CREDIBILITY: PRACTICAL APPLICATIONS

Howard Mahler

There are many fine papers on the theory behind credibility. However, today we are concentrating on the uses of credibility theory, but only if they are important for credibility practice. My talk will be from the point of view of a Bureau actuary or an actuary with a primary insurer.

Let X be the quantity we wish to estimate. For example, X might be the expected losses for a Workers' Compensation class relative to the statewide, i.e., X is the class relativity.

Let Y_1, Y_2, Y_3, etc. be various estimates for X. For example, Y_1 might be the current relativity, Y_2 might be the observed relativity for the latest data, Y_3 might be the relativity based on national data suitably adjusted, then we might estimate X by taking a weighted average of the different estimates Y_i.

$$X = \sum_{i=1}^{n} Z_i Y_i$$

where X = quantity to be estimated
Y_i = ith estimate of X
Z_i = weight assigned to ith estimate of X

Usually the weights Z_i are restricted to the closed interval between 0 and 1. In the most common situation we have two estimates, i = 2. In that case we usually write:

$$X = Z Y_1 + (1-Z) Y_2$$

where Z is called the credibility and 1-Z is called the complement of credibility. However, it is important to note that the usual terminology tempts us into making the mistake of thinking of the two weights and two estimates differently. The actual mathematical situation is symmetric.

We can now discuss those rules I think will aid you in using credibility for practical applications.

Rule 1A: Spend a lot of time and effort deciding on or choosing the Y_i. Each Y_i should be a reasonable estimate of X.

So for example, if trying to estimate a medical claim cost trend it may not make much sense to assign the complement of credibility to an estimate based on the general rate of inflation.

Rule 1B: Spend a lot of time and effort computing, collecting data on, or estimating each Y_i.

If you are going to include a value in your weighted average, it makes sense to try to carefully quantify that value.

Rule 2A: The procedure is generally forgiving of small "errors" in the weights. Therefore, do not worry overly much about getting the weights exactly right.

This is discussed in my paper "An Actuarial Note on Credibility Parameters" in PCAS 1986.

Rule 2B: The concept of credibility is a relative concept. A relative weight can only be assigned to any single estimator, if you know what all the other estimators are.

For example, assume you have two estimators each of which has been assigned "only" 50% credibility. This merely indicates that the two estimators are equally good or equally bad, not whether they are good or bad in some absolute sense.

Rule 2C: The less random variation in an estimate the more weight it should be given. In other words, the more useful information and the less noise, the more the weight.

Rule 2D: The more closely related to the desired quantity, the more weight an estimator should receive.

For example, observations more distant in time usually deserve less weight.

Rule 3: Cap the changes in relativities that result from the use of credibility.

A properly chosen cap may not only add stability, but may even make the methodology more accurate by eliminating extremes.

Let's go back to the question of choosing the estimators to use for a given task. The estimators Y_i can have many sources. For example:

1. The recent observation(s) of X.

2. The recent observation(s) of the same quantity as X, but for a superset.

3. The recent observation(s) of a similar quantity to X; there may be an adjustment necessary.

4. Past estimates(s) of X. There may be an adjustment for the intervening period of time.

5. The result of a model.

6. The result of judgement.

In the following real world example of the use of credibility to determine Workers' Compensation classification relativities, I used two estimators: the relativities in the current rates and the relativities indicated by the recent experience of each class. How much weight should be given to each estimator for each class?

In the actual methodology used, three pieces of the pure premium were each estimated separately and then added together. The three pieces were: serious, non-serious, and medical as per Roy Kallop's paper "A Current Look at Workers' Compensation Ratemaking" in PCAS 1975.

For each class, the relativity indicated by the most recent experience was given credibility $0 \leq Z \leq 1$

$$Z = \left(\frac{E}{F}\right)^{2/3} \qquad F > 0$$

where E was the expected losses for that class and F is the so-called standard for full credibility. (F would vary by serious, non-serious, and medical.) This is formula 1 on Exhibit 4.

The problem was to estimate an appropriate value of F to use for this purpose. For various values of F, the resulting estimates of the class relativities were compared to the observed future relativities.

In order to do this, I needed an extra year of data not used to make any of the estimates. To quantify how close a match was obtained between the predicted relativities and the relativities observed in this additional year, I calculated the mean squared error (using payrolls to weight the squared errors by class.)

The result for serious losses appears in a graph in Exhibit 2. F is in thousands of dollars of expected losses, and appears on a logarithmic scale. The minimum MSE occurs when the standard for full credibility is about $15 million.

There are a few points worth making:

1. Values of F between $10 million and $20 million all perform quite well.

2. For values of F within a factor of 10 of optimal, the graph of the mean square error is approximately symmetric when F is put on a log scale. For values of F differing from $15 million by the same factor less than or equal to 10, the MSE is roughly the same.

3. The reduction in mean squared error in this case due to the use of credibility is not that large, but neither is it atypical. (Not only is difficult to estimate the expected relative losses, but the actual observation varies randomly around the expected result.)

In this case, using the current relativities ($F=\infty$) gives a MSE of 2.385. Using the observed relativities (F=0) gives a MSE of 2.503. Thus in this case by this relatively simple use of credibility, we have reduced the mean square error by about 8% from .2385 to 2.200.

The basic reason the credibility procedure provides some improvement is that the "observed" relativities generally worked well for larger classes, while the current relativities were generally a better estimate for the smaller classes. The credibility method puts more weight on the generally better estimate.

The results of this study were used to modify the standards for full credibility.

	Standards for Full Credibility	
	Previous	New
Serious	25 average serious losses	175 average serious losses
Non-Serious	300 average non-serious losses	120 average non-serious losses
Medical	240 average non-serious losses	190 average non-serious losses

The credibilities using the previous and new full credibility standards are shown in Exhibit 3. As we expect, the change in full credibility standards substantially lowered the credibilities assigned to serious losses, raised the credibilities assigned to non-serious losses, and raised slightly the credibilities assigned to medical losses.

The MSE using the two sets of standards for full credibility are as follows:

	Mean Squared Error	
	Previous	New
Serious	2.267	2.200
Non-Serious	.337	.334
Medical	.454	.453

Thus we see that even for serious losses, where there was the most substantial revision to the standard for full credibility, the improvement in mean squared error is not overwhelming. While it is worthwhile to review the manner in which the credibilities are determined, provided the method automatically adjusts for inflation, such reviews need only be made infrequently, ex., every decade or two. Usually such reviews will show that the current method works reasonably well, but can be improved somewhat. It is generally more useful to spend your efforts on improving the estimates that are to be weighted together.

In this study, besides examining the parameter used in the formula for credibility, I also examined the use of other formulas. Some of these formulas shown in Exhibit 4 should be very familiar, while some may be new to some of you.

For practical uses of credibility, it's sufficient to know of the existence of these different formulas for the dependence of credibility on the size of the class or the risk. If one of them works significantly better, use it, without agonizing over the theoretical mumbo jumbo that lies behind the formula.

In the particular example here, using another formula produces no significant improvement. However in other situations, such as Workers' Compensation Experience Rating, it turns out that there is a significant improvement obtained by using formula 6.

Exhibit 5 shows an example of the different behavior you get with size of risk for formulas 3 through 6. (These four formulas are all based on the ideas of "Bayesian" as opposed to classical credibility.) Formula 3 goes to zero at zero and 1 at infinity. Formula 4 has some minimum credibility greater than zero for small risks. Formula 5 has some maximum credibility less than one for large risks. Formula 6 combines the behavior of Formulas 4 and 5.

A practical actuary might just select a curve and set of parameters that produce credibilities that seem reasonable.

The next example of a practical use of credibility involves revising the definitions of automobile insurance territories in Massachusetts. Each town's relative loss potential is determined based on 4 years of data and a relatively complicated credibility methodology. Then towns with similar loss potential are grouped together. Here we will ignore the details of the procedure which are explained in Robert Conger's paper in PCAS 1987, and focus on the results of the latest review conducted for 1989 rates.

The predictions of the methodology as used in the review of 1986 rates were compared with the subsequent observations. Using the methodology reduced the mean squared error to .0091 from .0117 if the observations had been relied on solely. Thus the credibility methodology performed its task of reducing the mean squared error, in this case by 22%.

However, such summary statistics do not tell the whole story. Credibility is a linear process, and thus the extreme cases may not be dealt with as well as they might.

For example, let's look at the results of applying the same methodology consistently over time to two small towns.

	Estimated Loss Potential Relative to Statewide Average			
	1984 Review	1986 Review	1988 Review	1989 Review
Acushnet	.84	.87	.88	.87
Brewster	.74	.84	.70	.61

	Indicated Territory (Prior to Capping)			
	1984 Review	1986 Review	1988 Review	1989 Review
Acushnet	5	6	6	6
Brewster	3	6	2	1

The results for the first town Acushnet, with 6000 exposures per year are typical. The relative loss potential varies somewhat from review to review, with a change in indicated territory of plus or minus one from time to time.

The results for the second town Brewster, with 5000 exposures per year, are not typical. In fact, Brewster was chosen as the most extreme case of fluctuating experience over this period of time. As you can see the estimated relative loss potential swung up and then down. This in turn resulted in large changes in the indicated territories. This occurred in spite of relying on four years of data, so that the data periods used in the reviews overlap. This occurred in spite of the use of credibility, which ameliorated the effect of the large fluctuations in the experience of this town.

Such large swings are unlikely. However, when dealing with 350 towns, something that only has only a .3% chance of happening per town, on average occurs for one town.

This problem is dealt with by capping movements. The actual cap chosen was to restrict movements to at most one territory either up or down. This is an example of the third rule I discussed earlier.

I have tried to describe a number of practical applications of credibility. I've given a number of general rules which you should find useful in your own work with credibility.

The theory behind the use of credibility can be complex. However, the use of credibility itself is set up precisely so that it can be understood by a layman. While ratemakers may differ in their knowledge of credibility theory, all ratemakers should be completely familiar with credibility practice.

Exhibit 1

<u>RULE 1A</u>: SPEND A LOT OF TIME AND EFFORT DECIDING ON OR CHOOSING THE Y_i. EACH Y_i SHOULD BE A REASONABLE ESTIMATE OF X.

<u>RULE 1B</u>: SPEND A LOT OF TIME AND EFFORT COMPUTING, COLLECTING DATA ON, OR ESTIMATING EACH Y_i.

<u>RULE 2A</u>: THE PROCEDURE IS GENERALLY FORGIVING OF SMALL "ERRORS" IN THE WEIGHTS. THEREFORE, DO NOT WORRY OVERLY MUCH ABOUT GETTING THE WEIGHTS EXACTLY RIGHT.

<u>RULE 2B</u>: THE CONCEPT OF CREDIBILITY IS A <u>RELATIVE</u> CONCEPT. A RELATIVE WEIGHT CAN ONLY BE ASSIGNED TO ANY SINGLE ESTIMATOR, IF YOU KNOW WHAT ALL THE OTHER ESTIMATORS ARE.

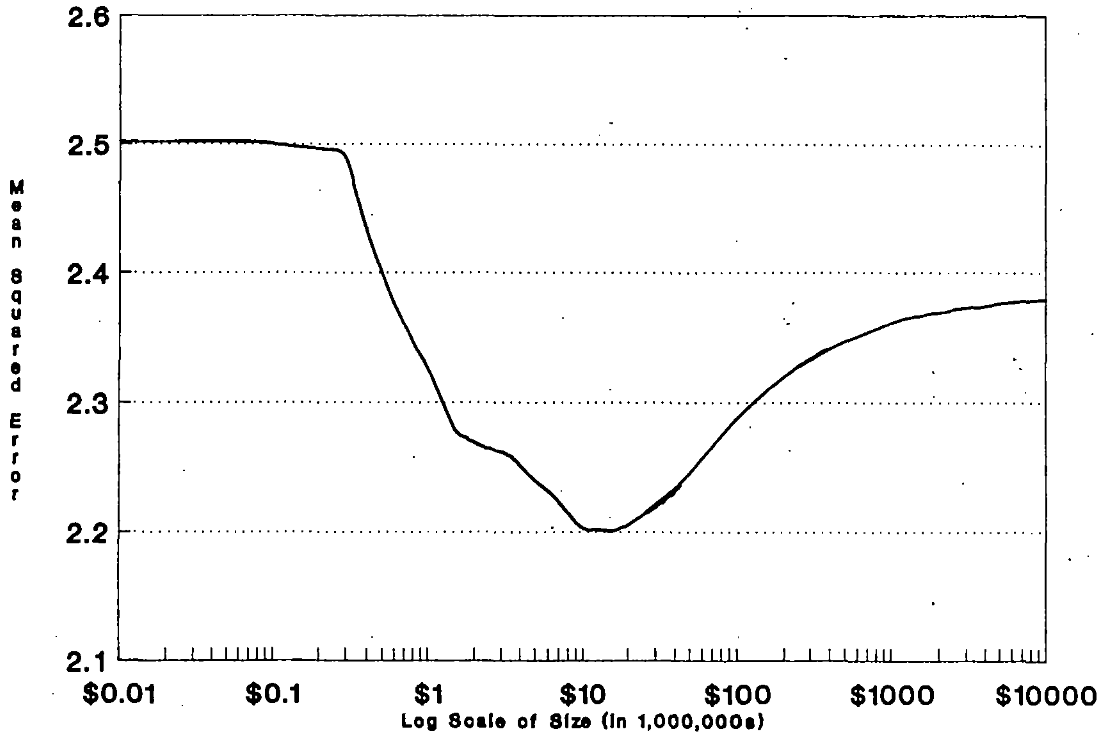
<u>RULE 2C</u>: THE LESS RANDOM VARIATION IN AN ESTIMATE THE MORE WEIGHT IT SHOULD BE GIVEN. IN OTHER WORDS, THE MORE USEFUL INFORMATION AND THE LESS NOISE, THE MORE THE WEIGHT.

<u>RULE 2D</u>: THE MORE CLOSELY RELATED TO THE DESIRED QUANTITY, THE MORE WEIGHT AN ESTIMATOR SHOULD RECEIVE.

<u>RULE 3</u>: CAP THE CHANGES IN RELATIVITIES THAT RESULT FROM THE USE OF CREDIBILITY.

Exhibit 2

Workers' Compensation Serious Pure Premiums



Standard for Full Credibility (F)

Exhibit 3

Credibility for Various Expected Losses

SERIOUS

Expected Losses (000's)	Current F 2,175,000	Indicated F 15,200,000
0	0%	0%
40	7	2
80	11	3
160	18	5
320	28	8
640	44	12
1,280	70	19
2,560	100	30
5,120	100	48
10,240	100	77
20,480	100	100

NON-SERIOUS

Expected Losses (000's)	Current F 1,260,000	Indicated F 496,000
0	0%	0%
10	4	7
20	6	12
40	10	19
80	16	30
160	25	47
320	40	75
640	64	100
1,280	100	100

MEDICAL

Expected Losses (000's)	Current F 1,008,000	Indicated F 800,000
0	0%	0%
15	6	7
30	10	11
60	15	18
120	24	28
240	38	45
480	61	71
960	97	100
1,920	100	100

$$Z = (E/F)^{2/3}$$

Exhibit 4

Various Credibility Formulas

(1)*	$Z = (\frac{E}{F})^{2/3}$	$F > 0$	Traditional Workers' Compensation
(2)**	$Z = (\frac{E}{F})^{\frac{1}{2}}$	$F > 0$	Classical/Limited Fluctuation
(3)***	$Z = \frac{E}{E+K}$	$K \geq 0$	Bayesian/Buhlmann
(4)****	$Z = \frac{E+I}{E+K+I}$	$I \geq 0$ $K \geq 0$	Risk Inhomogeneity
(5)****	$Z = \frac{E}{EJ+K}$	$K \geq 0$ $J \geq 1$	Parameter Uncertainty
(6)****	$Z = \frac{E+I}{EJ+K+I}$	$I \geq 0$ $J \geq 1$ $K \geq 0$	Risk Inhomogeneity and Parameter Uncertainty

* This formula is used for example in R. Kallop, "A Current Look at Workers' Compensation Ratemaking," PCAS LXII, 1975, p. 62.

** This formula is explained for example in L.H. Longley-Cook, "An Introduction to Credibility Theory," PCAS XLIX, 1962, p. 194.

*** This formula is explained for example in A.L. Mayerson, "A Bayesian View of Credibility," PCAS LI, 1964, p. 85.

**** These formulas are explained for example in H.C. Mahler, Discussion of G.G. Meyers, "An Analysis of Experience Rating," PCAS LXXIV 1987, p. 119.

Exhibit 5

Comparison of Credibility Formulas