Credibility for Pricing Loss Ratios and Loss Costs

Uri Korn, FCAS, MAAA

Abstract
This paper discusses how credibility can be applied to pricing loss ratios and loss costs. A method is also presented that can perform a credibility weighted allocation of losses without changing the overall average, which often occurs when applying credibility. Finally, it is shown how Generalized Linear Mixed Models can be used to credibility weight loss ratios while taking multiple dimensions into account. Workarounds are shown for some common pitfalls, and it is explained how to implement these models in spreadsheets.

Keywords. Bühlmann-Straub Credibility, Bayesian Credibility, Loss Ratios, Loss Costs, Generalized Linear Mixed Models

1. INTRODUCTION

When doing any type of actuarial analysis, credibility is an issue that must be frequently dealt with. However, many seemingly simple credibility applications are difficult to apply in practice. For loss ratios and loss costs, seemingly simple concepts such as how to calculate the Bühlmann-Straub parameters or how to perform a credibility weighted allocation are difficult to apply in practice.

In this paper, we discuss these and other practical issues that arise when using credibility with loss ratios and loss costs. For our discussion, we will use the term loss ratio for brevity, but everything mentioned is applicable to loss costs as well. For loss costs, "exposures" should be substituted for "premium" for everything below.

1.1 Outline

We will start our discussion with Bühlmann-Straub credibility and how to apply it to loss ratios and loss costs. The following section discusses the recommended method for calculating loss ratios for pricing studies when credibility is being performed. Section 4 introduces a credibility model that ensures that the credibility weighted results always tie to the original loss ratio. This method is especially useful for performing a credibility weighted allocation of a selected loss ratio. And finally, section 5 discusses the use of mixed models to perform the credibility weighting. It also discusses dealing with some common pitfalls and shows how to implement these models in a spreadsheet or other environment.

2. BÜHLMANN-STRAUB CREDIBILITY

The first topic that will be discussed is how to calculate the Bühlmann-Straub parameters. This includes calculation of the within variance and the between variance. The formulas for each are shown below (Dean

2005).

$$\hat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} W_{gn} (X_{gn} - \bar{X}_g)^2}{\sum_{g=1}^{G} (N_g - 1)} \quad \textbf{(2.1)}$$

$$\hat{VHM} = \frac{\sum_{g=1}^{G} W_g (\bar{X}_g - \bar{X})^2 - (G-1)\hat{EPV}}{W - \frac{\sum_{g=1}^{G} W_g^2}{W}} \quad \textbf{(2.2)}$$

Where EPV is the expected value of the process variance, or the "within variance", and VHM is the variance of the hypothetical means, or the "between variance". *W* is the weight, *G* is the number of groups, *N* is the number of periods, X_{gn} is the indication for group *g* in period *n*, $\bar{X}_g$ is the average for group *g* across all periods, and $\bar{X}$ is the average across all groups and periods.

2.1 The Within Variance

For loss ratios, we will assume that the variance of total losses is proportional to the premium, which implies that the variance of a loss ratio is proportional to the inverse of premium (since calculating the variance of the latter involves dividing the former by the square of the premium). A closer look shows that this must be the case, since the variance of total losses for two (uncorrelated) accounts is equal to the sum of the individual variances. Assuming any other relationship between premium and variance will not agree with this result and will lead to inconsistencies. Similarly, for loss costs, we will assume that the variance is inversely proportional to the exposures.

The next question is what data should be used for calculating this parameter. The answer is that it should be based off of the observed experience, although this is not as straightforward as it sounds. The variance should not be based off of the final selected estimates for each year by using a Bornhuetter-Ferguson method; doing so artificially reduces the variance since each year is moved closer to the a priori estimate and so does not represent the true volatility in the data. Instead, we recommend using an approach similar to the Cape Cod method that compares actual paid or reported losses to used premiums, which are premiums divided by the loss development factor. If the data is capped and excess ratios are used to produce final uncapped loss ratios, then the excess ratios should be applied to the premiums as well to produce used, capped premium, since this reflects the premium relevant to the capped losses. If we want, we can also reflect the fact that some of volatility observed in the loss ratios is due to yearly changes that are not captured in trend or rate changes. We can take this into account and give older, less predictive years less weight by applying an exponential decay factor to the weights as well. This will be discussed further later on. Doing this will reflect the level of credibility inherent in each year and group, and this is the weight that should be used in the formulas above. Dividing capped paid or reported losses by used, capped premium is mathematically equivalent to using the chain ladder estimates for the ultimate loss ratios multiplied or divided by one minus the excess ratios, depending on how the excess ratios are expressed. So, we are essentially using chain ladder ultimate loss ratios with weights for each year as described. Using this method, we can analyze the actual experience that has emerged and the volatility estimates will be appropriate.

Note that the within variance formula above multiplies the differences squared by the weights, but does not divide by the total of the weights afterwards. This is because the within variance used in the Bühlmann-Straub formula is really more accurately described as a within variance factor, and not the actual within variance for anything in particular. This can be seen from the Bühlmann-Straub credibility formula as well; rearranging the formula below shows that this parameter is divided by the dollar amount to come up with the final within variance.

$$Z = \frac{N}{N + W/A} = \frac{1}{1 + \frac{W/N}{A}} = \frac{1}{1 + V/A}$$

Where N is the weight, W is the within variance from the Bühlmann-Straub formula, or the within variance factor as we will call it, V is the actual within variance, and A is the between variance.

The within variance formula (2.1) assumes that the product of the weights and the square differences from the mean all have the same expected value. The square differences from the mean represent the variance component. So, by taking the average of these values as the within variance factor, this formula essentially assumes that the variances of each year multiplied by the weight are consistent, which is the same as assuming that the weights are proportional to the variance.

Lastly, we will note that formula (2.1) takes an average of the within variance factors by segment, only weighting by the number of years, but not the premium volume. If one wishes, one can modify the formula and use a weighted average by premium volume instead.

2.2 The Between Variance

The second parameter, the between variance, is even more volatile and difficult to calculate than the first. When constructing a hierarchical model with multiple levels, for smaller, lower down levels, if the estimates of this parameter appear unreasonable, assumptions can be made for how each level's between variance relates to the levels above it, and it can then be judgmentally selected accordingly. This parameter is easier to calculate with more groups, and so it can also be calculated between finer segmentations than being used, and then judgmentally adjusted as well. The formula shown above (2.2) can sometimes return negative values, which means that the indicated between variance is zero.

This formula also assumes that the within variance factor is the same for all groups. Using the logic we discussed, the following formula can be used when the within variance factor is assumed to differ among segments. Caution should be used when doing this however; the within variance is difficult to calculate due to data volatility. It is normally best to use an average across segments for everything. This should only be done in some special cases where the within variance is expected to be significantly different between groups, such as when working with primary and excess data together.

$$\hat{VHM} = \frac{\sum_{g=1}^{G} W_g [(\bar{X}_g - \bar{X})^2 - \frac{(G-1)}{G} \frac{\hat{EPV}_g}{W_g}]}{W - \frac{\sum_{g=1}^{G} W_g^2}{W}} \tag{2.3}$$

3. CALCULATION OF LOSS RATIOS

For the loss ratios used in any credibility method, we recommend using similar guidelines as mentioned, at least as a starting point. This is not essential however, except for when working with mixed models, which will be discussed later. To recap, the loss ratios for each year are equal to the capped paid or reported losses divided by the used, capped premium, which is the premium divided by the LDF and then multiplied or divided by one minus the excess ratio (ignoring trend and on-leveling). As we mentioned, this is equivalent to multiplying the capped loss ratios by the LDF and then multiplying or dividing by one minus the excess ratio.

If the losses are capped, the loss ratios produced from the credibility procedure should be adjusted to reflect the fact that we are only analyzing a portion of the losses. The final loss ratio should be taken as a weighted average of the credibility weighted result and the overall average loss ratio with weights of one minus the excess ratio and the excess ratio, respectively (assuming that the excess ratio is expressed as a percentage of total losses). This approach assumes that the excess portion that we are not analyzing is running the same as the average for all segments. (It is also acceptable, however, to assume that the excess portion for each segment is running the same as the capped portion and to skip this adjustment, if one desires.)

An additional factor should also be applied to the weight for each year so that more recent years which have more predictive power for the going forward loss ratio receive more weight. This factor is needed since a Bornhuetter-Ferguson method is normally done using the a priori loss ratio obtained from the Cape Cod method. This step uses the a priori loss ratio, but effectively gives even more weight to the recent years, which have the most predictive power for the going forward loss ratio. Since we are skipping this step, we need another way to give some more weight to the latest years. An exponential decay factor of 0.95 should give similar indications as the full Cape Cod/Bornhuetter-Ferguson method, depending on the LDFs, and a higher or lower factor can be used based on the perceived rate that the business is changing. So, to recap, the weights should be the used premium multiplied or divided by one minus the excess ratio, multiplied by the yearly decay factor. To keep the weights given to each segment appropriate, the total premium should be used as the weight when combining across segments, and the weights mentioned should only be used when aggregating results across years. If the same premium is used for multiple LDF segmentations, the used premium can be calculated using the implied LDF, that is, the total calculated chain ladder ultimate divided by the paid or reported losses.

There are many advantages to this approach. The first is that loss ratios produced in this fashion are a good representation of the actual experience for each year, and the weights correspond to the amount of credibility inherent in each year's estimate; this makes the data well suited for a credibility routine. Second, it is easier to streamline and automate than a Bornhuetter-Ferguson or other similar method, especially when there are many segmentations in the data. Third, it makes it easier to apply assumptions at finer levels of detail than the Bornhuetter-Ferguson method. Lastly, the final weights given to each year are more explicit instead of being implied from the loss development pattern.

There is sometimes some confusion that a Bornhuetter-Ferguson method already performs credibility weighting. This is only true from a reserving perspective, but not from a going forward profitability point of view. A Bornhuetter-Ferguson method gives more weight to the a priori loss ratio for more recent, greener years for which the IBNR for those years are more uncertain. But from a going forward perspective, even if all losses came in instantaneously and there was no need for any loss development, there would still be a need to credibility weight results because of the volatility inherent in the experience. For complete years, the amount of credibility for each year depends on the premium volume. For incomplete years, it is the premium multiplied by the percentage of the year that we have already observed. (The variance is really slightly higher because of the uncertainty in the estimation of the LDFs, but accounting for this would just give more weight to older years, which is counter-intuitive.) So, for a going forward, pricing perspective, if credibility is being applied, we recommend not using the Bornhuetter-Ferguson method and sticking with the original chain ladder method with weights as described. Doing this is non-essential, however, as we mentioned, except for mixed models. But regardless of which methods are used to calculate the actual loss ratios, the Bornhuetter-Ferguson results are <u>not</u> appropriate for the calculation of the within variance.

We will illustrate one way of performing this method with an example: we are developing a book of business that contains segments and sub-segments that we wish to perform credibility on in a hierarchical fashion. We group the data at the sub-segment level. We then calculate three values for each sub-segment for each year: the on-level premium; the trended, uncapped loss ratio; and the weight. The calculation of the on-level premium is straightforward. The calculation of the latter two is shown in these two formulas (assuming that our analysis is performed on capped, reported losses and that excess ratios are expressed as a percentage of total losses):

$$Loss\ Ratio = \frac{Capped\ Reported\ Losses}{On\text{-}Level\ Premium} \times LDF \times Trend\ Factor / (1 - Excess\ Ratio) \qquad \textbf{(3.1)}$$

$$Initial\ Weight = On\text{-}Level\ Premium / LDF \times (1 - Excess\ Ratio) \times Yearly\ Decay\ Factor \qquad \textbf{(3.2)}$$

Using these initial weights in the credibility calculation would cause improper weights being given to each segment and sub-segment that are not based on the total premiums of each. To use the total premium as the weights, but still perform the Cape Cod approach as we described above, we apply an off-balance factor for each sub-segment and calculate the final weights used as follows: (Subscripts are used in the below for added clarity; they were ignored in 3.1 and 3.2 for brevity.)

$$Off\text{-}Balance\ Factor_{\text{Sub-Segment}} = \frac{\sum^{All\ Years} On\text{-}Level\ Premium_{\text{Sub-Segment}}}{\sum^{All\ Years} Initial\ Weights_{\text{Sub-Segment}}} \qquad \textbf{(3.3)}$$

$$Final\ Weight_{\text{Sub-Segment}, Year} = Intial\ Weight_{\text{Sub-Segment}, Year} \times Off\text{-}Balance\ Factor_{\text{Sub-Segment}} \qquad \textbf{(3.4)}$$

The final loss ratio to use as the input for each sub-segment is calculated by taking the weighted average of the yearly loss ratios using this as the weight. With this approach, summing up the results by segment and year and then calculating the segment loss ratios will tie to the sum of the sub-segment loss ratios, which is clearly a desired condition. These final weights can also be used as the base for final weights in a Generalized Linear Mixed Model (GLMM) or a Bayesian credibility model and both the regression weights and the relative credibility by year will be appropriate. (A further step is really needed for GLMMs, which will be discussed later.) Once the credibility procedure is run, the final selected loss ratios are equal to:

$$LR_{\text{Sub-Segment}} = Credibility\ Loss\ Ratio_{\text{Sub-Segment}} \times (1 - XSR_{\text{Sub-Segment}}) + Average\ LR \times XSR_{\text{Sub-Segment}} \qquad \textbf{(3.5)}$$

Where *LR* is the loss ratio and *XSR* is the excess ratio. As mentioned, it is also acceptable to skip this last step. As a compromise, instead of using the overall average loss ratio, the loss ratio from the corresponding segment can be used as well.

As a slight alternative, it is also possible to develop losses and calculate the initial weights at the policy level. Results can then be rolled up into sub-segments by adding the ultimate losses and the initial weights from each policy. The final weights can then be calculated at this level, although it is possible to calculate them at the policy level as well. Doing this yields the same results, but allows for more flexibility in the segmentation structure used for credibility and also makes it easier to use assumptions, such as LDFs and excess ratios, at the policy level.

As mentioned, this approach produces data that fits very nicely into a credibility procedure. Another benefit is that the segmentation structure has less of an impact on the final results than a similar Bornhuetter-Ferguson method.

4. THE TUG-OF-WAR CREDIBILITY METHOD

We will introduce a credibility method that ensures that the average of the resulting credibility weighted results matches the original. This method is well suited for performing a credibility weighted allocation but has other uses that will be discussed. We will focus on loss ratios, although this method can be applied to other items as well.

A frequent problem with applying credibility to loss ratios, is that the average of the credibility weighted results often does not match the original. This causes practical issues since now we must either change our originally selected overall estimate or else the sum of the segments will not tie to the combined. A common solution is to apply an off-balance factor that forces the average of the credibility weighted loss ratios to equal the original overall average, but doing so often produces questionable results, especially when the segments are small and when this off-balance factor is large.

These problems will be demonstrated with the following example: We are analyzing a book of business with a total premium volume of $200 million, which consists of one very large segment with $100 million of premium and a bunch of smaller segments that in total make up the other $100 million. The total loss ratio is judged to be 70%, and we wish to produce credibility weighted loss ratios for each of the segments. The loss ratio of the large segment is 90% and is almost fully credible. The smaller segments have an average loss ratio of 50%, and because of their size, have almost no credibility. If we calculated credibility weighted loss ratios for each segment, the large segment would end up with a loss ratio close to 90%, and the smaller segments would be assigned loss ratios close to the overall mean, which is 70%. Each of these results seem to make sense at the individual level, but summing up all the parts, our average loss ratio for the book is now around 80%, much higher than the originally estimated 70%. If we applied an off-balance factor to each of the loss ratios, the factor would be equal to 0.7 / 0.8 = 0.875. The large segment would now have a loss ratio of 0.9 x 0.875 = 78.75%, and each of the smaller segments would have loss ratios of 0.7 x 0.875 = 61.25%. The combined average loss ratio is now 70%, as expected, but the results by segment are no longer reasonable. The large, almost fully credible segment is not given enough credibility, only around 50%, and the smaller segments are given way too much.

However, if we took a closer look at the above, the results before the off-balance factor may be problematic as well. If the total loss ratio is 70% and there is one large, nearly fully credible segment with a loss ratio of 90%, then this should imply that the total loss ratio of the smaller segments is 50%. In fact, if we conducted our analysis removing the large segment, this is what we would expect to see. The average loss ratio of the smaller segments can be deduced from what we know about the larger segment. Neither method above takes this into account since they both look at each segment individually, ignoring the results of the other segments.

4.1 Using Bayesian Credibility

To implement this method, we will be using a simple Bayesian credibility model that does not require any special software to run. The results of this model are also consistent with Bühlmann-Straub credibility as will be shown. The reason for using the Bayesian version is because the Bühlmann-Straub method only produces a point estimate, whereas we need to know the entire distribution so that we can find the most optimal solution subject to the constraint that the results must tie to the original overall number. This can only be done using the Bayesian version.

We will be using a normal distribution to model loss ratios, although with variances that differ for each observation. Note that this assumption is not the same as assuming that these items are normally distributed; we are only assuming that each individual loss ratio has a normal distribution on what its possible outcomes might have been. In this way, it is more similar to kernel smoothing than to assuming a distribution. Assuming normality with variances inversely proportional to the dollar amount also produces the same results as taking a weighted average by the dollar amounts, and so is consistent with traditional actuarial analysis.

We will also be assuming that the prior distribution (that is, the credibility complement, in Bayesian terms) is normal as well, which is the common assumption. This is a conjugate prior and the resulting posterior distribution (that is, the credibility weighted result) will also be normal. Only when we assume normality for both the observations and the prior, Bayesian credibility produces the same results as Bühlmann-Straub credibility. The mean of this posterior normal distribution is equal to the weighted average of the actual and prior means, with weights equal to the inverse of the variances of each. As for the variance, the inverse of the variance is equal to the sum of the inverses of the within and between variances (Bolstad 2007). The variance of the item being credibility weighted is comparable to the within variance, and the variance of the prior is comparable to the between variance. This means that the resulting credibility assigned is equal to the inverse of the within variance divided by the sum of the inverses of both the within variance and the between variance. Using some algebra:

$$Z = \frac{1/V}{1/V + 1/A} \times \frac{V}{V} = \frac{1}{1 + V/A}$$

Where V is the within variance and A is the between variance (or equivalently the variance of the prior distribution). Examining the Bühlmann-Straub credibility formula again, where W is the within variance factor:

$$Z = \frac{N}{N + W/A} = \frac{1}{1 + \frac{W/N}{A}} = \frac{1}{1 + V/A}$$

So, it can be seen that when using normal distributions, Bayesian credibility is equivalent to Bühlmann-Straub credibility. The likelihood formula for this Bayesian model is:

$$\begin{gathered} N(\mathit{Credibility\ Result, Actual\ Result, Within\ Variance}) \\ + N(\mathit{Credibility\ Result, Credibility\ Complement, Between\ Variance}) \end{gathered} \tag{4.1}$$

Where *N(A, B, C)* is the logarithm of the probability density function (PDF) of a normal distribution at *A* with a mean of *B* and variance of *C*. Maximizing the likelihood of this formula will produce the mentioned result. As an alternative, it is also possible to use the formulas for the mean and variance of the posterior normal distribution that we mentioned. (As a practical issue when programming, it may be necessary to set a minimum on the PDF values so that they are not too close to zero, which can cause problems with logarithms.)

This simple Bayesian model can be solved using only Maximum Likelihood Estimation (MLE). Since the resulting posterior distribution is normally distributed, the mode of this distribution is equal to the mean, as is known. This means that the MLE, which returns the mode, will also be returning the mean in this case.

4.2 Implementing the Method

To implement the Tug-of-War method, we maximize the likelihood of the credibility weighted loss ratios, while constraining the parameters so that the resulting average will match the original.

To do this, we start with initial parameters that represent the relative amount of the total losses allocated to each segment. We then use these initial parameters to calculate percentages that will always add up to one by taking the initial parameter for each segment and dividing by the sum of all of the initial parameters. We then convert this into a loss ratio by multiplying each percentage by the total amount of losses across all segments and then divide by the premium for each segment. We then calculate the likelihood for each loss ratio using the Bayesian credibility formula shown above. Since each loss ratio affects all of the others, we need to weight the likelihood of each segment to account for this. The weights used for each segment should be the premium. (To use weights in MLE, each log-likelihood should be multiplied by the weight.) The initial parameters are set using an optimization routine that maximizes the total likelihood.

In practice, it helps if the initial parameters are on a logarithmic scale so that negative numbers do not cause problems with negative loss ratios. The parameter of one of the segments can be fixed at a number such as zero or another value, since the real number of parameters is one less than the number of segments since the sum of the percentages must equal one. Also, to help ensure that the maximization routine converges to the correct solution, good starting values should be chosen; these can be obtained from the regular Bühlmann-Straub indications.

There are multiple ways to implement the above scheme. Another way is to set the percentages of one of the segments to one minus the sum of the rest, although this can sometimes result in negative percentages. Another version that is sometimes helpful is to use relativities instead of percentages. In this version, the initial parameters are the initial relativities (on a logarithmic scale). The average relativity is then calculated by taking a weighted average of these initial relativities using the premium as the weight. The final relativities are then set to the initial relativities divided by the average. This will ensure that the resulting average matches the original. Note that in both versions, credibility is calculated on the loss ratios themselves and not on the percentages or the relativities.

To review, the steps are as follows:

1) $\textit{Initial Parameters}\,(\textit{Set by Maximization Routine})$

2) $\textit{Relative Percent of Losses}_i = \exp(\textit{Initial Parameter}_i)$

3) $\textit{Percent of Losses}_i = \dfrac{\textit{Relative Percent of Losses}_i}{\sum \textit{Relative Percent of Losses}}$

4) $\textit{Loss Ratio}_i = \textit{Percent of Losses}_i \times \textit{Total Losses} / \textit{Premium}_i$

5) $\text{Log-}\textit{Likelihood}_i = \text{Log-}\textit{Likelihood}\,(\textit{Loss Ratio}_i) \times \textit{Premium}_i$

6) $\textit{Total}\ \text{Log-}\textit{Likelihood} = \sum \text{Log-}\textit{Likelihood}_i$

If implementing the relativities version, the steps are slightly different:

1) $\textit{Initial Parameters}\,(\textit{Set by Maximization Routine})$

2) $\textit{Initial Relativity}_i = \exp(\textit{Initial Parameter}_i)$

3) $Average\ Relativity = \frac{\sum Initial\ Relativities_i \times Premium_i}{\sum Premium_i}$

4) $Relativity_i = Initial\ Relativity_i\ /\ Average\ Relativity$

5) $Loss\ Ratio_i = Relativity_i \times Overall\ Loss\ Ratio$

6) $\text{Log-}\ Likelihood_i = \text{Log-}\ Likelihood\ (Loss\ Ratio_i) \times Premium_i$

7) $Total\ \text{Log-}\ Likelihood = \sum \text{Log-}\ Likelihood_i$

We named this method the Tug-of-War method because each loss ratio tries to maximize its own likelihood, and because of the constraint that the resulting average must equal the original, each loss ratio "tugs" on all of the others as they fight for the highest likelihood that they can achieve. This method produces better results than the application of a simple off-balancing factor, since the likelihood is maximized over all possible combinations that tie to the original average, and so the best tying result is selected. The complement for each segment is essentially revised based on available information from the other segments. It should be noted though that if the off-balance is small, there may not be much benefit to using this more complicated method, and the use of a simpler off-balancing factor may be preferable.

For both the overall loss ratio used as the complement of credibility as well as the individual segment loss ratios used in this model, they can be either actual loss ratios dictated solely from the experience, or they can be selected with some degree of judgment. If the overall loss ratio used is a selected loss ratio, and the segment loss ratios are from the experience, this method is essentially performing a credibility-weighted allocation of the selected loss ratio. A hierarchical model can also be built where the overall loss ratio used for each level is the credibility weighted result from the previous level. Alternatively, if the segment loss ratios are judgmentally selected, and the overall is set to the average of these loss ratios, then this method performs a credibility weighting on the selected loss ratios.

As another similar option, it is possible to use the actual, experience dictated loss ratios for both the overall and the segments and have this method take care of all the selections via credibility weighting, since with a good credibility method there is less need to manually select loss ratios. Adjustments can be made afterwards though, if needed. A hierarchical model can be built similar to the above, as well. It is suggested to use loss ratios and weights as was explained above, but any reasonable method can be used as long as the within and between variances are calculated correctly.

A couple of examples of applying this method are shown below for illustration. The first is very similar to the one given above but shows the actual estimate produced from applying this method in practice.

	Total	Segment 1	Segment 2	Segment 3	Segments 4 - 21
Total Premium	20M	10M	500K	500K	500K
Loss Ratio	70%	100%	40%	40%	40%
Within Standard Deviation	3.9%	31.6%	31.6%	31.6%	31.6%
Between Standard Deviation	10%				
Tug-of-War LR	70%	93.2%	46.8%	46.8%	46.8%
Implied Credibility		77.2%	77.2%	77.2%	77.2%
Bühlmann-Straub LR	81.7%	96.1%	67.3%	67.3%	67.3%
Bühlmann-Straub Credibility		87.0%	9.1%	9.1%	9.1%

Note that with this method, the large segment receives slightly less credibility than it does using the Bühlmann-Straub method. This is because the result of this large segment affects not only its own loss ratio, but all of the other segments as well.

The next example is nearly identical except that one of the smaller segments, segment 3, has a higher loss ratio of 80%. The details are shown below. The point of this example is to show that negative credibilities are possible since the large segment with ten million in premium and a very high loss ratio essentially lowers the complement of credibility for the remaining segments, since, as we have mentioned, we would expect to see an overall lower loss ratio if we performed the analysis without this large segment. Note though that the resulting Tug-of-War loss ratio for this segment still comes out higher than the other small segments, as expected.

	Total	**Segment 1**	**Segment 2**	**Segment 3**	**Segments 4 - 21**
Total Premium	20M	10M	500K	500K	500K
Loss Ratio	71%	100%	40%	**80%**	40%
Within Standard Deviation	3.9%	31.6%	31.6%	31.6%	31.6%
Between Standard Deviation	10%				
Tug-of-War LR	71%	93.4%	48.4%	**52.1%**	48.4%
Implied Credibility		77.2%	72.8%	**-210.3%**	72.8%
Bühlmann-Straub LR	82.3%	96.2%	68.2%	71.8%	68.2%
Credibility		87.0%	9.1%	9.1%	9.1%

These results and implied credibilities will be explained more in the next section as well.

4.3 Understanding the Results

The loss ratios resulting from this method can sometimes be difficult to interpret at first glance. Even though the correlation between the resulting loss ratios from this method and the Bühlmann-Straub method are usually very high, the relationship between the credibility numbers is less apparent at first. In the simple examples shown in the previous section, it was relatively easy to understand the results, but more realistic scenarios can be more difficult to interpret.

As we explained above, the complement of credibility is effectively changed with this method as it takes all of the information about the expected average loss ratio and the other segment's loss ratios into account. A segment's loss ratio is impacted by the other segments' loss ratios since they provide information and can be used to imply something about our current loss ratio. The amount of impact other loss ratios affect each other is related to how credible each loss ratio is. Using this logic, we can produce a formula to derive what the effective complement for each segment's loss ratio is. We do this by starting with the total losses for the entire book and subtracting out the amount of losses from all of the other segments using the Bühlmann-Straub derived loss ratios. But subtracting out all of these losses would be giving the effect that segments have on each other too much weight. To account for the partial credibility of these loss ratios, we subtract out only a portion of the losses; for this fraction, we use the calculated Bühlmann-Straub credibilities as an approximation. We then divide by the appropriate premium volume to convert these losses into loss ratios.

With this formula, each group receives a different effective complement based on the loss ratios and relative weights of all of the other segments. The formula is as follows:

$$Complement = \frac{Total\ Premium \times Average\ LR - \sum_{i=All\ Other\ Segments} Premium_i \times Cred\ LR_i \times Z_i}{Total\ Premium - \sum_{i=All\ Other\ Segments} Premium_i \times Z_i} \qquad \textbf{(4.2)}$$

Where *Cred LR* is the Bühlmann-Straub loss ratio and *Z* is the credibility. The implied credibility from this new effective complement can be calculated by inverting the credibility formula and solving for *Z*, which results in the following:

$$Z = \frac{LR_{Cred} - LR_{Complement}}{LR_{Segment} - LR_{Complement}} \qquad \textbf{(4.3)}$$

These resulting credibilities will not match the Bühlmann-Straub credibilities exactly, but the correlation is usually very high, and these can be used to help explain the results.

As mentioned, some of the resulting loss ratios may not fall in between the (original) complement and the initially indicated loss ratio. Even though we can understand and explain the results, this may still be undesirable. A simple solution is to just select different loss ratios for these segments. This occurs most often with smaller segments and so the impact to the overall average will be small. Another solution is to apply a penalty to the likelihood to help keep the results within range. One way to do this is to subtract from the likelihood the product of the amount that the loss ratio is out of the range by some small penalty constant. (This should be done within the parenthesis before the likelihood is multiplied by the premium so that the penalty is multiplied by the premium volume as well; this seemed to work best. Also, the penalty should usually be less than one or two.) This approach will not guarantee that the loss ratios remain within the range, but it will help push them closer and make being outside of the range less likely. Note, however, that using a penalty puts more constraints on the loss ratios and often lowers the correlation between the implied credibilities and the original and so may make the other loss ratios more difficult to explain.

4.4 Using Classical Credibility

Even though this method requires the within and between variance parameters, it can also be implemented in a classical credibility-like (or limited fluctuation) fashion, if desired. Even though classical credibility has some guidelines for selecting different credibility thresholds, such as having the estimate not deviate by more than 5% from the true value 90% of the time etc., any such selections for these parameters are mostly arbitrary. That is not to say that there are any problems with using classical credibility; it is just important to realize the need for judgmental estimates and not assume that the results are more objective than they really are. Classical credibility can provide reasonable credibility weighted results in a small amount of time, which in itself is a lot to say in its support.

The Bühlmann-Straub credibility formula is $N / (N + K)$. This formula will assign 50% credibility when $N = K$, and so K can be thought of as the criteria for half credibility. The premium threshold for this can be judgmentally selected. Alternatively, if one is more comfortable with choosing a full credibility threshold, a threshold can be selected for approximate full credibility, and we can then rearrange the classical credibility formula of $Z = \sqrt{X / K}$ to $K_{Half\ Credibility\ Criteria} = 0.25\, K_{Full\ Credibility\ Criteria}$ to convert this into a rough half credibility threshold, although of course, this will not be exact.

Using this, the within variance for a segment can be set as the inverse of the dollar amount being used as the weight multiplied by a factor. The between variance can be set to the inverse of the dollar amount that should receive half credibility multiplied by the same factor. The actual factor used has no impact; it is just needed to put the variances on an appropriate scale so that the method can converge.

5. USING GENERALIZED LINEAR MIXED MODELS

5.1 Credibility Weighting Loss Ratios and Loss Costs

As an alternative to the methods presented above, it is also possible to use a Generalized Linear Mixed Model (GLMM) to credibility weight loss ratios and loss costs. See Klinker (2011) for an introduction to these models. Using a GLMM with an identity-link and a normal distribution will produce the same results as applying Bühlmann-Straub credibility. Besides for the benefits it offers of easily allowing hierarchical and multidimensional models, using a GLMM automates the calculations of the within and between variances.

A problem, however, arises when using premiums as the base for the weights, since a GLMM assumes that a weight represents a number of observations. Because of this, using premium will almost always result in assigning full credibility to everything since each premium dollar will be counted as an observation and so the number of observations will be very high[1]. Using an alternative weight, such as claim counts, does not fulfill the desired objective of weighting by premiums, since GLMMs use the same weights for credibility as they do for the regression. Weighting by counts may also cause a bias if there are some segments with high frequency, low severity claims that have a high loss ratio and vice versa, for example. One solution is to multiply the weights by an additional constant equal to the total number of reported claims across all

1 The referenced paper actually shows an example using premium as weights but this appears to be an error.

segments divided by the sum of the original weights so that the new sum of the weights across all segments will be equal to the total number of reported claims. This will allow us to weight by premium volume but still keep the total weight consistent with the number of observations overall. This approach produces reasonable credibility estimates when applied in practice. To summarize, using this method in addition to what we discussed above, the weights should be equal to the following:

$$\textit{Premium} \;/\; \textit{LDF} \times (1 - \textit{Excess Ratio}) \times \textit{Yearly Weight Factor} \times K \qquad \textbf{(5.1)}$$

Where K is the factor that we mentioned[2]. Note that there is only one K factor for all of the data and it has the same value for every segment, regardless of the actual number of claim counts for each. Using this approach, it is possible to build hierarchical and multidimensional credibility models using GLMMs.

Using a GLMM also allows us to use a log-link when credibility weighting, which sometimes produces better results than an identity-link when there are extreme values, as there often are with volatile data, but not always; both ways can be tested to see which produces better results. To avoid errors caused from taking the logarithm of zero, observations with loss ratios of zero should be modified to a very small number slightly above zero, such as 0.00001. Also, even without a log-link, loss ratios with zero weights should be removed so as to not cause errors, which will occur with some GLMM implementations if left in.

5.2 Multidimensional Credibility Models

With GLMMs, it is also possible to build a multidimensional credibility model in which each dimension is assigned a relativity, and each relativity is credibility weighted back towards zero. For multidimensional models, multiplicative relativities usually behave much better and are recommended.

Assuming we have two dimensions and we wish to perform credibility weighting on the relativities of each, there are two main types of models we can build, and another that is a compromise of these two approaches, as will be explained. For this section, we will assume that the two dimensions we are dealing with are industry and territory.

The first type of model is a true two dimensional model where the resulting loss ratios are the product of the two relativities. This assumes that territory relativities are the same for each industry (and vice versa). So if a particular territory is higher than average overall, it will be higher for every single industry by the exact same amount. A positive of this model is that it leverages the credibility of each territory across all industries. But this is a down side as well since it assumes the relativities are always the same, which they will not always be.

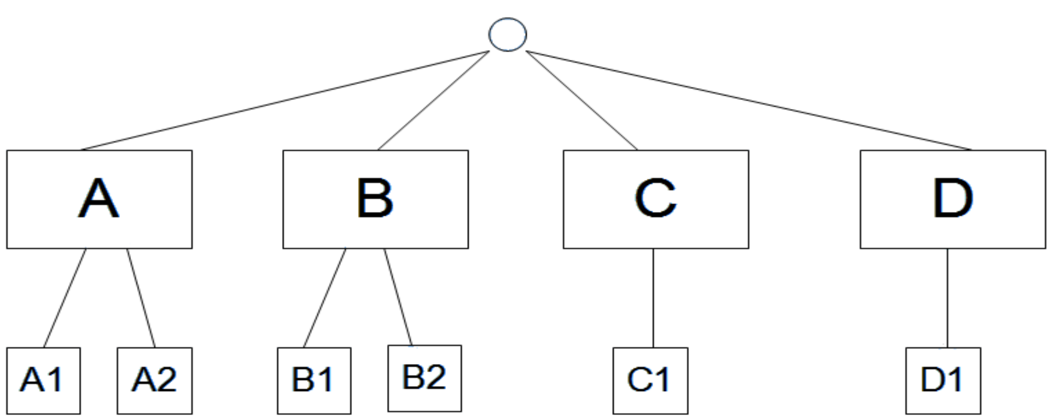
2 Note that this additional factor is not needed for Bayesian models.

The second type of model we can build is really a hierarchical model. For example, we can put territory under industry and first perform credibility at the industry level. We then perform a separate territory credibility calculation for each industry. This will allow territory relativities to differ by industry, but does not leverage the credibility of a territory across industries. Which of these two models to choose depends on our perception of how different the territory relativities are across industries and how volatile our data is. Both of these models can be calculated using GLMMs.

A compromise model can also be built that leverages the credibility of each territory across the industries, but also allows each industry's territory relativities to differ based on the amount of credibility within each cell. In this model, the territory relativities for each industry are effectively credibility weighted back towards the overall territory relativity, which itself is credibility weighted back towards zero. This is the ideal model that combines the best points of each of the above models. Such a model can be built using a GLMM with both industry and territory included as random effects (that is, included as components of the model that take credibility into account), and the interaction of these two added as a third random effect[3]. This type of model is very powerful at producing results at fine levels of detail even when the data is very thin and volatile.

5.3 Uneven Hierarchical Models

When building a model to perform credibility weighting, sometimes we can encounter a data structure where each group has a different number of levels. For example, suppose we are building a hierarchical model on groups and subgroups that looks like the following:



Groups A and B each have two children, while groups C and D only have one, and so really do not have

3 In R, an interaction is added by using a colon between variables. Using the lme4 package, a random effect has the syntax, *(1 | group)* where *group* is the variable we create a random effect on. To do as described, the syntax would be: *(1 | industry) + (1 | territory) + (1 | industry: territory)*

any sub-groupings. This can create problems when building a GLMM since if random effects are assigned to the subgroups C1 and D1, the groups C and D will each have two credibility coefficients that do the same thing, effectively giving double credibility to these groups. Really, we would want the coefficients for C1 and D1 to be given values of zero. This is in fact what happens when building a GLM on this type of data, but not with a GLMM.

This type of model can be built using a GLMM if we add the subgroup random effects as slope coefficients instead of regular intercept coefficients. To explain, most random effects modify the intercept and add or subtract an amount from the intercept, which is same as adding or subtracting this term from the entire equation. But it is also possible to have a random effect behave like a slope parameter instead[4]. Doing this, the coefficients of the random effects are multiplied by a data variable in the equation. Using this, we can create a new variable that is one if its subgroup has any siblings, meaning that it is not the only child of its parent, and zero if it has no siblings. If we create the subgroup random effect as a slope on this variable, it will not allow the nodes C1 and D1 to have non-zero values, and the model will behave as expected.

Similarly, if building the "compromise" model described in the previous section where we gave the example of constructing a model by industry and territory, this unevenness of levels may need to be accounted for as well. A regular model will give double credibility if there is a territory with only one industry, or an industry only under a certain territory. Instead of siblings, we refer to these relationships as cousins. To account for this, similar binary variables can be setup in the data that indicate whether any cousins exist, and the random effects can be added as slope parameters to these variables as described.

5.4 Implementing Mixed Models in Spreadsheets

GLMM credibility models that are either additive or multiplicative can also be implemented in spreadsheets fairly easily using maximum likelihood estimation. To do this, we first determine the formula of the loss ratios, such as *log(Fitted LR) = intercept + territory + industry*, which would create a multiplicative model with territory and industry relativities. To build a regular GLM without credibility weighting, the log-likelihood should be calculated as follows:

$$\sum N\,(\,Fitted\ LR_i\,,Actual\ LR_i\,,Within\,Variance\ Factor\,/\,Premium\ Base_i) \tag{5.2}$$

Where *N(A, B, C)* is the logarithm of the Normal PDF at *A* with a mean of *B* and a variance of *C*. The

4 The syntax shown in the previous footnote will create a random effect on the intercept. To create a random effect on the slope, the syntax is: *(0 + variable | group)* where *variable* is the variable we are creating the slope on. (The "*0 +*" is needed here to let it know not to create the random effect on the intercept as well. If we left this part out, random effects would be added both to the intercept and as a slope.)

intercept and territory and industry coefficients should all be determined using a routine that maximizes the total log-likelihood.

A GLMM can be implemented using the simple Bayesian model we described above since MLE parameters are assumed to be approximately normally distributed, and so the posterior distribution should be approximately normal as well. To calculate the log-likelihood for the GLMM, we add the following to formula (5.2):

$$\sum_{t=\,all\ territories} N\,(Coefficient_t\,, 0,\ Between\,Variance\ Territories) + \sum_{i=\,all\,industries} N\,(Coefficient_i\,, 0,\ Between\,Variance\ Industries) \tag{5.3}$$

Using zero as the mean for the prior distributions effectively weights everything back towards the intercept, which is what performs the credibility weighting. The between variances are difficult to calculate, limiting the advantage of this approach however. They can be estimated by looking at the variances of each parameter while controlling for all of the other parameters, possibly by using a GLM.

A plus side is that the Tug-of-War method can be implemented. We suggest using the relativities version shown above and implementing as follows, assuming a multiplicative model: The log-likelihood for the relativity coefficients should be calculated first using formulas similar to (5.3). The exponent of the log-relativities should be taken to calculate the actual relativities for each combination of dimensions and the weighted average overall relativity should be calculated. Revised relativities should then be computed by dividing each relativity by the average and the final loss ratios can be calculated by multiplying these relativities by the average loss ratio. The log-likelihood for each loss ratio can then be taken and added to the overall total. This method will ensure that the overall average of the credibility weighted results ties to the original. It is also possible to ensure that the average of each loss ratio across a particular dimension, industry for example, ties the original average industry loss ratios as well. This can be done by calculating the average relativities across each industry and dividing each relativity by the average relativity for each industry. Ensuring that the averages of more than one dimension tie to the originals puts too many constraints on the solution and is not recommended[5].

5 We ignored the log-likelihood weights in this discussion. One option is to leave out the weights even though this may cause the Tug-of-War method to not work as well. Another option is to apply weights to each relativity log-likelihood equal to the total premium for each item across all of the other dimensions, and weights to each loss ratio log-likelihood equal to the premium of each. This will help the Tug-of-War part of the method work better but slightly violates Bayes' formula which is the formula we are using for the credibility weighting.

6. ACCOUNTING FOR MIX CHANGES AND NON-RENEWALS

The last topic we will discuss is non-renewals and business mix changes. Very often, to improve a book of business, some accounts or segments perceived to be under-performing will be non-renewed, and actuaries are often asked to quantify the impact of these actions. One method (which is often favored by the underwriters) is to completely eliminate all non-renewed business from the experience and calculate predictions on this cleaned up data. But doing so does not account for the credibility of the non-renewed business. To give an extreme example, assume all policies have a loss on average of once every five years and are completely identical in terms of expected losses. If after a couple of years, all accounts with a loss are non-renewed, the historical loss ratio on the remaining business will clearly look much better, but the book really has not changed at all. The expected going forward loss ratio is exactly the same. The same example can be applied to business mix changes as well.

Instead, when calculating the benefit, we suggest incorporating credibility in most cases. (In some cases, however, a major change has truly been made and a unique segment has been non-renewed for which the overall loss ratio of the book does not serve as a good credibility complement; in these situations, it may not make sense to incorporate credibility.) If a particular segment is non-renewed, credibility weighted loss ratios can be produced by segment using one of the methods described above, and the difference to the total loss ratio can be calculated both with and without this particular segment to determine the effect. If accounts with the highest frequency or loss ratios are non-renewed, credibility weighted loss ratios can be calculated by frequency or loss ratio band and the effect can be determined. If just a bunch of poor accounts are non-renewed, a hierarchical model that properly reflects the segmentations in the book of business can be built that goes all the way down to the policy level, and the result of excluding these policies can be determined as well. Although this last case may be the most difficult to model. The same applies to mix changes. Credibility weighted loss ratios can be produced per segment and the total weighted average loss ratio can be calculated before and after the change to help judge the effect on the overall book.

7. CONCLUSION

As pricing actuaries, we are relied upon to help make many important strategic and quantitative decisions. Without a good credibility mechanism, a choice often needs to be made between not giving enough detail and giving enough detail but not accurately. Applying credibility allows us to balance these two demands and provide enough detail and do so accurately.

REFERENCES

Bolstad, W., "Introduction to Bayesian Statistics (Second Edition)," 2007, p.106-108, 207-209

Dean, C., "Topics in Credibility Theory," Education and Examination Committee of the Society of Actuaries, 2005, https://www.soa.org/files/pdf/c-24-05.pdf

Klinker, F., "Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting," Casualty Actuarial Society Forum, 2011, p. 1-25, http://www.casact.org/pubs/forum/11wforumpt2/klinker.pdf

Biography of the Author

Uri Korn is an AVP & Actuary at Axis Insurance serving as the Research and Development support for all commercial lines of insurance. Prior to that, he was a Supervising Actuary at AIG in the Casualty pricing department. His work and research experience includes practical applications of credibility, trend estimation, increased limit factors, non-aggregated loss development methods, and Bayesian models. Uri Korn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries.