Article from

Predictive Analytics & Futurism News

August 2018

Issue 18

How Credible is a Predictive Model?

By Eileen S. Burns

One of the outstanding questions the life insurance industry must face in the adoption of predictive models is how to translate the understanding stakeholders have with respect to current methodologies into this new framework. In this article, I enumerate several key reasons why companies struggle to gain comfort with new methods, note the open mathematical questions we face, and report on a few recent publications offering specific ideas to answer them.

WHY THE TROUBLE?

1. To start, there is confusion on the **terminology**. What does it mean to be a predictive model? I'll keep it simple. There is a denotation and a connotation for the term predictive model. I loved these terms when I learned them in school, and this is a perfect example of why they are both important. The denotation is that the model predicts the future. All actuarial projection models do this. So what's the difference? The connotation is that it is a model built on past experience, leveraging advanced analytical methods, to generate improved confidence in future predictions over less advanced methods.

   The "advanced" analytical methods can run the gamut from fairly simple to a place where a degree in complexity science is required. Logistic models have been around for ages, the benefit here is to use their ability to leverage smaller amounts of data as linear predictors, rather than keeping everything categorical, which slices data into tiny pieces that lack credibility. Or they can be as advanced as ensemble models built via machine learning. These are powerful methods, though they struggle with interpretability and potential for overfitting. Going another direction, there are agent-based models. These attempt to address the "why" which so often evades statistical analysis. Correlation is not causation—a statistical model can only confirm likely correlations. An agent based model aims to describe why agents (policyholders, agents, insurance companies, etc.) act as they do using causal relationships. They test these relationships on past data in order to parameterize a set of rules.

   All such models can offer improvements over traditional methods, assuming the model builders respect the requirements of stakeholders.

2. Then there is confusion on the term **predictive**. In name, it simply means estimate what will happen in the future. The trouble is when it is interpreted to mean more. For example, sometimes we lack the past experience to generate a model as described above. Can a predictive model solve this issue? Nope! No model of the past decade will be parameterized based on vast quantities of past data that includes rising interest rates. Any model parameterized on recent data that is used to predict responses when interest rates rise will be extrapolating. As with traders in the stock market, some of these models are likely to make accurate predictions, and some are likely to fail to do so. A modeler who guarantees accurate predictions is like the hedge fund guaranteeing a 15 percent return. But a modeler who tells you the underlying assumptions, and offers guidance for how to gain comfort in those predictions as well as in their uncertainty … they can allow you to face that uncertainty with eyes wide open, and isn't that what actuarial judgement is all about? Yes.

3. Finally, there is confusion around how **credible** predictions can be. Given the last few paragraphs, this should be an obvious concern. It is made worse by the fact that there is not a one-to-one comparison between "credibility" and "believability." That is, the credibility we are accustomed to quoting as actuaries, that is based on the quantity of observations in a given category, is not easily comparable to the believability of the prediction that comes from a predictive model. This question is different from the first two as it requires a mathematical answer.

So how do you decide to believe in a model that may be of any form, is based on past data and possibly a few educated assumptions, when your trusted forms of credibility aren't relevant? And secondly, if you determine that your assumption is not fully credible, what options do you have?

I'm so glad you asked!

The remainder of this article gathers industry commentary on two questions.

1. **Credibility measurement:** How do we quantify the believability of a data-based assumption?

2. **Credibility blending:** If we determine we don't have enough confidence in assumptions based on our own data and models, what options do we have for leveraging external data and models?

BACKGROUND

In 2013, the Actuarial Standards Board published a revision to "Actuarial Standard of Practice Number 25: Credibility Procedures," expanding practice areas covered to include life and pension. The standard addresses both of these questions. It describes the responsibility of actuaries to ensure that there is adequate care taken in assessing credibility or blending experience, areas of such procedures where an actuary may need to use judgement and related considerations, and in Appendix I lists several currently used methods for assessing credibility. Beyond the scope of practice areas, the notable addition to the latest draft includes a new category "Emerging Practice Involving Statistical Models."

RESOURCES

The guidance provided in ASOP 25 is intentionally minimal, merely allowing for the actuary to use judgment in deciding which methods are most appropriate for a given application and requiring adequate communication. There are two good resources (1 and 2 below) for actuaries to learn more about their options and see applications of a few methods, however they concern only older methods not those mentioned as "Emerging Practice." Luckily this topic has started to gain the attention of predictive modelers. I'm aware of four more recent publications (3–6 below) that offer motivation for addressing the issue of credibility, and/or possible solutions.

1. "Credibility Practice Note," American Academy of Actuaries, July 2008, Robert DiRico et al.

   The first two sections provide some motivation for revisiting credibility, and an amusing recap of state variations in requirements related to credibility. The third section discusses Limited Fluctuation and Greatest Accuracy (aka Buhlmann, aka Empirical Bayesian) credibility in detail, and addresses strengths and weaknesses of each. It also offers examples related to mortality, lapse, and reinsurance pricing, and a couple of cautionary tales, lest you start to think credibility can be straightforward. The last two sections can be seen as a resource—offering a short history of credibility theory and an extensive bibliography.

   Takeaway: This is a comprehensive resource for understanding how to apply two common types of credibility analysis (measurement and blending) and potential complications in applying them.

2. "Credibility Theory Practices" by Stuart Klugman et al. in December 2009.

   This was published in 2009, seemingly as an attempt to encourage more life companies to consider implementing credibility. "Statistical credibility's rigor can validate or improve actuarial judgment applied to company experience data."

   It presents thorough examples (with accompanying spreadsheets) for both limited fluctuation and Buhlmann credibility. The examples highlight the differences between the two methods when applied to A/E ratios for individual companies relative to the industry experience. The conclusion emphasizes that these differences stem from two important features of a block of business: the difference between its mean and the population mean, and the variation within the block about its own mean. The paper also includes a thorough bibliography.

   The publication consists of both the paper and a survey of 190 insurers "to find out the level of understanding in the industry, actuaries employed by U.S. insurance companies were surveyed to ascertain who uses credibility theory and how credibility theory is applied at responding insurers."

   Takeaway: This is a very practical article describing the same two common types of credibility analysis (again, both measurement and blending) with straightforward examples that allow easy comparison between the two.

3. "Is Credibility Still Credible?" Mark Griffin, *Risk Management*, August 2017.

   In the Joint Rrisk Management Section newsletter, Mark Griffin raised this question citing motivation from PBR, IFRS, Solvency II, and Embedded Value. He uses a simple example to highlight the need for a method that supports use of a company's data when it is the most relevant data available, explaining that some methods would argue otherwise. He rejects the out-of-the-box version of limited fluctuation credibility that would mandate a minimum of 1537 claims based on confidence of at least 95 percent and tolerance of at most 5 percent. He argues a hypothesis testing paradigm makes sense.

   Takeaway: If you need inspiration to reconsider how you are approaching credibility analysis, this is the article for you.

4. "Logistic GLM Credibility," Matthias Kullowatz, *Predictive Analytics and Futurism*, December 2017.

   My colleague Matthias Kullowatz notes that a predictive model such as a logistic GLM, generates probabilities, as well as confidence estimates, allowing him to reframe limited fluctuation credibility within the hypothesis testing framework. He laments it is still left to the actuary to set appropriate confidence and tolerance bounds, and discusses other issues such as the assumption of asymptotic normality and link function complications. He alludes to a method for determining credibility of an estimate relative to one from an industry population with "full credibility."

Takeaway: The article presents a proposal for using limited fluctuation credibility through a hypothesis testing framework to measure credibility in terms of statistical confidence. It is an easy extension to note that upon selecting confidence levels that constitute full and no credibility, this can then be used to blend models between company and industry experience.

5. "Calibrating Risk Score: Model With Partial Credibility," Shea Parkes and Brad Armstrong, *Predictive Analytics and Futurism*, July 2015.

   Shea Parkes and Brad Armstrong demonstrate a model for credibility that goes straight to blending of experience to calibrate risk scores for smaller blocks of policies. "Instead of estimating completely new weights, it is possible to use a technique known as ridge regression to only adjust the coefficients that are credibly different for the target population." They further describe that the method can be tuned to vary the weight given to each of the target and the reference. They discuss validation methods for such smaller blocks, and variations among ridge to lasso to elastic net regressions. The paper includes reference to a package in R.

   Takeaway: The article presents a proposal for using ridge regression to generate estimates for a small dataset that may differ from a larger reference set, but without losing the power of the reference dataset's credibility. Credibility measurement and blending is done implicitly through the model.

6. "Parameter Uncertainty," Brian Hartman et al., CAS, CIA, SOA Joint Committee, April 2017.

   This paper was published in 2017 by a cross-body joint effort of the CAS, CIA and SOA. In it Brian Hartman et al. give a comprehensive view of parameter uncertainty explaining "understanding the uncertainty associated with model estimates is essential to properly quantifying risk." While they don't mention credibility explicitly, the fundamental question addressed is the same—how much faith can you put in the estimates from your model? In the life context, they look at mortality rates, mortality curves, and single premium immediate annuity values. They propose a Markov chain Monte Carlo (MCMC) method to estimate the posterior variability of outcomes for a hypothetical block. The paper has additional examples pertaining to health and P&C that, as a non-practitioner, I will leave to you to explore.

   Takeaway: The paper proposes an MCMC method to estimate the likely breadth of possible futures, essentially, a confidence band around the best estimate. As with the Kullowatz article, the method could be used to blend models between company and industry experience, or alternatively could be adapted to consider company data as the sample data and consider the posterior estimate to be the final estimate.

CONCLUSION

You can see we're starting to chip away at the iceberg, but there's more to do. Specific topics to address include other ways to blend models, how to document actuarial judgment required, and how to determine when such judgments can be statistically tested. It would also do us well to standardize methods for the new options now listed in ASOP 25 for various emerging model forms, for which it states: "credibility can be estimated based on the statistical significance of parameter estimates, model performance on a holdout data set, or the consistency of either of these measures over time."

Our section is full of those who are interested in developing and applying new modeling methods, and as actuaries, we are still suited (and required) to explaining how the results should be understood and used. As we continue to push the envelope here, we'll need to continue to enhance our communication of what we've done.

Please send me a note if you are aware of publications on other methods for credibility analysis that we should add to the conversation, or if you want to write one of your own in an upcoming PAF newsletter!

Eileen S. Burns, FSA, MAAA, is a principal and consulting actuary with Milliman. She can be contacted at *eileen.burns@milliman.com.*