**Examination of the impact of using shorter time periods to collate data into General Insurance Run-off triangles and Comparison with Curve fitting techniques**

MSc Dissertation

**Rubhan Antony Xavier Raj**

**1 September 2017**

**Supervisor: Mr. David Hargreaves**

The dissertation is submitted as part of the requirements for the award of

the MSc in Actuarial Science

**Cass Business School**

**City, University of London**

Abstract

Reserving is one of the complex tasks performed in an insurance company. Usually, the data is collated into yearly bins, while calculating reserves by using Chain Ladder Method. In this project, the impact of using shorter time periods (half-yearly, Quarterly and monthly) to collate data into general insurance run-off triangles is examined. And then it is compared with the Actual Simulated Reserve and with the Curve fitting techniques and is critically analysed. Computer generated data from statistical distributions are collated into Yearly, Half-yearly, Quarterly and Monthly bins and reserves are calculated for each type of collation. On analysing the output, it is observed that the standard deviation of the reserve calculations increases, if we collate data into shorter time periods in Chain Ladder Method. But it gives a very clear picture of the progression of the claims by year segments. It is also inferred that the more the data available, the accurate the prediction is. And in the second part of the project, the same claims data that was generated early is used and a curve is fitted to it and its accuracy is examined. Curves are fitted to both Incremental and Cumulative claims data. It is observed that collation of data into shorter time periods provides more details for cure fitting and collation into shorter time periods provide the best fit. Fitting curves to cumulative data rather than incremental data provides better results.

Acknowledgements

This Dissertation is one of the most interesting works I have done in my course. I would like to thank **Mr. David Hargreaves** for guiding me all through the dissertation and for providing vital help to bring this dissertation into good shape. And I would like to thank **Professor. Jens Perch Nielson** for motivating and guiding me to choose this topic.

**Contents**

**List of Figures**

**List of Tables**

1 [Introduction 8](#_Toc492129062)

2 [Chain Ladder Method 8](#_Toc492129065)

2.1-[Data requirements for Chain Ladder Method 8](#_Toc492129066)

2.2-[Incremental Triangle 11](#_Toc492129067)

2.3-[Cumulative Triangle 12](#_Toc492129068)

2.4-[Forward Factors 12](#_Toc492129069)

2.5-[Determination of cumulative claim loss settlements 13](#_Toc492129070)

2.6-[Finding Reserves 14](#_Toc492129071)

2.7-[Comparison with simulated reserve 16](#_Toc492129072)

2.8-[Shortening the time periods to collate data in run of triangles 17](#_Toc492129073)

2.8.1-[Half – Yearly Reserve 17](#_Toc492129074)

2.8.2-[Quarterly Reserves 19](#_Toc492129075)

2.8.3-[Monthly Reserves 21](#_Toc492129076)

2.9-[Comparison of Reserves (10,000 Claims) 22](#_Toc492129078)

2.10-[Comparison of Reserves (1,000 claims) 25](#_Toc492129079)

2.11-[Inference 27](#_Toc492129080)

3 [Curve Fitting 28](#_Toc492129081)

3.1-[Fitting curves to the tail of Incremental Data 28](#_Toc492129082)

3.2-[Quadratic equations and Cubic equations can’t be used 30](#_Toc492129083)

3.3-[Craighead Curve Fitting Method 32](#_Toc492129084)

3.3.1-[Effect of varying “c” speed factor on claims ratio 32](#_Toc492129085)

3.3.2-[Effect of varying “b” tail factor on claims ratio 33](#_Toc492129086)

3.3.3-[Curve fitting process 34](#_Toc492129087)

3.4-[Fitting Curves to Cumulative data 35](#_Toc492129088)

3.4.1-[Algorithm (Curve Fitting - Cumulative claims) 36](#_Toc492129089)

3.5-[Limitations of Curve fitting 38](#_Toc492129090)

3.6-[Inference 38](#_Toc492129091)

4 [Conclusion 40](#_Toc492129092)

5 [Appendix 41](#_Toc492129094)

6 [Bibliography 42](#_Toc492129095)

List of Figures

[Table 1 Ten Claim Amounts and their Accident Year and Development Year 9](#_Toc492129096)

[Table 2 Data Generation-Technical Specifications 10](#_Toc492129097)

[Table 3 Projections in Incremental Data 11](#_Toc492129098)

[Table 4 Projections in Cumulative Data 12](#_Toc492129099)

[Table 5 Projections in Cumulative Data 13](#_Toc492129100)

[Table 6 Forward Factors 13](#_Toc492129101)

[Table 7 Projections in Cumulative Data (filled) 14](#_Toc492129102)

[Table 8 Projections in Incremental Data (filled) 15](#_Toc492129103)

[Table 9 Simulated Reserve 16](#_Toc492129104)

[Table 10 Half Yearly reserves 17](#_Toc492129105)

[Table 11 Quarterly Reserves 19](#_Toc492129106)

[Table 12 Average and Std. Deviation of 100 simulations (10,000 Claims) 23](#_Toc492129107)

[Table 13 Error in Reserves (10,000 Claims) 24](#_Toc492129108)

[Table 14 Average and Standard Deviation of 100 simulations (1,000 Claims) 26](#_Toc492129109)

[Table 15 Error in Reserves (1,000 Claims) 27](#_Toc492129110)

[Table 16 Paid Claims of 2002 (10,000 Claims) 29](#_Toc492129111)

[Table 17 Difference between values 29](#_Toc492129112)

[Table 18 Effect of Varying parameter “c” 33](#_Toc492129113)

[Table 19 Effect of varying parameter “b” 33](#_Toc492129114)

List of Tables

[Table 1 Ten Claim Amounts and their Accident Year and Development Year 9](#_Toc492129115)

[Table 2 Data Generation-Technical Specifications 10](#_Toc492129116)

[Table 3 Projections in Incremental Data 11](#_Toc492129117)

[Table 4 Projections in Cumulative Data 12](#_Toc492129118)

[Table 5 Projections in Cumulative Data 13](#_Toc492129119)

[Table 6 Forward Factors 13](#_Toc492129120)

[Table 7 Projections in Cumulative Data (filled) 14](#_Toc492129121)

[Table 8 Projections in Incremental Data (filled) 15](#_Toc492129122)

[Table 9 Simulated Reserve 16](#_Toc492129123)

[Table 10 Half Yearly reserves 17](#_Toc492129124)

[Table 11 Quarterly Reserves 19](#_Toc492129125)

[Table 12 Average and Std. Deviation of 100 simulations (10,000 Claims) 23](#_Toc492129126)

[Table 13 Error in Reserves (10,000 Claims) 24](#_Toc492129127)

[Table 14 Average and Standard Deviation of 100 simulations (1,000 Claims) 26](#_Toc492129128)

[Table 15 Error in Reserves (1,000 Claims) 27](#_Toc492129129)

[Table 16 Paid Claims of 2002 (10,000 Claims) 29](#_Toc492129130)

[Table 17 Difference between values 29](#_Toc492129131)

[Table 18 Effect of Varying parameter “c” 33](#_Toc492129132)

[Table 19 Effect of varying parameter “b” 33](#_Toc492129133)

# 

1 Introduction

Methods of calculating reserves in General Insurance are different from those used in Life Insurance, Health Insurance and Pensions and benefits since General Insurance contracts are generally for shorter duration. And the premiums are mostly paid only once at the start of the contract. Typical travel insurance policies last for only few days to few weeks (En.wikipedia.org, 2017). Thus, analysis of shorter time periods of policies becomes mandatory and the consequences of using shorter time data to predict far future must be clearly known. To meet these future liabilities and to generate consistent dividends for the shareholders, the insurance companies must hold adequate reserves and this helps the insurance company to remain solvent. The Solvency II, the directive in European Union Law for insurance regulation also demands that every company must hold a certain amount of reserve to do business in the European Market (En.wikipedia.org, 2017). To calculate these reserves, various companies use various reserving techniques. The most popular techniques among them are the Chain Ladder Method and Bornheutter-Ferguson Method.

2 Chain Ladder Method

The Chain Ladder Method is one of the most prominent actuarial loss reserving technique in the insurance industry. It estimates the Incurred But Not Reported Claims (IBNR) and projects Ultimate loss amounts. The chain ladder method is used property and casualty and health insurance fields. The main assumption in this method is that the historic loss development patterns follow in the future as well (En.wikipedia.org, 2017). Under Solvency II, the projection of Run-off triangles is one of the allowed methods for calculating reserves (www.fh-vie.ac.at, 2017).

2.1-Data requirements for Chain Ladder Method

The chain ladder method requires three main information. They are as follows:

1. Claim Amounts
2. Accident Year (Occurrence Year or Reporting Year)
3. Development Year (Settlement Year or Settlement Year)

A small sample for data requirement is shown in the table below:

|  |  |  |
| --- | --- | --- |
| **Accident Year or Reporting Year** | **Development Year** | **Claim Amounts (£)** |
| 29/10/2002 | 17/11/2002 | 2879.16623 |
| 5/1/2005 | 4/2/2005 | 2782.27107 |
| 25/1/2001 | 2/4/2001 | 2728.83535 |
| 26/3/2001 | 7/5/2001 | 2831.11756 |
| 26/12/2003 | 28/1/2004 | 2952.31339 |
| 26/11/2003 | 31/12/2003 | 3005.73209 |
| 10/1/2002 | 15/1/2002 | 2965.24890 |
| 15/10/2001 | 18/3/2002 | 2758.77006 |
| 31/3/2003 | 7/5/2003 | 2924.19438 |

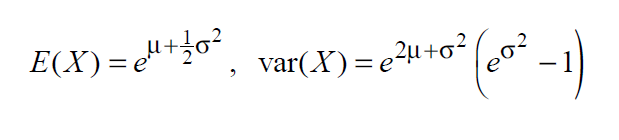
Table 1 Ten Claim Amounts and their Accident Year and Development Year

This is a computer-generated data. The technical specification of the data is as follow:

|  |  |
| --- | --- |
| Accident Year | Uniform Distribution U ~ (0,1). It is generated in excel by using the function RANDBETWEEN(startDate, endDate)  For Example: RANDBETWEEN (“1/1/2001”, “31/12/2005”). In this simulation, I have used a 5 year triangle. |
| Development Year | As the development Year is a date which should be higher than the accident date, a random date is generated from a LogNormal Distribution and is added to the Accident Year.  For Example: Accident Year + L  Where L ~ LogNormal(5.76957, 1.010768).  These values of µ and σ denotes a mean value of 1.5 years (534 days) with a standard deviation of with a standard deviation of 2 years (712 days). More randomness to the settlement date can be added by adjusting values of µ and σ. |
| Claim Amounts | This is generated form a lognormal Distribution.  For Example: Claim Amount C~LogNormal(µ, σ)  The number of data generated depends on a Poisson Random Variable N ~ Pois(λ). Using Poisson distribution for N makes more sense, rather than using a deterministic value, as the Number of Claims received in an insurance company is always random. |

Table 2 Data Generation-Technical Specifications

Thus, the data is being generated for more than 9 years and in those 9 years data, 5 years of data is used. (The 4 years of data is used to fill the lower triangle in the Simulated Data Rectangle, which is used to calculate the Actual simulated reserve). Depending on the requirement of mean and standard deviation, the corresponding µ and σ values can be generated using the formula of Log Normal Distribution,



Equation 1 Mean and Variance of Lognormal Distribution

Where,

E(X) represents Expected Value  
Var(X) represents Variance   
µ, σ represents parameters

2.2-Incremental Triangle

Typically, the claims that are reported are not paid on the same date. After background check and analysis, the insurance company settles the claims with a time delay. The day in which the claim was reported is called as the Occurrence Date or the reporting date and the date in which the claim amount is being settled is called as the Settlement date. The corresponding year of occurrence is called as the **Accident Year or Reporting Year** and the corresponding years taken settlement of the claim are called the **Development Years**. Thus, the generated data can be distributed in a triangular form based on the year of Accident and year of development as follows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 24,35,751 | 8,75,830 | 4,17,449 | 1,78,099 |
| **2002** | 16,23,278 | 25,23,247 | 9,12,021 | 3,59,375 |  |
| **2003** | 17,35,831 | 27,57,961 | 9,48,491 |  |  |
| **2004** | 16,96,850 | 26,81,896 |  |  |  |
| **2005** | 15,68,067 |  |  |  |  |

Table 3 Projections in Incremental Data

The value (1650302) in the row **2001** and the column **1** represents the Ultimate claim amounts reported in the year 2001 and settled in the same year. The value (912021) in the row **2002** and the column **3** represents the Ultimate claim amounts reported in the year 2003 and settled in 2004.

2.3-Cumulative Triangle

Then the cumulative triangle can be created by cumulating the data every year. For example, the claims settled in the year 2001 are added to the claims settled in the year 2002 and this amount is added to 2003 and so on.

In the below table, the value (40,86,053) in the row **2001** and the column **2** equals “1650302” and “4086053” in Table 3. The value (50,58,546) in the row **2003** and the column **3** equals “1735831”, “2757961” and “948491” in Table 3.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 40,86,053 | 49,61,883 | 53,79,332 | 55,57,432 |
| **2002** | 16,23,278 | 41,46,524 | 50,58,546 | 54,17,921 |  |
| **2003** | 17,35,831 | 44,93,792 | 54,42,284 |  |  |
| **2004** | 16,96,850 | 43,78,746 |  |  |  |
| **2005** | 15,68,067 |  |  |  |  |

Table 4 Projections in Cumulative Data

2.4-Forward Factors

Then next step is to find the forward factors using the cumulative triangle. The total cumulative claim amounts in the first column till the second last value is added and is divided the total cumulative claim amounts in the second column. This gives the forward factor f1.

Forward Factor Fj =

Equation 2 Forward Factor

Where,

**i, j** represents row and column  
**n** represents the number of years  
 represents the corresponding claim amount.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 40,86,053 | 49,61,883 | 53,79,332 | 55,57,432 |
| **2002** | 16,23,278 | 41,46,524 | 50,58,546 | 54,17,921 |  |
| **2003** | 17,35,831 | 44,93,792 | 54,42,284 |  |  |
| **2004** | 16,96,850 | 43,78,746 |  |  |  |
| **2005** | 15,68,067 |  |  |  |  |

Table 5 Projections in Cumulative Data

For Example,

Fj = = 2.288568

And thus, the forward factors are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **F1** | **F2** | **F3** | **F4** |
| **Forward Factors** | 2.550619 | 1.215014 | 1.077524 | 1.033108 |

Table 6 Forward Factors

2.5-Determination of cumulative claim loss settlements

The next step is to calculate the cumulative claim loss settlement amounts using the latest available claim loss settlement amount. This is done by using the below formula.

= x Fj

= x

Equation 3 Formula to find cumulative claim amounts

Where,

represents cumulative claim amount, i represents accident year and j represents development year.

represents cumulative claim amounts, i represents accident year and n represents development year.

**Fj**represents development factor or forward factor. The filled table looks like below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 40,86,053 | 49,61,883 | 53,79,332 | 55,57,432 |
| **2002** | 16,23,278 | 41,46,524 | 50,58,546 | 54,17,921 | 55,97,298 |
| **2003** | 17,35,831 | 44,93,792 | 54,42,284 | 58,64,192 | 60,58,344 |
| **2004** | 16,9,6850 | 43,78,746 | 53,20,237 | 57,32,683 | 59,22,481 |
| **2005** | 15,68,067 | 39,99,541 | 48,59,498 | 52,36,226 | 54,09,587 |

Table 7 Projections in Cumulative Data (filled)

2.6-Finding Reserves

To derive the estimated incremental claim settlement amounts, we must find the difference between the two consecutive settlement amounts (www.fh-vie.ac.at, 2017). Thus, by finding the difference back again from column 5 down to column 1, for every consecutive cell, we obtain the incremental settlement amounts as shown below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |  | **Projections by Year** |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 24,35,751 | 8,75,830.5 | 4,17,449.2 | 1,78,099.2 |
| **2002** | 16,23,278 | 25,23,247 | 9,12,021.7 | 3,59,375.6 | 1,79,376.8 | 3974,249 |
| **2003** | 17,35,831 | 27,57,961 | 9,48,491.8 | 4,21,908.2 | 1,941,51.9 | 14,66,554 |
| **2004** | 16,96,850 | 26,81,896 | 9,41,490.5 | 4,12,446.6 | 1,89,797.9 | 5,66,526 |
| **2005** | 15,68,067 | 24,31,474 | 8,59,956.3 | 3,76,728.2 | 1,73,361.2 | 1,73,361 |
|  |  |  |  |  |  |  | | |
|  |  |  |  |  |  | **Reserve** |  | **61,80,691** |

Table 8 Projections in Incremental Data (filled)

In this table, the yearly projections of reserves are obtained by adding the diagonal claim amounts. Thus, the reserve estimate for the year 2006 is 302436.6, 2007 is 75050.29 and so on. The overall reserve estimate is got by the addition of all the yearly projections and the Ultimate estimate is £ 61,80,691 in this case. The graphical representation is shown below:

Figure 1 Paid Claims (Yearly)

2.7-Comparison with simulated reserve

As we have used the Log normal distribution to simulate this data, we use the same parameters to calculate a simulated reserve. Thus, the we obtain a whole rectangle of simulated values, of which, the upper triangle represents the available data till date and the lower triangle represents the estimates for the future years. We can use this reserve to evaluate the efficiency of the Chain ladder method. An example of a fully simulated table is shown below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Development Year** | | | | | |
| **Accident Year** |  | **1** | **2** | **3** | **4** | **5** |
| **2001** | 16,50,302 | 24,35,751 | 8,75,830.5 | 4,17,449.2 | 1,78,099.2 |
| **2002** | 16,23,278 | 25,23,247 | 9,12,021.7 | 3,59,375.6 | 1,95,407.7 |
| **2003** | 17,35,831 | 27,57,961 | 9,48,491.8 | 3,35,465.8 | 2,48,577.9 |
| **2004** | 16,96,850 | 26,93,331 | 9,52,687 | 3,40,085.6 | 2,12,737.2 |
| **2005** | 15,74,518 | 25,74,123 | 8,31,646.9 | 3,45,929.2 | 1,73,195 |
|  |  |  |  |  |  | |
|  |  |  |  |  | **Reserve** | **62,09,856** |

Table 9 Simulated Reserve

In the above table both the upper triangle and the lower triangle are simulated values using the same Lognormal parameters (distribution used in this case). The value of the actual simulated reserve seems different from the CLM reserve. But to understand the true nature of the reserves we shall conduct the experiment 100 times and analyse the average output in later sections.

2.8-Shortening the time periods to collate data in run of triangles

2.8.1-Half – Yearly Reserve

Before repeating the simulation, we shall do the simulation for shorter time periods to collate data. So, we should put the data into bins like Half yearly, Quarterly and Monthly cumulative claim amounts. If the same Chain Ladder Method is followed, the incremental triangle for Half years after calculation of reserves will look like as shown in Table 9. Each cell in the table shows the cumulated claim amounts in the respective half years.

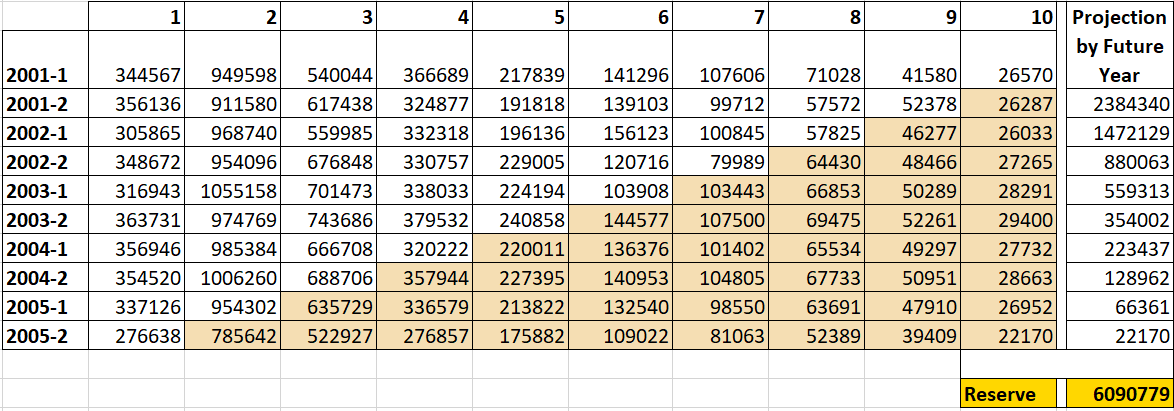


Table 10 Half Yearly reserves

The graphical representation of half-yearly reserves is shown below. This clearly shows the well spread figures in half yearly time span.

Figure 2 Paid Claims (Half Yearly)

**2.8.2-Quarterly Reserves**

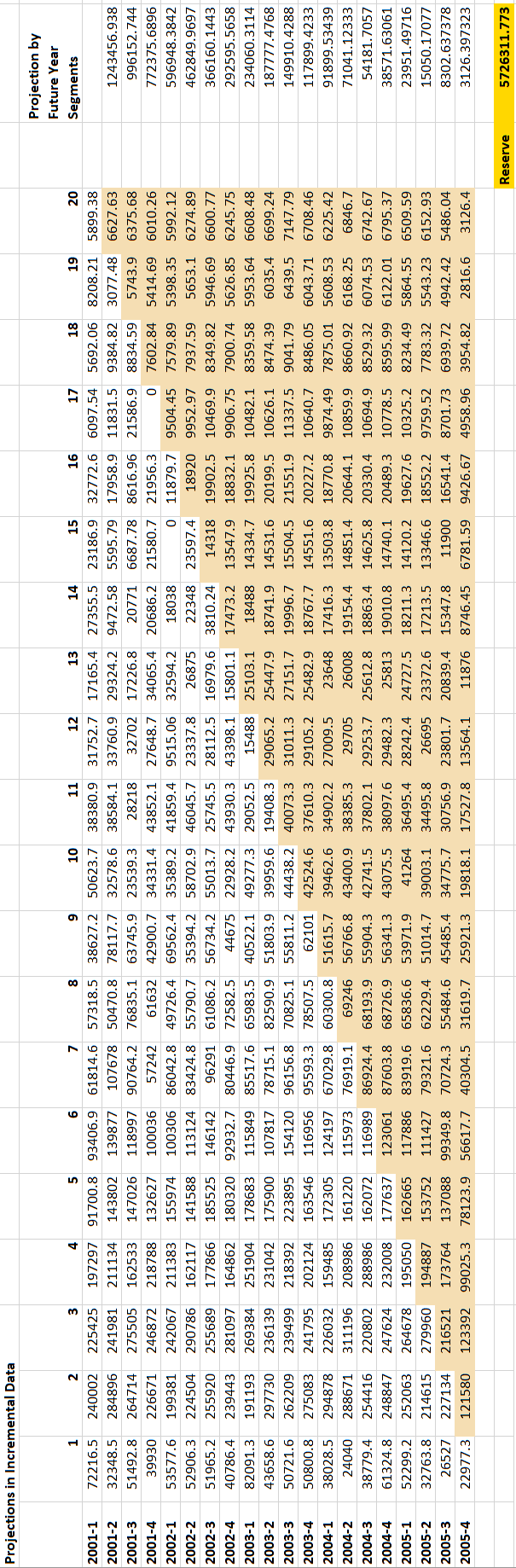


Table 11 Quarterly Reserves

Thus, when collated into Quarterly bins, the ripples in the graph are furthermore visible, which is shown in fig 3.

Figure 3 Paid Claims (Quarterly)

2.8.3-Monthly Reserves

The ripples in the incremental claims data increases as the time period of collation is decreased. The monthly data has the highest number of ripples compared to the other 4 types of collations.

Figure 4 Paid Claims (Monthly)

The graphs clearly show that the smoothness of the curve decreases as we separate the data and put into separate bins of Half-yearly, Quarterly and Monthly. The size of the table of Monthly reserves is 60 columns and its table is not shown here. Thus, on the same data, 5 types of simulations were done and the reserve values were calculated.

2.9-Comparison of Reserves (10,000 Claims)

On repeating the simulation 100 times and analysing the average and the standard deviation, the values of standard deviation seemed increasing from the yearly reserve to the Monthly reserve.

The error is calculated by the formula,

**Root Mean Square Error % = \* 100 %**

Equation 4 Root Mean Square Error

Where,

**Calculated reserve** is the reserve value form Chain Ladder Method  
**Simulated reserve** is the fully simulated reserve   
**n** is the number of data

The overall graph of the 100 runs is showing the highest ripples for Monthly Reserves and the lowest ripple for yearly reserves.

Figure 5 Reserves Calculated in 100 simulations (10,000 Claims)

On stacking the graphs one above the other the ripples are visible clearly.

Figure 6 Reserves Calculated in 100 simulations (10,000 Claims) (Stacked)

|  |  |  |
| --- | --- | --- |
|  | **Average** | **Std. Deviation** |
| **Yearly** | 63,15,795 | 2,68,395 |
| **Half-Yearly** | 65,13,743 | 3,30,067 |
| **Quarterly** | 65,36,373 | 4,30,918 |
| **Monthly** | 66,93,934 | 8,96,706 |
| **Simulated Reserve** | 63,35,699 | 1,64,340 |

Table 12 Average and Std. Deviation of 100 simulations (10,000 Claims)

Figure 7 Average and Standard Deviation of 100 simulations (10,000 Claims)

|  |  |
| --- | --- |
|  | **RMS Error % (Simulated Vs CLM)** |
| **Yearly** | 5.001257941 |
| **Half-Yearly** | 6.246861338 |
| **Quarterly** | 7.756131889 |
| **Monthly** | 14.47521815 |

Table 13 Error in Reserves (10,000 Claims)

Figure 8 Error in Reserves (10,000 Claims)

2.10-Comparison of Reserves (1,000 claims)

If the number of claims analysed (generated data) is decreased to 1000, the error and the standard deviation increases. This is shown below:

Figure 9 Reserves Calculated in 100 simulations (1,000 Claims)

On stacking the graph one above the other, the increased ripples due to the lesser available data. This can be compared with fig 5 and fig 6.

Figure 10 Reserves Calculated in 100 simulations (1,000 Claims) (Stacked)

Figure 11 Average and Standard Deviation of 100 simulations (1,000 Claims)

|  |  |  |
| --- | --- | --- |
|  | **Average** | **Std. Deviation** |
| **Yearly** | 6,49,377 | 79,874 |
| **Half-Yearly** | 6,76,532 | 1,00,539 |
| **Quarterly** | 6,74,634 | 1,32,451 |
| **Monthly** | 6,73,466 | 2,48,619 |

Table 14 Average and Standard Deviation of 100 simulations (1,000 Claims)

Figure 12 Error in Reserves (1,000 Claims)

|  |  |
| --- | --- |
|  | **RMS Error % (Simulated Vs CLM)** |
| **Yearly** | 14.75933303 |
| **Half-Yearly** | 18.17936778 |
| **Quarterly** | 21.98786401 |
| **Monthly** | 38.178164 |

Table 15 Error in Reserves (1,000 Claims)

Thus, if we use more data to analyse the claims, the error is significantly reduced.

2.11-Inference

1. Using shorter time periods to collate data increases the standard deviation in the reserve calculation.
2. But this can be used to get a clear picture of progression of claims by year segments.
3. Increasing the number of data (number of claims) used in the calculation of reserves increases the accuracy of the calculation. *It means, if the estimate is calculated using 10, 000 claims data, the results are accurate, when compared to an estimate calculated using 1,000 claims data.*

3 Curve Fitting

It is a process of fitting a curve to a series of points using a mathematical equation or a statistical distribution. The fitted curve can then be used to analyse the data. It can either involve Interpolation or Extrapolation. In Interpolation, a smooth curve is fitted to the data for data visualisation and statistical inference like relation between two random variables or uncertainty present in a curve is analysed. In Extrapolation, a curve is fitted beyond the observed range of data to analyse its future pattern. (En.wikipedia.org, 2017)

3.1-Fitting curves to the tail of Incremental Data

Likewise, in Reserving, a curve can be fitted to the pattern in the claims data and the progression of the reserving can be analysed. The type of curve fitted to the data depends on the nature and the shape of the data. In our case the data is in the shape of a normal distribution bell curve (from the second line of the Incremental triangle – Table 7).

Figure 13 Paid Claims (for the year 2002)

The points after year 2 can be fitted in an exponential curve. In this case we are considered about the tail of the graph to predict the year 5’s incremental claim amount.

|  |  |
| --- | --- |
| **Year Segment** | **Reserves** |
| 1 | 1623278 |
| 2 | 2523247 |
| 3 | 912021.7 |
| 4 | 359375.6 |
| 5 |  |

Table 16 Paid Claims of 2002 (10,000 Claims)

The next step is to find an equation to fit in the data. In this case I have taken the equation,

**f(x) = A\* e-t**

Equation 5 A simple equation with exponent to fit the tail

where,  
A is the Scaling factor  
t is time in years

On fitting a curve, the following curve is obtained. The converging algorithm found a value A=6856843. On substituting it in the equation, we get,

**f(x) = 6856843 \* e-t**

And thus, the Incremental Claim amount for the year 5 is **125587.46.** On comparing it with the value obtained using Chain ladder method,

|  |  |
| --- | --- |
|  | **Reserve** |
| **CLM** | 179376.8 |
| **Curve fitting** | 125163.5 |

Table 17 Difference between values

Figure 14 Curve fitted to the tail

As seen in table 16, the difference between the values is quite large. But fitting a better curve better results can be obtained.

3.2-Quadratic equations and Cubic equations can’t be used

To fit 3 to 4 points, quadratic equations or cubic equations can be used, which will give a better fit for the points. You can use an equation like,

**f(x) = a0 + a1 t + a2 t2 + a3 t3**

Equation 6 A cubic equation

where,

a0, a1, a2, a3 are coefficients  
t is time in years

But these equations can’t be extrapolated. If you try to extrapolate them, they do not converge. This is shown in fig 14 &15. So, we must fit the curve in an exponential equation.

A complex code can be written to fit the curve separately in different segments. For the starting part of the curve, quadratic splines or cubic splines can be fitted and the tail can be designed as an exponential equation as this depicts the real-world scenario of reserving.

Figure 15 On extrapolating the curve rises (Quadratic)

Figure 16 On extrapolating the curve rises (Cubic)

3.3-Craighead Curve Fitting Method

Besides using quantitative methods, IBNR (Incurred But Not Reported) claims can be estimated using Craighead Curve. This curve can only be used if there are more than two points. (www.actuaries.org.uk, 2017). This method requires premiums data to find the Loss ratio. (note: Chain Ladder method does not require collected Premiums data)

The formula used for the curve is:

**L (x + t) = A(x) \* {1 – }**

Equation 7 Craighead curve equation

Where,

L (x + t) is Actual Observed paid claims ratio at duration t for year of occurrence x

A (x) is Estimated Ultimate Claims Ratio for year x

t is time in years from year of occurrence x

b is parameter representing the tail of the curve

c represents the speed of claim settlement

If the tail of the paid claims curve is longer, parameter b will have higher values. If the claims take long time to settle, parameter c will have higher values.

3.3.1-Effect of varying “c” speed factor on claims ratio

The nature of the Craighead curve can be studied by varying the different factors in the curve. Keep A(x) and b as constants and varying c, we get,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **t** | **1** | **2** | **3** | **4** | **5** |
| **A(x)** | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| **b** | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
| **c** | 0.25 | 0.5 | 1 | 2 | 4 |

Table 18 Effect of Varying parameter “c”

Figure 17 Effect of Varying parameter “c”

3.3.2-Effect of varying “b” tail factor on claims ratio

On keeping A(x) and c as constants and varying b, we get,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **t** | **1** | **2** | **3** | **4** | **5** |
| **A(x)** | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| **b** | 1 | 1.5 | 2 | 3 | 4 |
| **c** | 2 | 2 | 2 | 2 | 2 |

Table 19 Effect of varying parameter “b”

Figure 18 Effect of varying parameter “b”

Thus, we can change the values of b and c iteratively, and fit a curve to the data as close as possible. This process requires the ultimate paid claims ratio for the particular line of business.

3.3.3-Curve fitting process

The following steps are mentioned in the Claims Reserving Manual – Vol 2 in the Institute and Faculty of Actuaries website and in the Insurance Regulatory and Development Authority of India (IRDA) website.

The algorithm for fitting the curve can be designed by the following procedure.

1. Use the most developed year as the base for the simulation and use the cumulative paid plus outstanding ratio (cumulative claims ratio) as the value of **A(x).**
2. Start with an initial guess for **b** and **c**. (b = 1, c = 0.5)
3. Calculate the value of **L (x + t).**
4. Run an iterative algorithm and minimise the square of the differences between, Observed and the calculated values of L (x + t).
5. The output of the code should give values of A(x), b, c which fits the graph.
6. Thus, by substituting the value of t (time in years) for the future years, the value of L(x+t) for future years can be obtained.

As the result of the simulation is cumulative data, the difference in value between two consecutive years gives the value of incremental data. (www.irdaindia.org, 2008)

3.4-Fitting Curves to Cumulative data

In generated data in Table 1, we don’t have the any information about premiums. So, the Craighead curve fitting method cannot be used. But a similar formula can be used to fit the data by iteratively finding the value of **A.**

The formula that can be used in this case is,

**C(t) = A \* {1 – }**

Equation 8 Better Equation to fit cumulative data based on Craighead equation

Where,

C (t) is the cumulative paid claims at time t

A is the ultimate cumulative claims at time n, where n is the last year considered.

b is the tail factor

c is the speed factor.

Before starting the curve fitting process, let us have a look at the cumulative claims data used in the Chain Ladder Method. The graph of the upper triangle is shown below. As you can see in fig. 17. the curve of the cumulative data is very smooth and it is suitable for curve fitting compared to complex shaped incremental data curve.

Figure 19 Cumulative claims (Monthly)

3.4.1-Algorithm (Curve Fitting - Cumulative claims)

1. Start with an initial guess for the maximum and the minimum values for A. *For Example, the maximum value can be chosen as £ 600,000 and the minimum value can be chosen as £ 100, 000. This is shown by a red line in the fig. 17.*
2. Start with an initial guess for values of b and c. Fix the min and max values. *say b(min) = 1, c(min) = 0.5, b(max) = 10, c(max) = 10.*
3. Use a **triple iteration algorithm** to find the values of A, b and c, by minimising the square of difference between the observed and calculated values of **C(t)**.

The final values of A, b and c are then used to find the cumulative claims for the future by substituting in the C(t) equation. The fit for the quarterly data is shown in fig 20.

Figure 20 Fitted curve

If the monthly claims data are used, then the fit is furthermore accurate. The fig.21 shows the increased accuracy of curve fitting. The extended orange line shows the prediction based on the A, b, c values obtained.

Figure 21 Fitted graph (Monthly Claims)

The prediction from 52nd month to the 60th month is very close to the pattern observed in the cumulative data.

3.5-Limitations of Curve fitting

* To fit a curve at least 3 points are needed. So, curves can’t be fitted where lesser details are available.
* For complex shaped curves, curve fitting becomes tedious.
* As we are extrapolating, it is subjected to higher degree of uncertainty.
* On extrapolating, as the time increases the uncertainty in result increases.

3.6-Inference

1. Using shorter time periods to collate data provides **more points to fit the data** and thus a better fit can be obtained. Because if the claims data is put into monthly bins, even 6 months data can give 6 points to fit a curve. While when considering yearly bins 3 years data give just 3 points for curve fitting.
2. Using shorter time periods to collate data increase the ripples in the incremental claims data and curve fitting becomes more complex (while fitting curve to Incremental data).
3. To fit the curve better and to obtain better results, complex equations must be used (in case of incremental claims data).
4. Best fit for the curve is obtained by **fitting curves to the cumulative data**, rather than incremental data.

4 Conclusion

It is concluded that the collation of the data into shorter time periods in Chain Ladder Method increases the standard deviation of the calculation. But if more data is used for estimation this standard deviation is significantly reduced. And collation of data provides clear picture of the progression of claims received. Thus, when new products are launched in an insurance company, and when the data is available for only few years, the shorter time periods collation can give a good picture of the claims progression. Also in curve fitting, the collation of data into shorter time periods provides more points to fit the curve. Thus, we can get better results in curve fitting as well. Fitting curves to cumulative data provides better results when compared to incremental data, as it has a smooth curve. Thus, collation of data into shorter time periods has good advantages at the cost of little variance in the estimated data.

5 Appendix

The VBA macro coded for this analysis is available in the following link:

<https://drive.google.com/open?id=0B6CZTL2bvynzeGZmWmV2aDNfcEk>

6 Bibliography

En.wikipedia.org. (2017). *Curve fitting*. [online] Available at: https://en.wikipedia.org/wiki/Curve\_fitting [Accessed 31 Aug. 2017].

En.wikipedia.org. (2017). *Loss reserving*. [online] Available at: https://en.wikipedia.org/wiki/Loss\_reserving [Accessed 2 Sep. 2017].

En.wikipedia.org. (2017). *Solvency II Directive 2009*. [online] Available at: https://en.wikipedia.org/wiki/Solvency\_II\_Directive\_2009 [Accessed 28 Aug. 2017].

En.wikipedia.org. (2017). *Chain-ladder method*. [online] Available at: https://en.wikipedia.org/wiki/Chain-ladder\_method [Accessed 28 Aug. 2017].

www.fh-vie.ac.at. (2017). *A practical guide to the use of the chain ladder method for determining technical provisions for outstanding reported claims in non-life insurance*. [online] Available at: http://www.fh-vie.ac.at/var/em\_plain\_site/storage/original/application/428aa414bf1ba9198dd84455133b4abd.pdf [Accessed 28 Aug. 2017].

www.actuaries.org.uk. (2017). *Claims Reserving Manual - Vol 2*. [online] Available at: https://www.actuaries.org.uk/documents/claims-reserving-manual-volume-2 [Accessed 31 Aug. 2017].