Case Study Part 3: Improving Financial Projections for Long-Term Care Insurance with Predictive Analytics

By Missy Gordon and Joe Long

Predictive analytics has significant potential to help long-term care (LTC) actuaries develop more accurate projections via an automated robust process. In our previous articles on this topic, we discussed the importance of giving the "right" amount of weight to a company's experience when adjusting an industry benchmark in order to produce a projection assumption that generalizes well to future data. Subsequently, we covered the use of penalized general linear models (GLMs) and gradient boosting machines (GBMs) to balance the trade-off between bias and variance without relying solely on actuarial judgment. In this article, we walk through an illustrative case study for one company (with its permission), call it Company Enlightened, that transitions from using traditional techniques (actual-to-expected studies) to using predictive analytics to develop a claim termination assumption.

Like most insurers providing LTC coverage, Enlightened did not have enough historical claim data to build an assumption completely from its own experience. Therefore, we used an industry benchmark as a starting assumption and adjusted it to better fit Enlightened's experience. This benchmark was developed from the *Milliman Long-term Care Guidelines*, which reflects industry experience that is tailored to this particular block of business by adjusting for demographics, product design, claim adjudication, and underwriting. One of Enlightened's initial requirements was that the new assumption be delivered in the same format as its existing assumptions to avoid modifying the projection system. This created a stepping stone approach, where progressing through the steps incrementally allows one to easily compare the approaches and gain comfort with using predictive analytics to develop the assumption. Furthermore, because of the flexibility that predictive analytics offers, it sets the stage for future assumption updates that consider new variables and interactions.

As actuaries, we are interested in more than just the single projection estimate that the assumption produces. Often we are required to conduct sensitivity tests or determine the amount of margin that should be included in an estimate. Monitoring the emerging experience is also important because we need to determine if our estimate is within a reasonable range of fluctuation or if it is a deviation due to a systemic shift underlying the experience that warrants investigation. Determining thresholds of reasonable fluctuation can be subjective in nature. Fortunately, with predictive analytics, we are able to remove some of this subjectivity by using techniques that estimate the uncertainty underlying the projection.

METHODS

The existing assumption was developed using a traditional actual-to-expected (AtoE) approach—combining credibility theory and actuarial judgment to adjust the benchmark. All calculations were performed using an Excel workbook to allow for a transparent avenue to make adjustments based on actuarial judgment from a seasoned actuary. Claim termination tables were developed for three sites of care: nursing home (NH), home care (HC), and assisted living facility (ALF). Each table varied by gender as well as by lifetime and non-lifetime benefit periods. This resulted in a total of 12 tables, each representing the benchmark with adjustments based on Enlightened's historical claim experience.

To isolate the incremental impact of shifting the assumption development following a traditional approach to one incorporating predictive analytics, we used the same historical data and benchmark. By using the same variables and assumption format, the new assumptions could be uploaded into the projection system for a direct comparison. After getting comfortable with the new predictive analytics approach, additional variables can easily be explored for the next assumption update.

Initially, we explored the use of a penalized GLM to update the projection assumptions. However, we found in this application, the parametric (user defined) formula for a penalized GLM created challenges due to the complex interactions underlying the data. For example, we would have had to make decisions concerning which claim duration months to band together or perhaps include higher-order terms to introduce a non-linear relationship. We also would have been required to determine the appropriate interactions among other driver variables in the starting assumption, such as incurred age, gender, claim situs, and benefit period. Given that one of our aims was to find a simpler, less time-consuming approach to expectations adjustment, we needed to identify an alternative method.

We decided to use a GBM algorithm. This allowed us to capture the complex interactions underlying the data in an automated fashion and also to determine the amount of credibility to give to the various data cuts. Although GBMs and machine learning models in general tend to have a "black box" quality—meaning it is not easy to parse exactly how the model arrived at a particular result—we were still able to produce an adjusted assumption that was in the same format as the current assumption. We did this by developing artificial observations for every cell in our base benchmark tables and then running them through the trained GBM model to produce the final adjusted assumption. As discussed in our prior article, we can gain more insight on how a model arrives at a prediction by looking at variable importance measures and partial dependence plots. There are emerging advancements and continuous research in this area, which is shedding light on these "black box" algorithms—making them more transparent.

COMPARING THE RESULTS

Figure 1 is an illustrative example for one of the 12 assumption tables we developed which compares the discounted average length of stay (ALOS) that is calculated from the benchmark, the traditional methodology (existing assumption), and the GBM approach (new assumption). As you can see, the traditional method and GBM produce a similar ALOS that is longer than the benchmark, which gives comfort that the GBM assumption is in a reasonable range of our prior developed assumption.

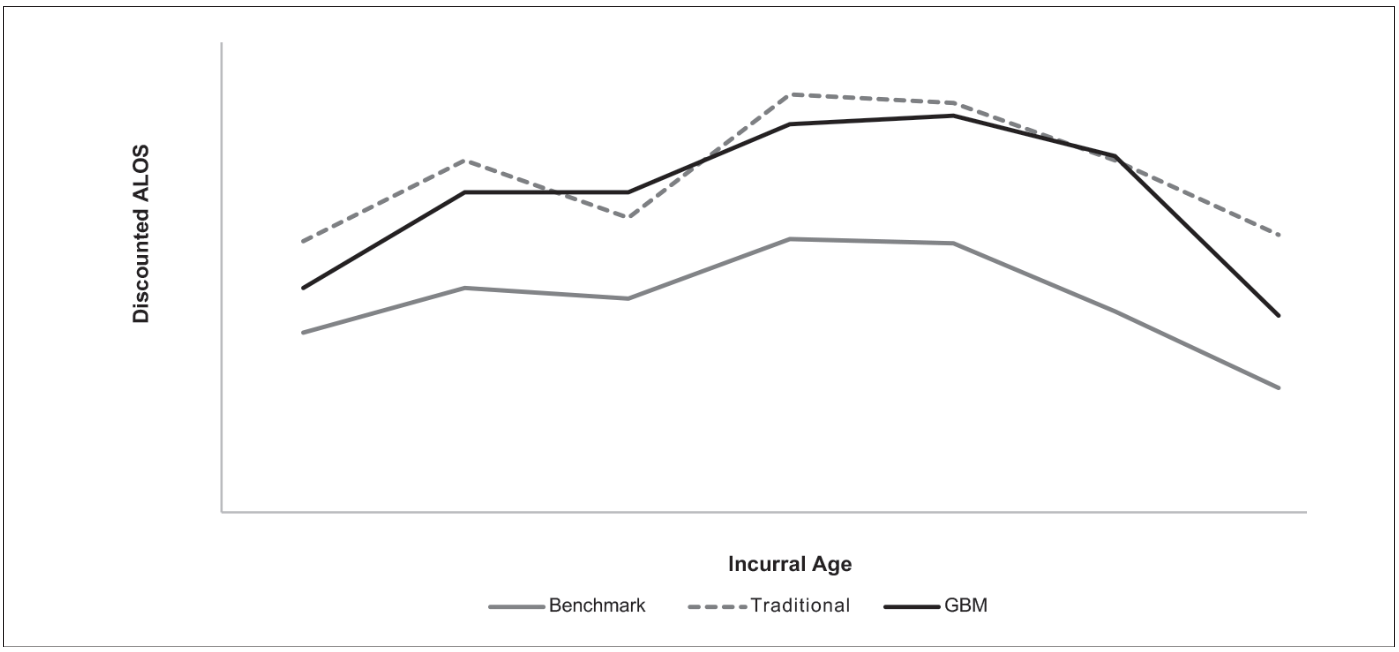
Figure 2 further illustrates the sensitivity of switching from using the traditional developed assumption to the GBM developed assumption shown with the calculation of future profit margin (as percent of premium). Again we see the GBM produced similar results with an impact of -0.3% from making the switch from the traditional to GBM method.

Figure 2
Comparisons of Different Approaches to Calculating Future Profit Margin

	Future Profit Margin		
	Traditional	GBM	Impact
Total	15.5%	15.2%	-0.3%

As we are only updating the claim termination assumption underlying the morbidity (i.e., no changes due to incidence or utilization), we would not expect wild deviations in future profits, but the impact is observable. This impact is for illustrative purposes and does not indicate the direction or magnitude that such a change might have for other companies and situations. Because the traditional study is highly dependent on actuarial judgement, the impact could be materially different for certain situations where the traditional approach is significantly over- or under-fitting the assumption.

Figure 1
Comparisons of Different Approaches to Calculating Discounted Expected Average Length of Stay



TESTING PERFORMANCE ON NEW DATA

The original study was performed on data gathered through 2014. Subsequently, new data was gathered, enabling us to test the predictive performance on the new two years of data. This allowed us to test how well each assumption development method performed on data that was not used to develop the original assumptions.

Figure 3 compares results on new claim experience data using the metrics of AtoE, mean squared error (MSE), and mean absolute error (MAE). The reason we use all three is that AtoE metrics can mask offsetting errors, which MSE and MAE measurements do not.

Figure 3
Actual-to-Expected Claim Termination Experience

Metric	Benchmark	Traditional	GBM
AtoE	0.90	0.93	0.93
MSE	72.4	63.6	54.3
MAE	6.5	5.5	5.2

The key takeaway from this table is that the GBM assumption produced similar results to the traditional assumption while having slightly better performance when looking at the MSE and MAE metrics. At first glance, it might seem that getting similar results is not that exciting. However, the important observation is that the GBM enables us to provide an automated process that does not demand the full labor of a seasoned actuary—making the results more reproducible (as opposed to many manual or judgement-based decisions). It also produces a better projection estimate as shown by the predictive performance metrics.

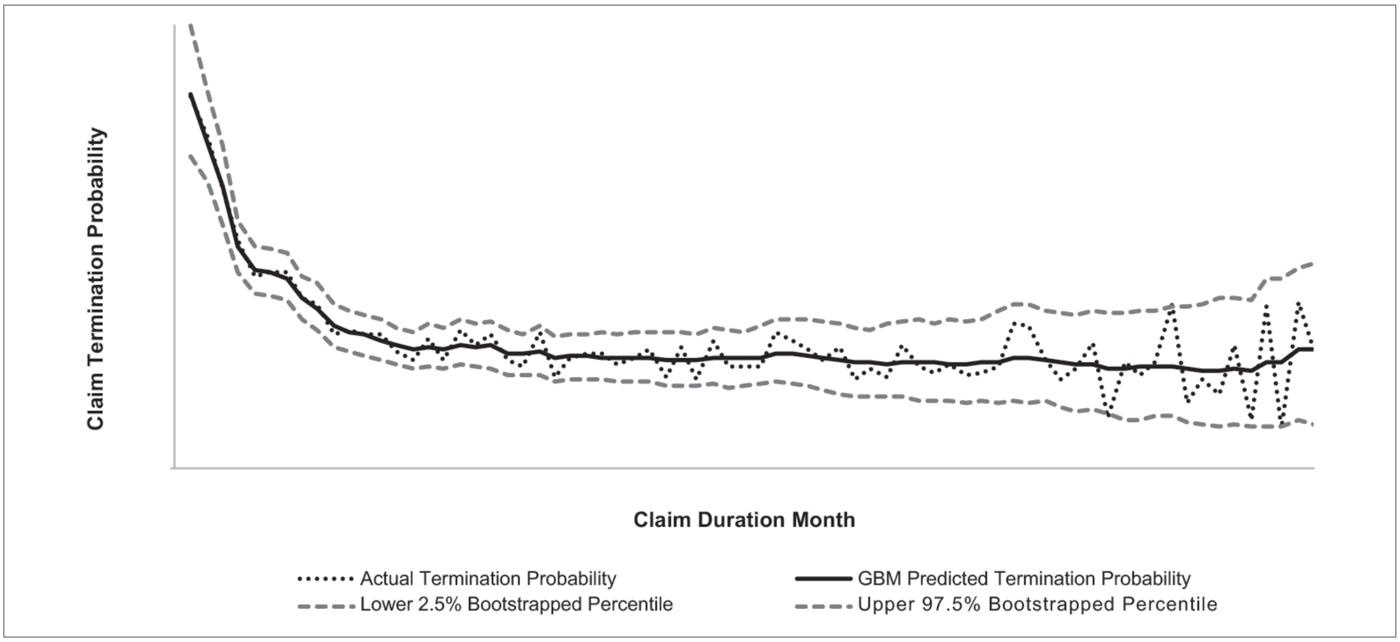
As discussed in our first article of the series, the traditional method requires a lot of judgment and uses a cumbersome Excel workbook that is difficult to update. These updates are also prone to human error. Predictive analytics automates the updating of assumptions, saving valuable time that can be used to solve new challenges and deliver value-added insight. The similarity in results also provides decision makers with comfort that the use of predictive analytics is not going to produce wildly different results from what a skilled actuary would provide.

More broadly, the automated nature of the GBM makes it easier to broaden the variables and interactions one can reasonably consider. For example, actuaries can explore adding new driver variables that were not historically included in the projection system. Predictive analytics can be used to efficiently assess whether these variables produce meaningful differences in outcomes, even if it was not feasible to incorporate them into the original assumption setting process. Adding third-party data also becomes much easier as does analyzing complex interactions such as morbidity improvements.

UNDERSTANDING UNCERTAINTY

After becoming comfortable with predictive analytics, we can use them to explore answers to additional questions. As experience emerges and deviates from that assumed—we can say with absolute certainty that it will happen–we might want to know if the emerging experience is an early detection of a new pattern or if it is within “normal” fluctuation. Often, we want to know how much we can anticipate actual

Figure 4
95% Credible Interval for the GBM Using Bootstrapping with 1,000 Replicates



experience to fluctuate around the model's estimate in order to aid in sensitivity testing or to determine how much margin to include in an estimate. With predictive analytics, we can do just that. There are techniques to estimate the amount of uncertainty in a model's estimation that helps us understand how the statistical noise inherent in historical experience data (or missing driver variables) affects our projection assumption. With a GLM, there is a predetermined theoretical formula that underpins the calculation of confidence intervals based on an assumed statistical distribution. A GBM, by contrast, is a machine learning technique that combines a large number of decision trees that makes it impossible to calculate a direct formulaic solution for model uncertainty. In such a case, we can pull ourselves over the fence of impossibility by using bootstrapping[1] paired with parallel cloud computing to estimate model uncertainty.

Bootstrapping uses "random sampling with replacement" to measure model uncertainty by providing a direct estimate of the requested distribution as opposed to assuming a parametric distribution from the outset. For instance, to better understand the plausible statistical fluctuation underlying the claim termination assumption, we conducted a bootstrap analysis on the GBM that was used to develop the claim termination assumption.

Saving you from the full and highly technical details, we accomplished this by creating 1,000 simulated data sets that were randomly sampled (picked) with replacement (can be picked again) from the original claim experience data set. For each simulated data set, we re-trained a GBM and used it to project claim terminations for the subset of original claim experience that was not picked for the simulated data set (i.e., out-of-bag[2] artificial terminations). These projections created a distribution of average claim termination rates by claim duration. From this bootstrapped distribution, we selected the 2.5 and 97.5 percentiles at each duration month to create the lower and upper bound for the 95% credible interval,[3] respectively. Figure 4 provides a visual depiction of the 95%-credible interval of claim termination probabilities that was created via the bootstrap analysis.[4] Other bootstrap analyses can be conducted to answer a variety of questions related to model uncertainty; this is only one example.

ADDITIONAL USES FOR PREDICTIVE ANALYTICS AND FUTURE EXPLORATION

These results point to a number of interesting areas for additional exploration in the field of predictive modeling for addressing the needs of the LTC community.

First, there is the possibility of updating additional assumptions. We have already used predictive analytics for morbidity incidence and incurred claims in developing the *Milliman Guidelines* industry benchmark and several company studies. Mortality also lends itself well to these techniques because one can use a standard table as the offset or starting expectation and then make adjustments to it. We have used predictive analytics to develop mortality assumptions for multiple companies. The same process can also be used for utilization and lapse, whether starting from scratch or adjusting an earlier benchmark.

Besides assumption development, predictive analytics can also be used in the field of fraud, waste, and abuse detection. These tools could be used to flag claims that might be fraudulent based on false diagnoses, falsified reports of resource use, overpricing, or waste. As claims age and blocks become more expensive to service, the detection of fraud, waste, and abuse becomes critical to reducing or preventing rate increases and maintaining plan solvency.

Finally, predictive analytics may be used to understand which care management approaches and specific interventions help reduce the incidence and severity of claims. Much of the LTC industry is closed block and faces significant challenges in managing this business. Rate increases can only go so far due to limitations and lack of consistency in the regulatory environment. Underwriting manages the risk on the front end, but without many new issues, companies need to look at managing the back end of blocks. Prescription drug history is a component of underwriting, but may also be useful in later years on the back end to identify insureds that may be most at risk for claim and allow a company to actively manage them.

At present, the potential of predictive analytics in the LTC industry is still in the early stages of being realized. We hope this article series has helped demonstrated some of the possibilities and provided an incentive for further exploration. With LTC being one of the most challenging lines of business, modern modeling methods provide great promise for better projecting anticipated performance and managing claims, enabling actuaries to provide greater value. ■

Missy Gordon, FSA, MAAA, is a principal and consulting actuary at Milliman. She can be reached at *missy.gordon@milliman.com.*

Joe Long is an assistant actuary and data scientist at Milliman. He can be reached at *joe.long@milliman.com.*

ENDNOTES

1 For more information on bootstrapping see section 5.2 on page 187 of the textbook *An Introduction to Statistical Learning.*

2 In resampling methods, out-of-bag refers to observations that were not selected in the resampled data. In this case, they were the observations that were not used to train the GBM within each bootstrapped replicate.

3 For more information on credible intervals see *CONFIDENCE VS. CREDIBILITY* INTERVALS. Retrieved Jun. 12, 2018, from *https://freakonometrics.hypotheses.org/18117*.

4 Special thanks to Shae Parkes, FSA, MAAA, a principal and consulting actuary at Milliman, for assisting us in the development of our methodology for bootstrapping a credible interval for a GBM.