

The Actuary and IBNR Techniques: A Machine Learning Approach

Caesar Balona* Ronald Richman†

14 August 2020

Abstract

Actuarial reserving techniques have evolved from the application of algorithms, like the chain-ladder method, to stochastic models of claims development, and, more recently, have been enhanced by the application of machine learning techniques. Despite this proliferation of theory and techniques, there is relatively little guidance on which reserving techniques should be applied and when. In this paper, we revisit traditional reserving techniques within the framework of supervised learning to select optimal reserving models. We show that the use of optimal techniques can lead to more accurate reserves and investigate the circumstances under which different scoring metrics should be used.

1 Introduction

Since Bornhuetter and Ferguson (1972) encouraged actuaries to pay attention to the topic of Incurred But Not Reported (IBNR) reserves, a vast literature on loss reserving techniques has emerged, see, for example Schmidt (2017) and Wüthrich and Merz (2012). Several main strands of this literature can be identified: improved reserving techniques for calculating best estimates, reformulating these techniques statistically to calculate measures of uncertainty, and, most recently, the application of machine learning techniques to the problem of IBNR reserving.

Many new IBNR reserving techniques have been developed since the development of the Chain Ladder (CL) and Bornhuetter-Ferguson (BF) techniques, with notable examples that are often used in common practice being the Cape Cod (CC) method from Bühlmann and Straub (1983) and its generalization due to Gluck (1997), and the Incremental Loss Ratio (ILR)

*QED Actuaries & Consultants, caesar.balona@qedact.com, caesar.balona@gmail.com

†QED Actuaries & Consultants, University of the Witwatersrand, ronald.richman@qedact.com, ronaldrichman@gmail.com

method, due to Mack (2002). These contributions provide new ways of determining IBNR reserves. Moreover, as reserving practice has developed, many smaller adjustments to the techniques have been suggested, for example, the summary measure to use when calculating chain-ladder factors, which may be estimated using weighted or simple averages, or medians. These smaller adjustments may sometimes not have a theoretical justification but are, nonetheless, often applied in practice.

Furthermore, reserving techniques have been enhanced through reformulation as statistical models, for example, two different models underlying the chain-ladder technique are given in the seminal contributions of Mack (1993) and Renshaw and Verrall (1998). Later contributions in this vein have extended the application of statistical techniques to the specific needs of insurance regulation, with a key result being the one-year viewpoint needed for Solvency II that is provided in Merz and Wüthrich (2008). These contributions allow for estimating the uncertainty contained in an IBNR analysis, for the purpose of quantifying capital requirements or producing risk adjusted cash-flows (see also England, Verrall, and Wüthrich (2019)).

Finally, recent developments have applied methodologies and techniques from the field of machine learning to the problem of IBNR reserving. Some of these applications have extended traditional techniques, see, for example, Wüthrich (2018) who develops a neural network model to predict loss development factors in the context of the chain-ladder method or Gabrielli, Richman, and Wüthrich (2020) who use a neural network to enhance the predictions of an Over Dispersed Poisson (ODP) generalized linear model. Alternatively, other novel approaches have departed from familiar loss reserving techniques, for example, Kuo (2019) applies recurrent neural networks to triangles of incurred and paid loss ratios to derive estimates of IBNR. These contributions often produce more accurate results than traditional techniques, but at the cost of applying relatively more complex models.

Despite the proliferation of techniques for estimating IBNR reserves, there is not much guidance on choosing between the various techniques. Instead, actuaries in practice often rely on heuristics to guide the choice of reserving model, for example using the BF technique to reserve for less developed accident years and the CL technique for more developed accident years. Since the choice of technique and the manner of its application is not always done systematically, it can sometimes be difficult to justify IBNR reserving figures to auditors and regulators. Also, claims experience contained in loss triangles often departs from the model examples found in the literature, for example, loss development factors may display trends relating to changing claims settlement delays or shifting loss recognition practices and outliers relating to large claims may be present in the triangles. Also in this case, heuristics are used to deal with these problems – for example, only the most recent diagonals in a triangle may be used to calculate loss devel-

opment factors and outliers may be excluded – but a systematic method for performing the analysis required to derive IBNR reserves on these types of triangles seems not yet to have been addressed in detail in the literature.

Furthermore, the development of Solvency II has complicated the issue of IBNR reserving further; Solvency II requires that reserves are set at the “best estimate” of the loss distribution, see, for example, European Union (2009) that requires that “the best estimate shall correspond to the probability-weighted average of future cash flows, taking account of the time value of money”. Without a systematic method of identifying the techniques that produce a best estimate reserve, actuaries are not only forced to rely on their professional judgement that a particular technique produces the best estimate, but must convince regulators of the validity of this judgement as well.

Finally, the context in which many IBNR techniques are presented in the literature is often restricted to a reserving analysis performed at a point in time to arrive at loss reserves, especially in the earlier IBNR literature. This is called the “static” perspective on reserving (Wüthrich & Merz, 2015). On the other hand, actuaries in practice must re-estimate loss reserves on a frequent basis (often quarterly), applying IBNR techniques to loss triangles that are augmented with new data and then reporting on the results. The accuracy of the actuary’s previous reserving exercise is often evaluated using an Actual versus Expected analysis (Bruce et al., 2015), which helps to guide the setting of reserves on the basis of emerging experience, but changes in loss reserves must be justified to the Board and senior management of the insurance operation in which the actuary is employed. This “dynamic” perspective of reserving has, more recently, begun to be addressed, see, for example, Wüthrich and Merz (2015), who provides mathematical definitions for the claims development process over time and shows how these results can be used for model selection and back testing. From a reserve risk perspective, Merz and Wüthrich (2014) shows how the total prediction error of IBNR in the the CL method can be split into the contribution of the reserve risk for each calendar year. Here, we focus on the dynamic perspective and discuss how reserving models can be selected and parametrized to produce good forecasting performance on unseen claims development experience.

In this paper we present a framework through which the actuary can analyse the properties of traditional reserving methods from the dynamic perspective of reserving, with the goal of assisting with model selection. We investigate how well the models selected using this framework perform for producing best estimate predictions of claims development. We do not depart from applying familiar techniques such as the CL and BF, but rather look to optimize the predictive accuracy of the reserves produced using these techniques. Since large reserving triangles are available to many actuaries working in companies with an extensive loss history, we propose that the actuary re-reserves consecutive calendar years, at each point adding a new

diagonal of experience, and then investigates the outcome of each previous reserving exercise based on the actual experience contained in the triangle. To measure the effect of different approaches used to re-reserve these prior years, we define an objective function to score the approaches using metrics familiar to reserving actuaries. The optimal set of techniques, and the adjustments to these techniques, is then selected as those minimizing the objective function.

The rest of the paper is organised as follows. Section 2 provides a notation for the framework proposed in Section 3. Two applications of the method are discussed in Sections 4 and 5. Finally, Section 6 concludes the paper with a discussion and avenues for future research.

2 Loss Triangle Notation and Methods

In this section, we provide a short mathematical definition of the claims reserving problem and common methods that are applied in practice. For a complete treatment of this topic, we refer the reader to Wüthrich and Merz (2012).

2.1 Loss Triangle Notation

IBNR reserving is usually performed on claims data aggregated into a triangle shaped array (although other configurations such as trapezoids are sometimes seen in practice, we do not address these here). Suppose that claims data have been observed at the end of every calendar year for the past K years, i.e. $k \in [0 \dots K]$. Claims (whether claim payments or changes in estimates of the payments for a particular claim), which are denoted as $X_{i,j}$, are allocated to cells of the triangle according to the accident year, in which a loss occurred, $i \in [0 \dots I]$, and the development year, $j \in [0 \dots J]$, in which the payments were made or the estimates changed. Note that one could also use quarters, months, or other time periods without loss of generality and that, depending on the purpose of the analysis, claims could also be allocated by reporting year or underwriting year, however for simplicity, we restrict our discussion in this section to accident years. In these definitions, we work with triangles in which the number of accident years are equal to the number of development years, i.e. $I = J = K$, and no development is expected after development period J (i.e. no tail factor is needed), but the method we propose is not limited to this case. Often, reserving methods are applied to the cumulative values of claims

$$C_{i,j} = \sum_{l=0}^j X_{i,l},$$

which can also be arranged in a triangular array. We define a reserving triangle observed in year K , Δ^K , as

$$\Delta^K = \{C_{i,j}; i + j \leq K\}.$$

The claims occurring in the most recent calendar year can be found on the largest diagonal of the triangle defined as the set

$$\{C_{i,j}; i + j = K\}.$$

To ensure that sufficient funds are retained to cover the claims liabilities of a company, liabilities equal to an estimate of all claims to be reported in the future are held. Since the final, or ultimate, amount of claims arising from each accident year i , $C_{i,J}$, is not known, IBNR techniques are applied to produce an estimate of these claims $\hat{C}_{i,J}^k$ in each calendar year k (in practice, this may be done more frequently than annually). Then, IBNR reserves for accident year i , R_{i,j^*}^k are calculated as $\hat{C}_{i,J}^k - C_{i,j^*}$, where $j^* = k - i$. This expression can be rearranged in terms of the forecast incremental claims such that $R_{i,j}^k = \sum_{l=j^*+1}^J \hat{X}_{i,l}$. Thus, the estimation of IBNR reserves depends on the estimate of ultimate losses.

Remark. *We have taken care to define estimates of ultimate claims and IBNR reserves as being at a point in time k , since the method proposed here tries to establish the accuracy of these estimates as more information is added to the triangle i.e. we are interested in the dynamic view of reserving.*

If a perfect forecast (in other words, incorporating knowledge of all of the future claims experience) of $C_{i,J}$ could be made by the actuary in a calendar period k , then there would be no fluctuation in the estimated ultimate claims over consecutive calendar periods. In reality, since perfect forecasts cannot be made, the actuary's forecasts of the ultimate claims may change as more information becomes available over subsequent calendar periods.

To this point, we have not defined any of the statistical properties of the estimates of claims to be reported in the future, $\hat{C}_{i,J}^k$. From some (currently less fashionable) actuarial perspectives, it might be considered optimal for an insurer to produce upwardly biased estimates i.e. for $E[\hat{C}_{i,J}^k - C_{i,J}] \geq 0$, thus producing reserve estimates that are expected to be more than sufficient to cover the actual claims. To make this fully rigorous, one needs to define the probability distribution of $C_{i,J}$ which is unknown, thus, in practice if conservative reserve estimates are being produced, these are evaluated against the reserve estimates produced by the IBNR reserving methods. Solvency II, IFRS 17 (IASB, 2020), and other similar regulations, require that the reserves be set as a best estimate. In the case of Solvency II, the best estimate is defined in the regulations on an assumption that, somehow, the future distribution of cashflows is known, see European Union (2009) which states that "the best estimate shall correspond to the probability-weighted

average of future cash flows, taking account of the time value of money". A more pragmatic definition of the best estimate should, at the minimum, include that at time k , the estimates of ultimate claims are expected to correspond to the actual claims that will be reported with minimal error, which could be quantified, for example, with the mean squared error by ensuring that $E[(\hat{C}_{i,J}^k - C_{i,J})]$ is as small as possible. The method that we describe later in this study can be used to provide some evidence that the selected reserving methods in fact produce estimates with minimal error on observed claims development and produce relatively low error on unobserved claim development.

2.2 Claims Development Result as an Optimization Objective

Differences between forecasts of the ultimate claims in consecutive periods give rise to the Claims Development Result (CDR) of Merz and Wüthrich (2008), which is a measure of the profit or loss incurred by a company due to the experience of past accident years. The CDR for accident year i in calendar year k is defined as

$$\begin{aligned}
 CDR_i^k &= C_{i,J}^k - C_{i,J}^{k-1} \\
 &= R_{i,j^*}^k + C_{i,j^*} - (R_{i,j^*-1}^{k-1} + C_{i,j^*-1}) \\
 &= \sum_{l=j^*}^J \hat{X}_{i,l}^k + C_{i,j^*} - \left(\sum_{l=j^*-1}^J \hat{X}_{i,l}^{k-1} + C_{i,j^*-1} \right) \quad (1) \\
 &= R_{i,j^*}^k - R_{i,j^*}^{k-1} + (X_{i,j^*} - \hat{X}_{i,j^*}^{k-1}) \\
 &= R_{i,j^*}^k - R_{i,j^*}^{k-1} + AvE_{i,j^*}^k,
 \end{aligned}$$

where AvE_{i,j^*}^k is the actual versus expected (AvE) result of the incremental claims on the next diagonal of the triangle. Thus, the CDR can be expressed as the sum of the difference between the sequential forecasts of the IBNR reserves for development period j^* and onwards plus the difference between the actual incremental claims experience in development period $j^* - 1$ and the expected claims experience. The latter part of the CDR, which is nothing more than the error of the forecast of the incremental claims made at time $k - 1$, is often inspected by actuaries in the form of an actual versus expected analysis, which is a diagnostic used to make changes to the IBNR reserves for the next calendar year. Changes made to the reserves on the basis of the CDR will reflect in the annual profit and loss statement of the insurance entity, and the key contribution of Merz and Wüthrich (2008) is to provide a distribution of the mean squared error of prediction (MSEP) of the CDR, which allows for the calculation of reserve risk capital under the one-year definition of Solvency II, see Ohlsson and Lauzenings (2009) for

more detail. Also, AvE_{i,j^*}^k forms the basis of the bootstrap methods proposed by England and Verrall (1999), which derives actual versus expected results from a comparison of a fitted CL model to an observed triangle, and then bootstraps these residuals.

At any calendar year k , the CDR for the subsequent year, CDR_i^{k+1} , is not yet known. However, with a suitably long claims history, the CDR that would have arisen in previous calendar years had a particular IBNR reserving method been followed can be calculated and can provide an indication to the actuary how suitable her reserving method is. In this paper, we use the CDR and AvE metrics assessed on historical data, in the manner just described, to assess the suitability of different applications of IBNR reserving techniques, and we select the optimal reserving method that minimizes the CDR in previous time periods. In other words, we use the CDR and AvE metrics as objective functions measuring the quality of a reserving methodology that we seek to minimize.

The CDR can be seen as composed of two components: first, the quality of the forecasts made is evaluated via the AvE metric, $AvE_{i,j^*}^k = (X_{i,j^*} - \hat{X}_{i,j^*}^{k-1})$ and second, the stability of the reserve estimates for future years, $R_{i,j^*}^k - R_{i,j^*+}^{k-1}$ is added. From an optimization perspective, it may be possible that some reserving techniques fit the observed data very well, i.e. minimize the AvE metric, but at the price of erratic and less accurate future reserving estimates. Thus, considering the CDR as an optimization objective, may maintain the stability of the reserve estimates and could help to regularize the models chosen.

Remark. *Other formulations of an objective function for optimization could be considered, based on the CDR. For example, more or less weight could be given to the stability of the reserving by adding a parameter α to the CDR*

$$CDR_i^{k,\alpha} = \alpha(R_{i,j^*}^k - R_{i,j^*+}^{k-1}) + AvE_{i,j^*}^k,$$

which could be optimized jointly with other hyperparameters. Alternatively, one could recognize that the AvE term in fact acts to minimize bias, and another term could be added to the objective function to penalize poor forecasts, for example the mean square error. We do not explore this further in this study.

In this study, we focus on minimizing the squared differences between the CDR and AvE metrics, and 0, which, according to the above, is the same as minimizing the difference between forecasts of ultimate claims made in consecutive years and the difference between actual and predicted claims in the next diagonal, respectively. In particular, based on a choice of reserving methodology, we calculate (note that we want to minimize the difference from zero and $(CDR_i^k - 0)^2$ reduces to what is shown as the numerator below):

$$CDR_{score} = \sqrt{\frac{\sum_{i=1}^I |X_{i,j^*}| (CDR_i^k)^2}{\sum_{i=1}^I |X_{i,j^*}|}}, \quad (2)$$

where we weight the score by the absolute value of the incurred claims in each year and analogously, we also calculate a score for the AvE metric, AvE_{score} . We refer to these scores as the root mean square error (RMSE) of the CDR, or the AvE, respectively.

When measuring the predictive performance of the reserving methodology, we base this on the actual claims development in several out of sample calendar years, as specified in each of the case studies below (which is the yellow area of Figure 1).

2.3 Methods

This section briefly reviews three of the most well-known IBNR reserving methods using triangles, that are used in the paper.

2.3.1 Chain Ladder

The Chain Ladder (CL) method produces forecasts for the unknown ultimate claims $C_{i,J}$ based on the key assumptions that future claims development will be in line with past claims development and that the current magnitude of the claims for an accident year $C_{i,j}$ can be used to predict the magnitude of future reported claims. More precisely, the key assumption of the CL method is that $E[C_{i,j+1}] = f_j C_{i,j}$, where the forecasts are made using so-called loss development factors (LDFs), f_j , multiplied by the current known claims amount. Since we consider several different methods in this study, we denote forecasts made using the CL method as $\hat{C}_{i,j+1}^{CL}$. The true LDFs, f_j , are not known and must be estimated from the available claims development data, thus, throughout, we work with estimated LDFs, \hat{f}_j . Forecasts of ultimate claims are made by $\hat{C}_{i,J}^{CL} = \hat{F}_{j,J} C_{i,j} = \prod_{k=j}^J \hat{f}_k C_{i,j}$.

The LDFs can be estimated as follows:

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

which shows that \hat{f}_j is a weighted average of the individual development factors $\hat{f}_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$.

Mack (1993) shows that under minimal assumptions, the estimated LDFs \hat{f}_j produced using Equation 3 are unbiased estimators of the true parameters f_j . In practice, actuaries often use variations on the CL method, in particular, by changing the manner in which the LDFs are calculated. For example, the weighted average in Equation 3 may be replaced with a simple (i.e. unweighted) average, or another summary statistic such as the median could be used. Alternatively, whereas the calculations in Equation 3 use all of the accident years in the triangle, in practice, actuaries often use data from only the most recent accident years to calculate \hat{f}_j , or exclude some of the factors, effectively re-weighting the calculation of \hat{f}_j with weights $w_{i,j}$, which are chosen subjectively, and are typically 1 for individual development factors to be included, and 0 for those to be excluded:

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j} \hat{f}_{i,j} w_{i,j}}{\sum_{i=1}^{I-j} C_{i,j} w_{i,j}} \quad (4)$$

A final option is to calculate accident year specific weights based on an exponential decay factor leading to individual development factors in closer proximity to accident year i contributing more weight.

2.3.2 Bornhuetter-Ferguson

The Bornhuetter-Ferguson (BF) method seeks to reduce the reliance on the current claims amount $C_{i,j}$ when projecting the ultimate claims, by assuming that future claims development will be proportional to the actuary's expectation of the ultimate loss, $\hat{C}_{i,J}^{BF} = C_{i,j} + (1 - \beta_j)E(C_{i,J}^k)$, where β_j is an estimate of the percentage of losses reported at development period j , and is usually taken as the inverse of $\hat{F}_{j,J}$, and $E(C_{i,J}^k)$ represents the expected ultimate losses for accident period i . If this expectation is set using the CL ultimate claims, $\hat{C}_{i,J}^{CL}$, then the BF method reduces to nothing more than CL method. Thus, to reduce the reliance on the current known claims $C_{i,j}$, an independent estimate of the ultimate claims is ordinarily used in the BF method, which is often expressed using an expected ultimate loss ratio, i.e. $\hat{C}_{i,J}^{BF} = C_{i,j} + (1 - \beta_j)\pi_i \widehat{ULR}_i^{k,BF}$, where π_i is the earned premium in accident period i and $\widehat{ULR}_i^{k,BF}$ is the actuary's expectation of the ultimate loss ratio at time k for accident year i , which has been derived independently of the claims experience to date.

In practice, these expected loss ratios may be set using input from pricing actuaries or underwriters, or on the basis of past experience from previous accident years, meaning that, the choice of expected ultimate loss ratio is subjective. In applying the BF method, in addition to needing to select

ultimate loss ratios, if the CL pattern is used as an estimate of β_j , then all of the subjective choices for applying the CL method also apply.

2.3.3 Cape Cod and Generalized Cape Cod

The Cape-Cod (CC) method of Bühlmann and Straub (1983) provides a way of inferring the ultimate losses from the loss data in an actuarially principled manner, by assigning greater credibility to accident years that are more developed. On the assumption that all accident years are expected to have the same ultimate loss ratio, the Cape Cod method calculates an estimate of the ultimate loss ratio:

$$\widehat{ULR}_i^{k,CC} = \frac{\sum_{n=1}^I C_{n,I-n+1}}{\sum_{n=1}^I \pi_n \beta_{I-n+1}} \quad (5)$$

This estimated ultimate loss ratio is then used in place of $\widehat{ULR}_i^{k,BF}$ in the BF method.

The assumption of a constant ultimate loss ratio for all accident years can be relaxed. One method of doing this is the Generalized Cape Cod method of (Gluck, 1997), which assigns greater credibility to the loss experience in years that are close to each other using an exponentially weighted average, as shown in Equation 6, where the decay factor γ weights the observed claims experience by time.

$$\widehat{ULR}_i^{k,GCC} = \frac{\sum_{n=1}^I C_{n,I-n+1} \gamma^{abs(i-n)}}{\sum_{n=1}^I \pi_n \beta_{I-n+1} \gamma^{abs(i-n)}} \quad (6)$$

It can be shown that setting the decay factor to zero produces the chain-ladder forecast and setting the factor to one produces the Cape-Cod loss ratio, therefore, the GCC method generalizes between the CL and CC methods.

In practice, the decay factor is often set equal to 75% based on Struzzieri, Hussian, and Plaza (1998), however, since the reserves can be quite sensitive to the choice of the decay factor, the choice usually requires quite some judgement on the part of the actuary.

2.4 Summary

In this section we have defined the key concepts that will be used throughout the rest of the paper, and defined three commonly used reserving methods.

Even with limiting the choice of methods to the CL, BF and GCC, nonetheless, there are a significant number of subjective choices that the actuary must make before applying the methods in practice, as detailed at the end of each of the three previous subsections. For example, some quite complicated ways of applying these techniques might be formulated, for example, the steps might consist of the following:

- Derive loss development factors for each development period using the weighted average estimator of the development factors as given in Equation 4, where the weights are set according to zero if any individual development factors are greater than or less than a certain threshold, which has been set subjectively.
- If working with an incurred triangle, decide on how much negative development to allow for and exclude any LDFs that give too much negative development.
- Using the development pattern as a plug-in estimator of β_j , apply the GCC method using a decay parameter set by judgement.
- In some circumstances, modify the reserves using a different loss ratio assumption.

This is a representative example of the many steps that might be undertaken in a practical reserving exercise.

3 Machine Learning Approach to IBNR Reserving

In the previous section, we have presented different methods that can be applied for reserving for IBNR claims, without much consideration of which method should be chosen and how the method should be applied to produce the most appropriate result. As mentioned in the introduction, in practice, these choices are often made on the basis of well-founded heuristics and experience, and the performance of these choices are monitored over time using metrics such as actual versus expected analysis.

In this section, we borrow methodology from the field of machine learning to select reserving models and make appropriate choices for the parameters that will help to ensure that accurate predictions are made by the models on unseen (i.e. out of sample) claims development data. Note that we do not attempt to give a full overview of machine learning here, thus, for a general introduction to different modelling approaches we refer the reader to Friedman, Hastie, and Tibshirani (2009) and Goodfellow, Bengio, and Courville (2016), and in the context of actuarial modelling, we refer the reader to Richman, von Rummell, and Wüthrich (2019) and Wüthrich and Buser (2018).

Similar to the field of reserving with different methods available to produce reserve estimates, in the field of machine learning many different types of models are available to produce predictions, see Friedman et al. (2009) for an overview. Furthermore, many of these models require the choice of hyperparameters which cannot be calibrated from the data, for example, the number of layers in a neural network. The machine learning approach to selecting among competing models and hyperparameter settings usually proceeds as follows. If enough data is available, then the data is partitioned into two disjoint subsets, with one, usually larger subset, used to train the model and the other subset used to test the model performance on unseen samples. If the data are small, then other schemes to test model performance might be applied, for example, cross-validation, see Friedman et al. (2009) for details. These techniques have in common that predictive performance is tested on out of sample data, and the model with the best predictive accuracy is usually the one selected for the modelling task, from a range of model types, formulations and parameters. Thus, different from traditional actuarial modelling which uses well founded expert judgement and professional knowledge and experience to select models subjectively, the machine learning approach selects the class of model, the parameters and the parameters based on their contribution to increasing the predictive accuracy of the model.

Applying this to reserving, we consider that to derive the claims reserve for the triangle Δ^k , the actuary adopts a predictive model M , such that $\hat{R}_{i,j}^{k,M} = M(X = \Delta^k)$, which is a selection of a model from some space \mathcal{M} containing all possible models. To perform this choice in a quantitative manner, then the choice of $M \in \mathcal{M}$ should be made on the basis of how well we expect the selected model to perform on predicting the unknown future claims amounts, i.e. it becomes necessary to score each of the potential models. Rather than use a likelihood function which must be based on assuming some statistical distribution for the claims amounts¹, we rather suggest that the score should reflect the goals of the reserving analysis, which, in the best estimate case, can be articulated as minimizing either the actual versus expected claims or the claims development results on out of sample data.

To estimate these scores on out of sample data, it is necessary to compare the actual claims experience to the expected, and we do this in a sequential manner to maximise our use of the available data. We proceed to find the optimal model using the following procedure, which references Figure 1 for clarity:

- Select a reasonably sized triangle of claims development experience

¹Whereas it is commonly assumed that incremental claims amounts are distributed according to the Over-dispersed Poisson distribution, distributional assumptions for the cumulative claims amounts are not made that often

which will provide data for fitting all of the models (shown as “Initial Triangle” in blue in Figure 1). Note that new diagonals of experience will be added to this triangle in the subsequent steps.

- 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 1).
- 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 2 .
 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 2).
 - (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 3.
 3. Calculate and store the average score across all of the calendar years in the training set, S^M .
- Select $M_{opt} = \operatorname{argmin}_{M \in \mathcal{M}}(S^M)$.

Note that the assessment of the CDR and AvE scores performed in this procedure is on unseen data, in other words, this is an assessment of out of sample predictive performance. We illustrate the data used by the procedure in Figure 1.

Practically, in this study we only work with a subset of \mathcal{M} that is defined by the CL, BF and GCC models, and some options for selecting the parameters for these models in different ways. We then set up a simple grid containing the different modelling options for each of the methods, and calculate the score for each option as described in the algorithm above. In the following case studies, we test the use of both the CDR and the AvE metrics as scoring functions for each model. As noted in Section 2.2, the CDR can be thought of as a regularized actual versus expected score where the penalty term added to the AvE result is the change in the ultimate reserve estimate between calendar years. This enforces that the optimal algorithm not only

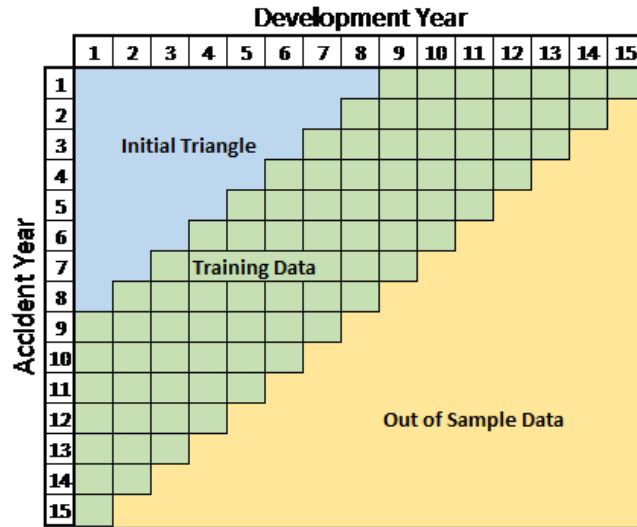


Figure 1: Illustration of how the scoring procedure splits up each triangle. All models utilize the blue initial triangle and then add new diagonals of experience from the training set, which is shown in green. The training set is comprised of several diagonals of the triangle on which firstly, the predictions of each reserving model are tested and then subsequently, the models are refit. The yellow area represents unknown claims development that is not used in the scoring procedure.

reduces predictive error by minimising AvE_{i,j^*}^k , but also stabilises reserves by minimising the change in reserves between calendar years. This is especially important for classes of business that take a long time to develop (so-called “long-tail” classes), since usually enough information to change reserve estimates only becomes available after several years of analysis. On the other hand, particularly in the case of lines that develop quickly (so-called “short-tail” classes) or in the case of rapidly changing calendar year trends in the triangles, maintaining stable reserves may not be a valid goal, thus, in the sections that follow, we also consider optimizing using the AvE metric.

In the following sections, which illustrate applications on several reserving triangles, we measure the performance of the optimal model M_{opt} by calculating the RMSE of the ultimate claims predicted by the model against the true ultimate claims in the triangle. To evaluate whether the scoring procedure selects a model with good predictive performance, we also calculate the RMSE of all of the other models and compare these to the RMSE of the optimal model.

The computations are performed in Python using the *chainladder* package (Bogaardt, 2020). The package allows for many reserving methods in-

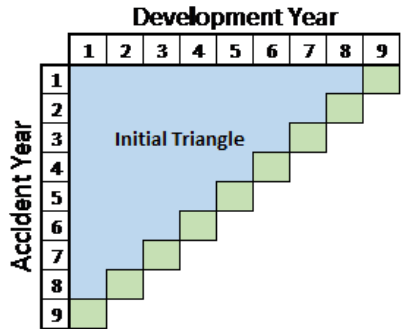


Figure 2: Illustration of the first iteration of the scoring procedure. The first diagonal of claims experience has been added to the initial triangle. The model will be fit to this augmented triangle, and the experience assessed against the second diagonal (which is not shown in this figure).

cluding the CL, BF, and GCC methods, as well as the selection of key parameters for each of these methods. In the tables defining the subset of models which we score below, we supply the name of the package parameters to enable easy reproduction of our work.

4 Case Study 1: Swiss Private Liability

In this section we present a detailed case study of applying the framework in Section 3 to a single large reserving triangle, using each of the CL, BF, and GCC methods. For the case study, we use a Swiss private liability claims triangle that was taken from Gisler (2015). The triangle contains claims development from 1979 until 2015 for a major Swiss insurer and Gisler (2015) notes that the figures have been adjusted for privacy reasons. The accident years range from 1979 until 1997, and the full development history is available until calendar year 2015. Since no premium data is provided for the triangle, premium data has been simulated by assuming that the a 60% loss ratio target as follows:

- Actual ultimate claims, $C_{i,J}$ are taken from the latest development year (recalling that we have the full triangle available).
- To provide some element of randomness, a linear regression model is fit to the ultimate claims using the accident years as a predictive variable, to generate a series of estimates $\hat{C}_{i,J}$.
- Residuals for each accident year $\epsilon_i = C_{i,J} - \hat{C}_{i,J}$, are calculated and are resampled using a single iteration of bootstrapping.

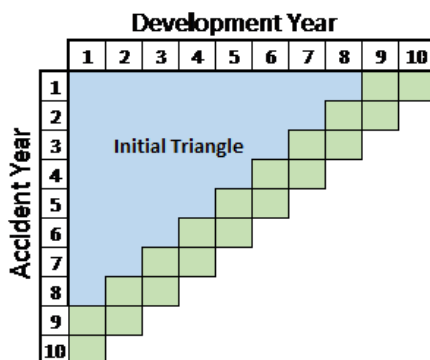


Figure 3: Illustration of the second iteration of the scoring procedure. The first two diagonals of claims experience have been added to the initial triangle. The model will be fit to this augmented triangle, and the experience assessed against the third diagonal (which is not shown in this figure).

- The bootstrapped residuals are used to calculate pseudo-ultimate claims, $C_{i,J}^* = C_{i,J} + \epsilon_i^*$, where ϵ_i^* represents the bootstrapped residual allocated to accident year i .
- Pseudo premium for each accident year is then calculated as $\frac{C_{i,J}^*}{0.6}$, i.e. we assume a 60% loss ratio.

A single iteration of this procedure results in a weighted-average ultimate loss ratio of 59.4%. We show the triangle and the simulated premium figures in Appendix A, which follows the same colouring scheme as used in the previous section..

The AvE_{score} and CDR_{score} were calculated across 13 calendar years from 1984 to 1996 for each parameter set. We chose these years for the training data to provide as much experience for the optimization process as possible, while not starting the optimization on a very small triangle. The test set was comprised of the future development of each of the accident years (1979 to 1997) over the entire lower triangle.

4.1 The Chain Ladder Method

Here, we consider 3 variations on the basic CL method, which uses all accident years for calculating the loss development factors and does not exclude any of the individual development factors. These variations allow for the varying of the number of accident years included in the calculation of the loss development factors \hat{f}_j , and for assigning a zero weight to one or both of the largest or smallest of the individual factors $\hat{f}_{i,j}$ when estimating \hat{f}_j , due to these potentially being outliers and not representative of future claims

development experience. In what follows, we refer to the application of the CL method without these variations as the “basic” CL method. We show these variations on the CL model in Table 1, which results in 36 unique combinations of parameters. Note that the search space could be expanded, for example, instead of choosing to drop the highest and lowest individual factors in all development periods, this choice could be made for each development period independently, or multiple calendar years could be excluded from the calculation. Since the number of possible combinations of parameters grows rapidly with each additional option added, practically one is forced to use a more limited search space, nonetheless, even searching over a limited number of combinations can result in some improvement over the standard CL method.

Parameter	Choice Set	Description
drop_high	[True, False]	Whether to drop the highest individual development factors, $\hat{f}_{i,j}$, in all development periods
drop_low	[True, False]	Whether to drop the lowest individual development factors, $\hat{f}_{i,j}$, in all development periods
n_periods	$k \in [10..19]$	Number of accident years, k , over which to calculate \hat{f}_j

Table 1: Search Space for CL Method

Using the framework defined in Section 3, the variations on the CL method with the parameters given in Table 2 were found to minimise the AvE_{score} and CDR_{score} . The table also shows the parameters used to define the basic CL method. Using a maximum of 11 of the most recent accident years (out of a total of 19) was found to be the optimal choice to minimise both scores. Dropping the highest individual development factor minimised the CDR_{score} , whereas not dropping any individual development factors minimised the AvE_{score} . Figure 4 shows the AvE_{score} and CDR_{score} metrics for each combination of the options in Table 1 on the training data. It can be seen that the model performance is much more sensitive to the choice of which development factors to drop than the number of periods of claims experience to include in the calculation of the development factors.

Parameter	Basic CL	Minimise AvE	Minimise CDR
drop_high	False	False	True
drop_low	False	False	False
n_periods	19	11	11

Table 2: Optimal Parameters Found for CL Method

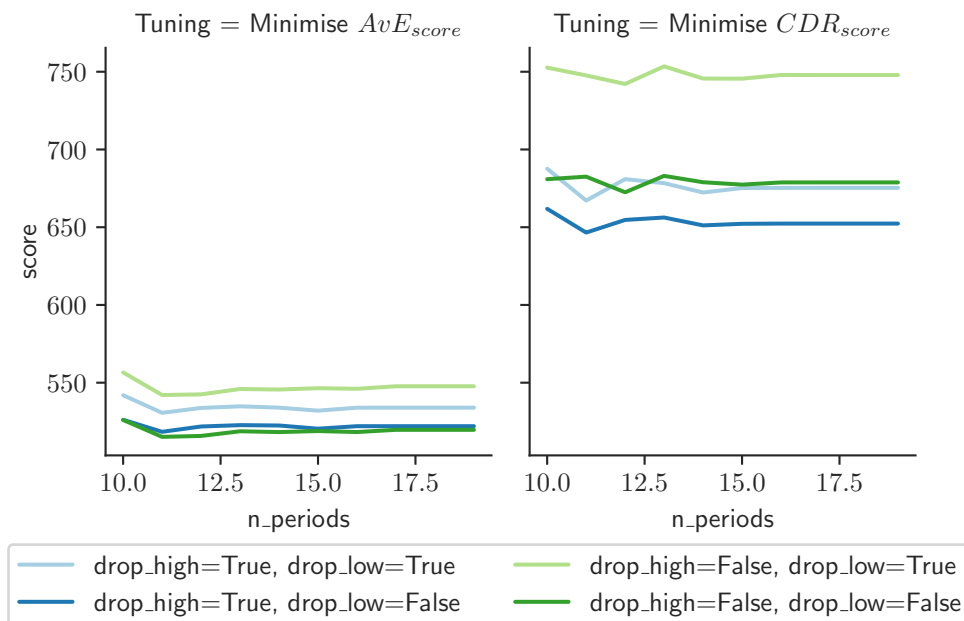


Figure 4: AvE and CDR score for CL Method on the training data

Table 3 shows the RMSE of the projected ultimate claims against the actual ultimate claims in the triangle, i.e. here we are testing how the selected models perform in predicting the final claims development of the triangle, which is the area in yellow in Figure 1. Despite the AvE_{score} being minimised on the training data with the given parameters, the resulting (out of sample) ultimate claim predictions are slightly worse than the basic CL model. On the other hand, the model selected by minimizing CDR_{score} performs better than the basic CL model. The rank column in Table 3 shows how well the selected model performed compared to all of the models considered. In this case, using CDR as the scoring metric results in a model in the top third of the search space. The RMSE scores achieved by the selected models are compared against all the other models in Figure 5.

Figure 6 shows an evaluation of the performance of the basic CL method,

and the two optimal models for each accident year, by subtracting the expected ultimate claims produced using each model from the actual ultimate claims $C_{i,J}$. For the model selected using the AvE score, which only reduced the number of accident years considered when calculating the LDFs, the figure shows that using a lower number of periods over which to calculate the development factors has minimal to no impact until the most recent accident years, where the ultimate claims are underestimated marginally more than the basic CL. Also, both the basic CL and minimising the AvE_{score} tend to overestimate development in most accident years. On the other hand, minimising the CDR_{score} reduces overestimation in these years at the expense of underestimating claims on the most recent accident years, and accident years 1988 through 1991. However, on average the estimates of the ultimate claims are closer to the actual ultimate claims than both the basic CL and the model found by minimising the AvE_{score} .

Model	RMSE	Delta	Rank
Basic CL	669.69	–	23 out of 40
Minimise AvE_{score}	675.38	+0.9%	27 out of 40
Minimise CDR_{score}	617.81	-7.8%	12 out of 40

Table 3: RMSE of Actual Ultimate Claims versus Predicted Ultimate Claims using the CL Method

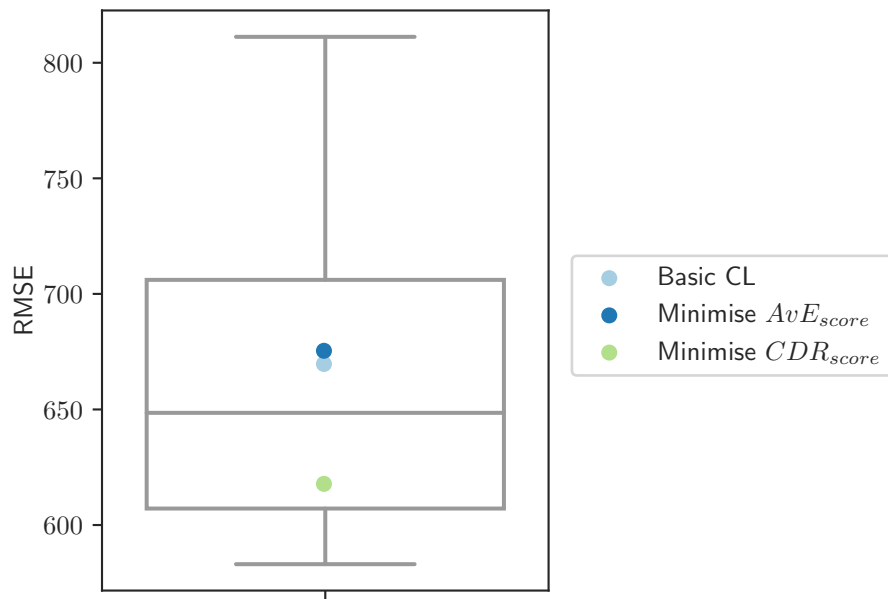


Figure 5: Boxplot of RMSE for all CL models

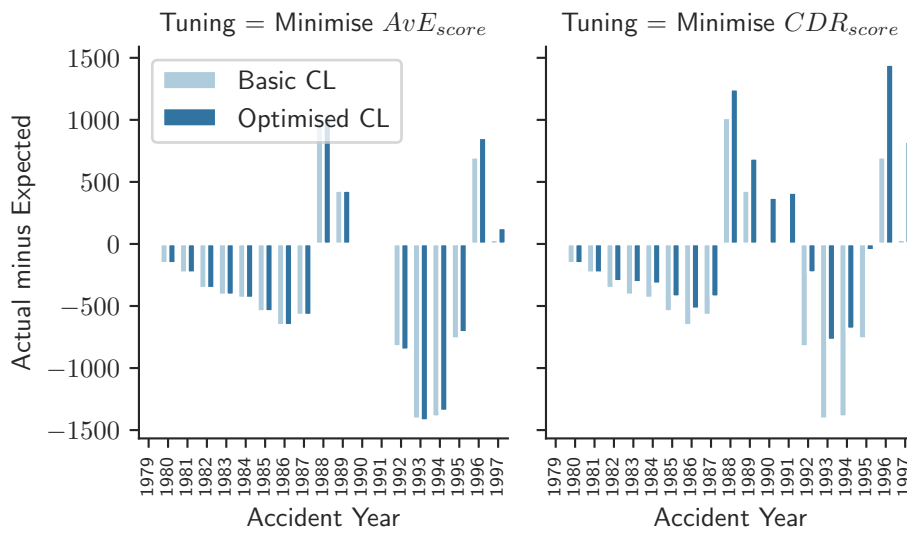


Figure 6: Actual Ultimate Claims minus Predicted Ultimate Claims by Accident Year, for each of the CL Models

Figure 7 shows the IBNR reserve by accident year for each approach

and Table 4 shows the total IBNR across all accident years. The basic CL IBNR results in the largest IBNR reserves. The optimal variations on the CL method both produce lower IBNR reserve than the estimate derived using the basic CL method, with the estimate derived using the model that minimizes the CDR_{score} metric 16.3% lower than the basic CL method respectively. Since, at least using the CDR_{score} metric, it is possible to find a variation on the CL method that results in more accurate predictions than the basic CL method, this indicates that the basic CL method is not a producing a best estimate on the Swiss triangle, i.e. the basic CL reserves are too conservative.

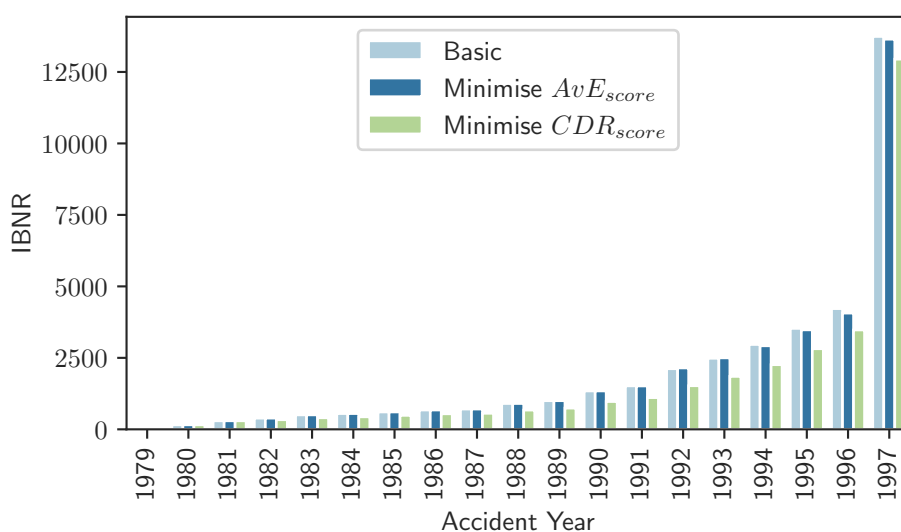


Figure 7: Comparison of IBNR Reserves by Accident Year

Model	Total IBNR	Delta
Basic CL	37 727	–
Minimise AvE_{score}	37 417	-0.8%
Minimise CDR_{score}	31 595	-16.3%

Table 4: Total IBNR Reserve for CL Tuning

4.2 The Bornhuetter-Ferguson Method

Applying the BF method requires the choice of an expected ultimate loss ratio in addition to the loss development factors from the CL method. We

consider the same variations on the CL method as in the previous subsection, but include in addition a search over possible choices for the expected ultimate loss ratio of the BF method. Since we expect the loss ratio to lie within the 50% to 70% interval (since we derived premiums on the basis of a 60% ultimate loss ratio), the search space for loss ratio is limited to 1% increments in this range, which expands the total search space, shown in Table 5, to 756 unique parameter sets.

Parameter	Choice Set	Description
drop_high	[True, False]	Whether to drop the highest individual development factors, $\hat{f}_{i,j}$ in all development periods
drop_low	[True, False]	Whether to drop the lowest individual development factors, $\hat{f}_{i,j}$ in all development periods
n_periods	$k \in [10..19]$	Number of accident years over which to calculate \hat{f}_j
apriori	$\alpha \in \{0.50, 0.51, \dots, 0.69, 0.70\}$	Apriori loss ratio for the BF method

Table 5: Search Space for BF Method

Table 6 shows the parameters that were found to minimise the RMSE of the AvE and CDR scores on the training data. Since we know the true loss ratio is 60%, we apply a basic BF method (i.e. using the CL development pattern without any of the variations considered previously) using a loss ratio of 60% for comparison purposes. Using 15 of the most recent accident years was found to minimise the AvE_{score} , whereas 11 accident years remained optimal for minimising the CDR_{score} . The optimal choices for dropping individual development factors remained the same as for the CL method for both scores. Minimising the the AvE and CDR scores led to a choice of 59.0% as the expected ultimate loss ratio (called “apriori” in the table), which is in line with the true value of 59.4%.

Figures 8 and 9 show the AvE_{score} and CDR_{score} metrics for each combination of the options in Table 5 on the training data. As would be expected, the BF method is very sensitive to the choice of loss ratio and when this is poorly chosen, the models are penalized quite severely.

Parameter	Basic BF	Minimise AvE	Minimise CDR
drop_high	False	True	True
drop_low	False	False	False
n_periods	19	15	11
apriori	0.60	0.59	0.59

Table 6: Best Parameters Found for BF Method

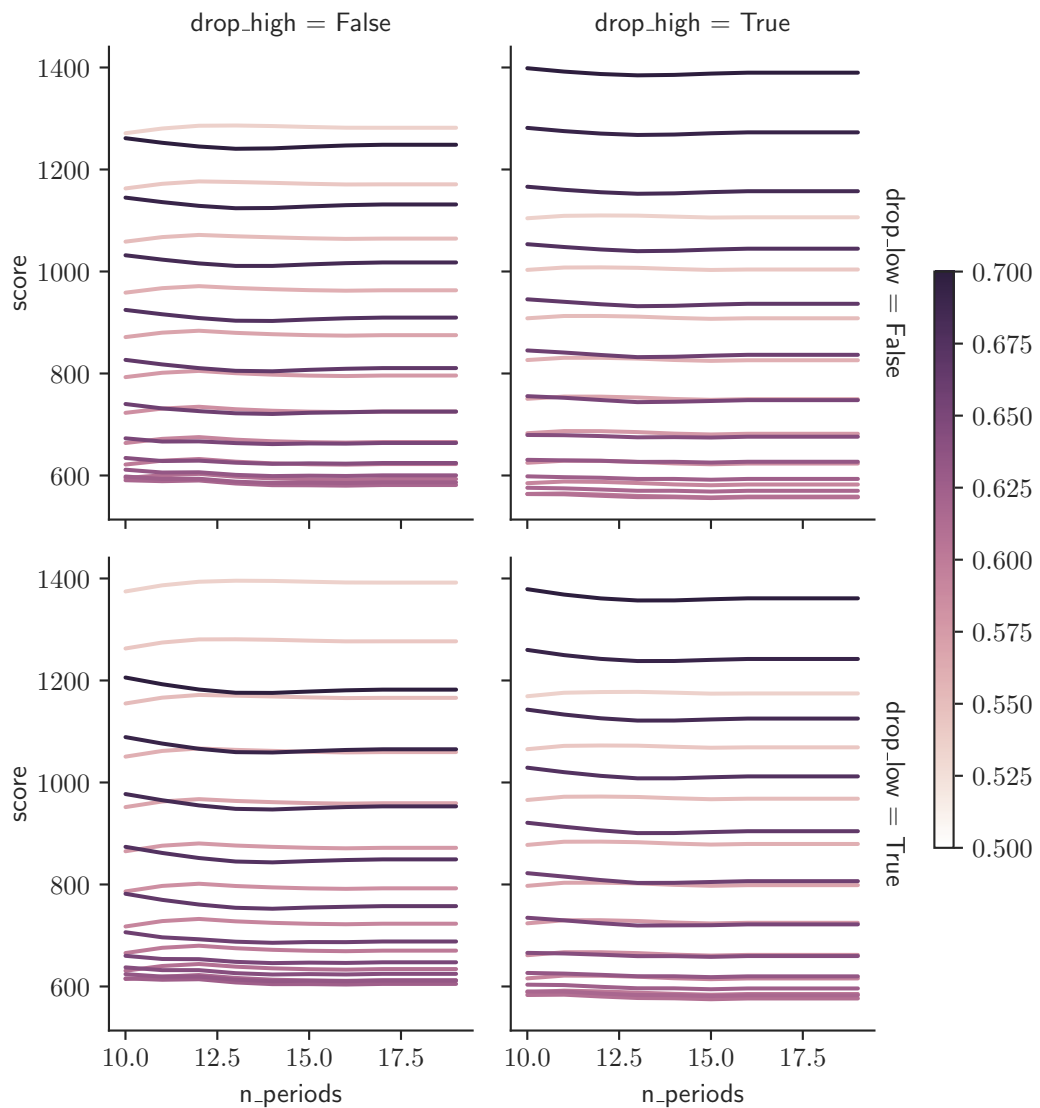


Figure 8: AvE score for BF method on the training data

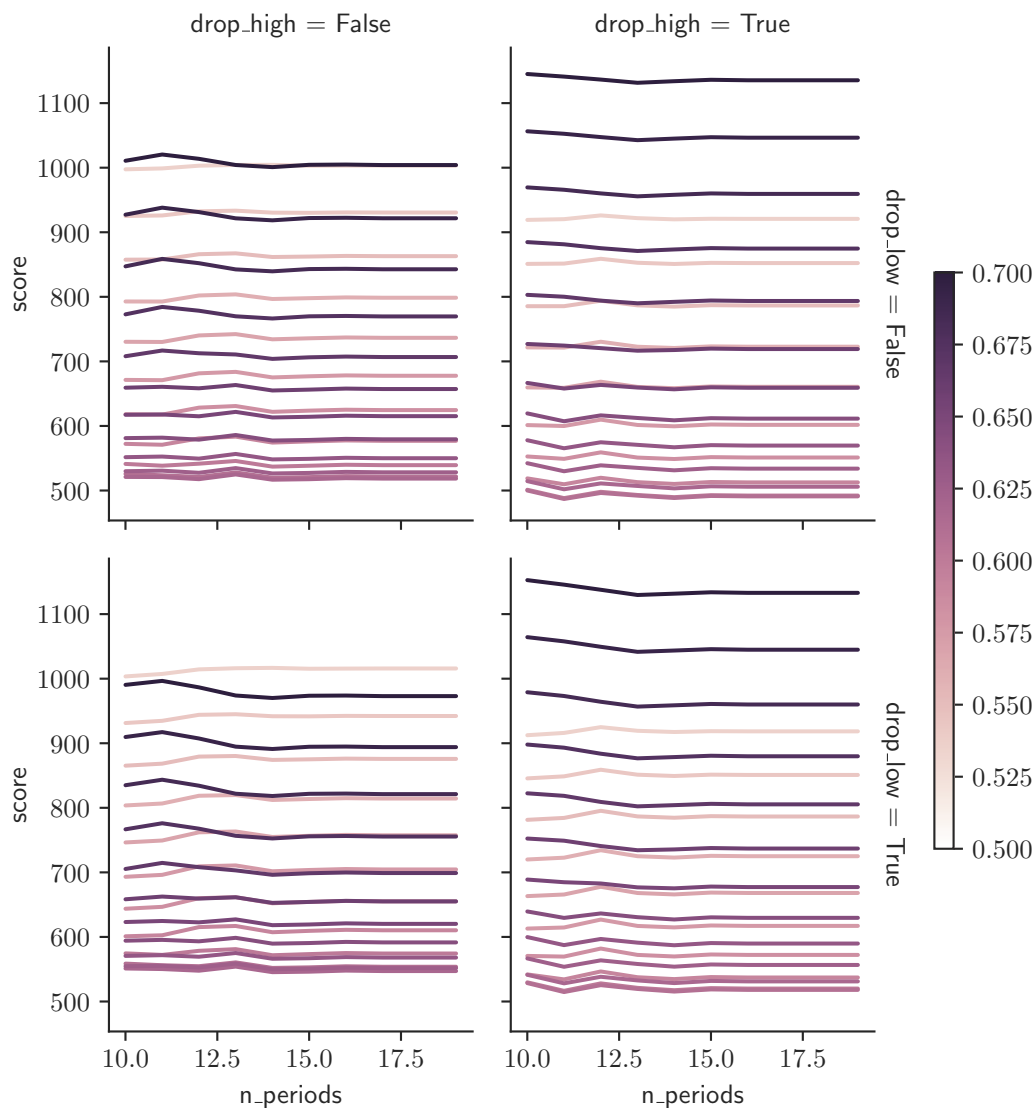


Figure 9: CDR score for BF method on the training data

We remark that, although we considered all parameter combinations when finding the optimal models, since the CL model specification underlying the BF model that was selected using the CDR score remained the same as in the previous section, it seems to be possible that one could select an optimal BF method in a two-step process: first by finding an optimal variation of the CL model and then by finding the optimal loss ratio. Following this two-step procedure would have produced the same model as that selected using the CDR score, but at much lower computational cost.

Table 7 shows the RMSE of the projected ultimate claims against the actual ultimate claims for the BF method. In terms of ranking, the models minimising the AvE_{score} and CDR_{score} metrics score in the top 20% of models considered here. The RMSE scores achieved by the selected models are compared against all the other models in Figure 10.

Model	RMSE	Delta	Rank
Basic BF	576.38	–	256 out of 840
Minimise AvE_{score}	537.27	-6.8%	159 out of 840
Minimise CDR_{score}	527.90	-8.4%	139 out of 840

Table 7: RMSE of Actual Ultimate Claims versus Predicted Ultimate Claims

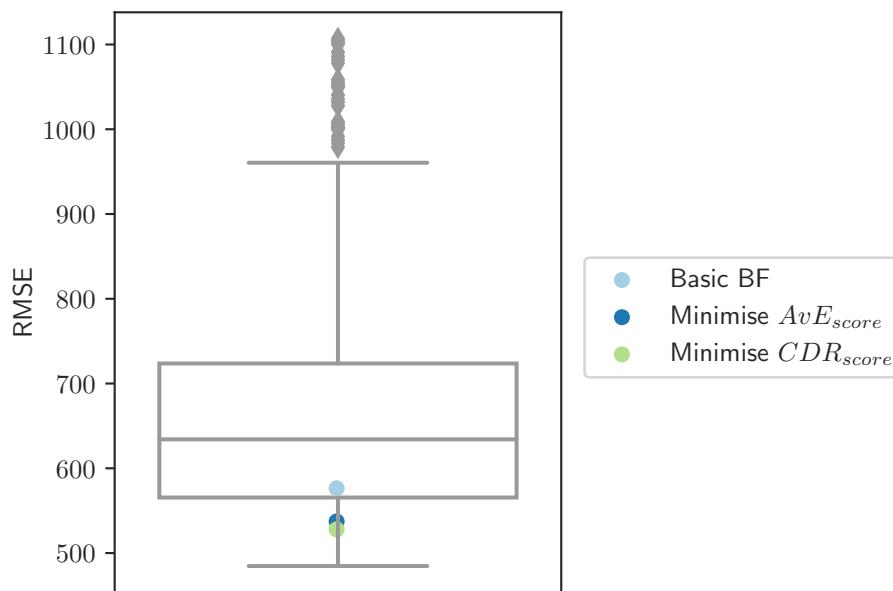


Figure 10: Boxplot of RMSE for all BF Models

The predicted claims from the models found by minimising the AvE_{score} and CDR_{score} resulted in lower RMSE scores compared to the base BF model with a 60% expected ultimate loss ratio assumption. However, as shown in Figure 11, both sets of predictions underestimate the ultimate claims in the most recent accident years, whereas the basic BF model displays almost a zero bias in these years. This suggests that adding a penalty for biased estimates might enhance the AvE_{score} and CDR_{score} metrics.

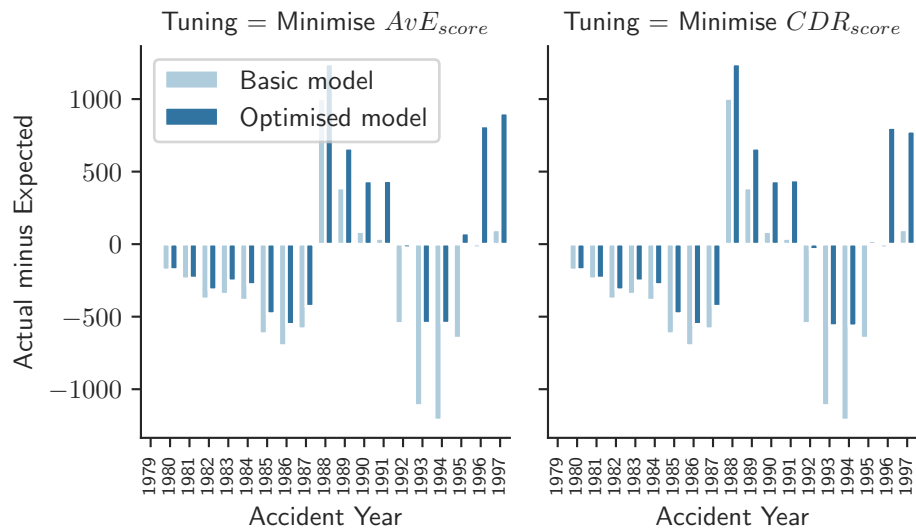


Figure 11: Actual Ultimate Claims minus Predicted Ultimate Claims by Accident Year, for each of the BF Models

Figure 12 shows the IBNR reserve by accident year for each approach under the BF method, and Table 8 shows the total IBNR reserve. The basic BF IBNR results in the highest IBNR reserves across all accident years, with minimising the AvE_{score} resulting in the lowest IBNR. Minimising the CDR_{score} results in IBNR reserves only slightly larger than minimising the AvE_{score} .

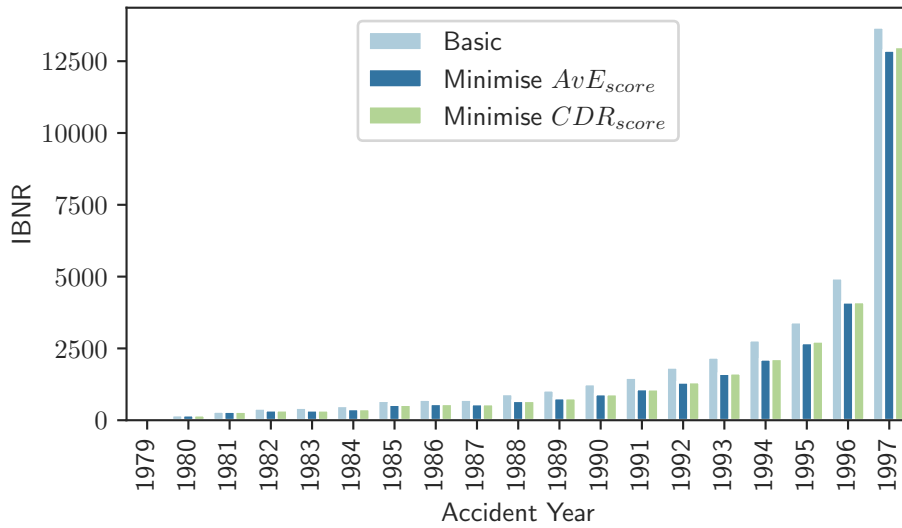


Figure 12: Comparison of IBNR Reserves by Accident Year

Model	Total IBNR	Delta
Basic BF	37 511	–
Minimise AvE_{score}	31 414	-16.3%
Minimise CDR_{score}	31 647	-15.6%

Table 8: Total IBNR Reserve for BF Tuning

Taking advantage of expert knowledge can be of significant value when tuning parameters, for example, by narrowing the search space to a range of 10% around the true value, the amount of computation would be reduced significantly. The search space could be restricted to a tighter region, or even to a single number, if the insurance book is very stable, or the reserving actuary has reason to believe the range to be narrower. As an example, Figure 9 shows the RMSE of the projected ultimate claims against the actual ultimate claims for the BF method when fixing the expected ultimate loss ratio to an estimate of 60%. The table shows that, in this case, slight accuracy gains are made compared to when the models searched over the wider range of 50% - 70%.

In terms of the ranking of the models, optimising over the smaller search space results in the model with median performance being selected when minimising the AvE score, whereas minimising the CDR score results in a model that performs better than the median. All scores of the models in

the smaller search space are shown in Figure 13.

Model	RMSE	Delta	Rank
Basic BF	576.37	–	24 out of 40
Minimise AvE_{score}	516.29	-10.4%	20 out of 40
Minimise CDR_{score}	509.74	-11.6%	15 out of 40

Table 9: RMSE of Actual Ultimate Claims versus Predicted Ultimate Claims

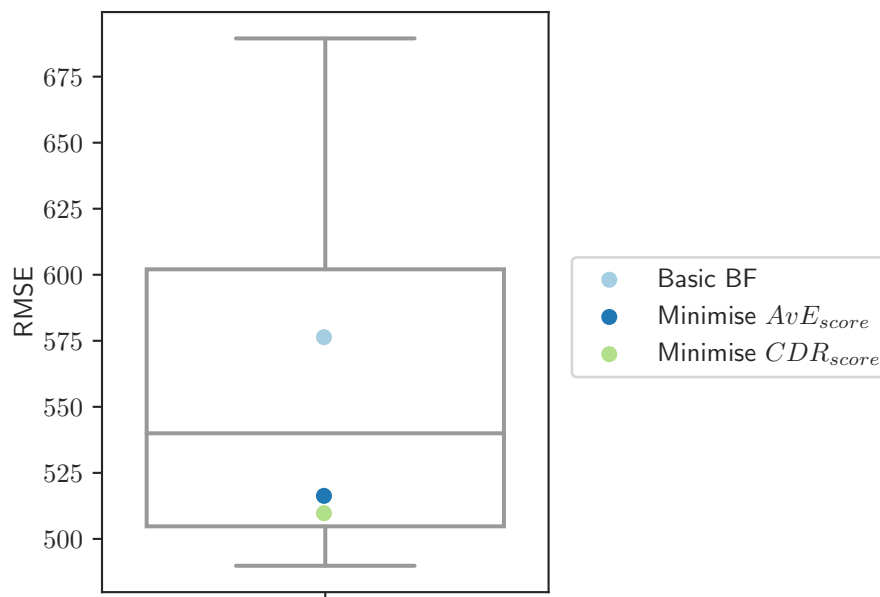


Figure 13: Boxplot of RMSE for Models in Narrow BF Search Space

4.3 The Generalised Cape Cod Method

The GCC method requires the choice of an exponential decay factor that is used to weight the contribution of the accident years in calculating the expected ultimate loss ratio. Table 10 shows the same search space as the CL method, expanded to include a search over possible choices for the decay parameter of the GCC method in 5% increments in the range 0% to 100% (i.e. including the CL method when the decay parameter is 0 and the GCC method when the decay parameter is 100%).

Parameter	Choice Set	Description
drop_high	[True, False]	Whether to drop the highest individual development factors, $f_{i,j}$ in all development periods
drop_low	[True, False]	Whether to drop the lowest individual development factors, $f_{i,j}$ in all development periods
n_periods	$k \in [10..19]$	Number of accident years over which to calculate f_j
decay	$\gamma \in \{0.00, 0.05, \dots, 0.95, 1.00\}$	Decay parameter for the GCC method

Table 10: Search Space for GCC Method

Table 11 shows the parameters that were found to minimise the AvE_{score} and CDR_{score} using the training data. Using a maximum of 11 of the most recent accident years was found to minimise both the AvE score and the CDR score. Minimising the AvE score resulted in the choice of a decay factor of 0%, in other words, the CL method was selected, whereas a parameter value of 95% was chosen when minimising the CDR score. Figures 14 and ?? show the AvE_{score} and CDR_{score} metrics for each combination of the options in Table 10 on the training data. The GCC method is very sensitive to the choice of decay parameter and it can be seen from these figures that the optimal range of decay parameters is strongly influenced by which of the AvE or CDR metrics is being considered, with the CDR metric producing a range of parameters much closer to what would be selected in practice by an actuary reserving for a long-tail line.

In the following, we also show results for the GCC with an heuristic assumption of a decay parameter of 75% for comparison purposes, as is sometimes the practice based on Struzzieri et al. (1998), and we refer to this model as the basic GCC model.

Parameter	Basic GCC	Minimise AvE_{score}	Minimise CDR_{score}
drop_high	False	False	True
drop_low	False	False	False
n_periods	19	11	11
decay	0.75	0.00	0.95

Table 11: Best Parameters Found for GCC Method

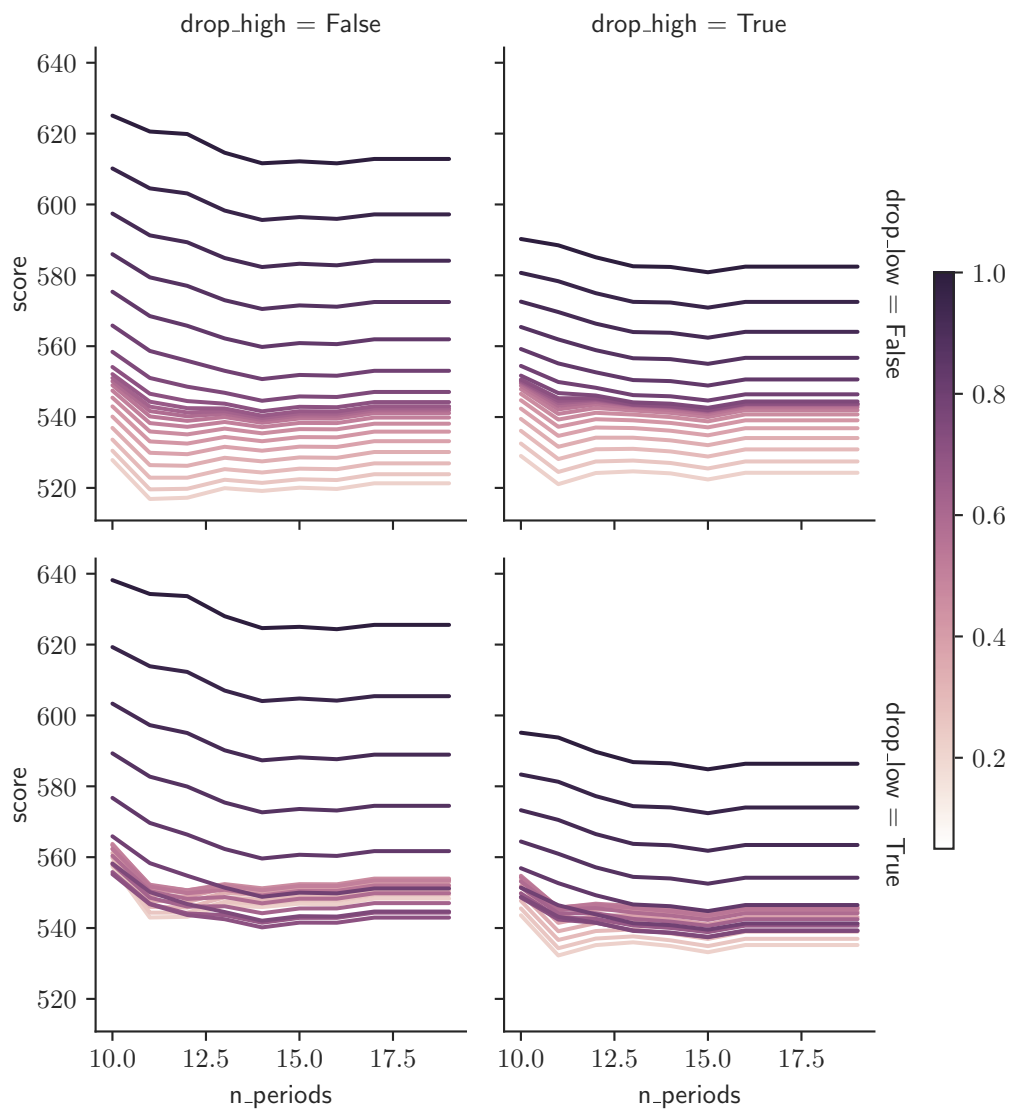


Figure 14: AvE score for CC method on the training data

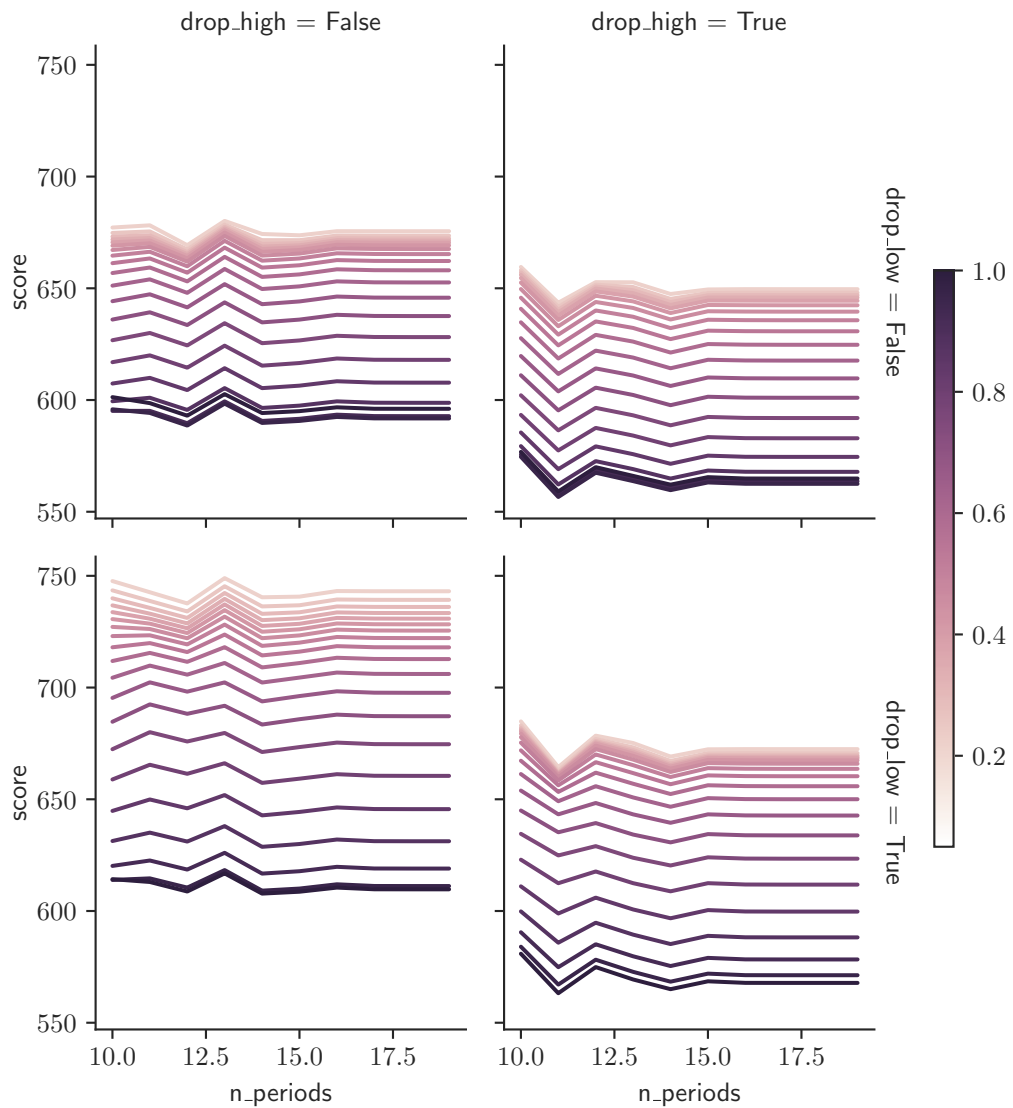


Figure 15: CDR score for CC method on the training data

Table 12 shows the RMSE of the projected ultimate claims using the GCC method against the actual ultimate claims. Minimising the AvE score results in a model that does not perform well, whereas minimising the CDR score once again selects a model with the best generalisation of the three models, with a score in the top 10% of all models. The scores for all models are shown in Figure 16.

Model	RMSE	Delta	Rank
Basic GCC	604.27	–	366 out of 800
Minimise AvE_{score}	670.69	+11.0%	595 out of 800
Minimise CDR_{score}	580.21	-4.0%	72 out of 800

Table 12: RMSE of Actual Ultimate Claims versus Predicted Ultimate Claims

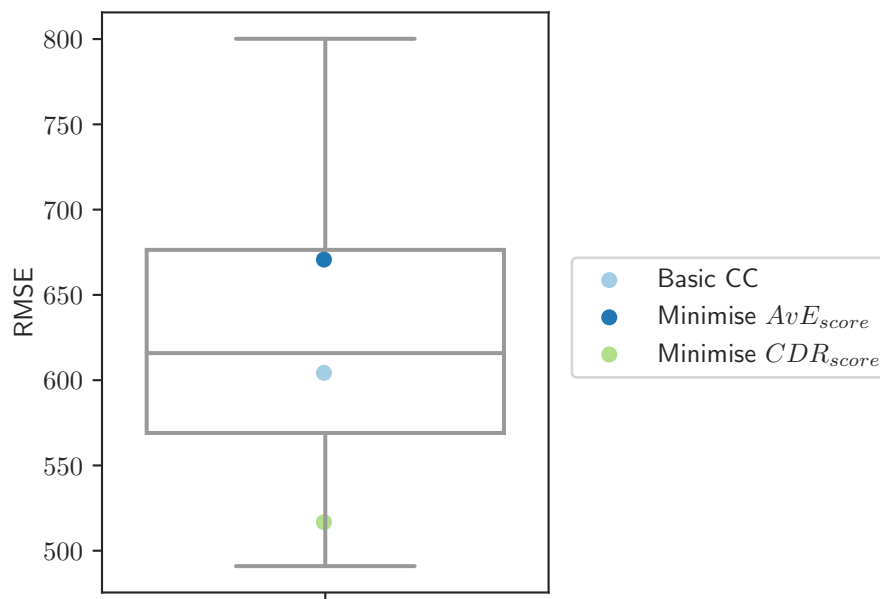


Figure 16: Boxplot of RMSE for all CC models

Minimising the AvE_{score} results in a higher RMSE score compared to the basic GCC with a 75% decay. This is due to the spoor predictions of the selected model for the accident years subsequent to 1991, as shown in Figure 17. Minimising the CDR_{score} provides similar and relatively accurate results in all accident years.

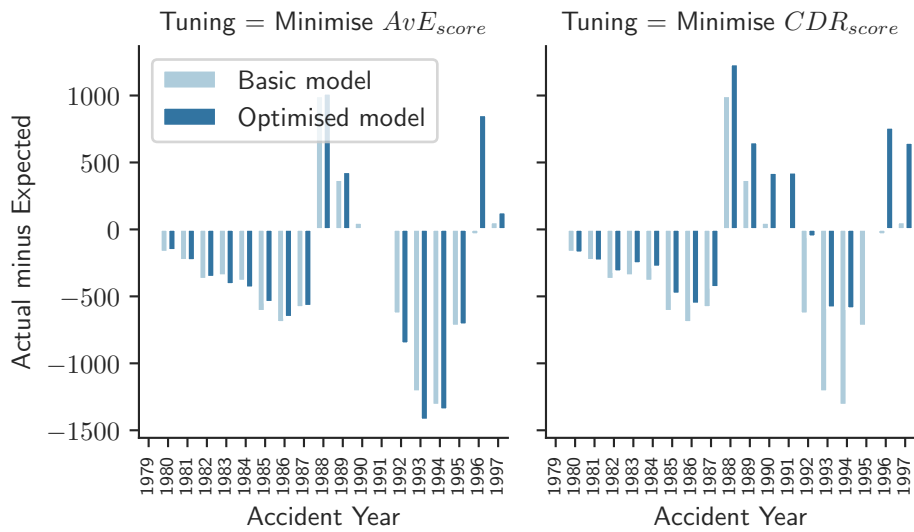


Figure 17: AvE by Accident Year for GCC models

Figure 18 shows the IBNR reserve by accident year for each approach using the GCC method, with the aggregate IBNR given in Table 13. The basic GCC method results in the highest IBNR reserves across all accident years, while minimising the CDR_{score} and the AvE_{score} results in similar figures that are lower than the reserves produced by the basic GCC method.

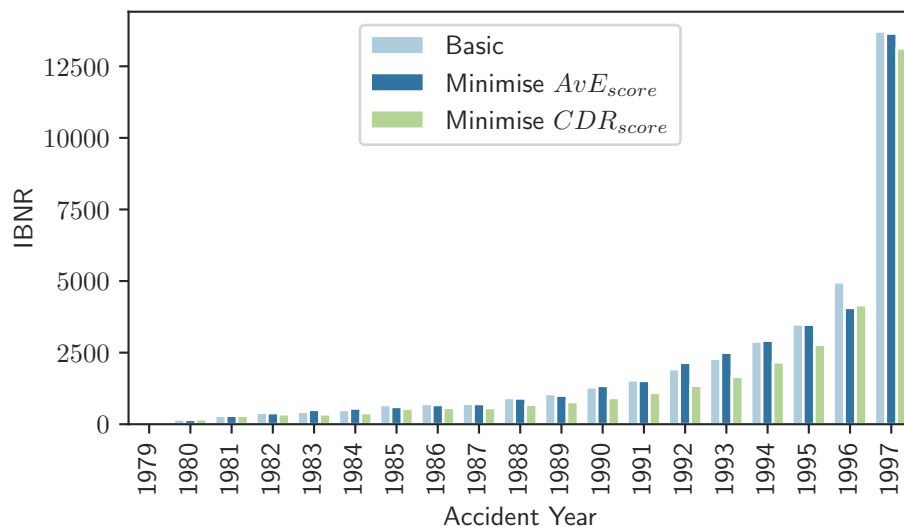


Figure 18: Comparison of IBNR Reserves by Accident Year

Model	Total IBNR	Delta
Basic GCC	37 511	–
Minimise AvE_{score}	31 414	-16.3%
Minimise CDR_{score}	31 647	-15.6%

Table 13: Total IBNR Reserve for GCC Tuning

4.4 Summary

We now assess which one of the three methods (CL, BF or GCC) we would have selected based on the AvE and CDR metrics calculated on the training data, which we show in Table 14. That is, we investigate which method produced the lowest AvE and CDR scores on the training data among the three methods analysed in the previous sections and what the implications of choosing that method would be for the out of sample predictive performance of that method.

The lowest CDR_{score} on the training data was produced using the BF method, which also produces the best results on the out of sample data. On the other hand, the AvE_{score} metric is minimized on the training data by the CL method. If presented with a choice between the CL and BF methods on a liability line of business, applying judgement would probably lead to the selection of the BF method, and selecting this method on the basis of the

low CDR score shown for the BF method in Table 14 would have provided good performance on the out of sample data. We discuss this more in the conclusion to the paper.

	CL	BF	GCC
AvE_{score}	515.16	555.73	516.87
CDR_{score}	646.56	486.88	556.60

Table 14: Lowest Training Set Scores for each Method and Score.

5 Case Study 2: Quarterly Incurred Triangles from Medium-size Insurer

We now apply the framework to two quarterly incurred triangles that, compared to the Swiss triangle in the previous section are much more unstable, due probably to a lower volume of business written by the medium sized insurer that supplied the triangles. These triangles exhibit some interesting characteristics that often occur in real-world exercises: individual accident quarters can be affected by single large claims, there is a non-constant loss ratio implied by the ultimate losses, and premium rates change over the period, although the premium rate information is not usually available to adjust the earned premium figures. Since these triangles are quite representative of some real-world reserving exercises performed by the authors for smaller insurers, if some extra accuracy relative to traditional methods can be achieved by applying the framework developed in this study to these triangles, then this should be a good indicator of potential gains from applying the framework in this paper in these cases.

The first triangle represents one class of short-tail property related perils, and the second triangle represents one class of long-tail liability related perils. The triangles were provided by a branch of a global insurer² and cover accident quarters from 2010Q1 through to 2014Q4. The figures have been adjusted to preserve confidentiality by dividing by a random number, and the earned premium has been normalised to an average loss ratio (across all development quarters) of 50%. Unlike the previous section, a full discussion of each reserving method will not be provided here, rather, the method that results in the lowest MSE (after finding each method's optimal parameter set) will be assessed. This adds an additional choice to the search space, i.e. which of the three methods (CL, BF, and GCC is) selected. Different models were assessed using the 2012Q1 to 2014Q3 accident quarters as a training

²The authors wish to thank the management and actuarial department of the insurer for supporting this research by providing data.

set, resulting in 11 calendar quarters on which these were assessed. The final calendar quarter, 2014Q4, was used for scoring the models which were fit on 2014Q3. The out of sample performance of each model was assessed on the subsequent development of all of the accident quarters.

Table 15 shows the sub-set of the search space, \mathcal{M} , of models considered. Note that the left column details the method the parameters apply to.

Method	Parameter	Choice Set	Description
All	drop_high	[True, False]	Whether to drop the highest individual development factors, $f_{i,j}$ in all development periods
All	drop_low	[True, False]	Whether to drop the lowest individual development factors, $f_{i,j}$ in all development periods
All	n_periods	$k \in [5..21]$	Number of accident years over which to calculate f_j
BF	apriori	$\alpha \in \{0.40, 0.41, \dots, 0.59, 0.60\}$	Apriori loss ratio for the BF method
GCC	decay	$\gamma \in \{0.00, 0.05, \dots, 0.95, 1.00\}$	Decay parameter for the GCC method

Table 15: Search Space for Quarterly Triangles

5.1 Long-tail Liability Line

It was found that the BF method resulted in the smallest AvE_{score} and the smallest CDR_{score} for the long-tail liability line of business. The parameters found for these respective methods are given in Table 16. In terms of loss development factors, the optimal models do not drop the highest and lowest individual development factors in the triangle. Minimising the AvE_{score} resulted in 10 accident periods being chosen to calculate the development factors, whereas minimising the CDR_{score} resulted in 11 periods being chosen. Minimising the AvE_{score} produced an expected ultimate loss ratio of 46%, and minimising the CDR_{score} produced an expected ultimate loss ratio of 41%. We compare these models to the basic GCC method with a decay parameter of 75%.

Table 17 shows the RMSE of the projected ultimate claims against the actual ultimate claims.

Parameter	Basic GCC	Minimise AvE	Minimise CDR
drop_high	False	False	False
drop_low	False	False	False
n_periods	21	10	11
apriori	n/a	0.46	0.41
decay	0.75	n/a	n/a

Table 16: Optimal parameters found for the long-tail line

Model	RMSE	Delta	Rank
Basic GCC	3 170.88	–	1073 out of 1 848
Minimise AvE_{score}	2 552.39	-19.5%	461 out of 1 848
Minimise CDR_{score}	2 893.23	-8.8%	832 out of 1 848

Table 17: RMSE of actual ultimate claims versus predicted ultimate claims for the long-tail line

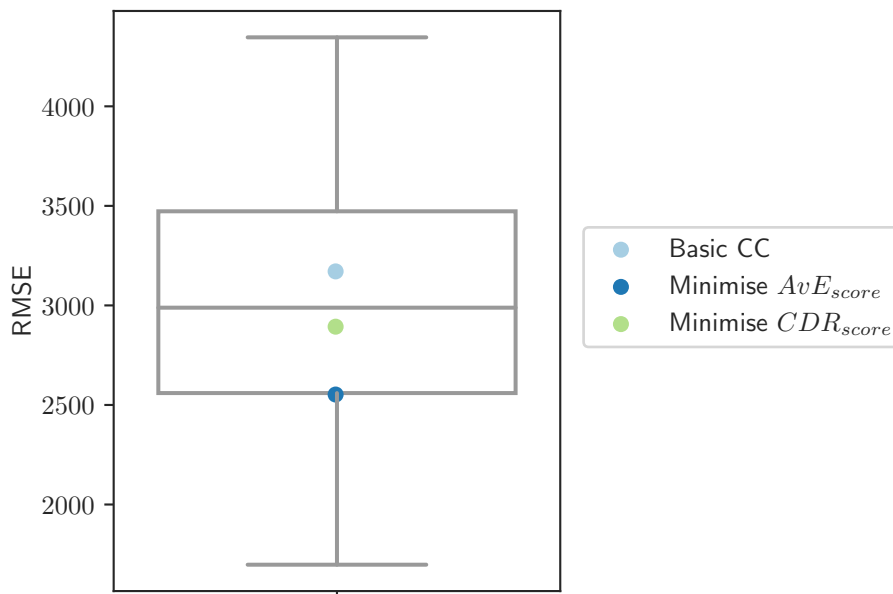


Figure 19: Boxplot of RMSE for Models in Search Space

Minimising the AvE_{score} results in the best prediction of ultimate claims when measured by RMSE, while minimising the CDR_{score} improves on the

basic CC method, but does not produce a model with better predictions than that produced by minimizing the AvE_{score} . This is perhaps due to changing calendar year trends within the triangle, which would somewhat invalidate the CDR_{score} objective of keeping reserve figures constant over time (we discuss this further in the conclusion). Minimising the AvE leads to the selection of a model that is in the top 25% of all methods and parameters considered, as measure by performance on the out of sample data.

Figure 20 shows that minimising both scores reduces the under-estimation of the ultimate claims for each accident quarter.

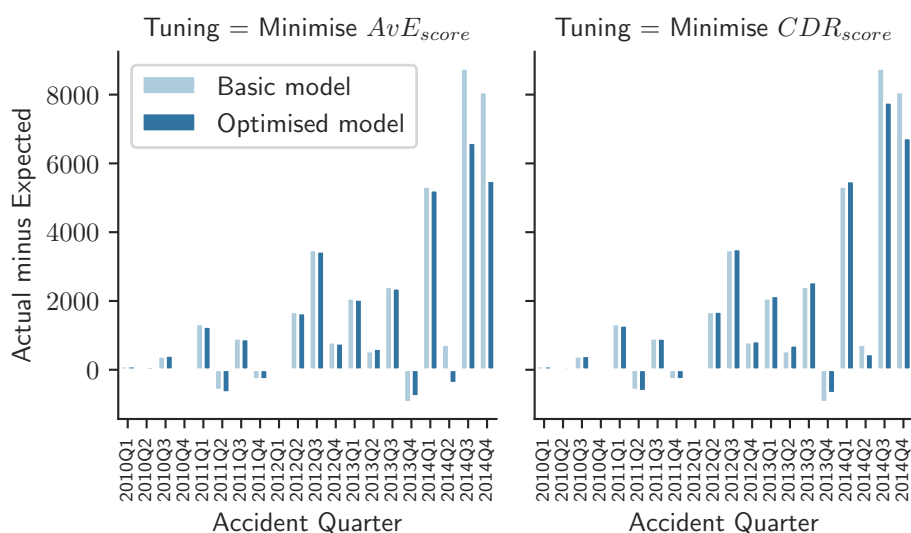


Figure 20: AvE by Accident Quarter for Long-Tail

Figure 21 shows the IBNR reserve by accident year for each approach and Table 18 shows the total IBNR for all accident years. Overall, the optimal methods resulted in a substantially higher IBNR, indicating that the basic GCC method underestimates the IBNR considerably.

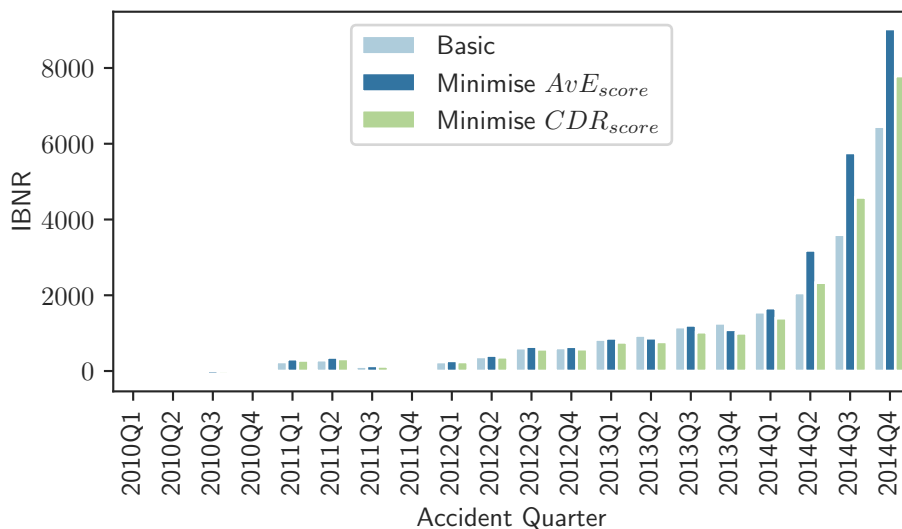


Figure 21: Comparison of IBNR Reserves by Accident Year

Model	Total IBNR	Delta
Basic GCC	20 859	–
Minimise AvE_{score}	44 731	+114%
Minimise CDR_{score}	37 155	+78%

Table 18: Total IBNR Reserve for Tuning: Long-Tail

5.2 Short-tail Property Line

Following the same procedure as for the long-tail line before, minimising the AvE_{score} and CDR_{score} on the data for the short-tail property line indicated that, for both metrics, the BF method is optimal. The optimal parameter sets are given in Table 19. It was found that dropping the lowest individual development factors minimised both the AvE and CDR scores, and that, for calculating the development factors, the use of 17 development periods minimised the AvE score whereas 11 development periods minimised the CDR score. Both the CDR and AvE metrics provide the same choice of expected loss ratio, only differing in the choice of the number of accident periods to consider.

Table 21 shows the RMSE of the projected ultimate claims against the actual ultimate claims, which provide some improvement over the basic GCC method used for comparison when the AvE metric is used, but no improve-

Parameter	Basic GCC	Minimise AvE	Minimise CDR
drop_high	False	False	False
drop_low	False	True	True
n_periods	21	17	11
apriori	n/a	60%	60%
decay	75%	n/a	n/a

Table 19: Optimal parameters found for the short-tail line

ment with the CDR score. These results are compared against all the models considered here in Figure 22.

Figure 23 shows that whereas the basic GCC method estimates the actual claims in the most recent accident quarter quite well, the models selected using the scoring procedure perform quite poorly, overestimating the IBNR required significantly. On the other hand, for older accident quarters, these models perform better than the basic GCC method. Inspecting the triangle for the short-tail line in Table 28 shows that the most recent accident quarter exhibits anomalously low development in the first development quarter; these LDFs are shown in table 20. Whereas the more recent accident periods since 2012Q3 exhibit LDFs generally above 1.5, in the case of 2014Q4, the LDF is only 1.39. Thus, given the information available when selecting the optimal models, this seems like a more reasonable outcome than the figures in Table 21 might indicate.

For earlier accident quarters, the optimal parameter set slightly underestimates the ultimate claims for each accident quarter, however, this is less than the underestimation produced by the basic GCC method. Note that the IBNR for these accident periods is estimated to be negative, as shown in Figure 24, thus, a lower underestimation results in higher IBNR.

In aggregate, the optimal models produce a smaller IBNR reserve in all but the latest accident quarter, as seen in Figure 24. Overall, Table 22 shows that the total IBNR is much larger when calculated using the optimal models.

	\hat{f}_1
2010Q1	1.585
2010Q2	1.316
2010Q3	1.196
2010Q4	1.362
2011Q1	1.268
2011Q2	1.371
2011Q3	1.345
2011Q4	1.532
2012Q1	1.636
2012Q2	1.367
2012Q3	1.431
2012Q4	1.525
2013Q1	1.512
2013Q2	1.316
2013Q3	1.470
2013Q4	1.611
2014Q1	1.808
2014Q2	1.591
2014Q3	1.542
2014Q4	1.392

Table 20: Loss Development Factors for first development period, short-tail triangle

Model	RMSE	Delta	Rank
Basic GCC	638.38	–	1 107 out of 1 848
Minimise AvE_{score}	626.73	-1.8%	1 050 out of 1 848
Minimise CDR_{score}	794.65	+24.5%	1 497 out of 1 848

Table 21: RMSE of actual ultimate claims versus predicted ultimate claims for the short-tail line

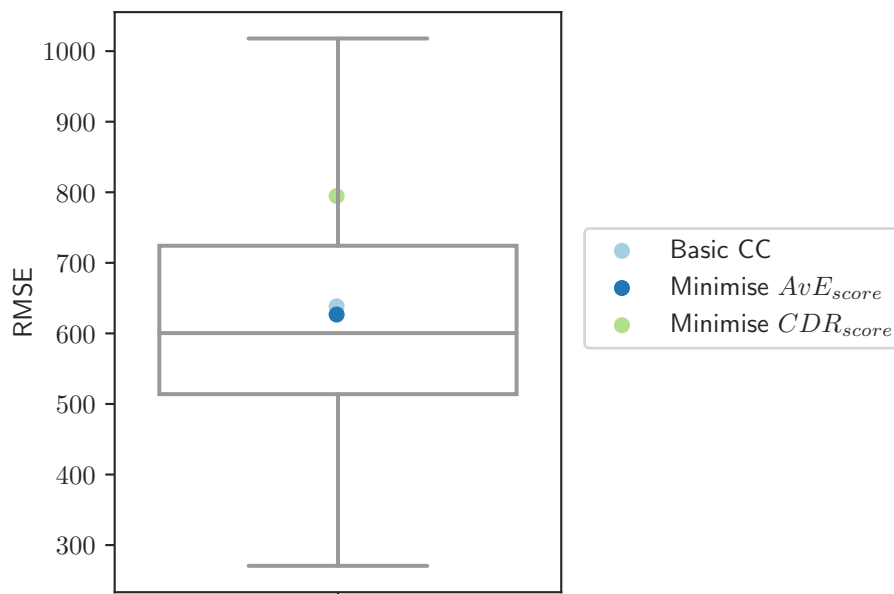


Figure 22: Boxplot of RMSE for Models in Search Space

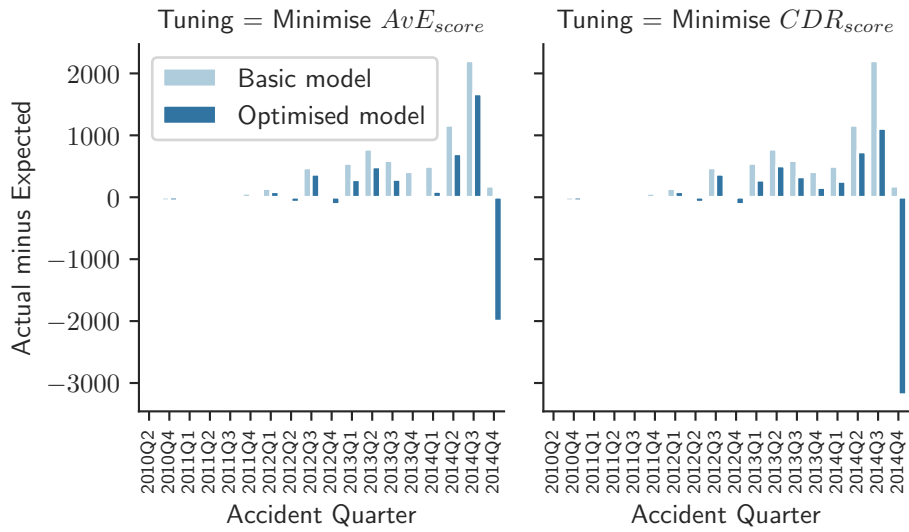


Figure 23: AvE by Accident Quarter for Short-tail line

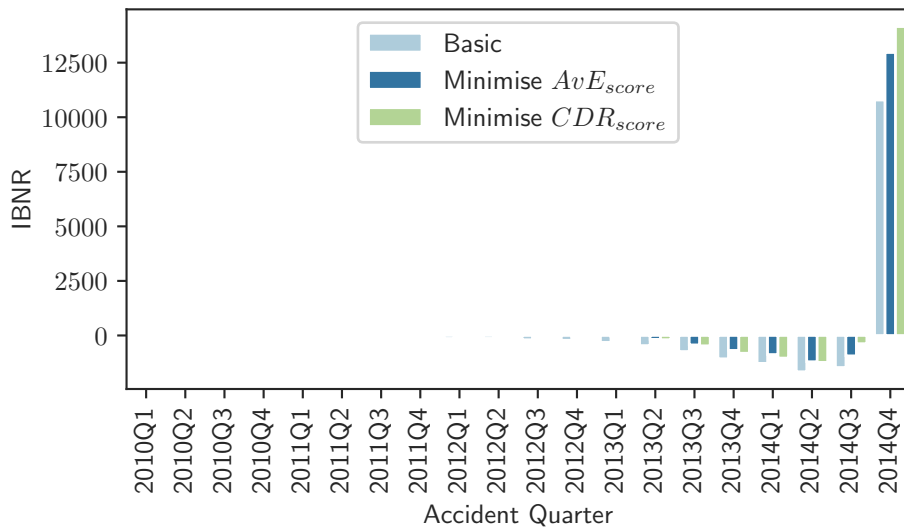


Figure 24: Comparison of IBNR Reserves by Accident Year

Model	Total IBNR	Delta
Basic GCC	2 971	–
Minimise AvE_{score}	8 226	+177%
Minimise CDR_{score}	9 602	+223%

Table 22: Total IBNR Reserve for Tuning: Short-tail

5.3 Summary

In this section, we have demonstrated that compared to the basic application of the reserving methods, the framework presented here works well on incurred triangles (the previous section demonstrated the method on paid triangles) and a more optimal amount of negative IBNR was calculated for the short-tail line. This is despite the quite messy triangles utilized in this section. We have also seen some limitations on the use of a single loss ratio for all accident periods in the BF method, and we discuss this point more in the last section. Also, compared to the findings on the Swiss triangle, where models selected using the CDR score were more optimal, on both triangles considered here, the models selected using the AvE score were better.

6 Discussion and Conclusion

In this paper, we have presented a framework for selecting reserving models that are expected to perform well in predicting out of sample claims development experience and demonstrated that, on three example triangles, our proposal performs relatively well. Thus, we can conclude that scoring reserving models based on historical claims development data provides a useful way of determining which models are likely to predict the future claims development experience well. Also, our framework provides a way to select reserving methods and the parameters of these methods in a relatively objective manner and to show that the selected methods produce a best estimate of IBNR reserves. This increased objectivity might be particularly useful when setting loss ratios for the BF method, or the decay parameter for the GCC method, both of which in practice are usually done in a subjective manner.

On a large and stable triangle comprised of Swiss private liability claims, the best models were identified by scoring the competing models using the CDR metric, whereas on two smaller and more volatile incurred triangles, the best models were identified using the AvE metric. Since there is much less certainty about the total IBNR required for the large triangle, it is more reasonable to use the CDR metric which penalizes models that do not produce stable reserve estimates at subsequent calendar periods. On the

other hand, on the more volatile triangles where the required IBNR is much more uncertain and could change significantly as more information becomes available, the AvE metric appears to be a better choice. Although we do not report on these calculations in detail, the coefficient of variation (CoV) of the CL reserves (measured using the Mack Standard Error (Mack, 1993)) for the Swiss triangle is around 0.135, whereas the CoV is approximately 0.67 and 1.9 for the long-tail and short-tail lines respectively, i.e. the uncertainty is almost as large as the reserve estimates. Thus, considering reserve volatility could be one way of determining which of the AvE or CDR metrics should be used. Nonetheless, more research into this is required.

We briefly consider some implications of applying the framework in a practical situation. It is unlikely that an actuary would accept automatic IBNR estimates without first interrogating these in some detail. One way in which this could be done is by comparing the estimated ultimate claims and LDFs produced by the models selected using the framework, with those selected using a more traditional actuarial analysis. By reconciling between the different models, we believe extra insight into the claims development experience can be generated and an opinion regarding the optimal model for a triangle can be formed. On the other hand, when reserving using triangles for the first time, applying the framework presented here could enable the actuary to produce quickly a baseline set of reserves estimates, which can then be modified according to a traditional analysis. Finally, by selecting reserving models that are expected to minimize the AvE or CDR metrics, the amount of effort necessary to explain reserve movements and adverse development should be reduced.

The framework presented in this study could be extended in several ways. Among the most important of these extensions is to search over different loss ratios for each accident period. To achieve this, a much larger search space must be considered and the simple grid search used in this study will no longer be feasible. A simple solution might be to consider a two-step procedure, as was suggested in Section 4.2, or use more advanced techniques such as Bayesian optimization. Another avenue for future research is to investigate the longer term impact of estimating reserves using an optimal model on reserve risk capital. Finally, more investigation into criteria for defining optimal predictions for IBNR reserves is needed. Here, we have evaluated the final models based on the MSE between the predicted and actual ultimate claims. However, we have not considered the different components of the MSE, which can be decomposed into bias and variance. Future research could consider whether minimizing the bias or variance component is more optimal for actuarial reserving purposes.

Acknowledgements

We wish to thank Shane O’Dea and Euodia Vermeulen for their valuable assistance in this study and Lisa Pines for her review and comments on the manuscript. We are very grateful to Mario Wüthrich for his comments which substantially improved an earlier version of the manuscript.

References

- Bogaardt, J. (2020). *chainladder: Actuarial reserving in Python* (Tech. Rep.). Retrieved from <https://chainladder-python.readthedocs.io/en/latest/>
- 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>
- Bruce, N., Lauder, A., Heath, C., Bennett, C., Wong, C., Overton, G., ... Jenkins, T. (2015). ROC Working Party: Towards the Optimal Reserving Process. The Fast Close process. *British Actuarial Journal*, 20(3), 491–511. doi: 10.1017/s1357321715000021
- Bühlmann, H., & Straub, E. (1983). *Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio* (Tech. Rep.). Unpublished.
- England, P., & Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics*, 25(3), 281–293. doi: 10.1016/S0167-6687(99)00016-5
- England, P., Verrall, R. J., & Wüthrich, M. V. (2019). On the lifetime and one-year views of reserve risk, with application to IFRS 17 and Solvency II risk margins. *Insurance: Mathematics and Economics*, 85, 74–88. doi: 10.1016/j.insmatheco.2018.12.002
- European Union. (2009). Directive 2009/138/ec of the European parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). *Official Journal of European Union*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The Elements of Statistical Learning The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition* (Vol. 27) (No. 2). New York: Springer New York. doi: 10.1007/978-0-387-84858-7
- Gabrielli, A., Richman, R., & Wüthrich, M. V. (2020). Neural network embedding of the over-dispersed poisson reserving model. *Scandinavian Actuarial Journal*, 2020(1), 1–29.
- Gisler, A. (2015). The Insurance Risk in the SST and in Solvency II: Modelling and Parameter Estimation. *SSRN Electronic Journal*. Re-

- trieved from <https://papers.ssrn.com/abstract=2704364> doi: 10.2139/ssrn.2704364
- Gluck, S. M. (1997). Balancing Development and Trend in Loss Reserve Analysis. *Proceedings of the Casualty Actuarial Society*, 84, 482–532. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.487.9282{\&}rep=rep1{\&}type=pdf>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- IASB. (2020). *IFRS - IFRS 17 Insurance Contracts*. International Accounting Standards Board.
- Kuo, K. (2019). Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3), 97.
- 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
- Mack, T. (2002). *Schadenversicherungsmathematik 2. Auflage*. Schriftenreihe Angewandte Versicherungsmathematik, DGVM.
- Merz, M., & Wüthrich, M. V. (2008). Modelling the claims development result for solvency purposes. *CAS E-Forum*, 542–568.
- Merz, M., & Wüthrich, M. V. (2014). Claims Run-Off Uncertainty: The Full Picture. *SSRN Electronic Journal*. Retrieved from <https://papers.ssrn.com/abstract=2524352> doi: 10.2139/ssrn.2524352
- Ohlsson, E., & Lauzenings, J. (2009). The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2), 203 - 208. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0167668709000638> doi: <https://doi.org/10.1016/j.insmatheco.2009.06.001>
- Renshaw, A., & Verrall, R. (1998). A Stochastic Model Underlying the Chain-Ladder Technique. *British Actuarial Journal*, 4(4), 903–923. doi: 10.1017/s1357321700000222
- Richman, R., von Rummell, N., & Wüthrich, M. V. (2019). Believing the Bot - Model Risk in the Era of Deep Learning. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3444833
- Schmidt, K. D. (2017). A Bibliography on Loss Reserving. Retrieved from <http://www.math.tu-dresden.de/sto/schmidt>
- Struzzieri, P. J., Hussian, P. R., & Plaza, T. P. (1998). Using best practices to determine a best reserve estimate. *CAS Forum*, 10121(212), 353–413. Retrieved from <http://www.casact.org/pubs/forum/98fforum/struhuss.pdf{\%}5Cnpapers2://publication/uuid/240855E4-35A3-4EB0-BBEA-DFB00A12D2A7>
- Wüthrich, M., & Merz, M. (2015). *Stochastic Claims Reserving Manual: Advances in Dynamic Modeling* (Tech. Rep.). Retrieved from <https://papers.ssrn.com/abstract=2649057> doi: 10.2139/ssrn.2649057

- Wüthrich, M. V. (2018). Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2), 407–436.
- Wüthrich, M. V., & Buser, C. (2018). *Data Analytics for Non-Life Insurance Pricing* (Vol. 2018) (No. 17 June). Swiss Finance Institute Research Paper. Retrieved from <https://ssrn.com/abstract=2870308> doi: 10.2139/ssrn.2870308
- Wüthrich, M. V., & Merz, M. (2012). *Stochastic Claims Reserving Methods in Insurance* (Vol. 435). Hoboken: John Wiley & Sons. doi: 10.1002/9781119206262

A Triangles

Note that the triangles shown in this appendix are coloured according to the scheme mentioned in Section 3.

	EP
1979	17 684
1980	18 762
1981	22 056
1982	25 489
1983	24 819
1984	26 887
1985	31 782
1986	32 585
1987	32 726
1988	36 372
1989	36 873
1990	38 938
1991	42 109
1992	40 818
1993	43 180
1994	47 061
1995	48 428
1996	52 565
1997	52 728

Table 23: Simulated Earned Premium for Swiss Triangle shown in Table 24

	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216	228	240
1979	3 670	5 817	6 462	6 671	6 775	7 002	7 034	7 096	7 204	7 326	7 395	7 438	7 468	7 541	7 580	7 677	7 794	7 919	8 047	8 051
1980	4 827	7 600	8 274	8 412	9 059	9 197	9 230	9 277	9 331	9 435	9 864	9 621	9 654	9 704	9 743	9 745	9 745	9 748	9 749	9 749
1981	7 130	10 852	11 422	12 024	12 320	12 346	12 379	12 426	12 460	12 500	12 508	12 547	12 557	12 624	12 639	12 644	12 647	12 666	12 711	12 713
1982	9 244	13 340	13 758	13 853	13 894	13 942	13 980	14 001	14 012	14 070	14 085	14 098	14 111	14 118	14 120	14 133	14 141	14 145	14 166	14 166
1983	10 019	14 223	15 403	15 579	15 732	15 921	16 187	16 420	16 531	16 559	16 568	16 585	16 597	16 614	16 617	16 619	16 681	16 681	16 687	16 711
1984	9 966	14 599	15 181	15 431	15 506	15 538	15 906	16 014	16 537	16 833	16 951	17 038	17 040	17 195	17 298	17 307	17 308	17 308	17 309	17 310
1985	10 441	15 043	15 577	15 784	15 926	16 054	16 087	16 107	16 311	16 366	16 396	16 414	16 419	16 426	16 480	16 480	16 480	16 480	16 482	16 482
1986	10 578	15 657	16 352	16 714	17 048	17 289	17 632	17 646	17 662	17 678	17 693	17 700	17 706	17 706	17 712	17 713	17 718	17 719	17 719	17 719
1987	11 214	16 482	17 197	17 518	18 345	18 480	18 505	18 520	18 578	18 633	18 671	18 689	18 746	18 772	18 774	18 804	18 804	18 804	18 806	18 806
1988	11 442	17 621	18 465	18 693	18 882	18 965	20 464	20 522	20 558	20 607	20 611	21 250	21 257	22 505	22 509	22 516	22 518	22 522	22 523	22 527
1989	11 720	17 779	18 655	18 940	19 098	19 368	19 970	20 162	20 195	20 415	20 510	20 594	20 657	20 752	20 823	20 899	20 983	21 124	21 622	21 627
1990	13 293	20 689	21 696	22 439	22 798	23 054	23 394	23 554	23 888	24 061	24 096	24 301	24 356	24 389	24 391	24 784	24 784	24 794	24 894	24 900
1991	15 063	22 917	23 543	24 032	24 156	24 232	24 360	24 410	24 884	24 896	24 968	25 031	25 172	25 459	25 460	25 470	25 734	25 774	25 877	25 883
1992	16 986	23 958	25 090	25 392	25 546	26 089	26 129	26 149	26 166	26 231	26 328	26 743	27 023	27 048	27 075	27 168	27 234	27 276	27 386	27 392
1993	16 681	24 867	25 871	26 463	26 941	27 120	27 164	27 183	27 250	27 490	27 497	27 561	27 565	27 582	27 706	27 787	27 855	27 898	28 010	28 017
1994	17 595	25 152	26 140	27 090	27 568	27 637	27 805	28 003	28 223	28 240	28 245	28 250	28 257	28 297	28 347	28 430	28 499	28 543	28 658	28 665
1995	16 547	25 396	26 506	27 043	27 866	27 882	27 903	27 933	28 492	28 545	28 564	28 686	28 695	28 897	28 948	29 033	29 104	29 149	29 266	29 273
1996	15 449	22 702	23 909	24 690	26 083	26 525	26 567	26 640	26 695	26 801	26 814	27 018	27 078	27 269	27 317	27 397	27 463	27 506	27 616	27 623
1997	18 043	26 918	28 256	28 569	28 964	29 268	29 344	29 393	30 159	30 936	30 966	31 122	31 190	31 410	31 466	31 558	31 635	31 684	31 811	31 819

Table 24: Swiss Private Liability Triangle

	EP
2010Q1	25 728
2010Q2	25 975
2010Q3	24 954
2010Q4	24 478
2011Q1	23 899
2011Q2	21 730
2011Q3	15 232
2011Q4	11 990
2012Q1	11 731
2012Q2	11 503
2012Q3	13 317
2012Q4	14 270
2013Q1	15 193
2013Q2	15 077
2013Q3	18 804
2013Q4	17 930
2014Q1	18 825
2014Q2	20 181
2014Q3	20 891
2014Q4	23 654

Table 25: Earned Premium for Long-Tail Liability Triangle shown in Table 26

	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
2010Q1	2 014	3 925	6 027	6 477	5 624	5 159	5 214	6 241	6 705	6 716	7 373	7 696	9 011	9 230	8 409	8 448	8 824	9 202	9 388	9 350
2010Q2	2 531	4 102	3 675	3 868	4 796	6 866	6 072	6 014	5 986	5 788	5 951	5 958	6 048	6 065	5 877	5 969	5 987	5 997	5 767	5 808
2010Q3	2 373	4 426	7 274	6 514	6 699	7 063	7 603	7 727	7 130	7 655	7 949	8 587	9 113	7 730	7 925	8 276	8 180	7 821	7 822	7 821
2010Q4	3 665	6 028	5 679	5 793	5 158	5 203	4 824	4 749	5 152	5 240	5 261	5 361	5 262	5 606	5 473	5 359	6 051	5 921	6 166	6 196
2011Q1	3 685	6 009	6 655	5 535	6 292	6 972	6 793	6 362	6 510	6 457	6 734	6 836	7 817	8 117	8 602	8 508	8 709	9 368	9 671	10 319
2011Q2	5 624	9 330	8 640	7 676	7 066	6 535	7 004	6 965	7 007	7 244	7 124	7 119	7 294	7 251	6 981	7 051	6 988	7 025	7 182	6 672
2011Q3	6 360	8 485	9 147	6 628	6 883	6 990	7 417	7 768	8 117	7 775	7 908	7 946	7 905	7 504	6 981	8 369	8 323	8 495	8 286	8 255
2011Q4	3 655	4 338	4 000	4 568	6 659	6 571	6 599	6 949	7 303	7 384	8 192	8 635	8 644	9 207	9 792	9 953	10 002	9 407	9 452	8 273
2012Q1	1 203	2 897	3 289	2 418	2 197	3 144	3 153	3 073	3 109	2 367	1 892	2 083	2 505	2 678	2 594	2 648	2 661	2 521	2 529	2 370
2012Q2	1 110	2 111	3 328	5 474	5 593	5 660	4 710	4 874	4 569	4 547	4 601	4 720	5 137	5 208	5 589	6 254	6 507	6 507	6 593	6 604
2012Q3	776	3 290	6 315	7 099	7 206	6 951	6 911	6 848	6 906	6 826	7 025	7 046	7 394	10 204	9 915	9 643	9 929	10 269	10 221	10 808
2012Q4	1 858	3 064	3 965	3 906	4 058	4 270	4 019	3 867	5 107	4 984	4 972	5 328	5 752	5 802	5 554	6 083	6 530	6 523	6 565	6 570
2013Q1	1 608	4 000	4 989	4 800	4 795	5 482	4 778	4 847	4 767	5 147	5 977	5 955	6 004	6 077	6 228	6 502	6 590	6 603	7 607	7 754
2013Q2	3 120	3 906	7 002	9 182	9 680	7 916	10 125	10 614	11 271	11 901	11 493	11 493	11 778	11 796	11 238	11 358	11 099	11 107	11 450	11 413
2013Q3	1 187	1 949	3 062	4 515	4 628	4 755	4 882	4 970	4 877	5 969	6 023	6 168	6 384	7 815	7 804	7 262	8 047	8 042	8 288	8 340
2013Q4	508	1 844	4 445	6 456	7 419	7 357	7 118	7 671	7 984	8 776	9 575	9 615	9 884	9 908	7 871	8 005	7 884	7 988	7 900	7 894
2014Q1	632	1 951	2 235	4 025	5 071	5 654	6 389	7 468	7 458	7 303	7 230	7 366	7 498	8 081	8 313	8 779	8 854	9 396	9 567	10 067
2014Q2	576	3 531	8 148	10 026	10 225	10 713	11 674	11 833	11 639	12 485	12 279	12 327	11 754	11 809	11 776	11 854	11 687	11 709	11 365	11 469
2014Q3	1 549	4 807	6 178	8 651	10 214	11 283	11 203	11 552	12 697	12 468	12 828	14 163	13 962	14 114	14 932	16 581	16 279	15 668	15 625	17 347
2014Q4	1 877	4 342	4 787	6 310	7 943	8 482	8 670	8 747	9 150	9 374	9 823	10 712	11 522	11 242	12 279	13 806	13 721	13 858	14 244	16 558

Table 26: Long-Tail Liability Triangle

	EP
2010Q1	51 347
2010Q2	53 877
2010Q3	54 577
2010Q4	55 527
2011Q1	56 768
2011Q2	57 244
2011Q3	58 405
2011Q4	59 912
2012Q1	60 339
2012Q2	61 308
2012Q3	60 262
2012Q4	60 513
2013Q1	62 668
2013Q2	63 109
2013Q3	62 934
2013Q4	63 694
2014Q1	64 467
2014Q2	66 843
2014Q3	68 594
2014Q4	69 621

Table 27: Earned Premium for Short-Tail Triangle shown in Table 28

	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
2010Q1	15 886	25 178	26 572	25 520	25 047	25 025	24 962	24 448	24 412	24 305	24 305	24 305	24 285	24 201	24 159	24 171	24 108	24 097	24 097	24 097
2010Q2	18 400	24 210	22 953	22 781	22 710	22 939	22 277	22 395	22 332	22 615	22 615	22 615	22 615	22 615	22 615	22 594	22 590	22 592	22 592	22 592
2010Q3	15 765	18 860	17 972	18 029	18 100	17 915	17 919	17 802	17 854	17 854	17 854	17 868	17 844	17 829	17 829	17 762	17 762	17 754	17 754	17 754
2010Q4	18 049	24 586	24 342	23 850	23 850	23 923	23 912	23 959	23 949	23 823	23 921	23 836	23 830	23 863	23 853	23 830	23 830	23 830	23 830	23 830
2011Q1	22 716	28 799	26 890	26 890	27 203	27 014	26 813	26 542	26 364	26 523	26 259	26 257	26 051	26 001	26 001	26 001	26 001	26 001	26 001	26 001
2011Q2	17 630	24 164	23 773	23 061	22 775	22 659	22 349	22 242	22 255	22 229	22 229	22 198	22 114	22 062	22 062	22 062	22 062	22 062	22 062	22 062
2011Q3	20 852	28 044	27 322	27 051	26 750	26 293	26 137	26 133	26 310	26 224	25 943	25 952	25 942	25 940	25 940	25 940	25 940	25 940	25 928	25 928
2011Q4	21 489	32 919	32 872	32 830	32 575	32 553	32 536	32 444	32 453	32 324	32 324	32 324	32 496	32 474	32 464	32 464	32 464	32 464	32 464	32 464
2012Q1	18 254	29 871	30 149	29 904	30 122	30 282	30 004	30 047	29 793	29 748	29 722	29 755	29 766	29 766	29 765	29 764	29 764	29 764	29 764	29 764
2012Q2	21 812	29 826	28 047	27 960	28 009	27 813	27 605	27 431	27 503	27 207	27 184	27 184	27 184	27 184	27 184	27 044	27 044	27 044	27 044	27 044
2012Q3	25 062	35 859	36 143	35 174	34 845	34 032	33 252	33 128	32 079	32 282	32 480	32 317	32 442	32 413	32 454	32 454	32 561	32 565	32 565	32 565
2012Q4	27 479	41 905	42 867	42 488	42 368	41 576	41 451	40 987	41 404	41 486	41 476	41 551	41 551	41 551	41 551	41 551	41 583	41 583	41 583	41 583
2013Q1	26 771	40 484	39 968	40 058	40 011	38 823	38 481	38 506	38 508	38 516	38 527	38 511	38 511	38 511	38 511	38 777	38 729	38 727	38 727	38 727
2013Q2	20 362	26 805	26 388	26 454	26 380	26 321	26 310	26 323	26 318	26 309	26 621	26 621	26 621	26 640	26 640	26 624	26 624	26 624	26 624	26 624
2013Q3	19 766	29 065	28 267	27 856	27 475	27 402	27 365	27 263	27 244	27 246	27 257	27 246	27 246	27 246	27 259	27 259	27 259	27 259	27 259	27 259
2013Q4	27 035	43 553	45 623	45 858	45 471	45 806	44 433	44 634	44 638	44 694	44 694	44 684	44 481	44 529	44 567	44 676	44 707	44 790	44 837	44 837
2014Q1	19 600	35 432	37 811	38 379	37 947	37 692	37 686	37 692	37 634	37 634	37 612	37 612	37 612	37 600	37 600	37 600	37 600	37 600	37 600	37 600
2014Q2	22 220	35 361	38 686	38 612	37 948	38 008	38 109	38 099	38 210	38 179	38 179	38 178	38 178	38 178	38 178	38 178	38 178	38 178	38 178	38 178
2014Q3	22 583	34 819	35 282	34 872	35 781	35 467	35 473	35 477	35 450	35 663	35 628	35 558	35 558	35 558	35 564	35 564	35 576	35 576	35 576	35 576
2014Q4	25 473	35 455	36 730	37 066	36 677	36 631	36 392	36 361	36 376	36 387	36 420	36 420	36 431	36 445	36 445	36 450	36 450	36 437	36 437	36 437

Table 28: Short-Tail Property Triangle

Actuarial reserving techniques have evolved from the application of algorithms, like the chain-ladder method, to stochastic models of claims development, and, more recently, have been enhanced by the application of machine learning techniques. Despite this proliferation of theory and techniques, there is relatively little guidance on which reserving techniques should be applied and when. In this paper, we revisit traditional reserving techniques within the framework of supervised learning to select optimal reserving models. We show that the use of optimal techniques can lead to more accurate reserves and investigate the circumstances under which different scoring metrics should be used