number of observations of N.

- ***Likelihood Ratio Test***

The likelihood ratio test is used to test if the sample of N follows a given distribution against an alternative distribution. In this study, we are interested in testing the following hypothesis:
H_0 : N follows a Poisson distribution vs H_1 : N follows a negative binomial distribution.
The likelihood ratio statistic is given by: $\mathbf{T} = \mathbf{2}(\mathbf{ln}(\mathbf{L_1}) - \mathbf{ln}\ (\mathbf{L_0})$
Where:

- L_0: The likelihood of the Poisson distribution
- L_1: The likelihood of the negative binomial distribution

The statistic T asymptotically tends to follow a chi-squared distribution with one degree of freedom.

2.5.2 THE COST DISTRIBUTION CASE:

Since we are dealing with continuous distributions, we test the fit using the following tests: Kolmogorov-Smirnov and Anderson-Darling.

- ***Kolmogorov – Smirnov Test***

The Kolmogorov-Smirnov goodness-of-fit test is a non-parametric test based on the maximum distance between a theoretical cumulative distribution function and the empirical cumulative distribution function of the sample. It tests the null hypothesis H_0 that the observed data are generated by a theoretical probability distribution considered to be a suitable model. We consider a random variable X with cumulative distribution function F_x, which we want to compare to a continuous theoretical cumulative distribution function F_0. We then test:

$H_0: F_x(x) = F_0(x)\ \forall\ x$ Versus $\exists x\ F_x(x) \neq F_0(x)$

Let $\{\boldsymbol{X_1}, \boldsymbol{X_2} \ldots . \boldsymbol{X_n}\}$ be a sample of size N from X, the empirical cumulative distribution function F_n associated with this sample is:

$$\boldsymbol{F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}\,(X_i)}$$

$\boldsymbol{F_n(x)}$ is the proportion of observations with values less than or equal to x.

The difference between the observed values and the theoretical values deduced from the distribution function F_0 can therefore be measured by the random variable:

$$\boldsymbol{D_n = sup_{x \in x}|F_n(x) - F_0(x)|}$$

The statistic D_n asymptotically converges to a chi-squared distribution with n-1 degrees of freedom.

- ***Anderson-Darling Test***

The Anderson-Darling test is based on calculating a distance between the empirical distribution function and the fitted function defined as follows:

$$\boldsymbol{{A_n}^2 = n * \int_{-\infty}^{+\infty} [F_x(x) - F_0(x)]^2 * \frac{1}{F_0(x) * [1 - F_0(x)]} * dF(x)}$$

While both the Anderson-Darling test and the Kolmogorov test serve the same purpose, the difference between them lies in the fact that for the latter, only the maximum deviation between the empirical distribution and the fitted distribution matters, whereas the discrepancy indicator in the former test captures the entirety of the data as the sum of all deviations is involved. Therefore, it can be inferred that the Kolmogorov test is much more sensitive to the presence of outliers in a sample compared to the Anderson-Darling test.

2.6 Monte Carlo Simulation

Simulation techniques are highly useful when the complexity of the model under consideration makes any analytical approach extremely challenging. Once the random variables are generated according to a given distribution, the Monte Carlo method involves approximating E(X) by the empirical mean of simulations of the random variable X, following the law of large numbers.

$$\boldsymbol{\bar{X}} = \frac{\mathbf{1}}{\boldsymbol{Nombre - simulations}} \sum_{i=1}^{Nombre-simulations} \boldsymbol{x_i} \xrightarrow[Nombre-simulations \to \infty]{} \boldsymbol{E(X)}$$

Monte Carlo simulation is a computational technique used to estimate the probability of outcomes by repeatedly sampling random values for uncertain parameters and running simulations. It is named after the Monte Carlo Casino in Monaco, renowned for its games of chance, to illustrate the random nature of the method.

In the context of risk assessment and decision-making, Monte Carlo simulation allows analysts to model the uncertainty and variability of input parameters and assess their impact on the outcomes of interest. It is widely used in various fields such as finance, engineering, physics, and statistics.

Here are the general steps of a Monte Carlo simulation:

1. Defining Input Parameters: Identify the uncertain parameters and their probability distributions. These could include variables such as interest rates, market prices, or project durations.

2. Generatinng Random Samples: Randomly sample values for each input parameter from its probability distribution. This can be done using various methods, such as pseudorandom number generators.

3. Running Simulations: For each set of sampled values, run a simulation of the model or system being analyzed. This could involve solving equations, running a financial model, or simulating a physical process.

4. Aggregating Results: Collect the results of each simulation run, such as key performance metrics or outcomes of interest.

5. Analyzing Results: Analyze the aggregated results to understand the distribution of possible outcomes and assess key risk factors. This may involve calculating statistics such as mean, variance, percentiles, or probability density functions.

Monte Carlo simulation offers several advantages, including its flexibility to model complex systems, its ability to incorporate uncertainty and variability, and its capability to generate probabilistic insights. However, it also has limitations, such as computational intensity, the need for large sample sizes for accurate results, and the requirement for appropriate validation and interpretation of results.

Overall, Monte Carlo simulation is a powerful tool for decision-making under uncertainty, providing valuable insights into the likelihood and consequences of different scenarios.

3 RESULTS

The objective of this work is to assess the probable overall burden that the insurer must anticipate regarding the homeowner's insurance portfolio. This coincides with the funds that the insurer must hold to honor its future commitments without jeopardizing its financial stability.

This portfolio consists of 1,763 policyholders with properties to cover, with an average value of $140,000 each.

To quantify the risk, it is necessary to analyze past claims experience in order to model future claims. In this context, we will rely on a claims database that includes the following information: an 8-year historical record from 2015 to 2022, including the date of occurrence, the claim identifier, and the claim amount.

However, to ensure comparability of historical data across different years of occurrence, we will proceed with an operation called "As If" scenario setting, which involves:

- Updating the claim amount based on the inflation rate between the year of occurrence and the reference year, which is 2023. We consider the following annual inflation rates recorded in Morocco:

Table 1. Annual Inflation Rates

Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Inflation Rate (%)	1,6	1,6	0,7	1,9	0,2	0,7	1,4	6,6

Based on the inflation rates provided above, we revalue each claim X as follows:

$$X^{As\ if} = X * \prod_{K=j}^{n}(1 + r_k)$$

With the following notations:

- n :: The reference year - 1 (2022 in our case)
- r_k: The inflation index for year k
- J : The occurrence year

- Updating the number of annual claims by assuming a proportional relationship between the number of claims occurring annually and the number of policyholders per fiscal year, hence the following expression:

$$n_k{}^{As\ if} = n_{adh_{2023}} * \frac{n_k}{n_{adhe_k}}$$

With :

- n_k : Number of claims related to the yeark
- $n_k{}^{As\ if}$: Number of claims As if
- n_{adhe_k} : Number of policyholders of the year k

The table below represents the frequencies and the total revalued claims between 2015 and 2022:

Table 2. Frequencies and the total revalued claims between 2015 and 2022

Occurrence Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Number of annual claims	62	74	92	79	57	83	74	68

Now that we have constructed our As If frequency vector and As If claims, the next step is to search for a model that fits the data well.

3.2 Model estimation

In the following, we will use the collective risk model:

$$S = \sum_{i=1}^{N} X_i$$

We use the following notations:

- N: the number of annual claims.
- X_i: The As If cost of the i^{th} claim.
- S: The total annual cost.

Based on the theoretical concepts explained earlier, we will model the annual cost. To do this, we will start by modeling the cost of claim Xi and then the number N.

3.2.1 Cost law determination

In the Cost study, we test the adequacy of our data with a set of commonly used insurance laws, namely: Log Normal, Pareto, Gamma, Exponential, and Weibull. To find the law that fits the data well, we use the criterion of minimizing the Kolmogorov distance. We compile a table consisting of various laws where we determine the corresponding Kolmogorov distance. Here are the results obtained:

Table 3. Kolmogorov distance for different cost laws

Models	**Log Normal**	**Gamma**	**Exponential**	**Weibull**	**Pareto**
Kolomogorov Test	0,063	0,789	0,387	0,278	0,184

According to the results, we observe that the Log Normal distribution is the most suitable for modeling the cost of the claim as it minimizes the Kolmogorov distance.

As for the parameters of the Log Normal distribution estimated by the maximum likelihood method, we obtain μ =11.97 and σ =0.84.

Next, we will verify if the fit by the Log Normal distribution is indeed supported by statistical tests.

3.2.2 FREQUENCY DISTRIBUTION DETERMINATION

The most commonly used distributions to model the frequency variable N are the Poisson distribution and the Negative Binomial distribution. To choose between them, we opt for the dispersion test.

Table 4. Dispersion ratio test

Surdispersion	$\overline{X}$	$\overline{S}^2$	**IF**
Indice Fisher (IF)	73,63	128,27	12,19

The Fisher index is compared with the 95% quantile of a Chi-square distribution with 7 degrees of freedom, which is equal to 14.07.

We observe that IF is 12.19, which is lower than 14.07. Therefore, we reject the hypothesis of over dispersion of the data, and consequently, the Poisson distribution is the most suitable.

Parameter Estimation:

To estimate the parameters, we use the maximum likelihood method. We obtain:

$$\lambda = \mathbf{73,63}$$

3.2.3 Model evaluation

To validate the quality of the fit, we will use the Kolmogorov and Anderson Darling tests for the claim cost and the likelihood ratio test for the number of claims.

As for the claim cost, we obtain the following results:

Kolmogorov Test:

Table 5. Kolmogorov Test

Model	**Dn**	**C**	**P-value**
Log Normal	0,063	0,096	**0,063**

At the 5% significance level, we accept the following hypothesis: The claim cost follows a LN (11.97; 0.84).

Indeed, the p-value obtained is greater than 0.05.

Anderson Darling Test:

Table 6. Anderson Darling Test

	A	**P-Value**
Test	0,13	0,075

The Anderson Darling test also confirms the quality of the fit, as the p-value associated with each test is greater than 0.05.

Likelihood Ratio Test:

Regarding the number of claims, we confirm the choice of the model through the likelihood ratio test presented within the following table :

Table 7. Likelihood Ratio Test

	$\boldsymbol{ln(L_0)}$	$\boldsymbol{ln(L_1)}$	**RV**
Ratio Test	- 43,97	- 30,42	**27,1**

Using the likelihood ratio test, we can conclude that the Poisson distribution is accepted since LR is greater than the 95th percentile of a chi-squared distribution with 1 degree of freedom, which is equal to 3.84, thus confirming the quality of fit.

3.2.4 SELECTED MODEL AND RISK ASSESSMENT

At the end of the obtained results, the overall cost S is modeled as follows:

$$S = \sum_{i=1}^{N} X_i$$

With $N \sim \text{Pois}\ (73.63)$ and $X_i \sim \text{LN}\ (11.97\ ;\ 0.84)$

Thus, E(S) can be calculated as E(N) * E(X) based on the theoretical notions explained earlier:

E(N) = 73.63 and E(X) = $ 22,476.2.

Therefore:

E(S) = 10.65 million MAD.

3.2.5 CALCULATION OF THE OVERALL RISK THROUGH MONTE CARLO SIMULATION

Knowing the frequency and severity distributions, we can consider applying the simulation method.

To apply this method, we have constructed a simulation algorithm that we have implemented. The following paragraph will outline the steps of the algorithm.

To apply Monte Carlo, we need to simulate a large number of values of the random variable S. Since this variable is derived from a compound distribution, we will follow these steps:

1. Set the number of simulations J for our case J = 100,000.

2. Create a vector N containing J simulations following the same distribution as N: $N = (n_1 \ldots n_j \ldots . n_J)$ with n_j being the jth realization of the simulation according to the distribution N.

3. For j = 1 to J:
   - Simulate $n_j$variables $x_1{}^{n_j}, x_2{}^{n_j} \ldots \ldots x_{n_j}{}^{n_j}$ from the random variable LN(11.97; 0.84).
   - For $i = 1$ to n_j :

$$s_j = \sum_{i=1}^{n_j} x_i{}^{n_j}$$

4. Calculate the average of the vector $\{s_j\}_{1 \le j \le J}$, which corresponds to E (S).

By applying the algorithm, we deduce that E (S) is equal to 10.42 million MAD, which is almost equivalent to the previous method.

This demonstrates the strength of this technique in approximating complex mathematical expectations.

4 DISCUSSION OF THE FINDINGS

In our study, the main objective is to assess the exact value of the actuarial risk associated with the homeowners insurance portfolio. At the beginning, we first updated the historical data to ensure comparability across different years of occurrence. This involved adjusting the claim amount based on the inflation rate and updating the number of claims by applying proportionality with the number of policyholders.

Furthermore, we adopted a statistical approach to estimate the frequency and severity of claims. The results obtained from the Poisson distribution models for frequency and Log-Normal distribution for severity allowed us to gain a more detailed understanding of the behavior of claims and their potential impact on the portfolio.

Indeed, Poisson and Log-Normal models are widely recognized for their ability to model frequency and severity of claims, respectively, in various insurance contexts. The choice of these models was based on their historical fit with claims data, as evidenced by the goodness-of-fit tests we conducted.

Specifically, in modeling severity, we tested the adequacy of our data against a set of commonly used insurance laws, namely: Log Normal, Pareto, Gamma, Exponential, and Weibull. To identify the law that best fits the data, we used the criterion of minimizing the Kolmogorov distance. According to the results, we observed that the Log Normal law is the most suitable for modeling claim severity as it minimizes the Kolmogorov distance. This choice was approved by both the Kolmogorov and Anderson tests.

Moreover, for modeling the frequency law, the most commonly used laws are Poisson and Negative Binomial. To choose between them, we opted for the dispersion test, which rejected the hypothesis of data overdispersion, thus retaining the Poisson law. This choice will be validated by the likelihood ratio test.

Based on the results obtained, the total aggregate cost S is modeled as follows:

$$S = \sum_{i=1}^{N} X_i$$

With $N \sim Pois(73.63)$ and $X_i \sim LN\,(11.97\ ;\ 0.84)$

Thus, this parametric approach enabled us to quantify the risk by calculating E(S), which amounts to 10.65 million MAD.

Furthermore, we compareed this result with the outcome obtained through the application of Monte Carlo simulation, which is one of the most effective methods for solving complex numerical problems. To apply this method, we constructed a simulation algorithm that we implemented.

Moreover, by using Monte Carlo simulation, we were able to deepen our understanding of the risk by simulating various scenarios and examining the results to obtain a more precise estimation of the probable financial burdens. This method confirmed the estimation of the average total cost of claims evaluated by the initial approach.

The results of our modeling have direct implications for the insurer's risk management strategy. They enable the insurer to better establish adequate reserves to cover future claims. This ability to forecast potential losses is essential for maintaining the insurer's solvency and complying with regulatory requirements.

5 Conclusion

In this study, we assessed the actuarial risk associated with a portfolio of homeowner insurance by modeling the annual number of claims and the average cost. The Poisson and Log-Normal distributions were found to be the most suitable models for claim frequency and cost, respectively. These models were validated using various goodness-of-fit tests, confirming their adequacy with historical data.

Additionally, by applying Monte Carlo simulation, we were able to enhance our understanding of the risk by simulating various scenarios and examining the results to obtain a more precise estimate of probable financial burdens. This method confirmed the estimation of the average total cost of claims, evaluated at approximately 10.65 million MAD.

The ability to reliably predict risk and financial burdens is crucial for insurance portfolio management, enabling the insurance company to maintain solvency and safeguard against significant adverse events. This study demonstrates the importance of advanced statistical methods in assessing actuarial risk and making informed decisions in the insurance domain.

In terms of future work, we aim to analyze the correlation within the portfolio in order to study the impact on the overall risk using different methods. In this context, we will resort to statistical methods derived from copula theory. Additionally, we will compute the Value at Risk for simulating extreme scenarios to quantify the probable exposure in such scenarios.

References

(1) Chiu, C.-Y., Dai, T.-S., Lyuu, Y.-D., Liu, L.-C., & Chen, Y.-T. (2022). Option pricing with the control variate technique beyond Monte Carlo simulation. *The North American Journal of Economics and Finance*, *62*, 101772.

(2) Dickson, D. (2016). Insurance Risk and Ruin. Cambridge University Press.

(3) Denuit, M., & Charpentier, A. (2004). Mathématiques de l'assurance non-vie. Economica.

(4) Ellili, N., Nobanee, H., Alsaiari, L., Shanti, H., Hillebrand, B., Hassanain, N., & Elfout, L. (2023). The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature. *The Journal of Finance and Data Science*, *9*.

(5) Kanji, Gopal K. (2006). *100 Statistical Tests* (3rd ed.). SAGE Publications.

(6) Kaas, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). Modern actuarial risk theory. Springer.

(7) Lefebvre, M. (2011). Probabilités, statistiques et applications. Presses Internationales Polytechnique.

(8) Melnikov, A. (2003). Risk analysis in finance and insurance. Taylor & Francis.

(9) Mun, J. (2010). Modeling Risk: Applying Monte Carlo Risk Simulation, Strategic Real Options, Stochastic Forecasting, and Portfolio Optimization (2nd ed.). John Wiley & Sons.

(10) Olofsson, P. (2005). Probability, Statistics, and Stochastic Processes. John Wiley & Sons.

(11) Peleckienė, V., & Peleckis, K. (2014). Solvency II Assumptions for Increasing the International Competitiveness of EU Insurance Industry. *Procedia - Social and Behavioral Sciences*, *110*, 822-831.