A Comparative Analysis of Stochastic Approaches for Claims Reserving in Private Health Insurance

ESMERALDA BRATI, ALMA BRAIMLLARI
Department of Statistics and Applied Informatics,
Faculty of Economy,
University of Tirana,
ALBANIA

Abstract: - To guarantee the fulfillment of all claims, an insurance company must allocate enough funds to cover both current and future claims for active policies. The application of stochastic models has found extensive use in various domains of insurance and finance. However, their application in the context of private health insurance has been somewhat limited. To address this gap in existing knowledge, this paper aims to explore the application of stochastic methods to disease portfolios. The study involves dividing the developmental periods into semi-annual intervals and determining the most appropriate model for forecasting claim reserves. The research objectives include evaluating the effectiveness of the Mack, Clark LDF, and Clark Cape Cod methods in predicting claim reserves within a disease portfolio. Moreover, the study intends to compare these models to identify the most appropriate method for claim reserve calculations. The data source employed for this study comprises private health insurance claims data covering the period from 2018 to 2022. When compared to alternative methods, The Clark LDF method with Weibull distribution is far more accurate in calculating the reserve of claims compared with other similar methods. From the pool of models analyzed, the Mack model and Clark LDF model with Weibull distribution show the lowest reserve claims and standard errors. Based on these result analyses, the insurance company is recommended to explore other strategies for its fund distribution as opposed to maintaining only one substantial claim reserve.

Key-Words: - LDF, Mack, claims, reserves, Cape Cod, PHI.

Received: April 13, 2024. Revised: September 14, 2024. Accepted: November 9, 2024. Available online: December 20, 2024.

1 Introduction

The physical and mental health consequences of air pollution must not be overlooked as they potentially have long-term harmful effects on human health which translate into healthcare costs and consequently lower life expectancy for the population. The rise in public concern about health among the population is a direct consequence of the increasing awareness of eco-environmental protection, [1]. Private health insurance (PHI) contributes significantly to healthcare costs worldwide. Due to the failure of markets and public policy deficiencies, the significant impact of the health system's performance is crucial. In countries with low incomes, private health insurance can serve as a transitional mechanism, aiding in the increase of prepaid income, [2]. The utilization of PHI is positively associated with financial development as well as social, cultural, and economic factors including education, income, and individualism. This implies that higher levels of income, education, and individualistic values, are associated with increased PHI consumption, [3].

Private health insurance is medical coverage that an individual purchases directly. This type of coverage can be obtained solely for the individual or can also include their family members and dependents. Depending on your policy, private health insurance can provide financial coverage for treatment in public or private hospitals as a private patient. In addition, PHI can provide faster access to certain hospital services by avoiding long waiting times in public hospitals. Individuals, who are both taxpayers and consumers of health services, play an important role in financing health services. Consequently, any escalation in their expenses will result in increased payments, subsequently reducing their quality of life. Since 2020, the COVID-19 pandemic has increased the number of healthcare protection and sickness preventive measures, [4], [5].

Within the framework of the Solvency II directive, the one-year reserve risk is defined as the risk of having insufficient loss reserves at the end of a one-year period, which is significant and different from the way insurance companies typically analyze the ultimate view. A claim event is the term used to describe an occurrence that prompts the insured to file a claim. To guarantee the fulfillment of all claims, an insurance company must allocate enough funds to cover both current and future claims for active policies. The insurance company is obligated to create a claim reserve to account for forthcoming policy claims, [6]. The investigation of claim reserves in general insurance is a highly researched topic in actuarial studies. Stochastic models have been widely adopted in numerous areas of insurance and finance, but their utilization in the realm of private health insurance has been relatively constrained.

The gap that exists in comprehensive research on stochastic methods in private health insurance can be attributed to various factors. Firstly, private health insurance is classified as a short-tail portfolio, implying that claims are typically resolved within a brief period. Secondly, the regulations and ethical guidelines that govern private health insurance companies can present difficulties in accessing and utilizing the data necessary for stochastic modeling. To bridge this gap in the literature, I employed stochastic methods in my research on disease portfolios, dividing the developmental periods into semi-annual intervals. This approach differs from previous studies that have applied these methods to annual development periods in other portfolios.

This study aims to explore the utilization of stochastic methods on claim reserves in PHI and to identify the most suitable model that can be employed for forecasting claim reserves.

The objectives of this research are as follows:

1. To evaluate the effectiveness of the Mack, Clark LDF, and Clark Cape Cod methods in calculating claim reserves within the disease portfolio.
2. To compare the different methods and determine which one is more suitable for calculating claim reserves.

The paper follows the subsequent structure: Section 2 provides a comprehensive analysis, with a particular emphasis on a recent study that explores the use of stochastic methods. The methodology utilized in this research, which encompasses Mack Chain Ladder and Clark's methods, is presented in Section 3. The results obtained from applying these methods to real data are discussed in Section 4. Lastly, Sections 5 and 6 conclude the paper by offering some final remarks.

2 Literature Review

Considering the implementation of the IFRS 17 (International Financial Reporting Standard) regulatory regime in Europe on January 1, 2023, insurers have been actively promoting the adoption of stochastic approaches to ensure precise forecasting of insurance reserves. Under the provisions of IFRS 17, companies are authorized to provide information regarding the effects of fluctuations in discount rates and other financial variables, which impact their profit. This flexible approach ensures the continuous evaluation of both the insurance contract portfolio and the assets that provide support to it, [7]. In the research, [8], is presented a comprehensive overview of diverse stochastic methods for loss reserving, which have already been developed. In comparison to deterministic methods, stochastic methods possess the advantage of being grounded in explicit statistical assumptions. Mack was the first to provide a thorough stochastic model for the chain-ladder technique. In the year 1993, his publication on claims reserving gained significant recognition for its in-depth analysis of the chain ladder model, specifically focusing on the precise calculation of the standard error.

The Mack model, first proposed by [9] and known as the Distribution-Free Chain-Ladder stochastic model, has been extensively study by various researchers. The research conducted in [10], compared different stochastic methods for claims reserving in the MTPL portfolio. It was found that the over-dispersed Poisson (ODP) method with bootstrapping was a more prudent choice compared to the Mack method. This decision was primarily influenced by the ODP method's ability to offer a broader confidence interval. In [11] authors emphasize the utilization of different stochastic models, such as distribution-free models, probability distribution-based models (including Normal, Poisson, Gamma, and Inverse Gaussian distributions), and the combination of these models with bootstrapping techniques to calculate reserves for life and non-life insurance datasets. The study concluded that Mack's model and the Gamma probability distribution-based model were determined to be the optimal choices for estimating reserves among the distribution-based methodologies. The research paper [12], evaluated different stochastic models, such as the Poisson, gamma, and log-normal models, through the analysis of real-life insurance data. The study emphasized the impact of model and residual choices on reserve estimates and prediction errors. The study [13], shows how the Mack Chain Ladder and its bootstrap projections may be used to estimate or anticipate reserves using actual non-life insurance data. The findings from this study indicate that Ultimate reserves obtained through the bootstrap technique closely align with those derived from the Mack model, but the prediction errors associated with the bootstrap technique are higher compared to those of the Mack model. The study [14], emphasizes the flexibility of Mack's distribution-free chain ladder model and its estimator in accurately predicting the ultimate claim amount. Through their research, the study [15], enhances Mack's model and introduces a fundamental framework. This framework enables the analysis of the efficiency of chain ladder forecasts for overall claims at the conclusion of the initial non-observable calendar year. The Cape Cod method, initially proposed by [16] and later extended by [17], presents an alternative stochastic model for estimating total loss reserves through growth curve modelling. The research conducted by [18], involved examination of the distribution of claims within the DMTPL portfolio through the utilization of both the Cape Cod and LDF techniques. The findings of the study indicated that the Cape Cod model exhibited lower variance and estimation error in comparison to the LDF method. The study [19], employed the Cape Cod method for estimating IBNR (Incurred But Not Reported) reserves of loss in non-life insurance. Estimation is done by fitting Weibull and log-logistic growth curve models. The research [20], also extends the analysis by adding various stochastic models to the Cape Cod model. These models are constructed to generate parameter estimates that approximate those obtained from the classical Cape Cod algorithm using maximum likelihood estimation. In the study [21], workers' compensation claims over 11 years were collected employing the Clark Cape Cod and Clark LDF models, which incorporate Weibull and log-logistic distribution functions. As demonstrated by its results, the Clark LDF model with log-logistic distribution performs better compared to other analyzed models within the data set since it is accurate and has less variability. The focus of the research [22], lies in examining different types of stochastic models, such as distribution-free models, probability distribution-based models, and their combinations with bootstrapping techniques when estimating reserves for both life and non-life insurance portfolios. The study concludes that methods incorporating bootstrapping demonstrate superior goodness-of-fit when compared to other models.

3 Materials and Methods

The data employed for this study comprises private health claims data obtained from an insurance company operating in Albania. A comprehensive analysis of 17,575 paid claims from the sickness portfolio covering the years 2018 to 2022 was done. The values are expressed in Albanian lek (ALL), and the cumulative losses are shown as a development triangle (run-off triangle). Each row in Table 1 corresponds to a specific accident year and each column represents a particular development year. The upper left triangle displays the observed data, while the lower right triangle encompasses the data for the future triangle that is yet to be estimated. Let $C_{i,j}$ be a random variable that represents the amount of the cumulative claim that has been incurred in period i and has been resolved after a specific delay period j. Development factors $\hat{\lambda}_j$ play an important role in predicting future claims. These factors are defined by the formula (1).

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j}}{\sum_{i=1}^{n-i} C_{i,j-1}} \quad (1)$$

Table 1. Run-off triangle

Accident year i	Development year j 1	2	3	…	J
1					
2					
3	(observations)				
⋮					(to be
I					predicted)

Source: Author

The outstanding claims liabilities for accident year i, denoted as Ri, where i e $\{1, \dots, I\}$, is given by formula (2).

$$R_i = C_{i,I} - C_{i,I-i+1} \quad (2)$$

Let R represent the total outstanding claims liabilities for accident years as defined by formula (3)

$$R = \sum_{i=1}^{I} R_i \quad (3)$$

The prediction of outstanding loss liabilities is referred to as claims reserves. Hence, we denote the claims reserves for accident year i as R_i, where $\hat{R}_i = \hat{C}_{i,I} - C_{i,I-i+1}$, $i \in \{1, \dots, I\}$ is used to denote the prediction of these reserves, and the total claims reserves is represented as $\hat{R} = \sum_{i=1}^{I} \hat{R}_i$. It is important to note that $\hat{C}_{i,I}$ serves as a predictor for $C_{i,I}$. Utilizing Mack and Clark's methodology, an assessment and analysis of the fluctuations in these reserves will be conducted.

3.1 Mack Chain Ladder Method

The Mack model, also known as the Distribution-Free Chain-Ladder stochastic model, was introduced by, [9], to forecast claims reserves in the insurance industry. The method can predict the standard errors of the chain-ladder estimation without assuming any distribution if it meets three conditions:

1. The expectation assumption, $E[\frac{C_{i,j+1}}{C_{i,k}}|C_{i,1}, C_{i,2}\ldots, C_{i,j}] = f_j$
   The coefficients f_j are commonly referred to as loss development factors (LDF) or age-to-age factors.
2. Variance assumption,
   $$Var[\frac{C_{i,j+1}}{C_{i,k}}|C_{i,1}, C_{i,2}\ldots, C_{i,j}] = \frac{\sigma_j^2}{w_{i,j}C_{i,j}^{\alpha}} \quad (4)$$
   There exist certain constants $\sigma_j^2 > 0$ from the formula 4, identified as the variance parameters for all values of $1 \leq i \leq I$ and $1 \leq j \leq I\text{-}1$, the weights $w_{i,j} \in [0,1], \alpha \in \{0,1,2\}$.
3. Independence assumption, $\{C_{i,1}, C_{i,2}\ldots, C_{i,n}\}$, $\{C_{j,1}, C_{j,2}\ldots, C_{j,n}\}$ are independent for $i \neq j$,

3.1.1 Conditional MSEP Estimators in the Mack Model

Mack method demonstrates that unbiased estimates for claims reserves can be obtained if these assumptions are upheld. The metric used to express the uncertainty and variability related to the projected future reserves is the mean squared error of prediction (MSEP), [11]. It is expressed by the formula (5), (6), (7) and (8) as: Prediction variance ≈Process variance +Estimation variance.

$$E[(C_{i,j} - \hat{C}_{i,j})^2 \approx E[(C_{i,j} - E(C_{i,j}))^2 + E\left[(\hat{C}_{i,j} - E(\hat{C}_{i,j})^2\right] \quad (5)$$

$$MSEP(\hat{C}_{i,j}) \approx Var(C_{i,j}) + Var(\hat{C}_{i,j}) \quad (6)$$

$$MSEP(\hat{R}_i) \approx MSEP(\textstyle\sum_{j\in\delta_i} \hat{C}_{i,j} \quad (7)$$

$$MSEP(\hat{R}) \approx MSEP(\textstyle\sum_{i=2}^{n} \hat{R}_i \quad (8)$$

This model estimates the mean squared prediction error (MSEP) of reserves (R_i) by using the mean and standard deviation of incremental claims $C_{i,j}$. The formula for calculating the expected cumulative reserves is $\widehat{D_{i,j}} = \sum_{l=n-i+2}^{n-i+j} \hat{C}_{i,l}$, where $2 \leq k \leq n$ and $\delta_i = n - i + 2, n - i + 3, \ldots, n$ represents the remaining development periods. The formula (9) is used to estimate the MSEP of reserves. $MSEP(\hat{R}) = MSEP(\sum_{i=2}^{n} \widehat{R_{i,j}}) = \sum_{i=2}^{n}[MSEP(\hat{R}_i)] + \hat{D}_{i,n}^2(\sum_{q=i+1}^{n} D_{q,n}) \cdot \sum_{j=n-i+1}^{n-1} \frac{2\hat{\sigma}_{j+1}^2}{\lambda_{j+1}^2 \sum_{q=1}^{n-1} D_{q,j}}$ (9)

3.1.2 Estimating the Mack Model's Parameters

The development factors are estimated by the formula (10) for $1 \leq j \leq I\text{-}1$

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} w_{i,j}C_{i,j}^{\alpha}}{\sum_{i=1}^{I-j} C_{i,j}} \cdot \hat{\lambda}_j \quad (10)$$

The formula (11) presented below outlines the calculation method for determining the standard deviation of the development period.

$$\hat{\sigma}_j^2 = \frac{1}{n-j}\sum_{i=1}^{n-j+1}(f_{i,j} - \hat{f}_j)^2 \quad (11)$$

3.1.3 Properties of the MCL Model Estimators

The properties of Mack model estimators are presented in the study, [13], as follows:

1. The $\hat{f}_j$ estimates provided in (10) are uncorrelated and unbiased.
2. The estimator $\hat{f}_j$ has the lowest variance wt6ygcompared to all other unbiased estimators of f_j that are weighted averages of the observed development factors.
3. The unbiased estimator of the parameter σ_j^2 is the estimator $\hat{\sigma}_j^2$
4. Under the model assumptions 1 and 3, the estimator $\hat{C}_{i,I}$ is an unbiased estimator of $E(C_{i,I} \epsilon D_I)$
5. For both the ultimate claims amount $\hat{C}_{i,j}$ and the true ultimate claims amount $C_{i,j}$, the expected values of the estimators are equal, $E(\hat{C}_{i,I}) = E(C_{i,I})$, where $2 \leq i \leq I$

3.2 Clark's Methods

The study conducted by [16], proposed two models, namely the Loss Developed Factor (LDF) model and the Cape Cod model. The LDF method operates under the assumption that the ultimate loss amount in each accident year `is not influenced by the losses incurred in other years. The Cape Cod method, on the other hand, assumes that there is a known correlation between the ultimate loss amount and the expected loss for each year during the historical period. This correlation is determined by an exposure base, which is typically measured using a level premium.

$$G(x) = \frac{1}{LDF_x} \quad (12)$$

The formula (12) represents the cumulative percentage reported (or paid) up to time x. It is significant to note that the time interval between the "average" accident date and the evaluation date is indicated by the time index "x". Our model estimation process begins with the evaluation of the emerging loss pattern. To analyze this pattern, we will utilize two different curves, namely the Weibull curve and the Log-logistic curve. The scale parameter θ and the shape parameter ω can be used to parameterize each of these curves. The form of the log-logistic curve is given by formula (13) as follow:

$$G(x|\omega,\theta) = \frac{x^{\omega}}{x^{\omega}+\theta^{\omega}} \quad (13)$$

The form of the Weibull curve is given by formula (14) as follow:

$$G(x|\omega,\theta) = 1-\exp\left(-\left(\frac{x}{\theta}\right)^{\omega}\right) \quad (14)$$

It is shown below how the expected rise in loss value, $\hat{\mu}_{i,j}$ is calculated for each of the two models.

3.2.1 "LDF" Method

The formula (15) provides an estimation of the expected loss by considering both the growth curve and the estimated ultimate loss for each accident year. To perform this calculation, n + 2 parameters need to be estimated, including n accident years, ω, and θ.

$$\hat{\mu}_{i,j} = Ult_i \cdot [G(j|\omega,\theta) - G(j-1|\omega,\theta)] \quad (15)$$

3.2.2 Cape Cod Method

To estimate the expected loss shown in the formula (16), three parameters need to be determined: ELR, ω, θ, considering the expected loss ratio (ELR) and the growth curve.

$$\hat{\mu}_{i,j} = Premium_i \cdot ELR \cdot [G(j|\omega,\theta) - G(j-1|\omega,\theta)] \quad (16)$$

Since we are dealing with data that is condensed into annual blocks as a development triangle, the model will have a limited number of data points (one for each "cell" in the triangle). However, the LDF method poses a significant challenge due to over-parameterization. Once the model for the expected loss emergence has been determined, it becomes essential to calculate the variance in the formula (17).

$$\frac{Variance}{Mean} = \sigma^2 \approx \frac{1}{n-p} \cdot \sum_{i,j}^{n} \frac{(c_{i,j}-\hat{\mu}_{i,j})^2}{\hat{\mu}_{i,j}} \quad (17)$$

where, p denotes the number of parameters, $c_{i,j}$ represents the actual incremental loss, and $\hat{\mu}_{i,j}$ represents the expected incremental loss.

3.2.3 Process Variance

The variance can be estimated into two components: process variance, which represents the random fluctuations, and parameter variance, which accounts for the uncertainty in our estimation. The predicted pattern of loss emergence is represented by the curve $G(x|\omega,\theta)$. The distribution of the actual loss emergence will revolve around this expectation. While the Cape Cod method may demonstrate a slightly higher estimated process variance, it tends to generate a significantly smaller estimation error. It is important to note that in order to handle claims, it is desirable to have more accuracy and less variability in the reserve estimate. The first step towards evaluating the variance of the parameters entails establishing the maximum likelihood function to estimate the best parameters.

$$\text{Likelihood}=\prod_i \Pr(c_i) = \prod_i \frac{\lambda_i^{\frac{c_i}{\sigma^2}} \cdot e^{-\lambda_i}}{\left(\frac{c_i}{\sigma^2}\right)!} = \prod_i \frac{\left(\frac{\mu_i}{\sigma^2}\right)^{\frac{c_i}{\sigma^2}} \cdot e^{-\frac{\mu_i}{\sigma^2}}}{\left(\frac{c_i}{\sigma^2}\right)!} \quad (18)$$

$$\text{LogLikelihood}=\sum_i \frac{c_i}{\sigma^2} \cdot \ln\left(\frac{\mu_i}{\sigma^2}\right) - \frac{\mu_i}{\sigma^2} - \ln\left(\left(\frac{c_i}{\sigma^2}\right)!\right) \quad (19)$$

$$l = \sum_i c_i \cdot \ln(\mu_i) - \mu_i, \text{ assuming, } \sigma^2 \text{ is known} \quad (20)$$

The loglikelihood function for the "LDF" model is given by formulas (21), (22), (23) as follows:

$$l = \sum_{i,j}\big(c_{i,j} \cdot \ln(Ult_i \cdot [G(j) - G(j-i)]\big) - Ult_i[G(j) - G(j-i)]) \quad (21)$$

$$\frac{\delta l}{\delta Ult_i} = \sum_i\left(\frac{c_{i,j}}{Ult_i} - [\,G(j) - G(j-i)]\right) \quad (22)$$

$$\text{For } \frac{\delta l}{\delta Ult_i} = 0,\ \widehat{Ult_i} = \frac{\sum_j c_{i,j}}{\sum_j [G(j)-G(j-i)]} \quad (23)$$

The loglikelihood function for cape Cod model is given by formulas (24), (25), (26) as follows:

$$l = \sum_{i,j}\big(c_{i,j} \cdot \ln(ELR \cdot P_i[G(j) - G(j-i)]\big) - ELR \cdot P_i[G(j) - G(j-i)]) \quad (24)$$

$$\frac{\delta l}{\delta ELR} = \sum_{i,j}\left(\frac{c_{i,j}}{ELR} - P_i[\,G(j) - G(j-i)]\right) \quad (25)$$

$$\text{For } \frac{\delta l}{\delta ELR} = 0,\ \widehat{ELR} = \frac{\sum_{i,j} c_{i,j}}{\sum_{i,j} P_i \cdot [G(j)-G(j-i)]} \quad (26)$$

The actual loss during the development period j and accident year i is denoted by $c_{i,j}$ and the premium for accident year i is denoted by P_i. Several authors have studied the methodology of the Clark LDF and Clark Cape Cod models, including, [19], and [21]. The calculations were performed using the R-Studio software ChainLadder package (R Development Core Team, 2023), [23].

4 Results

4.1 Fitting Distributions

The cumulative number of claims within the disease portfolio during the period i, which have subsequently been resolved by period j is illustrated in Table 2 (Appendix). The standard form of the development triangle is obtained by converting the occurrence period {2018, 2019, ...2022} into semi-annual accidents {2018S1, 2018S2, ...2022S1, 2022S2}.

To estimate the Clark LDF and Clark Cape Cod models, we employed the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) to analyze our data and determine the most suitable curve between the Weibull and Loglogistic distribution functions. The outcomes obtained through the utilization of the (fitdistrplus) package in R programming are displayed in Table 3 (Appendix). Through a comprehensive evaluation of the criteria (AIC and BIC), it is evident that the Weibull distribution is the optimal choice for modeling our data.

4.2 Results of Mack method

Verification of the Mack assumptions is illustrated in Figure 1 through the chain ladder diagnostic plot. Upon analysis, it is evident that none of the four residual plots display any significant trends. As a result, it can be concluded that Mack's supposition holds. Moreover, the chain ladder's development during the origin period demonstrates a similar trend from development period 1 to 10.

In Table 4 (Appendix), the second column presents the current values, with the third column indicating the percentage that these current values represent in comparison to the ultimate values. The ultimate value and future value are presented in the fourth and fifth columns, respectively. The overall ultimate value reaches a total of 306,589,764 ALL, whereas the projected reserve for the future stands at 10,073,000 ALL. The standard errors of the estimate and the coefficient of variation are shown in the sixth and seventh columns of the table. The findings are generated by employing the Mack function from the ChainLadder package in R.

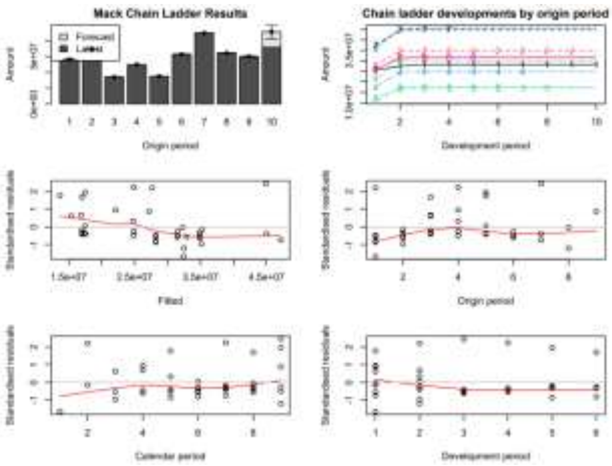

Fig. 1: Diagnostic Analysis of the Mack Chain Ladder Method

4.3 Results of Clark's LDF Method

In Clark's LDF method, it is assumed that the ultimate losses experienced in each successive period are distinct and independent from one another. The Clark's Cape Cod approach, on the other hand, operates under the assumption that the projected ultimate losses for each origin year can be calculated by multiplying the earned premium for that specific year with a theoretical loss ratio. The residual plots for both the LDF methods with Weibull distribution and with Log-logistic distribution are shown in Figure 2 and Figure 3.

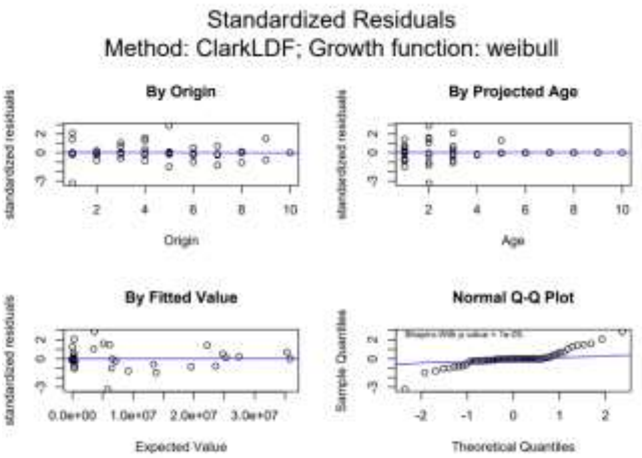

Fig. 2: Clark's LDF chart (Weibull)

When comparing the residuals of both models based on origin and projected age, it is evident that they do not show a similar pattern. Nevertheless, the LDF chart with Weibull distribution displays a more favorable distribution of fitted values and a slightly improved q-q plot. The outcomes are acquired through the utilization of the ClarkLDF function within the ChainLadder package in R.

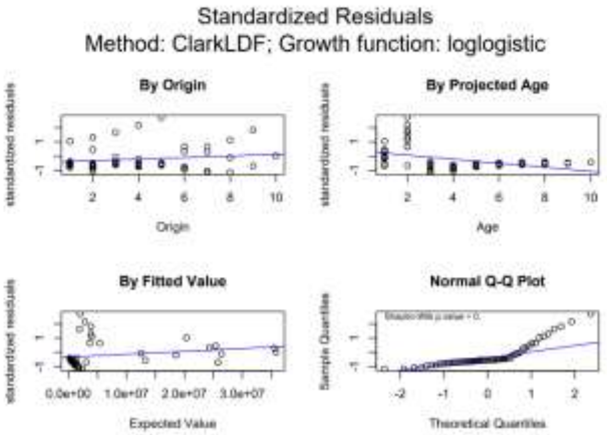

Fig. 3: Clark's LDF chart (Log-logistic)

The results obtained by utilizing the LDF method with Weibull distribution and with parameters θ=0.5 and ω=0.2772783 are presented in Table 5 (Appendix). The LDF method would provide an estimation of 306,594,932 ALL as the total ultimate losses, along with a reserve (future value) of 10,079,250 ALL for the paid claims.

The outcomes obtained for paid claims from the LDF method with Log-logistic distribution with parameters θ=0.5 and ω=0.2772783 are shown in Table 6 (Appendix). By employing the LDF technique, an approximation of 475,702,049 ALL can be determined as the total ultimate losses, together with a reserve (future value) of 179,186,367 ALL for the paid claims.

4.4 Results of Clark's Cape Cod Method

Figure 4 and Figure 5 present the residual plots for both the Cape Cod methods utilizing the Weibull distribution and the Cape Cod methods employing the log-logistic distribution. The graphs presented below indicate a slight upward trend in the development period and standardized waste.

In contrast to the LDF model, the Cape Cod method used to project future reserve values involves both the premium and the expected loss ratio. The results shown in Table 7 and Table 8 in Appendix were obtained by utilizing the ClarkCapeCod function included in the ChainLadder package of R. The outcomes obtained by employing the Cape Code method with Weibull distribution are illustrated in Table 7 (Appendix). The results obtained from this model anticipate a projected reserve amount 176,823,264 ALL.

This projection indicates an approximate total claim amount of 473,338,946 ALL. This model is projected to have an expected loss ratio of 0.945, a standard error of 17,630,357, and a variation coefficient of 10%. The data from Table 8 (Appendix) shows that the reserve (future value) is 180,026,706 ALL, implying a total claim amount of approximately 476,542,388 ALL. The standard error is 17,051,894 ALL, leading to a coefficient of variation of 9.5%.

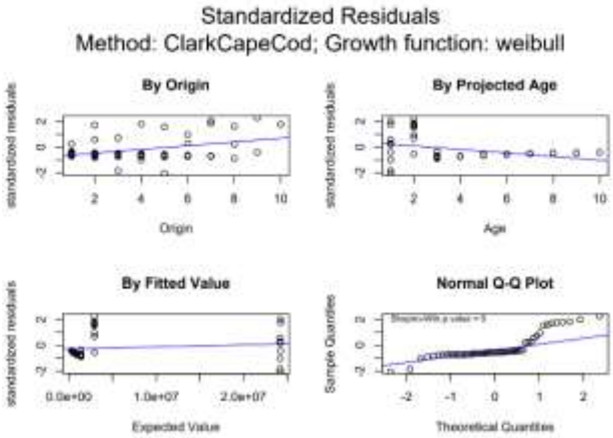

Fig. 4: Clark's Cape Cod chart (Weibull)

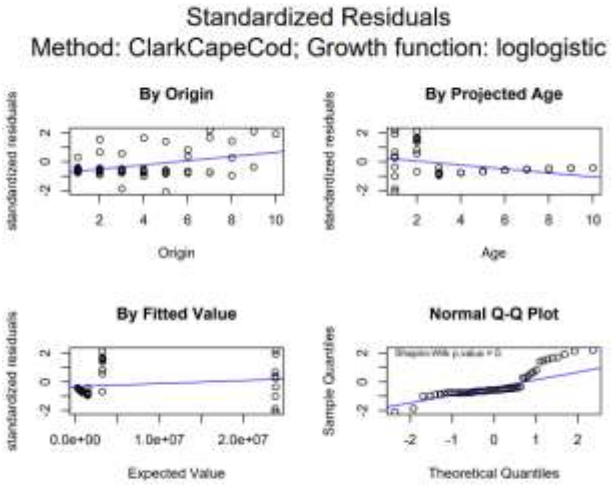

Fig. 5: Clark's Cape Cod chart (Log-logistic)

4.5 Comparison of the Methods

The models proposed initially are all included in summary Table 9, Table 10 and Table 11 in Appendix for the purpose of comparing them and selecting a model that fits our values. The ultimate claims, which have been determined using the methods described in the theory, are presented in the Table 9 (Appendix). Claims are considered fully paid when they reach 100% of the claimed amount. The reserve values, estimated through the methods outlined in section 3, are displayed in Table 10 (Appendix). Clark Cape Cod method demonstrates higher ultimate values and reserve amounts when compared to the Mack model and Clark LDF using the Weibull distribution. The valuation technique utilized by Clark Cape Cod deviates slightly from other methods by incorporating a fluctuating premium amount and a constant loss level ratio (ELR) during the analysis periods. Consequently, the reserves will be directly influenced by the projected premium amount in future periods. Furthermore, the model utilizing a log-logistic distribution predicts larger reserves in contrast to the model employing a Weibull distribution; these findings align with the results derived from the information criteria. Hence, it can be deduced that the model utilizing the Weibull distribution is the most appropriate for our dataset. The summary of the models, along with their respective standard errors and coefficients of variation, is presented in Table 11 (Appendix). Among the models analyzed, the Mack model and Clark LDF model with log-logistic distribution stand out for having the lowest coefficient of variation.

Specifically, the Mack model has a coefficient of variation of 0.42, while the Clark LDF model with Log-logistic distribution has a coefficient of variation of 8.6%. Additionally, their standard errors are 4,494,040 ALL and 15,447,440 ALL, respectively. The model with the lowest standard errors is the Clark LDF model with Weibull distribution, followed by the Mack model. The projected reserve amount differs by 6,250 ALL between these two models, with the Mack model showing the lowest reserves.

5 Discussion

To ensure the financial strength of an insurance company and meet solvency capital requirements, it is crucial to carefully assess the reliability of reserve valuation. Therefore, it is highly recommended to evaluate the data and decide regarding the most accurate reserve. The literature review encompasses various studies that indicate the superiority of stochastic methods, such as Mack's model and the Cape Cod method, in comparison to deterministic approaches in terms of precision and flexibility. Several studies emphasize the accuracy of specific models such as the Mack Chain Ladder, [13], [14], or the Clark LDF model, [21], while others concentrate on comparing multiple models and their suitability for diverse insurance datasets, [10], [11], [18], [21].

Various studies, including, [15], [20] and [22], contribute to the advancement of existing stochastic models through the introduction of improvements or the investigation of their suitability in novel contexts. For example, [20], improves parameter estimation by extending the Cape Cod technique with additional stochastic models, while, [22], explores the combination of different stochastic models into a unified model for reserve estimation.

In this study, the overall portfolio of the sickness insurer is subjected to analysis using three stochastic methods to predict a realistic reserve. The results derived from the Mack model demonstrate that the reserve estimation for paid claims is 10,073,000 ALL, with standard errors, 4,494,040 ALL. The application of the Mack model can yield results that apply to a wide range of companies operating in the private health insurance market in Albania. The findings of this paper align with the outcomes of previous studies, [11], [13], which determined that the Mack model is the most suitable approach for estimating reserves among distribution-based methodologies. This preference is based on the lower prediction errors observed in the Mack model compared to alternative models.

In contrast, the results of this study differ from the conclusions drawn by [10], [12] as they suggested that the ODP method with bootstrapping was a more prudent choice when compared to the Mack method. The results obtained by utilizing the LDF method with Weibull distribution estimate a reserve of 10,079,250 ALL for the paid claims with standard errors of 2,039,631 ALL. The application of the LDF method with the Log-logistic distribution has suggested a reserve of 179,186,367 ALL for the paid claims, with standard errors of 15,447,440 ALL. By implementing the Cape Code method, the estimated reserve (future value) for the paid claims amounts to 176,823,264 ALL with standard errors 17,630,357 ALL for the Weibull distribution, while the reserve (future value) for the Log-logistic distribution stands at 180,026,706 ALL with standard errors 17,051,894 ALL. Models utilizing the log-logistic distribution demonstrate a lower coefficient of variation, while those employing the Weibull distribution show a lower standard error. Consequently, it can be inferred that the Weibull distribution model is more suitable for fitting our data, despite displaying higher variability in the estimates. The results are in line with the research performed by [18] indicating that the optimal distribution for incurred and paid claims is the Weibull distribution. Furthermore, the Clark LDF method presents a lower standard error, compared to the Cape Cod method. It is crucial to emphasize that Mack's method provides a lower estimate for paid claims, resulting in a small difference of 6,250 ALL with the Clark LDF (Weibull) method. In comparison to alternative methods, the Clark LDF method accurately predicts the reserve of claims with the lowest value. The findings are in line with the study conducted by [21] recommending the Clark LDF model as a more reliable option for accurate claims reserving in insurance companies. This is because of its lower variability in reserve predictions compared to traditional models such as the Chain Ladder. The results of this study are inconsistent with the conclusions drawn by [18] and [19] who assert that the Cape Cod method offers a unique approach based on growth curve modeling and shows a smaller variance when contrasted with the LDF method.

The disparities in findings between previous studies and our study, which employed similar models but different datasets, highlight the existence of significant risk factors influencing the number of insurance claims across various datasets. Additionally, various models must be used in different datasets to capture the dataset characteristics and provide a more precise prediction.

6 Conclusion

This paper contributes to the growing body of literature on claim-reserving techniques in private health insurance by utilizing stochastic methods and conducting a comprehensive analysis. Researchers, insurers, and regulators will gain valuable insights from this study.

The study aims at analyzing the claims distribution and find the best forecasting approach for the loss reserve. The results of this research emphasize the importance of selecting an appropriate statistical distribution, where the Weibull distribution is the best choice to model claims data. Clark's method, which includes both LDF and Cape Cod techniques, can be effectively used to forecast disease portfolio claim reserves. Among other methods, the Clark LDF method with Weibull distribution stands out as it shows superior accuracy in claim reserves projection. Mack and Clark LDF models with Weibull distribution were shown to be a more accurate method than all other models demonstrating the lowest reserve claims and standard errors.

Additionally, the study examines different models to emphasize the need for insurers to make careful trade-offs between simplicity and complexity, while also taking into account supplementary aspects such as premium fluctuation. Furthermore, insurance companies must have accurate claim reserves to meet their financial obligations, and more importantly to maintain the confidence of policyholders and stakeholders. Consequently, alternative strategies for distribution of funds should be considered by the firm rather than only maintaining a large claims reserve.

Nonetheless, this research has its limitations. This study uses a given set of data from Albanian insurance companies that do not necessarily represent the whole population of private health insurance portfolios worldwide. Moreover, stochastic approaches are good at capturing risk but they require proper selection and validation of parameters in order to enhance reliability.

To address these limitations, future research should explore alternative methodologies and incorporate extensive datasets from diverse insurance markets. A potential area for future investigation is the use of random forest or artificial neural networks to estimate claim reserves within the private health insurance portfolio.

References:

[1] Jia, P., Yan, J. (2022). Effects of haze pollution and institutional environment on demand for commercial health insurance. *Frontiers in Psychology*, Vol.13, No.1002470. https://doi:10.3389/fpsyg.2022.1002470.

[2] Thomson, S., Sagan, A., & Mossialos, E. (Eds.). (2020). *Private health insurance: History, Politics and Performance.* Cambridge University Press. https://doi.org/10.1017/9781139026468.

[3] Trinh, C. T., Chao, C.-C., & Ho, N. Q. (2023). Private health insurance consumption and public health-care provision in OECD countries: Impact of culture, finance, and the pandemic. *The North American Journal of Economics and Finance*, Vol. 64, No.101849. https://doi:10.1016/j.najef.2022.101849.

[4] Letunovska, N., Saher, L., Syhyda, L., & Yevdokymova, A. (2023). Analysis of health care system development in the regions amidst the economic inclusiveness and social determinants of health. *WSEAS Transactions on Environment and Development*, Vol. 19, No. 14, pp. 163–173. https://doi:10.37394/232015.2023.19.14.

[5] Barzylovych, A., Ursakii, Y., Nadezhdenko, A., Mamatova, T., Chykarenko, I., & Kravchenko, S. (2021). The influence of medical services public management on the population’ life quality. *WSEAS Transactions on Environment and Development*, Vol. 17, No. 60, pp. 619–629. https://doi:10.37394/232015.2021.17.60.

[6] Szatkowski, M., & Delong, Ł. (2021). One-Year and Ultimate Reserve Risk in Mack Chain Ladder Model. *Risks*, Vol. 9, No. 9, pp. 152. http://dx.doi.org/10.3390/risks9090152.

[7] Arce, M., Giner, B., & Taleb, M. A. (2023). Due process as a legitimating mechanism: Participation and responsiveness in the development of IFRS 17: Insurance contracts. *Journal of Accounting and Public Policy*, Vol. 42, No. 107150. https://doi:10.1016/j.jaccpubpol.2023.107150.

[8] England, P. D., & Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, Vol. 8, No. 3, pp. 443–518. https://doi:10.1017/s1357321700003809.

[9] Mack, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance, Mathematics & Economics*, Vol. 15, No. 2–3, pp. 133–138. https://doi:10.1016/0167-6687(94)90789-7.

[10] Muja. A., Shehi. E., (2018). Actuarial models used in claims reserving. Fourth International Conference on: "Social and Natural Sciences– Global Challenge 2018" (ICSNS IV-2018) Vienna, ISBN: 978-9928-259-15-8, [Online].. https://www.iipccl.org/wp-content/uploads/2018/12/Proceedings-Vienna-3-November.pdf (Accessed Date: October 7, 2023).

[11] Tuysuz, S., & Pekel, P. (2019). A comparison of stochastic claims reserving methods. *European Journal of Business and Management Research*, Vol. 4, No.4. https://doi:10.24018/ejbmr.2019.4.4.43.

[12] Tee, L., Käärik, M., & Viin, R. (2017). On Comparison of Stochastic Reserving Methods with Bootstrapping. *Risks*, Vol. 5, No. 1, pp.1-21. https://dx.doi.org/10.3390/risks5010002.

[13] Abonongo John, Ida Anuwoje L. Abonongo, Loss Reserving -the Mack Method and Associated Bootstrap Predictions, *Applied Mathematics*, Vol. 11, No. 3, 2021, pp. 29-36. https://doi: 10.5923/j.am.20211103.01.

[14] Lindholm, M., Lindskog, F., & Wahl, F. (2020). Estimation of conditional mean squared error of prediction for claims reserving. *Annals of Actuarial Science*, Vol. 14, No. 1, pp. 93-128. https://doi:10.1017/S174849951900006X.

[15] Schmidt KD, Schnaus A. (1996). An Extension of Mack's Model for the Chain Ladder Method. *ASTIN Bulletin*. Vol. 26, No.2, pp. 247-262. https://doi:10.2143/AST.26.2.563223

[16] Clark, D. R. (2003). LDF curve-fitting and stochastic reserving: a maximum likelihood approach. *Casualty Actuarial Society E-Forum*, Vol. 03, Fall), pp. 41-91, [Online]. https://www.casact.org/sites/default/files/database/forum_03fforum_03ff041.pdf (Accessed Date: January 6, 2024).

[17] Zhang, Y., Dukic, V. & Guszcza, J. (2012). A Bayesian non-linear model for forecasting insurance loss payments. Journal of Royal Statistics, Vol. 175, No. 2, pp. 637-656. https://doi.org/10.1111/j.1467-985X.2011.01002.x.

[18] Raço, E., Haxhi, K., Llagami, E., & Zaçaj, O. (2022). Comparison of statistical methods for claims reserve estimation using R language. *WSEAS Transactions on Mathematics*, Vol. 21, No. 61, pp. 547–552. https://doi:10.37394/23206.2022.21.61.

[19] Wolny, A. (2013). On some modifications of Cape Cod method for loss reserving. *The 7th International Days of Statistics and Economics*, Prague, September 19-21, 2013, [Online]. https://msed.vse.cz/files/2013/198-Wolny-Dominiak-Alicja-paper.pdf (Accessed Date: December 12, 2023).

[20] G. Taylor, (2019). A Cape Cod model for the exponential dispersion family. *Insurance: Mathematics and Economics*, Vol. 85, pp. 126–137. https://doi.org/10.1016/j.insmatheco.2018.11.008.

[21] Rubio, R. D. R. (2023) The Clark LDF method: study and comparison with other claims reserving methods [Mastre's thesis, University of Barcelona]. Spain, [Online]. https://hdl.handle.net/2445/199820 (Accessed Date: November 22, 2023).

[22] Dal Moro, E. (2022). Unification of stochastic reserving models. *SSRN ElectronicJournal*. https://doi:10.2139/ssrn.4229217.

[23] Gesmann M, Murphy D, Zhang Y, Carrato A, Wüthrich M, Concina F, Dal Moro E. _ChainLadder: Statistical Methods and Models for Claims Reserving in General Insurance_. R package version 0.2.18.2023, [Online]. https://mages.github.io/ChainLadder/ (Accessed Date: January 15, 2024).

APPENDIX

Table 2. Run-off triangle data for the settlement of cumulative claims losses.

Accident Semiannually i	Development period j									
	6	12	18	24	30	36	42	48	54	60
2018S1	25,708,146	27,384,733	28,215,173	28,215,173	28,215,173	28,215,173	28,215,173	28,215,173	28,215,173	28,215,173
2018S2	27,982,595	34,791,276	34,877,058	34,877,058	34,877,058	34,877,058	34,877,058	34,877,058	34,877,058	
2019S1	12,210,107	16,717,660	17,022,202	17,022,202	17,022,202	17,022,202	17,022,202	17,022,202		
2019S2	17,552,062	24,535,120	24,859,991	24,859,991	24,875,511	24,875,511	24,875,511			
2020S1	10,876,183	17,319,770	17,505,677	17,505,677	17,505,677	17,505,677				
2020S2	26,105,381	31,222,767	31,469,533	31,469,533	31,469,533					
2021S1	37,449,430	44,562,562	44,891,943	44,899,543						
2021S2	25,624,861	32,177,447	32,180,247							
2022S1	21,596,157	29,670,863								
2022S2	35,799,875									

Source: Author

Table 3. Information Criteria

Information Criteria		
Distribution	AIC	BIC
Log-logistic	1914	1918
Weibull	1908	1912

Source: Author

Table 4. Mack 's model results

Times	Current Value	Dev. Date	Ultimate	Future Value	Mack.S.E	CV
2018S1	28,215,173	1	28,215,173	0	0	0
2018S2	34,877,058	1	34,877,058	0	0	0
2019S1	17,022,202	1	17,022,202	0	0	0
2019S2	24,875,511	1	24,875,511	0	0	0
2020S1	17,505,677	1	17,505,677	0	0	-0.637
2020S2	31,469,533	1	31,469,533	0	0	-0.592
2021S1	44,899,543	1	44,904,069	4,530	9,710	2.145
2021S2	32,180,247	1	32,184,721	4,470	8,330	1.862
2022S1	29,670,863	0.990	29,974,771	304,000	296,000	0.974
2022S2	35,799,875	0.786	45,561,049	9.760,000	4,180,000	0.429
Total	296,515,682		306,589,764	10,073,000	4,494,040	0.42

Source: Author

Table 5. LDF results (Weibull)

Origin	Current Value	LDF	Ultimate	Future Value	Std. Error	CV%
2018S1	28,215,173	1	28,215,173	0	1	22426594
2018S2	34,877,058	1	34,877,058	0	6	4323089
2019S1	17,022,202	1	17,022,202	8	21	1324932
2019S2	24,875,511	1	24,875,511	0	117	234911
2020S1	17,505,677	1	17,505,678	1	459	59992
2020S2	31,469,533	1	31,469,563	30	2,875	9589
2021S1	44,899,543	1	44,900,475	932	16,074	1725
2021S2	32,180,247	1	32,194,801	14,554	64,435	442
2022S1	29,670,863	1.01	29,965,898	295,035	303,882	103
2022S2	35,799,875	1.273	45,568,573	9,768,698	2,006,738	20.5
Total	296,515,682		306,594,932	10,079,250	2,039,631	20.2

Source: Author

Table 6. LDF results (Log-logistic)

Origin	Current Value	LDF	Ultimate	Future Value	Std. Error	CV%
2018S1	28,215,173	1.442	40,686,485	12,471,312	4,074,917	32.7
2018S2	34,877,058	1.456	50,775,813	15,898,755	4,601,471	28.9
2019S1	17,022,202	1.472	25,055,821	8,033,619	3,270,991	40.7
2019S2	24,875,511	1.491	37,090,688	12,215,177	4,033,357	33.0
2020S1	17,505,677	1.514	26,509,426	9,003,749	3,495,336	38.8
2020S2	31,469,533	1.544	48,581,487	17,111,954	4,773,379	27.9
2021S1	44,899,543	1.583	71,076,226	26,176,683	5,903,972	22.6
2021S2	32,180,247	1.640	52,776,119	20,595,872	5,237,175	25.4
2022S1	29,670,863	1.737	51,550,234	21,879,371	5,398,318	24.7
2022S2	35,799,875	2.000	71,599,750	35,799,875	7,558,486	21.1
Total	296,515,682		475,702,049	179,186,367	15,447,440	8.6

Source: Author

Table 7. Clark Cape Cod results (Weibull)

Origin	Current Value	Premium	ELR	Ultimate	Future Value	Std. Error	CV%
2018S1	28,215,173	50,020,000	0.945	43,386,060	15,170,887	5,164,127	34.0
2018S2	34,877,058	50,040,000	0.945	50,355,147	15,478,089	5,216,150	33.7
2019S1	17,022,202	50,060,000	0.945	32,846,212	15,824,010	5,274,116	33.3
2019S2	24,875,511	50,080,000	0.945	41,095,190	16,219,679	5,339,647	32.9
2020S1	17,505,677	50,100,000	0.945	34,187,300	16,681,623	5,415,151	32.5
2020S2	31,469,533	50,120,000	0.945	48,705,803	17,236,270	5,504,439	31.9
2021S1	44,899,543	50,140,000	0.945	62,829,444	17,929,901	5,614,103	31.3
2021S2	32,180,247	50,160,000	0.945	51,035,742	18,855,495	5,757,188	30.5
2022S1	29,670,863	50,180,000	0.945	49,920,806	20,249,943	5,966,276	29.5
2022S2	35,799,875	50,100,000	0.945	58,977,242	23,177,367	6,382,980	27.5
Total	296,515,682	501,100,000		473,338,946	176,823,264	17,630,357	10

Source: Author

Table 8. Clark Cape Cod results (Log-logistic)

Origin	Current Value	Premium	ELR	Ultimate	Future Value	Std. Error	CV%
2018S1	28,215,173	50,020,000	0.951	43,672,122	15,456,949	4,996,502	32.3
2018S2	34,877,058	50,040,000	0.951	50,629,874	15,752,816	5,044,095	32
2019S1	17,022,202	50,060,000	0.951	33,110,677	16,088,475	5,097,551	31.7
2019S2	24,875,511	50,080,000	0.951	41,351,210	16,475,699	5,158,532	31.3
2020S1	17,505,677	50,100,000	0.951	34,437,963	16,932,286	5,229,522	30.9
2020S2	31,469,533	50,120,000	0.951	48,956,513	17,486,980	5,314,490	30.4
2021S1	44,899,543	50,140,000	0.951	63,090,252	18,190,709	5,420,371	29.8
2021S2	32,180,247	50,160,000	0.951	51,327,479	19,147,232	5,561,055	29.0
2022S1	29,670,863	50,180,000	0.951	50,296,605	20,625,742	5,771,769	28.0
2022S2	35,799,875	50,100,000	0.951	59,669,693	23,869,818	6,209,101	26.0
Total	296,515,682	501,100,000		476,542,388	180,026,706	17,051,894	9.5

Source: Author

Table 9. Ultimate claims of Mack, LDF (Weibull), LDF(Log-logistic), Cape Cod (Weibull) and Cape Cod (Log-logistic) methods

Semiannually Accident	Mack Chain Ladder	Clark's LDF(Weibull)	Clark's LDF(Loglogistic)	Clark's Cape Cod(Weibull)	Clark's Cape Cod(Loglogistic)
2018S1	28,215,173	28,215,173	40,686,485	43,386,060	43,672,122
2018S2	34,877,058	34,877,058	50,775,813	50,355,147	50,629,874
2019S1	17,022,202	17,022,202	25,055,821	32,846,212	33,110,677
2019S2	24,875,511	24,875,511	37,090,688	41,095,190	41,351,210
2020S1	17,505,677	17,505,678	26,509,426	34,187,300	34,437,963
2020S2	31,469,533	31,469,563	48,581,487	48,705,803	48,956,513
2021S1	44,904,069	44,900,475	71,076,226	62,829,444	63,090,252
2021S2	32,184,721	32,194,801	52,776,119	51,035,742	51,327,479
2022S1	29,974,771	29,965,898	51,550,234	49,920,806	50,296,605
2022S2	45,561,049	45,568,573	71,599,750	58,977,242	59,669,693
Total	306,589,764	306,594,932	475,702,049	473,338,946	476,542,388

Source:Author

Table 10. Reserves estimation of Mack, LDF (Weibull), LDF (Loglogistic), Cape Cod (Weibull) and Cape Cod (Loglogistic) methods.

Semiannually Accident	Mack Chain Ladder	Clark's LDF(Weibull)	Clark's LDF(Loglogistic)	Clark's Cape Cod(Weibull)	Clark's Cape Cod(Loglogistic)
2018S1	0	0	12,471,312	15,170,887	15,456,949
2018S2	0	0	15,898,755	15,478,089	15,752,816
2019S1	0	0	8,033,619	15,824,010	16,088,475
2019S2	0	0	12,215,177	16,219,679	16,475,699
2020S1	0	1	9,003,749	16,681,623	16,932,286
2020S2	0	30	17,111,954	17,236,270	17,486,980
2021S1	4,530	932	26,176,683	17,929,901	18,190,709
2021S2	4,470	14,554	20,595,872	18,855,495	19,147,232
2022S1	304,000	295,035	21,879,371	20,249,943	20,625,742
2022S2	9.760,000	9,768,698	35,799,875	23,177,367	23,869,818
Total	10,073,000	10,079,250	179,186,367	176,823,264	180,026,706

Source:Author

Table 11. Model summary

	Mack Chain Ladder	Clark's LDF(Weibull)	Clark's LDF(Loglogistic)	Clark's Cape Cod (Weibull)	Clark's Cape Cod (Loglogistic)
Standard Error	4,494,040	2,039,631	15,447,440	17,630,357	17,051,894
CV%	0.42	20.2	8.6	10	9.5

Source: Author

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The author has no conflicts of interest to declare.