Stockholms universitet

Optimizing method selection for IBNR-reserve calculation using machine learning

Dennis Uygur Andersson

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2023:14
Försäkringsmatematik
Juni 2023

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2023:14**
http://www.math.su.se

Optimizing method selection for IBNR-reserve calculation using machine learning

Dennis Uygur Andersson*

June 2023

Abstract

This work implements an approach which were introduced by Caesar Balona & Ronald Richman in their article *The Actuary and IBNR Techniques: A Machine Learning Approach* (2021), that combines the strengths of both traditional and machine learning reserving methods. This approach is still based on the ordinary reserving methods available today, such as chain ladder and the Bornhuetter–Ferguson method, with the modification that we vary how the loss development factors are estimated and included/excluded. This is done using AvE and CDR as score tests. The estimated reserves using the machine learning approach were then compared to corresponding reserves using the standard methods. The outcome showed that the ordinary reserving methods, especially the chain ladder method, overall performed better, even though we sometimes gained better results using the new approach.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: dennisuygurandersson@gmail.com. Supervisor: Filip Lindskog.

Acknowledgments

I would like to thank my supervisor, Filip Lindholm, for the support during this thesis. I would also like to show my gratitude towards Fredrik Käll for his helpful feedback. Also, I would like to thank SiriusPoint for supporting me with data.

Lastly, I would like to thank the people near me for their mental support. You know who you are.

Contents

1 Introduction

As a reserving actuary at an insurance company, the primary responsibility is to calculate the Incurred But Not Reported (IBNR) reserve, which is a fund set aside for claims that the insurance company expects but has not yet been notified of. Today actuaries use standard reserving methods, such as the Chain-Ladder method, in software programs such as ”ResQ”. However, determining the reserve requires making a subjective decision about which loss development factors (LDF) to include in the calculation. This decision can be challenging because some LDFs may not accurately reflect reality and can be considered outliers. This subjective decision-making process is not foolproof and can lead to errors, as human judgment is a significant factor in the process today.

The methods for calculating reserves are constantly evolving. People are trying new ways to estimate reserves using statistical techniques like Bayesian methods and the double chain-ladder approach. However, machine learning is becoming more popular in this field, with mathematicians like Wüthrich implementing neural networks to improve the prediction of loss development factor (2018). While the existing machine learning approaches within reserving can produce better results, it often requires more complex models. Recently, Balona and Richman (2021) introduced a new technique that uses actual versus expected (AvE) and Claims Development Result (CDR) that combines the strengths of both traditional and machine learning methods.

The goal of this paper is to implement the machine learning approach from [1]. We will do this by using data based from a reinsurance company called Sirius-Point in form of case studies and compare this to ordinary reserving methods that exists today.

2 Theory & Methods

This section will, just as the title suggests, include the necessary theory and explanation of the machine learning approach that will be applied for trying to investigate the objective. Some of the theoretical framework will certainly be familiar, and we will therefore not go into too much detail, but is nevertheless important to the analysis. More important, we will highlight the parts that are essential to Balona & Richman method which is suggested in Section 2.3. We will however start to introduce loss triangles with corresponding notations that will be used throughout this thesis.

Loss triangles & notations

Following notations and theory from [1], when trying to determine the sought IBNR-reserve, we use known losses from observed years. That is, if we let $i \in [1, \ldots, I]$ denote a certain accident year and $j \in [1, \ldots, J]$ development year, we can write the incremental payment for when a loss occurred and were payed (or changed) a certain year as X_{ij}. Note that the incremental payments do not necessarily have to be based on years. The concretization is an arbitrary choice, and we could nevertheless assume that the payments occurred quarterly or monthly or according to any other time period. We will however stick to yearly payments during this paper, which will be motivated later on. Also, insurance companies usually separate the claims between underwriting-, reporting- and accident years, but for the sake of simplicity and uniformity, we will continue this paper using accident year. Furthermore, the reserving methods we will use is actually based on the so called cumulative claims,

$$C_{i,j} = \sum_{l=1}^{j} X_{i,l}, \tag{1}$$

which we mostly will focus on from now on. We can now set up triangles which we use as a basis to determine the IBNR-reserve. These triangles are called loss (development) triangles or reserve triangles and we will denote such a triangle observed as Δ^K, where K is how many calendar years of claim data that have been observed. We will assume that the observation is done at the end of the year, i.e. so that the periods for each calendar year is observed at the same moment. From this, we let $k \in [1, \dots, K]$ be the k:th calendar year and the reserving triangle can therefore be defined as

$$\Delta^K = \{C_{i,j} : i + j \leq K + 1\}$$

where equality corresponds to the most recent calendar year. Figure 1 illustrates how such a claim triangle could look. Note that the lower triangle, i.e. where $C_{i,j} : i + j > K + 1$ are events that has not yet been observed. Indeed, the goal is to estimate the final claim amount (ultimo), $C_{i,J}$, that further allows us to calculate the desired IBNR reserve. The ultimo claims for different calendar years k which we will try to estimate through the progress of the article will be denoted $\hat{C}^k_{i,J}$ and the corresponding IBNR reserve is calculated as

$$R_{i,j^*} = \hat{C}^k_{i,J} - C_{i,j^*}, \tag{2}$$

where $j^* = k - i + 1$, i.e. the latest observed development year based on calendar year k. Today there are several techniques that allow us to estimate ultimo such as (generalized) cape cod, the loss ratio method, chain ladder method and the Bornhuetter–Ferguson method, where we will focus on the latter two.

2.1 Chain Ladder Method

As mentioned earlier, there are several methods trying to determine the ultimate claim, but one of the most famous are the Chain Ladder (CL) method. There are varieties which merely depends on how we condition the expected value and

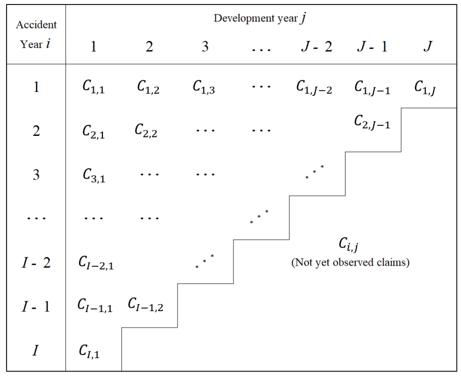

Figure 1: How the cumulative payments are illustrated in a reserve triangle, where $C_{i,j} : i+j \leq K+1$ are observed cumulative claim amounts, where equality corresponds to current year. These are represented in the upper triangle.

variance for $C_{i,j+1}$. In this paper, we will apply one of the most established CL methods available, namely Mack's [8] variant:

$$E[C_{i,j+1}|C_{i,1},\ldots,C_{i,j}] = C_{i,j}f_j$$
$$\text{Var}(C_{i,j+1}|C_{i,1},\ldots,C_{i,j}) = C_{i,j}\sigma_j^2$$

Simply put, we believe that the expected claim for accident year i and development $j+1$ depends on the claim from previous development year multiplied by a factor f_j, i.e. the loss development factor. We also assume that the claims for different accident years are independent. It can be shown [7] that the LDFs can be estimated accordingly

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}} = \frac{\sum_{i=1}^{I-j} C_{i,j}\hat{f}_{i,j}}{\sum_{i=1}^{I-j} C_{i,j}} \tag{3}$$

where

$$\hat{f}_{i,j} = \frac{C_{i,j+1}}{C_{i,j}} \tag{4}$$

And thus, the searched ultimo using Mack's Chain-Ladder method for accident year i is determined by

$$\hat{C}_{i,J}^{CL} = C_{i,j}\prod_{k=j}^{J}\hat{f}_k \tag{5}$$

2.2 Bornhuetter–Ferguson Method

The other method to be implemented is the Bornhuetter-Ferguson (BF) method [2], [10] & [1]. In contrast to Mack's Chain Ladder above, the BF method is based on our (the actuary's) expected ultimo. Letting β_j denote the proportion of losses which were reported in development period j, the ultimate claim is estimated according to

$$\hat{C}_{i,J}^{BF} = C_{i,j} + (1 - \beta_j)E[C_{i,J}^{k}] \tag{6}$$

Worth mentioning is that the β_j:s normally is taken as the inverse of the products from equations (5), i.e. $\prod_{k=j}^{J} \hat{f}_k$, which sometimes is called cumulative development factor (CDF). Also, on the right hand side of equation (6), the expected value of the ultimate claim for each accident year, i.e. $E[C_{i,J}^{k}]$ is the actuary's belief on what the ultimate claim amount will be. For instance, if we let $E[C_{i,J}^{k}] = \hat{C}_{i,J}^{CL}$, that is determining the ultimo using the chain ladder method, we see that equation (6) reduces to the ordinary CL-method described above. Moreover, it is not unusual that the ultimate claims using the BF-method is determined using the earned premium and corresponding ultimate loss ratio. This because of removing the dependence of already known claims. If we let π_i denote the earned premium and $\widehat{ULR}_i^{k,BF}$ the corresponding expected ultimate loss ratio for accident year i at time k, we can rewrite (6) as

$$\hat{C}_{i,J}^{BF} = C_{i,j} + (1 - \beta_j)\pi_i \widehat{ULR}_i^{k,BF} \tag{7}$$

Different insurance companies use different definition for the concept earned premium. We will for simplicity refer π_i as the single earned premium which is assumed to be contracted and expired during year i. With other words, premium earned on a policy that has been in force for only one year or less. Moreover, the ultimate loss ratio may also not be self-evident. This is the ratio between the actuary's belief of ultimate loss and the ultimate earned premium for accident year i. This ratio is obviously not known in practice, but instead based on previous years data and is also therefore selected subjectively. This is also the case in this paper, and as we will see further on be varied.

2.3 Machine Learning Approach

The field of machine learning is currently growing, including trying to implement its field to reserving. Hastie et al. [4] suggests different methods for so called supervised learning, which is some of the more established approaches today, such as neural network, gradient boosting and random forest. Even though some of this methods have been implemented to reserving (e.g. [5] & [12]) we will be attempting to use a different approach which were introduced by Balona & Richman [1].

One of the main features within supervised learning is dividing the data into two sets; training and test (validation) set. This is exactly what it sounds like, namely slicing the existing data into two components where first set trains the model and the second one test the trained model, especially if great amount of data is supplied. On the other hand, if we have a significantly smaller amount of data, other methods can be used such as cross validation [4]. Both these methods share a similarity in that they test how well the predictive model performs on data that has not been used in the initial training. The model that shows the best accuracy in predicting the outcome is typically chosen from a range of model types, formulations, and parameters. Unlike traditional actuarial modeling, which relies on expert knowledge and experience to select models subjectively, the machine learning approach selects models, parameters, and their values based on their ability to improve the accuracy of the predictions. How this will be implemented onto our reserve triangles will be explained later on. Moreover, another key component is how we test the trained models, and with our approach, will use two different score tests called Actual versus Expected (AvE) and Claims Development Result (CDR).

2.3.1 Score test: AvE & CDR

Actual versus expected (AvE) is a frequently used measurement in many insurance companies, and is merely the difference between the actual claim and what we expected it to be. The AvE-analysis also forms a large part of the insurance company's decision whether the IBNR should be changed or not. We define the AvE result of the incremental payments for upcoming calendar year as

$$AvE^k_{i,j^*} = X_{i,j^*} - \hat{X}^{k-1}_{i,j^*} \tag{8}$$

Moreover, the claim development result (CDR) is defined as the difference between best estimated ultimo now and the best estimated ultimo a future point in time. Here we will use one time point (year) and the CDR for accident year i in year k is defined as

$$\begin{aligned} CDR^k_i &= \hat{C}^k_{i,J} - \hat{C}^{k-1}_{i,J} \\ &= R^k_{i,j^*} + C_{i,j^*} - (R^{k-1}_{i,j^*-1} + C_{i,j^*-1}) \\ &= \sum_{l=j^*+1}^{J} \hat{X}^k_{i,l} + C_{i,j^*} - \left(\sum_{l=j^*}^{J} \hat{X}^{k-1}_{i,l} + C_{i,j^*-1} \right) \\ &= R^k_{i,j^*} - R^{k-1}_{i,j^*} + (X_{i,j^*} - \hat{X}^{k-1}_{i,j^*}) \\ &= R^k_{i,j^*} - R^{k-1}_{i,j^*} + AvE^k_{i,j^*}, \end{aligned} \tag{9}$$

where we see how AvE^k_{i,j^*} pops out in the last two rows. Indeed, we see how the CDR can be expressed in terms of the IBNR reserves, or more properly, as the sum of the difference between the estimated reserves for development j^* plus the actual versus expected analysis defined in (8). To break it down even more, we can see equation (9) as two components:

1. The precision of the predictions using the AvE measure, i.e. $(X_{i,j^*} - \hat{X}^{k-1}_{i,j^*})$ plus

2. $R^k_{i,j^*} - R^{k-1}_{i,j^*}$ which is merely the "stability" of our IBNR-reserves estimates.

The main thing here is that CDR gives an indication of how good the reserve method is, especially with enough information. We will therefore use both CDR and AvE in our analysis as measuring instruments for different methods, where the choice will land on the method that minimizes the score tests as an objective function.

The goal here is reducing the squared differences between both the metrics, AvE/CDR, and zero. This could be seen as minimizing $(AvE-0)^2$ and $(CDR-0)^2$. But also, since the claims for various calendar year differs, we will calculate the score by weighting on corresponding incremental payment which allows us to calculate the following equation:

$$CDR^k_{score} = \sqrt{\frac{\sum\limits_{i=1}^{I} |X_{i,j^*}|(CDR^k_i)^2}{\sum\limits_{i=1}^{I} |X_{i,j^*}|}} \tag{10}$$

In equation (10), we see how the score is weighted on the incremental payments made. We can use the same approach for determining the corresponding score function for AvE, namely

$$AvE^k_{score} = \sqrt{\frac{\sum\limits_{i=1}^{I} |X_{i,j^*}|(AvE^k_i)^2}{\sum\limits_{i=1}^{I} |X_{i,j^*}|}} \tag{11}$$

2.3.2 Algorithm

The algorithm to be used will now be introduced and explained, taken directly from [1], and we will also illustrate how the algorithm is used using figures. Here, we will also see how we will train the sliced data using triangles.

Recall from Section 2 that we denoted the reserve triangles for calendar year k as Δ^k. For such an arbitrary triangle, the actuary tries to implement a certain model in order to predict the future (ultimate) claims. Let us denote this model as M, which is from a certain model space $\mathcal{M}$, i.e. $M \in \mathcal{M}$. The estimated reserve based on a certain model can thus be denoted as

$$\hat{R}^{k,M}_{i,j} = M(X = \Delta^k)$$

To make a decision based on data, we need to choose a model $M \in \mathcal{M}$ that will perform well in predicting future claims amounts. This requires scoring each potential model. Rather than using a statistical distribution to determine likelihood, we will use a score that aligns with the goals of the analysis, which is to minimize the difference between AvE claims or CDR on so called ”out of sample” data. When calculating the scores on the out of sample data, we will make comparison between what actually have been paid and our estimates. Note that in normal cases the actuaries do not have fully developed claims triangle, but as we will see later, we will. This is to be able to estimate how well the model has performed. The following algorithm will be used:

- Select a reasonably sized triangle of claims development experience which will provide data for fitting all of the models (shown as ”Initial Triangle” in blue in Figure 2). Here, the new diagonals of experience will be added to this triangle in the subsequent steps. Note that Figure 2 uses 15 accident and development year, which necessary does not need to be the case for the data used in this paper.

- Select several of the most recent calendar years of the triangle as the training set with the first calendar year in the training set being k_{train} (shown as ”Training Data” in green in Figure 2).

- Select a reserving model for each M in a subset of $\mathcal{M}$. For each M perform the following steps:

    1. For the first calendar period k_{train} in the training set, find the reserves by calibrating all of the model parameters on $\Delta^{k_{train}}$, which is the blue area plus the first diagonal as shown in Figure 3.
    2. For each subsequent calendar year k in the training set:
        (a) Calculate the score for each accident year as CDR_i^k or AvE_{i,j^*}^k, based on the next diagonal of experience. Note that, at this stage, this next diagonal has not been used to fit the reserving model, i.e. it is out of sample data.
        (b) Calculate the weighted score across accident years, using the incremental claims as weights. See Equation (10) for example.
        (c) Re-estimate the reserves $R_{i,j}^k = M(X = \Delta^k)$ by refitting the model using the extra calendar year of data. This is shown for the first iteration in Figure 4.
    3. Calculate and store the average score across all of the calendar years in the training set, S^M

- Select $M_{opt} = \text{argmin}_{M\in\mathcal{M}}(S^M)$.

As suggested earlier, we will only use two different reserving methods (CL- and BF-method) which reduces the subset of $\mathcal{M}$ to these, where later on will see how we do a grid search using the algorithm above, which enables numerous different ways to calculate the score and suggestively select the method that minimizes the score. We will also use both of the metrics (CDR and AvE), much since CDR is like a more regularized version of AvE, where we penalize

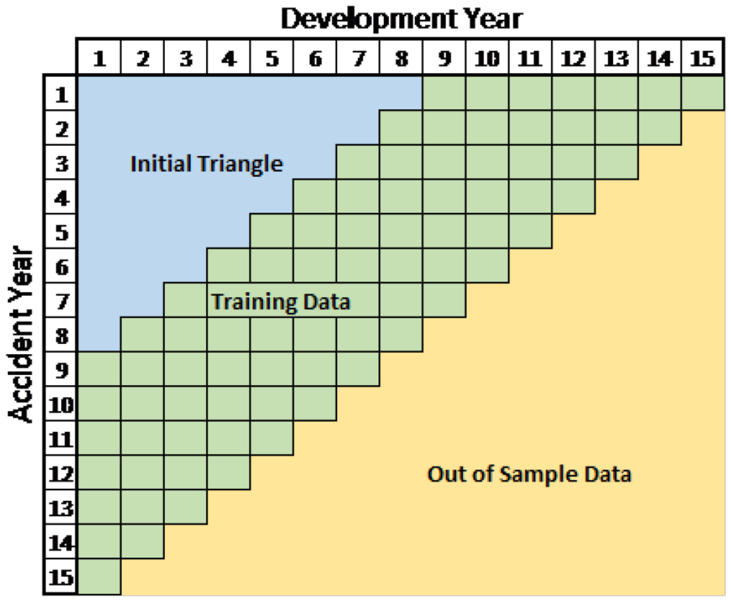

Figure 2: A general reserve triangle which is describing the procedure from above. The models start with the basic triangle shown in blue, and then additional information from a training set (green part) is added. The yellow part illustrates unknown claims, i.e. claims that has not yet been observed. Note that this figure is taken directly from [1].

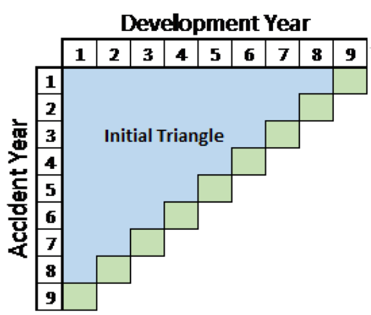

Figure 3: Illustration of the first iteration of the scoring procedure. In the first step, the model is fitted to the initial triangle, which is a starting point for analyzing the insurance data. Then, the experience of the insurance data is assessed against the first diagonal of the training data. This figure is taken directly from [1].

the difference in reserve estimates between calendar years. This ensures that the best algorithm not only makes accurate predictions but also keeps the reserve

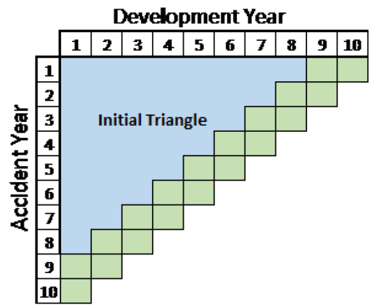

Figure 4: In the first iteration, the model is fitted to the initial triangle and compared to the first diagonal of the training data. In the second iteration, the first diagonal is added to the initial triangle and the model is fitted to this augmented triangle, then compared to the second diagonal of the training data. This figure is taken directly from [1].

estimates stable over time. This is according to Balona et al. important for long-tailed businesses, i.e. classes which take long time to develop. However, for short-tailed classes, it may not be practical to aim for stable reserve estimates. In these cases, the AvE metric seem to be a better way to optimize reserve estimates. Because of this, we will during our case studies use two different types of data which correspond to just such qualities, which we will see later.

2.3.3 Numerical example

We will try to further clarify how the different metrics work with the help of a simple numerical example that will hopefully also distinguish the difference between them. To help us, we will as support for the reader once again use Figure 3 during the example and the colours from Figure 2 represents the same part of the data. That is, the blue area is our initial triangle and the green part consists of our training set. The yellow part, which is the out of sample data, is merely there for the sake of completeness but will not be used here. We will use the same numerical example as Balona et al. for simplicity. They also take it one step further and illustrates the next step of the iterations with actual numbers and detailed calculation. We will however not do that since the methodology is the same as below.

Initially, we start of with the first iteration of the scoring procedure which is using the fully initial triangle (blue area) trying to determine the first diagonal of the training data. Assume that we have the following actual incremental payments and the corresponding expected incremental claims shown in Table 1. These numbers are obviously arbitrary but nevertheless sufficient considering the goal.

Moreover, to calculate the score for the first step here, we look back at equation

	Development year, j^*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A	Actual	15	12	10	8	6	4	2	0	0
B	Expected	18	10	9	10	4	2	1	0	0

Table 1: Numerical example of actual and expected incremental claim for the first diagonal shown in Figure 3. This table is taken directly from [1].

(11) which shows us how this is done. Table 2 illustrates the details on each of these steps.

	Development year, j^*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$C = A - B$	AvE^8_{i,j^*}	-3	2	1	-2	2	2	1	0	0
$D = C^2 \cdot A$		135	48	10	32	24	16	2	0	0
$\sum D / \sum A$	AvE_{score}					**2.16**				

Table 2: Continuing from Table 1 where the score is calculated for the metric AvE. This table is taken directly from [1].

The procedure for calculating the AvE_{score} is not harder than described above. We will continue using the same numerical example for calculating the score of the CDR-metric. Remember from equation (9) that we use the difference in the reserve estimates. In this case, this means $(\hat{R}^8_{i,j^*} - \hat{R}^7_{i,j^*})$ is added for each accident period in AvE^8_{i,j^*} which were calculated in Table 2. The details for calculating the CDR_{score} this is shown in Table 3.

	Development year, j^*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
F	R^7_{i,j^*}	24	7	4	3	2	1	1	0	0
G	R^8_{i,j^*}	26	8	4	3	2	1	1	0	0
$H = G - F$		2	1	0	0	0	0	0	0	0
$I = H^2 \cdot A$		44	4	0	0	0	0	0	0	0
$\left(\sum D + I\right) / \sum A$	CDR_{score}				**2.35**					

Table 3: Calculation steps when determining the CDR_{score} using our numerical example. This table is taken directly from [1].

We notice from Table 2 and 3 is that the AvE_{score} and CDR_{score} are 2.16 and 2.35 respectively. What can be seen from the detailed calculation of the CDR_{score}, it seems as suggested earlier that the CDR penalize for the difference in the reserve. For the latter development years, we see that the contribution

I is 0 which probably will be the same in most cases since we estimate the same reserve but for different calendar years. Moreover, the example above is only using the first diagonal of training set, e.g. the first iteration of the scoring process. To complete the procedure, the next step is to include another calendar year and repeat the same calculations as above, but now including the first diagonal from the training set. Finally, additional information from the training set is added until it is no longer possible, i.e. until development year 15 in this case. The final score is retrieved when taking the mean from the different scores.

2.3.4 Implementation of reserving methods

Lets recall the two methods from earlier, namely chain ladder method and Bornhuetter–Ferguson method, and the corresponding equations for estimating the ultimo, i.e. equation (5) and (7). The next step is to sort out how these will be implemented to the algorithm described in Section 2.3.2. As described in Section 1, the actuary makes a subjective choice on whether to include/exclude certain development factors (see equation 4) to be a part of the analysis and determination on the IBNR reserve. This mostly concerns the highest and/or lowest factor, since they could be outliers. We will therefore introduce the following equation for calculating the LDF:s

$$\hat{f}_j = \frac{\sum\limits_{i=1}^{I-j} C_{i,j}\hat{f}_{i,j}w_{i,j}}{\sum\limits_{i=1}^{I-j} C_{i,j}w_{i,j}} \tag{12}$$

where $w_{i,j}$ denotes the weights which equals 1 if we want to include a certain development factor, otherwise 0. As will be seen later, we will carry out the analysis by see how the CDR- and AvE-scores change when excluding the highest and lowest development factor compared. Moreover, for the BF-method, we will also vary the loss ratio parameter $\widehat{ULR}_i^{k,BF}$ within an arbitrary, but nevertheless reasonable range. Lastly, when calculating the loss development factors, the ordinary Mack's chain ladder method uses as many accident years as possible. We will however let these number of years vary for both the suggested methods.

Referring to Section 2.3.3 where a numerical example were illustrated, we will therefore get different scores from the different alternations of parameters. That is, depending on the choice of number of accident years, and whether we drop the highest and/or lowest individual development factor (and ultimate loss ratio) that gives the lowest score for the respective metric, is the method we will compare against the ordinary reserving method. This will hopefully also be clarified more during the actual case studies.

3 Data

The data that will be used is based on claims from the international reinsurance company SiriusPoint which has its head office in Pembroke, Bermuda. But when working with a single dataset, the outcome of a method may be influenced by factors such as luck or chance, making it difficult to interpret its true effectiveness. However, by simulating multiple datasets, we can obtain a range of outcomes and use statistical measures, which exactly what we will do. Note, however, that the important thing will not be how well we have recreated the real data, but only that we have captured the desired property - how long time it takes for the claims to be developed. The data from SiriusPoint is not picked at random, but here we have selected data that we know through experience have the desired characteristics. For privacy reasons, we will not disclose the type of data used. The claims are also manipulated by a certain factor.

Thus, we will use two data sets with one consisting of 10 accident and development years respectively and the other data set with 15. This choice is not random since we by experience know that these are fully developed after that many years. For the data set of 10 years, it will hopefully reflect the short-tailed business and vice versa for the other data set. We could have used even more years, but that is redundant and unnecessary for the analysis. As we discussed earlier, the time-concretization ”years” is an arbitrary choice and we could have just as easily do the analysis on a quarterly (or monthly) basis. However, since we merely are interested how well the ML-approach performs relative to the ordinary reserving methods, this is redundant. Also, using more time steps increases the computational power and is more time consuming.

The two data sets will be divided into two different case studies. For the first one, which will use the situation of 10 years, we will only do one simulation using a great number of claims for each accident year. Even though this gives us enough information to carry out the analysis, it could perhaps, just as for the case of the real insurance data, be influenced by factors such as luck or chance. Because of this, the other case study (that contains 15 accident and development years), will be simulated multiple times. 50 times to be exact. For each of this simulation, we will not only try to mimic the long-tailed property, but also use significantly fewer number of claims for each accident year. This will hopefully create randomness in the reserving triangle which the machine-learning approach will deal with.

3.1 Simulation

There are several ways to simulate claim data and triangles. One of the more commonly used is suggested by Gaberielli & Wüthrich [3], which is based on Swiss data and using a stochastic simulation machine, generates individual non-life insurance claims. Another one is from Lindholm et al. [6], that simulate according to a time series model which consistent is with CL. We will however take another approach that Verall et al. uses. Their method is convenient here since it is based on the same triangular form as our described reserve triangles. The mere difference is that instead of only using the claim amounts, we also use the underlying number of claims ([9] & [11]).

3.1.1 Claims

In order to simulate claim numbers, initially the number of claims that occurred for a certain accident year i will be simulated. We denote the number of incurred claims for accident year i fully paid in year j as N_{ij}. By using the following method we can simulate desirable data. For each of our data set which consists of dimension 10 and 15 respectively:

- Generate number of claims $N_{i,\cdot} = \sum_{j=1}^{J} N_{i,j}$ for each accident year using a Poisson distribution.
- For each number of claims given each accident year, simulate the delay using a multinomial distribution, that is

$$(N_{i,1}, \dots, N_{i,j}) \sim Multi(N_{i,\cdot}; p_1, \dots, p_j)$$

  where p_j denotes the delay probabilities. It can be shown [9] using maximum likelihood theory, how the delayed probabilities is calculated using the empirical data. This should encapsulate the desired tail-property.
- For each of the individual payments, simulate claim amount using a Gamma distribution where the SiriusPoint empirical data gives us the desirable parameters through ML-theory.

3.1.2 Premium

During the case studies, information about the premium is not provided from SiriusPoint, but is nevertheless needed to be able to use the BF-method. Indeed, the premium for accident year i denoted π_i (see equation (7)) needs therefore to be simulated in turn to be able to implement the desired reserving method. Note, we will use the same method as Balona et al. simulating the premium but also assume the same loss ratio as them, i.e. a 60% loss ratio. While it is true that the LR obtained through this simulation may not accurately represent the true loss ratio, it is not a concern as our primary objective is to compare the effectiveness of the traditional BF approach with our machine learning approach. Therefore, the exact loss ratio obtained is not crucial as long as it can serve the purpose of facilitating a comparison between the two methods. The simulation looks as follows:

1. The real ultimate claims, $C_{i,J}$, is taken from the most recent development year. That is, for each accident year, we use the actual ultimate claim amount, keeping in mind that we have access to the complete set of data for the entire triangle.

2. To introduce an element of variability in our analysis, we have fitted a linear regression model to the ultimate claims. That is, for each accident year i, let each cumulative claim $C_{i,j}$ be our responsive variable and development years j as covariate. Let the estimated ultimate claims using this method be denoted as $\tilde{C}_{i,J}$.

3. We calculate the residuals for each accident year as $\epsilon_i = C_{i,J} - \tilde{C}_{i,J}$ using single iteration of bootstrapping. This generates a vector of the same size as the number of accident years.

4. Using the residuals from the step 3, we calculate so called pseudo-ultimate claims that we denote $C^*_{i,J} = C_{i,J} + \epsilon^*_i$ for each accident year. This step could be questioned since this could mean taking residuals from a certain accident year and use it on a completely different one. However, since we want some variability in the loss ratio, this is an eligible property as long as the difference between the accident years is somewhat similar. For the two data sets, this is not an issue. Firstly, we will only use the BF-method for one of the case studies and thus only need to simulate the premium once. For that study, the claims are fairly similar between all the accident years and we therefore also retrieve stable simulated epsilons (and conclusively ultimate claims).

5. We finally calculate the pseudo premium based on a 60% loss ratio: $\frac{C^*_{i,J}}{0.6}$

3.1.3 Properties

We have now simulated both claim data and earned premium and can move forward with looking at the data. First of all, we are interested in whether we succeeded in capturing the desirable development characteristics of the respective data. As stated before, we will for the first case study only simulate once, and Figure 5 illustrates the loss development factors which were calculated using ordinary (Mack's) chain ladder over the different development year. Correspondingly, we determine the loss development factors for the long-tailed business which we see in Figure 6.

What we first notice is that we were able to encapsulate the properties from both classes which was what we wanted. It seems like the short-tailed class reaches its full development after around four years while it takes the second business varies between ten to fifteen years.

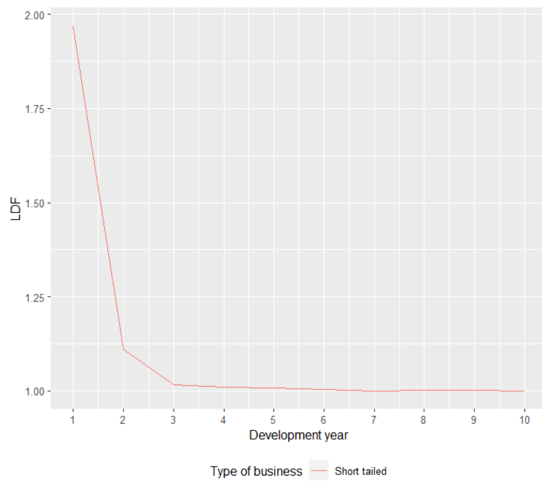

Figure 5: Loss development factors for the first case study using 10 development years, based on real data from the insurance company SiriusPoint. The line corresponds to the short-tailed business which in total has 10 years of accident and development years.

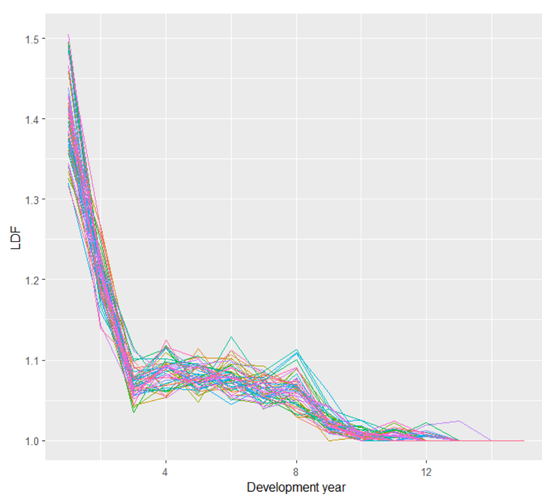

Figure 6: Loss development factors for the second case study using 15 development years, based on real data from the insurance company SiriusPoint. The lines corresponds to the long-tailed business which in total has 15 years of accident and development years. The number of simulated claims is significantly fewer and consequently creates randomness.

4 Case study 1: Short tailed business

For the first case study, we will try to implement what was discussed earlier. Recall from Section 3.1.3 that the short-tailed class seemed fully developed after approximately 4 years. The minor movements after that is marginal, but the approach should nevertheless encapsulate that. We will also during this (and the upcoming) case study use actual years. For this rapidly developing class, we have the accident years 1986 - 1995. Moreover, Figure 7 illustrates the loss development factors for the reserve triangle using Mack's CL-method, where the highlighted values corresponds to the highest and lowest individual development factor (IDF) respectively per development year. Note that all the accident years were used here during the calculations. The data used for this case corresponds to the years 1986 to 1995 (10 years).

We also have to make a choice on the number of years to train the model on. This choice should be based on us using as much training data as possible without starting with too small a triangle. We will therefore use the years 1986 - 1990 as initial triangle which results the years 1991 - 1995 as training set. We will now try to implement the two different reserving methods.

1,92207	1,11436	1,01568	1,01222	1,00695	1,00719	1,00053	1,00083	1,00204
1,96704	1,09589	1,01151	1,01235	1,00854	1,00475	1,00077	1,00231	
2,0249	1,11444	1,01948	1,01143	1,01074	1,00385	1,00127		
1,95941	1,11327	1,01741	1,00819	1,01108	1,00501			
1,91223	1,11466	1,01837	1,01275	1,00852				
1,96664	1,10657	1,01706	1,01122					
1,97622	1,10889	1,01635						
1,98873	1,11405							
1,98171								

Figure 7: Individual development factors for the short-tailed business where the red values corresponds to the highest IDF for each development year and the blue the lowest. The accident years corresponds to 1986-1995 and the development years 1 - 10.

4.1 The Chain Ladder Method

For the Chain Ladder approach, we will as discussed in Section 2.3.3, consider three variations for implementing the algorithm. Firstly, we will vary the number of years which the loss development factors is based on. Looking back at equation (3), we base the estimation for $\hat{f}_j$ on all accident years. The choice of parameter space here is semi-arbitrarily chosen in the sense that we want to include the highest amount of accident years possible since it equals ordinary chain-ladder estimation of the LDFs. Note that the IDFs illustrated in Figure 7 will look different for each choice of the number of accident years. The other thing we will vary is whether or not drop the highest and lowest loss development factor respectively. That is, for equation (12), we let $w_{i,j} = 0$ for the highest and lowest value $\hat{f}_{i,j}$ in all development years and see what happens to the score. Table 4 is a summary over how the search space looks and there are thus $2 \cdot 2 \cdot 7 = 28$ different combinations of parameters.

Parameter	**Possibilities**	**Description**
Drop $\max(\hat{f}_{i,j})$?	[True, False]	Vary the possibility to ignore the highest (individual) development factor in all development years.
Drop $\min(\hat{f}_{i,j})$?	[True, False]	Vary the possibility to ignore the lowest (individual) development factor in all development years.
Number of periods	$I \in [4, 10]$	How many accident years that should be used when calculating the loss development factor for each development year.

Table 4: Parameter space for CL-method. We vary the number of periods and whether to drop the highest and/or lowest individual development factor for all development periods.

It is obvious that the customary chain ladder method is retrieved by using all periods and not dropping either of the highest or lowest individual development factors. We will now implement the approach from Section 2.3, i.e. minimize AvE_{score} and CDR_{score} using grid search and the optimal parameters is found in Table 5. We notice that when minimizing the claims development result, we should use the lowest number of periods possible. Also, both performance metrics suggests that the we should ignore the lowest individual development factor in all development years. Indeed, looking at Figure 8 we notice that dropping $\hat{f}_{i,j}$ for all periods seems to reduce the score for both metrics. The scores does not also seem to be dependent on the number of periods, except perhaps when we drop both the highest and lowest individual development factor. In that case, we seem to see a somewhat linear trend for when the number of periods increases. This could also just be a random effect.

Parameter	**Ordinary CL**	$\mathbf{min}\left(\mathbf{AvE^{k}_{score}}\right)$	$\mathbf{min}\left(\mathbf{CDR^{k}_{score}}\right)$
Drop $\max(\hat{f}_{i,j})$?	False	False	False
Drop $\min(\hat{f}_{i,j})$?	False	True	True
Number of periods	10	6	4

Table 5: The choice of parameters which minimizes the respective metric for Chain-Ladder.

What is more interesting is that the ML-approach seem to overestimate the ultimate claim. This is perhaps not concerning in normal cases since it is a part of the reality, but it is nevertheless alarming since the actual claim amount versus the expected ultimo claim for each accident year is more off comparing to the ordinary chain ladder method which, according to Figure 9, seems to perform better. Before we conclude that the new method does not work, we need to investigate a few things. The first question is how big of a difference we actually are talking about in total. The IBNR reserve can be useful to us in this situation. Recall from Section 2, for each accident year, IBNR is calculated as the difference between the ultimate claims and the latest known claims. The total IBNR reserve is the sum of the reserves calculated for each accident year. In our case, this obviously corresponds to the estimated ultimate claim for each method. We start by looking at the total suggested reserve which is illustrated in Table 6, we see that even though the estimation when minimizing the metrics CDR and AvE is higher than both the ordinary CL and the actual IBNR, the difference in percent is not that high. That is, even though the normal chain ladder method seem to perform better in this case, the machine learning approach is not to be rejected. This is even clearer when looking at Figure 10 where we see that the difference in IBNR for accident years is relatively low when comparing the methods. Moreover, we could also question the choice of training set. Would it differ if we varied the choice on how many accident years we devote as initial triangle?

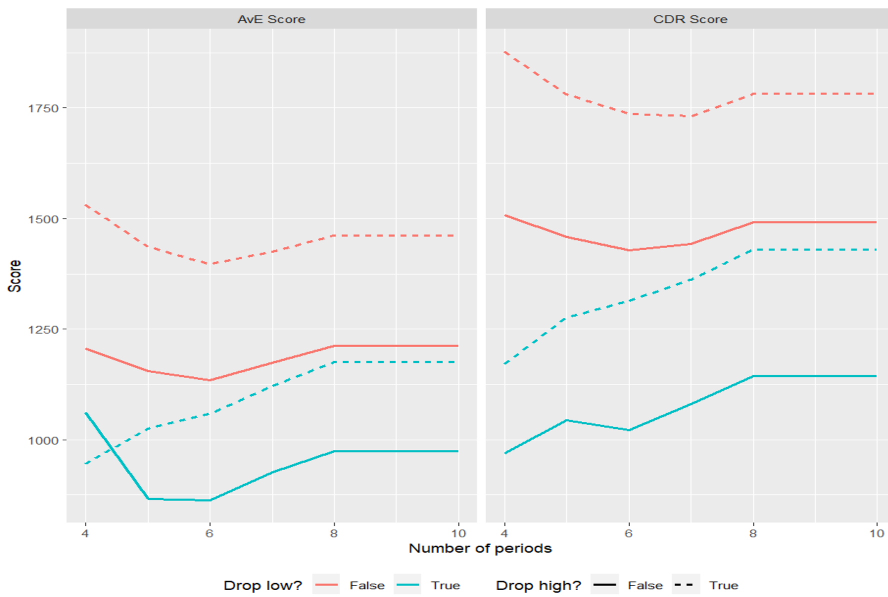

Figure 8: The score for both the metrics - AvE and CDR - when calculating the LDF based on different number of periods. The left plot illustrates the score for AvE and the right for CDR. Moreover, the dashed line corresponds to dropping the highest individual development factor, i.e. $\hat{f}_{i,j}$ for all development periods whilst the filled line is when we are not. Finally, the red line illustrates when dropping the lowest development factor and the blue when we are not. For example, the solid blue line on the left plot illustrates the AvE_{score} when dropping the lowest individual development factor in all development years. The score is also dependent on the number of periods (x-axis) where we for the same line see that the lowest score is when the number of periods equals six, which agrees to the results in Table 5.

4.2 The Bornhuetter–Ferguson Method

From previous sections we drew attention to the characteristic features of the BF-method. One of these was that an ultimate loss ratio was necessary (see equation (7)). The goal here is to use the machine learning approach with corresponding metrics and vary the ultimate loss ratio with a suitable choice. Apart from the ultimate loss ratio, we will vary the number of accident years to use which the loss development factors $\hat{f}_j$ will be calculated on together with the attempt to drop the highest and lowest individual development factor in all development periods respectively.

Recall from Section 3.1.2 where we simulated the corresponding premium to be used when calculating the ultimate claims using the basis of a 60% ultimate loss

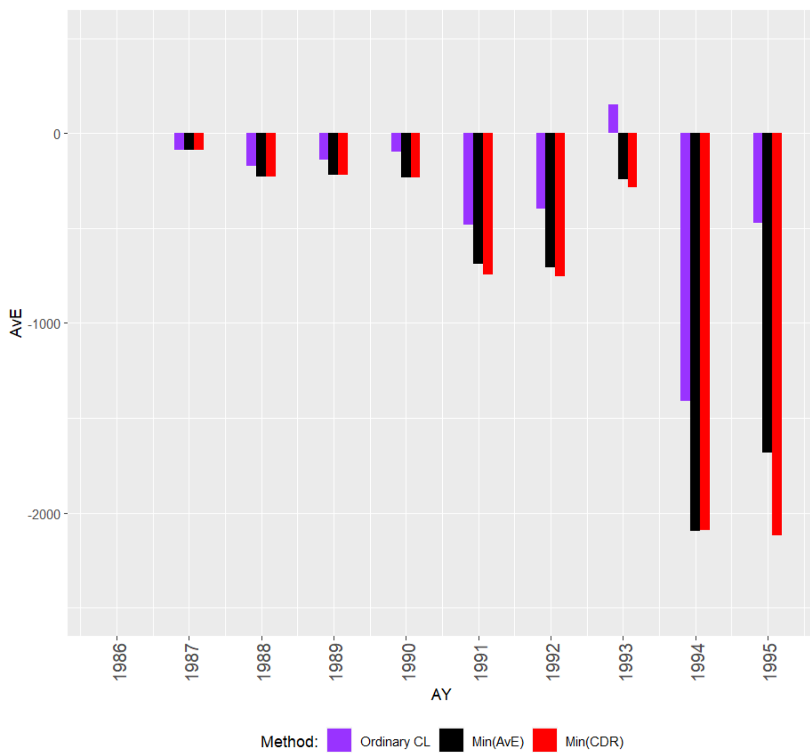

Figure 9: AvE for different metrics per accident year.

Method	**Total IBNR**	**Difference (%)**
Actual	101 071	-
Ordinary CL	104 177	3.07%
min(AvE)	107 512	6.12%
min(CDR)	107 252	6.69%

Table 6: Comparison between the estimated IBNR for our methods against actual total IBNR.

ratio. Because of this, we will vary the ultimate loss ratio, just as [1], between a 50% and 70% interval using step length 1%. The same number of periods will be used as for the chain ladder method above, and thus Table 7 illustrates

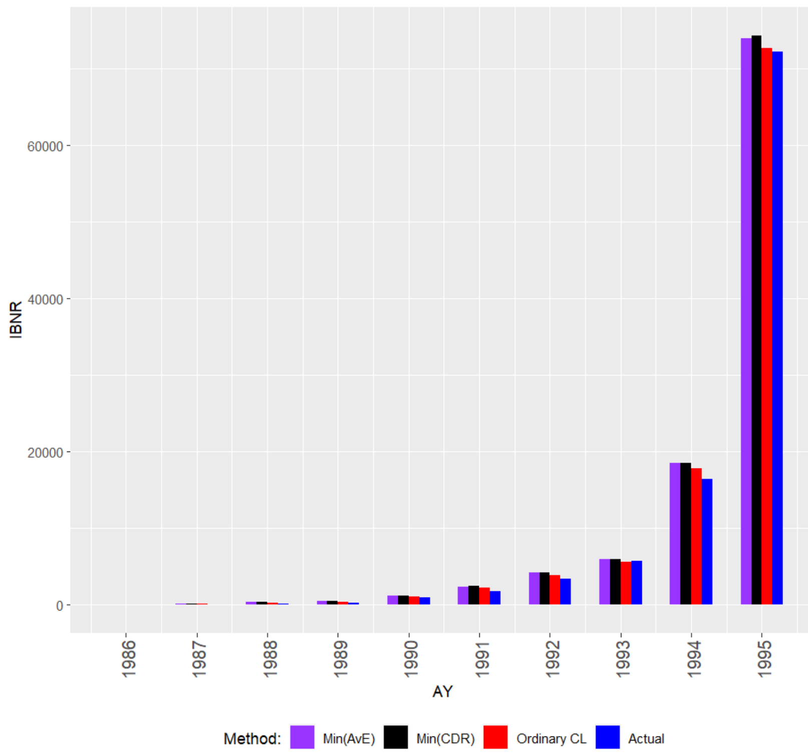

Figure 10: The IBNR-reserve when calculating the ultimo using the different models (min(CDR), min(AvE) and CL) as well as the actual reserve per accident year.

the possible combinations of parameters to search from. Note that the interval for choice of (apriori/expected) ultimate loss ratio and steps of course could be larger as well as smaller. This will however impact how many unique sets of search that will be done. Indeed, merely using the specified search grid as in Table 7 gives us $2 \cdot 2 \cdot 7 \cdot 21 = 588$ different combinations.

For the 588 different combinations, the parameters which minimize the two different performance metrics is shown in Table 8. Firstly, the ordinary BF-method is evidently formed using all the periods and 0.6 as apriori since that is our initial guess. The method also seems to use 5 respectively 4 most recent accident year for minimizing the performance metrics, where we also note that 4 years is the lowest possible amount. Also, the apriori parameter for the performance metrics is suggested to be the same, namely 0.58. Whether to drop the development factors is however different. The actual versus expected score seem to be lowest when dropping both the highest and lowest individual development factor in all development years. For the claims development result, the

Parameter	Possibilities	Description
Drop $\max(\hat{f}_{i,j})$?	[True, False]	Vary the possibility to ignore the highest (individual) development factor in all development years.
Drop $\min(\hat{f}_{i,j})$?	[True, False]	Vary the possibility to ignore the lowest (individual) development factor in all development years.
Number of periods	$I \in [4, 10]$	How many accident years that should be used when calculating the loss development factor for each development year.
Apriori	$\alpha \in \{0.5, \cdots, 0.7\}$	Testing different ultimate loss ratio for the BF-method.

Table 7: Parameter space for BF-method. We vary the number of periods and whether to drop the highest and/or lowest individual development factor for all development periods. Here, we will also vary the ultimate loss ratio parameter - also known as ”apriori”.

suggestion is to only drop the lowest one. This could be alarming since we from previous method saw that we overestimated the reserves and will perhaps see that once more. Further, Figure A.1 and A.2 (Appendix A) illustrates different scores and accident years for both performance metrics. Even though the lines looks somewhat ”lumpish”, you can still distinguish that the apriori parameter affects the scores significantly and that the number of periods does not seem to influence the score that much.

Just as were done in Section 4.1, using the Chain-Ladder method, we will compare the actual versus expected claims for different accident years to get an idea of how well the models perform. This is illustrated in Figure 11 and just as for the previous section, we seem to overestimate the ultimate claim. What is interesting here though is that both the minimized performance metric seem to perform much better than the ordinary BF-method and not only marginally. Using the parameters for when minimizing AvE_{score} is estimating the ultimate claim fairly well, especially for accident year 1993 which almost is on spot on correct claim amount. Comparing the nominal values in Figure 9 and 11 exhibits that the chain ladder method still seem to be closer to the actual value. The machine learning approach seems however to perform better in the case where we use the BF-method, especially for this short tailed business which is also supported when looking at Table 9 which illustrates the total IBNR for both the different methods but also the actual.

Parameter	**Ordinary CL**	$\mathbf{min}\left(\mathbf{AvE^{k}_{score}}\right)$	$\mathbf{min}\left(\mathbf{CDR^{k}_{score}}\right)$
Drop max($\hat{f}_{i,j}$)?	False	False	False
Drop min($\hat{f}_{i,j}$)?	False	False	True
Number of periods	10	5	4
Apriori	0.6	0.58	0.58

Table 8: The choice of parameters which minimizes the respective performance metric for Bornhuetter-Ferguson method.

Method	**Total IBNR**	**Difference (%)**
Actual	101 071	-
Ordinary BF	109 500	8.34%
min(AvE)	106 136	5.01%
min(CDR)	108 453	7.30%

Table 9: Comparison between the estimated IBNR for our methods against actual total IBNR and the ordinary BF-method.

5 Case study 2: Long tailed business

The second case study will as previously suggested use the data set which, if we recall from Section 3, reflect the long-tailed business that consists of 15 accident and development years respectively. We also saw that the claims were fully developed significantly later than in the first case study, approximately after 10 to 15 years. Just as before, we will now introduce the actual years and will look at the accident years 1986 - 2000. We can however not look at the individual development factors here like in Figure 7 since we have 50 different reserve

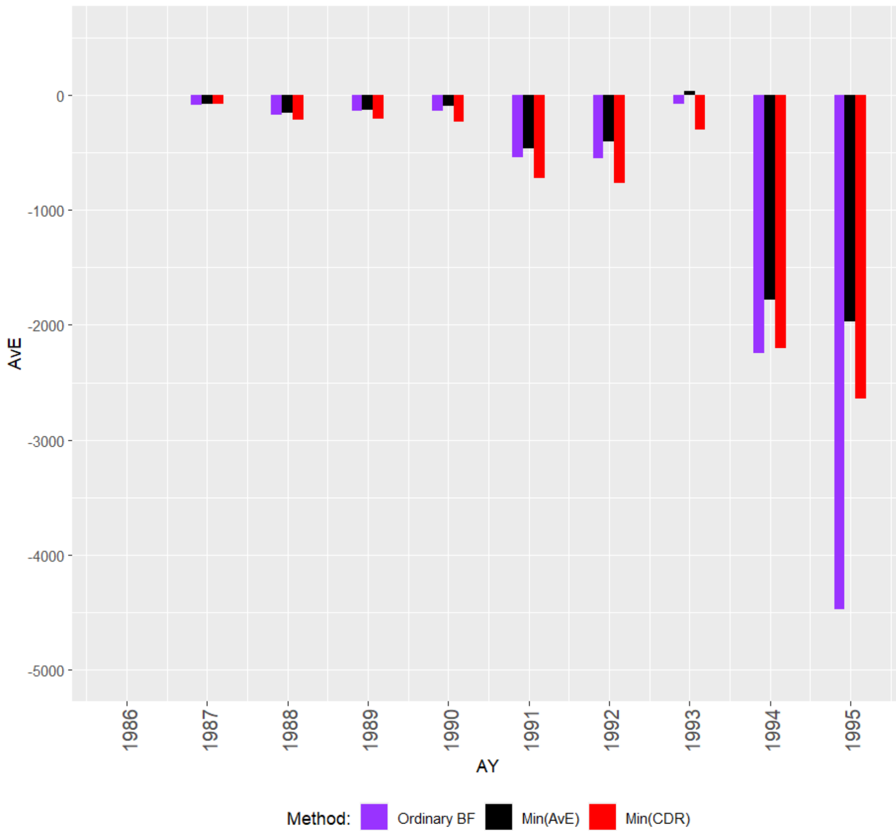

Figure 11: The actual ultimate claim minus the expected ultimo for the different variations of the BF-method.

triangles. One could examine some kind of average/median for the different IDFs, but that is redundant. We will during this case study only implement the machine learning approach with the chain ladder method.

The next thing that needs to be decided is the initial triangle and consequently the training set. By the same reasoning as before, the choice is based on using as much training data as possible without starting with a too small triangle. Referring back to Figure 2, we will choose 1986 - 1990 as initial triangle. This evidently implies that 1991 - 2000 will be used as our training set. The illustrations of the different results will be represented differently than the last case study since we have multiple reserving triangles.

5.1 The Chain Ladder Method

The three different variations of parameters will once again be varied together with the machine learning approach. To repeat, we will partly vary the number of accident year which the loss development factors $\hat{f}_j$ is calculated on. But also, we will also examine what happens when we drop the highest and lowest individual loss development factors in all development years. The number of periods is semi-arbitrarily chosen, but just as before, we will at least include all the accident years to see which score the ordinary Mack's CL-method perform in the metric scores. To begin with, Table 10 illustrates the search grid which in total consists of $2 \cdot 2 \cdot 10 = 40$ different combinations.

Parameter	**Possibilities**	**Description**
Drop max($\hat{f}_{i,j}$)?	[True, False]	Vary the possibility to ignore the highest (individual) development factor in all development years.
Drop min($\hat{f}_{i,j}$)?	[True, False]	Vary the possibility to ignore the lowest (individual) development factor in all development years.
Number of periods	$I \in [6, 15]$	How many accident years that should be used when calculating the loss development factor for each development year.

Table 10: Parameter space for CL-method in the second case study. We vary the number of periods and whether to drop the highest and/or lowest individual development factor for all development periods.

Unlike the first case study, it is not convenient to illustrate the parameters that minimize the metrics. Instead, we show that different combinations of the search grid from Table 10 arise only to get an overview. The different combinations that minimize the AvE_{score} and CDR_{score} respectively are illustrated in Figure 12 and 13, where we see that the algorithm suggest different models to estimate the IBNR reserve for the different simulated data sets. Even though it is indistinguishable to pair together respectively three parameters to each iteration, we still can see that the outcome varies.

We continue with the more interesting bits, namely the results, and it is not straight forward how this should be illustrated. We will nevertheless, like in the first case study, illustrate the difference between the actual ultimate claim minus the expected using the three different methods. Note, however, that we instead compare the difference of the summed ultimo. We will illustrate this

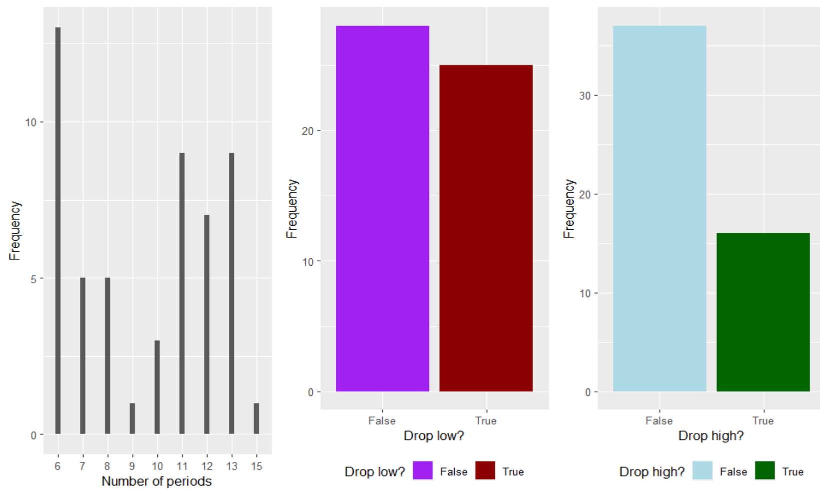

Figure 12: The different outcome of the search grid in Table 10 minimizing the AvE score. The left plot demonstrates the suggested optimal number of periods, the middle bar plot whether to drop the lowest individual development factor in all development years. The right illustration shows whether to drop the highest IDFs.

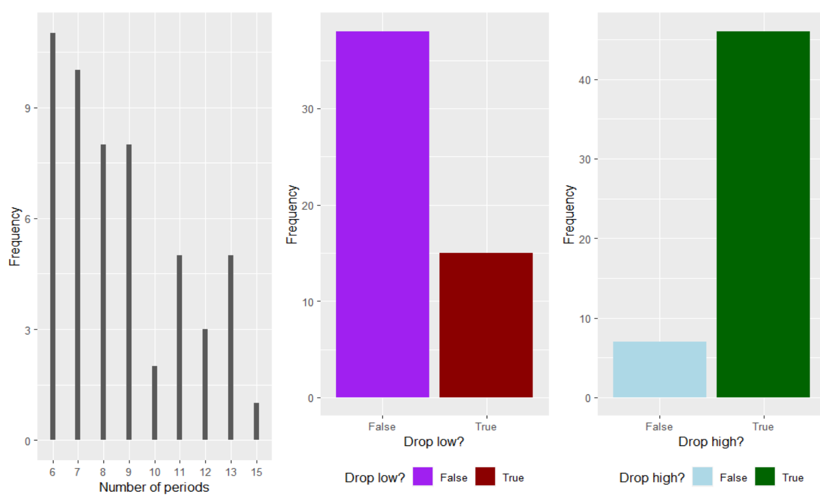

Figure 13: The different outcome of the search grid in Table 10 minimizing the CDR score. The left plot demonstrates the suggested optimal number of periods, the middle bar plot whether to drop the lowest individual development factor in all development years. The right illustration shows whether to drop the highest IDFs.

using two different figures which otherwise will look scribbled. Starting of with when minimizing the AvE metric, the result is shown in Figure 14. The difference between ultimate claims for the two methods do not differs that much. A larger deviation is around iteration 22, where both methods overestimate the reserve. Minimizing the AvE_{score} sometimes performs better than the ordinary chain ladder method and vice versa. Let us continue examine how well minimizing CDR_{score} estimates the ultimo. This is seen in Figure 15 where we see that the machine learning approach performs significantly worse than the ordinary CL.

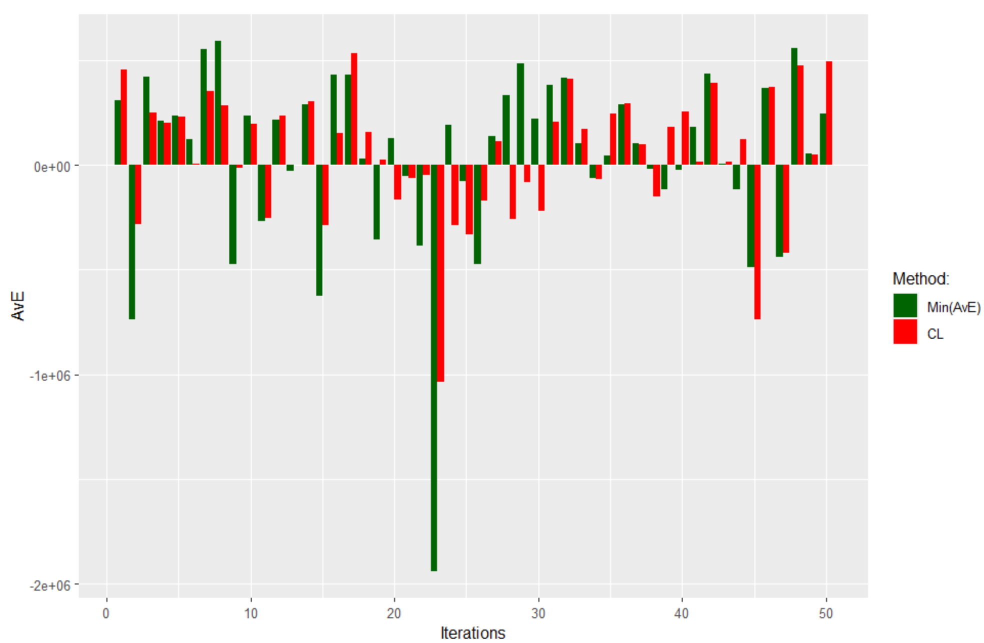

Figure 14: Actual ultimate payment minus the expected ultimo for both ordinary CL-method and when minimizing the AvE_{score}.

We will further look at the fraction between the total actual IBNR and the estimated one. This corresponds to the comparison we did in the first case study, see e.g. Table 6. The comparison of the total IBNR for the different iterations is illustrated in Figure 16. The interpretation is, the closer to 0 and more narrow the density plots are, the more accurate are the estimations of the total reserve. From this, we can first and foremost see that minimizing the CDR-metric confirms worse estimation of the IBNR compared to the chain ladder method. The calculations seems to underestimate the reserve, sometimes up to 70% less than the true. This is however not in alignment with what Balona et al. suggests. Indeed, as stated before, since the CDR keeps the reserve stable over time, the hypothesis were that determining the IBNR-reserve when minimizing the

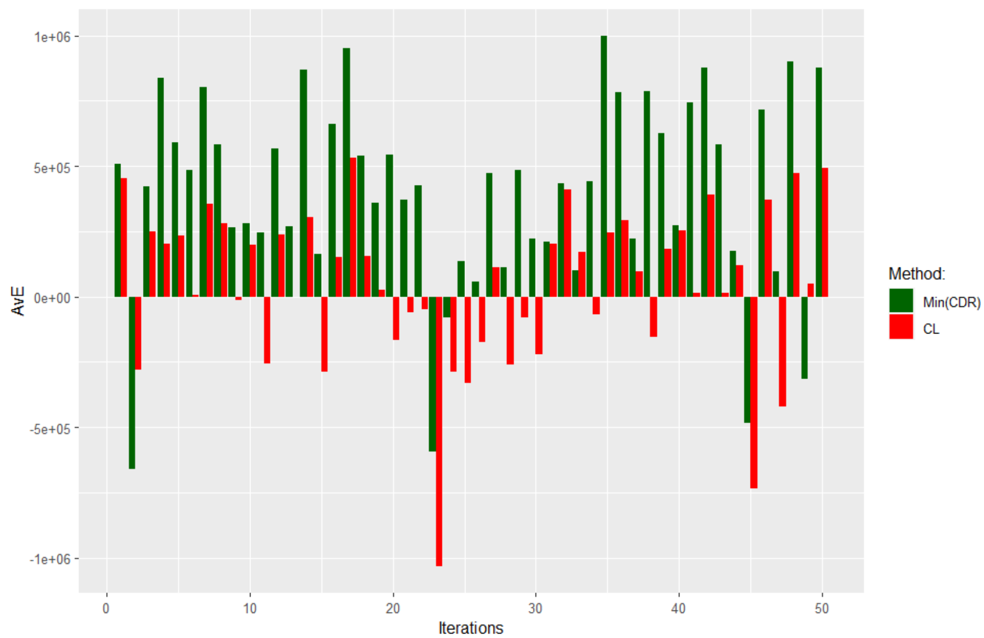

Figure 15: Actual ultimate payment minus the expected ultimo for both ordinary CL-method and when minimizing the CDR_{score}.

CDR_{score} should perform better.

Furthermore, looking at the second performance metric which is illustrated in the left density plot in Figure 16, we see that it more or less calculates the IBNR reserve just as good as the CL-approach. This was already implied from previous figures when illustrating the difference in ultimate claims calculations. A valid question from here is; when are the ML-approach performing better? In other words, can we see any pattern which indicates a better estimate in the ultimo for the AvE-metrics rather than the ordinary Mack's chain ladder method? If we revisit Figure 14 and focus on the reserve triangles where the ML-approach performs better in the sense that the absolute value of actual ultimate claim minus estimated ultimo when minimizing the AvE_{score} is less than the corresponding calculation for the customary CL modelling, we can directly compare how the data looked. We can therefore handpick those iterations, and illustrate how the loss development factors looks using all accident years. This is shown in Figure 17 together with the average LDF for all iterations shown in the blue line. We have also included the corresponding 95% confident interval which is shaded grey. From here we see that the iterations where the ML-approach performs better also seem to be when the data is somewhat "chaotic". Also, the LDFs in Figure 17 seem to be those that contribute to both the higher and lower "bounds" in Figure 6. Thus, the bottom line is that when the data seems changing, the ML-approach seems to handle the modelling better.

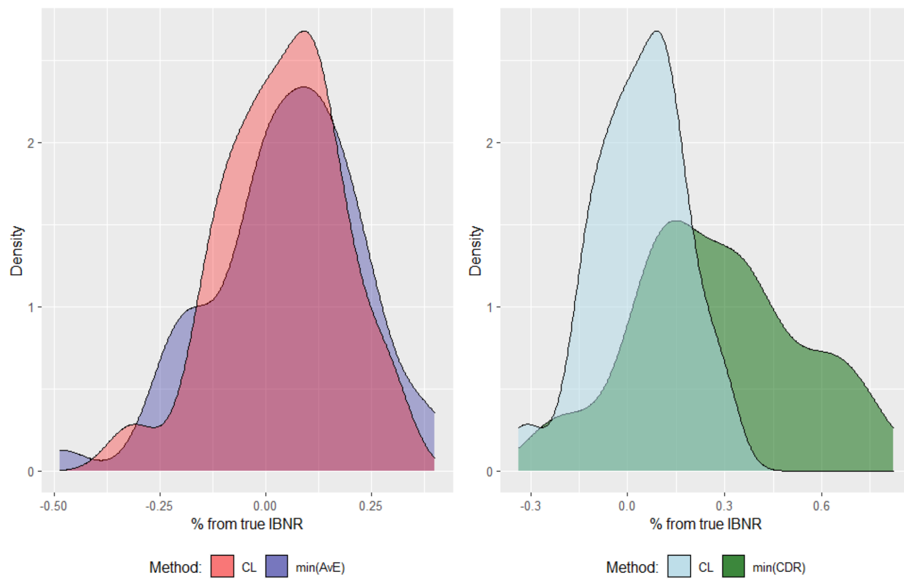

Figure 16: The fraction of the total actual IBNR and the estimated using ordinary CL and the ML-approach for both the metrics.

6 Summary

The goal with this thesis was to implement the machine learning approach that Caesar Balona and Ronald Richman introduced in their paper [1] for determining the IBNR-reserve. To our help, data were retrieved from the reinsurance company SiriusPoint. We established that when working with a single dataset, the outcome of a method may be influenced by factors such as luck or chance, making it difficult to interpret its true effectiveness. Because of this, we instead simulated new data, but where the insurance company information helped us determining parameters, especially the tail properties - which were essential for this paper. When implementing the machine learning approach, we used two of the most common reserving methods available as reference, namely Bornhuetter–Ferguson method and Mack's chain ladder method. In addition to the usual features found in machine learning, such as diving data into training and validation set, we used two performance metrics as score functions in order to evaluate the models. These metrics is called actual versus expected (AvE) and claims development result (CDR). The difference is that CDR is like a less variable version of AvE, where we penalize the difference in reserve estimates between calendar years. Finally, the algorithm that calculate the scoring procedure for each of the metrics is discussed in Section 2.3.2, where we also saw

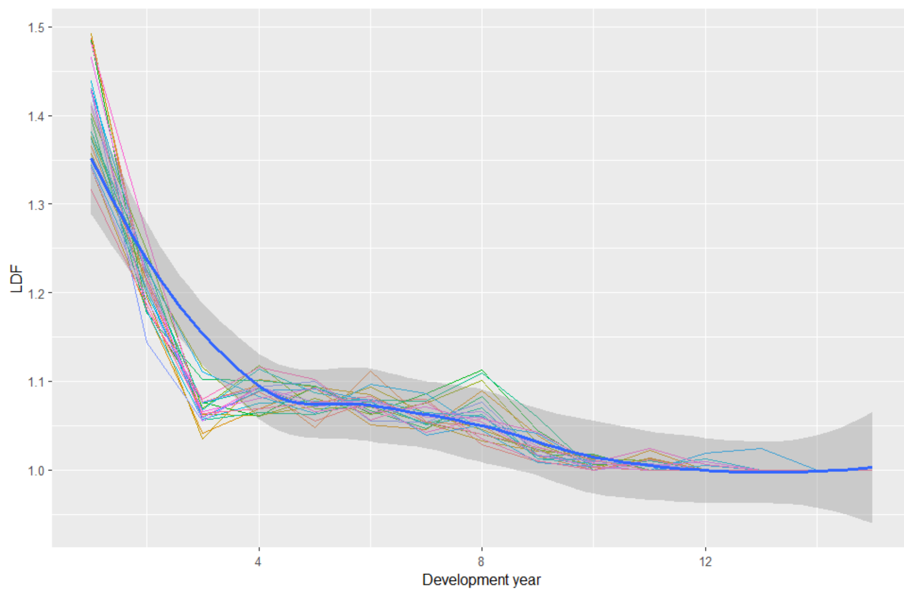

Figure 17: LDFs using ordinary Mack's chain ladder when the machine learning approach estimates the IBNR reserve better than the customary techniques. The blue line corresponds to all LDFs shown in Figure 6 with corresponding 95% confidence interval.

that the scores are weighted on the incremental claims. We then however chose the model which minimize each performance metric.

In order to draw conclusions, we used two case studies to our help. The first case study was based on a (one) simulated reserving triangle which had so called short-tailed property, i.e. a class/business which takes a short number until the full amount is payed for all accident years. This data is based on a run-off triangle using the accident years 1986 - 1995 (10 years) and we saw that it were fully developed after around four years. Moreover, to be able to compare the customary reserving method, we introduced some variations when determining the ultimate claims. These variations were:

- The number of accident years that should be used when calculating the loss development factor for each development year.
- The possibility to ignore the highest individual development factor in all development years.
- The possibility to ignore the lowest individual development factor in all development years.
- The ultimate loss ratio.

The first three parameters were varied when implementing our machine learning algorithm with the chain ladder method, and the latter variation was added when we used the BF-method. Using the CL-method, this resulted in a search grid of 28 different combinations for each of the metrics, and the parameters which minimized the AvE_{score} and CDR_{score} is reflected in Table 5. For the BF-method, we received a search space of 588 different combinations which resulted in the suggested parameters illustrated in Table 8. We thereafter calculated the ultimate claims for each model, and consequently the IBNR-reserve, which were compared to the corresponding ultimo and reserve using the ordinary reserving methods. This was achieved by partly subtracting the actual ultimo with the ultimate claim estimated for each model, but also by examining the difference for the total IBNR. For the Chain-Ladder method, we saw that the ML-approach did not perform better although the difference was marginal. However, for the BF-method, the new method seemed to estimate the reserves more accurately.

We approached the second case study a bit differently. Instead of merely simulating one reserving triangle, we repeated the simulation 50 times. This eliminates the risk of luck or chance when modelling the data and we can draw more conclusions. Not only that, in this case we also reduced the number of incurred claims per accident year to make sure that the analysed data is somewhat random. Moreover, in contrast to the first case study, we used insurance experience properties which were long-tailed that also had 15 accident and development years (accident years 1986 - 2000). We also ignored modeling according to the BF-method but focused on the CL-method.

Similar to the first case study we set up a search grid that entailed in 40 different combinations and conclusively spitted out 50 suggestions that minimized the respective performance metrics, one for each reserving triangle. We could conclude a couple of things here. Firstly, using the same comparison methods for the IBNR-reserve as before, minimizing the CDR_{score} performed worse than the customary chain-ladder method. We could however see that minimizing AvE_{score} sometimes estimated the IBNR-reserve better. This entailed in us examining when this was the case. A quick examination showed that when the reserving triangle was somewhat "unstable", the machine learning approach seemed to encapsulate those situations well.

7 Discussion

There are a couple of things that needs to be highlighted. Firstly, in the first case study when we only had one reserving triangle with each accident year containing a great number of claims, we noticed that the chain ladder method performed better than minimizing the scores of both the metrics, even though it was marginal. The question that arises thus becomes: why does the chain ladder method perform better? This is actually not surprising since the customary

Mack's chain ladder method captures all information from the reserving triangle. And since we established that the triangle is "stable", the CL-method will probably always outperform. Using the same logic for the second case study, we would perhaps hoped for better results, especially since we tried to simulate multiple "unstable" reserving triangles. We saw that minimizing the AvE_{score} sometimes estimated the IBNR reserve better than the ordinary CL-approach which could be traced back to the iterations where the data was most unstable. An improvement for future references is therefore to create even more scarce reserving triangle and do the same type of modelling. More number of triangles would also be in order. However, it is alarming that when minimizing the CDR_{score}, the estimated IBNR was significantly worse than the actual reserve. During the reserving process, actuaries do not have fully developed claims like in this paper, but instead have to use their best knowledge to ensure that the reserve will be as close to reality as possible.

Although this thesis may not confirm that Balona & Richman's approach should be implemented in insurance companies today, their methodology and algorithm should not be neglected. There is nothing that says we are limited to using the algorithm in the way we have done so, as it is ultimately a machine learning model that uses a scoring function, which can clearly be used for other purposes. One example is to combine the CL method and BF method together with minimizing the CDR and AvE metrics in the same way as in this article. This way, we can not only see which method estimates the reserve best, but also, and perhaps more importantly, weigh the different reserve estimation methods. We have also limited ourselves to only two reserve estimation methods, with a focus on CL, but there is nothing stopping us from introducing more methods such as Generalized Cape Cod (GCC). However, one thing is certain, machine learning methods for reserving will undoubtedly become the future in the insurance industry, and it is only a matter of time before it becomes the new "ordinary" method.

References

[1] Balona, C. & Richman, R. (2021). *The Actuary and IBNR Techniques: A Machine Learning Approach.* Variance, November.

[2] Bornhuetter, R. L., & Ferguson, R. E. (1972). The Actuary and IBNR. *Proceedings of the Casualty Actuarial Society, 59*, 181–195. Retrieved from https://www.casact.org/pubs/proceed/proceed72/72181.pdf

[3] Gabrielli, A and Wüthrich, M. (2018). *An Individual Claims History Simulation Machine*. Risks, 6(2):29. https://doi.org/10.3390/risks6020029

[4] Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, Ny Springer New York.

[5] Käll, F. (2022). *Gradient Boosted Trees Applied to Chain- Ladder Reserving*. Stockholms Universitet. Master Thesis in Actuarial Mathematics.

[6] Lindholm, M. Lindskog, F. & Wahl, F. (2019). *Estimation of conditional mean squared error of prediction for claims reserving*. Annals of Actuarial Science (2020), 14, pp. 93–128 doi:10.1017/S174849951900006X.

[7] Lindskog, F. (2022). *Risk Models, Claims Reserving and Solvency in Non-Life Insurance.*

[8] Mack, T. (1993). *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates.* ASTIN Bulletin, 23(2), 213-225. doi:10.2143/AST.23.2.2005092.

[9] Miranda, M. D. M. Nielsen, B. Perch, J. P. & Verrall, R. (2011). *Cash flow simulation for a model of outstanind liabilites based on claim amounts and claim numbers*. Astin Bulletin 41(1), 107-129. doi: 10.2143/AST.41.1.2084388.

[10] Radtke, M., Schmidt, D. S & Schnaus, A. (2016). *Handbook on Loss Reserving*. European Actuarial Academy. Springer International Publishing Switzerland 2012, 2016.

[11] Verrall, R., Nielsen, J.P. & Jessen, A.H. (2010). *Prediction of RBNS and IBNR claims using claim amounts and claim counts.* ASTIN Bulletin, 40(2), 871–887.

[12] Wüthrich, M. (2018). Neural networks applied to Chain–Ladder reserving. European Actuarial Journal, 8(2), 407-436. http://dx.doi.org/10.2139/ssrn.2966126

A Appendix

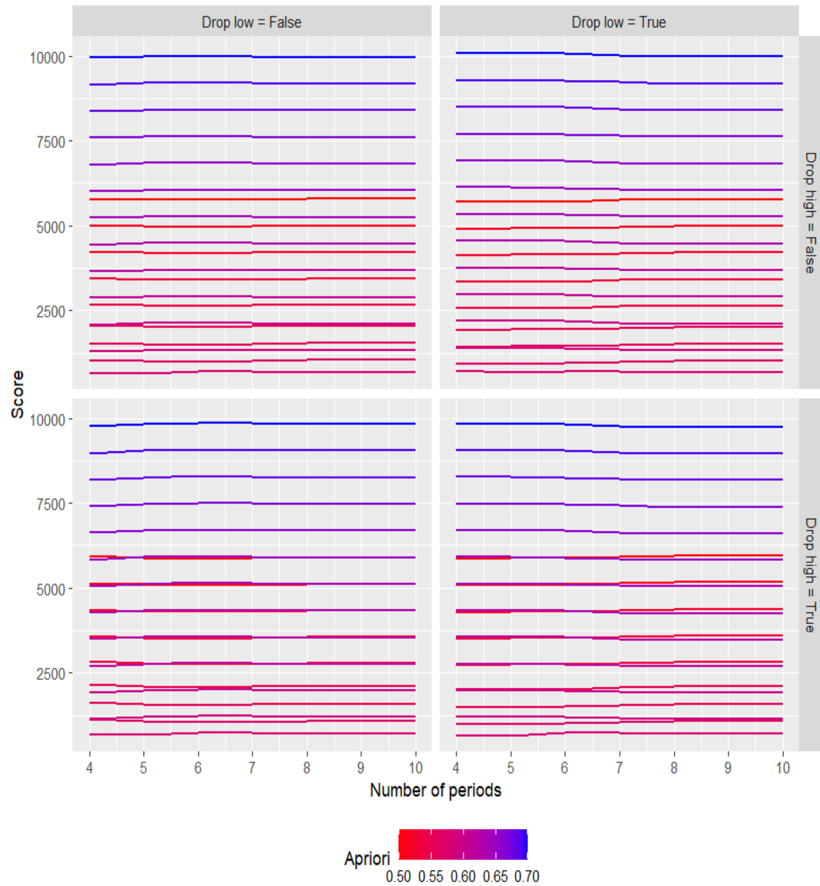

Figure A.1: Scores for actual versus expected performance metric using the search space for the BF-method which was illustrates in Table 7. The four different plots illustrates the four combinations of whether to ignore (drop) the highest and lowest IDF in all development years or not. For example, the upper right plot illustrates the score when including the highest IDF but dropping the lowest. Similarly, the colors correspond to variations of the apriori parameter, where a redder shade reflects a lower ultimate loss ratio and blue color a higher.

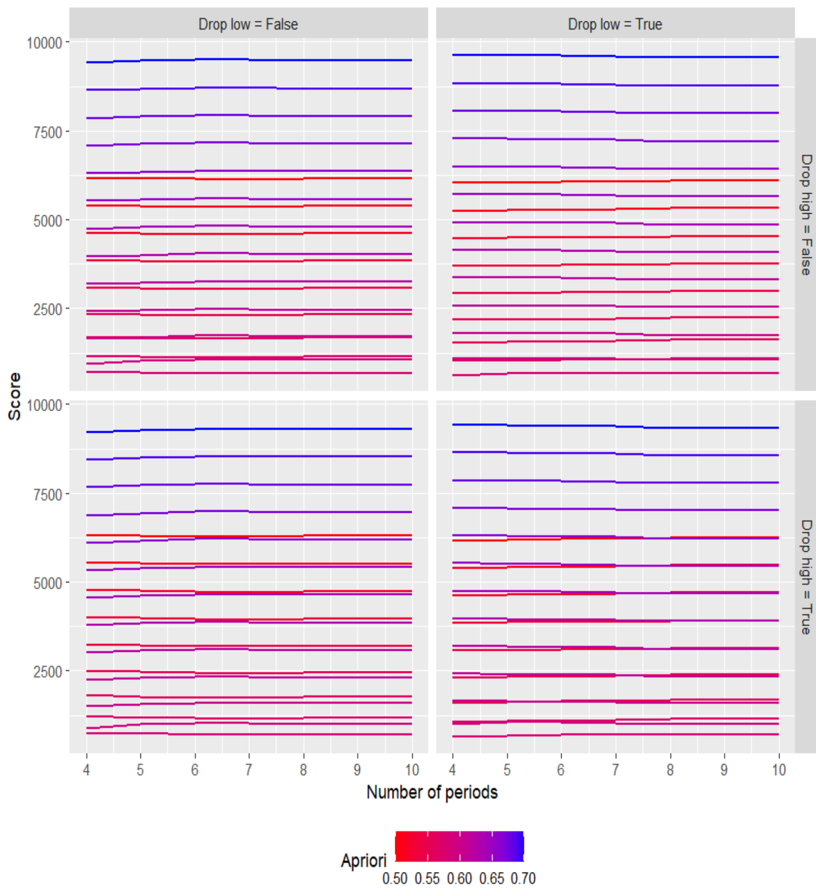

Figure A.2: Scores for claim development result performance metric using the search space for the BF-method which was illustrates in Table 7. The four different plots illustrates the four combinations of whether to ignore (drop) the highest and lowest IDF in all development years or not. For example, the upper right plot illustrates the score when including the highest IDF but dropping the lowest. Similarly, the colors correspond to variations of the apriori parameter, where a redder shade reflects a lower ultimate loss ratio and blue color a higher.