Actuarial-Research-Conference_2022_Chain-Ladder_implementation-comparison

Quantification of Variability of Chain Ladder Reserve Estimates: Comparison of Mack's Method vs Bayesian Simulation Method Regarding Implementation Difficulties.

By Wenyi (Roy) Lu and Bill Lu (co-author), UT Dallas   Mar 18, 2022.

Brief Abstract

This presentation can serve as a hands-on guideline for practicing P&C actuaries to build their own in-house models to quantify the range estimates of outstanding loss reserve in everyday work under the requirement of European Solvency II or for business plan purpose in North America.

This presentation shows explicitly how to use basic Excel functions to carry out quantifying variability of property/casualty insurance loss reserve estimates according to Thomas Mack's paper (1993) entitled "Measuring the Variability of Chain Ladder Reserve Estimates" formula by formula.

This presentation also provides concrete description of how to set up a practical Bayesian simulation-based model in R according to Professor de Alba's paper (NAAJ 2002) "Bayesian Estimation of Outstanding Claim Reserves" for the same task.

Quantification of Variability of Chain Ladder Reserve Estimates: Mack's method vs Bayesian simulation method regarding implementation difficulties.

Wenyi Lu, UT Dallas / Bill Lu, UT Dallas

Aug 06, 2022

Outline

- Motivation and Goals
- Mack's papers in chronological order
- Some background briefing (for reading by self)
- Mack's formulas in Excel
- de Alba's Simulation Methods (if time permits)
- Pros and cons of Mack's vs de Alba's methods
- References
- Future Talks. Qs and As

Motivation and Goals

- Motivation
  To make practicing actuaries' life easier if they want to set up Mack's formulas. I can reproduce every number in Mack's paper, so I decide to share this verifiable and easy Excel skills.

- Goals
  1. To explain explicitly how in Excel to calculate the Range Estimates of Loss Outstanding Reserves.
  2. To explain explicitly the counter method with Bayesian simulation in concrete language.

- The Format of this presentation (hands-on and down-to-earth)
  It is like actuarial team internal training to disseminate knowledge of established models and to assign tasks later.

- My Assumption on the Audience
  The audience know the idea of Chain-Ladder point estimate.

Mack's papers in chronological order

- [1] Mack, T. 1993. "Distribution-Free Calculation of the Standard Error of Chain-Ladder Reserve Estimates." ASTIN
- [2] Mack, T. 1993. "Measuring the Variability of Chain Ladder Reserve Estimates" Casualty Actuarial Society (CAS)
- [3] MACK, T. 1994. "Which Stochastic Model is Underlying the Chain-Ladder Method?" Insurance: Mathematics and Economics
- [4] Mack, T. 1999. "The Standard Error of Chain-Ladder Reserve Estimates: Recursive Calculation And Inclusion of A Tail Factor" ASTIN

Note:
1. All papers are available in .pdf via google.
2. Paper [2] has two versions. The version from Faculty and Insitute of Acturies Claims Reserving Manual v.2 (09/1997) is much more legible because it is re-edited.

Background (better for reading by self)

- General insurance claims data structure after full settlement:
  Assume:
  1. This line of business has been stable in the past k years.
  2. The claims incurred in any origin/accident year will be fully settled after s years.

Table 1

Matrix of Claims by Year of Origin and Development Year

Year of Origin	Development Year					
	1	2	...	t	...	s
1	X_{11}	X_{12}	...	X_{1t}	...	X_{1s}
2	X_{21}	X_{22}	...	X_{2t}	...	X_{2s}
3	X_{31}	X_{32}	...	X_{3t}	...	X_{3s}
⋮						
k	X_{k1}	X_{k2}	...	X_{kt}	...	X_{ks}

Background

- Observed insurance claims data up to most recently reported:
  1. Only the earliest origin/accident year has fully developed.
  2. The most recent accident year is still in first 12 months of development.
  3. So the right lower triangle of the matrix is blank.
  4. We need to project the entries in the right lower triangle.

Table 2

Observed Claims Data Summarized in a Runoff Triangle

Year of Origin	Development Year						
	1	2	...	t	...	$k-1$	k
1	X_{11}	X_{12}	...	X_{1t}	...	$X_{1,k-1}$	X_{1k}
2	X_{21}	X_{22}	...	X_{2t}		$X_{2,k-1}$	—
3	X_{31}	X_{32}	...	X_{3t}		—	
⋮						—	—
$k-1$	$X_{k-1,1}$	$X_{k-1,2}$				—	—
k	X_{k1}	—					—

Background

- Well-known examples:
  1. In the required textbook for short-term actuarial mathematics (STAM) exam for more than 30 years.
  2. Claims from accident year 1988 is fully developed as of 1992.

Table 3
Cumulative Loss Payments through Development Years

Accident Year	Cumulative Loss Payments (in thousands) Development Year 0	1	2	3	4	Ultimate Number of Claims
1988	2,000	6,000	9,000	11,200	14,000	1,000
1989	2,600	6,840	10,920	15,600		1,200
1990	2,380	8,960	14,400			1,400
1991	3,120	10,800				1,500
1992	3,800					1,500

Source: Brown (1993).

Background

- Well-known examples (continued):
  1. This table contains the numbers of closed claims up to respective development age/year.
  2. Estimated ultimate number of claims were provided without justification/explanation. These are not consistent with the results from the most widely used deterministic method (Chain-Ladder method).

Table 4
Cumulative Closed Claims through Development Years

Accident Year	Cumulative Closed Number of Claims					Estimated Ultimate Number of Claims
	Development Year					
	0	1	2	3	4	
1988	400	700	850	930	1,000	1,000
1989	480	790	1,000	1,140		1,200
1990	500	950	1,190			1,400
1991	570	1,050				1,500
1992	600					1,500

Source: Brown (1993).

An Established Method

- Well-known examples: Chain-Ladder development method

Table 1

Cumulative Closed Claims Through Development Year

Accident Year	Development Year 0	1	2	3	4	Estimated Ult count	
1995	400	700	850	930	1,000	1,000	
1996	480	790	1,000	1,140		1,200	60
1997	500	950	1,190			1,400	210
1998	570	1,050				1,500	450
1999	600					1,500	900

Table 1-2

Development Factor of Cumulative Closed Claims Through Development Year

Accident Year	Development Year 1/0	2/1	3/2	4/3
1995	1.7500	1.2143	1.0941	1.0753
1996	1.6458	1.2658	1.1400	
1997	1.9000	1.2526		
1998	1.8421			
Average	1.7897	1.2459	1.1189	1.0753

Table 1-3

Cumulative Closed Claims Through Development Year

Accident Year	Development Year 0	1	2	3	4	Estimated Ult Count	Expected Outstanding Count C-L
1995	400	700	850	930	1,000	1,000	-
1996	480	790	1,000	1,140	*1,226*	1,226	86
1997	500	950	1,190	*1,332*	*1,432*	1,432	242
1998	570	1,050	*1,308*	*1,464*	*1,574*	1,574	524
1999	600	*1,074*	*1,338*	*1,497*	*1,610*	1,610	1,010
						Total	1,861

An Established Method

- Well-known examples: Chain-Ladder (C-L) method
  We can project cumulative and incremental closed claim counts respectively by each development year.

Table 1-3

	Cumulative Closed Claims Through Development Year						Expected
Accident	Development Year					Estimated	Outstanding
Year	0	1	2	3	4	Ult Count	Count C-L
1995	400	700	850	930	1,000	1,000	-
1996	480	790	1,000	1,140	*1,226*	1,226	86
1997	500	950	1,190	*1,332*	*1,432*	1,432	242
1998	570	1,050	*1,308*	*1,464*	*1,574*	1,574	524
1999	600	*1,074*	*1,338*	*1,497*	*1,610*	1,610	1,010
						Total	1,861

Table 1-4

	Incremental Closed Claims By Development Year						Expected
Accident	Development Year					Estimated	Outstanding
Year	0	1	2	3	4	Ult Count	Count C-L
1995	400	300	150	80	70	1,000	-
1996	480	310	210	140	*86*	1,226	86
1997	500	450	240	*142*	*100*	1,432	242
1998	570	480	*258*	*156*	*110*	1,574	524
1999	600	*474*	*264*	*159*	*113*	1,610	1,010
						Total	1,861

An Established Method

- Well-known examples: C-L method: Outstanding liabilities

Cumulative Loss Payments (in '000) Through Development Year

Accident Year	Development Year 0	1	2	3	4
1995	2, 000	6, 000	9, 000	11, 200	14, 000
1996	2, 600	6, 840	10, 920	15, 600	
1997	2, 380	8, 960	14, 400		
1998	3, 120	10, 800			
1999	3, 800				

Accident Year	Development Year 1/0	2/1	3/2	4/3
1995	3. 0000	1. 5000	1. 2444	1. 2500
1996	2. 6308	1. 5965	1. 4286	
1997	3. 7647	1. 6071		
1998	3. 4615			
Average	3. 2277	1. 5743	1. 3454	1. 2500

Accident Year	Development Year 1	2	3	4	Paid-to-Date	Reserve for Year	
1996				*19, 500*	15, 600	3, 900	0
1997			*19, 373*	*24, 217*	14, 400	9, 817	1
1998		*17, 003*	*22, 875*	*28, 594*	10, 800	17, 794	2
1999	*12, 265*	*19, 309*	*25, 979*	*32, 473*	3, 800	28, 673	3
Total						60, 184	

An Established Method

- Well-known examples: C-L method: Average Claim Sizes
  Standard assumption (only good for some very friendly LOBs).
  The difficulty is that incremental average claim payments are from partial settlements of different claims.
- Well-known examples: C-L method: Average Claim Sizes
  We don't cover this here.
  It is easy to read Professor Brown's Textbook listed in References [7] .

New Task: Range Estimates of C-L Reserves

- The Chain-Ladder is the most widely used method to calculate outstanding general insurance liabilities (or called reserves).
- The results of Chain-Ladder method is point estimates of the outstanding unfulfilled liabilities. They do not provide information about uncertainty in these estimates by nature.
- Regulations on general insurance in the US so far do not ask for reporting range estimates of outstanding reserves.
- Regulations on general insurance in Europe started to require reporting range estimates of outstanding reserves in last decade.
- Range estimates of outstanding reserves will help companies to understand their financial position too in the US regarding risk-based capital.

Therefore, it is beneficial to quantify range estimates of outstanding reserves.

C-L Reserves' Point Estimates: Mack's Method in Excel

Mack 1993 [2] Reinsurance Association of America (RAA)

Cumulative claims amounts by DY

AY	1	2	3	4	5	6	7	8	9	10	Estimate Ultimate	Estimate O/S amount
1	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834		
2	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	16,858	16,858	154
3	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466	23,863	24,083	24,083	617
4	5,655	11,555	15,766	21,266	23,425	26,083	27,067	27,967	28,441	28,703	28,703	1,636
5	1,092	9,565	15,836	22,169	25,955	26,180	27,278	28,185	28,663	28,927	28,927	2,747
6	1,513	6,445	11,702	12,935	15,852	17,649	18,389	19,001	19,323	19,501	19,501	3,649
7	557	4,020	10,946	12,314	14,428	16,064	16,738	17,294	17,587	17,749	17,749	5,435
8	1,351	6,947	13,112	16,664	19,525	21,738	22,650	23,403	23,800	24,019	24,019	10,907
9	3,133	5,395	8,759	11,132	13,043	14,521	15,130	15,634	15,898	16,045	16,045	10,650
10	2,063	6,188	10,046	12,767	14,959	16,655	17,353	17,931	18,234	18,402	18,402	16,339
DFs											total	52,135

Cumulative claims amounts DFs by DY

AY	2/1	3/2	4/3	5/4	6/5	7/6	8/7	9/8	10/9
1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01
2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03	
3	2.6	1.54	1.16	1.16	1.19	1.03	1.026		
4	2.0	1.36	1.35	1.10	1.11	1.04			
5	8.8	1.66	1.40	1.17	1.01				
6	4.3	1.82	1.11	1.23					
7	7.2	2.72	1.12						
8	5.1	1.89							
9	1.7								
10									
mean wt avg	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009
CDFs	8.920	2.974	1.832	1.441	1.230	1.105	1.060	1.026	1.009

mature sizes stable

matched Mack's

volume-wted: good.

C-L Reserves' Point Estimates: Mack's Method in Excel

Mack 1993 [2] Reinsurance Association of America (RAA)

1. We match Mack's result (every number), say, Column M.
2. Row 86 the formula is wighted average mean (volume-wighted mean).
3. CDFs (cumulative development factors) more stable than simple average.

This is point estimate only so far. We will next look at range estimates.

Mack 1993 [1] paper has a bit different notation, for instance, some bold-face letter with a hat. Please read paper [1] and [2] side by side. This helps a lot.

C-L Reserves' Range Estimates: Mack's formulas

I : mature age

I = 10 in the example

$$(7) \quad (\mathrm{s.e.}(C_{iI}))^2 = C_{iI}^{\;2} \sum_{k=I+1-i}^{I-1} \frac{\alpha_k^2}{f_k^2} \cdot \left(\frac{1}{C_{ik}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}} \right)$$

key: Don't think it as double summations.

4 horizontal vectors

09/97 D6.11

PAPERS OF MORE ADVANCED METHODS

where

$$(8) \quad \alpha_k^{\;2} = \frac{1}{I-k-1} \sum_{j=1}^{I-k} C_{jk} \left(\frac{C_{j,k+1}}{C_{jk}} - f_k \right)^2, \; 1 \le k \le I-2$$

C-L Reserves' Range Estimates: Mack's formulas

We have now shown how to establish confidence limits for every R_i and therefore also for every $C_{iI} = C_{i,I+1-i} + R_i$. We may also be interested in having confidence limits for the overall reserve

$$R = R_2 + \ldots + R_I$$

and the question is whether, in order to estimate the variance of R, we can simply add the squares $(\text{s.e.}(\mathbf{R}_i))^2$ of the individual standard errors as would be the case with standard deviations of independent variables. But unfortunately, whereas the R_i's themselves are independent, the estimators $\mathbf{R}_i$ are not because they are all influenced by the same age-to-age factors $\mathbf{f}_k$, that is the $\mathbf{R}_i$'s are positively correlated. In Appendix F it is shown that the square of the standard error of the overall reserve estimator

$$\mathbf{R} = \mathbf{R}_2 + \ldots + \mathbf{R}_I$$

cov part : structure close to (7)

is given by

$$(11)\quad (\text{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I} \left\{ (\text{s.e.}(\mathbf{R}_i))^2 + \mathbf{C}_{iI} \left(\sum_{j=i+1}^{I} \mathbf{C}_{jI} \right) \sum_{k=I+1-i}^{I-1} \frac{2\alpha_k^2 / \mathbf{f}_k^2}{\sum_{n=1}^{I-k} C_{nk}} \right\}$$

C-L Reserves' Range Estimates: Mack's results

k	1	2	3	4	5	6	7	8	9
α_k^2	27883	1109	691	61.2	119	40.8	1.34	7.88	

A plot of $\ln(\alpha_k^2)$ against k is given in Figure 13 and shows that there indeed seems to be a linear relationship which can be used to extrapolate $\ln(\alpha_9^2)$. This yields $\alpha_9^2 = \exp(-.44) = .64$. But we use formula (9) which is more easily programmable and in the present case is a bit more on the safe side: it leads to $\alpha_9^2 = 1.34$. Using formula (11) for s.e.(**R**) as well we finally obtain

	$C_{i,10}$	R_i	$s.e.(C_{i,10}) = s.e.(R_i)$	$s.e.(R_i)/R_i$
i=2	16858	154	206	134%
i=3	24083	617	623	101%
i=4	28703	1636	747	46%
i=5	28927	2747	1469	53%
i=6	19501	3649	2002	55%
i=7	17749	5435	2209	41%
i=8	24019	10907	5358	49%
i=9	16045	10650	6333	59%
i=10	18402	16339	24566	150%
Overall		52135	26909	52%

no specific explanation by Mack. → but I got it.

C-L Reserves' Range Estimates: formula (8) in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
59			Cumulative claims amounts by DY									Estimate	Estimate
60					DY							Ultimate	O/S amount
61	AY	1	2	3	4	5	6	7	8	9	10		
62	1	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834		
63	2	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	*16,858*	16,858	154
64	3	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466	*23,863*	*24,083*	24,083	617
65	4	5,655	11,555	15,766	21,266	23,425	26,083	27,067	*27,967*	*28,441*	*28,703*	28,703	1,636
66	5	1,092	9,565	15,836	22,169	25,955	26,180	*27,278*	*28,185*	*28,663*	*28,927*	28,927	2,747
67	6	1,513	6,445	11,702	12,935	15,852	*17,649*	*18,389*	*19,001*	*19,323*	*19,501*	19,501	3,649
68	7	557	4,020	10,946	12,314	*14,428*	*16,064*	*16,738*	*17,294*	*17,587*	*17,749*	17,749	5,435
69	8	1,351	6,947	13,112	*16,664*	*19,525*	*21,738*	*22,650*	*23,403*	*23,800*	*24,019*	24,019	10,907
70	9	3,133	5,395	*8,759*	*11,132*	*13,043*	*14,521*	*15,130*	*15,634*	*15,898*	*16,045*	16,045	10,650
71	10	2,063	*6,188*	*10,046*	*12,767*	*14,959*	*16,655*	*17,353*	*17,931*	*18,234*	*18,402*	18,402	16,339
72	DFs											total	52,135
73			Cumulative claims amounts DFs by DY										
74					DY								
75	AY	2/1	3/2	4/3	5/4	6/5	7/6	8/7	9/8	10/9			
76	1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01			
77	2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03				
78	3	2.6	1.54	1.16	1.16	1.19	1.03	1.026					
79	4	2.0	1.36	1.35	1.10	1.11	1.04						
80	5	8.8	1.66	1.40	1.17	1.01							
81	6	4.3	1.82	1.11	1.23								
82	7	7.2	2.72	1.12									
83	8	5.1	1.89										
84	9	1.7											
85	10												
86	mean wt avg	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009			
91	f_k1	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009			
124	k	1	2	3	4	5	6	7	8	9			
125	alpha_k_sq	27883	1109	691	61.2	119.4	40.8	1.34	7.88	1.34			

C-L Reserves' Range Estimates: formula (7) in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
60				DY								Ultimate	O/S amount
61	AY	1	2	3	4	5	6	7	8	9	10		
62	1	5, 012	8, 269	10, 907	11, 805	13, 539	16, 181	18, 009	18, 608	18, 662	18, 834		
63	2	106	4, 285	5, 396	10, 666	13, 782	15, 599	15, 496	16, 169	16, 704	16, 858	16, 858	154
64	3	3, 410	8, 992	13, 873	16, 141	18, 735	22, 214	22, 863	23, 466	23, 863	24, 083	24, 083	617
65	4	5, 655	11, 555	15, 766	21, 266	23, 425	26, 083	27, 067	27, 967	28, 441	28, 703	28, 703	1, 636
66	5	1, 092	9, 565	15, 836	22, 169	25, 955	26, 180	27, 278	28, 185	28, 663	28, 927	28, 927	2, 747
67	6	1, 513	6, 445	11, 702	12, 935	15, 852	17, 649	18, 389	19, 001	19, 323	19, 501	19, 501	3, 649
68	7	557	4, 020	10, 946	12, 314	14, 428	16, 064	16, 738	17, 294	17, 587	17, 749	17, 749	5, 435
69	8	1, 351	6, 947	13, 112	16, 664	19, 525	21, 738	22, 650	23, 403	23, 800	24, 019	24, 019	10, 907
70	9	3, 133	5, 395	8, 759	11, 132	13, 043	14, 521	15, 130	15, 634	15, 898	16, 045	16, 045	10, 650
71	10	2, 063	6, 188	10, 046	12, 767	14, 959	16, 655	17, 353	17, 931	18, 234	18, 402	18, 402	16, 339
72	DFs											total	52, 135
91	f_k1	2. 999	1. 624	1. 271	1. 172	1. 113	1. 042	1. 033	1. 017	1. 009			
124	k	1	2	3	4	5	6	7	8	9			
125	alpha_k_sq	27883	1109	691	61. 2	119. 4	40. 8	1. 34	7. 88	1. 34			
126	b/4 curr CY expo by age	21829	60078	84426	94982	95436	80077	56368	34777	18662			
127													
128	i	c_i, 10	R_i	se(c_i, 10) =se(R_i)	se(R_i)/ R_i	Cov part							
129	2	16858	154	206	134%	650		650					
130	3	24083	617	623	101%	1463		1463					
131	4	28703	1636	747	46%	1495		1495					
132	5	28927	2747	1469	53%	2081		2081					
133	6	19501	3649	2002	55%	2308		2308					
134	7	17749	5435	2209	41%	2166		2166					
135	8	24019	10907	5358	49%	3483		3483					
136	9	16045	10650	6333	59%	2909		2909					
137	10	18402	16339	24566	150%	10290							
138													
139	Overall		52135	26909	52%	28809	55%	26909	52%				

C-L Reserves' Range Estimates: formula (11) in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
60				DY								Ultimate	O/S amount
61	AY	1	2	3	4	5	6	7	8	9	10		
62	1	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834		
63	2	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	*16,858*	16,858	154
64	3	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466	*23,863*	*24,083*	24,083	617
65	4	5,655	11,555	15,766	21,266	23,425	26,083	27,067	*27,967*	*28,441*	*28,703*	28,703	1,636
66	5	1,092	9,565	15,836	22,169	25,955	26,180	*27,278*	*28,185*	*28,663*	*28,927*	28,927	2,747
67	6	1,513	6,445	11,702	12,935	15,852	*17,649*	*18,389*	*19,001*	*19,323*	*19,501*	19,501	3,649
68	7	557	4,020	10,946	12,314	*14,428*	*16,064*	*16,738*	*17,294*	*17,587*	*17,749*	17,749	5,435
69	8	1,351	6,947	13,112	*16,664*	*19,525*	*21,738*	*22,650*	*23,403*	*23,800*	*24,019*	24,019	10,907
70	9	3,133	5,395	*8,759*	*11,132*	*13,043*	*14,521*	*15,130*	*15,634*	*15,898*	*16,045*	16,045	10,650
71	10	2,063	*6,188*	*10,046*	*12,767*	*14,959*	*16,655*	*17,353*	*17,931*	*18,234*	*18,402*	18,402	16,339
72	DFs											total	52,135
91	f_k1	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009			
124	k	1	2	3	4	5	6	7	8	9			
125	alpha_k_sq	27883	1109	691	61.2	119.4	40.8	1.34	7.88	1.34			
126	b/4 curr CY expo by age	21829	60078	84426	94982	95436	80077	56368	34777	18662			
127													
128	i	c_i,10	R_i	se(c_i,10) =se(R_i)	se(R_i)/ R_i	Cov part							
129	2	16858	154	206	134%	650		650					
130	3	24083	617	623	101%	1463		1463					
131	4	28703	1636	747	46%	1495		1495					
132	5	28927	2747	1469	53%	2081		2081					
133	6	19501	3649	2002	55%	2308		2308					
134	7	17749	5435	2209	41%	2166		2166					
135	8	24019	10907	5358	49%	3483		3483					
136	9	16045	10650	6333	59%	2909		2909					
137	10	18402	16339	24566	150%	10290		2909					
138													
139	Overall		52135	26909	52%	28809	55%	26909	52%				

Excel → not smart enough.

bingo!

C-L Reserves' Range Estimates: Mack's formulas in Excel

1. Formula (8): for k=8, for instance, cell "I125" has input of "I62:I63" as weights, "I76:I77" as individual ratios, and "I91" as the observed center.

2. Formula (7): for i=3, for instance, cell "D130" has input of rows 64, 91,125,126 and columns I,J. The input is 4 horizotal vectors of length of 2 elements.
The notation for (7) is not easy as C's can be either observed or estimated values in the table.

3. Formula (11): for i=3, for instance, the covariance part cell "F130" has input of rows 91,125,126, and columns I,J, plus "B130" with "B131:B137" pairs.

Key: Row 125 is Mack's contribution, and we learned from him to set up row 126 similarly. Then formula (7) becomes sumproduct of 4 vectors.
Cell "F137" needs to be 0, and Excel is not smart enough.

Counter Method: Bayesian simulation

- Assumptions implied in Chain-Ladder method

  1. Claim counts independent of average claim amounts (severity) in a period (a cell in the table).

  2. Development yeat t (column index t) behaves the same for each year of origin/accident.

  3. The average claim amounts (severity) in a period (a cell in the table) follows lognormal (reasonable start point.)

A New Bayesian Method: Data Claim Counts

$N_1 = n_1$, the total numbers of ultimately settled for year 1 (the earliest) of origin. This is known (the only known) as we assume the ultimate year of fully settled of every claim origin year is 5 (the end of index 4). Without loss of generality, assume k is the ultimate year of fully settled of all claims for every year of origin.

$\mathbf{x}_1 = (x_{11}, x_{12}, \ldots, x_{1,k-1}, x_{1k})$,
$\mathbf{x}_2 = (x_{21}, x_{22}, \ldots, x_{2,k-1})$,
$\vdots$
$\mathbf{x}_k = (x_{k1})$.

k fixed here

v.s. Mack's paper, k is index, varying.

Note that $\mathbf{x}_1$ is all known, $\mathbf{x}_2$ has x_{2k} unknown, and $\mathbf{x}_k$ has only x_{k1} known. As a result, $N_2, N_3, \ldots, N_k$ are unknown, with N_2 having most certainty and N_k least certainty.

Let $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ denote the vector of proportions of claims settled in the vector of the development years. This vector of parameters is stable (the same) for every year of origin.

We know that $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i,k-1}, x_{ik}) = Mult_k(N_i, \mathbf{p})$, * given $N_i, \mathbf{p}$.

* means hard for actuaries!

Bayesian: Priors/Posteriors about Claim Counts

Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, n_1 = sum(\mathbf{x}_1)\}$ denote the information available in the left upper triangle.

Let $x_i^* = sum(\mathbf{x}_i)$

Let $p_s^* = p_1 + p_2 + \cdots + p_s = 1 - p_{s+1} - \cdots - p_k$

Result1: Using non-informative priors for $(N_2 = n_2, N_3 = n_3, \ldots, N_k = n_k, \mathbf{p})$, we have $f(N_2 = n_2, N_3 = n_3, \ldots, N_k = n_k, \mathbf{p}|D) \propto$

$$[\prod_{i=2}^{k} C(n_i, n_i - x_i^*)(p_{k-i+1}^*)^{x_i^*}(1 - p_{k-i+1}^*)^{n_i - x_i^*}]\times$$

$$\{(\prod_{t=1}^{k} p_t^{x_{1t}})[\prod_{t=1}^{k-1}(\frac{p_t}{p_{k-1}^*})^{x_{2t}}]\ldots[\prod_{t=1}^{2}(\frac{p_t}{p_2^*})^{x_{k-1,t}}]\},$$

where $C(m, n)$ means m choose n combination formula. Result1 means the product of $(k-1)$ independent negative binomials for the $n_i, i = 2, 3, \ldots, k$, and a Dirichlet for $\mathbf{p}$. *

* hard for actuaries

Bayesian: Priors/Posteriors about Claim Counts

Result2:

$$f(\mathbf{p}|D) \propto \{(\prod_{t=1}^{k} p_t^{x_{1t}})[\prod_{t=1}^{k-1}(\frac{p_t}{p_{k-1}^*})^{x_{2t}}]\dots[\prod_{t=1}^{2}(\frac{p_t}{p_2^*})^{x_{k-1,t}}]\}$$

Proof (Result2): Directly taking summation on $n_2, n_3, \dots, n_k$ respectively will yield result2.
Note that even with $f(\mathbf{p}|D)$ given, we can't use it directly. Since we do not have the same information on $\mathbf{p}$ for each origin/accident year, we need to express this posterior *pdf* differently.

$$f(\mathbf{p}|D) \propto f(p_k|D)f(p_{k-1}|p_k, D)f(p_{k-2}|p_{k-1}, p_k, D)\dots f(p_2|p_3, p_4, \dots, p$$

$$\propto f(p_k|D)f(\frac{p_{k-1}}{p_{k-1}^*}|p_k, D)f(\frac{p_{k-2}}{p_{k-2}^*}|p_{k-1}, p_k, D)\dots f(\frac{p_2}{p_2^*}|p_3, p_4, \dots, p_k, D)$$

With the writing in the last 'proportional to', we can have Result3.

Bayesian: Priors/Posteriors about Claim Counts

Result3:

$$f(p_k|D) = Beta(x_{1,k} + 1, \sum_{t=1}^{k-1} x_{1,t} + 1)$$

$$f(\frac{p_{k-1}}{p^*_{k-1}}|p_k, D) = Beta(x_{1,k-1} + x_{2,k-1} + 1, \sum_{t=1}^{k-2}(x_{1,t} + x_{2,t}) + 1)$$

$$f(\frac{p_{k-2}}{p^*_{k-2}}|p_{k-1}, p_k, D) = Beta(x_{1,k-2}+x_{2,k-2}+x_{3,k-2}+1, \sum_{t=1}^{k-3}(x_{1,t}+x_{2,t}+x_{3,t}$$

$$\vdots$$

$$f(p_1|p_2, p_3, \ldots, p_k, D) = 1$$

With result3, I will simulate the unknown lower right triangle for numbers of claims as follows:

A New Method: A Bayesian model for the numbers of claims.

- Assumptions implied in Chain-Ladder method
  Let X_{it} = number of claims in $t-th$ development year.
  The available information is
  { X_{it}: i = 1, . . . , k; t =1, . . . , k; i+t ≤k+1 }
  Let N_i = total number of claims for origin year i, i = 1, . . . , k.
  The available information is
  $N_1 = n_1$ is observed, while $N_2, \ldots, N_k$ not observed yet.
- Theoretical results for N_i and X_{it}
  Assume $X_{it} \sim POI(\lambda_t)$ and X_{it} are independent.
  Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})'$, then Robert G.D. Steel 1953:

$$\mathbf{X}_i | N_i = n_i \sim Mult_k(n_i; p_1, \ldots, p_k), with\ p_t = \lambda_t / \sum_{t=1}^{k} \lambda_t$$

Simulating the unknown triangle claim counts.

- Simulation steps for the number of claims.
  1. To generate $p_k^{(j)}$, the proportion of k-th column from a

$$Beta(x_{1,k}+1, \sum_{t=1}^{k-1} x_{1,t}+1)$$

.

  2. To generate $\tilde{\theta}_{k-1}^{(j)}$, the relative proportion the (k-1)-th column out of the first (k-1) columns only, from a

$$Beta(x_{1,k-1}+x_{2,k-1}+1, \sum_{t=1}^{k-2}(x_{1,t}+x_{2,t})+1)$$

.

  3. Use the results of steps 1 and 2 to generate $p_{k-1}^{(j)} = \tilde{\theta}_{k-1}^{(j)}(1-p_k^{(j)})$
  4. To generate $\tilde{\theta}_{k-2}^{(j)}$ from $Beta(x_{1,k-2}+x_{2,k-2}+x_{3,k-2}+1, \sum_{t=1}^{k-3}(x_{1,t}+x_{2,t}+x_{3,t})+1)$.

Simulating the unknown triangle claim counts.

5. Use the results of steps 1-4 above to generate $p_{k-2}^{(j)} = \tilde{\theta}_{k-2}^{(j)}(1 - p_{k-1}^{(j)} - p_k^{(j)})$, and so on to $p_2^{(j)}$; the remaining proportion is $p_1^{(j)} = 1 - \sum_{i=2}^{k} p_i^{(j)}$. With this, we will have generated a vector $\mathbf{p}^{(j)} = (p_k^{(j)}, p_{k-1}^{(j)}, \ldots, p_1^{(j)})$.
6. Use this $\mathbf{p}^{(j)}$ and

$$n_i|\mathbf{p}, D \sim NB(x_i^*, p_{k-i+1}^*),$$

$i = 2, \ldots, k$, to generate an observation for each n_i. where $x_i^* = x_{i,1} + \cdots + x_{i,k-i+1}$, and $p_{k-i+1}^* = p_1 + \cdots + p_{k-i+1}$. Thus, $(n_2, n_3, \ldots, n_k) = \mathbf{n}^{(j)}$.
7. Use $\mathbf{n}^{(j)}$, $\mathbf{p}^{(j)}$ to generate observations for the unknown portions of $\mathbf{x}_i^{(j)}$ from each of $(k-1)$ multinomials (one for each year): $f(x_{i1}^{(j)}, x_{i1}^{(j)}, \ldots, x_{i1}^{(j)}|n_i^{(j)}, \mathbf{p}^{(j)})) = Mult_k(n_i^{(j)}; \mathbf{p}^{(j)})$, $i = 2, \ldots, k$. For the known part, we discard the generated values.

Bayesian: data/model about average claim amounts.

$\mathbf{M}_1 = (M_{11}, M_{12}, \ldots, M_{1,k-1}, M_{1k})$,
$\mathbf{M}_2 = (M_{21}, M_{22}, \ldots, M_{2,k-1})$,
$\vdots$
$\mathbf{M}_k = (M_{k1})$.

Note: * too hard for actuaries

Note that T_U=(k+1)k/2 is the number of cells of the left upper triangle having observed M_{it}.
Denote D' for the observed information collection of M_{it}.
Assume:
$\log(M_{it}) = y_{it} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2) \quad (5.1)$
This is an unbalanced ANOVA model. *
Using matrix notation, (5.1) can be written as follows:
$\mathbf{y} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$,
where $\mathbf{y}$ is a T_U-dimension vector that contains all observed y_{it}, $\boldsymbol{\theta}$=$(\mu, \alpha_2, \alpha_3, \ldots, \alpha_k, \beta_2, \beta_3, \ldots, \beta_k)'$ is the ((2k-1)$\times$1) vector of parameters, $\boldsymbol{\epsilon}$ is the ($T_U \times$1) vector of errors, and W is the ($T_U \times$(2k-1)) design matrix of the model. *

Bayesian: Priors/Posteriors about average claim amounts.

Here α_1=0 and β_1=0 is imposed to make sure W has full rank, meaning the estimability of the parameters. *

With non-informative priors* for independent $\boldsymbol{\theta}$ and σ, $f(\boldsymbol{\theta}, \sigma) \propto (1/\sigma)$. As a result, the posterior joint distribution is

$$f(\boldsymbol{\theta}, \sigma | D') \propto \sigma^{-(T_U+1)} \times \exp[-\frac{1}{\sigma^2}(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})]$$

$$= \sigma^{-(T_U+1)} \times \exp[-\frac{1}{\sigma^2} \times SST] \rightarrow$$

Note:

$SST = SSE + SSR = (\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}) + (\mathbf{W}\hat{\boldsymbol{\theta}} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{W}\hat{\boldsymbol{\theta}} - \mathbf{W}\boldsymbol{\theta})$, where $\hat{\boldsymbol{\theta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$.

Also note:

$f(\boldsymbol{\theta}, \sigma | D') \propto f(\boldsymbol{\theta} | \sigma, D') f(\sigma | D')$, where

$$f(\boldsymbol{\theta} | \sigma, D') \propto \sigma^{-(2k-1)} \times \exp[-\frac{1}{\sigma^2}(\mathbf{W}\hat{\boldsymbol{\theta}} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{W}\hat{\boldsymbol{\theta}} - \mathbf{W}\boldsymbol{\theta})]$$

Bayesian: Priors/Posteriors about average claim amounts.

Therefore

$$f(\sigma|D') \propto \sigma^{-(T_U-2k+2)} \times \exp[-\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})]$$

If we let $\lambda = 1/\sigma^2$, then $d\sigma/d\lambda \propto \lambda^{-3/2}$. As a result,

$$f(\lambda|D') \propto \lambda^{(T_U-2k+2)/2} \times \exp[-\lambda(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})] \times \lambda^{-3/2}$$

$$\propto \lambda^{[(T_U-2k+1)/2]-1} \times \exp[-\lambda(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})].$$

Therefore σ is from a "square-root inverted-gamma" distribution with parameters shape $\alpha = (T_U - 2k + 1)/2$ and rate $\beta = (\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{W}\hat{\boldsymbol{\theta}})/2$.

I will now simulate the unknown lower right triangle for severity of claims and compute reserves as follows:

Simulating the unknown triangle average claim amounts.

- Simulation steps for claim amounts.

  8. Generate an observation $\sigma^{(j)}$ from a "square-root inverted-gamma" distribution with parameters shape $\alpha = (T_U - 2k + 1)/2$ and rate $\beta = (\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})'(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})/2$. This can be done by first getting an observation $g^{(j)}$ from a $gamma(\alpha, rate = \beta)$ and then making $\sigma^{(j)} = 1/\sqrt{g^{(j)}}$.

  9. Generate an observation $\boldsymbol{\theta}^{(j)} = (\mu^{(j)}, \alpha_2^{(j)}, \ldots, \alpha_k^{(j)}, \beta_2^{(j)}, \ldots, \beta_k^{(j)})'$ from $N(\hat{\boldsymbol{\theta}}, \sigma^{(j)2}(\mathbf{W}'\mathbf{W})^{-1})$.
  This can be done in R by "mvnfast" package.

  10. Generate an observation from the predictive distribution $N(\mu_{it}^{(j)}, \sigma^{(j)}) \rightarrow Y_{it}^{(j)}$, with $\mu_{it}^{(j)} = \mu^{(j)} + \alpha_i^{(j)} + \beta_t^{(j)}$ for each (i, t) in the right lower triangle, and compute $M_{it}^{(j)} = \exp\{Y_{it}^{(j)}\}$

  11. $Z_{it}^{(j)} = X_{it}^{(j)} M_{it}^{(j)}$ for each $(i, t), i = 2, \ldots, k, t > k - i + 1$.

  12. To obtain the total reserves $R^{(j)} = \sum_{i,t} Z_{it}^{(j)}$.

Simulating the unknown triangle average claim amounts.

- Simulation steps for claim amounts when ~~seveirty~~ severity is not known or when negative incrementals are present.

  Intuitively just use aggregate cummulative claims amounts in step 8-12, meaning we set $M_{it}^{(j)} = 1$ for every cell in the table.

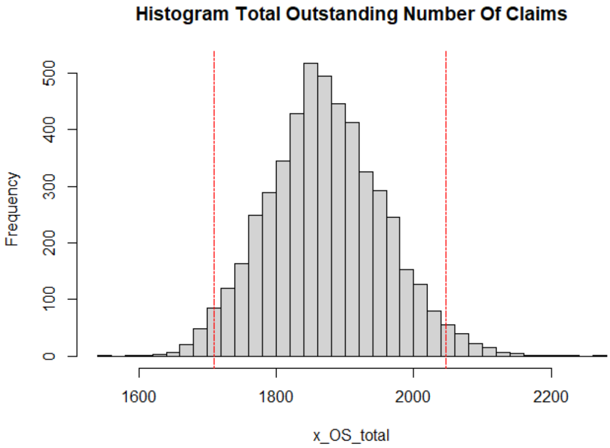
Bayesian models design matrix example.

- Make sure the design matrix is of full rank. Pay attention to index of the observation vector too.
  For k = 5, design matrix has shape of 15 x 9.

	A	B	C	D	E	F	G	H	I	J	K
1	k=5		1	2	3	4	5	6	7	8	9
2	formula-driven							2	3	4	5
3	1	1	1	0	0	0	0	0	0	0	0
4	1	2	1	0	0	0	0	1	0	0	0
5	1	3	1	0	0	0	0	0	1	0	0
6	1	4	1	0	0	0	0	0	0	1	0
7	1	5	1	0	0	0	0	0	0	0	1
8	2	1	1	1	0	0	0	0	0	0	0
9	2	2	1	1	0	0	0	1	0	0	0
10	2	3	1	1	0	0	0	0	1	0	0
11	2	4	1	1	0	0	0	0	0	1	0
12	3	1	1	0	1	0	0	0	0	0	0
13	3	2	1	0	1	0	0	1	0	0	0
14	3	3	1	0	1	0	0	0	1	0	0
15	4	1	1	0	0	1	0	0	0	0	0
16	4	2	1	0	0	1	0	1	0	0	0
17	5	1	1	0	0	0	1	0	0	0	0

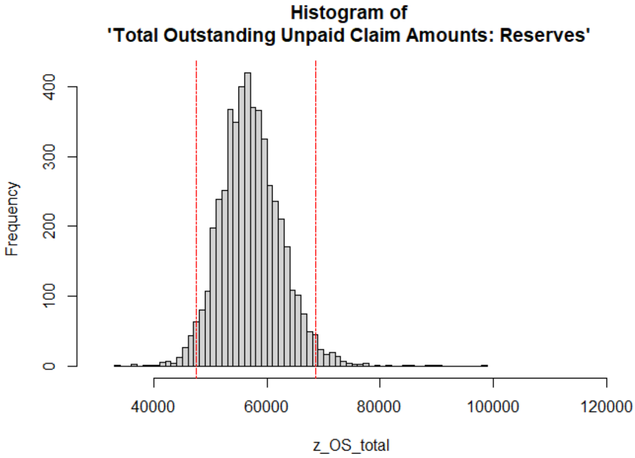
Bayesian models results for the unknown triangle.

- Simulation (5000 times) results for claim numbers.
  The mean of the predictive distribution is $1,872$, very close to C-L method's estimate. The 95% credible interval contains $1,861$ well in the center.



Bayesian models results for the unknown triangle.

- Simulation (5000 times) results for claim amounts.
  With severity information included, the mean of the predictive distribution is \$57, 158, very close to C-L method's estimate. The 95% credible interval contains \$60, 184 well in the center.



Comparison of Difficulties of Implementation

- Mack's method via Chain-Ladder:

  1. Pros: Easy to implement; can handle negative incremental payments; standard method for long time.

  2. Cons: No landscape of the distribution, say, skewed or not?

- Bayesian method:

  1. Pros: Can provide more information because simulation provides full landscape of the required estimates.

  2. Cons: Hard to implement because learning curve is very high to practicing actuaries; still more like blackbox,* e.g, first, the easiest part of design matrix really is not that easy for practicing actuaries already; second, in Professor de Alba's paper, he did not explain why in one of his example including severity information produced much worse result than ignoring severity informaion. Verral 1990

References

- [1] Mack, T. 1993. “Distribution-Free Calculation of the Standard Error of Chain-Ladder Reserve Estimates.” ASTIN Bulletin, Vol. 23 No. 2, 1993
- [2] Mack, T. 1993. “Measuring the Variability of Chain Ladder Reserve Estimates” Casualty Actuarial Society (CAS) prize-winning paper competition on 'Variability of Loss Reserves'.
- [3] MACK, T. 1994. “Which Stochastic Model is Underlying the Chain Ladder Method?” Insurance: Mathematics and Economics 15: 133–8.
- [4] Mack, T. 1999. “The Standard Error of Chain-Ladder Reserve Estimates: Recursive Calculation And Inclusion of A Tail Factor” ASTIN Bulletin, Vol. 29 No. 2, 1999
- [5] Mack, T. 2008. “The Prediction Error of Bornhuetter-Ferguson” Casualty Actuarial Society (CAS) E-Forum, Fall 2008.

References

- [6] DE ALBA, E. 2002. Bayesian Estimation of Outstanding Claim Reserves. North American Actuarial Journal, 2002. ISSN: 1092-0277
- [7] BROWN, R.L. 2001. Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance. Winsted, CT: ACTEX Publications.
- [8] RENSHAW, A.E., AND R. VERRALL. 1994. "A Stochastic Model Underlying the Chain-Ladder Technique," Proceedings of the XXV ASTIN Colloquium, Cannes.
- [9]
- [10]

Future Presentations

- Chain-Ladder Recursive way with tail ~~tactor~~ factor in Excel; Bornhuetter-Ferguson way in Excel (one separate talk).

- Thank You!

- Qs and As