Casualty Actuarial Society
E-Forum, Winter 2022

The CAS *E-Forum*, Winter 2022

The Winter 2022 edition of the CAS E-Forum is a cooperative effort between the CAS E-Forum Working Group and various CAS task forces, working groups and special interest sections. This E-Forum contains a report of the CAS Machine Learning Working Party and two papers submitted in response to the 2021 Ratemaking Call for Essays, the latter of which may be presented at the CAS 2022 RPM Seminar.

Machine Learning Working Group
Daniel W. Lupton, *Chairperson*

Syed Danish Ali	Gilbert Grady	Sharon D. Mott
Ka Lap Au	Lorenzo Invernizzi	Sandra Maria N M I Nawar
Christopher Blier-Wong	Navarun Jain	Marjan Qazvini
Weiwei Cao	Soo-Jin Lee	Seth Michael Roby
Samuel David Castillo	Ryan L. Liang	Leslie Alan Vernon
Marco De Virgilis	Nathaniel Moore Loughin	
James Ely	Liam H. McGrath	

Ratemaking Committee
Ronald S. Lettofsky, *Chairperson*

Morgan Haire Bugbee	Ray Yau Kui Ho	Harsha S. Maddipati
Caryn C. Carmean	Aditya Khanna	Joshua Jacob Newkirk
James Chang	Duke D Kim	Nadiya Rudomino
Brian Chiarella	Dennis L. Lange	Sandra J. Schrader
Donald L. Closter	Shan Lin	Allan I. Schwartz
Christopher L. Cooksey	Qianru Liu	Karen Sonnet, *Staff Liaison*
Greg Frankowiak	Daniel W. Lupton	

CAS *E-Forum*, Winter 2022

Table of Contents

Report of the CAS Machine Learning Working Party

2022 Essays on Ratemaking Topics

E-Forum Committee

For information on submitting a paper to the *E-Forum*, visit https://www.casact.org/publications-research/guide-submission-e-forum.

Machine Learning in Insurance

CAS Machine Learning Working Party

Abstract: The Machine Learning Working Party of the CAS identified one barrier to entry for actuaries interested in machine learning (ML) as being the fact that published research in an insurance context is sparse. The purpose of this paper is to provide references and descriptions of current research to act as a guide for actuaries interested in learning more about this field and for actuaries interested in advancing research in machine learning.

Keywords: machine learning, modeling, predictive analytics, telematics

1   INTRODUCTION

Bolstered by improvements in computing power and innovations in key algorithms, machine learning (ML) has been experiencing tremendous growth and expansion in many fields during the last decade. However, adoption of ML algorithms by the insurance industry has been comparatively slow, and the ultimate role of ML in actuarial practice has yet to be determined.

There are several reasons for this slow pace. Commonly cited issues include lack of available computing power, lack of knowledge, regulatory scrutiny, lack of adequate data, privacy challenges, difficulty of interpretation, and communication challenges. Nevertheless, the industry has begun to overcome these challenges in some areas, leading to increased use of ML algorithms in some domains.

ML is often discussed as a unified field, though it covers many diverse and technically distinct models and algorithms and draws from many areas such as computer science, statistics, mathematics, and bioinformatics. The methods at the center of ML are united by the concept of using a very flexible and general model, applying that model to some "training" data set, and then adjusting model parameters to find a suitable optimum for a given function (typically, though not always, with the goal of minimizing a "loss function"). Broadly speaking, ML may offer advantages over Generalized Linear Models (GLMs) or other traditional statistical models, both in terms of its ability to automatically capture non-linear relationships in data to produce more accurate models and in terms of its flexibility to take on many different functional forms.
Generally, uses for ML may include:

1. Developing predictive variables to use in other methods (often referred to as "feature engineering" in ML research); this includes clustering as well as development of non-linear transformations or combinations of data elements.

2. Determining appropriate binning or clustering of variables for use in other models.

3. Dimensionality reduction for high-dimensional data.

4. Identifying non-linear relationships between variables and a predictor with minimal modeling assumptions.

5. Prediction based on sparse data sets.

6. Development of computationally tractable approximations to intractable traditional models.

In recognition of the potential for ML to provide value to actuarial science, this document seeks to provide a survey of current research in actuarial applications of ML, with a particular focus on applications of ML to Property & Casualty (P&C) insurance. This paper may therefore serve as a guide to individuals and practitioners interested in applying machine learning to a particular area, or it may serve to help researchers identify areas in which additional studies would be of greatest benefit.

2 ENHANCING TRADITIONAL METHODS

In this section, we focus on the use of ML algorithms to enhance traditional models, particularly with respect to clustering and binning of variables. This partial reliance on machine learning has the distinct advantage of reaping many of the predictive benefits of machine learning while retaining many familiar statistical tools that are useful for diagnosing and understanding models.

For example, ML models are often deterministic and non-parametric, and it is not always straightforward to calculate a probability or information criterion with an ML model. Some popular GLM software packages on the market rely on binning continuous variables rather than directly modeling continuous variables. In some cases, it is also preferable to bin continuous variables in order to explore the existence of non-linear relationships.

To these ends, Henckaerts et al. (2018) use generalized additive models (GAM) to motivate binning of continuous variables for use in GLMs. In particular, they begin by developing a model of pure premium using the flexible GAM framework to model spatial and continuous variables. The next step is binning. For spatial effects, they explore the use of four different binning methods, and

test the impact of different numbers of bins on the goodness-of-fit of the GAM. For continuous variables, Henckaerts et al. (2018) apply evolutionary trees to test sequential splits of continuous variables to maximize goodness of fit subject to a constraint that no bin can be too small. These binned categories can then be applied in the context of a more familiar GLM.

Dai (2018) similarly uses tree-based models to perform spatial clustering for application in a GLM. The paper considers Gradient Boosting Machines (GBM) and random forests as options for clustering using tree-based algorithms. Both GBMs and random forests work by combining the predictions of multiple smaller models; however, GBMs work by iteratively generating many very small (or "weak") decision trees, whereas random forests work by simultaneously generating many independent, larger decision trees. The author notes that random forests are easier to train and tune, easier to parallelize, and more robust to overfitting, but GBMs tend to outperform them in prediction if carefully tuned. The author also notes that, compared to traditional GLMs, tree-based models have no assumption of model structure, simplify the use of interactions, assist in dealing with missing values, and provide for built-in variable ("feature") selection.

Both papers might be described as "feature engineering" - using machine learning to develop independent variables (e.g., binned categories or clusters) that can be useful in prediction tasks (e.g., ratemaking or reserving). In such applications, ML algorithms offer the advantage of being extremely flexible and, often, non-parametric. These features of ML algorithms assist in revealing unintuitive and non-linear predictive relationships. Such relationships may be missed by traditional GLMs, which typically require more direct effort on the part of the modeler to intentionally choose the specific transformations of a variable rather than searching over a broad function space to find what fits best.

3   LIFE INSURANCE

While the insurance industry has been tentative about the adoption of ML, it appears that there has been more adoption of ML within the finance sector. For this reason, many proposed applications of ML within the insurance industry come from life insurance, which has some features in common with finance that may make transfer of these techniques more natural. In some cases, these techniques may also be relevant to certain domains within property and casualty insurance.

One such area is mortality risk. Mortality risk changes over time for a given population, and multiple models exist for projecting mortality risk for individual populations. However, it is

reasonable to expect that related populations might have related mortality risks, or that changes affecting one population may be related to changes affecting others in a systematic, if non-linear, way. In some instances, modeling such multi-population mortality risks may present a challenging or intractable optimization problem that requires significant judgment. Neural networks are a natural fit to address these kinds of problems.

Richman and Wüthrich (2019) used neural networks to model general mortality risk. In particular, they explored extending the traditional Lee-Carter mortality model to multiple populations simultaneously by using a deep neural network, coded in R using the Keras package. Several extensions of the traditional model were tested, as well as several variants of the deep neural network architecture. The deep neural network significantly outperformed all other models considered in the paper.

For similar reasons, valuing large portfolios of variable annuities is problematic with traditional approaches. Determining the sensitivity of variable annuities to risk factors is time-consuming or even computationally intractable with a traditional Monte Carlo simulation approach. However, machine learning can approximate these computationally expensive methods with a high degree of precision and much greater speed.

To achieve this, (Gan 2013) uses a combination of k-prototypes clustering algorithms and Gaussian process models (also known as "kriging"). K-prototypes is a method that is algorithmically similar to the well-known k-means clustering algorithm, but with a more generalized measure of "distance" between two points that can reflect distances between continuous variables and categorical variables, making it slightly more flexible.

4   SOLVENCY MONITORING

Solvency monitoring is important for insurance companies and insurance regulators to make sure that companies can meet their obligations as they fall due. The National Association of Insurers Commissioners (NAIC) has developed the property-liability Risk-Based Capital (RBC) system to calculate the amount of capital that insurance companies need to hold relative to their retained risk. Similarly, the Solvency II Directive in the European Union requires EU insurers to hold a minimum amount of capital. RBC systems calculate the regulatory capital requirement by assessing risks such as credit risk, underwriting risk, market and operational risk, and allowing for inter-dependency

among these risks. Companies with RBC ratios below certain thresholds are subject to different degrees of regulatory intervention.

These solvency capital requirements may be re-framed in the context of machine learning as (linear) decision boundary problems. The amount of capital held by a company relative to their required solvency capital is a one-dimensional decision boundary. Similarly, systems like the Insurance Regulatory Information System (IRIS) might be seen as using a multi-dimensional decision boundary. By leveraging large amounts of data and capturing non-linearities, machine learning methods may be better able to make accurate predictions, even in consideration of the dimensionality problem for solvency modeling.

Support Vector Machines (SVMs) have been explored as one promising option for binary classification of companies based on solvency. SVMs work by automatically finding a dividing (hyper-)plane that separates solvent companies from insolvent ones. SVMs can have a linear decision boundary (i.e., a line above which companies are solvent and below which they are insolvent), or they can have an effectively nonlinear boundary by mapping inputs into a higher-dimensional feature space using a "kernel method" or "kernel trick" and finding a dividing hyperplane in the higher-dimensional space (which may correspond to a nonlinear boundary in the lower-dimensional space).

Salcedo-Sanz et al. (2004) use SVMs in combination with simulated annealing and Walsh analysis to perform feature selection on a set of 19 financial variables for predicting insurer insolvency. Their analysis suggests that a limited subset of just five financial ratios are sufficient to evaluate insurer solvency. Tian et al. (2019) rely on a different approach to SVMs, instead using a non-kernel fuzzy quadratic surface SVM applied to financial ratios and macroeconomic factors. This "non-kernel" method attempts to overcome limitations of kernel methods, particularly the fact that results of an SVM may be sensitive to the choice of kernel function, and no universal method exists for selecting the best kernel function. This method achieves promising results at the cost of higher dimensionality, which may make the problem computationally expensive depending on the number of variables used.

Random forest algorithms have been used to predict insolvency within the insurance industry (as in Kartasheva and Traskin 2013) and in more general business settings (as in Behr and Weinblat 2017). Compared to SVMs, random forests can handle binary data without additional assumptions, whereas SVMs require the user to specify a notion of "distance" between the binary data points.

Random forests can be designed to provide probability of insurer default rather than a binary classification of "default" or "healthy." Random forests can also be used to automatically rank the importance of variables to determining insurer solvency. They have an advantage over statistical approaches like logistic regression in their ability to automatically detect and model highly non-linear relationships. However, random forests may be challenging to interpret, and may produce more highly variable results, particularly with sparse data sets.

5 RESERVING

5.1 Individual Claim Reserving

Estimating future claim payments is one of the main tasks performed by actuaries on a daily basis. Such estimates are of high value for the insurance companies because they constitute one of the largest liabilities on the balance sheet. It follows that the accuracy and timing of these figures is of primary concern for all stakeholders.

Traditionally, actuaries employed classical methodologies to perform this task. Such methodologies of estimating claim liabilities rely on aggregate data of insurance claims. These approaches have the advantage of relative simplicity, making them easy to communicate to stakeholders. In addition, by aggregating loss information from many claims, these methods may provide more stable results; however, this stability belies the uncertainty inherent in the classical loss projections. Specifically, less mature years may be subject to considerable uncertainty. Accurate estimates of loss reserves may not be available for months or years after they are incurred. This contributes to uncertainty in reserve estimates and, consequently, in the profit of the company and in the available funds. This could lead to delay in important strategic decisions, which could significantly harm profitability and market share.

Machine Learning methods could fill this gap, providing an accurate estimate at a very early stage of the claim indemnification process for individual claims. In addition, ML methods make it possible to take advantage of all available claim information, unlike standard triangle-based methods that only employ information about the timing and amount of claims. Using this additional information can reduce uncertainty in claim estimates, particularly for immature claims where triangle-based methods have comparatively little data on which to base their estimates.

Moreover, ML methodologies are fully flexible, and allow actuaries to consider (almost) any kind of feature information. We are, in fact, not limited to fixed data structures (e.g., triangles, which only

provide insights about claim amount, timing, and development). As an example, ML can mine claim description text data to generate new features that can improve model predictiveness.

Another advantage of ML techniques is that such algorithms can operate without extensive user assumptions, as ML techniques can estimate most parameters of interest from the data. In addition, they can update/retrain themselves automatically. ML algorithms can also be deployed in an automatic way in order to achieve an instant estimated ultimate amount when the claim is first reported.

As a result, ML methods can provide considerable savings in terms of both time and money. Processes such as claim triage can be performed automatically and in a fraction of the time. It is also important to note that ML methodologies can adjust and adapt to changes in observations. ML methods can assist in identifying, studying, and reacting to trends more quickly than traditional methods because such trends can be discovered automatically.

Common methods and algorithms that hold promise for individual claim reserving include Neural Networks, Gradient Boosting, and Classification and Regression Trees (CART).

Wüthrich (2016) is a good introduction to individual claims reserving with machine learning. This paper provides a toy example using regression trees, with several simplifications, in order to introduce and prove the concept of ML in individual claim reserving. The author notes that although this paper presents a very simplified model, there would be little difficulty in generalizing it to the full real world problem.

Jamal et al. (2018) presents five different ML techniques applied to forecast individual claim development. They were implemented in a cascading triangular way similar to triangular reserving methods, and the prediction results were compared with results achieved by classical reserving methods. The findings offer a better understanding of the possible complexity of the nature of the claims, point out some weaknesses that traditional methods might have, and indicate a strong potential for machine learning algorithms.

De Virgilis and Cerqueti (2020) analyze a set of fully developed claims in order to predict the ultimate cost with the aid of ML techniques. The authors employ several ML methods highlighting strengths and weaknesses of each one. On a high level the target is to predict the ultimate cost of claims on the individual level at the moment they are reported to the insurance company, alongside the intermediate cashflows before the claim is closed. The paper also contains a section in which

methods to predict claims with no payments (CNP) are implemented. A final paragraph is dedicated to the topic of predicting the amount of Incurred But Not Yet Reported claims (IBNYR), a concept often overlooked in the context of individual claim reserving.

5.2 Aggregate Reserving

Notwithstanding the promise of individual claims reserving, machine learning can also provide improvements to aggregate reserving methods by contemplating additional claims information and by capturing uncertainty in claim reserves. Machine learning models that have been used for this purpose include neural networks, random forests, gradient boosting machines, boosted Tweedie models, and Gaussian process regression models.

5.2.1 Neural Networks

Neural networks can be implemented to enhance classical methods such as Mack's chain ladder model and the Over-Dispersed Poisson generalized linear model or to directly model incremental paid losses and outstanding case reserves.

Classical aggregate claims reserving assumes a homogeneous claims portfolio. (Wüthrich 2018) starts with Mack's chain ladder model of cumulative paid claims by accident year and further refines it for heterogeneity and individual claims features using a neural network. The neural network incorporates claim information including line of business, labor sector, accident quarter, age, and affected body part. The neural networks employed are relatively simple and contain only one hidden layer, with the specific parameters and architecture differing for each development period. The paper helpfully discusses important data pre-processing steps and methods of tuning hyper-parameters - in this case using out-of-sample validation. The paper demonstrates the promise of ML techniques for individual reserving. Neural networks are more accurate in three out of the four lines of business tested.

The model presented by (Gabrielli, Richman, and Wüthrich 2019) embeds a cross-classified Over-Dispersed Poisson ("ODP') reserving model of incremental paid claims into a neural network architecture using the Keras package in R. This approach goes beyond individual portfolios to learn representations simultaneously over several portfolios. The resulting model for each individual line of business is more accurate in all lines of business at the price of slightly higher prediction uncertainty as measured by bootstrapping. Multi-triangle representation learning requires an additional line-of business-dependent embedded layer. The multi-triangle model gives the best

results with respect to the true reserves, and prediction uncertainty is reduced over individual lines of business models.

(Kuo 2019) considers another approach. In Kuo's paper, the DeepTriangle model combines paid losses and claims outstanding to estimate loss reserves using a deep neural network, also coded using Keras in R. The neural network architecture uses incremental paid and outstanding case reserves by accident year as inputs and outputs incremental paid and outstanding for cells in the run-off triangle. Between the input vector and output vector lie hidden layers of the neural network with associated intermediate values, activation functions, biases, weights, and a mean absolute percentage error loss function. Fifty companies are modeled simultaneously on Aggregate Schedule P data between 1988 and 1997, covering 10 years of development for four lines of insurance: commercial auto liability, personal auto liability, workers compensation, and other liability.

Validation of the model across lines of business shows it meets or exceeds the predictive accuracy of existing stochastic methods. Out of time performance is compared to the Mack chain-ladder method, the bootstrap overdispersed Poisson method, and a selection of Markov Chain Monte Carlo models. DeepTriangle achieves the lowest mean absolute prediction error and root mean squared prediction error except for the commercial auto line. Future extensions of the model may include outputting distributions of loss reserves, rather than point estimates.

5.2.2 Gaussian Process Regression

A common criticism of link ratio and regression based models is that they tend to be heavily parameterized for a problem with few degrees of freedom.

(Lally and Hartman 2018) propose a hierarchical Bayesian Gaussian process model, a flexible nonparametric statistical/machine learning method which provides a robust and smooth fit to a wide variety of data types, structures, and distributions. Gaussian process regression is a method of smoothly interpolating and extrapolating while capturing uncertainty in predicted data points based on parameters inferred from the data points' relationship to each other.

Gaussian process regression has many favorable features: it is non-parametric, it has implementations in many standard software packages (for instance, Stan), it has a probabilistic interpretation, and input warping automates feature engineering. The existence of a probabilistic interpretation allows one to develop a posterior predictive distribution for predicted data points and

thereby determine the uncertainty in estimated values (or the uncertainty arising from potential measurement error in observed values).

6   INSURANCE FRAUD DETECTION

Fraud is dramatically increasing with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year.

Although prevention technologies are the best way of reducing fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed.

Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunication fraud, and computer intrusion, to name but a few.

We describe the tools available for statistical fraud detection and the areas in which fraud detection technologies are most used.

Fraud detection is particularly important in health insurance as it can significantly impact healthcare costs. Bauder and Khoshgoftaar (2017) compare the performance of supervised, unsupervised and hybrid machine learning approaches to predict fraud for the 2015 Medicare data. While the conclusion is not surprising - that supervised learning outperforms the other approaches - labeled fraud datasets may not be widely available at insurance companies, and label quality may be poor. As a result, unsupervised methods may provide at least some degree of useful information for motivating further investigation. In another paper, Bauder, Herland, and Khoshgoftaar (2019) evaluate the predictive power of ML methods using both two separate data sets, i.e., training and validation data sets and cross-validation sets in which a single data set is divided into smaller training and test subsets and find that the former provides a more realistic picture of the real-world model. For a review of statistical fraud detection and other applications of ML in health insurance, see, for example, (Bolton and Hand 2002) and (Mehta, Katz, and Jha 2020), respectively.

Similar to the application of ML in insurance that raises concern regarding regulation, the application of ML in health care raises ethical issues which are addressed in (Char, Shah, and Magnus 2018).

7   TELEMATICS

The use of telemetry devices has been growing in auto insurance as more carriers introduce telematics products and as more data is collected. Data collected through telematics devices promises to be useful in a variety of insurance applications: pricing, underwriting, claim response and handling, fraud detection, maintenance recommendations, and more.

Pricing is an especially promising application, as telematics variables can accurately capture a driver's behavior (presumably the proximate cause of many claims). This could potentially replace proxy variables like gender or household size. In reality, many programs have yet to actually incorporate telematics into their pricing; gathering the data and analyzing it have been an important first step. Because raw telematics data must be organized into useful predictors, some early research has focused on using machine learning techniques to extract the most information from this raw data.

There are two main flavors of telematics that have emerged – pay-as-you-drive (PAYD) and pay-how-you-drive (PHYD). The main distinction is that PAYD focuses simply on driving habits (e.g., distance, time of day, location), while PHYD adds information about driving style (e.g., speed, braking). Depending on the specific hardware/software used, the raw data collected varies. In general, it includes basic PAYD information such as number of trips, distance and duration of trips, and time of day. It may also collect data about GPS location, speed, acceleration, and turns every second; or it may gather information on road types and conditions. Insurers are challenged to turn this high-frequency, high-dimensional data into useful covariates for predicting loss costs. Several studies have experimented with different techniques for developing these covariates.

Verbelen, Antonio, and Claeskens (2018) is an early study focused on PAYD variables. The authors compared several approaches for modeling claims frequency. Their goal was to evaluate the predictive power of telematics variables as well as compare traditional (time) and telematics (distance) exposure measures. They created four different datasets using third-party liability data from a Belgian insurer: one with only traditional rating variables (e.g. driver age, gender, postal code, and vehicle age) and time as exposure; one with only telematics data (e.g. yearly distance, number of trips, distance on road types, time of day, and day of week) and distance as exposure; one with a combination of traditional and telematics variables and time as exposure; and one with a combination of traditional and telematics variables and distance as an exposure.

The authors used generalized additive models (GAMs) as their framework. GAMs allow for more flexible relationships between continuous predictors and response, i.e., non-linear effects. Since some of the telematics variables are compositional (i.e., proportions of different categories that sum to 1), the authors developed a novel approach to include these as predictors. Based on the model results, the authors found: that including telematics variables improved the model (but time as exposure was preferred); that differences in gender were explained by driving habits (women drove fewer miles per year); and that time of day and type of road were predictive – driving in the evening is riskier and driving on urban roads or motorways was riskier than other roads.

A subsequent series of papers investigates PHYD variables and various approaches to corral this high-dimensional data. (Wüthrich 2016) started with an idea of visualizing a driver's style by plotting speed-acceleration (v-a) heatmaps. These heatmaps show the distribution of time spent at each combination of speed and acceleration. Drivers can be compared by calculating the dissimilarity between their heatmaps (each pixel of the heatmap has a value that can be used to calculate distance). The author used K-means to categorize similar drivers.

Because the categorical variable resulting from K-means is less desirable than a low-dimensional continuous variate, (Gao, Meng, and Wüthrich 2019) explored two additional techniques for turning the v-a heatmaps into useful predictors. The authors compared two Principal Components Analysis approaches to the original K-means approach. The first PCA approach was Singular Value Decomposition, which is restricted linearly; the second, based on a neural network, gives a non-linear analog to PCA. The authors found the SVD and NN approaches were able to provide sufficient representations of the information in the v-a heatmaps. The benefit of using these techniques in place of K-means is that continuous predictors require fewer parameters than categorical ones, which leads to less over-parameterization; and with continuous variates, new data could be simulated.

(Gao, Meng, and Wüthrich 2019) compared the predictive power of the three techniques described above (K-means, PCA, and bottlenecked neural network) in predicting claims frequency. The authors used a Poisson Generalized Additive Model (GAM), which allows for non-linear covariate effects. All three approaches improved the out-of-sample deviation of a Poisson GAM predicting claims frequency with initial covariates of driver age and vehicle age, with PCA and the neural network outperforming K-means. The authors believe PCA and NN approaches were more successful because they are numeric instead of categorical and therefore more granular. They found

that the three approaches accounted for some of the same information. Because most accidents occur at low speeds, the authors focused on low speed intervals and longitudinal (straight-line) acceleration rates. The resulting principal component and bottleneck activation features correspond to safer driving (lower acceleration) at low speeds (5-10 km/hr or about 3-12 mph). For severity modeling, the authors note that high speed intervals and lateral acceleration rates should be investigated.

8   SUMMARY

In this report, we considered applications of ML in different areas of insurance. All papers that we discuss confirmed the strong predictive power of these tools compared with traditional methods. However, there are still some concerns among practitioners. These include the interpretability of the results, regulatory, and ethical issues. Although these techniques attracted the attention of many researchers, there is not much research that addresses these concerns. It is the view of this working party that working in these areas can facilitate the use of such strong tools in the insurance industry.

9   DISCLAIMER

While this paper is the product of a CAS working party, its findings do not represent the official view of the Casualty Actuarial Society. Moreover, while we believe the approaches we describe are very good examples of the use of machine learning techniques in various insurance contexts, we do not claim they are the only acceptable ones.

10   REFERENCES

[1]   Bauder, R. A., M. Herland, and T. M. Khoshgoftaar. 2019. "Evaluating Model Predictive Performance: A Medicare Fraud Detection Case Study." In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (Iri)*, 9–14. https://doi.org/10.1109/IRI.2019.00016.

[2]   Bauder, Richard A., and T. Khoshgoftaar. 2017. "Medicare Fraud Detection Using Machine Learning Methods." *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 858–65.

[3]   Behr, A., and Jurij Weinblat. 2017. "Default Patterns in Seven Eu Countries: A Random Forest Approach." *International Journal of the Economics of Business* 24: 181–222.

[4]   Bolton, Richard J., and David J. Hand. 2002. "Statistical Fraud Detection: A Review." *Statist. Sci.* 17 (3): 235–55. https://doi.org/10.1214/ss/1042727940.

[5]   Char, Danton S., Nigam H. Shah, and David Magnus. 2018. "Implementing Machine Learning in Health Care - Addressing Ethical Challenges." *New England Journal of Medicine* 378 (11): 981–83. https://doi.org/10.1056/NEJMp1714229.

[6]   Dai, Jie. 2018. "Enhancing the Generalized Linear Modeling Approach with Machine Learning Technique." *Casualty Actuarial Society E-Forum* 2: 1–9.

[7]   De Virgilis, Marco, and Pierluigi Cerqueti. 2020. "Estimation of Individual Claim Liabilities." *CAS Working Party.*

[8]   Gabrielli, Andrea, Ronald Richman, and Mario Wüthrich. 2019. "Neural Network Embedding of the over-Dispersed Poisson Reserving Model." *Scandinavian Actuarial Journal* 2020 (July): 1–29. https://doi.org/10.1080/03461238.2019.1633394.

[9]   Gan, Guojun. 2013. "Application of Data Clustering and Machine Learning in Variable Annuity Valuation." *Insurance: Mathematics and Economics* 53 (November): 795–801. https://doi.org/10.1016/j.insmatheco.2013.09.021.

[10] Gao, Guangyuan, Shengwang Meng, and Mario V. Wüthrich. 2019. "Claims Frequency Modeling Using Telematics Car Driving Data." *Scandinavian Actuarial Journal* 2019 (2): 143–62. https://doi.org/10.1080/03461238.2018.1523068.

[11] Henckaerts, Roel, Katrien Antonio, Maxime Clijsters, and Roel Verbelen. 2018. "A Data Driven Binning Strategy for the Construction of Insurance Tariff Classes." *Scandinavian Actuarial Journal* 2018 (February): 1–25. https://doi.org/10.1080/03461238.2018.1429300.

[12] Jamal, Salma, Stefano Canto, Ross Fernwood, Claudio Giancaterino, Munir Hiabu, Lorenzo Invernizzi, Tetiana Korzhynska, Zachary Martin, and Hong Shen. 2018. "Machine Learning and Traditional Methods Synergy in Non-Life Reserving." http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/539b8ff8c7848d2ac12582d6001e53b8/$FILE/ASTIN_MLTMS%20Report_SJAMAL.pdf.

[13] Kartasheva, A., and Mikhail Traskin. 2013. "Insurers' Insolvency Prediction Using Random Forest Classification." In.

[14] Kuo, Kevin. 2019. "DeepTriangle: A Deep Learning Approach to Loss Reserving." *Risks* 7 (September): 97. https://doi.org/10.3390/risks7030097.

[15] Lally, Nathan, and Brian Hartman. 2018. "Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping." *Insurance: Mathematics and Economics* 82 (C): 124–40. https://doi.org/10.1016/j.insmatheco.2018.

[16] Mehta, Mehul C., Ingrid T. Katz, and Ashish K. Jha. 2020. "Transforming Global Health with Ai." *New England Journal of Medicine* 382 (9): 791–93. https://doi.org/10.1056/NEJMp1912079.

[17] Richman, Ronald, and Mario V. Wüthrich. 2019. "A Neural Network Extension of the Lee-Carter Model to Multiple Populations." *Annals of Actuarial Science*, 1–21. https://doi.org/10.1017/S1748499519000071.

[18] Salcedo-Sanz, Sancho, Mario DePrado-Cumplido, María Jesús Segovia-Vargas, Fernando Pérez-Cruz, and Carlos Bousoño-Calzón. 2004. "Feature Selection Methods Involving Support Vector Machines for Prediction of Insolvency in Non-Life Insurance Companies." *Intelligent Systems in Accounting, Finance and Management* 12 (4): 261–81. https://doi.org/https://doi.org/10.1002/isaf.255.

[19] Verbelen, Roel, Katrien Antonio, and Gerda Claeskens. 2018. "Unravelling the Predictive Power of Telematics Data in Car Insurance Pricing." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67 (5): 1275–1304. https://doi.org/https://doi.org/10.1111/rssc.12283.

[20] Wüthrich, Mario. 2018. "Machine Learning in Individual Claims Reserving." *Scandinavian Actuarial Journal* 2018 (January): 1–16. https://doi.org/10.1080/03461238.2018.1428681.

[21] Wüthrich, Mario V. 2016. "Machine Learning in Individual Claims Reserving." Swiss Finance Institute Research Paper Series 16-67. Swiss Finance Institute. https://ideas.repec.org/p/chf/rpseri/rp1667.html.

Biographies of Working Party Members

Marco De Virgilis is a Senior Actuarial Data Scientist working for Allstate in the Chicago Office where is currently working on developing ML solutions to actuarial problems in both ratemaking and reserving. After achieving his MSc in Actuarial Sciences, he worked as a reserving actuary for Direct Line Group in the London office reporting on both personal and commercial lines. Following this experience, he worked as a consultant for Deloitte. During this time, he developed a more thorough knowledge of the actuarial market and industry practices. He then worked in the R&D department as a Data Scientist for Unipol, one of the biggest insurance Italian players. During this experience he developed analytics skills linking standard actuarial practices and modern frameworks.

Daniel Lupton is Vice President and Consulting Actuary at Taylor & Mulder, Inc, where he is responsible for reserving, pricing, and risk-focused financial examination work. Mr. Lupton is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, a Certified Specialist in Predictive Analytics, and holds a Master's in Business Administration from the Robert H. Smith School of Business. Mr. Lupton is Chair of the Machine Learning Working Party and sits on the Ratemaking Committee and the Committee on Risk.

Liam McGrath is a Senior Consultant and Data Scientist in Willis Towers Watson's Insurance Consulting and Technology (ICT) practice. His focus is helping insurers build predictive modeling and artificial intelligence solutions for pricing, underwriting, and claim functions. Liam received a Bachelor of Science degree in mathematics from Carnegie Mellon University. He is an Associate of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and a Chartered Property Casualty Underwriter. He currently serves on the Machine Learning Working Party. Liam has contributed to presentations and publications in various insurance forums on topics including natural language processing, model interpretability, and ethics of artificial intelligence.

Marjan Qazvini graduated with a PhD from the University of Melbourne. She is a lecturer since 2017 and currently is working at Heriot-Watt University, Dubai. Her research areas are Actuarial Modelling, Risk Theory, Long-term care, Copula Modelling and climate change.

Seth Roby is a Pricing Actuary at Berkley Mid-Atlantic Group (a WR Berkley Company), where he is responsible for ratemaking, state filings, and operational data analysis. Prior to joining Berkley, he worked for Markel Corporation in a variety of pricing, reserving, and strategic roles. Mr. Roby is an Associate of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and holds a Master's in Business Administration from the Virginia Commonwealth School of Business.

Leslie ("Les") Vernon is a Senior Consulting Actuary with Perr&Knight, telecommuting from Atlanta, Georgia as an extension of their Santa Monica, California office. He has 30 years of experience in the property-casualty insurance industry in the areas of ratemaking and reserving. Mr. Vernon graduated from the University of California at Los Angeles with a Bachelor of Science in Mathematics and earned a Master of Science in Statistics degree from the University of California at Irvine. He is a Fellow of the Casualty Actuarial Society and a Member in the American Academy of Actuaries. He has served on the Exam Committee of the Casualty Actuarial Society, and he

currently serves on the Machine Learning Working Party. He is also presently a mentor in the Physical Sciences Undergraduate Mentoring program at the University of California at Irvine.

Decision Trees and Categorical Independent Variables with Many Levels

Chao Guo

Abstract

This paper discusses a common concern regarding decision tree models, which is categorical independent variables with many levels. By dissecting the decision tree algorithm and running numerical simulations in R, we conclude that categorical variables with many levels do not always cause trouble, and any preprocessing on them should consider sample size, correlation among response and independent variables, and many other factors.

Keywords. decision tree; machine learning; categorical variables; credibility.

1. INTRODUCTION

Decision trees and other tree-based machine learning algorithms have gained popularity in the actuarial community. Due to their complexity and opaqueness, many practitioners rely on certain "rules of thumbs" to build their models. In this essay, we examine one particular well-circulated advice: categorical variables with many levels can cause trouble for tree-based models, so practitioners have to pre-process their data and trim down the number of levels. It turns out that there is much subtlety to this advice.

2. BACKGROUND AND METHODS

Decision trees and their related machine learning algorithms (e.g., random forest, gradient boosting tree) have gained popularity in the actuarial community in the past few years. This is partly due to the fact that tree-based models are capable of modeling more complicated relationship between the response and the independent variables. Also, computing power has become more accessible, and open source software has become more mature and reliable.

However, unlike generalized linear models (GLMs), tree-based models are often viewed as "black-box models". Indeed, tree-based models do not have coefficient estimates, so it is tricky to examine marginal effect of each individual independent variable. Their statistical properties are not as well-studied as GLMs either, making it difficult to perform statistical inference using tree-based models.

Perhaps because of the scarcity of reliable resources, practitioners often rely on and, in some cases, follow well-circulated "rules of thumb" without any question. This habit has two issues. First, many of the "the rules of thumb" simply do not hold. Second, even if they do hold, their context often gets lost. As a result, practitioners might be misguided and create a biased rating plan that is hard to explain

to business partners as well as regulators.

One of the common beliefs is this: categorical variables with many levels can cause trouble for tree-based models. As a result, practitioners often feel compelled to do some kind of pre-processing, such as one-hot encoding. But how many levels is too many? Why does large cardinality cause trouble? Does large cardinality always cause trouble?

2.1 Decision Trees Revisited

To answer these questions, we first need to understand how a decision tree makes a split. For a detailed description, please refer to section 9.2 of reference 2. We only provide a sketch here.

For each variable X_i in $X_1, X_2, \cdots, X_p$:
1. If X_i is numeric and has distinct values $x_1, x_2, \ldots, x_{i_m}$
    a. Examine all partitions $L(x_{i_k}) = \{y_i \mid x_i \le x_k\}$ and $R(x_{i_k}) = \{y_i \mid x_i > x_k\}$
    b. Choose the x_{i_k} that minimizes the impurity of partition $L(x_{i_k}), R(x_{i_k})$ to make a split.
2. If X_i is categorical and has levels a, b, c
    a. Generate a subset $\{a\}$ and its complement $\{b, c\}$ from all levels. Examine partition on the response $L = \{y_i \mid x_i \in \{a\}\}$ and $R = \{y_i \mid x_i \in \{b, c\}\}$. Compute the impurity of this pair L, R.
    b. Generate another subset $\{b\}$ and its complement $\{a, c\}$ and do the same thing.
    c. Repeat the process with all possible subset and its complement.
3. Choose the variable that gives you the partition with the smallest overall impurity.

If we examine the process of growing a decision tree, we see that, in fact, it is not that the cardinality of X_i that matters. Rather, it is the number of different partitions on the response variable the tree can induce matters. Specifically, if X_j is a categorical variable with q levels, then it can induce 2^{q-1} different partitions on the response y_i's. On the other hand, if X_j is numeric with q distinct values, then it induces only q partitions. As a side note, that is why randomForest package in R cannot handle categorical variables with more than 32 categories. A categorical variable with 32 categories induces 2,147,483,647 partitions on y_i's. See reference 3.

From this perspective, if X_i has enough levels, the tree may tend to include X_i to the model's detriment. However, that is not the full story. We also need to consider how strong the relationship is between the response and other variables, say, X_j.

2.2 A Simulation Study

For example, assume that the true model is $Y = 3X_1$ without any noise term. In this case, even if we feed the decision tree with an irrelevant X_2 with 100 levels, we will reach a tie and the decision tree will split using X_1 50% of the time. On the other hand, if the true model now includes a noise term $Y = 3X_1 + \epsilon$ where $\epsilon \sim N(0, 0.5)$, the tree will always favor the irrelevant X_2.

```
set.seed(45)    # R version 4.1.2
library(rpart)   # Version 4.1-15
library(rpart.plot) # Version 3.1.0

par(mfrow = c(2, 1))
x1 <- rnorm(mean = 1, n=100)
y <- 3*x1
x2 <- as.factor(seq(1, 100))
df <- data.frame(x1, x2 = sample(rep_len(x2, length.out = 100)), y)

fitted_cart <- rpart(y ~ x1 + x2, data=df, method = "anova",
        control = rpart.control(minsplit = 2, maxdepth = 1))
rpart.plot(fitted_cart, main = "Case 1: 100 levels with no noise. Tie.")
print(fitted_cart$splits)
```

Output:
```
 count ncat  improve  index adj
x1  100  -1 0.5994785 1.552605  0 # Decision tree chose x1 to split
x2  100 100 0.5994785 1.000000  0
```

```
# Number of level equals to sample size, with a little bit of noise
x1 <- rnorm(mean = 1, n=100)
y <- 3*x1 + rnorm(n = 100, sd = 0.5)
x2 <- as.factor(seq(1, 100))
df <- data.frame(x2 = sample(rep_len(x2, length.out = 100)), x1, y)

fitted_cart <- rpart(y ~ x2 + x1, data=df, method = "anova",
        control = rpart.control(minsplit = 2, maxdepth = 1))
rpart.plot(fitted_cart, main = "Case 2: 100 levels with very little noise")
print(fitted_cart$splits)
```

Output:
```
 count ncat  improve   index    adj
x2  100 100 0.6414503 1.0000000 0.0000000 # Decision tree chooses x2 to split
x1  100  -1 0.6371685 0.9917295 0.0000000
x1   0  -1 0.9800000 0.9079153 0.9565217
```

3. RESULTS AND DISCUSSION

The key takeaways from the simulation are the following:

1. Even though we might have a categorical variable with lots of levels, when the noise level is

small, the tree can still choose the right variable X_1 to make a split.

2. When the noise level increases, it becomes more and more difficult to make a "good split" using X_1. Since X_2 has many levels and can be used to create more partitions on y_i's, the decision tree tends to use X_2 to make a split.

To better understand this behavior, we should realize that decision trees have a very interesting property: Although a decision tree forces all variables to interact with each other (see footnote 1), it essentially does a sequence of "univariate analysis" at each node.

For example, at any given node, it might be the case that two variables, *gender* and *year_of_education*, combined can explain the response pretty well and lead to a good partition on the response. But when we look at them individually, none of them is highly correlated with the response In such a situation, a third unrelated categorical variable with lots of level might be selected. This type of situation is probably very common in practice. As a result, it is very understandable that many practitioners believe that categorical variables with many levels tend to cause problems.

Another thing to consider is the sample size. For example, a categorical variable might have 20 levels. If we have 10 million data points, then 20 levels might not be "too many". However, in practice, even though we have a large data set, we might still run into trouble. Because as the tree grows deeper, the sample size within each partition becomes smaller and smaller. For example, if we start with 100,000 data points and we grow our tree 3 levels deep perfectly balanced, we end up with only 12,500 data points.

4. CONCLUSIONS

In conclusion, a categorical variable's influence over the decision tree depends on many factors, including its number of levels, its "correlation" with the response variable, sample size of the data, etc. We advise against rushing to dropping certain levels of a categorical variable, or dropping the variable itself completely. We suggest understanding the model's behavior related to the categorical variables using tools such as variable importance plot (i.e., VIP), preprocess the variables using business context if interpretability is desired, and clarify the goal of the model.

5. REFERENCES

[1]   https://chaoguo14.github.io/2021/02/27/Decision-tree-and-variables-having-lots-of-levels/
[2]   Hastie, T. and Tibshirani, R. and Friedman, J. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition". 2009. Page 305 – 310.
[3]   https://stats.stackexchange.com/questions/49243/rs-randomforest-can-not-handle-more-than-32-levels-what-is-workaround

Biography of the Author

Chao Guo is a data scientist at Amazon. Before that, he worked at Munich Re as a senior data scientist. He has a master's degree in statistics from Rutgers University as well as a bachelor's degree in mathematics from University of California-Los Angeles.

Alternative to Tweedie in Pure Premium GLM

David R. Clark, FCAS, MAAA

Abstract

The Tweedie distribution provides a variance structure that is widely used in GLM for pure premium ratemaking. This essay suggests the quasi-Negative binominal (QNB) as an alternative. Both can be interpreted as collective risk models but the QNB has a variance structure that is more commonly used in other actuarial applications.

Keywords:. Tweedie Distribution, Negative Binomial, Generalized Linear Model, GLM

1. INTRODUCTION

"Statisticians, like artists, have the bad habit of falling in love with their models."

—George E. P. Box

The introduction of Generalized Linear Models (GLM) in ratemaking in the 1990's was a great step for Actuarial techniques and brought greater computational efficiency and flexibility to ratemaking. Extensions for mixed models (GLMM), generalized estimating equations (GEE) and additive models (GAM) have expanded the usefulness of the technique even further.

However, one assumption made early on has never really been challenged. The general approach is that a Tweedie variance structure is recommended for pure premium (i.e., aggregate loss rather than separate frequency and severity models). The main argument seems to have been that the Tweedie, when constraining the variance parameter $1<p<2$ can be interpreted as a collective risk model with Poisson frequency and Gamma severity.

True enough.

But there is a rival to the Tweedie variance structure that also has a collective risk model interpretation: the quasi-negative binomial (QNB). The QNB can be interpreted as a Poisson frequency with a [discretized] logarithmic severity. The logarithmic distribution is derived by mixing the mean parameter of a geometric random variable using an upper-truncated Pareto as the mixing distribution; it can be viewed as a discrete version of a mixed exponential model for severity.

Like the Tweedie, the QNB is a three-parameter curve that behaves like a compromise between the Poisson and Gamma variance structures. The difference between them is only in the behavior of the variance function. And the QNB follows the assumptions of collective risk loss models more closely.

For more details on the Negative Binomial, including code, interested readers are referred to the book-length treatment by Hilbe.

2. VARIANCE OF COLLECTIVE RISK MODELS

In a traditional collective risk model, we have a random variable N representing frequency, and a random variable X representing severity. The aggregate loss Z is the sum of N independent draws from the severity distribution.

$$Z = \sum_{j=1}^{N} X_j \tag{2.1}$$

The variance of the aggregate loss distribution is given by:

$$Var(Z) = Var(X) \cdot E(N) + E(X)^2 \cdot Var(N) \tag{2.2}$$

If we further assume that the frequency is negative binomial, then the variance of Z can be written as the sum of two components: first a multiple of the expected loss, second a multiple of expected loss squared.

$$Var(N) = E(N) + c \cdot E(N)^2 \tag{2.3}$$

$$Var(Z) = \left(\frac{E(X^2)}{E(X)}\right) \cdot E(Z) + c \cdot E(Z)^2 \tag{2.4}$$

This structure can also be interpreted as a process variance component plus a non-diversifiable risk component. That is, the coefficient of variation (CV) will asymptotically approach some finite percent rather than going to zero as the portfolio gets very large.

The Tweedie distribution does not share this behavior. Its variance function is defined as:

$$Var(Z) = \phi \cdot E(Z)^p \tag{2.5}$$

In this form, the dispersion parameter ϕ is estimated in the GLM but is considered a "nuisance parameter" because it does not affect the estimate of the expected loss. For the Tweedie, it is more difficult to validate the GLM results against other collective risk model work.

The QNB also includes a dispersion parameter ϕ estimated by the model. But the variance function more closely mimics the collective risk model used in other applications. The parameters ϕ and k are both in the same units as the response variable of the model.

$$Var(Z) = \phi \cdot \left(E(Z) + \frac{1}{k} \cdot E(Z)^2 \right) \tag{2.6}$$

3. COMPARISON OF TWEEDIE AND QNB

As mentioned above, both the Tweedie and the QNB can be interpreted as collective risk models and as compromises between the variance structures of Poisson and Gamma. The table below compares some of the differences in moments, especially as they change with the expected value μ.

If the loss cost is similar across all classes in the rating plan, then these alternative structures will give virtually identical results. However, the variance structure for Tweedie will tend to fit better for the larger loss cost classes. This is mainly because the CV decreases faster for the Tweedie than for the QNB. Secondarily, if the Tweedie shape parameter is constrained as 1<p<2, then it will have slightly lower kurtosis, leading to more reaction to observations in the "tail" of the distribution.

	Tweedie	Quasi-Negative Binomial
Mean	μ	μ
Variance	$\phi \cdot \mu^p$	$\phi \cdot \left(\mu + \frac{1}{k} \cdot \mu^2 \right)$
Coefficient of Variation	$CV = \sqrt{\dfrac{\phi}{\mu^{2-p}}}$	$CV = \sqrt{\dfrac{\phi}{\mu} + \dfrac{\phi}{k}}$
$\lim\limits_{\mu \to \infty} CV$	0	$\sqrt{\dfrac{\phi}{k}}$
Skewness	$p \cdot CV$	$\left(1 + \dfrac{\mu}{\mu + k} \right) \cdot CV$
Kurtosis	$3 + p \cdot CV \cdot (Skew - CV) + Skew^2$	$3 + 2 \cdot CV \cdot (Skew - CV) + Skew^2$

The behavior of the CV can be illustrated graphically to show that the CV is generally higher at the ends of the range of expected loss costs. The results being less sensitivity to the extremes.
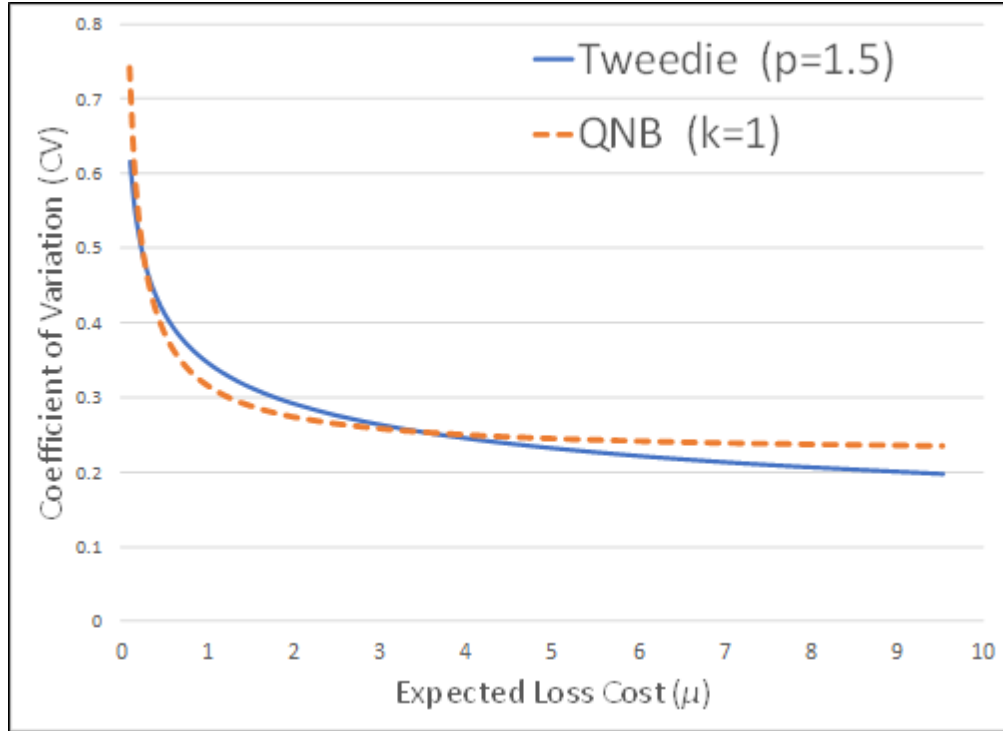
If we extend the curves further to the right, the Tweedie will approach a CV of zero whereas the QNB would approach a finite number greater than zero.

4. CONCLUSIONS

The selection of a variance structure in a GLM model will not materially change the fitted values in most cases. However, what will change is the significance measures (t-statistics, p-values, etc.) used to evaluate the model parameters. Changing the variance structure could change how we view the importance of different rating variables, or even the choice of which rating variables should be included.

All of this suggests investigating use of the QNB as an alternative to the Tweedie when modeling pure premium in GLM.

5. REFERENCES

[1]   Andersen, Duncan, Sholom Feldblum, Claudine Modlin, Doris Schirmacher, Ernesto Schirmacher, and Neeza Thandi, "A Practitioner's Guide to Generalized Linear Models," *CAS Study Note*, 2007.

[2]   D. R. Clark, "A Note on the Upper-Truncated Pareto Distribution," *CAS Forum* Winter 2013, Vol. 1, 1-22.

[3]   Goldburd, Mark, Anand Khare, Dan Tevet, and Dmitriy Guller, "Generalized Linear Models for Insurance

Ratemaking," *CAS Monograph* No. 5.

[4]   Hilbe, Joseph M., "Negative Binomial Regression," Second Edition 2011, Cambridge University Press.

[5]   Jorgenson B, de Souza MCP, "Fitting Tweedie's compound Poisson model to insurance claims data," *Scandinavian Actuarial Journal*, 1994. 69-93.

[6]   R. W. M. Wedderburn, "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika* 1974, Vol. 61 No. 3, 439-447.

Abbreviations and notations

CV, coefficient of variation (standard deviation divided by mean)

GLM, Generalized Linear Model

QNB, Quasi-Negative Binomial (Negative Binomial with dispersion parameter to rescale)

Biography of the Author

Dave Clark is a senior actuary with Munich Reinsurance America, working with Corporate Risk and Underwriting. He is best known within the actuarial community for his study note on "Basics of Reinsurance Pricing" on the CAS examination syllabus.

He can be contacted by email at: dclark@munichre.com