

# The Analysis of Motor Vehicle Insurance Claim Reserve Using Robust Chain Ladder

Desnu Anggara Suwardi <sup>1,\*</sup> Yogo Purwono <sup>2</sup>

<sup>1</sup> Universitas Pendidikan Indonesia

<sup>2</sup> Universitas Pendidikan Indonesia

\*Corresponding author. Email: [desnu.anggara@gmail.com](mailto:desnu.anggara@gmail.com)

## ABSTRACT

In a non-life insurance business, an insurer often needs to build up a reserve to ensure the company can fulfill its obligation. Chain ladder is one of the most widely used methods in claim reserving. However, the chain ladder method is very vulnerable to an outlier. This study focused on claim reserving that was resistant to outlier data by using a robust chain ladder. There are two steps to robustify the chain ladder method. The first step is to detect outlier by using the median as a development factor to compute the residual and adjust the outlying values. The second step is to apply a classic chain ladder method to the adjusted data. This study shows that a robust chain ladder has a better result than a standard chain ladder method.

**Keywords:** Insurance, Insurance Claim, Robust Chain Ladder.

## 1. INTRODUCTION

As the recipient of the risk, the insurance company needs to ensure that the company can fulfill its obligations in paying the claims of its customers and cover the company's daily operating costs. General insurance companies have long been interested in the valuation of liabilities in long-term business, as in the quantification of liabilities in general. In almost all general insurance business lines, coverage is usually limited to a period of 12 months where the problem of premium rates tends to be associated proportionally with the experience of claims so that claims reserves are a direct and continuous concern.

There are two types of outstanding claims, namely Incurred but Not Reported (IBNR) and Reported but Not Settled (RBNS) [1]. If the claim data is compiled based on the time of occurrence as a column and the delay time as a row, then the upper triangular matrix form is obtained, which contains information on the cumulative claim data observed. This matrix is commonly called the run-off triangle.

The claim reserves discussed in this study focused on reserves for claims occurring but not yet reported (Incurred But Not Reported: IBNR). Each of these reserves is under the heading "technical reserves" and can

be referred to as "provisions" in terms of accounting. The reserves are intended to represent liabilities of insurance companies arising from events that have occurred within the scope and during the insurance coverage period.

In this study, we chose to calculate the reserve of motor vehicle insurance. We chose motor vehicle insurance because, based on Otoritas Jasa Keuangan (OJK) data, motor vehicle insurance is the insurance product that contributes the most significant premium to the general insurance sector, especially in PT QWE.

One of the widely used methods in calculating claim reserve is Chain Ladder Method. Reference [2] first discovered Chain-Ladder, then developed by [3]. However, The classic chain-ladder method is theoretically susceptible to outliers because it uses mean as a development factor. To overcome this, robustification methods can be applied so that the chain ladder becomes more resistant to the outliers data. The robust method applied to the Chain ladder was introduced by [4]. This method has been developed to minimize the impact of these outliers on the Chain Ladder Method.

### 1.1 Claim

A claim is a guarantee of the risk or damage that occurs by the insurance company to the insurance

participant following the agreement of the policy [5]. Determining the estimated amount of money that an insurance company must prepare in fulfilling its obligations to pay claims that occur in the future is an essential task for insurance companies. This obligation is often referred to as claim reserves [6].

There are several stages in the claim reserving process. This procedure's initial stage is a simplification, where data sets from various dimensions are lowered to a much simpler dimension. The next stage is a statistical estimation. When the claim reserving model matches the model structure and parameter estimation used. If the model becomes complex, the initial process of simplification should be reviewed. The last stage is the calculation of estimated reserve reserves (loss reserve).

**1.2 Classic chain ladder methods**

Generally, the interpretation of outstanding claim liability for insurance long tail is based on the triangle run-off data. Run-off triangle data contains an overview of the overall claim (aggregate) and summarizes the data of a collection of individual claims. Data in the run-off triangle is usually the size of claims and the number of claims, both of which are presented in incremental or cumulative forms equation 1 are follow:

$$1 \leq i \leq n \quad 1 \leq j \leq n \tag{1}$$

Illustration of the run-off triangle and future triangle can be seen in Table 1, where the data is incremental. The lines in the table show the event period, and the column shows the period of delay, while the main diagonal shows the limitation of claims in each payment period, Table 1 as follow.

**Table 1.** Incremental run-off triangle

Accident year	Development year				
	1	2	...	n-1	N
1	$D_{1,1}$	$D_{1,2}$	...	$D_{1,n-1}$	$D_{1,n}$
2	$D_{2,1}$	$D_{2,2}$	...	$D_{2,n-1}$	$D_{2,n}$
M	M	M	O	M	M
n-1	$D_{n-1,1}$	$D_{n-1,2}$	...	$D_{n-1,n-1}$	$D_{n-1,n}$
n	$D_{n,1}$	$D_{n,2}$	...	$D_{n,n-1}$	$D_{n,n}$

From the incremental data table, it can be formed into a table with cumulative data with the following equation (2).

$$CD_{i,n} = \sum_{k=1}^n \tag{2}$$

Outstanding claims liability for the year i was defined by the following equation (3)

$$R_i = \sum_{k=n+2-i}^n D_{i,k} \tag{3}$$

Chain Ladder techniques consist of estimation of  $\lambda_j$  by using equation (4).

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{i,j}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} \tag{4}$$

and the estimated total amount of the claim is equation (5) as follow.

$$C_{n-j+1,n} = C_{n-i+1,j} \lambda_{j+1} \lambda_{j+2} \dots \lambda_n \tag{5}$$

**1.3 Sensivity chain ladder methods to outlier**

In real data, one or several observations are often found to deviate from other observations called outliers. This outlier can occur because of an error in recording data, or the value is indeed true. Table 2 shows the illustration that the chain ladder method is susceptible to an outlier.

**Table 2.** Chain ladder without outlier

Development year						R
	1	2	3	4	5	
1	357,848	766,940	610,542	4,829,400	527,326	
2	352,118	884,021	933,894	1,183,289	269,362	269,362
3	290,507	1,001,799	926,219	3,415,637	452,576	3,868,213
4	310,608	1,108,250	959,653	3,661,813	485,194	5,106,571
5	443,160	1271263	1,159,452	4,424,613	586,266	7,441,595
<b>Total Claim Reserving</b>						16,685,741

Suppose there is one outlier, that is, in the year of the 1st event and the 5th development period by multiplying 10 claim values, the following results are obtained. Table 3 show chain ladder with outlier as follow.

**Table 3.** Chain ladder with outlier

Development year	Development Period					R
	1	2	3	4	5	
1	357.848	766.940	610.542	4.829.400	5.273.260	-
2	352.118	884.021	933.894	1.183.289	2.693.628	2.693.628
3	290.507	1.001.799	926.219	3.415.637	4.525.761	7.941.398
4	310.608	1.108.250	959.653	3.661.813	4.851.947	9.473.324
5	443.160	1271263	1,159,452	4,424,613	5,862,667	12,717,995
<b>Total Claim Reserving</b>						32,826,344

The calculation results in Table 3 show that the outliers have a tremendous impact on the results of Chain Ladder's calculation. This is because the calculation of Chain Ladder calculations is based on cumulative claims data so that if there are outliers in the 1st or 2nd year of delay, it will have a very large impact because the cumulative data in the 3rd, 4th, and so on delays will increase significantly so that the value of the development factor increases considerably. Increasing

the value of this development factor also increases the estimation of claim reserves.

**2. METHODS**

Unlike the Classic Chain Ladder, this development factor in the robust chain ladder is estimated using the equation (6) introduced by [4].

$$f_j^{\%} = median \left\{ \frac{X_{i,j}}{X_{i,j-1}} \mid 1=1, \dots, n-j+1 \right\}, 2 \leq j \leq n \tag{6}$$

Median is used to estimate development factors because, unlike the mean, the median is not affected by outliers.

After obtaining a development factor, a residual Pearson is formed. This Pearson residual will be used to detect which data from the triangle run-off is an outlier. Residual Pearson can be obtained by an equation (7).

$$r_{ij} = \frac{X_{ij} - E(X_{ij})}{\sqrt{Var(X_{i,j})}} \tag{7}$$

This Pearson residual will then be tested whether it contains an outlier or not. Boxplot used to detect outlier in the Pearson residual, show by equation (8).

$$[Q_1 - 3IQR, Q_3 + 3IQR] \tag{8}$$

If there is a Pearson residual outside the interval, then the claim data that conforms to the row and column with the Pearson residual must be modified so that the calculation of claim reserves is not biased. Testing outliers using this boxplot must meet the normality requirements of residual peers. If the Pearson residual does not have a normal distribution, then the outlier test is used adjusted boxplot introduced by [7]. Shapiro Wilk test or Kolmogorov-Smirnov test can be used to test the normality of Pearson Residual.

Suppose there is data in development period 1, which is an outlier, modification of the data needs to be done. This modification depends on the column of data in the second column on the same row. If the data in column 2 in the same row is also an outlier, then the data modified using this equation (9).

$$C_{k,1} = median \{ C_{i,j} \mid 1=1, \dots, n-j+1, 2 \leq j \leq n \} \tag{9}$$

However, if the data in column 2 in the same row, not an outlier, then this equation (10) is used.

$$C_{k,1} = \frac{X_{k,2}}{median \left\{ \frac{X_{i,j}}{X_{i,j-1}} \mid 1=1, \dots, n-j+1 \right\}}, 2 \leq j \leq n \tag{10}$$

Pearson residuals will be formed like the previous process to detect outliers in other columns, but the incremental fitted value obtain by using the equation (11).

$$\hat{X}_{ij}^1 = X_{i1} \hat{f}_j^1, \quad 2 \leq i \leq n, n-i+2 \leq j \leq n \tag{11}$$

With equation (12)

$$f_j^{\%} = median \left\{ \frac{X_{i,j}}{X_{i,1}} \mid 1=1, \dots, n-j+1 \right\}, 2 \leq j \leq n \tag{12}$$

This incremental fitted value is used to calculate the Pearson residual. Then we applied the normality test to the Pearson residual to determine which method will be used to detecting the outlier.

If there are no outliers detected, then the incremental data that has been modified in the previous stage can be directly used to find claims reserves with the Classic Chain Ladder method. If there are still outliers, the Pearson residual, which is the outlier, is modified to the residual median. Then the modified residual is used to find incremental value X using the equation (13).

$$X_{ij} = r_{ij} \sqrt{Var(X_{i,j})} + E(X_{ij}) \tag{13}$$

By applying Classic Chain Ladder Method to These new incremental values, we can find the Claim Reserve resistant to the outlier.

**3. RESULT AND DISCUSSION**

In this study, the data used are data on motor vehicle insurance claims that have been paid by the insurance company "QWE" from 2010 to 2016. The company "QWE" is a general insurance company with many products from various insurance business lines. The data used is data that has been cleared so that the data can be directly used for this research. To maintain company data confidentiality, the author does not specify the amount of claims that have occurred and paid by the company. The data used are then grouped based on the accident year and development year to form an incremental run-off triangle.

To determine the claim reserves that PT QWE must prepare, two methods will be used, including the Chain Ladder method and Robust Chain ladder method. To determine the method to be used for claim back-up, these methods are tested on the existing claim reserve data sample by cutting PT QWE data taken from the year 2010 1st quarter up to the year of 2013 2nd quarter and 10 years delay. The sample data is then divided into only the top triangles, namely as many as 55 data, while the other 45 data will be used as a reference to determine which method the results are closest to the total of the 45 data.

The Incremental run-off table used as a sample to test the method will be shown in the Table 4.

**Table 4. Incremental Run-off triangle Incurred claim**

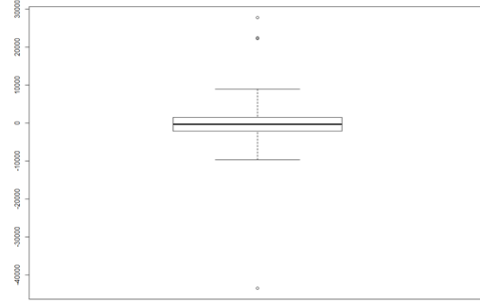
Accident years	Development										
	1	2	3	4	5	6	7	8	9	10	
1	5.524. 559.71 8	16.69 8.631. 367	20.07 9.217. 436	20.06 0.684. 229	20.07 6.173. 699	20.15 4.001. 504	20.19 3.967. 298	20.19 3.967. 298	20.26 1.367. 298	20.25 1.367. 298	
2	5.856. 832.84 6	35.83 6.639. 991	25.23 6.800. 025	15.81 1.716. 187	15.83 5.152. 276	15.84 8.232. 926	16.11 3.501. 036	16.11 3.501. 036	16.14 4.882. 936		
3	4.311. 565.58 6	14.60 8.485. 231	17.46 2.438. 299	18.06 9.380. 453	18.44 4.882. 089	18.73 2.725. 837	18.85 2.882. 762	18.84 2.882. 762			
4	7.347. 554.04 1	15.64 1.841. 432	18.15 2.291. 772	19.43 0.888. 989	19.62 4.465. 103	19.95 7.222. 298	20.07 7.099. 467				
5	3.764. 845.98 1	24.66 0.875. 072	17.98 4.785. 343	18.87 8.404. 444	19.30 3.613. 087	19.75 2.959. 230					
6	4.219. 898.68 2	33.726.6 70.415	19.320.6 87.326	21.617.5 04.394	22.481 932.830						
7	4.050. 236.93 7	19.704.0 26.086	25.639.9 76.189	18.048.0 81.828							
8	8.074. 779.96 5	22.400.9 38.575	29.496.5 56.389								
9	5.934. 357.01 8	24.526.6 82.977									
10	5.617. 944.96 8										

By using the median as the development factor, the Pearson residual is obtained. Table 5. shown pearson residual 1<sup>st</sup> iteration.

**Table 5. Pearson residual 1<sup>st</sup> iteration**

Accident years	Development										
	1	2	3	4	5	6	7	8	9	10	
1	1995 6.37	1439. 45	- 1521. 73	- 2012. 97	- 2517. 81	- 522.2 6	- 1658. 37	- 2156. 63	- 958.3 5	0.00	
2	1410 2.04	- 1287. 92	- 3164. 34	- 1676. 17	- 1803. 88	- 390.7 8	322.1 6	- 120.3 8	844.1 2		
3	- 4611. 43	1366. 56	- 4665. 52	- 2813. 10	- 551.9 1	3559. 91	2995. 89	2521. 85			
4	3874 73.35	- 1182. 74	- 1002 9.35	- 4591. 81	- 3556. 42	98.51 -	- 496.0 3				
5	- 1208 3.69	- 2428. 01	- 4202. 58	- 1149. 67	3391. 61	7891. 15					
6	- 9457. 65	- 1678 2.17	- 2996. 36	8834. 87	1387 5.40						
7	- 2469 1.96	- 4536. 96	- 2787. 71	1374 6.49							
8	7659. 20	- 8014. 64	3344. 92								
9	- 1643 1.52	9260. 04									
10	0.00										

The normality test for these development factors is rejected. Then we used the adjusted boxplot to find the outlier data. Fig 1 show the adjusted boxplot first iteration residual as follow.



**Figure 1. Adjusted boxplot first iteration residual**

We found five outliers from this adjusted boxplot. These outliers located in r3,1,r7,1,r9,1, and r9,2.

Note that in column 1, there are three outliers, then the incremental data in column-1 that corresponds to the outlier needs to be modified. This modification depends on the residual data in column 2 in the same year of occurrence, in the outliers detected there are one outliers located in column-2 that has a same row with outlier in column 1, so the out-liers in column 1 will be modified into X<sub>9,1</sub> = 6,547,442,053.25, X<sub>3,1</sub> = 5,868,897,676.53, and X<sub>7,1</sub> = 6,893,901,405.26.

Then we calculate the Incremental value based on column 1, which the outlier already modified. Using these incremental fitted values then we can find the Pearson residual for the second iteration. Table 6. shown pearson residual 2<sup>nd</sup> iteration.

**Table 6. Pearson Residual 2<sup>nd</sup> iteration**

Accident years	Development									
	1	2	3	4	5	6	7	8	9	10
1	0.00	0.00	8580. 64	0.00	- 2240. 26	0.00	9183. 75	1908 8.56	3640 8.84	0.00
2	0.00	7475 3.12	- 1304 9.64	- 7780 6.33	- 7985 5.41	- 7820 3.76	- 6838 2.86	- 5927 1.17	- 4383 5.95	
3	0.00	- 2405 5.70	- 1905 4.03	- 2281 4.62	- 2250 7.28	- 1891 7.54	- 9487. 01	0.00		
4	0.00	0.00	3372. 94	4667. 07	3755. 83	7850. 76	1760 8.00			
5	0.00	7295 7.89	2816. 70	1356. 63	2162. 54	7100. 72				
6	0.00	1233 18.03	- 2982. 07	4595. 46	8079. 76					
7	0.00	0.00	2235 8.17	- 3654 8.15						
8	0.00	- 6006 0.94	- 4257 8.99							
9	0.00	0.00								
10	0.00									

The normality test for these residuals was also rejected. The outlier was then detected by using the adjusted boxplot shown in Fig. 2

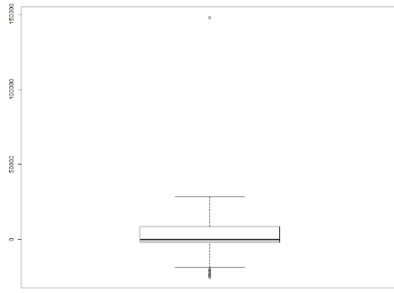


Figure 2. Adjusted boxplot second iteration residual

From figure 2, we found six outlier  $r^{1,2}$ ,  $r^{1,9,2}$ ,  $r^{1,3}$ ,  $r^{1,4}$ ,  $r^{1,5}$ ,  $r^{1,7}$  and  $r^{1,8}$ . These Pearson residuals are then replaced by the median of the Pearson residual. By converting these Pearson residuals into incremental values, we can find the claim reserve by applying the Chain Ladder method to these modified incremental values.

The results of the robust chain ladder will be compared with the chain ladder method and real data. The result of the comparison is shown in Table 7.

Table 7. Comparison Chain Ladder, Robust Chain Ladder, and Real Data

Real Data	Classic Chain Ladder	Robust Chain Ladder
32,001,106,631.00	31,651,136,663.01	28,432,098,867.44
68,926,943,232.00	70,516,346,137.42	64,084,470,057.88
101,029,864,851.00	105,220,984,910.79	99,928,645,850.22
150,822,591,121.00	152,675,614,422.43	147,225,075,266.14
166,258,503,800.00	168,923,409,912.77	162,809,379,781.11
216,518,587,325.00	234,841,920,026.12	224,129,932,902.15
269,882,898,038.00	257,148,302,994.74	251,669,323,633.66
322,360,127,065.00	502,025,588,968.78	284,977,540,515.80
351,104,243,013.00	431,327,416,229.92	424,569,769,753.46
1,678,904,865,076.00	1,954,330,720,265.99	1,687,826,236,627.85



Figure 3. Comparison Chain Ladder, Robust Chain Ladder, and Real Data

To further ascertain whether it is better between the classic chain ladder and the robust chain ladder, the baseline estimation value of RMSE and MAPE will be calculated from the results of the two methods compared to the actual data in each year of occurrence, and the following result is shown in Table 5.

The RMSE and MAPE calculation, the two methods show that the difference in the classical chain ladder method and the real data smaller than the robust chain ladder in the second accident year and the third accident year, for the 4th accident year to the 10th accident year and total, the difference in robust chain ladder and actual data is smaller than the classic ladder chain. Overall, the RMSE value and MAPE value of the robust chain ladder are smaller than the classic ladder chain method. This means that based on RMSE and MAPE, the robust chain ladder method is better. Table 8. RMSE and MAPE classic chain ladder and robust chain ladder as follow.

Table 8. RMSE and MAPE Classic Chain Ladder and Robust Chain Ladder

	Chain Ladder	Robust Chain Ladder
RMSE	6852972685	5862932123
MAPE	0.02003	0.01725

From the comparison of the two methods, both graphically and RMSE & MAPE values, it can be seen that the Robust chain ladder method has a better result than the classic chain ladder. At first glance, it appears that the difference between RMSE and MAPE is not too large. However, if we look at the total reserve claims that PT QWE must prepare through the classic chain ladder method and the robust chain ladder against the real data, the total error of the reserve claim error from the two methods has a big difference shown in the following Table 9.

Table 9. The error of total claim reserving

Method	Error
Classic Chain ladder	16.41%
Robust chain ladder	0.53%

#### 4. CONCLUSION

The Robust Chain Ladder method is quite appropriate for use in estimating IBNR claim reserves in PT QWE. The Robust Chain Ladder method has a better result than the classic Chain Ladder method. It can be seen from the results of claim reserve using the Robust Chain Ladder closer to the actual value compared to the Classic Chain Ladder method, which means that the Robust Chain Ladder method can be used as an alternative reserving method, especially PT QWE.

Outlier detection in the Robust Chain Ladder Method can be used as a reference to verify and check the data. By knowing where the outliers are, it easier and more efficient to find out the cause of the outliers, whether there are errors in inputting data or indeed the data outliers are true because there are certain phenomena in certain years and development year such as natural disasters that impact on claim data.

Other methods can be used to detect outliers for further research related to the claim reserving with

outliers. Besides using the median, other robust statistical methods can be used for further research, including other robust estimators such as Mestimators.

## REFERENCES

- [1] I. B. Hossack, J. H. Pollard, and B. Zehnwirth, *Introductory Statistics with Applications in General Insurance*. 1999.
- [2] E. Kremer, "IBNR-claims and the two-way model of anova," *Scand. Actuar. J.*, 1982.
- [3] P.D. England, and R.J. Verrall, "Analytic and bootstrap estimates of prediction errors in claim reserving", *Insurance: Mathematics and Economics* Elsevier., vol. 25, no. 3, pp. 281-293, 1999.
- [4] T. Verdonck, M. Van Wouwe, and J. Dhaene, "A robustification of the chain-ladder method," *North Am. Actuar. J.*, 2009.
- [5] J. Bower, "Effective Public Management.," *Harvard Business Review.*, 1977.
- [6] M. Olofson, "Stochastic loss reserving testing the new guidelines from the australian prudential regulation authority (apra) on swedish portfolio data using a bootstrap simulation and the distribution-free method by Thomas Mack," Sweden, 2006.
- [7] M. Hubert and S. Van Der Veeken, "Outlier detection for skewed data," in *Journal of Chemometrics*, 2008.