Claim reserving prediction with Bornhuetter-Ferguson method

I G C D Saputra, S Nurrohmah and S F Sari
Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: snurrohmah@sci.ui.ac.id

Abstract. In the long-tail business insurance class, claim settlement takes at least one year, since the claim was filed. This is because insurance companies must investigate in advance of the incident. The problem in the long-tail business can be attributed to the large or unresolved number of claims, called outstanding claims. The prediction of outstanding claims has an important role considering insurance companies are required to allocate sufficient reserves for future payment of claims. Generally, the problem of prediction of outstanding claims utilizes run-off triangle. Run-off triangle is an instrument, which is used to analyze the relationship between claims payments that have been made over the past few years. One of the prediction methods that commonly used is the Bornhuetter-Ferguson method. This method will be applied to the paid loss and ALAE Commercial Auto/Truck Liability/Medical insurance data, which is obtained from Highline Annual Statement Schedule P Data from all industry in the United States for the period 1999–2008, where the claim payment is stated in USD (000's).

Keywords: Bornhuetter-Ferguson method, long-tail business insurance, outstanding claims, run-off triangle

1. Introduction

The problem in the long-tail business can be attributed to the large or unresolved number of claims, called outstanding claims. The prediction of outstanding claims has an important role considering insurance companies are required to allocate sufficient reserves for future payment of claims. If the prediction of outstanding claims is poor, the insurance companies may be collapse. Generally, the problem of prediction of outstanding claims utilizes run-off triangle. Run-off triangle is an instrument, which is used to analyze the relationship between claims payments that have been made over the past few years.

In the run-off triangle, the data usually can be either amount of claims or number of claims, where both of them are presented in the form of incremental losses or cumulative losses. The large or number of claims between development years is assumed to follow a certain pattern in each accident year. These patterns are so called development patterns. The prediction methods differ from one to another depending on the patterns formed. One of the prediction methods that commonly used is the Bornhuetter-Ferguson method. This method will be applied to the paid loss and ALAE Commercial Auto/Truck Liability/Medical insurance data, which is obtained from Highline Annual Statement Schedule P Data from all industry in the United States for the period 1999–2008, where the claim payment is stated in USD (000's) [1].
2. Run-off triangle
Suppose there is a portfolio of risks and it is assumed that each claim of portfolio is settled either in the accident year or in the following n development years. The portfolio may be modelled either by incremental losses or cumulative losses.

Let there is a family of random variables \(X_{i,k}\) and we interpret the random variable \(X_{i,k}\) as the loss of accident year \(i\) which is settled with a delay of \(k\) years and hence in development year \(i+k\) and in calendar year \(i+k\) and in calendar year \(i+k\). We refer to \(X_{i,k}\) as the incremental loss of accident year \(i\) and development year \(k\) [2]. The incremental losses are represented by the following run-off triangle in table 1. As can be seen in Table 1, the values of \(X_{i,k}\) are observable for calendar years \(i+k\) \(\leq n\) (development triangle), whereas the values of \(X_{i,k}\) are non-observable for calendar years \(i+k > n\) (future triangle) [3].

The development triangle can be denoted as,

\[
D_n^D = \{X_{i,k} ; i + k \leq n, k \in \{0, ..., n\}\} \tag{1}
\]

and for the future triangle can be denoted as,

\[
D_n^F = \{X_{i,k} ; i + k > n, k \in \{0, ..., n\}\} \tag{2}
\]

Let there is a family of random variables \(Y_{i,k}\) and we interpret the random variable \(Y_{i,k}\) as the loss of accident year \(i\) which is settled with a delay of at most \(k\) years and hence not later than in development year \(k\). We refer to \(Y_{i,k}\) as the cumulative loss of accident year \(i\) and development year \(k\) [2]. The cumulative losses are represented by the following run-off triangle in table 2. As can be seen in table 2, the values of \(Y_{i,k}\) are observable for calendar years \(i+k\) \(\leq n\) (development triangle), whereas the values of \(Y_{i,k}\) are non-observable for calendar years \(i+k > n\) (future triangle) [3].

The development triangle can be denoted as,

\[
D_n^{\text{CD}} = \{Y_{i,k} ; i + k \leq n, k \in \{0, ..., n\}\} \tag{3}
\]

and for the future triangle can be denoted as,

\[
D_n^{\text{CF}} = \{Y_{i,k} ; i + k > n, k \in \{0, ..., n\}\} \tag{4}
\]

Table 1. Run-off triangle in the form of incremental losses.

| Accident year | Development year |
|---------------|-----------------|
| \(0\)         | \(X_{0,0}\) \(X_{0,1}\) ... \(X_{0,k}\) ... \(X_{0,n-i}\) ... \(X_{0,n-1}\) \(X_{0,n}\) |
| \(1\)         | \(X_{1,0}\) \(X_{1,1}\) ... \(X_{1,k}\) ... \(X_{1,n-i}\) ... \(X_{1,n-1}\) \(X_{1,n}\) |
| \(i\)         | \(X_{i,0}\) \(X_{i,1}\) ... \(X_{i,k}\) ... \(X_{i,n-i}\) ... \(X_{i,n-1}\) \(X_{i,n}\) |
| \(n-k\)       | \(X_{n-k,0}\) \(X_{n-k,1}\) ... \(X_{n-k,k}\) ... \(X_{n-k,n-i}\) ... \(X_{n-k,n-1}\) \(X_{n-k,n}\) |
| \(n-1\)       | \(X_{n-1,0}\) \(X_{n-1,1}\) ... \(X_{n-1,k}\) ... \(X_{n-1,n-i}\) ... \(X_{n-1,n-1}\) \(X_{n-1,n}\) |
| \(n\)         | \(X_{n,0}\) \(X_{n,1}\) ... \(X_{n,k}\) ... \(X_{n,n-i}\) ... \(X_{n,n-1}\) \(X_{n,n}\) |
Table 2. Run-off triangle in the form of cumulative losses.

| Accident year | Development year |
|---------------|------------------|
| 0             | $Y_{0,0}$ $Y_{0,1}$ ... $Y_{0,k}$ ... $Y_{0,n-i}$ $Y_{0,n-1}$ $Y_{0,n}$ |
| 1             | $Y_{1,0}$ $Y_{1,1}$ ... $Y_{1,k}$ ... $Y_{1,n-i}$ $Y_{1,n-1}$ $Y_{1,n}$ |
| ...           | ...              |
| n-k           | $Y_{n-k,0}$ $Y_{n-k,1}$ ... $Y_{n-k,k}$ ... $Y_{n-k,n-i}$ $Y_{n-k,n-1}$ $Y_{n-k,n}$ |
| ...           | ...              |
| n-I           | $Y_{n-1,0}$ $Y_{n-1,1}$ ... $Y_{n-1,k}$ ... $Y_{n-1,n-i}$ $Y_{n-1,n-1}$ $Y_{n-1,n}$ |
| n             | $Y_{n,0}$ $Y_{n,1}$ ... $Y_{n,k}$ ... $Y_{n,n-i}$ $Y_{n,n-1}$ $Y_{n,n}$ |

$Y_{i,n-i}$ ~ cumulative losses of the present calendar year $n$ or current (cumulative) losses. $Y_{i,n}$ ~ ultimate (cumulative) losses.

The problem is to predict the non-observable cumulative losses (outstanding claims) and the claim reserve. The claim reserve for accident year $i$ is defined by,

$$R_i = \sum_{k=n-i+1}^{n} X_{i,k} = Y_{i,n} - Y_{i,n-i}, \quad i \in \{1, \ldots, n\}. \tag{5}$$

Therefore, the total claim reserve is defined by,

$$R = \sum_{i=1}^{n} R_i. \tag{6}$$

3. Bornhuetter-Ferguson method

The Bornhuetter-Ferguson method is one of the most popular claim reserving techniques with distribution-free. The distribution-free Bornhuetter-Ferguson method is based on known cumulative losses $Y_{i,k}$ in development triangle. The focus in this method are to predict outstanding claims $Y_{i,k}$ and the claim reserve for accident year $i$. The Bornhuetter-Ferguson method is based on the following assumptions.

**Assumption 1.** \([4]\) Cumulative losses $Y_{i,k}$ of different accident year $i$ are independent random variables. That is,

$$\{Y_{i,0}, \ldots, Y_{i,n}\}, \{Y_{j,0}, \ldots, Y_{j,n}\} \text{ for } i \neq j \text{ are independent.} \tag{7}$$

There exist parameters $\mu_0, \mu_1, \ldots, \mu_n > 0$ and development pattern for cumulative quotas $\gamma_0, \gamma_1, \ldots, \gamma_n > 0, \gamma_n = 1$, such that,

$$\mathbb{E}(Y_{i,0}) = \gamma_i \mu_i.$$
The Assumption 1 (equation 8 and
implies that, 

\[ E(Y_{i,k+j}|Y_{i,0}, \ldots, Y_{i,k}) = Y_{i,k} + (Y_{k+j} - Y_{k})\mu_i \] (8)
holds for all \( i \in \{0, \ldots, n\}, \ k \in \{0, \ldots, n - 1\}, \ j \in \{1, \ldots, n - k\}. \)
The Assumption 1 (equation 8 and \( n = 1 \)) implies that,

\[ E(Y_{i,k}) = \gamma_k \mu_i \text{ dan } E(Y_{i,n}) = \mu_i. \] (9)
The unknown values \( Y_{i,k} \) are then predicted by a predictor,

\[ \hat{Y}_{i,k} = \hat{\gamma}_k \hat{\mu}_i, \quad i + k > n \] (10)
and calculated as shown on table 3, where \( \hat{\gamma}_k \) is so called prior estimator of development pattern for cumulative quotas \( \gamma_k \) and \( \hat{\mu}_i \) is so called prior estimator of expected ultimate losses \( \mu_i \) [5]. These prior estimators can be obtained from internal, external, or mixed information [6]. In this case, both prior estimators are obtained from external information [6].

Schmidt [7] states that in the modification Bornhuetter-Ferguson method for these prior estimators holds true:

\[ \hat{\mu}_i = \pi_i \hat{\kappa}^{CC} \text{ dan } \hat{\gamma}_{n-i} = \hat{\gamma}_{n-i}^{CL} \] (11)
where \( \pi_i \) is premium income of accident year \( i \) and \( \hat{\kappa}^{CC} \) is cape-cod loss ratio which is estimator of the expected loss ratio \( \kappa_i = E\left[\frac{Y_{i,n}}{\pi_i}\right] \) of the accident year \( i \); \( \hat{\gamma}_{n-i}^{CL} \) is the chain ladder estimator of parameter \( \hat{\gamma}_{n-i} \). In this case, \( \hat{\mu}_i^{CC} \) is obtained from mixed information and \( \hat{\gamma}_k^{CL} \) is obtained from internal information.

Thus, the Bornhuetter-Ferguson predictor of outstanding claims is defined by,

\[ \hat{Y}_{i,k}^{BF} = Y_{i,n-i} + (\hat{\gamma}_k^{CL} - \hat{\gamma}_{n-i}^{CL})\pi_i \hat{\kappa}^{CC}, \quad \text{for } i + k > n \text{ and } i, k \in \{0, \ldots, n\} \] (12)
where \( \hat{\gamma}_k^{CL} = \prod_{j=k}^{n-1} \frac{1}{\pi_j} \) [4], \( \hat{\kappa}^{CC} = \frac{\sum_{i=0}^{n-k} Y_{i,n-i}}{\sum_{i=0}^{n-k} \pi_i} \) [7] and \( \hat{\gamma}_{n-i}^{CL} = \frac{\sum_{i=0}^{n-k-1} Y_{i,k+i}}{\sum_{i=0}^{n-k-1} \pi_i} \) [4]. \( \hat{\gamma}_k^{CL} \) is the chain ladder prior estimator of development pattern for factors \( \lambda_k \).

Therefore, the claim reserve for accident year \( i \) (equation 5) and the total claim reserve (equation 14) is defined by,

\[ \hat{R}_{i}^{BF} = \hat{Y}_{i,n-i}^{BF} - Y_{i,n-i} \] (13)
\[ \hat{R}^{BF} = \sum_{i=1}^{n} \hat{R}_{i}^{BF} \] (14)

4. Results and discussion

4.1. Data description
The data used in this paper is from [1] that cumulative paid loss and ALAE (part 3) Commercial Auto/Truck Liability/Medical data, which is all industry in the U.S. for the period 1999–2008, where row represents accident year and column represents development month (in 12 months). The cumulative paid loss and ALAE is stated in USD (000’s). This paper will predict the claim reserve for next year, which is 2009.
Table 3. The calculated $\hat{\gamma}_k^{CL}$ and $\hat{\mu}_i^{CC}$

| $\hat{\gamma}_k^{CL}$ | $\hat{\mu}_i^{CC}$ |
|------------------------|---------------------|
| 0.2251                 | 7992240             |
| 0.4720                 | 8600937             |
| 0.6810                 | 9386804             |
| 0.8305                 | 10604584            |
| 0.9178                 | 11776474            |
| 0.9594                 | 12565933            |
| 0.9799                 | 12878831            |
| 0.9893                 | 12943178            |
| 0.9956                 | 12827070            |
| 1.0000                 | 12287162            |

Table 4. Claim reserving prediction

| Accident year $i$ | Claim reserve |
|------------------|---------------|
| 0                | -             |
| 1                | 37795         |
| 2                | 100244        |
| 3                | 213102        |
| 4                | 477786        |
| 5                | 1033195       |
| 6                | 2182351       |
| 7                | 4128768       |
| 8                | 6773078       |
| 9                | 9521097       |
| Total            | 24467415      |

4.2. Processed data
The data is development triangle in the form of cumulative losses. First of all, to obtain the standard form of the development triangle, it will be changed accident year \{1999, 2000, ..., 2008\} to accident year \{0, 1, ..., 9\}. Then, it will be changed development month \{12, 24, ..., 120\} to development year \{0, 1, ..., 9\}. Finally, the modified-data will be used to calculation of claim reserve.

4.3. Claim reserving prediction
The Bornhuetter-Ferguson method is applied for the data. We can see in the table 3, the results of the prior estimator of development pattern for cumulative quotas \($\hat{\gamma}_k^{CL}$\) and the prior estimator of expected ultimate losses \($\hat{\mu}_i^{CC}$\) are presented. The results of the claim reserve for accident year $i$ and the total claim reserve are presented in the table 4. Thus, the total claim reserve that must be prepared by all industry in the U.S. for claim payments in 2009 is $24,467,415 \text{(in 000's).}$

5. Conclusion
This paper concludes that the problem of prediction of outstanding claims utilizes run-off triangle. Run-off triangle is a very useful instrument, which is used to analyze the relationship between claims payments that have been made over the past few years. By using this instrument, we applied the
Bornhuetter-Ferguson that is one of the most popular claim reserving techniques with distribution-free. According to the result of paid loss and ALAE (part 3) Commercial Auto/Truck Liability/Medical data, the total claim reserve that must be prepared by all industry in the U.S. for claim payments in 2009 is about USD 24.4 billion.

References
[1] Shapland M R and Xiao P 2016 Using the Hayne MLE models: A practitioner’s guide The Casualty Actuarial Society (CAS) Forum Summer 2016 (Arlington: Casualty Actuarial Society) Supplements Industry Data pp 1–116
[2] Schmidt K D and Zocher M 2008 Variance 2 85–110
[3] Olofsson M 2006 Stochastic Loss Reserving Testing the New Guidelines from the Australian Prudential Regulation Authority (APRA) on Swedish Portfolio Data Using A Bootstrap Simulation and The Distribution-Free Method by Thomas Mack Master Thesis (Stockholm: Stockholm University)
[4] Zboňáková L 2014 Generalized Linear Models in Reserving Risk Master Thesis (Prague: Charles University)
[5] Linda B and Kubanova J 2012 Proc. 30th International Conference on Mathematical Methods in Economics September 11-13 Karviná, ed J Ramík and D Stavárek (Karviná,: Silesian University in Opava) pp 524–9
[6] Schmidt K D 2006 Methods and models of loss reserving based on run-off triangles: A unifying survey Casualty Actuarial Society (CAS) Forum Fall 2016 (Arlington: Casualty Actuarial Society) 269–317
[7] Schmidt K D and Zocher M 2008 CAS: Variance 2 85–110