Non-Life Insurance Pricing using R

Allan Engelhardt*, Suresh Gangam

1. CYBAEA Limited, London, England
2. 64 Squares, Maharashtra, India
*Contact author: Allan.Engelhardt@cybaea.net

Keywords: Insurance, GLM, Ensemble models

Insurance have greatly benefited from adopting the R platform and leading companies are already reaping the rewards. We will show one example from non-life insurance pricing which will cover both technical implementation and business change, and we will share information on the commercial benefits obtained. By using a specific example we can keep the presentation concrete and the benefits real; however, the applicability of the approach is general and we will touch on this in the discussion.

There are many advantages of R. We will focus on two. First, R is finely balanced to allow exploratory data analysis and interactive model development while also being a platform for statistical computing and data mining. As we will show, this is key for productivity and an element to set up (bit-perfect) reproducible models.

Second, it is comprehensive in the sense that most approaches to statistics and data mining are included in the tool or its contributed packages. Among other benefits, this allows you to easily run multiple model types on your data, ensuring compatibility with classic and often robust approaches while at the same time taking advantage of the latest developments and emerging industry standards.

Non-life insurance pricing is a well-known and well-established process and yet still a critical business issue. The standard for tariff analysis is generalised linear models. We first show how to develop such a model in R, including model selection and validation. We touch upon how to deploy the model (both scoring using the model and updating the model itself) while ensuring the results remain validated and reproducible.

Next we show how easy it is to extend the model to more complex techniques. In the interest of time we jump over intermediate approaches and go straight to ensemble models, which are possibly the state-of-the-art for high-performance models.

We are in no way advocating wholesale abandonment of classical approaches for modern techniques, "black-box" or otherwise. Rather, we propose that you make use of both: continuity and understanding tempered with the results from the latest up-to-date methods. In the final part we cover some of these business issues to show how other insurers resolved them and what commercial benefits resulted. Examples include using the advanced models to restrict the validity domain of the classical approach ("risk we do not understand and will not insure") and using them to create derived variables, such as interaction variables, to extend the domain of the GLM ("understanding complex risk").