Skippie Insurance NCD Simulation

Background

Our team was asked to perform an analysis on the current No Claims Discount (NCD) scheme Skippie Insurance's car insurance customers. They have a simple discount/markup scheme dependent on the insured customers driving record – one demotion per claim, or a promotion in the absence of claims.

NCD systems were introduced in Europe in the 1960s, following the works of German Actuaries Bichsel[4], and Bühlmann[5]. They are designed to assist with customer retention, as a safe customer becomes more likely to remain with the insurer they have a NCD with. They also allow the insurer to have more capital, as the discount is retroactive.

Analysis

We conducted our simulation on R, the company's standard computation package. The actual methodology of the simulation is clearly outlined in the R notebook (Appendix 1). We instead focus on evaluating the method.

The first step involved splitting the data to get n_{ij}. This represents the number of policy holders that transitioned from state $i \to j$, where i was the NCD level in 2018, and j in 2019. Using table in R, we efficiently derived $\{X(2018) = i\} \cap \{X(2019) = j\}$ for all combinations i and j.

To use the formula for $\hat{p}_{0,1}$ from the reference paper, we initially used a nested `for` loop. This calculated the intersect of each state and output the Transition Probability Matrix (TPM). A much more readable and computationally effective way was found using the `rowsums` function.

The most critical component of our analysis was defining the `Run_MC_Sim` function and its use in simulation. This was done with high efficiency – running 1000 simulations on 10,000 observations took only a couple of minutes. The function output the mean premium directly, so we didn't need to think about the individual values of each simulation.

Efficiency was mainly from the `which(rmultinom())` function which allowed the output of a random multinomial RV to be indexed immediately. A potential source of ineffectiveness was in the next step – running the simulation on the data. We created a loop to run it on each row, rather than using parallel execution to run it on the whole column simultaneously. For more than 100k observations, we should consider using this method.

Recommendations to Improve the Scheme

The NCD system is widely used in the industry and has proven to be very effective in increasing customer retention rates. The issue here is not with NCD as a whole, but Skippie Insurance's specific implementation. With modifications to the way we forecast future claims, we can tailor the NCD levels more closely with the expected risk.

1. **Consider more than one feature**

Several studies have shown that if an insurer can only use one factor, it should be frequency of past claims. Fortunately, we are not restricted to this and so should be using a variety of significant factors in forecasting future claims.

Consider the TrueRisk Credit model[1] used by the Reinsurance Group of America (RGA). It uses credit history to model the risk of death, with high accuracy. The creators of this model demonstrate the use of factors not directly related to health that are significant in predicting the riskiness of a customer. There are examples of insurers using financial, biometric, and social factors in predicting risk of accident.

Further, we are currently only moving drivers up or down a level based on their claims on the previous year. Should we limit ourselves to past claims, we should at least use more than two years of history.

2. Different parameters for Markov Chain

The $\hat{p}_{ij}$ estimator models transition probability based on one year of transitions. A more accurate model might use the number of claims.

Consider this paper[2] which uses accidents from the past 2 years in their Markov Chain. Like us, the author simulated the number of people in each NCD level, then used the sum of their premiums to determine expected premium. Unlike us, the author modelled the *number of claims using a Poisson distribution*.

We tried a similar method; 1) Use the mean number of claims to estimate the parameter lambda; 2) Use this probability function to estimate the number of claims. Our code is in the appendix. Using this method, the expected premium is $3.33M, or ~$640k more than the Markov Chain. The conservative insurer would prefer to use the Markov Chain estimator to predict expected premium.

A further improvement would be to use each year's claims to estimate the parameter lambda for the next. If we do this however, we would need to note that a Markov Chain is no longer an appropriate model. As the probability distribution would change each year, the increments wouldn't be independent or stationary. In this case, a neural network would be more accurately able to predict the number of accidents.

In fact, a neural network would currently be a much better method. Since a customer's driving habits in one year are not independent of the previous, the *independent increments* assumption already fails.

3. Change discount levels

The current model punishes people for making claims by increasing premiums at levels -1, -2. This will lead to attrition for the same reason that discounts lead to retention. Customers on negative levels can find lower premiums elsewhere.

The insurer studied in the above paper has five discount levels up to 50%. Should the customer make a claim, they are reset to the base level. We believe this policy will be much more effective in rewarding safe drivers. Further to the above point, it would also be wise to factor customer attrition and acquisition in our model.

4. Customer Lifetime Value

Using Limiting Probabilities, we can determine that the long-term expected annual premium income for a customer will be $246. Let us also determine the expected claims cost. Where X_i is the individual claims cost, and N is the number of claims in a year:

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N] \cdot E[X] = 0.1452 \cdot \$2000 = \$290$$

This means our average costs will actually exceed the premium income we will receive. As risky customers continue to get more discounts. In practice, as customers on negative discount levels leave Skippie for other insurers, limiting probabilities will be even more skewed to the left.

Our recommendation in this case is to model the risk of each individual discount level. Then we can calculate the expected claims cost of a customer in that level. Our premium should be based on this figure, rather than an arbitrary linear discount as it is currently.

5. **Claims Expenses**

Gross premium income isn't the important factor, but rather, net premiums after claims. In 2018, we received a premium income of $ 3,004,680 and expenses of $2,888,000. Calculations (under *2018 Profitability* in the R Code in the appendix), show a 2018 profit of $116,680.

Therefore, the loss ratio for Skippie is 96% - compare this with IAG in 2009 was 74.2% and QBE with 60.3%[7]. Low profit margins like this are indicative that we are pricing our insurance policy too cheaply.

Further, we must look at the frequency and size of claims to make decisions about our premium policy. Consider the *Modern claim frequency and claim severity models* [6] article on the Journal of Cogent Economics & Finance. Findings include:

1) The cost of an individual claim is Gamma distributed
2) The wait time between claims is exponentially distributed

What we can conclude is that while premium income may be discrete, claims expenses are charged continuously. With a better understanding of this expense, we can determine probability of ruin. Premiums should be charged based on these factors.

Conclusions

The main learnings from this NCD simulation are about the importance of different risk levels in an insurance portfolio. Many of the faults with our current model come down to the fact that the current discount levels are too simplistic for accurate pricing. Most of our suggested improvements are related to understanding exactly how much a certain customer is likely to cost, and basing the premium they pay on this figure.

Specific improvements would involve capturing more data from the customers, such as behavioral, financial, and social. Running these through a regression algorithm such as a neural network would then provide a much more accurate prediction of number of claims. This will help us price our premiums and maintain adequate capital levels.

Bibliography

1. Rgare. 2019. Predictive Nature Credit Data. [ONLINE] Available at: https://www.rgare.com/knowledge-center/media/videos/predictive-nature-credit-data. [Accessed 17 July 2019].
2. Nath, Dilip & Prasenjit Sinha(2014). *A Markovian Study of NCD System of Insurance Regulatory and Development Authority and its Application*. Thailand Statistician. 12. 223-236.
3. Anderson, Theodore W., and Leo A. Goodman. "*Statistical inference about Markov chains.*" Annals of Mathematical Statistics (1957): 89-110
4. Bichsel, F., Erfahrungs-Tarifierung in der Motorfahrzeug-halfplichtversicherung, Milleilungen der Vereinigung Schweizerischer Versicherungsmathematiker, 1964; 119-129.
5. Bühlmann, H., Optimale Prämienstufen systeme, Milleilungen der Vereinigung Schweizerischer Versicherungsmathematiker, 1964; 193-213.
6. Gilenko, E., 2019. Modern claim frequency and claim severity models. Cogent Economics & Finance, IV, 51.
7. Intelligent Investor. 2019. Financial Ratios for Insurance. [ONLINE] Available at: https://www.investsmart.com.au/investment-news/top-5-financial-ratios-insurers/57809. [Accessed 17 July 2019].