An approach to the analysis of claims experience in motor liability excess of loss reinsurance

PAPER REPORT

INTRODUCTION
This paper mainly deals with the problem of extrapolating the number of claims occurred in the past and the current year but would be reported in the future. The approach to solve this problem is Maximum Likelihood Principle, which is very familiar to us. In this report, the main assumptions and principle of the approach will be reviewed, then an example will be provided to see how the method works in practice.

It is sincerely welcome of readers of this report to point out my mistakes if there is any and make any comment on this report.
PART 1

Revision of the assumptions and method used in this paper

We know there is always a time-lag between the occurrence of a claim and the time it is reported to the insurer. Then there arises a problem of estimating the amount of claims occurred and have not been reported but will be reported in the future.

Here we discuss the problem from the prospect of the reinsurer. Since the accumulation of the size of the claim will have deeper effect on the reinsurer than the insurer because the reinsurer pays the claim if the size exceeds an agreed limit.

In the model we consider the claim occurred and reported in the past $k$ years (which we have already known) as well as the claims occurred in the past $k$ years but will be reported in the future $k$ years.

Then we have the claim table

Table 1

<table>
<thead>
<tr>
<th>year occurred</th>
<th>year reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>$n_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$k$</td>
<td>$n_{k1}$</td>
</tr>
</tbody>
</table>

Here the entry $n_{ij}$ refers to the number of claims occurred in the year $i$ but reported in the year $j$. In this case, we require $i + j \leq k + 1$, since we are standing at the time spot $T=k$.

We make an important assumption here that the claim counting variables have a Poisson distribution, which allows the Maximum Likelihood to be deduced.

Then we introduce two sets of parameters into the model, namely

$n \quad r_i, i = 1, 2, \ldots, k$, the probability of a claim being reported $i$ years after its
occurrence. Note that $\sum_{i=1}^{k} r_i = 1$, which means all claims will be eventually reported.

$\lambda_i, i = 1, 2, L k$, the distribution parameter of the Poisson variable.

Another important issue here is that in this paper we are interested in the number of excess claims: $n_i$. In fact, what makes sense is the total excess claim amount rather than the claim number. But in this case, we could use the expected numbers of excess claims instead of the total claim amount, since the mean amounts of excess claims for fixed excess will always show a STABLE pattern, which is strictly true in the Pareto case. This can be proved as follows:

We donate the amount of a single claim to be a random variable $X$, which has a Pareto distribution, i.e. $X \sim Pa(\alpha, \lambda)$, the density function of $X$ is given by

$$f_X(x) = \frac{\alpha \lambda^\alpha}{(x + \lambda)^{\alpha + 1}}, \quad x > 0, \alpha > 0, \lambda > 0$$

For a reinsurer, he will make payment to a claim if the claim amount exceeds a certain point, for instance $M$, then we can donate the amount of a claim paid by the reinsurer to be another random variable $Z$, where

$$Z = \begin{cases} 
0 & (X < M) \\
X - M & (X \geq M) 
\end{cases}$$

Hence the expected value of $Z$ is

$$E = E(Z) = \int_{M}^{\infty} (x - M) \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha + 1}} \, dx$$

$$= \alpha \lambda^\alpha \int_{M}^{\infty} \frac{x - M}{(\lambda + x)^{\alpha + 1}} \, dx$$

$$= \alpha \lambda^\alpha \left( \int_{M}^{\infty} \frac{x + \lambda}{(\lambda + x)^{\alpha + 1}} \, dx - \int_{M}^{\infty} \frac{\lambda + M}{(\lambda + x)^{\alpha + 1}} \, dx \right)$$

$$= \alpha \lambda^\alpha \left( \frac{1}{(\alpha - 1)(\lambda + M)^{\alpha - 1}} - \frac{\lambda + M}{\alpha (\lambda + M)^{\alpha}} \right)$$

$$= \frac{\lambda^\alpha}{(\lambda + M)^{\alpha-1}(\alpha - 1)}$$
In fact what we want to obtain is the conditional expected amount of a claim paid by the reinsurer. For this purpose, we introduce a new random variable \( W = Z \mid X > M \)

\[
E(W) = \frac{E(Z)}{P(X > M)} = \frac{\lambda^a}{(\lambda + M)^{\alpha-1}(\alpha - 1)} \times \frac{1}{1 - F(M)}
\]

\[
= \frac{\lambda^a}{(\lambda + M)^{\alpha-1}(\alpha - 1)} \times \left( \frac{\lambda + M}{\lambda} \right)^\alpha
\]

\[
= \frac{\lambda + M}{\alpha - 1} \quad \text{the mean value of Pareto}(\alpha, \lambda + M)
\]

It can be shown that if there are \( s \) homogeneous claims, then the random variable is \( sX \), then \( sX \sim \text{Pareto} (\alpha, s\lambda) \). We can see the sum of excess points of the \( s \) claims is \( sM \). The new expected value of \( W \) is

\[
E(W') = \frac{s\lambda + sM}{\alpha - 1} = sE(W) = E(sW).
\]

Hence we can say for a fixed excess point \( M \), the mean amounts of excess claims show a stable pattern.

**In this paper, we can see our aim is to determine the 2k variables, and further estimate another k lambdas.**

Since we have made the Poisson assumption, let \( N_{ij} \) be the entry variable of the number of the claim occurred in year \( i \) and reported in year \( j \), then we have

\[
P(N_{ij} = n_{ij}) = \frac{(r_j \lambda_{i+j-1})^{n_{ij}} \exp(-r_j \lambda_{i+j-1})}{n_{ij}!}.
\]

Then there must be

\[
P(n_{11}, L n_{1k_1}, L n_{k_1}, L r_1, L r_k, \lambda_1, L \lambda_k)
\]

\[
= \prod_{i=1}^{k} \prod_{j=1}^{k-i+1} \frac{(r_j \lambda_{i+j-1})^{n_{ij}} \exp(-r_j \lambda_{i+j-1})}{n_{ij}!}
\]

subject to \( \sum_{i=1}^{k} r_i = 1 \)

Hence the log-likelihood function

\[
L = -\sum_{i=1}^{k} \sum_{j=1}^{k-i+1} r_j \lambda_{i+j-1} + \sum_{i=1}^{k} \sum_{j=1}^{k-i+1} n_{ij} \ln r_j \lambda_{i+j-1} - \sum_{i=1}^{k} \sum_{j=1}^{k-i+1} n_{ij}!
\]
We need to find the maximum of $L$, so we use Lagrange function, which is

$$\Gamma = L - s\left(\sum_{i=1}^{k} r_i - 1\right)$$

to ensure the values we will obtain satisfy $\sum_{i=1}^{k} r_i = 1$.

The maximum value of $L$ is found where

$$\frac{\partial \Gamma}{\partial r_i} = 0 \quad \text{and} \quad \frac{\partial \Gamma}{\partial \lambda_i} = 0 \quad \text{for} \quad i = 1, 2, \ldots, k$$

which is equivalent to the equations in the paper:

$$\frac{\partial L}{\partial r_i} = -\sum_{t=i}^{k} \lambda_t + \sum_{t=1}^{k-i+1} \frac{n_{ti}}{r_i} + s = 0$$

and

$$\frac{\partial L}{\partial \lambda_i} = -\sum_{t=1}^{i} r_t + \sum_{t=1}^{i} \frac{n_{i-t+1,i}}{\lambda_i} + s = 0$$

It is easy to prove that $s = 0$.

We define the following

$$v_j = \sum_{i=1}^{k-i+1} n_{ij}, \quad d_j = \sum_{i=1}^{j} n_{i,j-i+1}$$

We can have the recursive equations

$$d_k = \lambda_k \quad v_k = r_k \lambda_k$$

$$d_{k-1} = \lambda_{k-1} (1 - r_k) \quad v_{k-1} = r_{k-1} (\lambda_{k-1} + \lambda_k)$$

$$L \ L \ L \ L \ L \ L \ L \ L \ L$$

$$d_1 = \lambda_1 (1 - \sum_{j=2}^{k} r_j) \quad v_1 = r_1 \sum_{j=1}^{k} \lambda_j$$

From these equations we can obtain the parameters of $\lambda$s and $r$s.

Still we have further $k$ parameters to determine:

$$\lambda_i, \quad i = k + 1, k + 2, L \ L \ 2k$$

These can be determined using exponential estimation function.

The following example shows how the approach works.
PART II

An example

Notes:
1. The data used in this example is given with no other reference, namely solely by myself. In this case, the aim of the example is just to show how this approach works.
2. The paper suggests that the approach will perform a better result if \( k \) is around 10, for simplicity, we just examine the case which \( k = 6 \).

Suppose we have the following data, expressed using the previous claim table:

<table>
<thead>
<tr>
<th>year occurred</th>
<th>year reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
</tr>
</tbody>
</table>

Then we have the values of \( v_i \) and \( d_i \):

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_i )</td>
<td>276</td>
<td>159</td>
<td>91</td>
<td>29</td>
<td>9</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_i )</td>
<td>34</td>
<td>66</td>
<td>88</td>
<td>115</td>
<td>124</td>
<td>140</td>
</tr>
</tbody>
</table>

Using the recursive equations, we have:

\[
\lambda_6 = d_6 = 140
\]

\[
r_6 = \frac{v_6}{\lambda_6} = \frac{3}{140} = 0.02142857
\]
Perform the calculation one by one,

\[
\begin{align*}
\lambda_3 &= \frac{d_3}{1 - r_6} = \frac{124}{1 - 0.02142857} = 126.71533 \\
r_3 &= \frac{v_3}{\lambda_6 + \lambda_5} = \frac{9}{266.71533} = 0.03374384 \\
\lambda_4 &= \frac{d_4}{1 - r_5 - r_6} = \frac{115}{0.9448276} = 121.715328 \\
r_4 &= \frac{v_4}{\lambda_6 + \lambda_3 + \lambda_4} = \frac{29}{388.43066} = 0.0746594
\end{align*}
\]

Then we have the 12 values,

<table>
<thead>
<tr>
<th>(i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_i)</td>
<td>82.33215</td>
<td>96.45073</td>
<td>101.12987</td>
<td>121.71532</td>
<td>126.71533</td>
<td>140</td>
</tr>
<tr>
<td>(r_i)</td>
<td>0.4129614</td>
<td>0.2713528</td>
<td>0.185881</td>
<td>0.0746594</td>
<td>0.03374384</td>
<td>0.0214286</td>
</tr>
</tbody>
</table>

In the paper, it is suggested that an exponential extrapolation will be probable to give the other \(k\) lambdas.

We may assume that the relation between \(i\) and \(\lambda_i\) is

\[
\lambda_i = e^{a+bi} \quad \text{for } i=1,2,3\L
\]

We donate \(y_i = \ln \lambda_i, \quad x_i = i\). Then we will have

<table>
<thead>
<tr>
<th>(i=x_i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_i)</td>
<td>82.33215</td>
<td>96.45073</td>
<td>101.12987</td>
<td>121.71532</td>
<td>126.71533</td>
<td>140</td>
</tr>
<tr>
<td>(\ln \lambda_i = y_i)</td>
<td>4.41076</td>
<td>4.56903</td>
<td>4.61641</td>
<td>4.80168</td>
<td>4.84194</td>
<td>4.94164</td>
</tr>
</tbody>
</table>
We have
\[ \sum x = 21, \quad \sum x^2 = 91, \quad \sum y = 28.18146, \quad \sum y^2 = 132.5624, \quad \sum xy = 100.46431 \]
\[ S_{xx} = 91 - \frac{21^2}{6} = 17.5, \]
\[ S_{yy} = 132.5624 - \frac{28.18146^2}{6} = 0.19411871 \]
\[ S_{xy} = 100.46431 - \frac{21 \times 28.18416}{6} = 1.8292 \]
\[ R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} = 0.9849562, \text{ which indicates strong linear relationship.} \]

Then the parameters, \( a \) and \( b \) are given by:
\[ \hat{b} = \frac{S_{xy}}{S_{xx}} = 0.10452571, \]
\[ \hat{a} = \bar{y} - \hat{b}\bar{x} = 4.33107 \]
then the linear equation \( y = 4.33107 + 0.10452571x \)

namely \( \hat{\lambda}_i = \exp(4.33107 + 0.10452571i) \)

We are ready to give the rest 6 lambdas

<table>
<thead>
<tr>
<th>( i )</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_i )</td>
<td>158.0</td>
<td>175.4</td>
<td>194.8</td>
<td>216.2</td>
<td>240.1</td>
<td>266.5</td>
</tr>
</tbody>
</table>

Everything is ready, now we can estimate the number of claims we want

See table 3 and table 4.
### Table 3

<table>
<thead>
<tr>
<th>year occurred</th>
<th>year reported</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>30</td>
<td>21</td>
<td>16</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>25</td>
<td>22</td>
<td>4</td>
<td>2</td>
<td>r_6 \lambda_7</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>27</td>
<td>20</td>
<td>9</td>
<td>r_5 \lambda_7</td>
<td>r_6 \lambda_8</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>39</td>
<td>28</td>
<td>r_4 \lambda_7</td>
<td>r_5 \lambda_8</td>
<td>r_6 \lambda_9</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
<td>38</td>
<td>r_4 \lambda_7</td>
<td>r_5 \lambda_8</td>
<td>r_6 \lambda_9</td>
<td>r_6 \lambda_{10}</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>r_5 \lambda_7</td>
<td>r_6 \lambda_8</td>
<td>r_5 \lambda_9</td>
<td>r_6 \lambda_{10}</td>
<td>r_6 \lambda_{11}</td>
</tr>
<tr>
<td>7</td>
<td>r_6 \lambda_7</td>
<td>r_5 \lambda_8</td>
<td>r_5 \lambda_9</td>
<td>r_6 \lambda_{10}</td>
<td>r_6 \lambda_{11}</td>
<td>r_6 \lambda_{12}</td>
</tr>
</tbody>
</table>

Substituting the shaded area with the appropriate values:

### TABLE 4

<table>
<thead>
<tr>
<th>year occurred</th>
<th>year reported</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>30</td>
<td>21</td>
<td>16</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>25</td>
<td>22</td>
<td>4</td>
<td>2</td>
<td>3.39</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>27</td>
<td>20</td>
<td>9</td>
<td>5.33</td>
<td>3.75</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>39</td>
<td>28</td>
<td>11.80</td>
<td>5.92</td>
<td>4.17</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
<td>38</td>
<td>29.37</td>
<td>13.10</td>
<td>6.57</td>
<td>4.63</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>42.87</td>
<td>32.60</td>
<td>14.54</td>
<td>7.30</td>
<td>5.14</td>
</tr>
<tr>
<td>7</td>
<td>65.25</td>
<td>47.60</td>
<td>36.21</td>
<td>16.14</td>
<td>8.10</td>
<td>5.71</td>
</tr>
</tbody>
</table>
Final remark and papers for extensive reading

This paper gave an approach of estimating the number of claims occurred in the past but will be reported in the future. In insurance, we call this IBNR (incurred but not reported). Here are some papers relevant to IBNR. They are for extensive reading if you have any interest in the field.

- The Actuary And IBNR
- A Bayesian Credibility Formula For IBNR Counts
- A Simple Parametric Model For Rating Automobile Insurance Or Estimating IBNR Claims Reserves
- IBNR Reserve Under a Loglinear Location-Scale Regression Model
- A Probabilistic Model For IBNR Claims