Deloitte.

P&C Pricing in the Age of Machine Learning

Actuarial and Insurance Solutions

January 2024

MAKING AN IMPACT THAT MATTERS since 1845

Introduction

As machine learning and artificial intelligence (ML/AI) proliferate throughout the insurance industry, applications in actuarial science are becoming increasingly popular topics. Actuaries involved with rate making have the seemingly impossible task of predicting the future, including in regard to claims and customer behaviour. Additionally, the current best prediction methods involve complex algorithms that make use of an insured party's known characteristics. ML/AI provides an avenue to build higher predictive power into pricing models. Actuaries have begun to research and experiment with these complex algorithms, but are faced with barriers such as the trade-off between predictive accuracy and model explainability, learning curves associated with new software, and rating-implementation costs.

In this document, we will explore some practical applications of ML/AI for pricing actuaries in property and casualty (P&C) insurance, as well as different types of models that can be used. We highlight some key findings of existing research and important considerations for actuaries building and testing new models.

Practical applications of ML/AI in P&C pricing

ML/AI algorithms are powerful tools that property and casualty (P&C) actuaries can use in a variety of contexts to optimize and enhance the rate-making process. Some practical applications include:

1. Pricing individual risk exposures

   ML/AI algorithms can be used to model claims frequency, severity, and/or pure premium based on a policyholder's characteristics. Studies such as Jain (2018), Colella and Jones (2023), and Gustafsson and Hansén (2021)—further explored in the "Existing research" section of this paper—have shown that ML/AI algorithms have the potential to predict future claims more accurately than traditional actuarial techniques such as one-way analyses and generalized linear models (GLMs).

2. Treating missing values

   ML/AI algorithms can help actuaries fill in variables for which values are missing in a data set. Some of these algorithms can automatically handle missing values; for example, Chen and Guestrin (2016) designed XGBoost to automatically assign a missing value to a default direction that had been optimally learned from available data. For models that don't automatically complete this task, such as GLMs, an actuary could instead use a more sophisticated imputation method, such as predictive mean matching or regression imputation.

3. Addressing feature engineering

   Actuaries can use ML/AI to select and/or modify existing predictors to better capture the relationship between predictor and response. These updated predictors can then be used in more traditional actuarial models, such as GLMs.

   For pricing actuaries, binning continuous variables might be an important use of ML/AI for feature engineering. For example, when using age as a predictor of pure premium, young and old drivers are generally seen as more risky than middle-aged drivers. However, ML/AI algorithms might identify a more optimal way to bin age groups, where younger and older groups are segmented into smaller increments than middle-aged groups.

   Additionally, a subset of addressing feature engineering could be employing interaction terms—using ML/AI to identify significant interaction effects between predictors in order to determine optimal combinations of variables for rating.

4. Using retention and conversion modelling

   Understanding policyholders' retention and conversion trends is complementary to loss-cost modelling. ML/AI algorithms can produce customer retention and conversion formulas, which can then be used to perform price-optimization exercises.

Popular ML algorithms

Machine learning is a broad term that encompasses a wide variety of algorithms, each with its own advantages, disadvantages, and scenarios for optimal use. The following table summarizes the pros, cons, and key details of the ML/AI algorithms that are being considered more broadly by P&C pricing actuaries.

Algorithm	Pros	Cons	Notes	Resources
Generalized linear model (GLM)	• Widely adopted • Simple to implement in most rating engines, and easy to translate into base rate and/or differentials • High level of interpretability; useful to explain models to non-actuarial stakeholders	• Does not predict claims as well as modern ML/AI techniques • Actuaries are often constrained to the error distributions of their chosen software; those seeking to model a more uncommon choice of distribution will be limited	• Standard practice for decades in P&C rate making • Extensive supportive literature exists	• Generalized linear models for insurance rating, Second edition
Generalized additive model (GAM)	• Widely accepted • Many actuaries have experience with GAMs • Straightforward to use for prediction, inference, confidence intervals, etc. • Functionality available in existing software	• Adds smoothed terms to GLM, reducing interpretability • May be less predictively accurate than tree-based models • Compared with other models, output is less interpretable and may be more difficult to explain to stakeholders • Risk for overfitting as the model becomes more flexible	• Used in practice for decades • Extensive supportive research and existing literature • Goldburd et al. (2020) liken a GAM to a GLM that inherently handles non-linearity	• "GLM, GAM, and more": Interpretable machine learning • Generalized additive models • Generalized linear models for insurance rating, Second edition
Extreme gradient	• Very high predictive accuracy	• Output is less interpretable and	• Uses existing gradient tree	• XGBoost: A scalable tree

boosting (XGBoost)	when compared with traditional techniques (e.g., GLMs) • Improves upon the gradient tree boosting algorithm, with increased speed and model performance (i.e., often faster than training a neural network) • Less susceptible than neural networks to over-fitting	more difficult to explain to stakeholders • Sensitive to hyperparameter tuning • Susceptible to overfitting if not using appropriate training/testing data sets	boosting techniques to create a faster, highly scalable, and better-performing ML algorithm	boosting system • Fitting data with XGBoost
Neural networks	• Higher predictive accuracy than traditional techniques	• Susceptible to overfitting • Hyperparameters are not intuitive • Depending on the software used, neural networks can sometimes take a long time to train models • Compared with others, output is less interpretable and may be more difficult to explain to stakeholders	• Feed-forward artificial neural network (ANN) is commonly used • Feed-forward ANNs have an input layer, one or more hidden layers, and an output layer, which transmit data through a model via neurons (basic unit of ANN)	• Deep learning with H2O, Sixth edition • Towards machine learning: Alternative methods for insurance pricing—Poisson-gamma GLMs, Tweedie GLMs, and artificial neural networks
Single tree models	• Interpretable and explainable to stakeholders • Numerical/categorical predictors do not need to be preprocessed • Models easily perform feature selection and	• Can be unstable, where a small change in the data can cause a large change in the model (i.e., susceptible to overfitting) • Other algorithms such as XGBoost and random forest address the shortcomings of		• What is random forest?

	handle missing values	general tree models, thus creating more powerful algorithms than these models		
Random forest	• Lower risk for overfitting than decision tree • Can be used for various purposes • Can handle missing data	• Computational time can be long with large data sets, as the algorithm computes many decision trees • Increased complexity compared with general single tree • Output is less interpretable and more difficult to explain to stakeholders, compared with other algorithms	• Improves upon decision trees through bagging and feature randomness	• What is random forest?
GLM with regularization	• Regularization ensures that only variables that contribute a certain level of predictive accuracy are included • Strong method for controlling overfitting while maintaining a high level of predictability	• Even though it improves upon GLM via penalization methods, this adaptation makes it more complex	• Regularization methods: ridge, lasso, elastic net	• Generalized linear models for insurance rating, Second edition
Combined actuarial neural network (CANN)	• Better predictive accuracy compared with a standard GLM • Computational time can be relatively quick for a well-fit GLM	• Output is less interpretable and may be more difficult to explain to stakeholders	• Embeds a classic GLM and neural network • Wüthrich and Merz (2018) liken CANN to “neural net boosting” of a GLM	• Yes, we CANN! • CANNs in actuarial rate making • Nesting classical actuarial models into neural networks
Generalized linear mixed	• GLMMs provide a way to incorporate	• A GLMM equation usually does not have	• The model shrinks predictor	• Generalized linear mixed

model (GLMM)	credibility into a GLM, working to modify coefficients based on data credibility	a closed-form solution, and may instead require an iterative solution that is more computationally intensive • Some distributional properties that held under a GLM are not maintained, thus GLMM outputs will be comparatively less interpretable	coefficients closer to the mean if there is low data credibility	models for rate making • Generalized linear models for insurance rating, Second edition

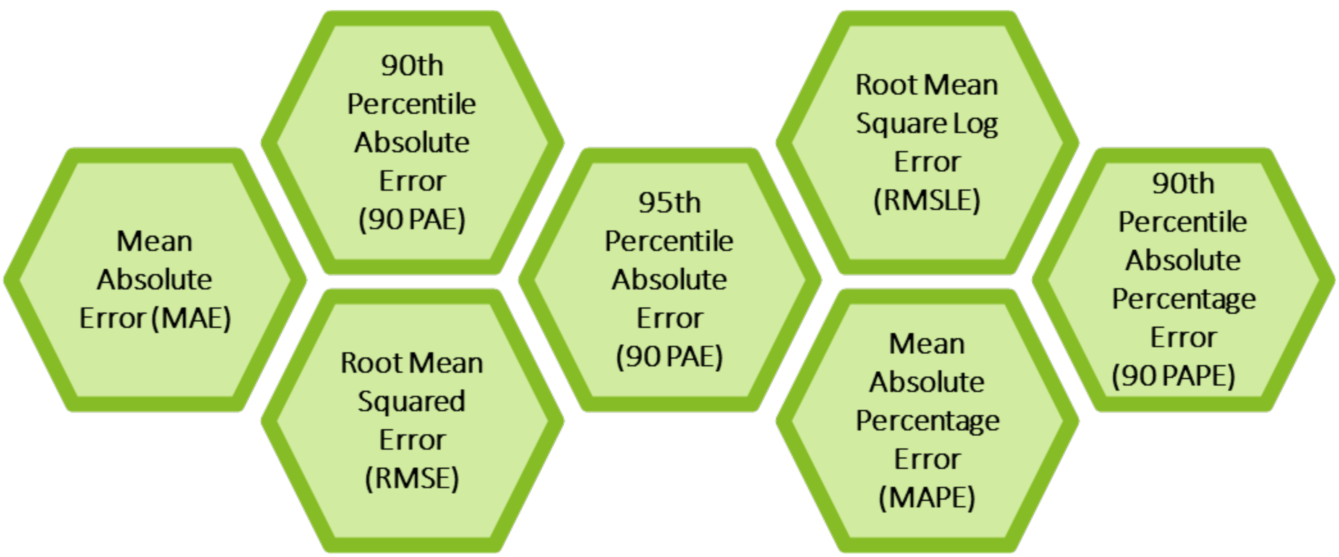
Accurate generalized linear model (AGLM)	• Aims to maintain a one-to-one relationship between predictor and response
	• High predictive accuracy through discretization of numerical features, coding of numerical features with dummy variables, and regularization
		• The AGLM approach was published relatively recently, in 2020, so literature is limited
		• Limitations with the AGLM R-documentation package when implementing AGLM for pure premium prediction:
		  ○ Supports only Gaussian, binomial, and Poisson error distributions
		  ○ In P&C pricing, pure premium, frequency, and severity are assumed to follow Tweedie, Poisson, and gamma error distributions, respectively
		• Actuaries may be required to build separate frequency and severity models
			• As per Fujita et. al (2020), AGLM is based on GLM, but equipped with more recent data-science techniques	• AGLM: A hybrid modelling method of GLM and data-science techniques

Importance of evaluation metrics

As ML/AI models are implemented in practice, actuaries will likely seek to define the "optimal model"—either in terms of choice of algorithm or hyperparameter selection for a specific algorithm. Of course, optimality is subjective, so conclusions will differ depending on evaluation criteria.

For rate making, the scenario is unique in the sense that future claims must be non-negative, and we would expect the chosen model to neither underestimate nor overestimate claims—this is to ensure that an insurer can be competitive but still profitable. As such, assessing a few common quantitative performance metrics such as mean squared error (MSE) and mean absolute percentage error (MAPE) may not provide actuaries with a comprehensive view of a model's suitability. Additionally, certain algorithms are more susceptible to overfitting, so even though some quantitative measures of predictive accuracy may show a model to be a good fit, it may, in reality, be overfit.

Colella and Jones (2023) highlight the importance of using a combination of both quantitative and qualitative performance metrics to assess model optimality. Some *quantitative* evaluation metrics for continuous variables (e.g., pure premium) are shown in the following graphic:



Examples of *qualitative* evaluation criteria include:

1. Actual versus predicted plot

   Actuaries can produce scatter plots with actual responses on the x-axis and predicted responses from a fit model on the y-axis. It is often helpful to superimpose a $y = x$ reference line to indicate a perfectly accurate model, and thus help illustrate where a given model may be overestimating and/or underestimating claims, as well as whether (and where) it is struggling to predict claim sizes.

For example, Colella and Jones (2023) showed neural networks had strong quantitative metrics, but the actual vs. predicted plots showed that models were overfit to policies with zero claims, predicting a value of 0 in most instances.

2. Decile charts

As demonstrated in Goldburd et al. (2020), decile charts can be used to predict pure premium by plotting quantiles of average predicted premiums against actual pure premiums. This can then provide key insights into how well a model can identify generalized groupings of risks.

3. Lorenz curves

Again described in Goldburd et al., Lorenz curves result from plotting the cumulative percentage of a given monetary variable against the cumulative percentage of the population in question. The researchers further explained how to calculate the corresponding Gini coefficient of a rating algorithm that can "quantify the ability of the rating plan to differentiate the best and worst risks."

Jain's 2018 study highlights additional metrics that can be used to assess model suitability: For qualitative metrics, the researchers drew attention to a scatter plot of a model's residual versus fitted values, where residual values in a well-fitting model would be small (close to 0).

For quantitative metrics, Jain compared models based on an array of metrics, including:

1. Akaike information criterion (AIC)

AIC penalizes complex models with many parameters. A lower AIC for one model than for a second model indicates that the first can better capture variability in the data.

2. Cross-validation (CV)

This method, also used by Colella and Jones (2023), involves splitting an entire data set into k-folds, where *k* is the number of subsets/models (folds). One iteratively runs the model on k–1 of the data groups, thereby leaving one data group out to serve as the testing set. Actuaries can then calculate the average metric of all the folds as a performance metric, such as k-fold CV MSE.

CV is an industry-standard technique for evaluating a model's performance, and can help actuaries to identify and remediate model overfitting. Additionally, it can help tune hyperparameters and effectively show model suitability on different testing data sets.

3. Risk premium ratio

Jain defined the risk premium ratio as $\frac{Observed\ claims\ cost}{Expected\ risk\ premium}$, where expected risk premium equals the model's predicted claims cost. This metric can aid actuaries in assessing the aggregate profitability and premium adequacy of a given rating model. The Jain study also suggested segmenting the data and calculating the risk premium ratio on groups of claims based on a chosen risk factor, such as vehicle age.

Doing so can highlight if the model exposes an insurer to premium inadequacy and possible adverse selection for a particular group of risks.

Existing research

ML/AI adoption continues to progress in the actuarial world; yet, existing research has only begun to scratch the surface for possible applications and best practices. The following are key conclusions of a few pivotal studies that explore ML/AI algorithms in P&C pricing. These findings—and the studies themselves, which can be accessed via the corresponding links—could help actuaries integrate these algorithms into their work.

1. Towards machine learning: Alternative methods for insurance pricing—Poisson-gamma GLMs, Tweedie GLMs, and artificial neural networks

   Jain (2018) used three models to predict claims costs of one-year auto insurance policies: Poisson-gamma GLM, Tweedie GLM, and ANN. The study concluded that, while all three demonstrated good performance metrics, Tweedie GLM and ANN were the most "actuarially fair" overall, as they neither overpriced customers in aggregate nor undercharged riskier segments such as young drivers—unlike the Poisson-gamma GLM. These findings highlight that, unlike traditional GLMs, ANNs do not require distributional assumptions, though they are less interpretable than more traditional models.

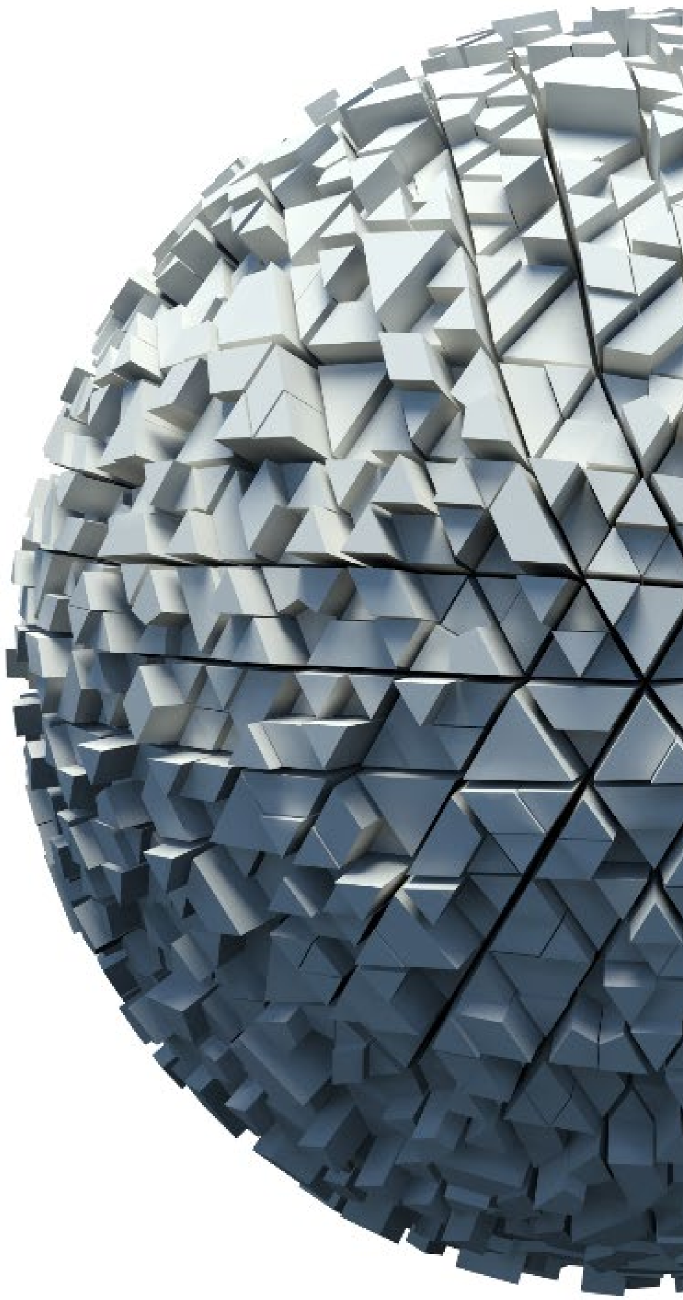
2. Machine learning and rate making: Assessing performance of four popular algorithms for modelling auto insurance pure premium

   Colella and Jones (2023) used four models to predict pure premium of auto insurance policies: GLM, AGLM, XGBoost, and neural network algorithms. They concluded that XGBoost is a promising model that can offer high predictive accuracy if built and tuned properly. However, the study also found that GLMs continue to be a valuable addition to the pricing actuary's tool kit. Additionally, the researchers highlighted the importance of actuarial models producing reasonable predictions. As such, using a variety of both quantitative and qualitative performance metrics can offer a more comprehensive view of model suitability.

3. Combined actuarial neural networks in actuarial rate making

   Gustafsson and Hansén (2021) compared CANNs and traditional GLM performances to predict pure premium. They developed five different CANN models with varying combinations of neural network hyperparameters, and with a Poisson GLM with log-link function. The study found that CANN models outperform the corresponding GLM. The researchers also noted that another benefit of CANN is in capturing the relationships between features and responses, which GLM cannot do. Still, actuaries will need to investigate different methods for interpreting CANNs and other complex models.

Conclusion



ML/AI algorithms are powerful tools that can be used to enhance the predictive accuracy of actuarial rate-making models. However, actuaries must be cognizant of the trade-off between predictive accuracy and model interpretability, as the latter is important when presenting models to stakeholders and/or regulators who may lack a technical understanding of the subject matter.

There exists a wide variety of ML/AI practical applications in actuarial rate making, with various ML/AI algorithms that can be utilized. Each algorithm has its pros and cons, but ultimately, the choice will depend on the situation. To help determine the optimal model in a given scenario, actuaries should use an array of both quantitative and qualitative performance metrics. Lastly, there are several industry-research studies that can be referenced to better understand the methodology and constraints of implementing an ML model. With the information and insights presented here, we hope to have offered a glimpse into the rapidly evolving world of machine learning and encouraged actuarial teams to consider ML/AI applications in their own work.

To learn more about how Deloitte's Pricing Centre of Excellence can help your organization use machine learning to enhance pricing capabilities, please contact:

Antonio Ferreiro
Partner, National Leader P&C
Audit & Assurance
aferreiroc@deloitte.ca

Shayan Sen
Senior Manager
Audit & Assurance
shasen@deloitte.ca

We'd like to thank **Harrison Jones** and **Sofia Colella** for their help with this publication.

Sources

Balasubramanian, Ramnath, Ari Libarikian, and Doug McElhaney, "Insurance 2030—The impact of AI on the future of insurance," *McKinsey Insights*, March 12, 2021, https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance.

Berk, Michael, "How to use elastic net regularization with any GLM," *Medium*: *Towards Data Science*, July 5, 2021, towardsdatascience.com/how-to-use-elastic-net-regularization-with-any-glm-6eab524bbcc6.

Candel, Arno, et. al, *Deep learning with H2O, Sixth edition*, H2O.ai, Inc., Oct. 2023, https://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/DeepLearningBooklet.pdf?_ga=2.173504483.1750675754.1668797671-1172431931.1666288634.

Chen, Tianqi, and Carlos Guestrin, "XGBoost: A scalable tree-boosting system," KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, dl.acm.org/doi/10.1145/2939672.2939785.

Colella, Sofia, and Harrison Jones, "Machine learning and rate making: Assessing performance of four popular algorithms for modelling auto insurance pure premium," *CAS E-Forum*, March 24, 2023, eforum.casact.org/article/73245-machine-learning-and-ratemaking-assessing-performance-of-four-popular-algorithms-for-modeling-auto-insurance-pure-premium.

De Virgilis, Marco, et al., "Machine learning in insurance," *CAS E-Forum*, Aug. 10, 2022, eforum.casact.org/article/36305-machine-learning-in-insurance.

Fujita, Suguru, et al., "AGLM: A hybrid modelling method of GLM and data-science techniques," Institut des Actuaires, March 15, 2020, www.institutdesactuaires.com/global/gene/link.php?doc_id=16273&fg=1.

Goldburd, Mark, et al., "Generalized linear models for insurance rating, Second edition," *CAS monograph series, Number 5*, Casualty Actuarial Society, 2020, www.casact.org/sites/default/files/database/monographs_papers_05-goldburd-khare-tevet.pdf.

Gustafsson, Axel, and Jacob Hansén, "Combined actuarial neural networks in actuarial rate making," KTH Royal Institute of Technology, School of Engineering Sciences, May 2021, www.diva-portal.org/smash/get/diva2:1596326/FULLTEXT01.pdf.

IBM, "What is random forest?" www.ibm.com/topics/random-forest, accessed Oct. 19, 2023.

Jain, Navarun, "Towards machine learning: Alternative methods for insurance pricing—Poisson-gamma GLMs, Tweedie GLMs, and artificial neural networks," Institute and Faculty of Actuaries, 2018, www.actuaries.org.uk/system/files/field/document/F7%20Navarun%20Jain.pdf.

Klinker, Fred, “Generalized linear mixed models for rate making: A means of introducing credibility into a generalized linear model setting,” *Casualty Actuarial Society E-Forum*, 2011, www.casact.org/sites/default/files/2021-02/pubs_forum_11wforumpt2_klinker.pdf.

Kloese, Sam, “Generalized additive models,” NAIC, Dec. 2, 2022, https://content.naic.org/sites/default/files/call_materials/GAM%20Background%20Info.pdf.

Labram, Alex, “Article: Fitting data with XGBoost,” Institute and Faculty of Actuaries, Oct. 11, 2019, www.actuaries.org.uk/news-and-insights/news/article-fitting-data-xgboost.

Molnar, Christoph, “GLM, GAM, and more,” *Interpretable Machine Learning*, Aug. 21, 2023, christophm.github.io/interpretable-ml-book/extend-lm.html.

Panlilio, Alex, et al., “Practical application of machine learning within actuarial work,” Institute and Faculty of Actuaries, Jan. 30, 2018, www.actuaries.org.uk/system/files/field/document/Practical%20Application%20of%20Machine%20Learning%20within%20Actuarial%20Work%20Final%20%282%29_feb_2018.pdf.

Schelldorfer, Jürg, and Mario V. Wüthrich, “Nesting classical actuarial models into neural networks,” SSRN, January 22, 2019, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3320525.

Spedicato, Giorgio, et al., “Machine learning methods to perform pricing optimization: A comparison with standard generalized linear models,” *Variance, Vol. 12. Issue 1*, Casualty Actuarial Society, July 2021, www.casact.org/sites/default/files/2021-07/Machine-Learning-Methods-Spedicato-Dutang-Petrini.pdf.

Wühritch, Mario V., and Michael Merz, “Editorial: Yes, We CANN!: Astin Bulletin: The Journal of the IAA,” *Cambridge Core*, Cambridge University Press, Dec. 7, 2018, www.cambridge.org/core/journals/astin-bulletin-journal-of-the-iaa/article/editorial-yes-we-cann/66E8BEC373B5CCEF3BF3303D442D6B75.