RELATIVITY PRICING THROUGH ANALYSIS

OF VARIANCE

by CARL E. CHAMBERLAIN

Reviewed by DANIEL C. GODDARD

Mr. Chamberlain's paper is the first in years to address the problem of calculating class relativities for a two-way (or n-way) classification system. He proposes a new model that offers more flexibility (and more complexity) than previous ones. Essentially, his approach is to fit an additive model to the data, and then fit a multiplicative model to the residuals. As he explains, this approach is suggested by Analysis of Variance theory.

I have a few technical comments on Chamberlain's model. I also will discuss some problems with the model fitting approach in general, which suggest some areas for future research.

In fitting the additive model, Chamberlain minimizes

$$Z=\sum_{i,j} n_{ij}(r_{ij} - \hat{r}_{ij})^2$$

which is "the squared /absolute/ error with appropriate weights." For n_{ij}, he uses exposure in one case, and premium in another. Bailey and Simon, in "Two Studies in Automobile Insurance Rate-making" (PCAS XLVII,1960) suggest minimizing

$$X^2 = \sum_{i,j} n_{ij}\hat{r}_{ij}\left(\frac{r_{ij} - \hat{r}_{ij}}{\hat{r}_{ij}}\right)^2$$

Here n_{ij} is exposure; so $n_{ij}\hat{r}_{ij}$ is proportional to the expected losses. Their formula amounts to the squared relative error, weighted by expected losses.

The choice of absolute or relative error is probably a toss-up. But I believe that expected losses are clearly the better choice for weights. Using just exposures can bias the results. In Chamberlain's auto example, Class 1A accounts for 66% of the exposure but only 52% of the losses.

Also, Bailey and Simon are minimizing Chi-squared, which is used to test how well the model fits. So, their approach will always do better than Chamberlain's on the Chi-squared test. Minimizing Z does have one practical advantage in the additive case, however. The equations can be solved explicitly, while minimizing X^2 requires an iterative approach.

An important purpose of the weights is to act as a "surrogate for credibility". Bailey and Simon explain why expected losses can be used to reflect the relative credibility of the squared error. This is based on what Insurance Services Office in their research calls classical credibility (characterized by a fixed standard for full credibility, and a square root formula for partial credibility).

However, I question whether this weighting really does replace credibility. Consider what happens if we apply any of the models - additive, multiplicative, or Chamberlain's - to a one-way classification system. Consider Bailey and Simon's equations (6) or (9), or Chamberlain's (14) for the case where j has only one value (i.e. the data has only one column). They all reduce to $\hat{r}_i = r_i$. There is no use of credibility left here. The accepted solution for the one-way problem is $\hat{r}_i = Zr_i + (1-Z)\ r$, where Z is the credibility, and r is the value for a larger group. So, our models for n-way class relativities do not work for the special case of n=1. It appears that the weighting by expected losses functions as credibility only to the extent that there are interactions between the dimensions. We are not really using the weighting as a surrogate for credibility; the surrogate is actually the structure of the model which defines the interactions we will consider. In other words, we decide on a model and data is "credible" to the extent that it fits the model.

To get a further sense of what is happening, let us look at the four criteria Bailey and Simon give for an acceptable set of relativities:

Criterion 1. It should reproduce the experience for each class and merit rating class and also the overall experience; i.e., be balanced for each class and in total.

Criterion 2. It should reflect the relative credibility of the various groups involved.

Criterion 3. It should provide a minimal amount of departure from the raw data for the maximum number of people.

Criterion 4. It should produce a rate for each subgroup of risks which is close enough to the experience so that the differences could reasonably be caused by chance.

In the one-way case, balance is taken care of with a balancing or "test correction" factor; it is not a consideration in calculating the relativities. In the two-way case, balance for each class is desirable to insure the model structure is reasonable, but it assumes each class is fully credible in total.

Where criterion 2 calls for reflecting the relative credibility of the groups, criterion 4 in effect calls for reflecting the absolute credibility of each class. These are familiar criteria in the one-way case; in fact, they are the only ones used.

Criterion 3 is a test of how well the model fits. It is irrelevant in the one way case, because we assume no model. So, as we go from the one-way case to the two-way, we add an important and fundamental assumption: there is some rational structure to the interactions between the classes. Do we need this assumption? Should we make it? Presumably, we are trying to get the best estimate of each r_{ij}. So why not calculate the classical credibility Z_{ij} for cell ij, and then

$$\widehat{r}_{ij} = Z_{ij}\, r_{ij} + \left(1 - Z_{ij}\right) r$$

There are at least three problems with this formula. First, we need a standard for full credibility. This has been discussed extensively elsewhere.

Second, what do we use for r? If we set r=r.. (using Bailey and Simon's notation) we are ignoring the information we have about other risks in row i or column j. We have reduced the problem to a one-way classification scheme. Another choice is to use some combination or average of $r_{i.}$ and $r_{.j}$. This is what the NCCI does with the national relativity and the pure premium on level in computing their class relativities.

Third, the classical credibility is independent of our choice of r. This is one of the chief arguments for Bayesian credibility. However, Bayesian credibility has only been developed for the case where each class is a member of only one group. That is, it is just for a one-way classification scheme. ISO has done considerable work on how to group classes where there are several different criteria that could be used. They have suggested using multi-dimensional scaling to reduce all the criteria to a one-way scheme.

It appears what we need is a multi-dimensional credibility theory (which I will leave to more mathematical actuaries than me to develop). Such a theory would solve ISO's grouping problem; it would give us the best estimate for each class relativity; and it would avoid having to guess at an appropriate structure.

There are practical problems with such an approach. A set of relativities so calculated would have to be published as a table, not as a few parameters and a formula. In some cases this is no hardship. In Chamberlain's property example, he used twelve parameters and a fairly complicated formula to fit a table of nine numbers. Other cases are not so easy. For example, multi-dimensional credibility would give us a different set of class factors for each auto rate territory. Apart from adding many pages to the rating manual, such a change would require major changes in most automated systems. So, we would probably want to select one or two sets of class factors that are "close enough" for most territories. In other words, we would fit a model to simplify the structure. Having already credibility weighted the data, we can attribute any residual error to the choice of the model. The present procedure cannot distinguish between errors due to the choice of model, and errors due to statistical fluctuations in the data.

This leads to a new perspective on Chamberlain's model. We start by assuming that there is a pattern to the relativities, and our estimates should reflect as much of this pattern as possible. So, we start by fitting a first-order additive model. We then test whether a second-order multiplicative model shows any significant remaining pattern in the residuals. In theory, we could go on fitting higher order models, until the F value is no longer significant. In practice, with the size of the data sets usually encountered, I would expect two stages to be sufficient. This procedure should extract the maximum possible pattern from the data, in the same sense that polynomial stepwise regression does for a time series. Just as with polynomial stepwise regression, Chamberlain's procedure does not necessarily give the simplest or most efficient model. And the fact that it detects a pattern is no assurance that the pattern is reasonable.

This brings me back to my main point. Our procedures for two-way relativities are based on a very different point of view than those for one-way relativities. I believe we need a multi-dimensional credibility theory to reconcile the two.