Technical reserving in health insurance: A bootstrap approach

Ayyoub SAOUDI[1], Fatima EL KASSIMI[2], Jamal ZAHI[3]

[1] PhD student, University Hassan 1st, Faculty of Economics and Management, LM2CE, Settat, Morocco

[2] PhD, University Hassan 1st, Faculty of Economics and Management, LM2CE, Settat, Morocco

[3] Full Professor, University Hassan 1st, Faculty of Economics and Management, LM2CE, Settat, Morocco

E-mail : a.saoudi@uhp.ac.ma, f.elkassimi@uhp.ac.ma, zahi71@hotmail.com

Article history

Received Jun 19, 2023
Revised Jun 26, 2023
Accepted Jul 01, 2023
Published Jul 06, 2023

ABSTRACT

Bootstrap methods have long been used by actuaries to estimate technical reserves. The present work aims to illustrate the use of this method in health insurance, by exploiting the portfolio of claims paid by the insureds of a Moroccan mutual health insurance company over 12 months. This article aims to highlight the advantages of using the Bootstrap algorithm to predict the distribution of future claims expenses and estimate the standard error, to guarantee the solvency of the mutual in question and maintain its ability to meet these future commitments. The probability distribution used in this work is the Log-Normal, as it is best suited to the claims experience studied. Using this distribution, the ultimate cost of claims, the claims reserve, and its standard error have been estimated.

Keywords: Bootstrap, Technical reserves, Health insurance.

I. INTRODUCTION

Health insurance is a key sector of the insurance industry, protecting individuals against the financial risks associated with medical expenses. Health insurers face several challenges when it comes to technical reserving, *i.e.*, estimating expected losses and the reserves needed to cover these losses. The uncertainties associated with healthcare costs, disease trends, and the use of healthcare services make this estimation even more difficult [1].

One of the most commonly used methods for estimating insurance losses is the Bootstrap method. This algorithm is a statistical approach that is based on the classic Chain Ladder (CL) method, it enables us to estimate the distribution of losses using random samples with replacement. By applying this method to aggregated health insurance data, it is possible to obtain a more precise estimate of expected losses and the reserves needed to cover these losses.

This article proposes an innovative approach to health insurance reserving using stochastic modeling *via* the Bootstrap method. In the following section, we present the research methodology adopted, the third section provides the results obtained through our model, the fourth section consists of a discussion of the advantages and disadvantages of the algorithm, and a summary conclusion is given in the fifth section.

II. METHODOLOGY

Technical reserving in health insurance is a complex process involving the collection, aggregation, and analysis of healthcare data. The aggregate data used to estimate reserves is generally collected from a variety of sources, such as healthcare providers, healthcare networks, billing systems, and insurance plans.

The aggregate data used may include information on healthcare costs, frequency of healthcare use, presence of chronic illnesses (long-term conditions) among insureds, and risk factors: demographic and socio-economic characteristics of insureds.

These data are then processed and analyzed to calculate the technical reserves required to maintain the insurance company's solvency, thereby improving its ability to meet its future commitments to its policyholders. The estimation of technical reserves relies on statistical and actuarial models to predict future healthcare costs, as well as adjustments to take account of population health trends.

In short, the aggregate data used in this respect is a crucial factor in the reserving process, insofar as it is essential in assessing risks and estimating the provisions needed to ensure adequate healthcare coverage for policyholders.

Actuaries often structure this aggregate data in the form of a triangular table. This data, illustrated in the runoff triangle, represents claims payments (healthcare costs) for each year of occurrence and each year of development.

A. *Presentation of Bootstrap method*

Invented by [2], the Bootstrap algorithm makes it possible to estimate the variability of a parameter in particular, by carrying out resampling with a discount using Monte-Carlo simulations [3]. Statisticians often use this method when the sample is the only information available, and its distribution law is unknown. The aim is to iteratively resample the same data to determine its distribution. If the sample available is not representative, the Bootstrap samples will deviate from the real values. Hence the need for a representative sample of the population.

Bootstrapping has become increasingly popular among actuaries when it comes to calculating technical reserves [5], [6], and [7]. The approach of this algorithm is to resample, with replacement, the diagonal residuals and create pseudo data sets. To estimate residual values, the Bootstrap algorithm uses the CL method to calculate adjusted values.

Not only does this method make it possible to evaluate the uncertainty of the reserves, as allowed by the formula in [4], but it also makes it possible to estimate the distribution of the reserves. In fact, by resampling, the Bootstrap generates new runoff triangles, and thus a distribution of reserves is calculated based on these new triangles.

B. *Model hypotheses*

For the Bootstrap model to function properly, it must satisfy the condition that the data in the starting sample are independent and identically distributed *(i.i.d)*. However, this assumption is rarely met. Indeed, non-cumulative payments ($Y_{i,j}$) are generally not identically distributed. For this reason, sampling is not based on increments, but on the Pearson, residuals calculated from these data.

In the case of reserving, the initial sample is the upper triangle of settlements, and the aim is to estimate a confidence interval for the total technical reserve estimate.

In what follows, the following notations will be used:

- $\hat{X}_{i,j}$: Estimated incremental payment for the year in which a claim occurs and the year in which it develops;
- $\hat{C}_{i,j}$: Estimated cumulative payment for the year in which a claim occurs and the year in which it develops;
- $r_{i,j}$: Pearson residual calculated for the year in which a claim occurs and its year of development;
- $r^*_{i,j}$: New Pearson residual calculated after resampling for the year of occurrence i and year of development j.

C. *Upper triangle prediction*

Before embarking on estimating the triangle, it should be noted that the Log Normal model will be used, as it gives identical results to those provided by the CL method.

The estimation of the upper part of the triangle is based initially on the selected age-to-age factors, which are applied to the last cumulative diagonal of the initial triangle. This involves applying the following formula:

$$\forall i = 0,\dots,1 \quad \hat{C}_{i,j-1} = C_{i,j-1} \qquad (1)$$

After estimating the diagonal, we proceed by applying age-to-age factors in the opposite direction of the development of the claims of the lower triangle in CL, according to the following formula:

$$\forall i + j < 1 \quad \hat{C}_{i,j} = \frac{\hat{C}_{i,j-1}}{\lambda_i} \qquad (2)$$

Since the first column of incremental and cumulative settlements are identical. The prediction of the triangle of incremental settlements will thus be made based on the predicted triangle of cumulative settlements, by applying the following formula to the first column:

$$\forall i = 0,\dots,1 \quad \hat{X}_{i,0} = \hat{C}_{i,0} \qquad (3)$$

The remaining incremental payments in the upper triangle are predicted using the following formula:

$$\forall 0 < i + j \le 1 \quad \hat{X}_{i,j} = \hat{C}_{i,j} - \hat{C}_{i,j-1} \qquad (4)$$

D. *Bootstrap Mechanism*

At each Bootstrap iteration, the Pearson triangle of residuals is resampled: for each cell, a random draw is made from the non-excluded residuals. As a result, the same value may appear several times in the triangle.

New upper triangles of incremental amounts are then calculated at each iteration. Each term is estimated using the following formula:

$$\forall i + j \le 1 \quad X^*_{i,j} = r^*_{i,j} \times \sqrt{\hat{X}_{i,j}} + \hat{X}_{i,j} \qquad (5)$$

For each iteration, starting from the upper incremental triangle, the corresponding upper cumulative triangle can be calculated and expanded using the classic CL method. This then yields a vector of reserves per year of occurrence, as well as the total reserve for each iteration.

The two methods detailed above have common drawbacks since they both rely not only on the underlying CL assumptions but also on cumulative amounts. The disadvantage of having to apply it to cumulative amounts is that, for triangles with unknown increments for the earliest years of occurrence A1, ..., Ax, it is not possible to calculate the cumulative values on the right-hand side of the triangle for rows 1 to x, which are crucial for the valuation of the total loss.

Conversely, incremental methods can be applied to the right-hand side of the triangle, as the corresponding data are available from recent accounting years.

TABLE I. ADVANTAGES AND DISADVANTAGES OF BOOTSTRAP

Advantages	Disadvantages
▪ Gives a confidence interval. ▪ No assumptions about the distribution of variables.	▪ Does not work if distribution tails are too thick. ▪ Does not work if statistics involve extreme values.

III. DATA SET

Our database has been collected from a private-sector mutual health insurance company and concerns health care PEC reimbursements for the 2022 year. The variables available to us are the claim ID, the insurance policy, the PEC reimbursement date, the PEC declaration date, the reimbursement date, the reimbursement rate, and the reimbursement amount.

For confidentiality reasons, the amounts of reimbursements have been multiplied by a coefficient x, so that the results obtained during the applications are not biased. These data are not directly exploitable, and specific processing is required at this stage.

IV. RESULTS AND DISCUSSION

The Bootstrap algorithm belongs to the so-called black-box models, as it does not display intermediate results such as the coefficients of the runoff or the completed triangle. However, it does give the final result in the form of monthly IBNRs accompanied by the corresponding error margins.

TABLE II. BOOTSTRAP RESULTS

Month	Latest Mean	Ultimate Mean	IBNR	IBNR .S.E	IBNR 75%	IBNR 95%
1	128 256	128 256	0	0	0	0
2	123 335	124 405	1070	3 681	1 102	8 509
3	102 427	106 476	4 049	5 620	5 868	15 063
4	108 887	113 739	4 852	6 255	7 027	18 083
5	79 427	84 953	5 526	6 347	8 170	17 934
6	114 299	124 247	9 948	8 656	14 457	26 435
7	131 194	149 216	18 022	12 163	24 874	40 772
8	118 242	140 076	21 834	12 043	28 838	43 352
9	120 949	149 105	28 156	14 607	35 749	57 028
10	74 717	95 352	20 635	11 888	27 328	41 469
11	46 885	64 758	17 873	11 222	22 62	40 007
12	6 367	35 895	29 528	28 635	41 172	87 042

This table shows the average of the latest settlements (Latest Mean) and the average of the ultimate charges (Ultimate Mean) from which the corresponding PSAP value for each year is calculated.

TABLE III. ESTIMATING IBNR USING THE BOOTSTRAP METHOD

	Totals
Latest:	1 154 985
Average ultimate charges:	1 316 476
Mean PSAP:	161 492
IBNR. S. E:	54 205
Total IBNR 75%:	196 583
Total IBNR 95%:	254 885

The amount is obtained by subtracting final settlements from average final expenses:

161492= 1 316 476 – 1 154 985

- This table summarizes the totals obtained by adding IBNRs, final expenses, and final settlements for each year.

Total IBNR 95% = 254 885 represents the amount required to cover 95% of the claims experience of the insurance company in question, for which it must set aside 161,492 Dhs in reserves.

TABLE IV. LOG-NORMAL DISTRIBUTION

Log-normal distribution	
Meanlog	***Sdlog***
11.9009136	0.4936881
(0.1561179)	(0.1103920)

- The mean equals 11.9009136
- Standard deviation equals 0.4936881

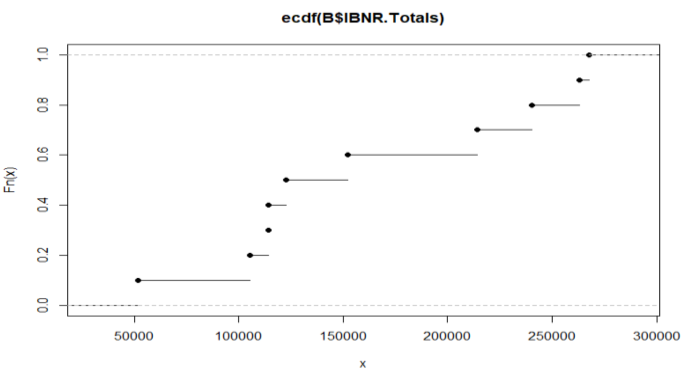

Figure 1. Empirical distribution of total IBNR

The present figure illustrates the distribution function of the reserves estimated using the Bootstrap method.

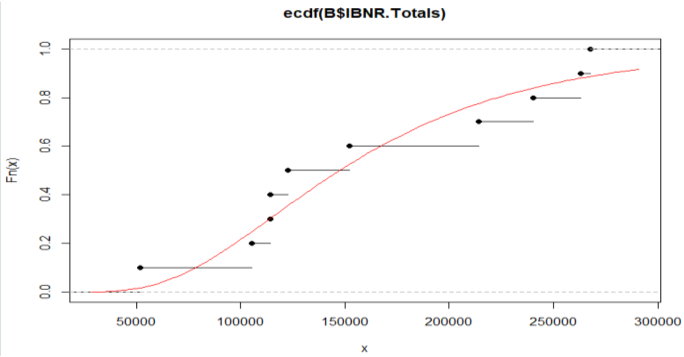

Figure 2. Log-normal fitting of the empirical distribution of total IBNR

This figure shows that the estimated provisions fit a lognormal distribution, represented by the red curve.

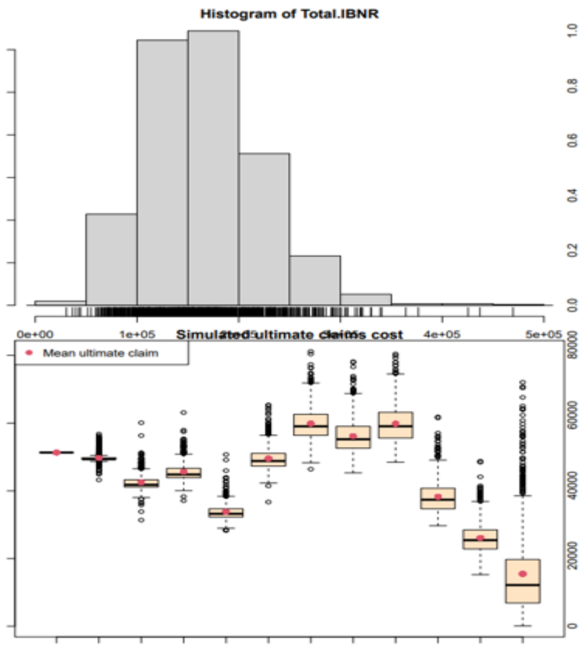

Figure 3. Results obtained using the plot function in the Chain Ladder library.(First)

The figure above illustrates the histogram of simulated total reserves (IBNR) (Above), Below: ultimate charge by year of origin.

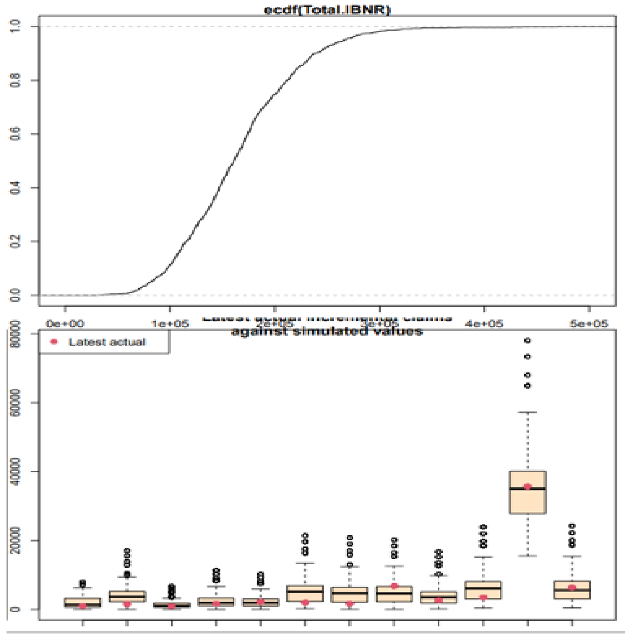

Figure 4. Results obtained using the plot function in the Chain Ladder library.(Second)

The present figure illustrates the empirical distribution function of total IBNR (Above). And a comparison between actual payments and simulated values (Below).

V. CONCLUSION

In conclusion, this article highlights the use of the stochastic Bootstrap method in health insurance to estimate technical reserves. The results obtained show that the probability distribution best suited to the claims experience studied is the Log-Normal distribution. Using this distribution, the Bootstrap algorithm was used to predict the distribution of future claims expenses and estimate the standard error. The results enabled us to estimate the ultimate cost of claims and the claims reserve with a high degree of accuracy, thereby improving the ability of the mutual in question to meet its future commitments to its policyholders.

VI. REFERENCES

[1] A. SAOUDI, F. EL KASSIMI et J. ZAHI, «Technical reserving in nonlife insurance: A literature review of aggregated,» Journal of Integrated Studies In Economics, Law, Technical Sciences & Communication, 2023.

[2] B. Efron, « Bootstrap Methods: Another Look at the Jackknife,» the annals of Statistics, pp. 1 - 26, 1979.

[3] M. Rizzo, «Bootstrap and Jackknife. In Statistical computing with R,» Boca Raton: Chapman & Hall/CRC, pp. 183-211, 2008.

[4] T. Mack, «The standard error of Chain Ladder reserve estimates: recursive calculation and inclusion of a tail factor,» Astin Bulletin, p. 361–366, 1999.

[5] Z. ORIANA, R. ENDRI, H. KLEIDA, L. ETLEVA et H. KOSTAQ, «Bootstrap Methods for Claims Reserving: R Language Approach,»

WSEAS TRANSACTIONS on MATHEMATICS, pp. 252-259, 2022.

[6] S. Björkwall, H. Ola et O. Esbjörn, «Non-Parametric and Parametric Bootstrap Techniques for Age-to-Age Development Factor Methods in Stochastic Claims Reserving,» Scandinavian Actuarial Journal, pp. 306-331, 2009.

[7] Shapland et R. Mark, Using the ODP Bootstrap Model: A Practitioner's Guide, Arlington, Virginia: Casualty Actuarial Society, 2016.