Written By: Kemi Akinyemi and Benjamin Leiser

Published In: Actuarial Technology Today, May 2020

PUBLISHED ARTICLES

The Use of Advanced Predictive Analytics for Rate Making in Insurance

By Kemi Akinyemi and Ben Leiser

The Society of Actuaries Actuarial Innovation & Technology Program Steering Committee engaged Risk & Regulatory Consulting LLC (RRC) to conduct research on the intersection of U.S. insurance ratemaking and analytics techniques in advanced modern rating systems and associated regulatory considerations in the U.S. RRC performed research on the current regulatory environment and emerging issues concerning the use of advanced analytics techniques for rating and summarized that information within a report located at https://www.soa.org/globalassets/assets/files/resources/research-report/2019/insurance-regulatory-issues-us.pdf. The contents of this paper are based on the research within the report.

Companies use basic and complex modeling tools and techniques to set rates for their products within the constraints of applicable regulations, statutes and laws. The complex modeling tools rely on advanced predictive modeling techniques. The property and casualty industry appears to be further along in its use of advanced predictive analytics than the life and health industries.

Actuaries currently rely on Actuarial Standards of Practice (ASOP) and laws and regulations such as state and federal statutes as their guidelines for developing advanced models, but they are looking to regulators to provide additional comprehensive guidance. Companies are looking to understand what documentation regulators need when rates are developed using innovative methods and advanced predictive techniques.

State regulators are tasked with ensuring that the rates for the insurance products are adequate, not excessive, and not unfairly discriminatory, among other responsibilities. Regulators understand that advanced modeling techniques could vary across companies and are looking to companies to appropriately demonstrate that their use of advanced predictive analytics in determining rates are appropriate for the products offered and are not unfairly discriminatory to policyholders. Some companies that are exploring advanced predictive analytics are using them in ratemaking, while others are limiting their reliance on advanced predictive analytics until comprehensive regulatory guidance becomes available.

For any model used for rate making, the company must understand how the data is being used, how the model relies on the data for rate making, and be able to communicate this clearly to all key stakeholders, including policyholders, management and the public.

MODELING TECHNIQUES

The following are some modeling techniques that are used for rate making. Companies need to assess the risk and reward tradeoff between the tools that they consider or use.

1. Basic modeling techniques such as trending and linear regression, are still in use in some companies, especially within the health industry. In the property and casualty industry, the following are two basic univariate approaches for determining an overall rate level:
   a. **Pure-premium method:** The pure-premium method determines an indicated average rate and involves projecting the average loss and loss adjustment expenses per exposure and the average fixed expenses per exposure to the period that the rates will be in effect. The sum of those two is then adjusted for variable expenses and the target profit percentage by dividing by one minus the sum of the variable expense provision and target profit percentage to get the indicated average rate. In other words, the average

rate is the average premium per exposure. To derive the current premiums, one would multiply the average rate by the current exposures. This method is used with new lines of business where there aren't any current rates to adjust.

b. **Loss-ratio method:** The loss-ratio method compares the estimated percentage of each premium dollar needed to cover future losses, loss adjustment expenses, and other fixed expenses to the amount of each premium dollar that is available to pay for such costs. The sum of the projected loss and the loss adjustment expense ratio, and the fixed expense ratio is divided by one minus the sum of the variable expense provision and the target profit percentage to get the indicated change factor. The change factor represents the indicated adjustment to the current rates. For example, if the change factor is 1.10, this indicates that the current rates are 10 percent too low and need to be increased by a factor of 1.10. The major difference between the pure-premium and loss-ratio approaches is that the loss-ratio approach uses premium as opposed to exposure.

Basic models are still in use primarily because of their ease of design, use and explanation. Univariate methods are limited because they do not account for the effect of other rating variables. There are multivariate classification rate making techniques that consider multiple rating variables simultaneously and automatically adjust for exposure correlations between rating variables and allow for the interaction and interdependency between two or more rating variables.

2. Complex models are increasingly being used and are multi-variate in nature. Multivariate methods also attempt to remove unsystematic effects in the data (noise) and capture only the systematic effects in the data (signal) as much as possible. Multivariate techniques also produce model diagnostics and information about the appropriateness of the model fitted. The property and casualty industry appears to be further along in their use of advanced predictive analytics than the life and health industries. The following are some of the more complex models being used:

   a. **Generalized linear models (GLM):** The GLM models are one of the most common tools used because it is easier to design, has a wide range of applicability, and is relatively easier to explain to regulators and other stakeholders. The GLM is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. GLMs are based on a single regression equation whose predictions are easier to explain through regression coefficients. GLMs and other linear models focus on expressing a relationship between an observed response variable and several predictor or explanatory variables. For example, actuaries can use GLMs to determine base rates and proposed rate changes for pricing analyses. As an example, GLMs can be helpful for auto rate making because there are many factors that can be predictive of an adequate rate such as the number of accidents, credit scores, geographic location, vehicle type, age, gender and years of driving experience. GLMs can account for multiple variables at a time and more accurately than more simple models. GLMs can also be used in a reserving capacity to estimate future loss reserves for non-traditional exposures such as loyalty programs for airlines and hotels.

   b. **Machine learning:** Machine learning is an application of artificial intelligence that provides systems with the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it to learn for themselves. The results obtained from machine learning are likely to be more accurate than the results from GLM, but the machine learning model is not as easy to explain as the GLM model.

   Machine learning techniques, particularly artificial neural networks (ANNs), are increasingly popular in several disciplines, but not necessarily insurance due to a lack of interpretability. The explosion in the variety and volume of available data, coupled with cheap data storage and fast computing power, have made ANNs a key tool of data scientists. ANNs allow modeling of nonlinear processes and can be a useful tool for solving problems such as retention modeling, fraud detection, claims triage, property/casualty reserving using individual claim detail and traditional pricing models. ANNs attempt to digitally replicate the biological mechanism of the human brain and its neurons. Although ANNs have been on the upswing in a variety of fields, the insurance sector has not utilized these "brain-like" techniques on a large scale. The insurance field is still heavily skewed in favor of the more familiar and traditional data-mining techniques, such as GLMs and rule-based algorithms.

Other tools used include classification and regression trees, random forests, and neural networks, but these tend to be less transparent. Machine learning is increasingly being considered in the insurance industry. GLMs and linear models are very transparent, making it easier to explain to regulators and stakeholders than machine

learning. Although machine learning and other advanced modeling techniques could be more accurate than GLMs and linear models, it could be difficult to determine the extent to which each variable is contributing to the determined rate. As insurers consider the use of predictive analytics in the determination of rates, regulators want to ensure that the determined rates are appropriate and not unfairly discriminatory. Because the regulatory space is evolving, there is limited guidance on how the use of predictive analytics in the determination of rates could inadvertently lead to unfair discrimination and how companies can check their models and demonstrate that their models do not produce unfairly discriminatory results. Situations could arise where machine learning models unfairly discriminate against classes of individuals or promote institutional bias. Companies will need to understand what variables could be unfairly discriminatory and should be able to demonstrate that their models are compliant and not unfairly discriminatory.

MODEL VALIDATION

Effective insurance rate making heavily relies on models working effectively to capture risk and appropriately account for the key elements of rate determination. Model validation is the process of performing an independent challenge and thorough assessment of the reasonableness and adequacy of a model based on peer review and testing across multiple dimensions, including design, data, assumptions, results, and governance. Proper model validation and governance are necessary to mitigate financial model risk due to the potential negative impact of models failing to appropriately price products as designed. Through model review and validation, companies can confirm that their models are working as designed, identify limitations of their models, and manage the associated risks in the models. An effective model validation process should be periodic and include a review of the following: data, application code, assumptions and methodologies, plan code mapping, product features, controls, and model performance and outcome analysis. Some companies review their model through a "Train/Test/Validate" approach, where three different datasets are used to train the model, test the model, and then validate the model. Through out-of-time validation, the model can be validated based on a different time period to evaluate its robustness.

DATA SOURCES AND TYPES OF VARIABLES ALLOWED

The types of variables allowed to be used in rating could vary depending on the line of business. The rating variables are used to segment the insured population into different groups of similar risks for rating purposes. The criteria for selecting variables may be summarized into the following criteria categories: actuarial or statistical, operational, social, and legal.

Actuarial or Statistical Criteria: Companies typically consider the following actuarial or statistical criteria to help ensure the accuracy and reliability of the potential rating variable:

- Statistical significance,
- homogeneity, and
- credibility.

Operational Criteria: Companies also consider practical and operational constraints after identifying the statistical criteria of their variables. Practical rating variables should have the following qualities:

- Objective,
- inexpensive to administer, and
- verifiable.

Social Criteria: Companies may want to consider public perception as they identify the rating variables to be used. The following items affect social acceptability of using a particular risk characteristic as a rating variable:

- Affordability,
- causality,
- controllability, and
- privacy concerns.

Legal Criteria: The risk classification may be affected by state and federal statutes and regulations. Generally, constitutions govern statutes and statutes govern regulations. The rate classification system must comply with the applicable laws and regulations of each jurisdiction in which a company is writing business. Actuaries need to be familiar with the laws and regulations of each jurisdiction in which their company writes insurance and assure that the classification rating complies with that jurisdiction's laws and regulations. This usually requires working with other professionals, such as lawyers or regulatory compliance experts, in determining what is acceptable and what is not.

BENEFITS AND DRAWBACKS

Advanced predictive analytics techniques enable insurers to better understand their data. Proper implementation of predictive analytics techniques can improve an insurer's consistency and efficiency in product pricing and product development. The use of predictive analytics in rate making has several benefits:

- Improve pricing by increasing the number of rate segments and price points,
- efficient underwriting and pricing,
- improve competitive advantage, and
- provide insurers with a better understanding of the risks and key drivers.

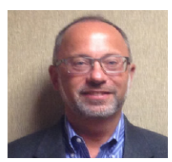
Some drawbacks of the use of advanced modeling techniques in insurance could include the following:

- Companies could have unrealistic expectations that machine learning and artificial intelligence can deliver solutions to every business problem.
- Internal models could be built by people who do not understand sound actuarial practices. Insufficient understanding of internal or vendor models could lead to setting unfairly discriminatory rates. Companies need to be able to explain their models such that the ability of consumers and regulators to understand the rating is not compromised.
- Some companies may rush to be first, while utilizing poor due diligence in the process. This can lead to unintended consequences of volatility, regulatory compliance issues, or bad headlines. Making too big of a mistake too early in the process affects the stability of the model and lengthens the time to adoption.
- Certain benefits of risk pooling could be diminished if there is over-segmentation of risks.
- Consumers may be unjustifiably penalized in the form of higher rates for irrelevant factors.

CONCLUDING REMARKS

The insurance industry appears to be trailing other industries in the use of advanced predictive models to set rates, with the property and casualty industry leading the pack. Although these complex techniques have been embraced by other industries and other sections of insurance industries, insurers appear to be cautious in the use of advanced predictive analytics for setting rates. Reasons for this may be due to limited comprehensive regulation on the use of predictive analytics for rate making and how the insurer can demonstrate that their rates are not unfairly discriminatory, as well as the resource constraints of budget, time and staff. Some of the complex models being used by insurance companies include GLMs and machine learning, with the machine learning being more precise than GLM. GLM is generally understood by the industry, but the machine learning models would typically require more explanation. The increased complexity of the models implies that companies need to be able to demonstrate that their rates are not unfairly discriminatory, and regulators need to be prepared to understand and review multiple variations of these advanced modeling techniques. Companies need to validate the models used and ensure that the data used in the model meet certain required criteria. In addition, companies will also decide their risk and reward threshold from using certain data and advanced predictive models and if the if the potential benefits outweigh the potential costs.

Kemi Akinyemi, FSA, EA, MAAA, is an actuarial consultant at Risk and Regulatory Consulting. She can be contacted at *kemi.akinyemi@riskreg.com.*



Benjamin Leiser, FSA, MAAA, is an actuarial manager at Risk and Regulatory Consulting. He can be contacted at *ben.leiser@riskreg.com.*

REFERENCE

This article is based on the Society of Actuaries report on "Insurance Regulatory Issues in the United States," and source documents are contained within that report. Society of Actuaries. 2019. "Insurance Regulatory issues in the United States" May 2019 *https://www.soa.org/globalassets/assets/files/resources/research-report/2019/insurance-regulatory-issues-us.pdf*