Volume LXXIX, Part 2 No. 151

PROCEEDINGS
November 15, 16, 17, 18, 1992

CREDIBILITY BASED ON ACCURACY

JOSEPH A. BOOR

Abstract

This paper shows that the optimal credibility split between two estimators is related to how well each estimator predicts the underlying experience. First, an equation is shown which expresses the credibility assigned to each estimator in terms of the average prediction error of the other estimator and the average squared difference between the two estimators. That equation is verified using the classic Bayesian credibility $n/(n+k)$ formula and a formula for weighting prior observations of time series that was developed by the author. An enhancement to the classic Bayesian method for class credibilities is shown. Finally, the author shows that optimal credibility is proportional to the accuracy of each estimator, less the extent to which both estimators make the same errors.

1. INTRODUCTION

Much of historical credibility emanates from one of three philosophies:

1. The square root rule and its cousins.
2. The Bayesian $n/(n+k)$ formula.
3. Alternative Bayesian philosophies that assume that losses are distributed according to some member of a family of distributions, and assign some judgmental probabilities to the distributions within the family.

Each of these approaches has its own set of problems.

Approach 1, the square root rule, apparently is little more than an ad hoc formula to graduate credibility from 0 to 1 in a fashion that:

a) tends to assign relatively high credibilities to small samples; and

b) achieves full credibility at some point.

Approach 2, the $n/(n+k)$ formula, is based upon a presumption that the sample mean is from a distribution chosen from a set of distributions. The complement of the credibility is to be assigned to the grand mean of all possible distributions. However, many credibility situations are not characterized by the process of first choosing a distribution randomly and then sampling from that distribution. For example, consider the case where a rate change indicated by a state's data is credibility-weighted against straight trend. While some might argue that the choice between the state's data and straight trend is just such a "distribution of distributions," clearly straight trend applied to the last rate indication is not the grand mean of that family of distributions. Further, when those assumptions do apply, such as in the class ratemaking problem, the grand mean must also be estimated.

Approach 3, the alternative Bayesian approach, relies on a presumption that the distribution of potential losses is a member of some family of distributions. The major problem with this approach is that the typical real-world loss distribution is not a precise mathematical curve. Further, this approach usually imposes some second probability distribution upon the choice of a distribution. An ideal method should be distribution-free.

To avoid these difficulties, it is worthwhile to list some of the attributes of a good approach to credibility.

1. Since the purpose of credibility is to hone an estimate of losses, it should do so in the best fashion possible. Specifically, it should produce optimum estimators of the unknown mean expected loss.
2. It should work in a wide variety of situations; e.g., when the complement of the credibility is assigned to trend, econometric projections, alternative methods of estimating the sample mean, or a sample of a larger but related distribution.
3. It should be distribution-free. It should not work solely when losses approximate some specific mathematical probability distribution.
4. Intuitively, it seems that the credibility should, in some sense, be related to how effectively the observed sample losses approximate the underlying propensity for loss. Further, whatever statistic receives the complement of credibility should receive greater weight as its effectiveness in estimating the underlying propensity for loss increases.

One method that meets all these criteria involves minimizing the expected squared error in estimating the propensity for loss. Specifically, if we seek to credibility-weight two statistics x_1 and x_2 to approximate Y, and if:

$\tau_1^2 = \mathrm{E}[\,(x_1 - Y)^2\,]$; i.e., τ_1^2 is the expected error of x_1 as an estimator;

$\tau_2^2 = \mathrm{E}[\,(x_2 - Y)^2\,]$; i.e., τ_2^2 is the expected error of x_2 as an estimator;

$\delta_{12}^2 = \mathrm{E}[\,(x_1 - x_2)^2\,]$; i.e., δ_{12}^2 is their expected squared difference;

then

$$Z(x_1, x_2) = \frac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}$$

produces the estimate

$$Zx_1 + (1 - Z)\,x_2\,,$$

which minimizes the error in estimating Y; i.e.,

$$\mathrm{E}[\,(Zx_1 + (1 - Z)x_2 - Y)^2\,] = \min.$$

The proof is provided in Appendix A.

A quick review of the above criteria will show that this approach does indeed fulfill all four: it is optimal by design, works with a wide variety of estimators, is distribution-free, and assigns credibility in accordance with predictive accuracy.

2. USING OBSERVED ACCURACY

Before going further, it might be worthwhile to note that this method produces the same value of Z when we attempt to predict an observed value Y' rather than the actual propensity for losses Y.

In particular, if Y is the propensity for loss, then let $Y' = Y + \Delta$ be the observed losses. In this case, Δ would be independent of Y, x_1, and x_2, and have a mean of 0 and a variance of S^2.

Then, since the x and Δ are independent,

$$\tau_1'^2 = \tau_1^2 + S^2,$$

$$\tau_2'^2 = \tau_2^2 + S^2,$$

and

$$Z(x_1, x_2, Y') = \frac{\tau_2'^2 - \tau_1'^2 + \delta_{12}^2}{2\delta_{12}^2}$$

$$= \frac{\tau_2^2 + S^2 - \tau_1^2 - S^2 + \delta_{12}^2}{2\delta_{12}^2}$$

$$= \frac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}$$

$$= Z(x_1, x_2, Y) .$$

Hence, computing Z for actual observed losses produces the same Z as that appropriate for predicting the underlying propensity for loss. Further, since $\tau_1'^2$, $\tau_2'^2$, and δ_{12}^2 may then be estimated from historical observed losses, Z may be estimated from actual observed losses.

The Pitfalls of Skew

One aspect of using historic predictive accuracy must be noted—it is impractical with the highly-skewed distributions typically associated with individual risks and small pools of losses.

The classic example is the credibility of a medium-size commercial insured's own experience relative to industry experience. Even when the insured has average exposure (i.e., the unmodified manual rate is right for the insured), the insured will typically experience loss ratios in the 40% to 50% range year after year. However, every five to 10 years it will experience a loss ratio in excess of 100%. That is because the loss size distribution, and hence the insured's aggregate loss distribution, is highly skewed. Simply stated, the insured is exposed to very large, but relatively infrequent, losses. Because those losses are so large, they represent a disproportionately large part of the insured's exposure to loss. Because they are so infrequent, they do not show up in the insured's loss experience every year.

The naive observer might conclude after viewing several years of low loss ratios that the insured's own low loss experience is a much better predictor of that insured's loss experience than the manual rate. The previous section of this paper would seem to support that statement.

Actual observed accuracy is misleading, however, because of the partial observation of prediction error. Over the last four years or so, only a portion of the distribution of prediction errors (specifically the low side) has been observed. When a large loss occurs, the full range of prediction error can be glimpsed. Depending on how long it takes for that large loss to occur, the manual rate may then appear to be either too high or too low.

Many of the standard actuarial treatments for skewed distributions correct this problem. For instance, one could compare an insured's historical loss experience with the following year's basic limits losses and an industry provision for excess losses. One could also replace the industry excess loss provision with, say, 30 years of the insured's own trended excess losses. Alternatively, one could compare the estimating accuracy of a large body of similar insureds. Although each member insured's observed losses may not be fully representative of the full error distribution, a large enough body of insureds should approximate all probable prediction errors.

Before proceeding further with an analysis of how this method may be applied in practice, it is worthwhile to investigate whether it reproduces some of the credibility formulae that are already known.

Example 1. Classic Bayesian Credibility

Let x be the result of a two-stage process. First, a mean $\theta \in \omega$ is selected randomly from a set of means ω with (grand) mean M. Then x_1, x_2, ..., x_n are selected from a distribution with mean θ and variance S^2. Their mean $\bar{x}$ is, of course, dependent on θ, but each $x_i - \theta$ is independent, not only of the other $x_j - \theta$, but also of θ. Further, let the $\theta \in \omega$ have variance σ^2. Then, classic credibility [1] says

$$Z(\bar{x}, M, \theta) = n/[n + (S^2/\sigma^2)].$$

To prove this, note that independence of the x_i and θ implies:

$$\tau_{\bar{x}}^2 = (S^2/n),$$

$$\tau_M^2 = \sigma^2, \text{ and}$$

$$\delta_{\bar{x},M}^2 = \sigma^2 + S^2/n.$$

Then, the formula states:

$$Z(\bar{x}, M, \theta) = \frac{\sigma^2 - (S^2/n) + \sigma^2 + (S^2/n)}{2(\sigma^2 + (S^2/n))}$$

$$= \frac{2\sigma^2}{2[\sigma^2 + (S^2/n)]}$$

$$= \frac{n\sigma^2}{n\sigma^2 + S^2} = \frac{n}{n + (S^2/\sigma^2)}.$$

In other words, the classic Bayesian credibility is reproduced.

Example 2: Weighting of Trended Time Series

Let $x(t)$ begin at some $x(0)$ and then be changed infinitesimally often by infinitesimally small, but variable, perturbations. In other words, the change in $x(t)$ from time to time is a result of a very large number of very slight random occurrences, just like those affecting most econometric time series. The random nature of the perturbations assures us that $x(t)$ will be distributed normally about $x(0) + t\mu$ (where μ is the average rate of change), and the variance of $x(t)$ is $t\sigma^2$ (where σ^2 is the unit variance).[1] In fact,

$$E[x(a) - x(b)] = (a - b)\mu$$

and

$$\mathrm{Var}[x(a) + (b - a)\mu - x(b)] = |a - b|\sigma^2.$$

To simplify matters, let us consider the case where $\mu = 0$. Further, as in most practical problems, when x(1), x(2), ..., $x(n)$ are estimated by $\hat{x}(1)$, $\hat{x}(2)$, ..., $\hat{x}(n)$, there should be some estimation errors Δ_i. So, $\hat{x}(i) = x(i) + \Delta_i$, and the Δ_i are independent and identically distributed with mean 0 and variance S^2. Then, when weighting $\hat{x}(1)$, $\hat{x}(2)$, $\hat{x}(3)$, etc. to estimate $x(n)$, the weights

[1] The limited reading of this author on Poisson processes suggests that some of the conclusions on the mean and variance may be common knowledge of statisticians.

$$W_i = S^{2(n-i)}F_i ,$$

where the F_i are defined recursively by

$$F_1 = 1;$$

$$F_2 = S^2 + \sigma^2 \text{ ; and}$$

$$F_{i+1} = (2S^2 + \sigma^2)\, F_i - S^4 F_i - 1$$

produce the optimal estimate. To test this, only the two-point estimate will be verified. Unfortunately, similar credibility estimators for three or more estimators are very unwieldy. The two-point estimate involves an estimate for $\hat{x}(3)$ given $\hat{x}(2)$ and $\hat{x}(1)$.

According to our formula, the result should be

$$\hat{x}(3) \approx \frac{S^2\hat{x}(1) + (\sigma^2 + S^2)\,\hat{x}(2)}{\sigma^2 + 2S^2} ;$$

i.e., it should be true that

$$Z\,[\hat{x}(2), \hat{x}(1), \hat{x}(3)\,] = \frac{S^2 + \sigma^2}{\sigma^2 + 2S^2} .$$

Note that, since $\hat{x}_i - x_i$ is orthogonal to $\hat{x}_j - x_j$ and $x_i - x_j$:

$$\begin{aligned}\tau_1^2 &= \mathrm{E}[\ (\hat{x}(2) - \hat{x}(3)\ )^2\ ] \\ &= \mathrm{E}[\ (\hat{x}(2) - x(2)\ )^2\ ] + \mathrm{E}[\ (x(2) - x(3)\ )^2\ ] + \mathrm{E}[\ (x(3) - \hat{x}(3)\ )^2\ ] \\ &= S^2 + \sigma^2 + S^2 = 2S^2 + \sigma^2 \ .\end{aligned}$$

$$\begin{aligned}\tau_2^2 &= \mathrm{E}[\ (\hat{x}(1) - \hat{x}(3)\ )^2\ ] \\ &= \mathrm{E}[\ (\hat{x}(1) - x(1)\ )^2\ ] + \mathrm{E}[\ (x(1) - x(3)\ )^2\ ] + \mathrm{E}[\ (x(3) - \hat{x}(3)\ )^2\ ] \\ &= S^2 + 2\sigma^2 + S^2 = 2S^2 + 2\sigma^2 \ .\end{aligned}$$

$$\begin{aligned}\delta_{12}^2 &= \mathrm{E}[\ (\hat{x}(1) - \hat{x}(2)\ )^2\ ] \\ &= \mathrm{E}[\ (\hat{x}(1) - x(1)\ )^2\ ] + \mathrm{E}[\ (x(1) - x(2)\ )^2\ ] + \mathrm{E}[\ (x(2) - \hat{x}(2)\ )^2\ ] \\ &= S^2 + \sigma^2 + S^2 = 2S^2 + \sigma^2\ .\end{aligned}$$

Thus,

$$Z\,(\hat{x}(2), \hat{x}(1), \hat{x}(3)\ ) = \frac{2S^2 + 2\sigma^2 - (2S^2 + \sigma^2) + 2S^2 + \sigma^2}{2\ (2S^2 + \sigma^2)} = \frac{S^2 + \sigma^2}{2S^2 + \sigma^2}$$

So, in at least two cases, this approach does reproduce credibilities already known from other analyses.

To be truly useful, an approach should yield new methods. The true classification ratemaking problem will now be addressed.

3. CLASS CREDIBILITIES

The objective is to produce a rate for a subgroup α of a large group Γ. The means of α and Γ are unknown, but have been estimated using:

$$\bar{a} = (1/n)\sum_i^n a_i\ ;\ \ a_i \in \alpha\ \text{ for } \mu_\alpha \text{ and}$$

$$\bar{g} = (1/(m+n)\ )\left(\sum_i^n a_i + \sum_j^m b_j\right);\ \ b_j \in \beta = \Gamma - \alpha\ .$$

Further, Var $(a_i) = \sigma_\alpha^2$ and Var $(b_j) = \sigma_\beta^2$ have been estimated from actual data.

Before proceeding, note that weighting $\bar{a}$ with $\bar{g}$ using $Z\,(\bar{a}, \bar{g}, \mu_\alpha)$ effectively assigns a portion of the complement of the credibility back to $\bar{a}$ since $\bar{g} = (n\bar{a} + m\bar{b})/(n+m)$. So, it may be more worthwhile to evaluate

$$Z\,(\bar{a}, \bar{b}, \mu_\alpha) = \frac{mZ\,(\bar{a}, \bar{g}, \mu_\alpha) + n}{n+m}\ ;$$

and then use

$$Z(\bar{a}, \bar{g}, \mu_\alpha) = \frac{Z(\bar{a}, \bar{b}, \mu_\alpha)(n+m) - n}{m}.$$

Note that

$$\mathrm{Var}(\bar{a}) = (\sigma_\alpha^2)/n\ ;\ \ \mathrm{Var}(\bar{b}) = (\sigma_\beta^2)/m\ ;\ \text{ and}$$

$$\tau_1^2 = \mathrm{E}[\,(\bar{a} - \mu_\alpha)^2\,] = \sigma_\alpha^2/n\ .$$

Since the b_j are independent, and the $b_j - \mu_\beta$ have mean 0,

$$\tau_2^2 = \mathrm{E}[\,(\bar{b} - \mu_\alpha)^2\,] = \sigma_\beta^2/m + (\mu_\beta - \mu_\alpha)^2\ ;$$

and

$$\begin{aligned}\delta_{12}^2 &= \mathrm{E}[\,(\bar{a} - \bar{b})^2] = \mathrm{E}[\,(\bar{a} - \mu_\alpha)^2] + \mathrm{E}[\,(\mu_\alpha - \mu_\beta)^2] + \mathrm{E}[\,(\bar{b} - \mu_\beta)^2] \\ &= (\sigma_\alpha^2/n) + (\mu_\alpha - \mu_\beta)^2 + (\sigma_\beta^2/m)\ .\end{aligned}$$

Hence

$$\begin{aligned}&Z(\bar{a}, \bar{b}, \mu_\alpha) \\ &= \frac{(\sigma_\beta^2/m) + (\mu_\alpha - \mu_\beta)^2 - (\sigma_\alpha^2/n) + (\sigma_\alpha^2/n) + (\mu_\alpha - \mu_\beta)^2 + (\sigma_\beta^2/m)}{2\,(\,(\sigma_\alpha^2/n) + (\mu_\alpha - \mu_\beta)^2 + (\sigma_\beta^2/m)\,)} \\ &= \frac{\sigma_\beta^2/m + (\mu_\alpha - \mu_\beta)^2}{(\sigma_\alpha^2/n) + (\sigma_\beta^2/m) + (\mu_\alpha - \mu_\beta)^2} \\ &= \frac{n}{n + [\sigma_\alpha^2/(\,(\sigma_\beta^2/m) + (\mu_\alpha - \mu_\beta)^2\,)\,]}\ .\end{aligned}$$

Or, if m is much larger than n

$$= \frac{n}{n + [\sigma_\alpha^2/(\mu_\alpha - \mu_\Gamma)^2]}\ .$$

Of course, σ_α^2 and $(\mu_\alpha - \mu_\Gamma)^2$ are unknown, but they may be estimated using the sample variance of $a_i \in \alpha$ and the difference between the existing rate for α and the overall average rate.

This formula illustrates several key points:

1. Highly heterogeneous classes (high σ_α^2) should receive lower credibilities.
2. Extremely high or low rates [high $(\mu_\alpha - \mu_\Gamma)^2$] should be based more heavily on their own experience than on the overall average rate.
3. For classes that form a statistically large proportion of the group, the complement of the credibility should be assigned to the mean of the remainder of the group, not the group mean.

4. A NUMERICAL EXAMPLE

Suppose that one is attempting to find the underlying mean, μ_x, of a distribution given last year's observation $X_{1,i-1}$ of the distribution, and last year's observation of $X_{2,i-1}$, a related statistic. Further, assume $X_{2,i}$ is thought to be cyclic and biased as a predictor of $X_{1,i}$, and its year-to-year variations are thought to be independent of those of X_1. Whether X_1 is cyclic or stationary is not known. The observations are shown in Table 1. Of course, the values $X_{1,i+1}$ and $X_{2,i+1}$ are unknown at time i, but the goal is to find the Z_i such that $Z_i X_{1,i} + (1 - Z_i) X_{2,i}$ is an optimum estimator of μ_x at time $i + 1$.

In this case, since X_1 and X_2 are independent predictors of μ_x, $E[(X_1 - X_2)^2]$ reduces to $\tau_{x,1}^2 + \tau_{x,2}^2$, so Z becomes $Z = \tau_{x,2}^2/(\tau_{x,1}^2 + \tau_{x,2}^2)$. X_1's error, $\tau_{x,1}^2$, may be estimated by $S_{x,1}^2$, the sample variance of the $X_{1,i}$ seen to date. Since $X_{2,i}$ is biased, $\hat{\tau}_{x,2}^2$ will be estimated by $S_{x,2}^2 + (X_{2,i+1} - \hat{Y}_i)^2$, where $\hat{Y}_i$ is the last estimate of μ_x. This method recognizes both the cyclic nature of X_2 (by using $X_{2,i+1} - \hat{Y}_i$) and a potential cyclic pattern in X_1 because it considers just the last observed values, not the history of X_1 and X_2. Arbitrarily, the first credibility was chosen at 50%. Results are shown in Table 2.

TABLE 1

Year	$X_{1,i}$	$X_{2,i}$
1	72.44	104.15
2	79.06	114.66
3	72.98	112.75
4	79.74	99.01
5	66.69	103.04
6	86.38	102.80
7	68.97	106.23
8	78.61	97.79
9	88.97	101.63
10	74.97	102.83

TABLE 2

OBSERVED HISTORY

Year	X_1	X_2	$\tau^2_{x,1} = S^2_{x,1}$	$S^2_{x,2}$	$(X_2 - \mu_x)^2$	$\tau^2_{x,2}$	Z	$\hat{Y}$
1	72.44	104.15	—	—	—	—	50%	88.30
2	79.06	114.66	43.82	110.46	694.85	805.31	95	80.84
3	72.98	112.75	13.51	31.34	1018.25	1049.59	99	73.38
4	79.74	99.01	15.04	54.02	656.90	710.92	98	80.07
5	66.69	103.04	28.82	44.75	527.70	572.45	95	68.51
6	86.38	102.80	47.86	38.36	1175.98	1214.34	96	87.04
7	68.97	106.23	47.38	31.97	368.38	400.35	89	73.07
8	78.61	97.79	42.08	36.02	611.08	647.10	94	79.76
9	88.97	101.63	56.66	32.82	478.30	511.12	90	90.24
10	74.97	102.83	50.81	29.51	158.61	188.12	79	80.82

In this case $\hat{Y}_i$ looks like a poor predictor of μ_x at $i + 1$.

However, when the true μ_x is considered, $\hat{Y}$ is a very good estimator of μ_x. The $X_{1,i}$ were generated using a normal distribution with mean $\mu_x = 80$ and variance $\sigma_x^2 = 100$. The $X_{2,i}$ were generated using a normal distribution with mean 100 and variance 100 for $X_{2,1}$, and successively generating each new $X_{2,i+1}$ using a normal distribution with mean $100 + .8\,(X_{2,i} - 100)$ and variance 36. One can show that the resulting $X_{2,i}$ all have a marginal distribution that has mean 100 and variance 100 since $36 = 100\,(1 - (.8)^2)$. Therefore, a priori, the credibility of X_i should be $(100 + 400)/(100 + 400 + 100) = .83333$. However, that credibility should vary with where X_2 happens to be in its cycle.

It just happened that the $X_{1,i}$ tended to fall on the low side of the distribution, and that the $X_{2,i}$ began on the high side of the distribution and tended to stay there.

In any event, the average squared prediction error of $\hat{Y}$ is 44.86, roughly a 20% reduction in the prediction error of X_1 alone (55.46 as a predictor of the value $\mu_x = 80$). In fact, the error of $\hat{Y}$ is even below 45.6, which results from what retrospectively turns out to be the best possible fixed credibility (96%). That is because this method gives greater weight to X_2 when it is close to $\hat{Y}$ in the cycle, and less weight when it is further away. So, in this example, the theory works.

5. CREDIBILITY DEMYSTIFIED

One of the side benefits of this approach is that it offers an explanation of credibility that can be understood by laymen. The credibility of each estimator is proportional to its accuracy as an estimator, less the extent to which the two estimators say the same thing. Clearly this explanation is much more desirable than "we've always done it this way," and more understandable to lay people than "we look at the process variance and the variance of the hypothetical means." But it has yet to be shown that the above explanation is true.

Note that

$$Z(x_1, x_2, Y) = \frac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}.$$

Further

$$\begin{aligned}\delta_{12}^2 &= \mathrm{E}[\ (x_1 - x_2)^2\ ] \\ &= \mathrm{E}[\ (\ (x_1 - Y) - (x_2 - Y)\ )^2\ ] \\ &= \mathrm{E}[\ (x_1 - Y)^2\ ] + \mathrm{E}[\ (x_2 - Y)\ )^2\ ] - 2\mathrm{E}[\ (x_1 - Y)\,(x_2 - Y)\ ]\ .\end{aligned}$$

Further, assuming that x_1 and x_2 are unbiased estimators, $\mu_1 = \mu_2 = Y$, so

$$\delta_{12}^2 = \tau_1^2 + \tau_2^2 - 2\mathrm{Cov}(x_1, x_2)\ .$$

And when that is included in the formula for Z,

$$\begin{aligned}Z(x_1, x_2, Y) &= \frac{\tau_2^2 - \tau_1^2 + \tau_1^2 + \tau_2^2 - 2\mathrm{Cov}(x_1, x_2)}{2\tau_1^2 + 2\tau_2^2 - 4\mathrm{Cov}(x_1, x_2)}\ ; \\ &= \frac{2\tau_2^2 - 2\mathrm{Cov}(x_1, x_2)}{2\tau_1^2 + 2\tau_2^2 - 4\mathrm{Cov}(x_1, x_2)}; \\ &= \frac{\tau_2^2 - \mathrm{Cov}(x_1, x_2)}{\tau_2^2 + \tau_1^2 - 2\mathrm{Cov}(x_1, x_2)}\ .\end{aligned}$$

Dividing top and bottom by $\tau_1^2\tau_2^2$ yields

$$Z(x_1, x_2, Y) = \frac{(1/\tau_1^2) - (R_{12}^2/\tau_1\tau_2)}{(1/\tau_1^2) + (1/\tau_2^2) - 2\,(R_{12}^2/\tau_1\tau_2)}\ ;$$

where R_{12}^2 is the correlation of x_1 and x_2. Clearly, if τ_1^2 is the error of x_1, then $1/\tau_1^2$ must be x_1's accuracy. Further, R_{12}^2 is the extent to which x_1 and x_2 vary together, and the division by $\tau_1^2\tau_2^2$ normalizes it relative to the inverse squared errors. Hence, the credibility of each estimator is proportional to its accuracy, less the extent to which the estimators say the same thing.

6. PRACTICAL APPLICATIONS AND FORMULAE

One criticism of this approach is that, like optimum credibility, the appropriate credibility formula depends on the circumstances. For instance, if the two estimators are not heavily skewed, their historic accuracy $n/\sum(x_{1,j} - Y')^2$ and $n/\sum(x_{2,j} - Y')^2$ may be used to derive the optimum credibility as shown in Section 2. Per Example 2, the formula $(1 + (\sigma^2/S^2))/(2 + (\sigma^2/S^2))$ may be used with σ^2/S^2 estimated using the historic year-to-year changes in $\hat{x}$. As shown in Section 3, the credibility of a class's own experience should be $n/(n + (S_\alpha^2/(r_\alpha - r_\Gamma)^2))$, where n is the number of exposure units, S_α^2 is estimated by comparing each year's class experience to a long-term average, and $r_\alpha - r_\Gamma$ is the difference between the current rate for class α and the current average rate. Alternatively, S_α^2 could be presumed to be constant across all classes and one could then find the S_α^2 that minimized the average squared error (weighted by exposures) when such a formula uses last year's experience to predict this year's data.

The most important results are:

1. The credibility of a piece of data and the formula used to derive it vary with the specific situation.
2. Using the formulae in this paper, one may derive credibility formulae that, up to determining a constant or two, represent the best credibility formulae. The constants can then be determined using historic data to find the constants that minimize the average squared error.
3. The fundamental truth of this paper, that credibility should be based on accuracy, makes intuitive sense and can be understood by laymen.

7. SUMMARY AND CONCLUSIONS

In summary, this approach seems to hold promise and appears to offer opportunities to improve the accuracy of loss estimates. However, it will only truly be useful when the estimation errors τ_1^2 and τ_2^2 are evaluated. Whether one believes in this approach or not, this author believes that the large ratemaking organizations should collect statistics on the effectiveness of the various loss estimators they use. Even if other credibility procedures are used, it only makes sense that their effectiveness be monitored. Futher, this author believes that greater understanding of how credibility should work can only improve the actuarial work product.

REFERENCES

[1] Venter, Gary G., et al, *Foundations of Casualty Actuarial Science,* New York, Casualty Actuarial Society, 1990, pp. 375-483.

[2] Efron, Bradley, "Controversies in the Foundations of Statistics," *American Mathematical Monthly,* Vol. 85, No. 4, April 1978, pp. 231-246; reprinted in the *Casualty Actuarial Society Forum,* Fall 1991, pp. 259-275.

APPENDIX A

PROOF OF $Z(x_2, x_1, Y) = \dfrac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}$

If $\tau_1^2 = \mathrm{E}[\,(x_1 - Y)^2\,]$, $\tau_2^2 = \mathrm{E}[\,(x_2 - Y)^2\,]$, and $\delta_{12}^2 = \mathrm{E}[\,(x_1 - x_2)^2\,]$,

the goal is to set

$$\Phi(Z) = \mathrm{E}[\,(Zx_1 + (1 - Z)\,x_2 - Y)^2\,] = \text{ min} .$$

But algebra gives

$$\begin{aligned}
&\mathrm{E}[\,(Zx_1 + (1 - Z)\,x_2 - Y)^2\,] \\
&\quad = \mathrm{E}[\,(Z\,(x_1 - Y) + (1 - Z)\,(x_2 - Y)\,)^2\,] \\
&\quad = Z^2\,\mathrm{E}[\,(x_1 - Y)^2\,] + (1 - Z)^2\,\mathrm{E}[\,(x_2 - Y)^2\,] \\
&\quad\quad + 2Z\,(1 - Z)\,\mathrm{E}[\,(x_1 - Y)\,(x_2 - Y)\,] \\
&\quad = Z^2\tau_1^2 + (1 - Z)^2\tau_2^2 + 2Z\,(1 - Z)\,\mathrm{E}[\,(x_1 - Y)\,(x_2 - Y)\,] .
\end{aligned}$$

Using $2(A - C)\,(B - C) = (A - C)^2 + (B - C)^2 - (A - B)^2$,

$$\begin{aligned}
&2\,\mathrm{E}[\,(x_1 - Y)\,(x_2 - Y)\,] \\
&\quad = \mathrm{E}[\,(x_1 - Y)^2\,] + \mathrm{E}[\,(x_2 - Y)^2\,] - \mathrm{E}[\,(x_1 - x_2)^2\,] \\
&\quad = \tau_1^2 + \tau_2^2 - \delta_{12}^2 .
\end{aligned}$$

Substituting that back in the overall squared error (Φ) equation,

$$\begin{aligned}
\Phi\,(Z) &= \mathrm{E}[\,(Zx_1 + (1 - Z)\,x_2 - Y)^2\,] \\
&= Z^2\tau_1^2 + (1 - Z)^2\tau_2^2 + Z(1 - Z)\,[\tau_1^2 + \tau_2^2 - \delta_{12}^2] \\
&= Z\tau_1^2 + (1 - Z)\tau_2^2 + (Z^2 - Z)\,\delta_{12}^2 .
\end{aligned}$$

This is minimized when

$$\frac{d\Phi}{dZ} = 0 = \tau_1^2 - \tau_2^2 + (2Z - 1)\delta_{12}^2$$

or

$$\tau_2^2 - \tau_1^2 + \delta_{12}^2 = 2Z\,\delta_{12}^2\,,$$

or

$$Z = \frac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}\,.$$

Now, this is only a minimum when

$$\frac{d^2\Phi}{dZ^2} = 2\delta_{12}^2 \geq 0\,.$$

However δ_{12}^2 is always non-negative.

Note that Z may be negative; i.e., when

$$\tau_1^2 \geq \tau_2^2 + \delta_{12}^2\,.$$

But, fortunately, that occurs only where $(x_2 - Y)$ and $(x_1 - x_2)$ tend to have the same sign overall; i.e., x_2 is generally between x_1 and Y. Thus, there may be cases where zero credibility is warranted; i.e., x_1 is not a useful predictor. Alternatively, where

$$\tau_2^2 \geq \tau_1^2 + \delta_{12}^2\,,$$

full credibility should be assigned to x_1.

APPENDIX B

JUST HOW DISTRIBUTION-FREE IS DISTRIBUTION-FREE?

As noted in the Section 1 of this paper, the credibility, Z, produced by this method produces the lowest expected squared error attainable using the $Zx_1 + (1 - Z)\,x_2$ (additive weighting) formula. That credibility does not depend on the particular characteristics of each estimator's distribution, but only on how well each estimator predicts the unknown quantity Y, and on the average squared difference between the two estimators. However, it does assume that the best estimator of the unknown Y is the one that minimizes the expected squared estimating error, and that it uses a $Zx_1 + (1 - Z)\,x_2$ formula. One should consider whether each of these implicit assumptions really produces the best estimator of the unknown Y.

Aside from the fact that the expected squared error function is ubiquitous in mathematics and related disciplines, there is a practical reason for using it as a penalty function whose minimum defines the best estimator. Conceptually, one might begin by viewing the expected absolute error $E[\,|\text{estimator} - Y|\,]$ as the best penalty function, since it measures the actual error of the estimator. That approach, however, has one considerable drawback. An extremely large error receives the same weight as a small error, even though extremely large errors may have catastrophic consequences. For example, if a rate would produce precisely the required profit 19 years out of 20 but threaten the company's solvency one year out of 20, prudence would dictate that the one year out of 20 receives a disproportionate share of attention. One logical approach is to weight each absolute error with the size of that absolute error—in effect to use $E[|\text{estimator} - Y| \cdot |\text{estimator} - Y|]$ or $E[(\text{estimator} - Y)^2]$.

The use of an additive weighting is less supportable. This author knows of no reason why an estimator of the form, say $x_1^Z \cdot x_2^{(1-Z)}$ would not be a better estimator. It is clear that if x_2 is biased, some $Zx_1 + (1 - Z)\,(x_2 - C)$ formula is better. Superficially, it appears that a rigorous analysis, perhaps using calculus of variations, could produce different formulae for different families of distributions.

On a more positive note, there are two reasons for using an additive weighting formula. When the two estimators x_1 and x_2 are known to be normally distributed about unknown means, but τ_1^2, τ_2^2, and S_{12}^2 can be estimated, the additive weighting formula is best.[2] Also, additive weighting has a long history in ratemaking.

[2] The author doubts he truly discovered this. Witness the argument in [2].