M2 Reserving Claim Counts

Topics in Insurance, Risk, and Finance

Professor Benjamin Avanzi, Miranda Zhai

2025-08-08

Table of contents I

Table of contents II

General idea

Assume that we there exists $e(i)$ so that we can write

$$E[N(i,j)] \equiv e(i)\mu(j),$$

where (this is important), $\mu(j)$ does **not depend on** i.

- $e(i)$ will be called the **exposure** in period of occurrence i (for instance, "car-years").
- $\mu(j)$ may be interpreted as **relative claim frequency** in period j (per unit of exposure).
- The cumulative version

$$\frac{E[N(i,\cdot)]}{e(i)} = \mu(\cdot)$$

is the **relative frequency of claim occurence per period**.

In practice, this is not always achievable, that is,

$$\frac{E[N(i,j)]}{e(i)} = \mu(i,j),$$

with only *weak dependency* of $\mu(i,j)$ on i.

1. Exposure (2.1)

2. IBNR claims - Exposure based methods (2.2.2)

3. IBNR claims - Normalised methods (2.2.3) ✠

4. IBNR claims - Chain Ladder (2.2.4)

5. Claim frequency (2.3)

2 IBNR claims - Exposure based methods (2.2.2)

- Allowing for an i effect in μ
- Examples of $f(i)$
- Estimating $N(i, j)$
- Example
- Extrapolation Notes ✠

Allowing for an i effect in μ

Assume

$$\mu(i,j) = f(i)v(j),$$

where $f(i)$ is some known function (otherwise determined).

We have then

$$E[N(i,j)] = e(i)f(i)v(j).$$

We will discuss how to work with this.

2. IBNR claims - Exposure based methods (2.2.2)

- Allowing for an i effect in μ
- Examples of $f(i)$
- Estimating $N(i,j)$
- Example
- Extrapolation Notes ✠

Examples of $f(i)$

- A simple example could be:

$$f(i) = \alpha + \beta i,$$

  which leads to a linear adjustment across rows (periods of origin i) to "base frequency" of development $v(j)$ (for given development period j).
- It is unlikely to hold across all columns without modification, so one could extend this to

$$\mu(i,j) = f_j(i)v(j)$$

  so that α and β will (potentially) depend on j as well
- This could lead to a highly over-parametrised model.

- In practice, changes in "development speed" often occur in the first two development periods mostly, and in opposite direction (justifying a separate α and β).
- We could then use

$$
\begin{aligned}
\mu(i,0) &= f_0(i) = \alpha_0 + \beta_0 i, \\
\mu(i,1) &= f_1(i) = \alpha_1 + \beta_1 i, \\
\mu(i,j) &= v(j), \ j = 2, 3, \dots.
\end{aligned}
$$

2 IBNR claims - Exposure based methods (2.2.2)

- Allowing for an i effect in μ
- Examples of $f(i)$
- Estimating $N(i,j)$
- Example
- Extrapolation Notes ✠

Estimating $N(i,j)$

Now, assume

$$N(i,j) \sim \mathsf{Poisson}(e(i)f(i)v(j))$$

so that (assuming independence across periods of origin)

$$N(\cdot,j) \sim \mathsf{Poisson}\left(v(j)\sum_i e(i)f(i)\right).$$

Then we can show that

$$\hat{v}(j) = \frac{N(\cdot,j)}{\sum_i e(i)f(i)}$$

(over existing data) is a maximum likelihood, minimum variance, unbiased, consistent estimator (in short, a good one!).

In the end, our estimator for $E[N(i,j)]$ is

$$E[\widehat{N(i,j)}] = e(i)f(i)\hat{v}(j)$$

with IBNR

$$\text{IBNR}(i) = e(i)f(i) \left[\sum_{j=I-i+1}^{I} + \sum_{j=I+1}^{\infty} \right] \hat{v}(j).$$

- the first $\sum$ include $\hat{v}$'s estimated from available data - the second $\sum$ cannot be estimated from data, and will be extrapolated from the former set (e.g. linear regression or more typically, log-linear regression as exemplified later).

Example

Taylor (2000), Table 2.1 and Table 2.2 provide an example of such calculations, where $f(i) = 1$.

See the spreadsheet `Module2.xlsx`.

2   IBNR claims - Exposure based methods (2.2.2)

- Allowing for an i effect in μ
- Examples of $f(i)$
- Estimating $N(i, j)$
- Example
- Extrapolation Notes ✠

Extrapolation Notes ✠

- We fit a linear regression to $\ln \hat{v}(j)$ for $j = 4, \ldots, 10$ and tell Excel that $x \equiv j - 5$.
- Using SLOPE and INTERCEPT functions we get β_1 and β_0, so that

$$\tilde{v}(j) = \exp[\beta_0 + \beta_1(j-5)] = e^{\beta_0} \times \left(e^{\beta_1}\right)^{j-5} = 3.87 \times (0.54)^{j-5} \quad (2.14$$

- Note you can also get $\beta + \beta_1(j-5)$ with the Excel function FORECAST.
- Now, for the tail estimate ("11 and later"):

$$\sum_{j=11}^{\infty} \tilde{v}(j) = e^{\beta_0} \left(e^{\beta_1}\right)^6 \sum_{j=0}^{\infty} \left(e^{\beta_1}\right)^j = e^{\beta_0} \left(e^{\beta_1}\right)^6 \frac{1}{1 - e^{\beta_1}} \quad (2.15).$$

3 IBNR claims - Normalised methods (2.2.3) ✠
- Motivation
- The idea
- Generalisation
- Example

Motivation

- In the previous section, one hoped that claim notification rates μ (as proportions of exposure e) would be consistent (constant) across periods of origin i, or at least approximatively or predictively so.
- There may not always exist such an exposure.
- For instance, consider Public Liability of a manufacturer of toys:
  - Would time a good measure of exposure? or revenue?
  - This would unlikely to be satisfactory as the mix of business (which toys are sold and in what quantities) is likely to change all the time.

3 IBNR claims - Normalised methods (2.2.3) ✠

- Motivation
- The idea
- Generalisation
- Example

The idea

- We keep the idea of multiplicative structure, but μ would multiply something else than $e(i)$, say $\alpha(i)$:

$$E[N(i,j)] = \alpha(i)\mu(j).$$

- Then we automatically have

$$\frac{E[N(i,j)]}{E[N(i,0)]} = \frac{\mu(j)}{\mu(0)}$$

which is independent of i.

- Now Jensen's inequality teaches us that

$$\frac{E[N(i,j)]}{E[N(i,0)]} \neq E\left[\frac{N(i,j)}{N(i,0)}\right];$$

Nevertheless Taylor (2000) argues that the difference is small and that we can use the ratios $N(i,j)/N(i,0)$ to estimate the μ's.

- Taking a weighted average, we get

$$\widehat{v}(j) = \frac{\sum_i N(i,j)}{\sum_i N(i,0)}$$

  as an (approximately) unbiased estimator of $\mu(j)/\mu(0)$.

- Hence,

$$E[N(i,j)] \text{ is estimated by } N(i,0)\widehat{v}(j)$$

  and

$$\text{IBNR}(i) = N(i,0) \left[\sum_{j=I-i+1}^{I} + \sum_{j=I+1}^{\infty} \right] \widehat{v}(j).$$

  This is analogous to the previous "exposure" based formula.

- Note that even if we don't have $E[N(i,j)] = \alpha(i)\mu(j)$ but rather $E[N(i,j)] = \alpha_j(i)\mu(j)$, the triangle of normalised counts $N(i,j)/N(i,0)$ are useful to investigate, to look for trends.

3 IBNR claims - Normalised methods (2.2.3) ✠

- Motivation
- The idea
- Generalisation
- Example

Generalisation

- We will now "anchor" our prediction not just on $N(i,0)$, but on on a subset S of existing data.
- This generalises what we had before to

$$\frac{E[N(i,j)]}{E\left[\sum_{m \in S} N(i,m)\right]} = \frac{\mu(j)}{\sum_{m \in S} \mu(m)}.$$

where S is any subset of $\{0, 1, ..., I\}$, but typically the first m development periods for each period of origin i.
- It is assumed that this ratio is independent of i.
- This should work well if S is deemed to be a good indicator of the propensity to claim in a given period i.

- The prediction of $N(i,j)$ becomes

$$\left(\sum_{m \in S} N(i,m) \right) \hat{v}(j)$$

  with $\hat{v}(j)$ the estimate of the the ratio on the previous slide.

- Note that the most recent origin years may require special treatment (see example), as we do not have observations for all $m \in S$.

Example

- Table 2.3 and 2.4 revisit the previous example with $S = \{0, 1\}$.
- This can be useful if some disturbance to the claim notification process has caused a movement between development periods 0 and 1 but has had no other effect.
- See the spreadsheet `Module2.xlsx`.

Notes on treatment of the last row:

- The last row in the triangle requires special treament, as only $N(I,0)$ is available. We extrapolate

$$N(I,0) + N(I,1) \text{ as } N(I,0)\frac{\hat{v}(0) + \hat{v}(1)}{\hat{v}(0)}.$$

Now, because $\hat{v}(0) + \hat{v}(1) = 1000$ (by definition since we take them from the same type of average - "orange" here) we end up with

$$N(I,j) \text{ is estimated by } \left(N(I,0)\frac{\hat{v}(0) + \hat{v}(1)}{\hat{v}(0)}\right)\frac{\hat{v}(j)}{1000} = N(I,0)\frac{\hat{v}(j)}{\hat{v}(0)}.$$

1. Exposure (2.1)

2. IBNR claims - Exposure based methods (2.2.2)

3. IBNR claims - Normalised methods (2.2.3) ✠

4. IBNR claims - Chain Ladder (2.2.4)

5. Claim frequency (2.3)

4. IBNR claims - Chain Ladder (2.2.4)

- Introduction
- Derivation 1 - Heuristic approach
- Derivation 2 - Poisson approach
- Derivation 3 - Non-parametric approach
- The Chain Ladder setting

Introduction

- Probably the most famous (and basic) loss reserving technique.
- Quoting Taylor (2000): "The name is understood to refer to the chaining of a sequence of ratios (the age to age factors below) into a ladder of factors (the age to ultimate factors below) which enable one to climb (i.e. project) from experience recorded to date to its predicted ultimate value."
- We will see three derivations, which are informative about the meaning of the method.

4 IBNR claims - Chain Ladder (2.2.4)

- Introduction

- Derivation 1 - Heuristic approach

- Derivation 2 - Poisson approach

- Derivation 3 - Non-parametric approach

- The Chain Ladder setting

Derivation 1 - Heuristic approach

- We start from

$$E[N(i,j)] = \alpha(i)\mu(j)$$

as before, but consider

$$\frac{E[A(i,j+1)]}{E[A(i,j)]} = \frac{\sum_{m=0}^{j+1} \mu(m)}{\sum_{m=0}^{j} \mu(m)} \equiv v(j)$$

instead, where

$$A(i,j) = \sum_{m=0}^{j} N(i,m)$$

are the **cumulative** claim counts for period of origin i, as of development period j.

- This leads to an estimator

$$\hat{v}(j) = \frac{\sum_i A(i, j+1)}{\sum_i A(i, j)},$$

where the summations run over those values of i where we have both A's available.

- These are called **chain ladder factors**, **age to age factors**, or **link ratios**.

- Future values of $A(i,j)$ are predicted by

$$\widehat{A}(i,j) = A(i, I-i)\widehat{v}(I-i)\widehat{v}(I-i+1)\cdots\widehat{v}(j-1),$$

  where $A(i, I-i)$ is the latest observation for period i (on the diagonal).

- In particular, ultimate counts are

$$\widehat{A}(i,\infty) = A(i, I-i)\widehat{\pi}(I-i),$$

  where

$$\widehat{\pi}(I-i) = \prod_{m=0}^{\infty} \widehat{v}(I-i+m),$$

  called **age to ultimate factors**.

Example

- Table 2.5, 2.6 and 2.7 revisit the previous example using the chain ladder example.
- See the spreadsheet `Module2.xlsx`.
- ✠ Notes on the extrapolation: assuming that it is the excess over 1 that decays exponentially, we fit $\ln\left[\hat{v}(j-6)-1\right]$ for $j=6,\ldots,9$. The extrapolation to infinity is then added to 1 to get an age to ultimate factor. This is an ad hoc approach, but it's the one's taken in the book.

4 IBNR claims - Chain Ladder (2.2.4)

- Introduction
- Derivation 1 - Heuristic approach
- Derivation 2 - Poisson approach
- Derivation 3 - Non-parametric approach
- The Chain Ladder setting

Derivation 2 - Poisson approach

- Assume that

$$N(i,j) \sim \text{Poisson}(\alpha(i)\mu(j)),$$

  and that they are mutually independent.

- Because

$$\alpha(i)\mu(j) = [k \cdot \alpha(i)] \left[\frac{1}{k} \cdot \mu(j) \right],$$

  α and μ are not uniquely defined unless we introduce an extra
  condition, which we conveniently choose to be

$$\sum_{j=0}^{J} \mu(j) = 1,$$

  (J being the largest value of j for which any $N(i,j)$ is observed) so
  that $\mu(j)$ can be interpreted as the proportion of claims being notified
  in development year j.

- It can be shown (see Taylor (2000)) that the estimators seen
  before are the MLE estimators in this context.

Because

$$\frac{1}{\pi(j)} = \frac{A(i,j)}{A(i,\infty)}$$

is the proportion of claims notified **by** period j,

$$\frac{1}{\pi(j)} - \frac{1}{\pi(j-1)} \equiv \mu(j)$$

is the proportion of claims notified **in** development period j (assume $\pi(-1) = 1$).

- Example: The age to ultimate factors and corresponding proportions $\mu(j)$ are calculated under Table 2.5.

4. IBNR claims - Chain Ladder (2.2.4)

- Introduction
- Derivation 1 - Heuristic approach
- Derivation 2 - Poisson approach
- Derivation 3 - Non-parametric approach
- The Chain Ladder setting

Derivation 3 - Non-parametric approach

- The Poisson derivation is a parametric one, and starts with rather strong assumptions.
- What assumptions do we need to make *at a minimum*, in order to retrieve the chain ladder estimators according to some rigorous calculations? (in particular, do we need the Poisson assumption?)
- It turns out that if we assume

$$E[N(i, j+1)|A(i, j)] = x(j)A(i, j)$$

for some $x(j)$, and that

$$Var(N(i, j+1)|A(i, j)) = \sigma^2(j)A(i, j),$$

for some constant $\sigma^2(j) > 0$, then minimising *weighted least squares* between this model and the data yields $x(j)$ such that we get the chain ladder estimators again.

- This does not require any distributional assumption!

4. IBNR claims - Chain Ladder (2.2.4)

- Introduction
- Derivation 1 - Heuristic approach
- Derivation 2 - Poisson approach
- Derivation 3 - Non-parametric approach
- The Chain Ladder setting

The Chain Ladder setting

- Chain ladder won't always work.
- In particular, it requires a **fixed notification pattern** - the $\mu(j)$ can't vary across origin periods.
- One can think of many reasons why notifications may come quicker (e.g. IT systems) or slower (COVID-19), either permanently or temporarily (same examples, respectively).
- As a solution, one can either "tweak" the chain ladder procedure, or proceed with an entirely different approach, which allows for such changes. This is (mostly) out of scope of this introductory treatment of reserving techniques.

1 Exposure (2.1)

2 IBNR claims - Exposure based methods (2.2.2)

3 IBNR claims - Normalised methods (2.2.3) ⊞

4 IBNR claims - Chain Ladder (2.2.4)

5 Claim frequency (2.3)

5. Claim frequency (2.3)
   - Claim frequency
   - Notified claim frequency
   - Example

Claim frequency

- If an exposure exists, it is possible to estimate a **claim frequency**, that is, the number of claims per unit of exposure.
- This is mathematically defined as

$$f(i) = \frac{A(i, \infty)}{e(i)}$$

  for period of origin i, and is estimated by

$$\hat{f}(i) = \frac{\hat{A}(i, \infty)}{e(i)}.$$

- Examination of such claim frequencies may help with model validation, and may help identify anomalies.

5. Claim frequency (2.3)
   - Claim frequency
   - Notified claim frequency
   - Example

Notified claim frequency

- This is defined as

$$\frac{A(i, I - i)}{e(i)}$$

  for period of origin i.
- This corresponds to the frequency of claims that have already been notified (obviously, a decreasing number in i).
- Comparing this with $\hat{f}(i)$ gives an idea of the proportion of ultimate claims that have already been notified.

5 Claim frequency (2.3)

- Claim frequency
- Notified claim frequency
- Example

Example

- Table 2.8 (see `Module2.xlsx`) for calculations displays estimated claim frequencies from the three methods presented in the previous section.
    - 1991 and 1992 seem to reverse a strong downward trend from 1984 to 1990.
    - is this due to a method issue (introduced when "completing the rectangle"), or due to a real effect?
- Table 2.9 (see `Module2.xlsx`) for calculations and associated Figure 2.2 compares the notified claims frequency with the estimated claims frequencies of Table 2.8.
    - It appears that the reversal of trend is already present in the actual data, and is not a result of the modelling.
    - This may reassure the actuary that the reversal is real. She would need to gather more data from the claims department to ascertain whether that is a reasonable feature of the model.

References

Selected references:

Taylor, Greg. 2000. *Loss Reserving: An Actuarial Perspective*. Huebner International Series on Risk, Insurance and Economic Security. Kluwer Academic Publishers.