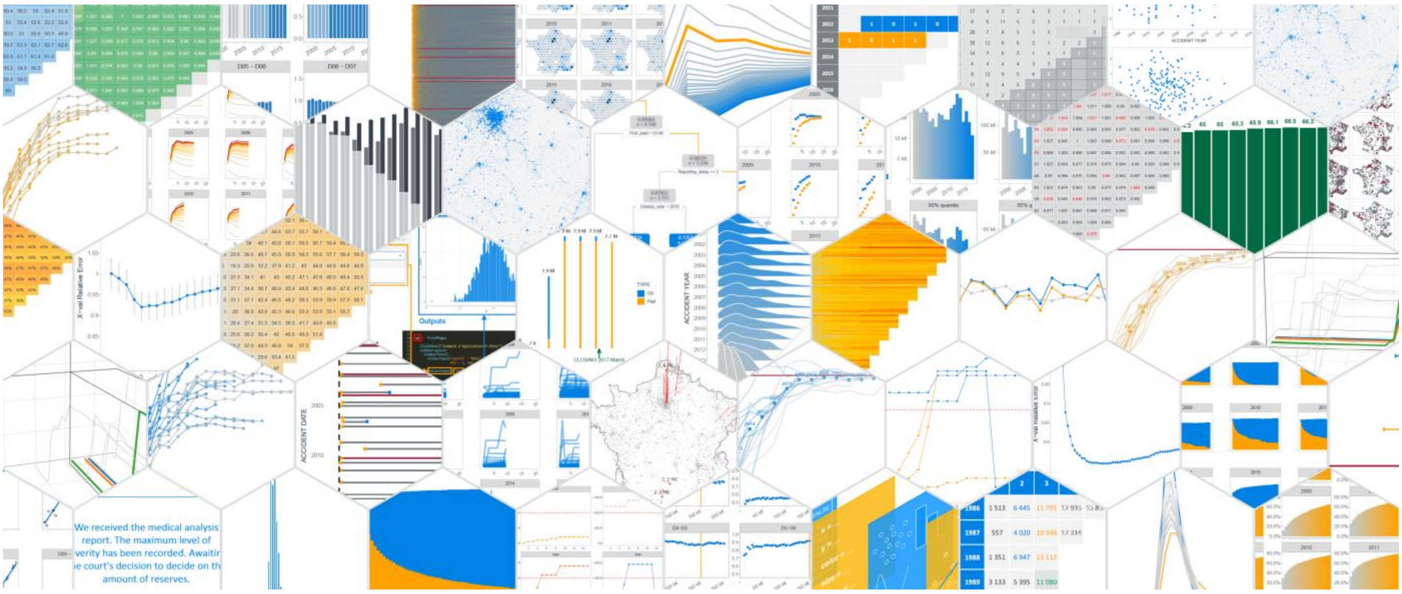
MILLIMAN WHITE PAPER

Applications of Data Science to Non-Life Reserving

Rémi BELLINA, IA
Loup ORTIZ, IA



Reserving is at the heart of the role of the actuary. Due to the strategic and financial importance that it holds in insurance companies, reserving is the focus of constant innovation and development.

The traditional techniques for the calculation of reserves, of which the Chain Ladder method is the figurehead, rely only on aggregated claims data. However, as volumes of data collected grow, innovative calculation methods which can be applied at more detailed levels of granularity are brought into focus. These innovative developments have been widely proposed by the actuarial community.

At this stage, many of these innovations are not yet sufficiently mature in the view of companies to merit implementation, particularly in view of operational constraints such as reporting deadlines and quality of data. In practice, classical methods such as the Chain Ladder, Bornhuetter-Ferguson and pure exposure methods based on loss ratios remain the key components of the reserving toolkit.

However, with the use of traditional methods, a number of challenges remain such as obtaining a deep understanding of the risk, understanding risk at the claim or policyholder level, estimation of individual large claims and automation of reserving processes (analyses of actual vs expected, sensitivity analyses or delivering results).

In this paper we examine ways in which the relatively new tools of data science can be used in conjunction with individual claim and policy data to extract more value from reserving analyses.

Data Science: A New Toolbox

Actuaries can exploit Analytics and Data Science in many ways. These can include innovative visualisations, automation of repetitive tasks or even using Machine Learning algorithms to model complex processes. These developments can benefit stakeholders throughout insurance companies.

One example of such applications is in the estimation of claims reserves. Traditionally, the techniques used to calculate outstanding claims reserves rely on data aggregated in the form of development triangles, and on the application of traditional methods such as the Chain Ladder to these triangles. Data Science shines a new light on all these analyses, all the way from the capture and management of data to the calculation of ultimate claims estimates.

How can Data Science and other new technologies bring a new reserving perspective to insurance companies ?

There is a challenge to overcome for the whole insurance industry, and its scope goes beyond the typical issues of the calculation of ultimate totals. There is no out-of-the-box Data Science solution capable of manipulating data in a variety of forms, extracting relevant information from this data, automating calculations, supporting the expert in decision making and presenting results in a clear and understandable manner. Therefore, it is up to the insurance companies to innovate.

How can the data collected at the level of individual claims be leveraged?

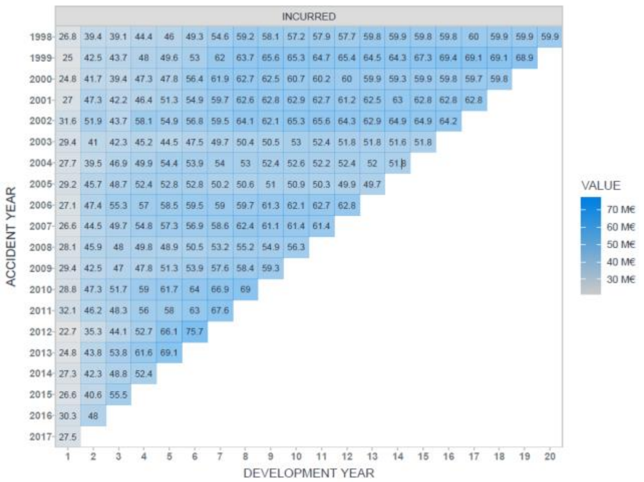
Programming languages such as R and Python offer tools to achieve these benefits. Being open-source, they are easy to grasp and make use of, and as such they have proliferated throughout insurance companies. In fact, for many companies, the tools developed in these programming languages continue the evolution of traditional practices and are now used to support decision making in a robust and pragmatic way.

Data Visualisation and Initial Diagnostics

The management and the in-depth analysis of data are essential to the reserving process. Typically, there is a wealth of claims characteristics that can be leveraged. However, extracting the value in this data is dependent on both skilful data manipulation and on the development of graphics and visualisations, which are crucial to appreciating data in both its granular and aggregated formats within reserving.

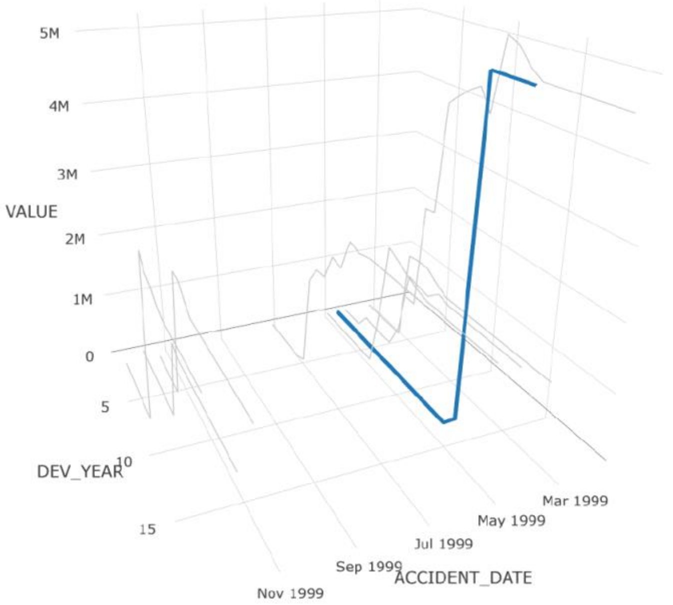
For example, visualisations of triangulated aggregated payment developments can give insight into development patterns, the underlying risks and the specific nature of the losses to the portfolio. In addition, graphical representations at the level of the individual claim make the most of the granularity of the data and allow the identification of unusual features within the portfolio. These visualisations can be built into the automated processes which exist for the production of triangles and other data extracts. Figure 1 shows how colour can add value to a traditional triangle-based analysis.

FIGURE 1 VISUALISATION OF DEVELOPMENT TRIANGLE



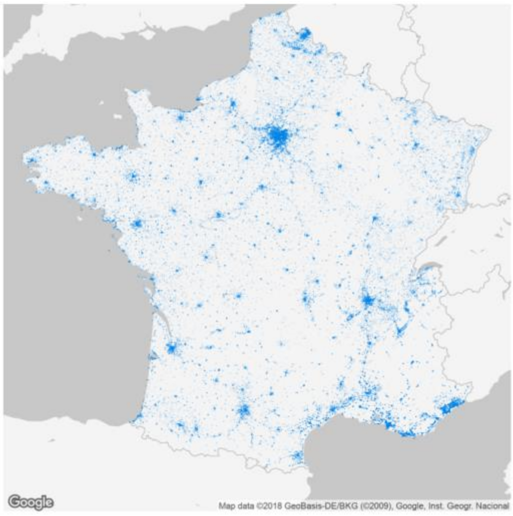
Visualisations of data at the individual policy or claim level can broaden the scope of discussions between stakeholders. One example of visualisations at granular detail can be seen in Figure 2, which shows developments in claims estimates over time.

FIGURE 2 MOVEMENTS IN INDIVIDUAL CLAIMS



Another illustration of this is the use of mapping techniques to illustrate the breakdown of losses in a portfolio within a geographical area and to identify areas of heavy concentration. An example of this is losses related to liability classes being linked to the legal system of regional courts. These geographical representations can be combined with historical data of catastrophic events, to measure their impact. Aggregating all the information on a map provides a summary from a different angle and hence represents a benefit to managers and underwriters insofar that underwriting policy may be updated in light of new findings. Figure 3 overlays claims data onto a map for this purpose.

FIGURE 3 GEOGRAPHICAL REPRESENTATION OF LOSSES



Identifying Unusual Claims and Deriving Ultimates

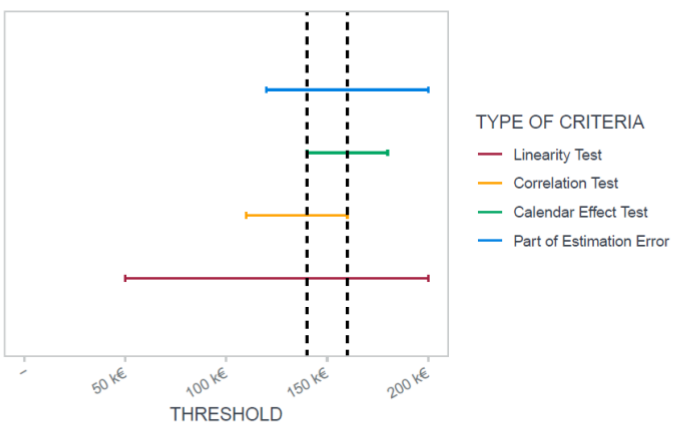
One key step in any reserve analysis is the segmentation of claims, the aim of which is to identify groups of homogeneous claims so that traditional techniques may be applied to the resulting data. The requirement for homogeneity and stability places great importance on segmentation to ensure well-fitting models.

One very common segmentation is the classification of claims between attritional losses and large losses. These two classes of claims may feature different characteristics and are, in general, treated differently when analysed. Typically, the segmentation of claims is based on a large claims threshold selected by the reserving actuary, with all claims exceeding the threshold being considered as large. In practice, there are many ways to implement such a segmentation.

Often, the selection of a large claims threshold is based on arbitrary or externally imposed constraints. However, it is possible to make the choice of threshold more objective by basing it on diagnostics which are representative of the stability of the resulting data. These diagnostics can be based on the Chain Ladder method and its stochastic variant, the Mack model.

Consequently, for a given set of data and choice of threshold, we can analyse these diagnostics to validate the assumptions of stability. These could include: linearity of losses compared to premiums or exposures, non-correlation of link ratios, calendar effects, development of the prediction error of the Mack model etc. The amalgamation of these tests identifies a range of suitable thresholds for which the required hypotheses are met. Such an approach provides a simple and objective framework as illustrated in Figure 4 overleaf.

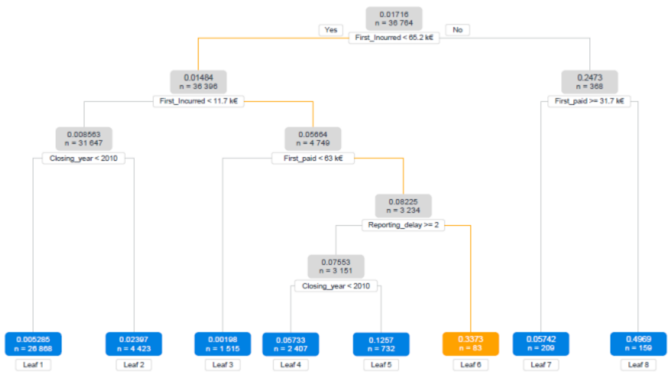
FIGURE 4 POSSIBLE RANGES OF THRESHOLDS



It should be noted, however, that any segmentation of data using a large claims threshold introduces heterogeneity towards recent attritional claims. It is possible that these could become large claims over the course of their future development. The implementation of predictive Machine Learning models with a view to identifying the claims most at risk of becoming large may help to refine the segmentation.

An example of such Machine Learning algorithms is the Classification and Regression Trees (CART) algorithm. When trained on developed historical claims, this algorithm is a flexible tool providing a segmentation which is easy to interpret. This can easily help stakeholders to understand claims with the greatest levels of uncertainty. Figure 5 below shows an example of such a CART algorithm for the segmentation of claims data.

FIGURE 5 CART ALGORITHM FOR ANTICIPATING LARGE CLAIMS



Combining the two approaches to segmentation definitively augments the information available on the risk profiles of a portfolio, taking advantage of the data on individual claims.

Similarly, the popular methods of Machine Learning allow the actuary to manage individual claims data in a flexible way to extract as yet underused information in an efficient manner. For example, claims descriptions, recorded in free text by claims managers for the most unusual claims, can be leveraged to enhance the understanding of large, uncertain claims. Sophisticated text mining methods, or simply pragmatic approaches based on regular expressions, can be used to extract pertinent information and promote dialogue between actuaries and claims managers. Figure 6 below illustrates text mining methods used to extract valuable information from unstructured text.

FIGURE 6 TEXT MINING

Beyond this, Figure 7 below illustrates how a decision tree algorithm may be able to identify homogeneous groupings of link ratios among heterogeneous data.

FIGURE 7 ANALYSIS OF LINK RATIOS BY ACCIDENT YEAR

Data Science techniques complement the reserving actuary's estimations in many ways. For example, it could be in a totally automated manner for a second opinion, or used to support modelling of segments with long, complex development patterns using individual claims data.

Artificial Intelligence: From Data to Reporting

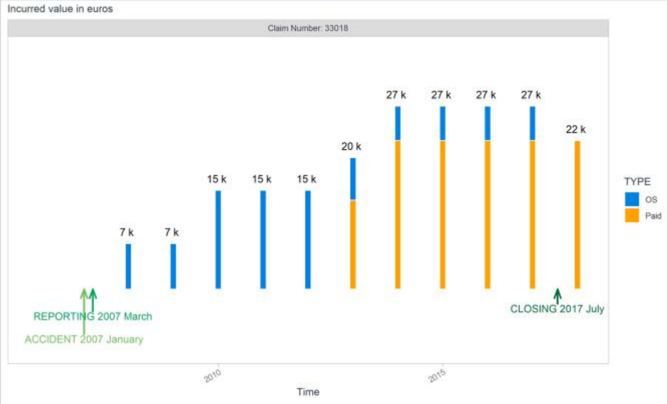
The Oxford Dictionary defines Artificial Intelligence (AI) as the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages. We will see how this theory can be applied to the challenges of selection and implementation of models for estimating claims reserves. From the point of view of Data Science, these applications can be broken down into different areas, related to automation and taking advantage of individual claims data.

The creation of diagnostics based on claims information as well as estimating ultimate claims are routine tasks within insurance companies. In this context, automation of these two procedures through decision-assisting AI represents a definite addition of value for actuaries. We note in passing that the standardisation of data which is necessary for the execution of AI provides other benefits such as reducing the amount of time spent on the processing of data so that more time can be spent on valuable analysis.

Once Data Science methods have been applied to the visualisation, data manipulation and modelling of reserves, it is essential to enrich regular reporting by presenting and interpreting the results in an effective way. For example, the inclusion of commentary or conclusions via automatic natural language generation provides an important saving of time and streamlines the description and interpretation of results.

A great deal of work is also being done on visualisations which can be applied at both the portfolio level and the individual claim level. One popular approach is the development of interactive web applications using applications such as Shiny and Dash, which have been adapted to R and Python respectively. Such applications are especially appropriate in the scope of reserving as they can bring together a range of analyses into one interface while enabling the actuary to react directly to the modelling assumptions. Figure 8 below shows an example of an automated visualisation of an individual claim with commentary generated via an algorithm.

FIGURE 8 AUTOMATIC REPORTING ON INDIVIDUAL CLAIMS



- Claim number 33018 relates to an incident in January 2007, reported in March 2007 - 2 months later.
- It has incurred €22k to date and has been subject to 2 payments for a total cumulative payment of €22k.
- Compared to the last review, there has been a decrease in the incurred of €5k.
- This claim was closed in 2017, 10 years after it was reported.
- Due to the payments and incurred amounts involved and its characteristics, it has been classified as an attritional claim.

Conclusion

Benefits of data science are typically aligned along two complementary dimensions: streamlining of processes and deep understanding of risks. Each stage of the reserving process can be reimagined from the point of view of Data Science, with support from existing actuarial techniques and expert knowledge.

The increasing volumes and granularity of data collected uncover promising opportunities. Although still being at an early stage, Data Science has the potential to transform many aspects of claims reserving.

CONTACTS
Rémi BELLINA, IA, Paris
remi.bellina@milliman.com

Loup ORTIZ, IA, Paris
loup.ortiz@milliman.com

George BARRETT, London
george.barrett@milliman.com

© 2019 Milliman, Inc. All Rights Reserved. The materials in this document represent the opinion of the authors and are not representative of the views of Milliman, Inc. Milliman does not certify the information, nor does it guarantee the accuracy and completeness of such information. Use of such information is voluntary and should not be relied upon unless an independent review of its accuracy and completeness has been performed. Materials may not be reproduced without the express consent of Milliman.