The Actuarial Profession
making financial sense of the future

Non parametric IBNER projection

Claude Perret
Hannes van Rensburg
Farshad Zanjani

GIRO 2009, Edinburgh

Agenda

- Introduction & background
- Why is IBNER important?
- Method description
- Issues
- Examples

Introduction

- Original intention for a paper but further research in progress
  - What to call it??? Kernel regression, GLM proxy, CHF method, Generalized Chain-Ladder?
  - Further research
  - The road to hell is paved with good intentions
- Instead, you get this presentation and a few freebies (codes in SAS, Excel VBA and Excel)
- This presentation proposes a new approach for projecting individual claims to an ultimate position.

Introduction

- Extensive literature for projecting aggregated claims
  - methods for separating the IBNYR from the IBNER
  - hardly any for estimating individual claims IBNER
  - Individual claims uncertainty range associated
- Why no interest?
  - difficult to get individual claims triangles (systems + data size)
  - IBNER more an issue for pricing than reserving

Background

Employers Liability Project for London Market portfolio

- Subject to deductibles, large losses and poor exposure information e.g. location of the risk
- Scope was purposely very broad and included the following investigations:
  - Historical claims severity inflation by year
  - Increased Limits Factor for pricing
  - Prepare claims data for predictive modelling (GLM)
  - Derive IBNER development factors for pricing

Background

- Needed method to project individual claims
- Several approaches investigate, none gave convincing results on individual claim basis
- Overall IBNER amount credible, not at the individual claim level
- General weakness of methods:
  - heavy reliance on the last known position of claim
  - Not allowing for differences in small, medium or large development patterns
- We required ultimate claims distribution to be dispersed in a realistic way and not form “blobs” of data

Background

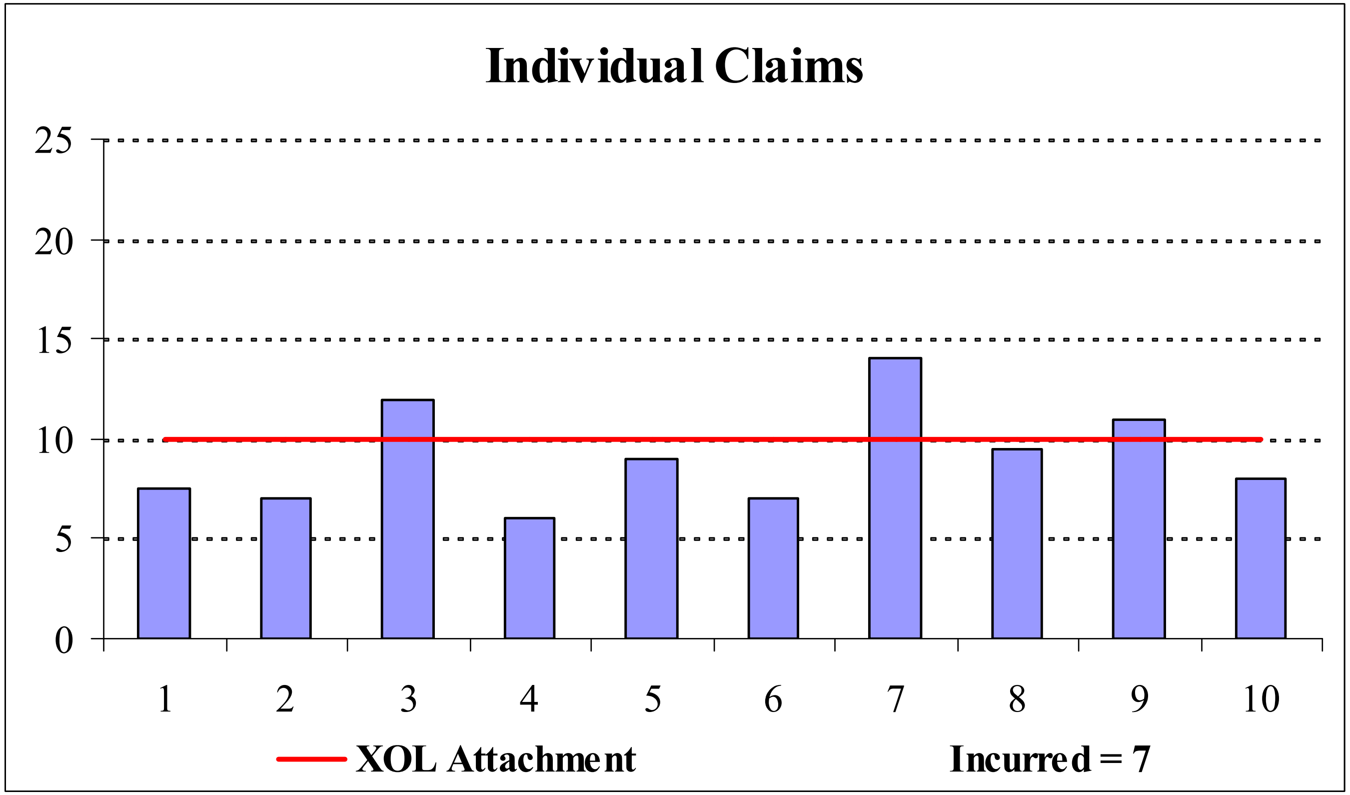
- The words "realistic" and "credible" can be quite subjective
- Definition for credible is based on historical experience, which is the usual approach adopted for most actuarial work (e.g. chain-ladder)
- Implies that we would like to see individual claims projected in line with other comparable claims that are more mature
- This is the key requirement that led us to this method

Background

Below is a recap of the methods we tried:

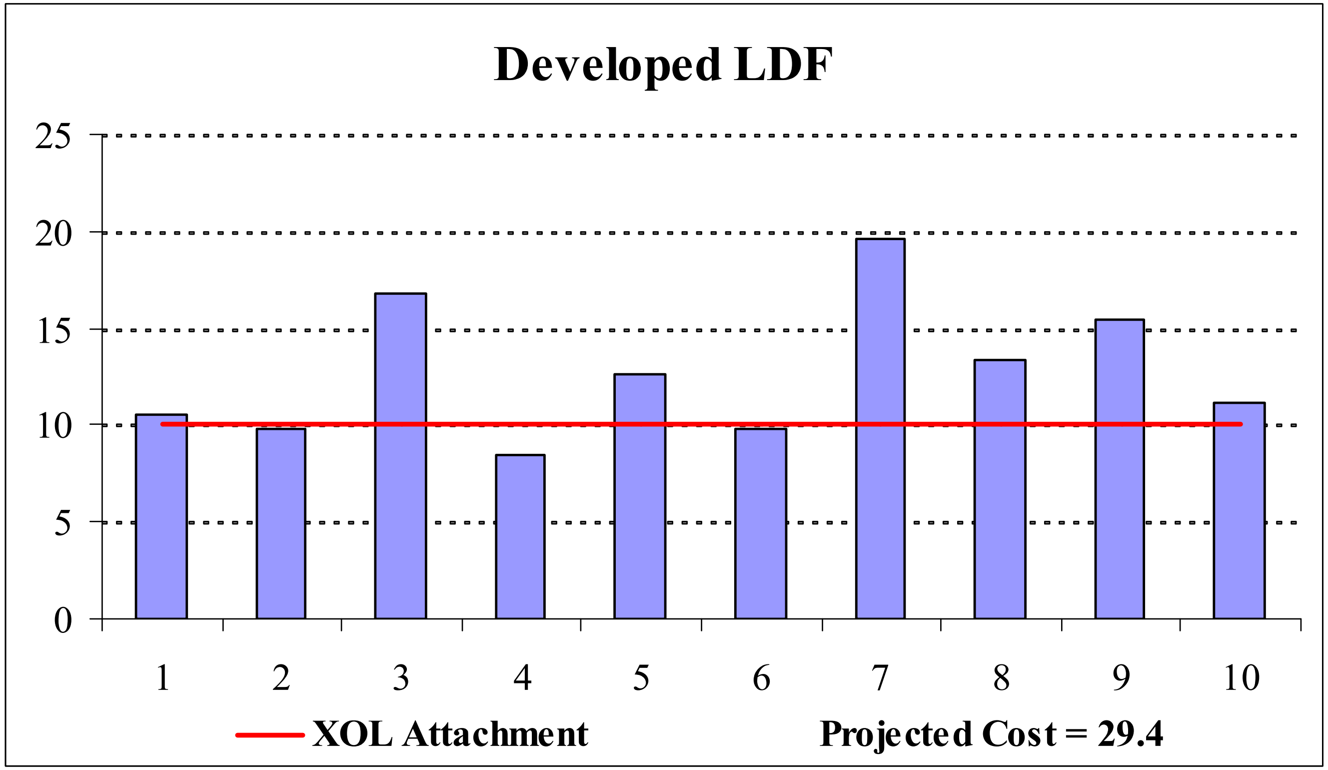
- Band age-dependent LDF
- Percentile age-dependent LDF
- Stochastic LDF approach
- And these methods applied to various types of triangles:
  - accident date, reported date, booked date
  - incurred, paid, settled only
  - quarterly, annual
  - development columns as fixed valuation date or fixed maturity
- The CRUX

IBNER projection problem



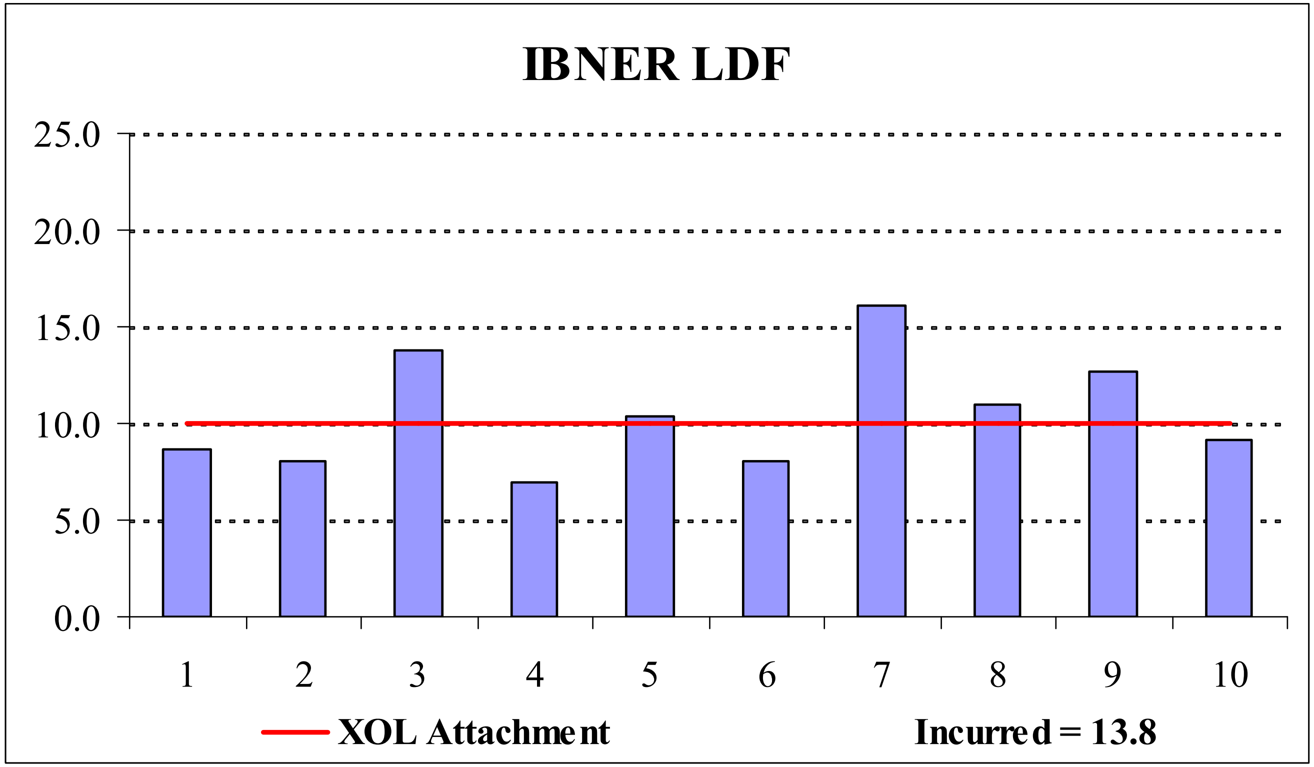
- Large claims listing for individual year
- Estimate the cost to the £10 xs £10 layer

Example 1 – Use Chain Ladder Ldf



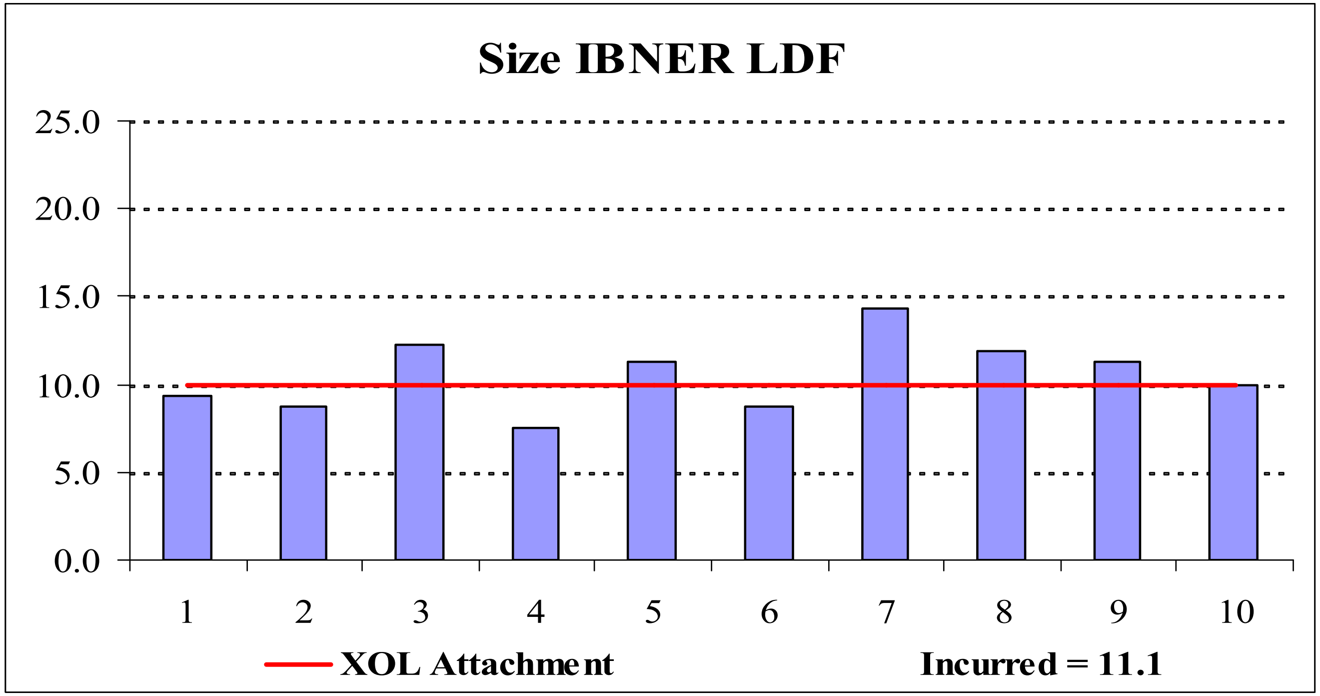
- Use chain ladder factor from incurred triangle, estimated as 1.4
- Problem - most of this development due to new claims, so cost to layer over estimated

Example 2 – Estimate IBNER factor



- Remove pure IBNR effect to estimate average IBNER factor as 1.15
- Might be good reasons why development should differ by size, e.g. sum insured or market precedents

Example 3 – Size dependent IBNER factor



- Estimate separate IBNER factor for below and excess £10m cause frequency to increase within layer
- Estimate factor as 1.25 for below £10m and 1.025 for above £10m
- Problems with fixed threshold and Ldf dependent on selection of threshold

Applications of IBNER

Many instances where it is useful to split numbers (IBNR) and movements in case reserves (IBNER):

- Pricing Excess of loss contracts
- Pricing for changes in deductibles or limits
- Stochastic claims severity modelling i.e. fitting statistical distributions to individual inflated projected claims
- Pricing aggregate deductibles and stop losses
- Projecting reinsurance recoveries for long tail classes
- Reserving applications
- Deriving claims inflation for a portfolio

Other applications of method

- Development factors can allow for other factors such as claim type, accident year, claims handler etc.
- Reserving for heterogeneous portfolios
- Reserving for claims made policies
- Win factors used in setting case reserves
- Identify claims to reserve separately
- Reinsurance projections allow for factors such as cedant and report delay

Data requirement

- Minimum data:
  - Transaction description (paid, reserve…)
  - Claims transactional amounts (all in one currency)
  - Transaction dates

- Additional useful data (if available and not exhaustive):
  - Claims reporting date
  - Claims date of loss
  - Indemnity type (BI, PD, injury type…)
  - Claims headers (indemnity, fee, recovery…)
  - Claim status (open, close, reopened)
  - Claims handler
  - Deductible applied if any

Data preparation

- Select most appropriate cohort (report or booked date) and frequency of development (quarterly or annually)
- If second booked date is available, when claim has actually been assessed, this could be best for initial comparison
- Produce appropriate development data from transactional database
- Run some data clean up algorithms e.g. remove "Phantom" movements

Data preparation
Claims inflation

- Claims need to inflated to consistent basis in order to compare claims across years
- Inflation to be applied “vertically”, i.e. inflate every development period with same factor
- Many different approaches, could be a flat rate or index
- Index can vary by accident year, claim header and claim size
- For pricing, inflation should be applied up to middle of exposure period to be priced
- For reserving, need to reverse out inflation after development to get back to reserve in monetary terms

Outline of method

- The development of a claim will be based on the development of other comparable claims more mature
- How claims are comparable is measured by calculating a distance
- This distance can be as complex as desired depending on the number of parameters considered and could include:
  - Time Weights
  - Paid to Incurred ratios
  - Claim type
  - First booked date and 2nd booked date
  - Reporting lag
  - Open / closed
  - Claims handler

Outline of method

Distance Calculation

- Calculate distance between projection and comparison claim at each comparable development period
- The age weights ω_a are applied to each development period in relation to “importance” of period on likely ultimate cost
- The total weighted distance is the sum over all development periods up to maturity of projection claim

$$D = \sum_a \frac{\left| Inc_{P,a} - Inc_{C,a} \right|}{Inc_{C,a}} \times \omega_a$$

- The distance is mapped to calculate a “likeliness” factor for each claim

Outline of method

Weight calculation example

- The weights determine the importance of the distance at each point in time
- We used formula $\omega_a = a^{0.75}$ to give more weight to more recent incurred positions
- Using a power of 0 assumes all development periods have same relevance in predicting the ultimate cost
- Using a power of 1 linearly increases the importance of development periods
- Could use the average payment pattern for ω_a

Outline of method

Likeliness calculation example (power = 0)

Power =	0								
Weigth	1.00	1.00	1.00	1.00	Cumulative Distance				
Distance	**at age 1**	**at age 2**	**at age 3**	**at age 4**	**at age 1**	**at age 2**	**at age 3**	**at age 4**	**Likeliness**
0%	0%	0%	0%	0%	0%	0%	0%	0%	**100%**
2%	2%	2%	2%	2%	2%	4%	6%	8%	**96%**
4%	4%	4%	4%	4%	4%	8%	12%	16%	**92%**
6%	6%	6%	6%	6%	6%	12%	18%	24%	**88%**
8%	8%	8%	8%	8%	8%	16%	24%	32%	**84%**
10%	10%	10%	10%	10%	10%	20%	30%	40%	**80%**
12%	12%	12%	12%	12%	12%	24%	36%	48%	**76%**
14%	14%	14%	14%	14%	14%	28%	42%	56%	**72%**
16%	16%	16%	16%	16%	16%	32%	48%	64%	**68%**
18%	18%	18%	18%	18%	18%	36%	54%	72%	**64%**
20%	20%	20%	20%	20%	20%	40%	60%	80%	**60%**

Age	1	2	3	4	5
Comparison	3,516	7,112	7,112	12,000	17,000
Projection	2,500	7,000	7,675		
distance	29%	2%	8%		
D =	38%				
L =	74%				

Outline of method

Likeness calculation example (power = 0.75)

Power =	0.75								
Weigth	1.00	1.68	2.28	2.83	Cumulative Distance				
Distance	**at age 1**	**at age 2**	**at age 3**	**at age 4**	**at age 1**	**at age 2**	**at age 3**	**at age 4**	**Likeliness**
0%	0%	0%	0%	0%	0%	0%	0%	0%	**100%**
2%	2%	3%	5%	6%	2%	5%	10%	16%	**96%**
4%	4%	7%	9%	11%	4%	11%	20%	31%	**92%**
6%	6%	10%	14%	17%	6%	16%	30%	47%	**88%**
8%	8%	13%	18%	23%	8%	21%	40%	62%	**84%**
10%	10%	17%	23%	28%	10%	27%	50%	78%	**80%**
12%	12%	20%	27%	34%	12%	32%	60%	93%	**76%**
14%	14%	24%	32%	40%	14%	38%	69%	109%	**72%**
16%	16%	27%	36%	45%	16%	43%	79%	125%	**68%**
18%	18%	30%	41%	51%	18%	48%	89%	140%	**64%**
20%	20%	34%	46%	57%	20%	54%	99%	156%	**60%**

Age	1	2	3	4	5
Comparison	3,516	7,112	7,112	12,000	17,000
Projection	2,500	7,000	7,675		
distance	29%	3%	18%		
D =	50%				
L =	80%				

Outline of method

Using additional factors

$$L = \frac{L_D - \sum_k L_k \beta_k}{1 + \sum_k \beta_k}$$

- L_D is the distance likeliness
- L_k is the likeliness for each of the k other factors
- β_k is the weight given to likeliness of factor k in relation to the distance likeliness
- Future research includes converting this formula into a multivariate model where interactions between distance and factors are taken into account

Comparison to Chain Ladder

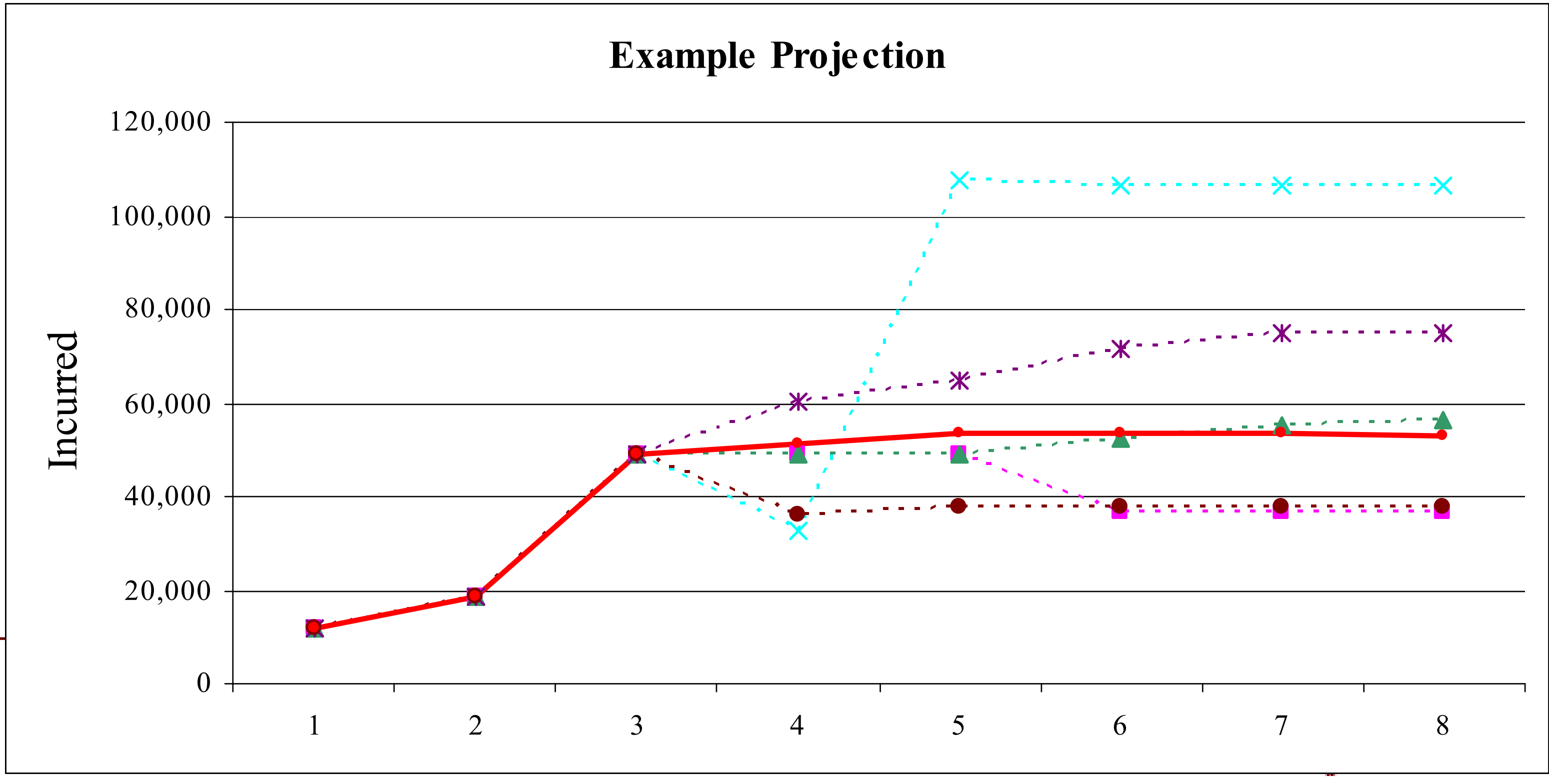
- Chain ladder is a special case where weights calculated purely on size of claim, irrespective of differences in claim size at each point in time.

Stochastic application

- Output from method is a matrix of possible development factors with associated likeliness for each projection claim
- Weights can be scaled to sum to 1
- This naturally gives an empirical distribution of possible outcomes with associated probabilities

Stochastic example

	1	2	3	4	5	6	7	8	Likeliness
Projection	**12,068**	**18,566**	**48,855**	51,444	53,257	53,424	53,411	52,987	
Claim #1	**12,068**	**18,566**	**48,855**	48,855	48,855	36,641	36,641	36,594	55%
Claim #2	**12,068**	**18,566**	**48,855**	48,855	48,855	52,654	55,164	56,298	14%
Claim #3	**12,068**	**18,566**	**48,855**	32,425	107,524	106,323	106,323	106,323	12%
Claim #4	**12,068**	**18,566**	**48,855**	60,331	64,856	71,342	74,909	74,909	9%
Claim #5	**12,068**	**18,566**	**48,855**	36,217	37,778	37,778	37,778	37,778	25%



Hurdles

- Large data sets: use a chain-ladder on smaller claims and develop individually the other claims using this method. Linkage between the 2 analysis needs to be done carefully.
- Inflation and development factor vicious circle
- Model calibration
- Processing lags at the beginning of the claims development: adjust weight given to development pattern
  - Paid to Incurred Ratios
  - Significant time in life cycle index
- Stochastic modelling: issue of large amount of data to store
- Impact of systemic changes to claims development pattern (regulatory or legal change, reserving philosophy...)

Examples – actual case study

- Below follows an actual case study on Bodily injury claims data, based on report year of claims
- Slight issue with nil values for claims in early development periods
- The method was applied to annual data in order to derive IBNER factors

Examples – individual claims

Age	Incurred	Ultimate	CDF
1	3,918	15,544	3.967
1	7,278	20,710	2.846
1	4,085	15,597	3.818
1	1,485	9,436	6.354
1	45,034	89,621	1.990
1	57,902	150,729	2.603
1	4,850	18,517	3.818
1	601,647	1,016,723	1.690
1	1,107	15,458	13.970
1	53,500	127,660	2.386

- *This shows an example of the IBNER projection for the most recent year of data (less than one year mature)*
- *The likeliness are calculated on only one quarter comparison*
- *There is a wide range of outcomes depending on the size of claim*

Examples – individual claims

Year	Incurred	Ultimate	CDF
3	4,778	6,030	1.262
3	14,564	16,672	1.145
3	118,732	134,362	1.132
3	40,432	47,555	1.176
3	1,947	3,028	1.555
3	7,347	11,320	1.541
3	641,327	637,525	0.994
3	20,439	23,003	1.125
3	11,665	13,333	1.143
3	1,271	2,142	1.685

- *This shows an example of the IBNER factor for three year maturity*
- *Again, this shows a wide spread of development factors by claim size*

Examples – cumulative factors by band

Claims band by age		1	2	3	4	5	6
0	10,000	4.722	2.021	1.369	1.238	1.064	1.076
10,000	20,000	2.249	1.527	1.181	1.097	1.041	1.020
20,000	50,000	2.109	1.533	1.235	1.193	1.089	1.071
50,000	100,000	3.402	1.534	1.235	1.167	1.156	1.235
100,000	250,000	2.497	1.584	1.138	1.144	1.191	1.116
250,000	500,000	1.688	1.464	1.206	1.215	1.116	0.980
500,000	1,000,000	1.482	1.194	1.071	1.197	1.526	0.983
1,000,000	10,000,000	1.191	1.075	1.132	1.564	1.052	1.007

- *This shows the best estimate cumulative development factor for each age*
- *It shows that smaller claims are subject to a higher average development factor than large claims*

Examples – development factors by band

Claims band by age		1	2	3	4	5	6
0	10,000	1.743	1.109	1.058	1.086	1.016	1.021
10,000	20,000	1.443	1.080	1.046	1.023	1.012	1.035
20,000	50,000	1.456	1.085	1.051	1.020	1.015	1.043
50,000	100,000	1.406	1.089	1.061	1.018	1.012	1.072
100,000	250,000	1.409	1.093	1.041	1.015	1.006	1.032
250,000	500,000	1.410	1.123	1.035	1.029	0.987	1.000
500,000	750,000	1.388	1.094	1.028	1.005	0.975	0.981
750,000	1,000,000	1.353	1.105	1.038	0.988	0.963	0.959

Coefficients of variation		1	2	3	4	5	6
0	10,000	59%	40%	22%	15%	17%	9%
10,000	20,000	51%	27%	23%	7%	9%	14%
20,000	50,000	43%	20%	21%	7%	12%	17%
50,000	100,000	40%	24%	17%	7%	5%	23%
100,000	250,000	34%	16%	12%	6%	5%	13%
250,000	500,000	27%	29%	14%	12%	2%	5%
500,000	750,000	18%	12%	14%	3%	3%	4%
750,000	1,000,000	18%	17%	5%	2%	2%	1%

Examples – chain ladder comparison

Individual	Inc_d1	Inc_d2	Inc_d3	Inc_d4	Inc_d5	Inc_d6	Inc_d7	Inc_d8	Inc_d9
2000	75,722	147,336	177,764	205,459	217,470	221,463	217,941	218,666	217,321
2001	69,085	139,931	161,904	171,185	177,288	180,515	192,047	191,965	180,114
2002	50,465	135,835	169,878	188,700	218,330	226,209	226,495	226,464	226,169
2003	37,558	88,218	103,227	116,781	128,758	129,539	129,422	129,446	134,164
2004	43,314	92,717	119,761	134,645	145,675	147,849	149,181	149,399	164,914
2005	53,623	125,178	145,484	166,512	175,857	178,628	180,699	180,494	205,929
2006	48,648	105,503	130,983	139,821	146,811	148,691	149,622	149,988	149,866
2007	53,065	105,305	123,144	132,783	137,591	139,402	149,493	150,031	149,975

Chain ladder	Inc_d1	Inc_d2	Inc_d3	Inc_d4	Inc_d5	Inc_d6	Inc_d7	Inc_d8	Inc_d9
2000	75,722	147,336	177,764	205,459	217,470	221,463	217,941	218,666	217,321
2001	69,085	139,931	161,904	171,185	177,288	180,515	192,047	191,965	190,784
2002	50,465	135,835	169,878	188,700	218,330	226,209	226,495	226,851	225,455
2003	37,558	88,218	103,227	116,781	128,758	129,539	131,250	131,456	130,647
2004	43,314	92,717	119,761	134,645	145,675	148,794	150,759	150,995	150,066
2005	53,623	125,178	145,484	166,512	180,935	184,808	187,249	187,543	186,389
2006	48,648	105,503	130,983	146,687	159,393	162,805	164,955	165,214	164,198
2007	53,065	105,305	127,292	142,552	154,901	158,216	160,306	160,558	159,570

Examples

- Pre-simulated case study

  Back-test of the method

Freebies provided…
(disclaimer: use it at your own risk)

- SAS code
- Excel spreadsheet
- VBA code in Excel
- Any further development, please do share it with the community

Contact Details

Claude Perret

CPERRET2@travelers.com

Hannes van Rensburg

Hannes.van.rensburg@watsonwyatt.com