Article from

Predictive Analytics and Futurism

May 2019

Issue 20

All About Them Curves: Ordered Lorenz Curves and Lift Curves

By Michael Niemerg

Often, models are crafted to optimize some objective function. However, in the real world, the quality of a model is multidimensional and can't really be summarized in a single metric. One of the goals in any thorough model validation process is the evaluation of a model with an entire array of diagnostics. The more perspectives we can judge a model from, the better sense we can get of what the model is truly achieving.

Two very closely related validation measures I've found myself employing continually more often are Lorenz curves and Gini gain. Both give a sense of how well a model is able to stratify risk in the sense of rank ordering. The Lorenz curve is a visualization of this stratification, while the Gini gain is a way to transform such a visualization into a single summary statistic.

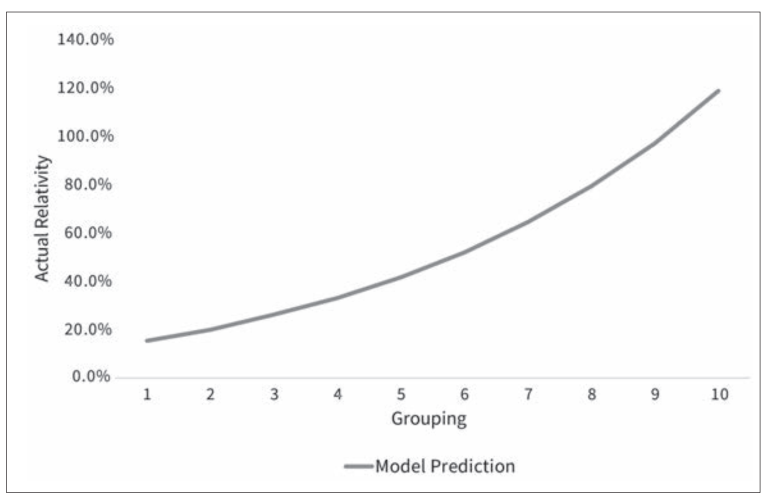
These two model validation diagnostics are great alternatives and complements to lift curves, which are another commonly utilized visualization technique used to measure model stratification. The term "lift curve" has multiple meanings, so I'm going to define precisely what I mean.

The best way to define a lift curve is to explain how it is created. To start, you need three elements for each observation in your dataset: the predicted value of an outcome (coming from either a manual rate, mortality table or any other type of predictive model), the actual value of that outcome, and a grouping by which to bin the observations. The chosen grouping can vary; for example, quantiles or prediction ranges are often used.

For consistency and concreteness, I will call predicted outcomes "claim costs." From these claim costs, I will also frequently refer to as "actual relativities" and "predicted relativities"—where both of these values are expressed as ratios relative to some baseline expectation. Using health rating as an example, we may reference a manual rate as our expected value. Then, if the actual experience of one member was $500 per member per month (PMPM) and their manual rate was $400, the actual relativity is 1.2 ($500/$400). If we were trying to build a predictive model to adjust this manual rate, and for this member the predicted outcome was $600, the predicted relativity would be 1.5. In certain situations, using these relativities instead of the claim costs themselves can be simpler. For instance, if we want to see how much a single variable from our rating formula stratifies risk while holding everything else constant, we can use the baseline in the relativity to account for all the variables in the rating process.

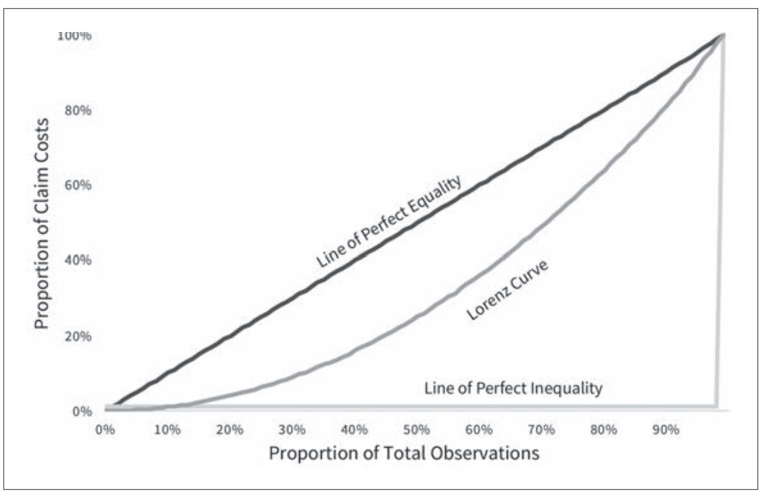
Now, with all the data in hand, we create our graphic to display our lift curve. Taking our observations sorted into groupings, we calculate the actual relativity for each grouping and plot the results (with predicted relativities on the x-axis and actual relativities on the y-axis). Figure 1 offers an illustrative example. For the example in Figure 1, as the grouping increases, so does the actual relativity. If we made this example more concrete and said that the grouping represented predicted relativities, then this chart would be demonstrating that as our predicted relativity increases, the actual relativity also increases (exactly what we were hoping for).

Figure 1
An Illustrative Lift Curve



Now, let's define Lorenz curves and Gini gain. When one typically thinks of Lorenz curves, the immediate association is with economics, since Lorenz curves have a storied history of being used to measure income inequality. In a typical construction of Lorenz curves, the function measures the cumulative proportion of individuals on the x-axis (I'm going to instead use the more neutral lingo described earlier and call these observations) and the cumulative proportion of income on the y-axis where the observations are ordered on income (this will be replaced by our claim costs). The basic idea is to measure how unequal the distribution is. See Figure 2 for an example. The line of perfect equality is what we would expect if we had a perfectly equal distribution (everyone has the exact same claim cost) and the line of perfect inequality is the line we would get if we had a perfectly uneven distribution (one observation generates 100 percent of the claim costs).

Figure 2
An Illustrative Lorenz Curve



What I'm going to talk about from this point on are ordered Lorenz curves, which are very similar conceptually to Lorenz curves, but now we use a predicted relativity to order our observations. Once ordered in this fashion, we graph the cumulative portion of actual claim costs and the cumulative proportion of observations based on this ordering. In building our curve this way, we are able to see whether our predicted relativity is distributing actual claim costs unevenly (which is desirable—ideally, our predictions have some value beyond pure randomness). For instance, suppose we were using an ordered Lorenz curve to test the impact of a new manual rating adjustment factor where a 1.00 signifies no adjustment to the manual rate. Here, we would hope that those observations earning a score below 1.00 would be associated with a disproportionally small amount of losses and that those with an adjustment greater than 1.00 would be associated with a disproportionately high share of losses. This would result in an ordered Lorenz curve with a bowed shape, similar to that in Figure 2.

The Gini statistic allows us a succinct way to summarize a Lorenz curve or ordered Lorenz curve with a single metric. It is equal to the total area between the Lorenz curve and the line of perfect equality divided by the total area between the Lorenz curve and the line of perfect inequality (since the total area under the line of perfect equality will be equal to 0.5, this is also equivalent to two times the area between the Lorenz curve and the line of perfect equality). Having this single summary metric is nice when values are either close and the Lorenz curves are hard to visually distinguish from the line of equality or if you want to summarize an entire array of model validation metrics in a single table to compare them simultaneously.

What are lift curves and ordered Lorenz curves achieving for us? Remember that our end goal is to create a visualization that helps us see risk stratification in a meaningful way. With the alternative approach of plotting every single data point (comparing predicted versus actual), the result would be a line bouncing all over the place, because insurance claim costs are highly volatile. By way of contrast, lift curves and ordered Lorenz curves are both employing a form of smoothing to make results visually interpretable. An ordered Lorenz curve smooths out the variance in actual results between different ranges by showing a cumulative value (any single observation will only contribute a small amount to the cumulative distribution resulting in a curve that appears smooth), whereas a lift curve is doing that with its grouping.

One thing that needs to be kept in mind is that by its very nature, smoothing removes granularity to make underlying trends more visible. One caution when using lift curves here is, therefore, that the break points in a lift curve are arbitrary, and slightly different break points can result in massively different looking curves. Compare Figure 3 (Pg. 12) and note how different these two models' results are. Now look closer—Figure 3(a) is based on the same model (I pulled a sleight of hand). The only difference is that the range on the x-axis in Figure 3(b) is shifted slightly—by a mere 0.01. This slight difference in binning created a massively different looking result. This isn't even a contrived example. I was able to create these graphs using data from a real project and without much experimentation. Moreover, the model represents several hundred thousand lives, so it isn't just a consequence of using a small sample. Most lines of business in insurance are volatile, and this variability impacts the lift curves.

This underscores the fact that you can't focus too much on small perturbations in the lines of lift curves.

Figure 3
Lift Curve Comparison for Slightly Different Ranges of Lift Curves

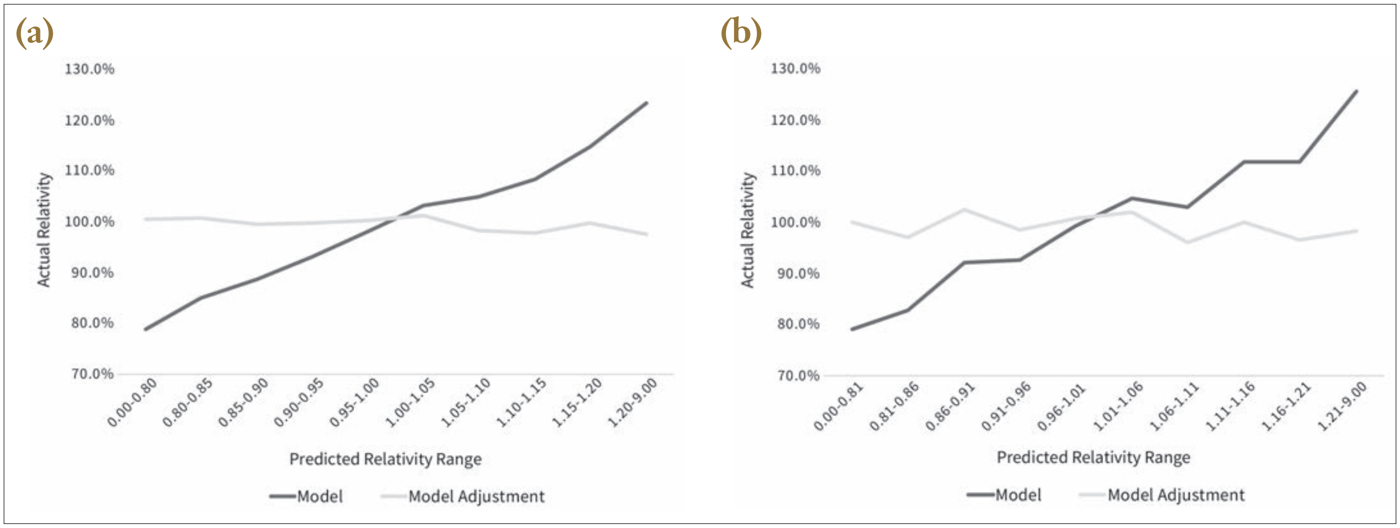
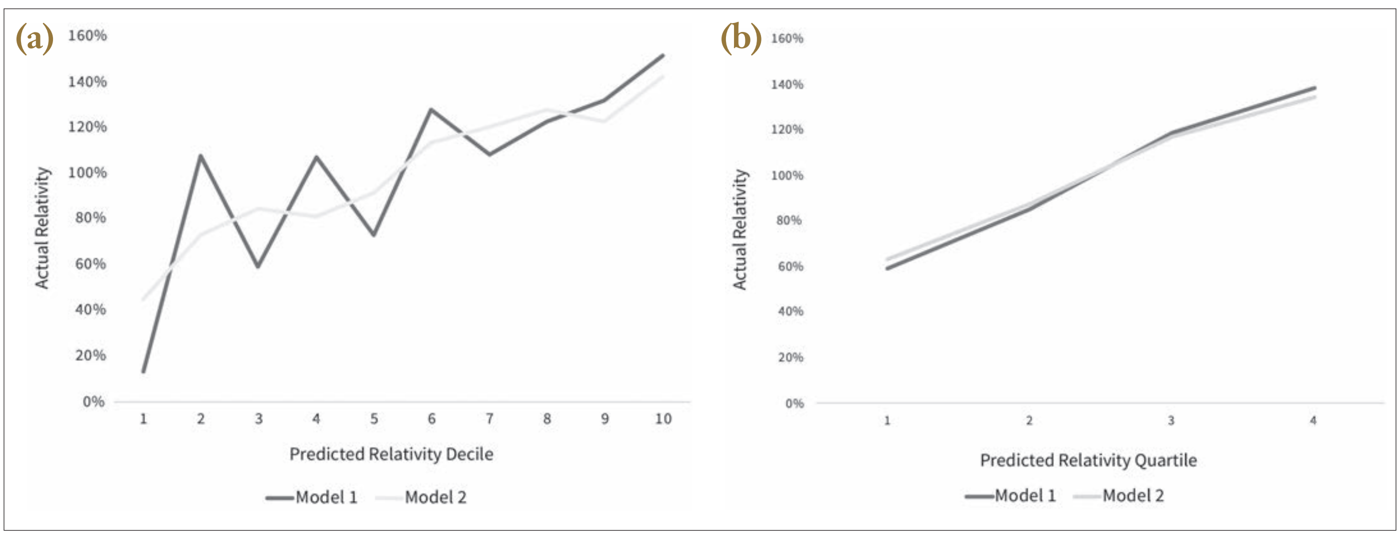


Figure 4
Lift Curve Comparison With Two Different Bin Sizes: Fine (a) and Coarse (b)



Although lift curves are great visual tools, they can be over-interpreted because the binning will always be somewhat arbitrary. For high-variance events such as health care claims or mortality, these small perturbations in the predicted model results can move the predicted value for an observation with a large claim from one bin to the next, and this can result in a significantly different lift curve even if the prediction itself isn't meaningfully different. For instance, a single million-dollar claimaint may have a predicted relativity of 0.899 in one iteration of a predictive model and 0.902 in another version of a predictive model. Because of this, a Lorenz curve is much more invariant than a lift curve to these small perturbations of the scores since it does not include any binning.

For another example, see Figure 4, where I compare two different models (Model 1 and Model 2) against one another. The difference between the two figures here is that I am simply using different bin sizes for my predicted relativities (deciles and quartiles). Although both figures show consistency in the sense that Model 1 stratifies risk better in both figures, the smoothing in Figure 4(b) masks the fact that there appears to be greater volatility in the values for Model 1. However, in this instance, the Lorenz curve may not save us because it also

LIFT CURVES VS. ORDERED LORENZ CURVES

LIFT CURVES

+ Easier to tell how predicted values change as the predicted relativity changes
+ Simple interpretation
+ Easier-to-visualize model bias
- Hard to summarize with a single metric
- Hard to visually distinguish differences between two competing models when the changes in predictive power are marginal
- Small changes in how groupings are determined can correspond to large visual changes

ORDERED LORENZ CURVES

+ Shape is more stable when large claims are present
+ Gini gain summarizes risk stratification in a single statistic that is easily compared across models
+ Not subject to arbitrary groupings
- Hard to tell if the model predictions are biased without using other validation metrics
- Less intuitive and harder to explain

won't obviously show the higher volatility of Model 1 when it is visualized either.

Another word of caution: Although both lift curves and ordered Lorenz curves do a great job of displaying the rank ordering of a model, they don't tell you whether your model is getting you close to your target. For an illustrative example, see Figure 5. If you compare the predicted relativity versus the target for the five observations here, you see a large deviance. The model significantly underpredicts low target values and significantly overpredicts very high model values. However, merely comparing rank orders would suggest a quality predictive model. The lowest actual relativity value corresponds to the lowest predicted relativity, and the highest actual relativity corresponds to the highest predicted relativity.

Figure 5
An Example of a Model That Rank Orders Well but Whose Predicted Values are Biased

Observation	Predicted Relativity	Actual Relativity
1	20%	100%
2	50%	120%
3	120%	140%
4	200%	160%
5	500%	250%

For a lift curve, it is easy to add a quick modification to make this discrepancy easy to visualize. Simply include another line in the graphic that includes the actual values after adjusting for the impact of the predicted relativity (as shown in Figure 3(b) and 4(a)). If the model is unbiased, its predictions will adjust the results to be close to a 100 percent relativity (there will always be some noise as well, so pay attention to sample size). Other validation metrics, such as mean absolute error and root mean-squared error, can be useful here as well.

I use lift curves daily; by pointing out these weaknesses of lift curves, I am not hating on them. However, they can be fickle, so it's important to always interpret them with caution. These pathologies of lift curves are intuitive when you think about them, but it is easy to get careless. That's why I like using Lorenz curves as additional side information. They respond differently to changes in predicted relativities, and when you calculate the Gini gain from them, you are able to summarize the stratification of the model in a single statistic that isn't hampered by the subjectivity of visual interpretation.

In summary, Lorenz curves and Gini gain are good alternatives and complements to lift curves. Together, they form a dynamic combination of ways to measure risk stratification. Don't use just one of them. Use all of them. ■

Michael Niemerg, FSA, MAAA, is a predictive modeling manager for Milliman IntelliScript. He can be contacted at *Michael.Niemerg@milliman.com.*

REFERENCE

Frees, Edward (Jed) W., Glenn Meyers, and A. David, Cummings. 2014. Insurance Rate-making and a Gini Index. *Journal of Risk and Insurance* 81, no. 2:335–366.