Comparison of Statistical Methods for Claims Reserve Estimation using R Language

ENDRI RAÇO, KLEIDA HAXHI, ETLEVA LLAGAMI, ORIANA ZAÇAJ
Department of Mathematical Engineering,
Polytechnic University of Tirana,
ALBANIA

Abstract: - Stochastic methods of reserves estimation serve to assess the technical provisions of outstanding claims and forecast cash payments of claims in the coming years. The chain ladder model developed by Mack is the more prevalent model. The main deficiency in the chain-ladder model is that the chain-ladder model depends on the last observation on the diagonal. If this last observation is an outlier, this outlier will be projected to the ultimate claim. One of the possibilities to smooth outliers on the last observed diagonal is to robustify such observations, making use of the maximum likelihood estimation along with the common Loss Development Factor (LDF) curve fitting and Cape Cod (CC) techniques. This paper aims to highlight the advantages of using these methods for the best estimate of claims reserves in the Domestic Motor Third Party Liability portfolio. The maximum–likelihood parameter estimation and Chi-square test, are used to specify the probability distribution that best fits the data. Using the Standard Chain Ladder method, LDF, and CC method the claims reserve is calculated based on the run-off triangles of paid claims or the run-off triangles of the incurred claims. Many times, the projections based on the paid claims are different than the projections based on the incurred claims. The solution for this problem is the Munich Chain Ladder method.

Key-Words: - Chain ladder method, Maximum likelihood estimation, Loss Development Factor, Cape Cod, Munich Chain ladder, Run off triangles

Received: September 23, 2021. Revised: May 19, 2022. Accepted: June 22, 2022. Published: July 14, 2022.

1 Introduction

The technical claims reserves, as all technical reserves directly affect the profit loss statement, as well as the technical balance of the insurance company, it is required as a fair evaluation of them. The correctness and dependability of data have a significant impact on the outcomes of stochastic techniques applications [6]. The reliability of the data has a direct impact on the assessment of the claims reserve. The absence of this reliability can alter the results, either underestimating or overestimating the final estimation. The actuary knowing the progress and history of claims in a portfolio, the market where are developed claims payments over the years, the values of outstanding claims, and claims in process court, decides which values estimates are more appropriate to establish technical reserves [1]. In addition, the insurance company must hold sufficient assets to cover technical reserves. The value of assets covering technical provisions must at all times be not less than the gross amount of technical reserves. For this purpose, the estimation of claims reserves is a very important issue. Since the Domestic Motor Third Party Liability (DMTPL) is the most important portfolio of general insurance in Albania, we used the data of this product to apply Loss Development Factor (LDF) curve fitting and Cape Cod (CC) techniques. R programing languages facilitate the comparison of methods.

2 Data and Fitting Distributions

We take into consideration DMTPL claims paid and incurred from 2015 to 2021. The claims amounts are in Albanian currency, Lek.

2.1 Paid Data Claims Distribution

We studied 9,262 Domestic Motor Third Party Liability claims paid.

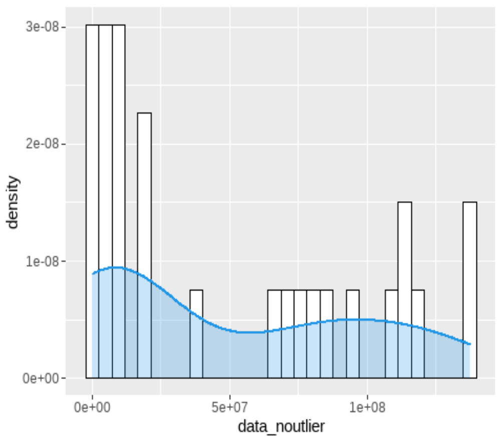

Fig 1: The empirical density of paid claims 2015-2021

Based on the Chi - square tests and information criteria the best distribution that fits the paid claims data is the Weibull distribution [8].

Table 1. Chi-square test

Goodness of fit statistics	Weibull	Gamma	Lognormal
Kolmogorov-Smirnov	0.19160	0.20402	0.27387
Cramer-von Mises	0.15910	0.19672	0.66494
Anderson-Darling	1.26022	1.33719	3.54576

Table 2. Information Criteria

Goodness of fit criteria	Weibull	Gamma	Lognormal
Akaike Criterion AIC	-21.5501	-28.2695	10.6406
Bayesian Criterion BIC	-18.8857	-25.6051	13.3050

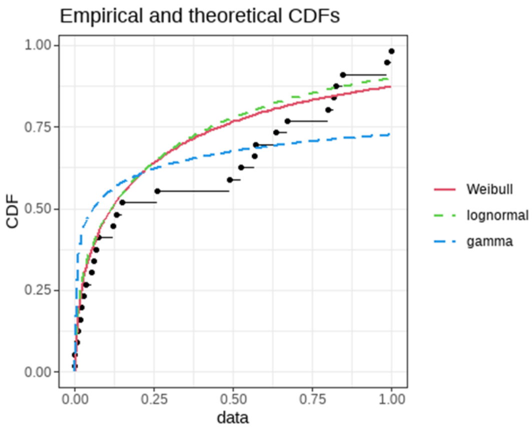

Fig. 2: The empirical and theoretical cumulative density function of paid claims data

2.2 Incurred Data Claims Distribution

We analyzed 8,413 incurred claims from 2015 to 2021.

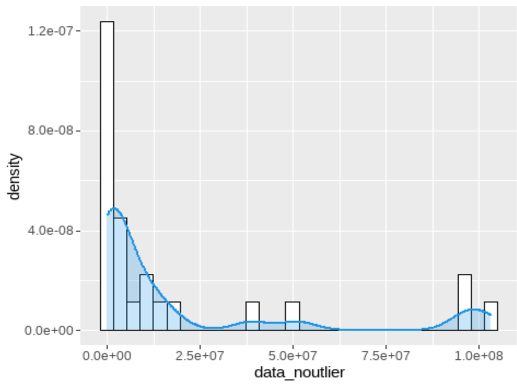

Fig. 3: The empirical density of incurred claims 2015-2021

For the incurred claims too, if we go through the distribution parameters for each theoretical distribution using the fitteR library, it results that the Weibull distribution fits better [8].

Table 3. Chi-square test

Goodness of fit statistics	Weibull	Gamma	Lognormal
Kolmogorov-Smirnov	0.12030	0.14716	0.24891
Cramer-von Mises	0.04982	0.06765	0.35897
Anderson-Darling	0.55078	0.58125	2.35675

Table 4. Information Criteria

Goodness of fit criteria	Weibull	Gamma	Lognormal
Akaike Criterion AIC	-81.4963	-83.9391	-61.0574
Bayesian Criterion BIC	-79.0585	-81.5014	-58.6195

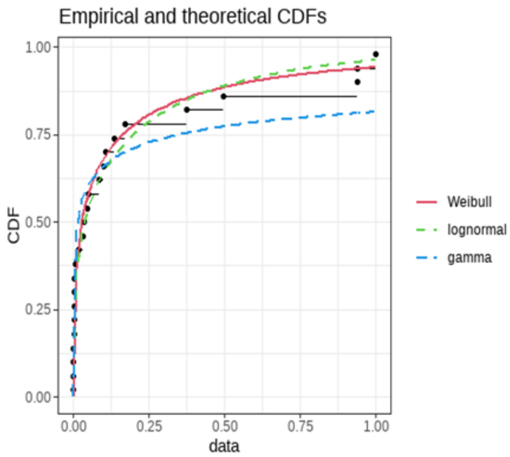

Fig. 4: The empirical and theoretical cumulative density function of incurred claims data

3 Methods

3.1 Clark LDF and Cape Cod Methods

According to LDF e CC techniques to create a proper model of claims reserving we have considered the basic objectives as follows: [2]

• Describing the loss emergence in easier mathematical terms as a track for the selection of the amounts related to carried reserves

• Providing estimating strategies for the predicted reserve's range of probable outcomes.

As a main aspect of the LDF technique, the ultimate loss amount in each accident year is independent of claims in prior years. Instead, the Cape Cod technique assumes that there is a connection between the amounts of final loss projected in each of the years in the prior documented period and that an exposure base accurately detects that relationship. Earned level premium is frequently used as the exposure base. Although both of the above methods can be used for the estimation of the reserves, Cape Cod will be preferable. Because we are dealing with data aggregated into annual blocks as a development triangle, there will be relatively few data points in the model,one data point for each "cell" in the triangle.When the LDF method is used, there is a real issue with over-parameterization. [2]

	1	2	3	4	5	6	7
2015	92,415,152	78,075,705	16,871,685	2,825,878	1,470,000	39,200	102,436
2016	109,733,734	78,493,045	10,240,515	2,615,660	18,050,000	1,064,382	
2017	116,159,869	71,964,169	8,203,749	3,800,000	7,497,779		
2018	112,281,029	67,384,960	20,844,118	4,990,223			
2019	137,313,347	87,328,648	9,487,241				
2020	113,195,095	35,929,511					
2021	135,245,016						

Fig. 5: The incremental triangle of paid claims

The statistical claims reserving model has two primary elements: the emerging of the expected value of the losses in specific periods and the distribution of actual emerging regarding the expected value. The projected amount of loss will emerge based on the estimate of the ultimate loss by year and the estimate of the sample of loss emergence, according to this model. Based on an estimate of the ultimate claim by year and an assessment of the pattern of loss emergence, the model calculates the predicted amounts of claims to emerge. [2]

G(x) is the cumulative percentage of claims paid in time x. The time index "x" represents the time from the "average" accident date to the paid date.

$$G(x) = 1/LDF_x \tag{1}$$

The model will include a Weibull curve, parameterized with a scale θ and a shape ω.

$$G(x|\omega,\theta) = 1 - \exp(-\left(\frac{x}{\theta}\right)^{\omega}) \tag{2}$$

The next step in estimating the amount of loss emergence by period is to apply the emergence pattern G(x), to an estimate of the ultimate claim by accident year. [2]

The final step is the estimation of the variances. It comes because of process variance (random amount) and parameter variance, which is the uncertainty in the estimation. The assumption is that the claim in each period has the same variance/mean ratio and the incremental claims are independent and identically distributed [2]

$$\frac{Variance}{Mean} = \sigma^2 \approx \frac{1}{n-p}\sum_{AY,t}^{n}\frac{\left(c_{AY,t}-\mu_{AY,t}\right)^2}{\mu_{AY,t}} \tag{3}$$

where p is the number of parameters

$c_{AY,t}$ is the actual incremental loss emergence, and $\mu_{AY,t}$ is the expected incremental loss emergence. This corresponds to the chi-square error term. [2].

Usually, the CC method is preferred since the LDF method requires an estimation of parameters, one for each accident year (AY) loss, as well as ω and θ. Due to the additional information given by the exposure base and the fewer parameters, the Cape Cod method has a smaller parameter variance. The process variance can be higher or lower than the LDF method. Generally, the Cape Cod method produces a lower total variance than the LDF method. [2]

Results of the methods θ=0.50000 ω=1.459283

Table 5. LDF technique

Year	Current value	LDF	Utimate value	Future value	Standard Error	CV%
2015	191,800,056	1.002	192,152,611	352,555	1,703,131	483.1
2016	220,197,336	1.004	221,093,168	895,832	2,797,751	312.3
2017	207,625,566	1.009	209,555,566	1,930,000	4,209,341	218.1
2018	205,500,330	1.022	210,062,885	4,562,555	6,672,370	146.2
2019	234,129,236	1.057	247,455,742	13,326,506	11,948,301	89.7
2020	149,124,606	1.168	174,235,699	25,111,093	16,273,130	64.8
2021	135,245,016	1.815	245,498,753	110,253,737	42,637,449	38.7
Total	*1,343,622,146*		*1,500,054,424*	*156,432,278*	*52,873,148*	*33.80*

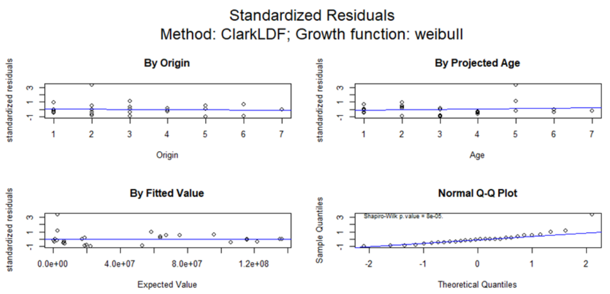

Fig. 6: LDF standardized residuals

Table 6. Cape Cod method

Year	Current value	ELR	Earned Premiums	Utimate value	Future value	Standard Error	CV%
2015	191,800,056	0.233	891,779,071	192,235,619	435,563	1,843,393	423.2
2016	220,197,336	0.233	855,071,845	221,103,668	906,332	2,712,980	299.3
2017	207,625,566	0.233	892,394,125	209,737,359	2,111,793	426,579	202.0
2018	205,500,330	0.233	821,252,582	209,999,188	4,498,858	6,323,195	140.6
2019	234,129,236	0.233	888,137,830	245,960,473	11,831,237	10,444,722	88.3
2020	149,124,606	0.233	1,060,131,446	186,144,645	37,020,039	18,464,852	49.9
2021	135,245,016	0.233	1,094,152,645	251,586,915	116,341,899	31,766,420	27.3
Total	*1,343,622,146*		*6,502,919,544*	*1,516,767,867*	*173,145,721*	*45,984,943*	*26.50*

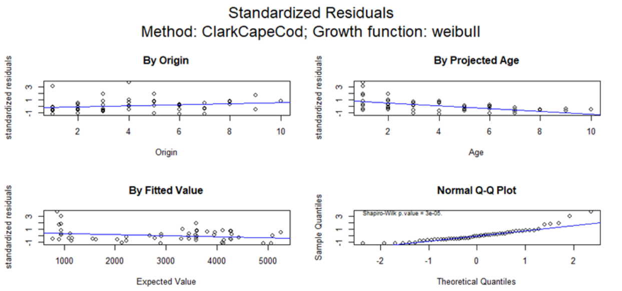

Fig. 7: Cape Cod standardized residuals

3.2 Munich Chain Ladder Method

We usually detect a considerable correlation between the paid-to-incurred ratios and the corresponding paid and incurred individual development factors [3]. Consider a fixed development year of the data triangle; in accident years with a previous paid-to-incurred ratio that is below average, we always see above-average paid development factors and/or below-average incurred development factors. With below-average paid and above-average incurred development variables, accident years with an above-average paid-to-incurred ratio indicate the reverse pattern. This is to be expected, and residual charts may be used to confirm it [3].

The triangles used are the cumulative triangle of the paid claims (Fig. 8) and the cumulative triangle of incurred claims from 2015 to 2021 (Fig. 9)

	1	2	3	4	5	6	7
2015	92,415,152	170,490,856	187,362,541	190,188,420	191,658,420	191,697,620	191,800,056
2016	109,733,734	188,226,779	198,467,294	201,082,954	219,132,954	220,197,336	
2017	116,159,869	188,124,038	196,327,787	200,127,787	207,625,566		
2018	112,281,029	179,665,989	200,510,107	205,500,330			
2019	137,313,347	224,641,995	234,129,236				
2020	113,195,095	149,124,606					
2021	135,245,016						

Fig. 8: Triangle of cumulative claims paid

	1	2	3	4	5	6	7
2015	161,556,506	200,198,640	214,379,971	218,999,971	219,907,171	220,344,171	220,783,371
2016	96,662,685	106,874,242	124,512,808	128,312,808	128,612,808	128,634,408	
2017	96,992,904	105,695,304	106,178,304	106,351,104	106,552,304		
2018	103,104,902	114,068,621	117,428,621	117,548,621			
2019	51,274,330	52,954,330	52,965,488				
2020	128,792,520	133,817,587					
2021	211,380,853						

Fig. 9: Triangle of cumulative claims incurred

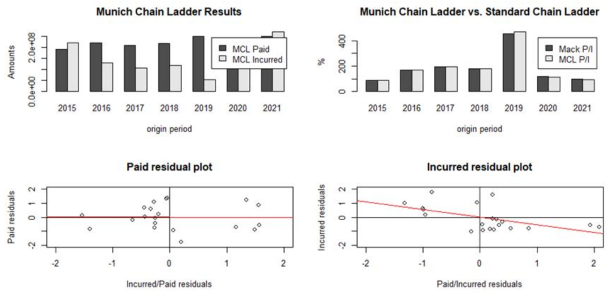

Fig. 10: Paid and incurred residual plots

The Munich chain ladder (MCL) approach integrates paid claims P and incurred claims I data by projecting P/I ratios. The proportion of paid and incurred claims, as well as the fraction of incurred claims that have been paid at the time of calculation, is referred to as the P/I ratio. [3]

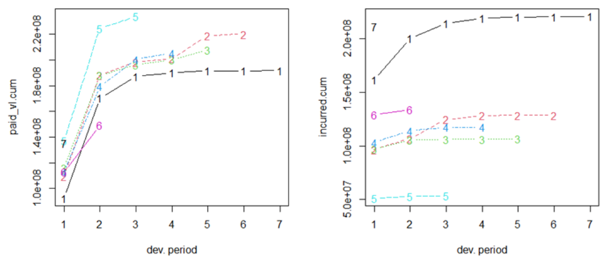

Fig. 11: Paid and incurred development period

The procedure of the Munich Chain Ladder is as follows [3]:

- Plotting the residual plots of paid development and incurred development factors for all development years
- Drawing a regression line from the origin
- We calculate the residual and read the accompanying development factor determined from the average development factors for a specific P/I ratio.

The P/I ratio of the year of the accident to the year of development is [3]:

$$(P/I)_{i,t} = \frac{P_{i,t}}{I_{i,t}} \tag{4}$$

The average ratio for all years of accidents in year t is [3]:

$$(P/I)_{i,t} = \frac{\sum_{j=1}^{n} P_{j,t}}{\sum_{j=1}^{n} I_{j,t}} = \frac{1}{\sum_{j=1}^{n} I_{j,t}} \sum_{j=1}^{n} I_{j,t}\,(P/I)_{j,t} \tag{5}$$

which is the weighted average of the reports (P/I) in the year of development t with the value of claims incurred. The development factors for paid claims and incurred claims are respectively:

$$f_{s\to s+1}^{P} = \frac{\sum_{j=1}^{n-s} P_{j,s+1}}{\sum_{j=1}^{n-s} P_{j,s}} \quad and \quad f_{s\to s+1}^{I} = \frac{\sum_{j=1}^{n-s} I_{j,s+1}}{\sum_{j=1}^{n-s} I_{j,s}} \tag{6}$$

For the projected amounts, we will have [3]:

$$P_{i,s+1} = P_{i,s} f_{s\to s+1}^{P} \quad and \quad I_{i,s+1} = I_{i,s} f_{s\to s+1}^{I} \tag{7}$$

The MCL technique takes advantage of the historical connection between paid and incurred claims to determine the extent to which they have happened. It generates a paid and incurred prognosis based on the information available. The MCL approach gives the same result as the Standard Chain Ladder when the correlations between paid and incurred claims are not substantial [3]. The results of the projections are the paid and the incurred quadrangle.

	1	2	3	4	5	6	7
2015	92,415,152	170,490,856	187,362,541	190,188,420	191,658,420	191,697,620	191,800,056
2016	109,733,734	188,226,779	198,467,294	201,082,954	219,132,954	220,197,336	220,315,008
2017	116,159,869	188,124,038	196,327,787	200,127,787	207,625,566	208,186,234	208,297,489
2018	112,281,029	179,665,989	200,510,107	205,500,330	214,916,453	215,496,337	215,611,498
2019	137,313,347	224,641,995	234,129,236	23,839,545	249,396,961	250,075,102	250,208,752
2020	113,195,095	149,124,606	159,391,590	162,285,116	169,670,161	170,124,557	170,215,465
2021	135,245,016	218,318,876	33,318,391	237,548,803	248,309,405	248,971,100	249,104,134

Fig. 12: Projection of claims paid

	1	2	3	4	5	6	7
2015	161,556,506	200,198,640	214,379,971	218,999,971	219,907,171	220,344,171	220,783,371
2016	96,662,685	106,874,242	124,512,808	128,312,808	128,612,808	128,634,408	128,890,314
2017	96,992,904	105,695,304	106,178,304	106,351,104	106,552,304	106,614,464	106,826,378
2018	103,104,902	114,068,621	117,428,621	117,548,621	117,854,568	117,941,622	118,176,158
2019	51,274,330	52,954,330	52,965,488	53,069,706	53,115,742	53,183,960	53,494,330
2020	128,792,520	133,817,587	144,219,209	147,067,485	14,758,500	147,817,574	148,112,242
2021	211,380,853	241,683,162	262,293,591	267,915,617	268,940,932	269,440,368	269,977,927

Fig. 13: Projection of claims incurred

Estimation of the MCL parameters

Table 7. MCL method latest and ultimate claims and ratios

Year	Latest Paid	Latest incurred	Latest P/I Ratio	UltimatePaid	Ultimate incurred	Ultimate P/I Ratio
2015	191,800,056	220,783,371	0.8687	191,800,056	220,783,371	0.8687
2016	220,197,336	128,634,408	1.7118	220,315,008	128,890,314	1.7093
2017	207,625,566	106,552,304	1.9486	208,297,489	106,826,378	1.9499
2018	205,500,330	117,548,621	1.7482	215,611,498	118,176,158	1.8245
2019	234,129,236	53,494,330	4.3767	250,208,752	53,069,706	4.7147
2020	149,124,606	133,817,587	1.1144	170,215,465	148,112,242	1.1492
2021	135,245,016	211,380,853	0.6398	249,104,134	269,977,927	0.9227
Total	*1,343,622,146*	*972,211,474*	*1.3820*	*1,505,552,402*	*1,045,836,096*	*1.4396*

4 Conclusion

The scope of this paper is the analysis of distribution for the claims data and the best estimates method for the claims reserves. The data analyzed are the claims incurred and paid by the DMTPL Albanian market. We noted that the best distribution that fits incurred and paid claims are the Weibull distribution. Usually, the Cape Cod method has a smaller variance than the LDF method.

Table 8: Summary of Clark's techniques

Clark's	Current value	Utimate value	Future value	Standard Error	CV%
LDF	1,343,622,146	1,317,409,914	156,432,278	52,873,148	33.80
Cape Cod	1,343,622,146	1,516,767,867	173,145,721	45,984,943	26.50

The Cape Cod method requires the estimation of three parameters. The LDF method requires the estimation of n+2 parameters. As a result of this, CC method is easier even sometimes may have a higher process variance estimated, but it will produce a smaller estimation error.

The Munich Chain Ladder method considers the correlation between paid claims and incurred claims. The Munich chain ladder seeks to resolve the differences that arise between the standard paid claims and the incurred chain ladder indications. MCL provides separate estimations for paid and incurred, but they are closer to one another. In the cases where the correlations are not significant, the MCL method provides the same results as the Standard Chain Ladder method.

Table 9. Summary of MCL method

MCL	Paid	Incurred	P/I Ratio
Latest	1,343,622,146	972,211,474	1.38203
Ultimate	1,505,552,402	1,045,836,096	1.43957

References:

[1] T. Mack, [1997] Measuring the variability of chain–ladder reserve estimates. In: Claims Reserving Manual, vol. 2. London: Institute of Actuaries.

[2] David R. Clark, [2003] LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach

[3] G. Quarg, T. Mack [2008] Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses. In: CAS, Volume 2, Issue 2, pp. 266 -299

[4] P. England, R.Verrall, [1999] Analytic and Bootstrap estimates of prediction error in claims reserving. Insurance: Mathematics and Economics 25, 281-293

[5] M.V. Wüthrich, M. Merz [2013] Financial Modeling, Actuarial Valuation and Solvency in Insurance, Springer

[6] M.V. Wüthrich, M. Merz [2008] Stochastic Claims Reserving Methods in Insurance, Wiley

[7] Moro, M. G., ChainLadder: Statistical Methods and Models for Claims Reserving in General, Retrieved from https://CRAN.R-project.org/package=ChainLadder,2022

[8] Boenn, M. fitteR: Fit Hundreds of Theoretical Distributions to Empirical Data, 2017

[9] Team, R. C, R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved 01 24, 2022, from https://www.R-project.org/

[10] "Handbook on Loss Reserving", Springer, Science and Business Media LLC, 2016

[11] Martinek. L.: Analysis of Stochastic Reserving Models by Means of NAIC Claims Data, Risks, 2019

[12] Gesmann, M.: Claims Reserving and IBNR. Computational Actuarial Science with R (2014) Chapman and Hall/CRC.

[13] www.casact.org

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

-Endri Raço carried out the simulations in R and graph plotting.
-Kleida Haxhi has worked on paper structure, algorithm choice, and conclusions.
-Etleva Llagami has organized and executed the experiments of Section 3.1.
-Oriana Zaçaj has organized and executed the experiments of Section 3.2.