Project 2

Jan Alexandersson, Max Sjödin & Anton Stråhle

15 februari 2020

Introduction

The purpose of this project is to examine how one can predict future expenses in order to be able to reserve against them. Specifically we wish to examine both the expected future payments as well as the variance within these payments. The data which we have been given is collected over several years and divided between two vastly different insurance products, henceforth refereed to as product, or branch, 1 or 2. It is important to mention that we use Jan Alexanderssons data in this project.

Exercise 1

Since the claims triangle is limited to 10 development years the data has been truncated to accommodate this limit. The final 10 years of the data can be aggregated in order to create the following two paid claims triangles.

Table 1: Paid claims triangle for product 1

	1	2	3	4	5	6	7	8	9	10
1	14002196	20146708	21434023	22027415	22151985	22213380	22265789	22265789	22265789	22265789
2	7814759	11756938	12506719	12719773	12846004	12846004	12846004	12846004	12846004	
3	5181897	7401722	7820233	7922290	7922290	7940404	7940404	7940404		
4	5037120	7327216	7944307	8104325	8117003	8117003	8117003			
5	8042298	11453010	12085662	12166515	12212464	12274792				
6	6752949	10210348	10890964	11255347	11302949					
7	3715909	5176779	5580922	5689709						
8	6507705	9460226	10056047							
9	8386236	11910073								
10	6407931									

Table 2: Paid claims triangle of product 2

	1	2	3	4	5	6	7	8	9	10
1	3380661	11917434	19149141	23136198	25676759	26747939	27363824	27499046	27701642	27860156
2	4125276	14152392	21368586	25851401	28418708	29548445	30201830	30681144	30772972	
3	4388321	16013036	24759810	30271447	32403142	33733300	34205903	34665267		
4	3275256	11368463	16545811	19797983	21403379	22569878	23045976			
5	5981591	19410790	29558338	34979542	38445302	40182223				
6	4140485	13329441	20440916	24176081	26578838					
7	4282806	16150541	23746116	28208154						
8	3958824	13791706	20831932							
9	3775045	15531614								
10	4358136									

Using Mack's non-parametric CL approach we can predict the total cost of future payments per year as follows in Table 3 and Table 4.

Table 3: Full claims triangle of type 1 predicted with CL

	1	2	3	4	5	6	7	8	9	10
1	14002196	20146708	21434023	22027415	22151985	22213380	22265789	22265789	22265789	22265789
2	7814759	11756938	12506719	12719773	12846004	12846004	12846004	12846004	12846004	12846004
3	5181897	7401722	7820233	7922290	7922290	7940404	7940404	7940404	7940404	7940404
4	5037120	7327216	7944307	8104325	8117003	8117003	8117003	8117003	8117003	8117003
5	8042298	11453010	12085662	12166515	12212464	12274792	12287377	12287377	12287377	12287377
6	6752949	10210348	10890964	11255347	11302949	11328296	11339910	11339910	11339910	11339910
7	3715909	5176779	5580922	5689709	5717088	5729908	5735783	5735783	5735783	5735783
8	6507705	9460226	10056047	10264529	10313922	10337051	10347649	10347649	10347649	10347649
9	8386236	11910073	12683551	12946506	13008805	13037977	13051345	13051345	13051345	13051345
10	6407931	9286944	9890068	10095108	10143686	10166433	10176857	10176857	10176857	10176857

Table 4: Full claims triangle of type 2 predicted with CL

	1	2	3	4	5	6	7	8	9	10
1	3380661	11917434	19149141	23136198	25676759	26747939	27363824	27499046	27701642	27860156
2	4125276	14152392	21368586	25851401	28418708	29548445	30201830	30681144	30772972	30949061
3	4388321	16013036	24759810	30271447	32403142	33733300	34205903	34665267	34840692	35040057
4	3275256	11368463	16545811	19797983	21403379	22569878	23045976	23315657	23433647	23567739
5	5981591	19410790	29558338	34979542	38445302	40182223	40973727	41453197	41662973	41901376
6	4140485	13329441	20440916	24176081	26578838	27747438	28294003	28625096	28769955	28934582
7	4282806	16150541	23746116	28208154	30831459	32187035	32821051	33205119	33373155	33564122
8	3958824	13791706	20831932	24963281	27284818	28484458	29045541	29385428	29534135	29703135
9	3775045	15531614	23591639	28270288	30899371	32257933	32893345	33278259	33446666	33638054
10	4358136	15380393	23361944	27995040	30598525	31943860	32573085	32954252	33121019	33310543

Lastly we wish to predict the future payments for each of the claims years, i.e. $R_{2,i}, ..., R_{10,i}$, and then combine them into a the total chain ladder reserves, i.e. R_i for $i = 1, 2$.

Table 5: Total Chain Ladder Reserve

Product 1	5297420
Product 2	66433558

Exercise 2

We now want to check whether or not Mack's underlying assumptions are met in our case. The assumptions are as follows.

1. $E[C_{i,k+1}|C_{i,1}, ..., Ci, k] = f_k C_{i,k}$
2. Independent accident years
3. $Var(C_{i,k+1}|C_{i,1}, ..., Ci, k) = \sigma_k^2 C_{i,k}$

We begin by examine whether or not the we have an approximate linear relationship between $C_{i,k}$ and $C_{i,k+1}$ for $i = 1, ..., 10$ for the two branches.

Figure 1: Linear approximation between C_i,k and C_i,k+1 for product 1

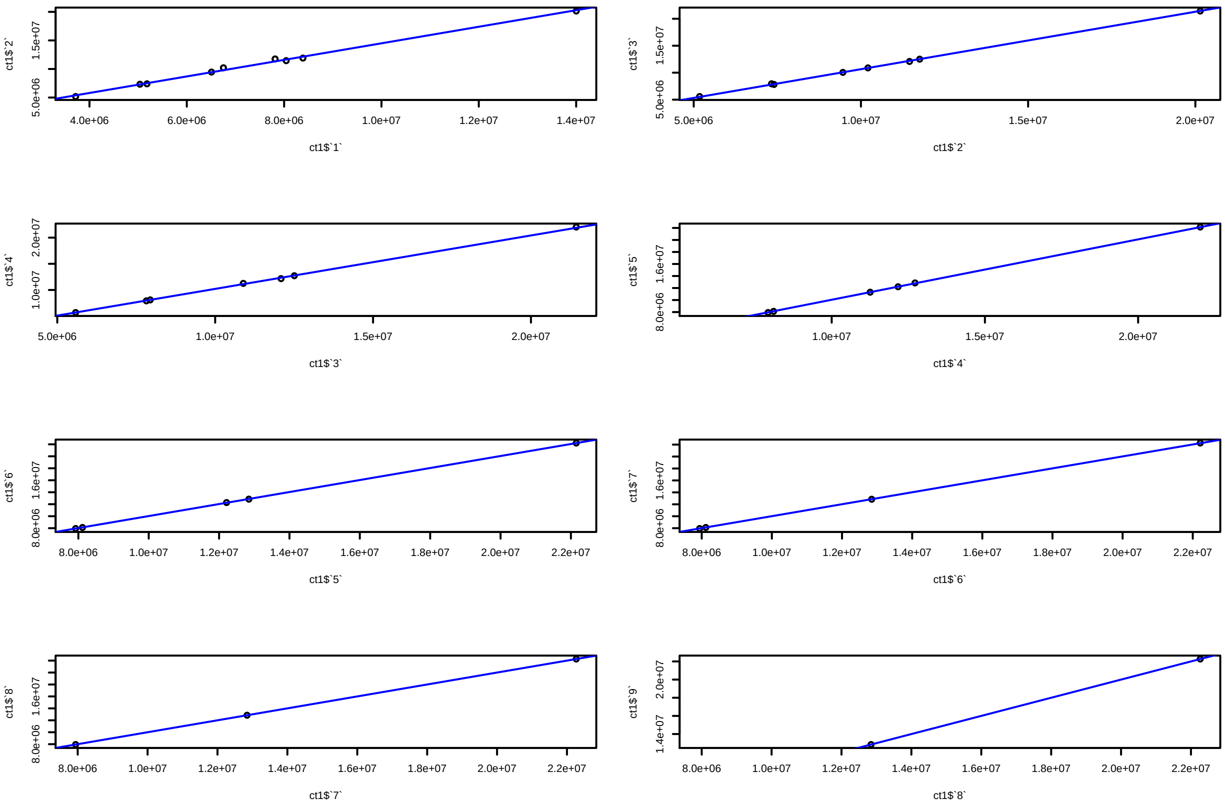
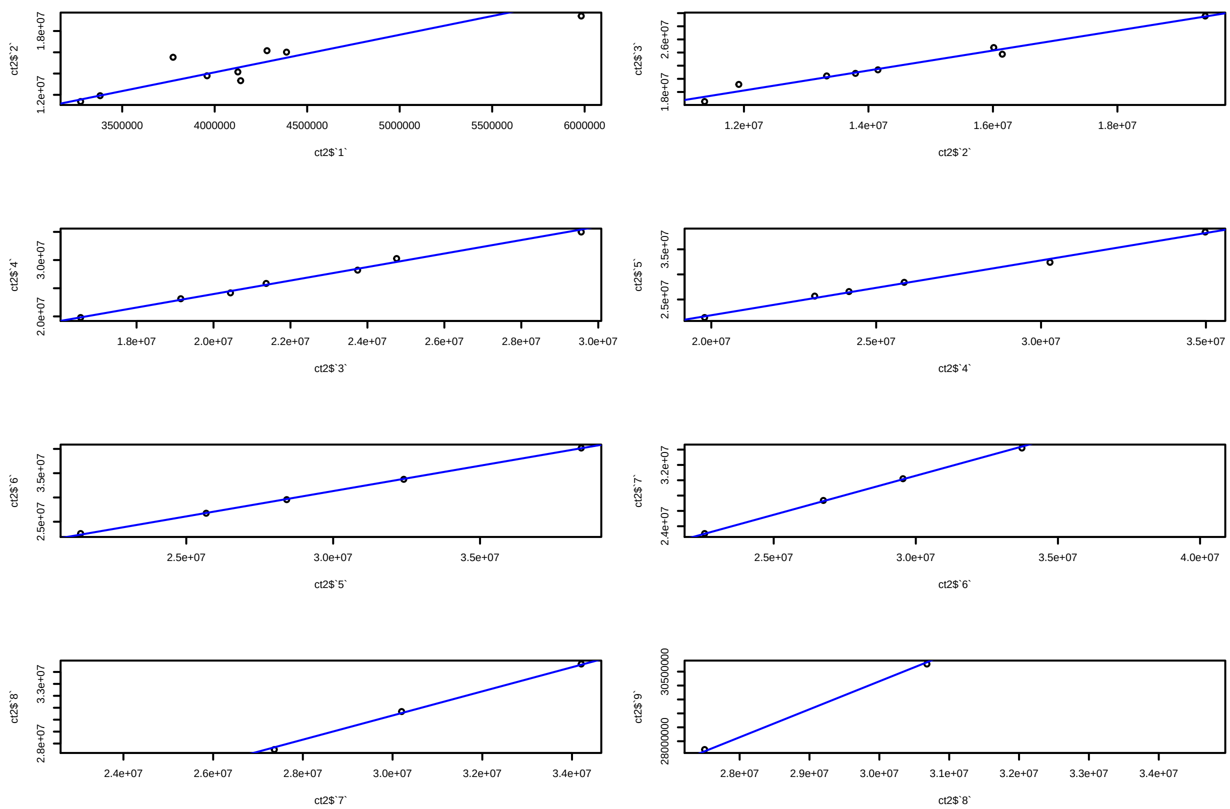


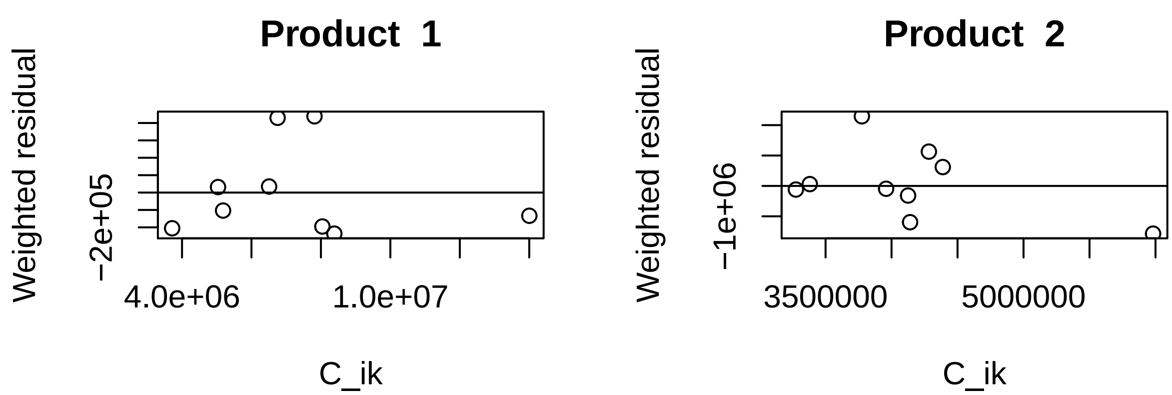
Figure 2: Linear approximation between C_i,k and C_i,k+1 for product 2



We note from the previous figures that the assumption of linearity seem to hold for both insurance branches.

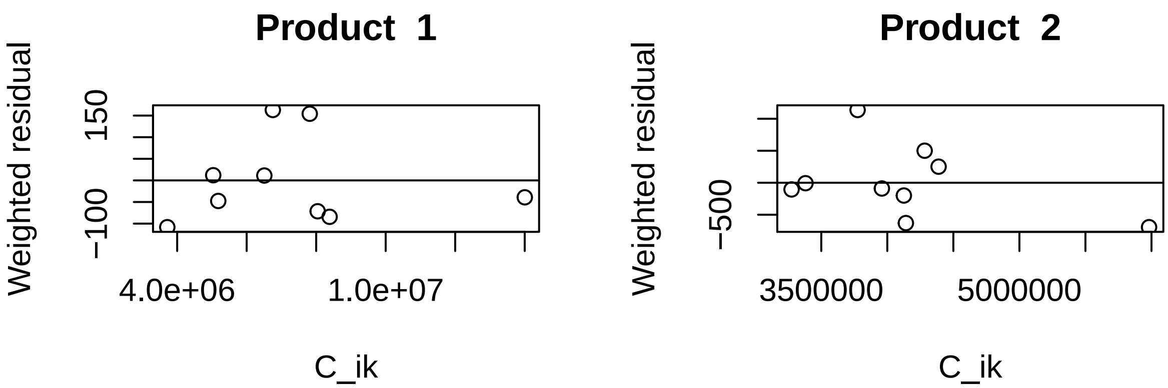
We continue by investigating the third chain ladder assumption. According to Mack's article it is suggested to plot three different plots of residuals. First we plot expression 1: $C_{i,k+1} - C_{i,k} f_{k0}$ against $C_{i,k}$, where $f_{k0} = \sum_{i=1}^{I-k} C_{i,k} C_{i,k+1} / \sum_{i=1}^{I-k} C_{i,k}^2$.

Figure 3: Expression 1 against C_i,k for product 1 and 2



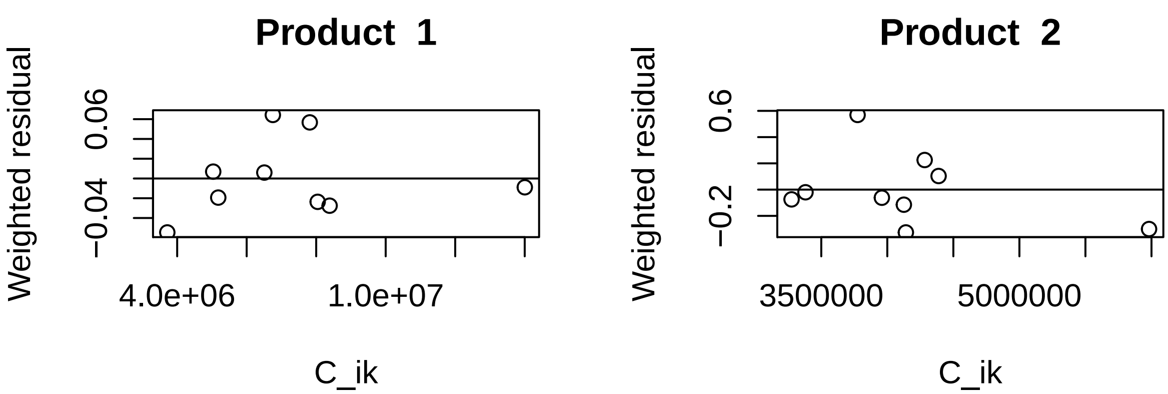
Then we plot expression 2: $(C_{i,k+1} - C_{i,k} f_{k1})/\sqrt{C_{i,k}}$ against $C_{i,k}$, where $f_{k1} = \sum_{i=1}^{I-k} C_{i,k+1} / \sum_{i=1}^{I-k} C_{i,k}$.

Figure 4: Expression 2 against C_i,k for product 1 and 2



Lastly, we plot expression 3: $(C_{i,k+1} - C_{i,k}f_{k2})/C_{i,k}$ against $C_{i,k}$, where $f_{k2} = \frac{1}{I-k}\sum_{i=1}^{I-k}(C_{i,k+1}/C_{i,k})$.

Figure 5: Expression 3 against C_i,k for product 1 and 2



Mack suggest that all this should be done for every development year with at least 6 data points, i.e for $k \leq I - 6$, however we chose do only include plots for $k = 1$ since we could not see any systematic residuals or non-random pattern for any development years. We can therefore conclude that the third chain-ladder assumption is satisfied.

Lastly we want to examine whether or not we have any calender year effects, the second chain-ladder assumption. To do this we follow the procedure explained in Appendix H in Mack's article. We begin by creating a triangle for the development factors which we can see in Table 6 for branch 1 and in Table 9 for branch 2. We continue by computing the column means in this triangle. Now, if the development factor for a cell is

- equal to the mean, we assign "-" to this cell,
- smaller than the mean, we assign "S" to this cell,
- larger than the mean, we assign "L" to this cell.

The result of this is shown in Table 7 and Table 10 respectively for branch 1 and 2. Following the article, we now want to compute the number of "S" and "L" for each diagonal A_j for $2 \leq j \leq 9$. We denote the number of "S" in a diagonal by S_j and the number of "L" by L_j. Furthermore, we compute

- $Z_j = \min(S_j, L_j)$,
- $n = S_j + L_j$,
- $m = [(n-1)/2]$, which denotes the largest integer $\leq (n-1)/2$,

- $E[Z_j] = \frac{n}{2} - \binom{n-1}{m}\frac{n}{2^n}$,
- $Var(Z_j) = \frac{n(n-1)}{4} - \binom{n-1}{m}\frac{n(n-1)}{2^n} + E[Z_j] - E[Z_j]^2$.

The result of this computation can be seen in Table 8 and Table 11 respectively for branch 1 and 2. We will now perform a test which test the null-hypothesis of no calendar year effects. We use the test statistic $Z = \sum_j Z_j$, and if Z is not in the following 95% confidence interval, $(E[Z] - 2\sqrt{Var(Z)}, E[Z] + 2\sqrt{Var(Z)})$, we can reject the null-hypothesis of no calendar year effects. Furthermore, we have that $E[Z] = \sum_j E[Z_j]$ and $Var(Z) = \sum_j Var(Z_j)$.

For branch 1 we got $Z = 16$ which is inside the corresponding confidence interval (8.4496157, 16.1753843). For branch 2 we got $Z = 17$ which is also inside the confidence interval (10.3296148, 18.3578852) and thus we do not reject the null-hypothesis of no calendar year effects for branch 2. We can thus go on with the chain-ladder method.

Table 6: Developement Factor Triangle for Branch 1

	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	1.438825	1.063897	1.027685	1.005655	1.002771	1.002359	1	1	1
2	1.504453	1.063773	1.017035	1.009924	1.000000	1.000000	1	1	
3	1.428381	1.056542	1.013050	1.000000	1.002287	1.000000	1		
4	1.454644	1.084219	1.020142	1.001564	1.000000	1.000000			
5	1.424097	1.055239	1.006690	1.003777	1.005104				
6	1.511984	1.066659	1.033457	1.004229					
7	1.393139	1.078068	1.019493						
8	1.453696	1.062982							
9	1.420193								

Table 7: Developement Factors as S or L if smaller or larger than column mean for Branch 1

	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	S	S	L	L	L	L	-	-	-
2	L	S	S	L	S	S	-	-	
3	S	S	S	S	L	S	-		
4	L	L	L	S	S	S			
5	S	S	S	S	L				
6	L	L	L	L					
7	S	L	S						
8	L	S							
9	S								

Table 8: Some results for Branch 1

Sj	Lj	Zj	n	m	EZj	VarZj
1	1	1	2	0	0.5000	0.2500000
2	1	1	3	1	0.7500	0.1875000
2	2	2	4	1	1.2500	0.4375000
2	3	2	5	2	1.5625	0.3710938
3	3	3	6	2	2.0625	0.6210938
4	2	2	6	2	2.0625	0.6210938
3	3	3	6	2	2.0625	0.6210938
4	2	2	6	2	2.0625	0.6210938

Table 9: Developement Factor Triangle for Branch 2

	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	3.525179	1.606818	1.208211	1.109809	1.041718	1.023025	1.004942	1.007367	1.005722
2	3.430653	1.509892	1.209785	1.099310	1.039753	1.022112	1.015870	1.002993	
3	3.649012	1.546228	1.222604	1.070419	1.041050	1.014010	1.013429		
4	3.471015	1.455413	1.196556	1.081089	1.054501	1.021094			
5	3.245088	1.522779	1.183407	1.099080	1.045179				
6	3.219295	1.533516	1.182730	1.099386					
7	3.771019	1.470298	1.187906						
8	3.483789	1.510468							
9	4.114286								

Table 10: Developement Factors as S or L if smaller or larger than column mean for Branch 2

	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	S	L	L	L	S	L	S	L	-
2	S	S	L	L	S	L	L	S	
3	L	L	L	S	S	S	L		
4	S	S	S	S	L	L			
5	S	L	S	L	L				
6	S	L	S	L					
7	L	S	S						
8	S	S							
9	L								

Table 11: Some results for Branch 2

Sj	Lj	Zj	n	m	EZj	VarZj
1	1	1	2	0	0.50000	0.2500000
1	2	1	3	1	0.75000	0.1875000
1	3	1	4	1	1.25000	0.4375000
3	2	2	5	2	1.56250	0.3710938
4	2	2	6	2	2.06250	0.6210938
4	3	3	7	3	2.40625	0.5537109
4	4	4	8	3	2.90625	0.8037109
3	5	3	8	3	2.90625	0.8037109

Exercise 3

In this exercise we wish to examine the variance parameter for the last development year (i.e. development year 10). We can do this by implementing the formulas presented in Mack's paper. First we have that

$$\widehat{\text{s.e.}}(R)^2 = \sum_{i=2}^{I} \widehat{\text{s.e.}}(C_{i,I})^2 + C_{i,I}\left(\sum_{j=i+1}^{I} C_{j,I}\right) \sum_{k=I+1-i}^{I-1} \frac{2\hat{\sigma}_k^2}{\hat{f}_k^2 \sum_{n=1}^{I-k} C_{n,k}}$$

where

$$\widehat{\text{s.e.}}(C_{i,I})^2 = C_{i,I}^2 \sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{C_{i,k}} + \frac{1}{\sum_{j=1}^{I-k} C_{j,k}} \right)$$

and where we estimate σ_k^2 by

$$\hat{\sigma}_k^2 = \frac{1}{I-k-1} \sum_{j=1}^{I-k} C_{j,k} \left(\frac{C_{j,k+1}}{C_{j,k}} - \hat{f}_k \right)^2$$

and

$$\hat{f}_k = \frac{\sum_{j=1}^{I-k} C_{i,k+1}}{\sum_{j=1}^{I-k} C_{i,k}}$$

as previously.

Table 12: Reserve Risk

Product 1	491194
Product 2	4586266

We see in Table 12 that there is vastly more risk ($\widehat{s.e}(R)$) in Product 2 than Product 1 which seems reasonable if we note that there seems to be no new claims during the last four development years, for branch 1, meaning that the risk only propagates in the initial six development years, at least according to our observed years. Another factor that leads to this difference is the fact that the ultimate claims reserve is greater for the first insurance product compared to the second.

Exercise 4

In this exercise we aim to predict the ultimate claim amounts R_1 and R_2 for insurance product 1 and 2 respectively. We then want to update our predictions assuming that we access to more data. In this case we use the first 10 years of data which have been fully developed whilst assuming that we only have access to a fraction of those years. Initially we assume to know the outcomes of the first 5 years which can be presented in the following paid claims triangles (Table 13 and 14).

Table 13: Paid claims triangle for product 1

	1	2	3	4	5	6	7	8	9	10
1	5781300	8973331	9562950	9685920	9743639	9778360	9778360	9778360	9778360	9778360
2	5130963	7663255	8150914	8294074	8406248	8406248	8406248	8406248	8406248	
3	12740801	18837941	20322702	20829689	20888367	20951190	20975966	20975966		
4	9746002	14596749	15847533	16170742	16248357	16248357	16303659			
5	11144988	16137484	17413958	17653527	17760105	17789609				
6	10803156	15227483	16522220	16776514	16859250					
7	11713414	17248401	18385807	18971032						
8	7175107	10668895	11470221							
9	8769619	13031521								
10	8265826									

Tables 13 and 14 can then be filled out using Mack's non-parametric CL approach, giving us the ultimate claim amounts for the two branches. What we then want to do is to iteratively assume additional years of data as known. In the first step this corresponds to assuming that the data present in the subantidiagonal is known and so on. When assuming this new data as know we also take it into account when estimating

Table 14: Paid claims triangle for product 2

	1	2	3	4	5	6	7	8	9	10
1	4027571	14282790	21133709	24754908	27106842	28366730	28804464	29258721	29475130	29475130
2	3446148	12180366	17506235	21515847	23349275	24127647	24693375	25037741	25340380	
3	4801933	14993873	22625793	27098683	29166886	30826270	31287603	31530472		
4	4418263	14979610	22820566	27052709	29126525	30193891	30936711			
5	3453026	11427253	18071289	22207881	24318283	25144653				
6	4392510	14600036	21689333	25874390	27948630					
7	3697717	12077496	17117187	20878180						
8	3052243	9903821	15291235							
9	5166485	17131735								
10	4191222									

the development factors f_k which we then use to fill out the claims triangle (trapezoid). From this filled out claims triangle we can then extract the predicted ultimate claims amount for each iteration. These values are presented in Table 15 below.

Table 15: Development of ultimate claim predictions given additional years of data

Branch	0	1	2	3	4	5	6	7	8
1	149073425	149623460	149699237	149687965	149613291	149669595	149659761	149652191	149652191
2	286830634	289014336	287807470	289468788	288615446	289142567	289111674	288964958	288761436

The main thing to take away from this is that the initial predictions seem to be, in retrospect, quite good predictors of the ultimate claim amounts.