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

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,

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}, \quad \mathrm{var}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

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				
		1	**2**	**3**	**4**	**5**
Accident Year	**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				
		1	**2**	**3**	**4**	**5**
Accident Year	**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 f_1.

$$\text{Forward Factor } F_j = \frac{\sum_{i=1}^{n+1-j} C_{ij}}{\sum_{i=1}^{n+1-j} C_{i,j-1}}$$

Equation 2 Forward Factor

Where,

i, j represents row and column

n represents the number of years

C_{ij} represents the corresponding claim amount.

		Development Year				
		1	**2**	**3**	**4**	**5**
Accident Year	**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,

$$F_j = \frac{1650302+1623278+1735831+1696850}{4086053+4146524+4494792+4378764} = 2.288568$$

And thus, the forward factors are:

	F_1	F_2	F_3	F_4
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.

$$\widehat{C_{ij}} = \widehat{C_{i,j-1}} \times F_j$$

$$\widehat{C_{in}} = C_{i,n+1-i} \times \prod_{j=n+2-i}^{n} F_j$$

Equation 3 Formula to find cumulative claim amounts

Where,

$\boldsymbol{C_{ij}}$ represents cumulative claim amount, i represents accident year and j represents development year.

$\boldsymbol{C_{in}}$ represents cumulative claim amounts, i represents accident year and n represents development year.

F$_j$ represents development factor or forward factor. The filled table looks like below.

		Development Year				
		1	2	3	4	5
Accident Year	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
		1	2	3	4	5		
Accident Year	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:

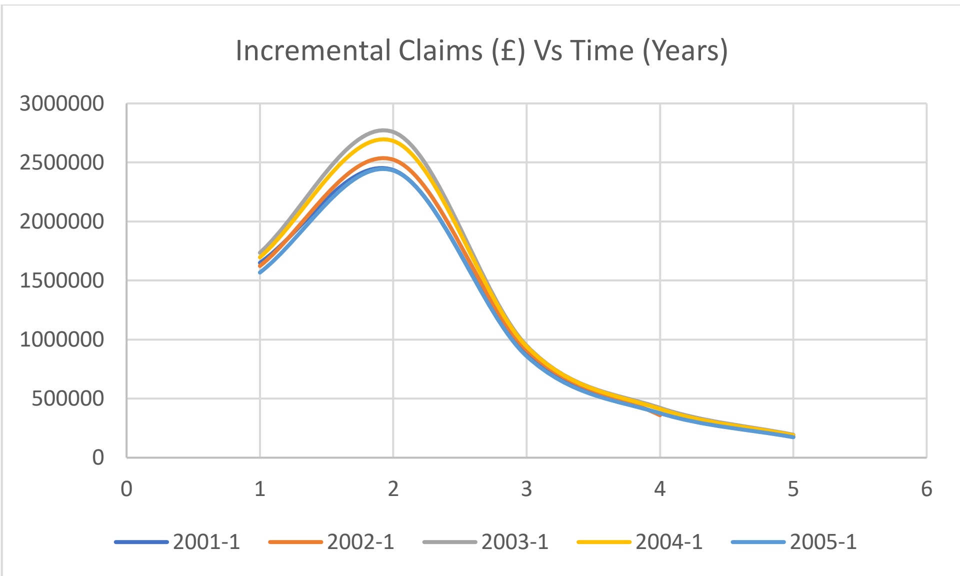

Figure 1 Paid Claims (Yearly)

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				
		1	**2**	**3**	**4**	**5**
Accident Year	**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.

	1	2	3	4	5	6	7	8	9	10	Projection by Future Year
2001-1	344567	949598	540044	366689	217839	141296	107606	71028	41580	26570	
2001-2	356136	911580	617438	324877	191818	139103	99712	57572	52378	26287	2384340
2002-1	305865	968740	559985	332318	196136	156123	100845	57825	46277	26033	1472129
2002-2	348672	954096	676848	330757	229005	120716	79989	64430	48466	27265	880063
2003-1	316943	1055158	701473	338033	224194	103908	103443	66853	50289	28291	559313
2003-2	363731	974769	743686	379532	240858	144577	107500	69475	52261	29400	354002
2004-1	356946	985384	666708	320222	220011	136376	101402	65534	49297	27732	223437
2004-2	354520	1006260	688706	357944	227395	140953	104805	67733	50951	28663	128962
2005-1	337126	954302	635729	336579	213822	132540	98550	63691	47910	26952	66361
2005-2	276638	785642	522927	276857	175882	109022	81063	52389	39409	22170	22170
										Reserve	**6090779**

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.

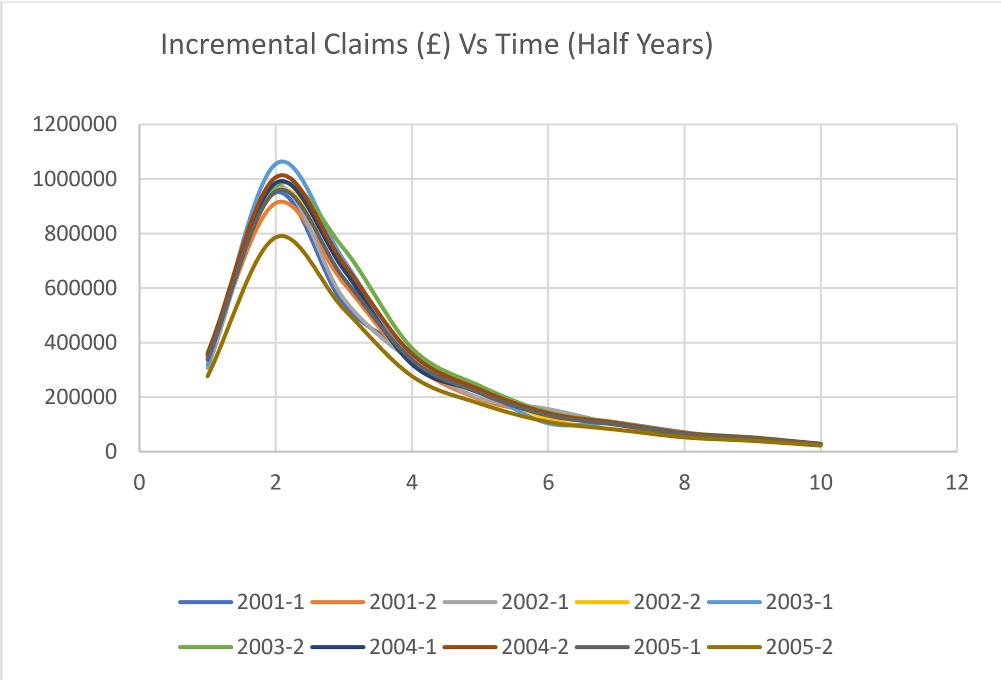

Figure 2 Paid Claims (Half Yearly)

2.8.2-Quarterly Reserves

Projections in Incremental Data																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		Projection by Future Year Segments
2001-1	72216.5	240002	225425	197297	91700.8	93406.9	61814.6	57318.5	38627.2	50623.7	38380.9	31752.7	17165.4	27355.5	23186.9	32772.6	6097.54	5692.06	8208.21	5899.38		
2001-2	32348.5	284896	241981	211134	143802	139877	107678	50470.8	78117.7	32578.6	38584.1	33760.9	29324.2	9472.58	5595.79	17958.9	11831.5	9384.82	3077.48	6627.63		1243456.938
2001-3	51492.8	264714	275505	162533	147026	118997	90764.2	76835.1	63745.9	23539.3	28218	32702	17226.8	20771	6687.78	8616.96	21586.9	8834.59	5743.9	6375.68		996152.744
2001-4	39930	226671	246872	218788	132627	100036	57242	61632	42900.7	34331.4	43852.1	27648.7	34065.4	20686.2	21580.7	21956.3	0	7602.84	5414.69	6010.26		772375.6896
2002-1	53577.6	199381	242067	211383	155974	100306	86042.8	49726.4	69562.4	35389.2	41859.4	9515.06	32594.2	18038	0	11879.7	9504.45	7579.89	5398.35	5992.12		596948.3842
2002-2	52906.3	224504	290786	162117	141588	113124	83424.8	55790.7	35394.2	58702.9	46045.7	23337.8	26875	22348	23597.4	18920	9952.97	7937.59	5653.1	6274.89		462849.9697
2002-3	51965.2	255920	255689	177866	185525	146142	96291	61086.2	56734.2	55013.7	25745.5	28112.5	16979.6	3810.24	14318	19902.5	10469.9	8349.82	5946.69	6600.77		366160.1443
2002-4	40786.4	239443	281097	164862	180320	92932.7	80446.9	72582.5	44675	22928.2	43930.3	43398.1	15801.1	17473.2	13547.9	18832.1	9906.75	7900.74	5626.85	6245.75		292595.5658
2003-1	82091.3	191193	269384	251904	178683	115849	85517.6	65983.5	40522.1	49277.3	29052.5	15488	25103.1	18488	14334.7	19925.8	10482.1	8359.58	5953.64	6608.48		234060.3114
2003-2	43658.6	297730	236139	231042	175900	107817	78715.1	82590.9	51803.9	39959.6	19408.3	29065.2	25447.9	18741.9	14531.6	20199.5	10626.1	8474.39	6035.4	6699.24		187777.4768
2003-3	50721.6	262209	239499	218392	223895	154120	96156.8	70825.1	55811.2	44438.2	40073.3	31011.3	27151.7	19996.7	15504.5	21551.9	11337.5	9041.79	6439.5	7147.79		149910.4288
2003-4	50800.8	275083	241795	202124	163546	116956	95593.3	78507.5	62101	42524.6	37610.3	29105.2	25482.9	18767.7	14551.6	20227.2	10640.7	8486.05	6043.71	6708.46		117899.4233
2004-1	38028.5	294878	226032	159485	172305	124197	67029.8	60300.8	51615.7	39462.6	34902.2	27009.5	23648	17416.3	13503.8	18770.8	9874.49	7875.01	5608.53	6225.42		91899.53439
2004-2	24040	288671	311196	208986	161220	115973	76919.1	69246	56766.8	43400.9	38385.3	29705	26008	19154.4	14851.4	20644.1	10859.9	8660.92	6168.25	6846.7		71041.12333
2004-3	38779.4	254416	220802	288986	162072	116989	86924.4	68193.9	55904.3	42741.5	37802.1	29253.7	25612.8	18863.4	14625.8	20330.4	10694.9	8529.32	6074.53	6742.67		54181.7057
2004-4	61324.8	248847	247624	232008	177637	123061	87603.8	68726.9	56341.3	43075.5	38097.6	29482.3	25813	19010.8	14740.1	20489.3	10778.5	8595.99	6122.01	6795.37		38571.63061
2005-1	52299.2	252063	264678	195050	162665	117886	83919.6	65836.6	53971.9	41264	36495.4	28242.4	24727.5	18211.3	14120.2	19627.6	10325.2	8234.49	5864.55	6509.59		23951.49716
2005-2	32763.8	214615	279960	194887	153752	111427	79321.6	62229.4	51014.7	39003.1	34495.8	26695	23372.6	17213.5	13346.6	18552.2	9759.52	7783.32	5543.23	6152.93		15050.17077
2005-3	26527	227134	216521	173764	137088	99349.8	70724.3	55484.6	45485.4	34775.7	30756.9	23801.7	20839.4	15347.8	11900	16541.4	8701.73	6939.72	4942.42	5486.04		8302.637378
2005-4	22977.3	121580	123392	99025.3	78123.9	56617.7	40304.5	31619.7	25921.3	19818.1	17527.8	13564.1	11876	8746.45	6781.59	9426.67	4958.96	3954.82	2816.6	3126.4		3126.397323
																					Reserve	5726311.773

Table 11 Quarterly Reserves

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

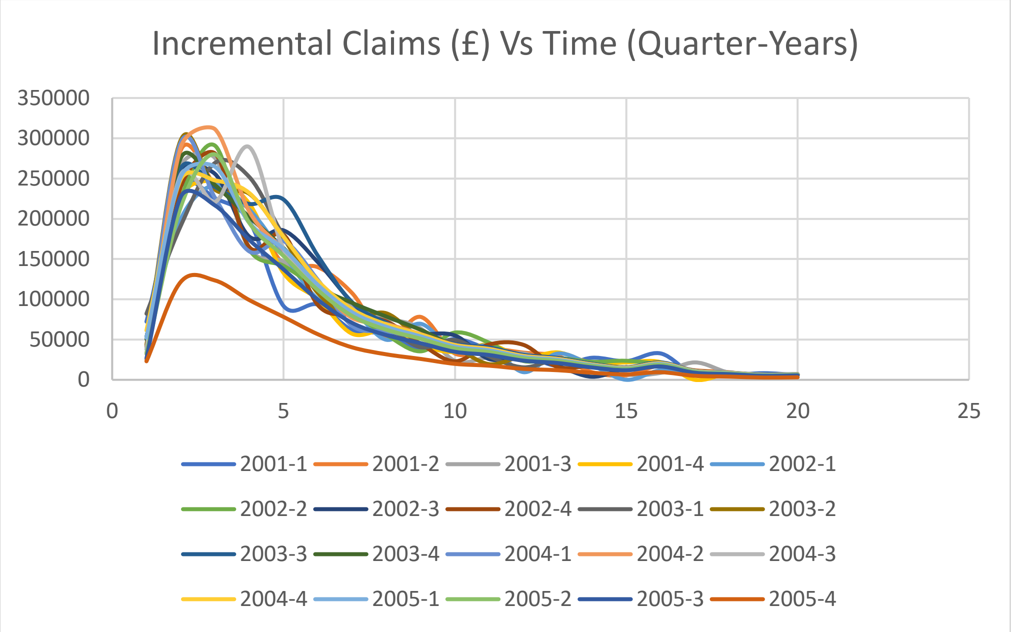

Figure 3 Paid Claims (Quarterly)

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.

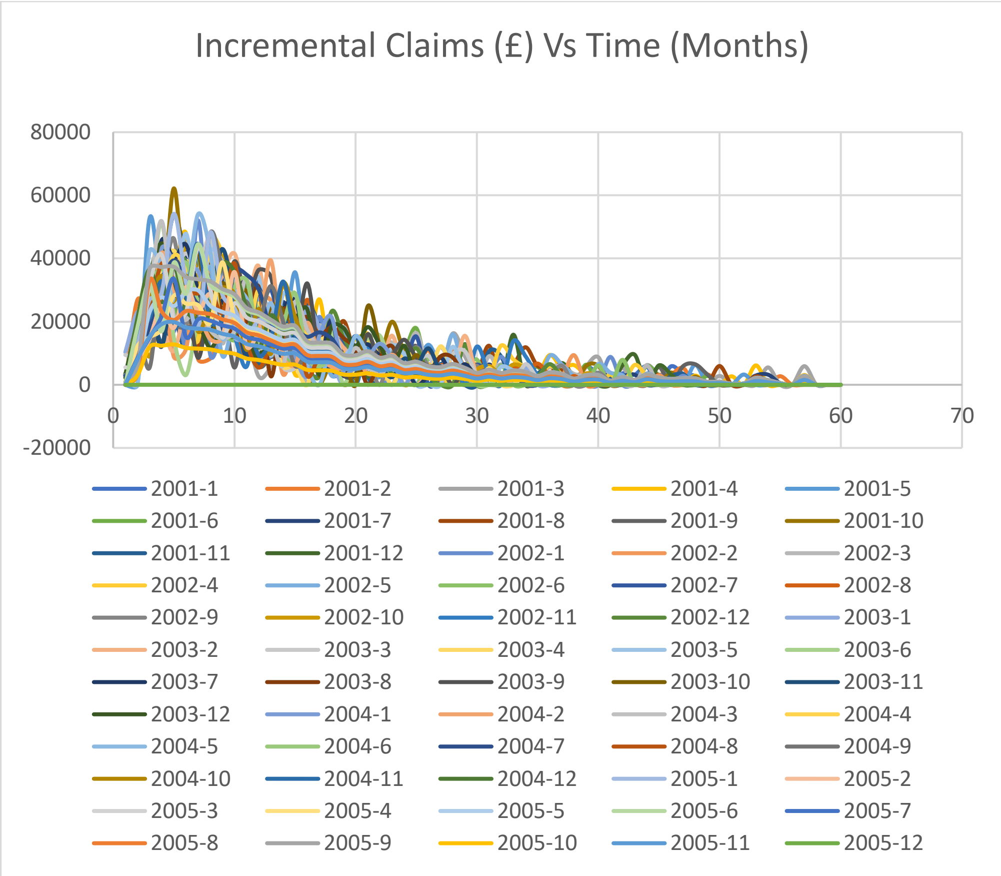

Figure 4 Paid Claims (Monthly)

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,

$$\textbf{Root Mean Square Error \%} = \frac{\sqrt{\sum(\textit{Calculated Reserve} - \textit{Simulated Reserve})^2/n}}{\textit{Calculated Reserve}} * 100\ \%$$

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.

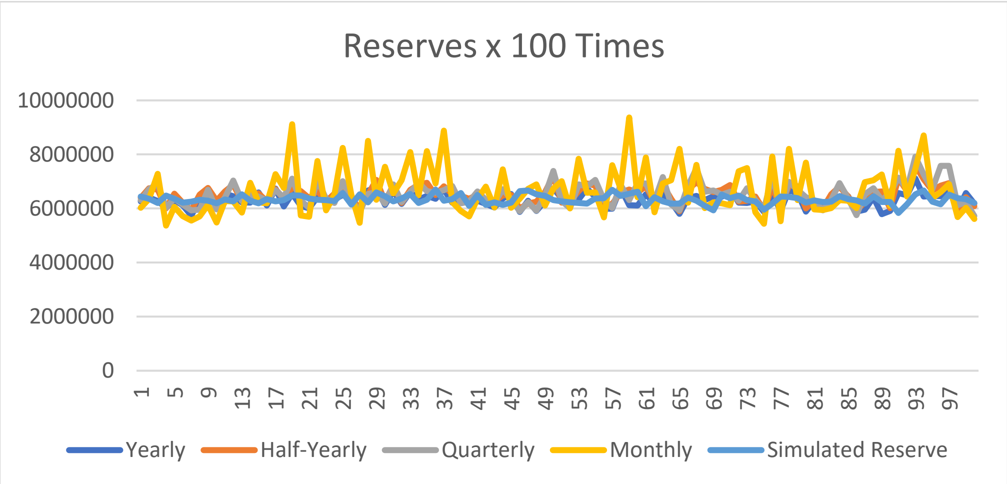

Figure 5 Reserves Calculated in 100 simulations (10,000 Claims)

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

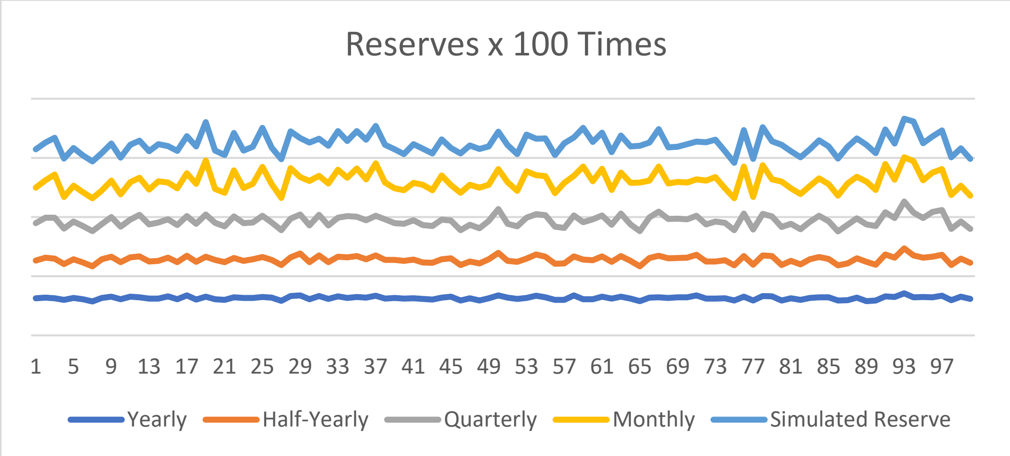

Figure 6 Reserves Calculated in 100 simulations (10,000 Claims) (Stacked)

	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)

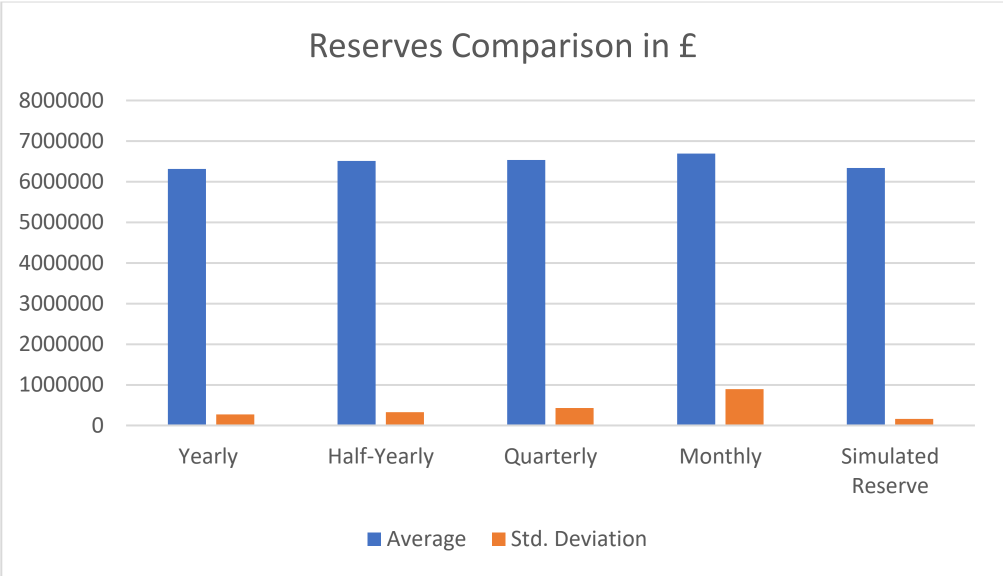

Figure 7 Average and Standard Deviation of 100 simulations (10,000 Claims)

	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)

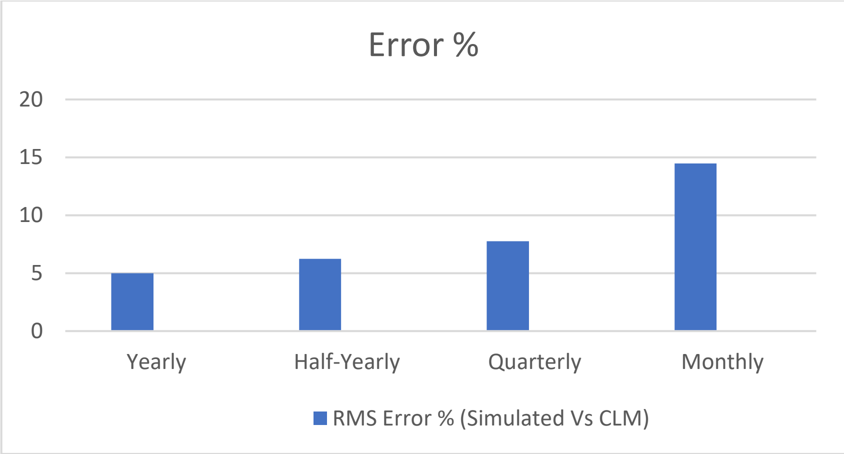

Figure 8 Error in Reserves (10,000 Claims)

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:

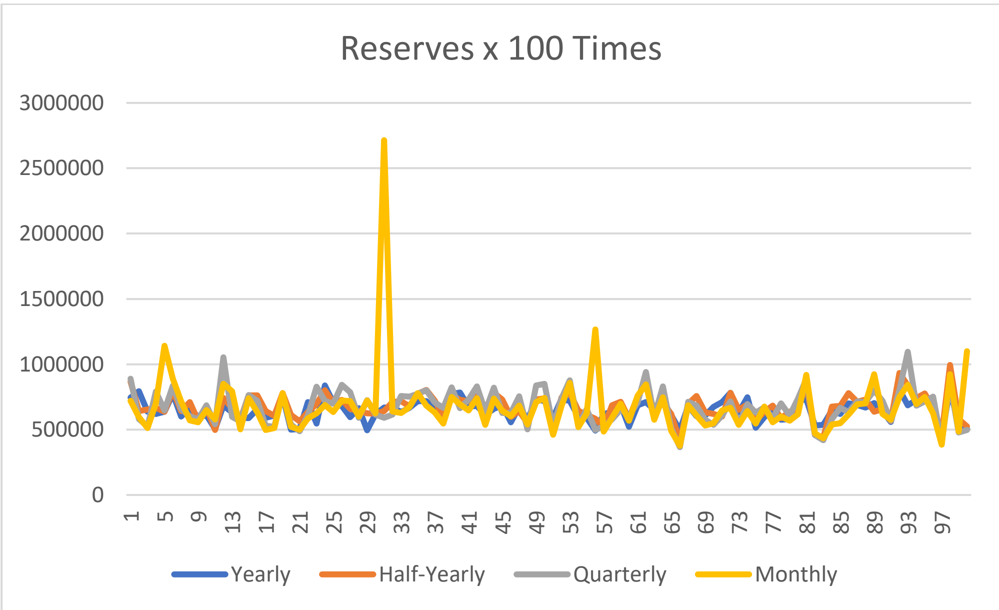

Figure 9 Reserves Calculated in 100 simulations (1,000 Claims)

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.

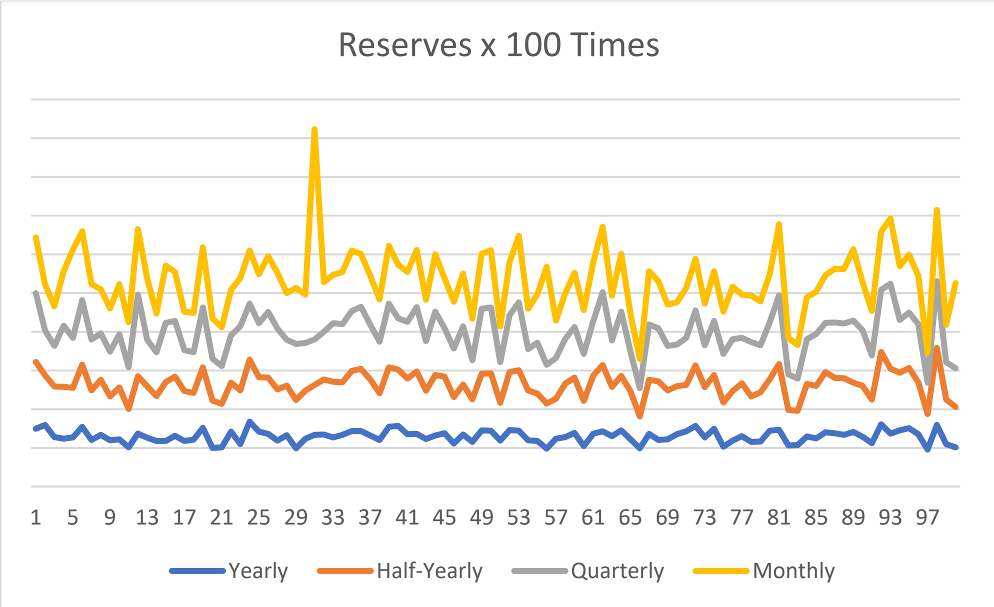

Figure 10 Reserves Calculated in 100 simulations (1,000 Claims) (Stacked)

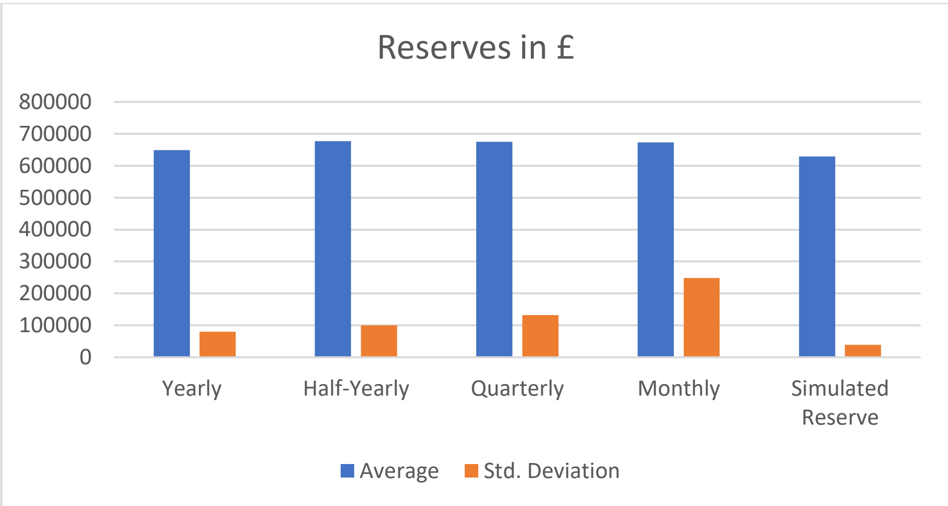

Figure 11 Average and Standard Deviation of 100 simulations (1,000 Claims)

	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)

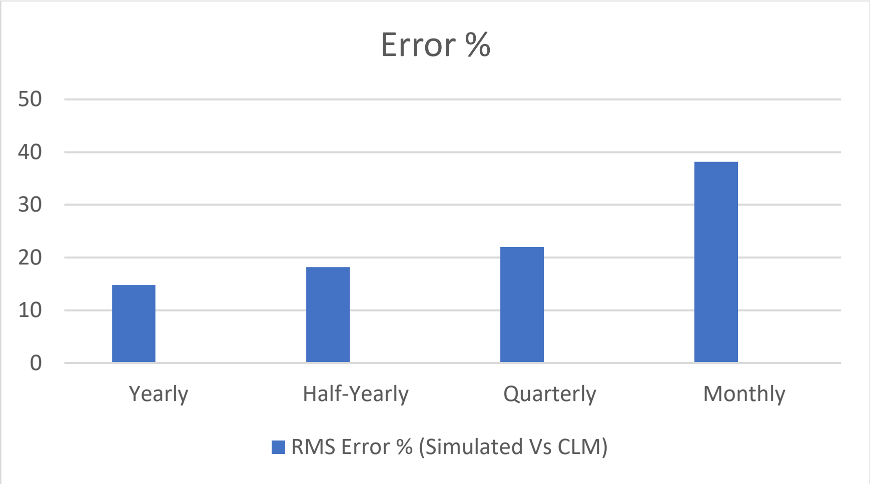

Figure 12 Error in Reserves (1,000 Claims)

	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).

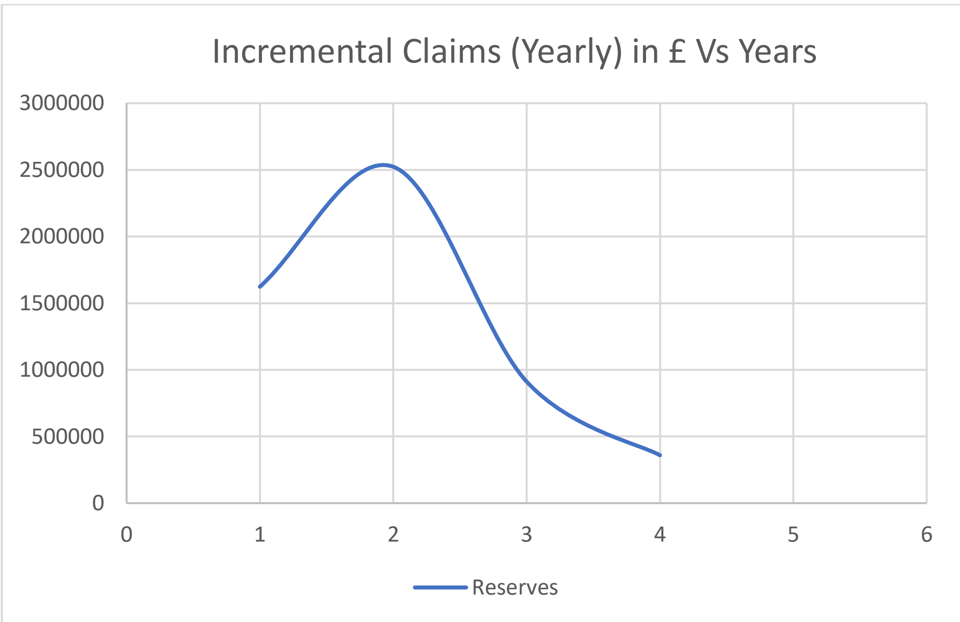

Figure 13 Paid Claims (for the year 2002)

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,

$$\mathbf{f(x) = A^* \, e^{-t}}$$

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,

$$\mathbf{f(x) = 6856843 * e^{-t}}$$

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

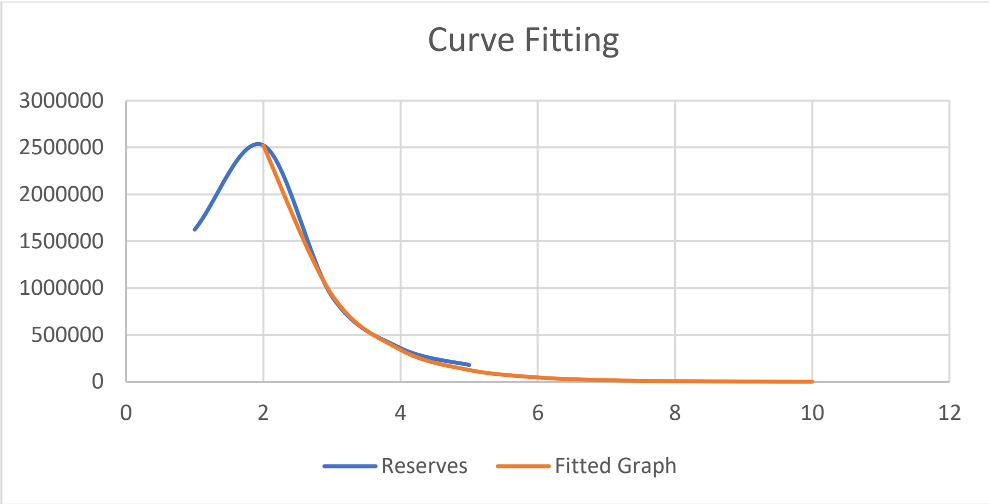

Figure 14 Curve fitted to the tail

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,

$$f(x) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Equation 6 A cubic equation

where,

a_0, a_1, a_2, a_3 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.

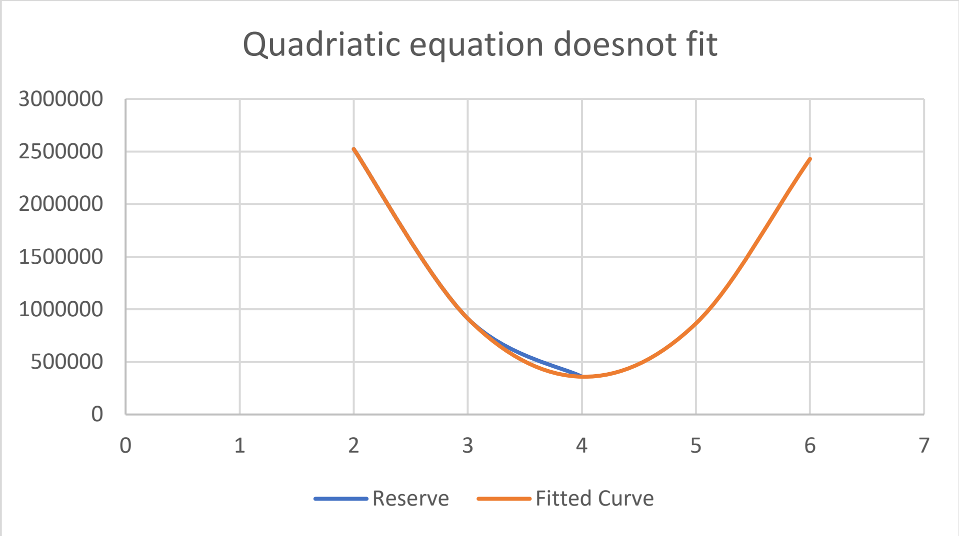

Figure 15 On extrapolating the curve rises (Quadratic)

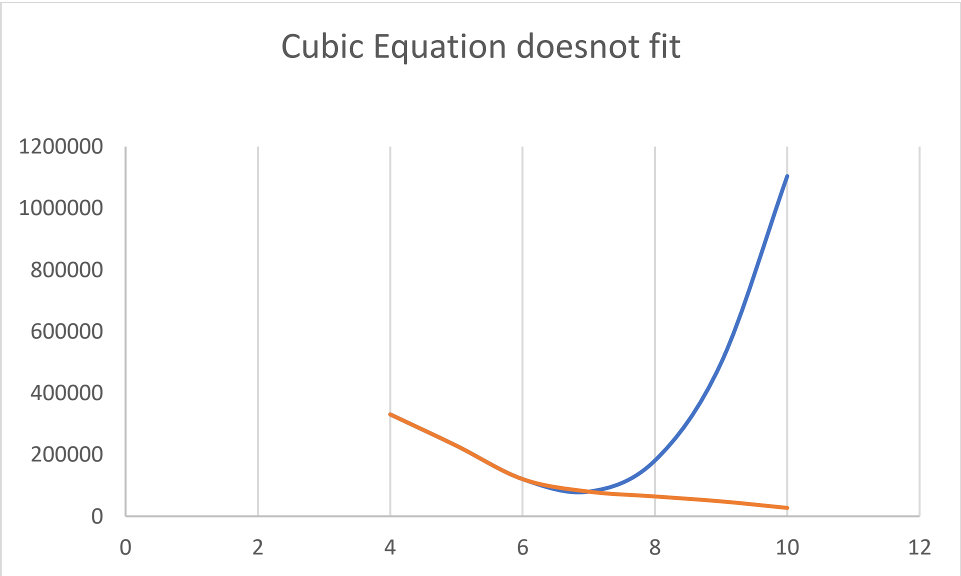

Figure 16 On extrapolating the curve rises (Cubic)

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:

$$\mathbf{L\ (x + t) = A(x) * \{1 - e^{(\frac{-t}{b})^{\wedge}c}\}}$$

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"

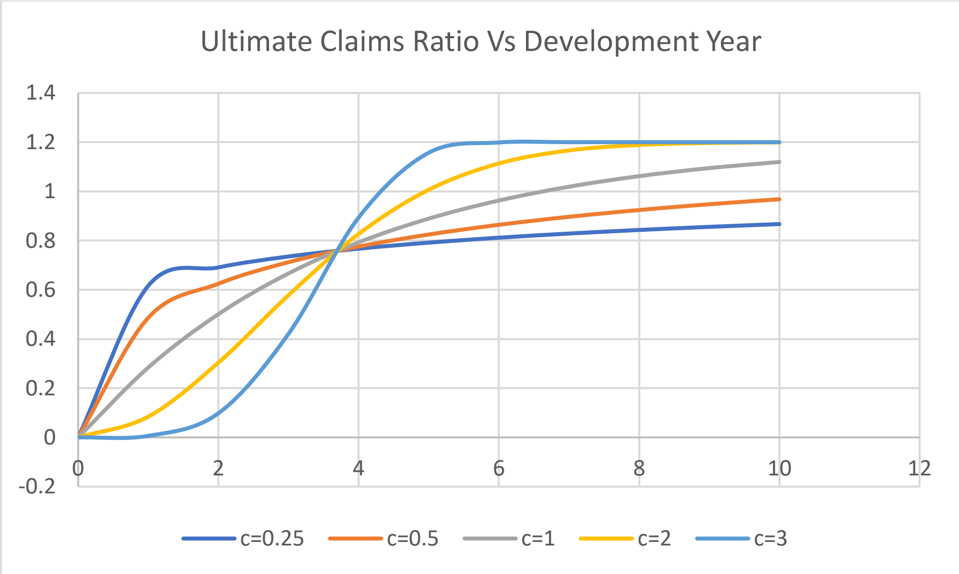

Figure 17 Effect of Varying parameter "c"

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"

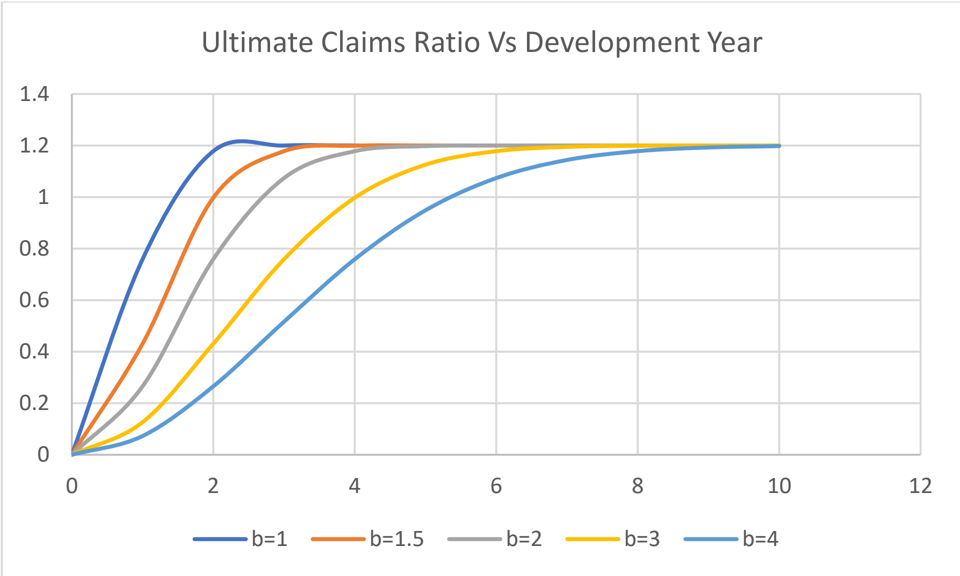

Figure 18 Effect of varying parameter "b"

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,

$$\mathbf{C(t) = A * \{1 - e^{(\frac{-t}{b})^{\wedge}c}\}}$$

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.

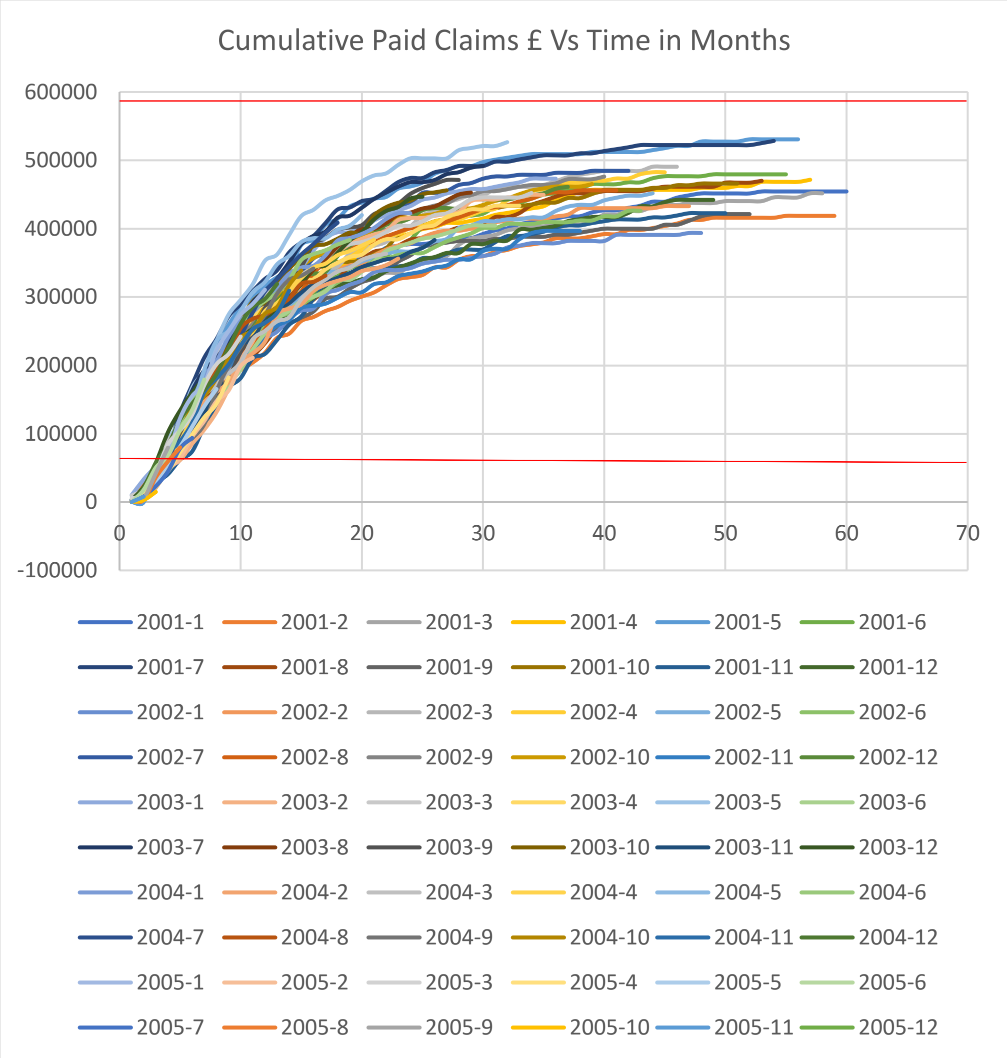

Figure 19 Cumulative claims (Monthly)

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.

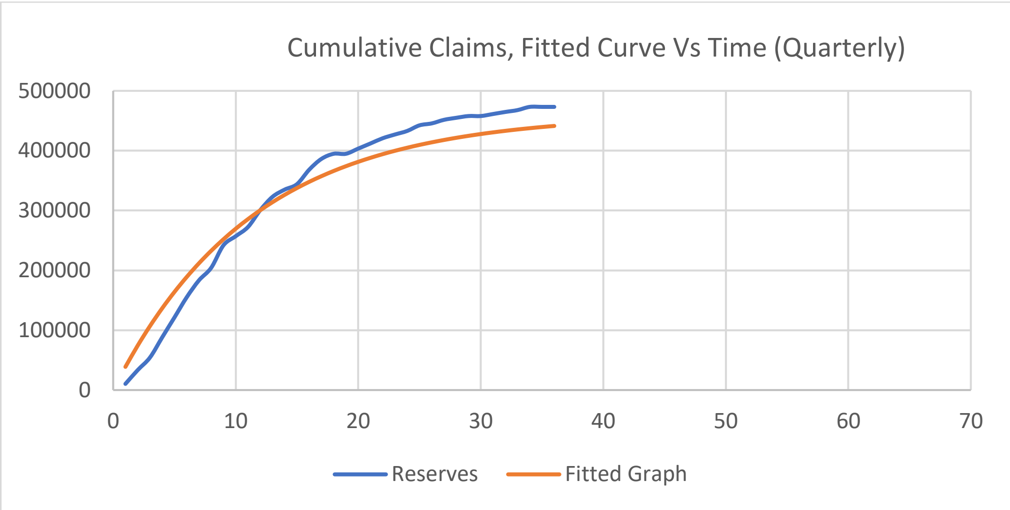

Figure 20 Fitted curve

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.

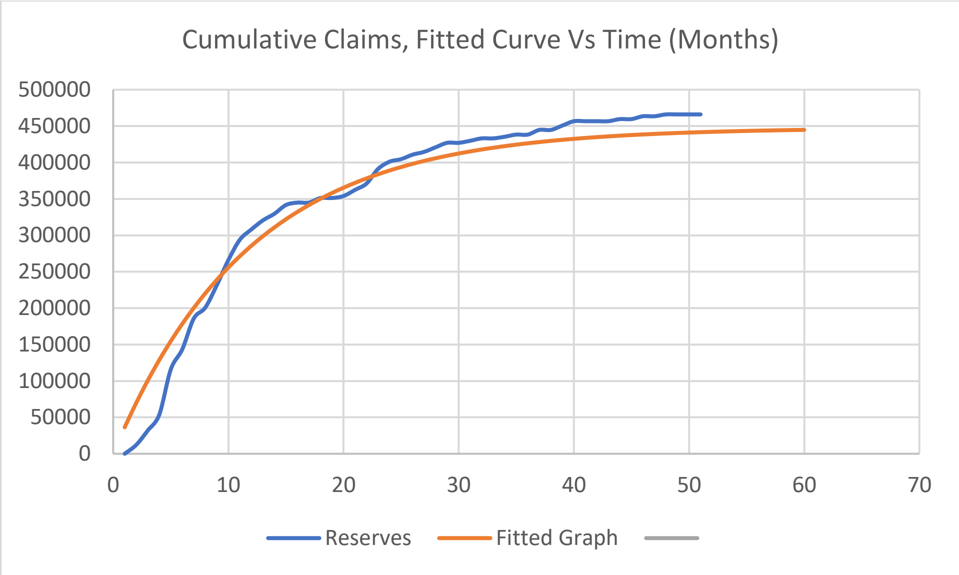

Figure 21 Fitted graph (Monthly Claims)

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].