The Credibility of the Overall Rate Indication: Making the Theory Work

by Joseph Boor

ABSTRACT

Actuaries have used the so-called "square root rule" for the credibility for many years, even though the "F" value can take any value, and its assumption that the data receiving the complement of credibility is stable is often violated. Best estimate credibility requires fewer or no assumptions, but often requires certain key constants. This paper provides a variety of methods for estimating the key constants needed to implement best estimate credibility formulas, especially those arising from the Gerber-Jones formula. As such, this paper provides the tools needed to implement key theoretical formulas in practical actuarial work.

KEYWORDS

Credibility, best estimate credibility, overall rate indication, Brownian motion, process variance, parameter variance

1. Introduction

It is well established that the limited fluctuation or "square root" credibility has limitations. Since it is designed to produce stable estimates, not best estimates, it does not provide the most accurate rates. Further, since any conceivable combination of the fluctuation that may be acceptable and probability of a chance violation of the accepted fluctuation is a priori no different than any other, it is challenging[1] to show that any particular full credibility standard is better than any other. Lastly, the square root rule relies on an assumption that the statistic receiving the complement of credibility is stable. When the complement of credibility is, say, three years of 15% trend, that assumption is clearly violated. So there is a strong need[2] for best-estimate credibility.

Some time ago (1967) Hans Bühlmann developed a formula[3] for the best estimate credibility of a single risk or a single class when the complement of credibility is assigned to the large group that the risk or class is part of. His $P/(P+K)$ formula[4] is well known and represents a truly optimal (in the sense of making the best predictions) credibility formula. But a formula is also needed for the credibility of the overall rate change for a product or line of business. It is quite common in actuarial work to develop a rate indication for such a group, realize that supplemental data is needed, and credibility weight the overall indicated change with something such as the inflationary trend since the last rate change.[5] Considering that the overall rate change affects every rate for every class and every risk, this author believes that the credibility of the overall rate indication deserves as much attention as the credibility of the class data within it.

A solid theoretical background has been laid for the credibility of this overall rate indication. Credibility is by nature a process that is designed to update an estimate of loss costs. A paper by Jones and Gerber (1975) provides formulas for the weights in updating formulas (to be discussed later) in terms of the covariances of the historical data points.[6] This formula, in fact, provides the optimum linear estimate of future costs given all the prior data, not just the data used in the current rate update.

Nevertheless, knowing the mathematical form of the credibility is not the same thing as being able to compute the credibility. As will be shown, standard credibility formulas derived from the Gerber-Jones approach use values for the Brownian motion variance in year-to-year trend, plus values for the "observation error" variances between observed data points and the true expected costs that underlie them.[7] To compute the credibility, it is necessary to estimate those variance parameters. This paper provides techniques designed to do just that.

2. The theory—Key credibility formulas for the overall rate indication

In this section the key theoretical results from the Jones and Gerber (1975) paper are presented. This should provide the practitioner a summary of the key formulas that create best-estimate credibility. Likely none of the material is new.

2.1. The general Gerber-Jones formulas

The goal is to apply the Gerber-Jones formulas to a realistic model (ultimately, geometric Brownian motion for trend, and observation error with a constant coefficient of variation) of the relationship between historical data and the unknown future loss

[1] A logical resolution of this would be to invoke best estimate principles.

[2] While the approach of this paper is that of best estimate credibility, it is important to note that this approach involves much more analysis than limited fluctuation credibility. So, there may certain cases where cost limitations support the use of limited fluctuation credibility. If the complement of credibility involves small rate changes, limited fluctuation credibility may serve the goal of rate stability. If the best estimate credibility does not serve that goal effectively, there may be some circumstances where limited fluctuation credibility is preferable.

[3] See Bühlmann (1967).

[4] To avoid confusion between the variable names used in various formulas, this will generally be referenced with "U" used in lieu of "P" within this article. The reader is advised to be aware of the alternate notation.

[5] See Boor (1996).

[6] Other relevant papers include those by Mahler (1998) and Ledolter, Klugman, and Lee (1991).

[7] In this context, both variances are intended to have meaning in a broad sense rather than the mathematically narrow definition of variance.

cost. So, to facilitate the reader's understanding, the key Gerber-Jones formulas are shown below.

The first statement that must be made is that the Gerber-Jones formula, and, unless stated otherwise, all other formulas, assume that any necessary trend and current level adjustments have already been made to the data. For example, although the prior data used in a credibility formula involves trending and current level adjustments, those adjustments are assumed to have been done[8] in the background, so all that is involved is determining the optimum credibility weights for the previous years.

With that background, a credibility formula[9] and data pattern is of the updating type[10] through the $n+1^{st}$ projection (e.g., the optimum[11] estimate of future loss costs P_{n+1} is a credibility weighted average $P_{n+1} = Z_n S_n + (1 - Z_n) P_n$ of the previous estimate of loss costs P_n and the new data S_n) if there is a constant μ and sequences $V_1, V_2, \ldots, V_n$ and $W_1, W_2, \ldots, W_n$ such that

$$E[S_i] = \mu \qquad \text{for each of the } S_i, \tag{2.1}$$

$$Cov[S_i, S_j] = V_i + W_i$$

$$\text{for each case where } i = j, \text{ and} \tag{2.2}$$

$$= W_i \qquad \text{if } i < j. \tag{2.3}$$

Further, when the credibility formula and data pattern are of that updating type, then the optimum credibilities are

$$Z_i = \frac{W_i - W_{i-1} + Z_{i-1}V_{i-1}}{W_i - W_{i-1} + Z_{i-1}V_{i-1} + V_i} \tag{2.4}$$

and

$$Z_1 = \frac{W_1}{W_1 + V_1}. \tag{2.5}$$

2.2. The linear updating-type formulas

As a first step towards understanding the notation, it is helpful to introduce the credibility under a standard linear Brownian motion with a drift (T), variance parameter "δ^2" for the Brownian motion, and a constant error variance "σ^2" between each trended data point $S_i = S_i^* + (n+1-i)T$ and the trended underlying expected cost at period i, or $L_i = L_i^* + (n+1-i)T$. Logically, the actual deviations from the expected loss ($S_i - L_i = E_i$ per this linear model) could be expected to be independent from both each other and the L_i's. Of note, this treatment is not new, but is presented so that the reader may understand the process.

Then, if we take "μ" to be the true mean expected loss[12] at time[13] $n+1$, $\mu = L_{n+1} = E[P_{n+1}]$ then the underlying prior expected loss follows a Brownian motion. Further, since $Cov[A + \alpha B, C + \beta B] = \alpha\beta Var[B]$ when A, B, and C are mutually independent,

$$Cov[S_i, S_j] = Cov[L_i + E_i, L_i + (L_j - L_i) + E_j]$$

$$= Var[L_i] = i\delta^2 \; (i < j) \tag{2.6}$$

(noting that L_j is further along in the Brownian motion than L_i, the random motion between L_i and L_j is independent of L_i). Further,

$$Cov[S_i, S_i] = i\delta^2 + \sigma^2 \tag{2.7}$$

So, in the Gerber-Jones formula

$$V_i = \sigma^2; \text{ and} \tag{2.8}$$

$$W_i = i\delta^2. \tag{2.9}$$

[8]A discussion of the rationale for separating the trend estimation from the variance parameter estimation will be presented later.

[9]This presentation of the Gerber-Jones formulas is slightly weakened to simplify the presentation for a more general audience. The interested reader is encouraged to review the original article for the broader result.

[10]Any credibility process on any set of data points can be designed so that the rate indication is the optimum combination of the new data point and the current rate. Credibility formulas of the updating type note that the current rate is merely a prior credibility-weighted combination (in practice, usually with trending) of the prior data points. For a credibility formula to be of the updating type, the credibility-weighted combination of the new point and the current rate is also the optimum combination of all the prior data points if they were considered individually.

[11]Least squared error, considering all possible linear combinations of the adjusted prior year data.

[12]Here μ is not used in the same sense as in the Jones and Gerber (1975) paper, where it is L_0.

[13]The resulting credibility is identical if some $L_{n+\Delta t}$ is desired instead of L_{n+i}, as one may affirm by plugging in the alternate covariance structure. $W_n - W_{n-1}$ is unaffected.

Hence, per formula (2.4),

$$Z_i = \frac{\delta^2 + Z_{i-1}\sigma^2}{\delta^2 + Z_{i-1}\sigma^2 + \sigma^2}, \tag{2.10}$$

where each Z_i is the optimum credibility to use when combining the new data (S_i) with the prior estimate (P_i) to produce the optimum estimate P_{i+1} of L_{i+1}. Further, the resulting combination of all the prior data points $S_{j<i+1}$ that P_{i+1} represents is the optimum estimate of L_{i+1} given the available data.

Jones and Gerber (1975) also show that the successive Z_i's converge to a limit (which could conceivably be used as a proxy for the credibility Z_i when i is large). In this scenario, setting $Z_i = Z_{i-1}$ in formula (2.10) and solving for Z_i gives

$$Z = \frac{\delta^2\left(\sqrt{1 + 4\dfrac{\sigma^2}{\delta^2}} - 1\right)}{2\sigma^2}. \tag{2.11}$$

2.3. The geometric Brownian motion formulas

The linear model has a key weakness—it assumes that the growth in losses is linear. In fact, it is well-established that most insurance lines of business suffer inflation that causes loss costs to grow exponentially rather than linearly. That reality requires an adjustment to the Brownian motion model. Instead of having $E[L_i - L_{i-1}] = 0$ for each i, we should expect zero growth ($E[L_i/L_{i-1}] = 1$). Instead of expecting the $L_i - L_{i-1}$'s to have identical and independent normal distributions, one would expect the L_i/L_{i-1}'s to have independent identical lognormal distributions, with the aforementioned mean of unity and some common variance of δ^2. So if one begins with unadjusted data points, each denoted as S_i^*, the points used to estimate $L_{n+1} = E[P_{n+1}]$ are the inflated values $S_i^*(1+T)^{n+1-i} = S_i$'s.

Lastly, a model for the differences between the observed S_i's and the true expected costs, the L_i's, must be included. In this model, the ratios S_i/L_i are assumed to have independent, identical lognormal distributions with a mean of unity and a constant variance of σ^2. These distributions are also expected to be independent from those of the year-to-year drifts (L_i/L_{i-1}'s). The common observation variance of the trended values assumption is consistent with roughly equal numbers of claims from year to year with severity inflation affecting the loss sizes. It would be less proper for an increasing book of business that encompasses more and more expected claims from year to year with consequent reductions in the coefficient of variation of the process variance.

In any event, the covariance structure, using the identity[14] $Cov[AB, CB] = E[A] \times E[C] \times Var[B]$, is[15]

$$\begin{aligned}
Cov[S_i, S_j] &= Cov[L_i \times E_i, L_i \times (L_j/L_i) \times E_j] \\
&= E[L_j] \times E[L_{j-1}/L_j] \times, \ldots, \times E[L_i/L_{i\mp1}] \\
&\quad \times E[E_i] \times Var[L_i]\,(i > j) \\
&= 1 \times 1 \times, \ldots, \times 1 \times Var[L_i] \\
&= (\delta^2 + 1)^1 - 1\,(i > j),
\end{aligned} \tag{2.12}$$

Further, by the identity $Var[AB] - Var[A]Var[B] + E[A]^2 Var[B] + E[B]^2 Var[A]$,

$$Cov[S_i, S_i] = \sigma^2(\delta^2 + 1)^i + (\delta^2 + 1)^i - 1. \tag{2.13}$$

So, the key values for the Gerber-Jones formula in this case are

$$W_i = (\delta^2 + 1)^i - 1; \tag{2.14}$$

$$V_i = \sigma^2(\delta^2 + 1)^i; \text{ and,} \tag{2.15}$$

$$Z_i = \frac{\delta^2 + Z_{i-1}\sigma^2}{\delta^2 + \delta^2\sigma^2 + Z_{i-1}\sigma^2 + \sigma^2}. \tag{2.16}$$

A comparison to equation (2.10) shows that this is identical to the formula for the linear case, except for the additional $\delta^2\sigma^2$ term in the denominator.

[14]This holds when A, B, and C are independent.

[15]The last step in this calculation, computing the final variance, involves properties of the lognormal distribution. Details of the mathematics are not presented, as there is an opportunity for confusion between the σ used in this paper to denote observation variance and the "sigma" parameter used in specifying lognormal distributions.

But, one should consider that when at least one of the values δ^2 and σ^2 is very small, the combination term $\delta^2\sigma^2$ should be a small part of the denominator. Thus, one might say that, for the case of geometric Brownian motion,

$$Z_i \cong \frac{\delta^2 + Z_{i-1}\sigma^2}{\delta^2 + Z_{i-1}\sigma^2 + \sigma^2}. \tag{2.17}$$

Further, the steady-state credibility may be approximated as

$$Z \cong \frac{\delta^2\left(\sqrt{1+4\dfrac{\sigma^2}{\delta^2}}-1\right)}{2\sigma^2} \tag{2.18}$$

As a relevant side note, the summands involved in equations (2.13) and (2.14) would inflate uniformly as the losses are projected ahead more than one year, to some $n + \Delta t$ instead of to time $n + 1$, and the credibility equation would remain unchanged.[16]

3. Multi-year formulas and best estimate credibility for the overall rate indication

The approach outlined earlier involves updating a rate with a single new year of data. But it is very common to see rate indications that update a rate with, say, the weighted average of the data from the last five years. The role of this multi-year data in a best estimate credibility formula merits discussion.

3.1. Reasons not to reuse older years

Updating formulas that use multiple years reuse data from prior estimates. So, the reuse of data should be evaluated. The first point to be made is that using multiple years is perfectly appropriate when limited fluctuation credibility is involved. Limited fluctuation credibility deals solely with the extent to which the body of data receiving credibility can be relied on to not create unwarranted increases or decreases of some specified size. It does not purport to create a best estimate of the future costs. It has been stated, though, by the well-respected Howard Mahler in 1986 that this method often produces future loss estimates that are comparable to those of best estimate credibility.

To state it simply, re-using prior years in a Gerber-Jones formula unduly complicates the computations. For example, assume an estimate has been continually updated over 14 years from P_1 and S_1 to P_{15} with rolling five-year averages[17] $Q_1, \ldots, Q_{14}$ of the data points $S_1, \ldots, S_{14}$. Logically, the step is to produce the estimate P_{16} using Q_{15}. Note, though, that the covariance between Q_{15} and Q_{14} is fairly high, since they have the points S_{11}, S_{12}, S_{13}, and S_{14} in common. However, Q_{15} and Q_1 have no common components.[18] Generally,[19] $Cov[Q_{15}, Q_{14}] \neq Cov[Q_{15}, Q_1]$. Therefore, the Gerber-Jones formula cannot be used when multiple years are combined.[20] Therefore, the practice of combining multiple years of data in this context is suboptimal.

That conclusion has a very relevant corollary. If the exposures most useful for limited fluctuation credibility stem from five or even ten years, but best estimate credibility is only based on the most recent year, the resulting credibilities should by nature be different. Therefore, there are circumstances where limited fluctuation credibility is not a good substitute for best estimate credibility.

3.2. Correcting the prior estimate for changes in ultimate loss estimates

There is, however, one respect in which the use of multiple years could improve the estimate. The existing rate is based on the data available earlier, when the various years' losses were less mature than they are at the time of the updated rate indication.

[16]For reference, this is also true for the linear case.

[17]Or an average of the available years, where fewer than five are available.

[18]Likely they do have some indirect common components stemming from the dependence of the backwards drift per a Brownian motion of some type as per sections 2.2 and 2.3.

[19]A more detailed analysis of this may be found in Appendix B.

[20]It may, alternately, work when rates are made biannually and two years of updating data are combined.

So, it makes sense to update the existing rate for the additional development before using it in the credibility formula. Of course, the existing rate is a multiple credibility weighted average of many years. Further, it is not just an average of many years of loss ratios or pure premiums, it is rather either an average of trended loss ratios brought to the current rate level or trended pure premiums. So, some calculations must be done to include this additional loss development in the prior rate that is used as the complement of credibility. Due to the requirement to use current level data, the correction process for loss ratio ratemaking is slightly more complex than that of pure premium ratemaking. Therefore, Table 1 shows how the calculations needed to update a loss ratio at present rates for loss development might flow.

The references to "Prior" and "Last Prior" refer to the data used in computing the loss ratio estimate that was used in the last rate change. The "First Assigned" values refer to what was used the first time the specific year of data was used. Also, note that although the loss ratios of many years are likely embedded in the prior loss ratio, only the last five were revised. That is because more mature years see fewer year-to-year revisions in ultimate losses, and contribute a diminishing portion after credibility (see column 7).

Table 1. Sample update of prior rate review loss ratio information for ultimate loss changes

Year	(1) (Data) Loss Ratio at Charged Rates @12/31/10	(2) (Data) Loss Ratio at Charged Rates @12/31/11	(3) (2)–(1) Absolute Loss Ratio Change	(4) (Data) Last Prior Current Level Factor
2006	75%	74%	–1%	1.390
2007	98%	90%	–8%	1.300
2008	32%	40%	8%	1.210
2009	75%	70%	–5%	1.150
2010	64%	52%	–12%	1.100
Year	**(5) (Data) Credibility First Assigned**	**(6) 1.0–(5) Complement of Credibility**	**(7) (6)∗[Next (7)] Complement of Credibility**	**(8) (5)∗[Next (7)] Credibility in Last Prior**
2006	45%	55%	9%	7%
2007	32%	68%	16%	8%
2008	38%	62%	24%	15%
2009	35%	65%	39%	21%
2010	40%	60%	60%	40%
Year	**(9) (Data) Trend Rate First Assigned**	**(10) [1.0+(5)]∗[Next (10)] Total Trend Factor in Last Prior**	**(11) (3)∗(7)∗(10)/(4) Change to Prior Estimate**	**(12) (Selected) Change to be Reflected**
2006	6%	1.469	–0.08%	–0.08%
2007	7%	1.386	–0.66%	–0.66%
2008	8%	1.295	1.27%	1.27%
2009	9%	1.199	–1.09%	–1.09%
2010	10%	1.100	–4.80%	–4.80%

A. Total Change to Prior	–5.36%
B. Prior Loss Ratio for Ratemaking	65.72%
C. Last Rate Change Taken	–5.00%
D. Trend Factor for this Filing	1.12
E. = (B.+A.)8D./(1.0+C.) New "Prior" Value to which Complement of Credibility is Applied	71.16%

It is also worth mentioning that in this example the current level factors could be updated for the next rate review by simply multiplying column (4) by unity plus item "C". Similar adjustments could be made for the "Credibility in Last Prior" and "Total Trend Factor in Last Prior" columns.

Of course, this example mirrors the calculations in the theoretical literature—the data is assumed to be collected at midnight of December 31, 2011, then used to make rates that are effective at 12:01 a.m. of January 1, 2012. However, the corrections needed to reflect practical realities would appear to be straightforward.

3.3. Updated ultimate losses and updating-type credibility

It could be expected that the process of updating prior year ultimate losses could distort the optimum credibility. In lines such as excess casualty reinsurance, the ultimate loss estimates S_n, S_{n-1}, etc., for the most recent years could have a very high observation error, and those five or so years back could be much closer estimates of the true expected loss L_i's within their respective years. So, on that basis the true optimum credibility could be expected to be larger for some of the "older" years than the most recent year. However, that would clearly not create an "update."

Some perspective can be provided about this situation. First, when prior year estimates are not corrected, the formulas of section 2 do provide the optimum credibility. Further, updating the prior year ultimate losses can only be expected to improve the accuracy of the resulting loss prediction. So, this approach can be expected to produce a high quality estimate of future costs, up to any distortion due to lengthy loss development.

If loss development uncertainty is expected to significantly distort the credibility, it may well be preferable to simply start from scratch each year with the ultimate loss estimates for, say, the last twenty years. One may then compute estimates of the process variance in each year, estimates of the loss development error variance in each year, and the Brownian motion-type variance parameter.[21] It is not difficult to see that, under the linear model (possibly the geometric as well), an updating formula can be derived for the assignment of weights to the various years. It should be clear that the resulting credibility weights may differ greatly between years. However, it does not involve the sort of updating of the prior rate that is part of the typical actuarial application. Rather it involves simply computing a rate from scratch.[22] Since the focus of this paper is on updating an existing rate with new data, this situation will not be analyzed further in this paper.

4. Estimating the parameters: Z, K, B, δ^2 and σ^2

The section will give the reader some tools for creating estimates of the key variances, and thus help create better loss cost projections. It is not intended to be a survey on the subject. Rather it is intended to give the practitioner the tools needed to implement best estimate ratemaking. The interested reader may review some of the ideas in De Vlyder 1981 and Hayne 1985, to get two other perspectives on this subject.

First, a few quick notes are in order:

Note 1. In many situations, it is not necessary to estimate both δ^2 and σ^2. Key formulas can be converted to a function of $K = \delta^2/\sigma^2$, so K is all one needs to estimate.

Note 2. When estimating δ^2 and σ^2 for geometric Brownian motion, note that they are functions of δ'^2 and σ'^2 from the logarithmic transform to a linear Brownian motion, $\exp(\delta'^2) - 1 = \delta^2$, and $\exp(\sigma'^2) - 1 = \sigma^2$. So, once one determines how to estimate the constants of variance (or even just their ratio) in a linear Brownian motion, one may estimate the credibility for the geometric Brownian motion.

Note 3. The observation errors (with variance σ^2) consist logically of a combination of the sample variance (i.e., the limitations of the law of large

[21]A further description is beyond the scope of this paper, but once the concept of an updating formula is abandoned, it may be preferable to use a model such as integrated Brownian motion to better mirror reality.
[22]See Appendix B for a deeper discussion of the initial credibility Z_1.

numbers due to the high skew in insurance statistics and inability of "small" claim samples to fully estimate the true expected losses each year) and the loss development uncertainty between the early data we base our projections on and the final actual claims costs in each year. Further, the sample variance and development variance are independent and so may be added to determine σ^2.

Note 4. (Subtraction of Two Estimated Quantities) If we subtract one highly uncertain "large" number from another "large" number, and the difference is "small," the result has a "large" variance most of the time. When estimating a small number, that "large" variance typically overwhelms the true "small" value one seeks to estimate.

Note 5. (Common Additive Error in all the Data) If all the historical data points are affected equally and simultaneously by a common error that is independent of all the other error terms (for example, all the data is biased by addition of a single, uniform, unknown, amount "ε" from some distribution with a zero mean), then the optimal solution may be estimated by disregarding this error. Logically, this may be converted algebraically to a situation where one is estimating a future value that contains ε, with ε removed from all the historical data. Since the variance of ε is independent of all aspects of variance in the historical data, the ε component of the costs being predicted is not susceptible to estimation using the historical data. Hence, it may be disregarded in optimizing the estimate of future costs. A similar result holds when ε is a constant error multiplier with a mean of one within the data, except that one must consider that the mean of the inverse of ε may not be unity.

With those concerns in mind, a few methods for estimating the key parameters follow.

4.1. Method 1: The credibility that would have worked in the past.

This approach actually involves no estimation of δ^2 or σ^2; rather, it estimates Z directly. Since estimating Z directly removes the barriers to implementing best estimate credibility for the overall rate indication, it merits discussion (even though it does not involve δ^2 and σ^2). The basic methodology involves assuming some credibility value Z, then using all the data but the last year to estimate the last year given. Assume that one has, say, ten years of on-level, appropriately trended[23] loss ratios. Then, one could note that the fifth year's value could be estimating by first applying some unknown credibility factor Z to the fourth[24] year's data, $Z(1-Z)$ to the third year's data, $Z(1-Z)^2$ to the second year's data, etc., then dividing by the sum of the credibilities, $1-(1-Z)^4$, to correct for the off-balance. In effect, a single credibility value is assumed to have been proper for all four updates.

Once that equation is established, one could vary Z in order to find which Z minimizes the squared difference between the fifth year's data and the credibility-weighted average. Most modern spreadsheet programs contain solution-generating capabilities that make it straightforward to find such a solution. Then, one may also construct similar equations to solve for a common credibility of Z that use the first five values to predict the sixth, the first six values to predict the seventh, etc. The last step involves replacing the individual solutions of Z that each minimize the squared error of a single predictive step with a solution of a single Z that minimizes the sum of all the squared errors of all the predictive steps simultaneously.

The resulting Z is arguably the best estimator of the credibility in the data, at least as long as a single credibility is appropriate for all the years.

Table 2 illustrates how this process would work with ten years of essentially random sample data. The shaded boxes show the inputs and outputs to the solution process (note that the "Target" box pulls

[23]This may involve, since the time period is so long, using different trends for different prior periods.

[24]This seemingly contemplates the same "stroke of midnight" ratemaking issue discussed earlier. However, note that since the random variation between say, the fifth year and the sixth year affects all the historical year estimate errors identically; note 5 indicates that the Z best suited to estimate the sixth year from the first four years is the same Z that is best suited to estimate the fifth year from the first four years.

Table 2. Sample calculation of *Z* from initial reported data and final cost of ten years of data—when data has zero trend

Input/Output for Solution Function			
Value to minimize	**Target =**	0.046	
= Value to vary to minimize Target is		***Z* =**	0.366

Part 1. Data and Estimation of Older Years

Accident Year	(1) Data Initial Data Values	(2) Data Final Ultimate Value	(3) *Z*((1−*Z*)^k) All Estimating Weights	(4) [5 Later (3)] Weights for Estimating 1995	(5) (1)∗(4) 1995 Estimate	(6) [4 Later (3)] Weights for Estimating 1996	(7) (1)∗(5) 1996 Estimate
1991	1.023	1.070	0.010	0.093	0.095	0.059	0.061
1992	0.991	1.107	0.015	0.147	0.146	0.093	0.092
1993	1.209	1.022	0.024	0.232	0.280	0.147	0.178
1994	0.576	0.923	0.038	0.366	0.211	0.232	0.134
1995	0.886	0.769	0.059		0.000	0.366	0.324
1996	0.858	0.907	0.093		0.000		0.000
1997	0.810	0.880	0.147		0.000		0.000
1998	1.061	0.871	0.232		0.000		0.000
1999	0.891	0.767	0.366		0.000		0.000
2000	0.967	0.826	0.000		0.000		0.000
A. Column Sums				0.838	0.732	0.897	0.788
B. (A./[A. in Prev. col.] Loss Ratio Est.					0.874		0.879
C. (from (1)) Actual Loss Ratio Values					0.769		0.907
D. (B-C.)^2 Squared Error in Estimate					0.011		0.001

Part 2. Estimation of Remaining Years and Total Prediction Error (Target)

Accident Year	(8) [3 Later(3)]∗(1) 1997 Estimate	(9) [2 Later (3)]∗(1) 1998 Weights	(10) [Next Row(3)]∗(1) 1999 Estimate	(11) (3)∗(1) 2000 Estimate	
1991	0.038	0.024	0.015	0.010	
1992	0.059	0.037	0.024	0.015	
1993	0.113	0.072	0.045	0.029	
1994	0.085	0.054	0.034	0.022	
1995	0.206	0.130	0.083	0.052	
1996	0.314	0.199	0.126	0.080	
1997	0.000	0.296	0.188	0.119	
1998	0.000	0.000	0.388	0.246	
1999	0.000	0.000	0.000	0.326	
2000	0.000	0.000	0.000	0.000	
A. (as above)	0.814	0.812	0.903	0.899	
B. (as above)	0.871	0.847	0.928	0.914	**Sum of Est. Errors = Target**
C. (as above)	0.880	0.871	0.767	0.826	
D. (as above)	0.000	0.001	0.026	0.008	0.046

up the "Target" value computed at the bottom of the spreadsheet).

This method has good utility as long as δ^2 and σ^2 are stable over time and the data is not prone to very rare large losses.[25] It is reasonable to expect δ^2 to be stable as long as the average trend factor is stable, but often that does not occur. Further, it would be reasonable to expect σ^2 to be fairly stable as long as the premium volume in the line, adjusted for trend, is stable.

What must be said. This approach has nothing to do with the formulas stated earlier. However, it does address the key question in this paper, determining the optimum credibility. Further, since Z has a formula in δ^2 and σ^2, it may also used to determine a second variance constant once a first variance constant is known. Then, one might possibly revise the estimate of σ^2 (derived from Z and δ^2) to better account for process variance due to large losses, and consequentially revise the estimate of Z.

4.2. Method 2: Fitting *K* and *B* across a large number of similar datasets

In this case, one might assume that the ratemaker is computing rates for a single line of business in 50 U.S. states, or some other situation where there is a fairly large number of segments, and all the segments have approximately the same trend and observation-error-variance-per-unit-of-exposure characteristics. One would also have to assume that the complement of credibility is still supposed to be assigned to the existing rate plus trend, not some amalgam of all the segments. One must also assume that the old premium/exposure and loss data using in pricing the last, say, twelve years of rates are available for each of the segments. And lastly, it would help if the second-to-last data point for each segment, possibly the last data point, is developed enough that each value $L_{n+1,s}$, for each class (s) is as close an estimate of the expected costs $E_{n+1,s}$ as is reasonably possible.

[25]On the other hand, if one converts to basic limits ratemaking with (necessarily) reliable increased limits factors, then that problem may be significantly mitigated.

Just like the estimation of Z in the previous subsection, K and B may be estimated from the data by solving for the values that would produce the best estimates of the most recent costs in the various segments. In the previous subsection the total squared differences between the credibility-weighted average of various sets of years and the future years they project were minimized. In this case, for each segment "s," one must construct the credibility-weighted average $P_{n,s}$ of the last n (= 10, or 5, or whatever is most feasible) years of data (the $S_{i,s}$'s) in order to estimate each $L_{n+1,s}$. In doing so, the credibilities should be computed using formula (A.7)

$$Z_{i,s} \cong \frac{U_{i,s} + Z_{i-1,s}\left(K + BU_{i,s}\right)}{U_{i,s} + \left(1 + Z_{i-1,s}\right)\left(K + BU_{i,s}\right)}. \tag{4.1}$$

Per the solution routine, K and B should then be modified so that the squared errors the resulting $P_{n+1,s}$'s make in estimating the $L_{n+1,s}$'s are minimized. Crucially, K and B are not to vary from segment to segment. Rather, a single pair of K and B that minimize the sum of all the squared prediction errors is to be found via the solution algorithm.

So the weight assigned to the year $n-i$ data for the line s data, $S_{n-i,s}$, is

$$M_{n-i,s} = \left(1 - Z_{n,s}\right)\left(1 - Z_{n-1,s}\right)\ldots\left(1 - Z_{n-i+1,s}\right)Z_{n-i,s}. \tag{4.2}$$

The resulting predictions[26] of the $L_{n+1,s}$'s are then the various values of

$$P_{n+1,s} = \sum_{i=1}^{n} M_{i,s}S_{i,s} + \prod_{i=1}^{n}\left(1 - Z_{i,s}\right)S_{0,s} \tag{4.3}$$

(where each $S_{0,s}$ represents the rate or rating information in effect just before the experience period).

As before, the sum across all the s's of the squared estimating errors $\Sigma_s(P_{n,s} - L_{n,s})^2$, or perhaps a premium

[26]The first credibility formula (2.5) from Jones and Gerber (1975) of $Z_{1,s} \cong \frac{W_1}{W_1 + V_1} = \frac{U_{1,s}}{U_{1,s} + K + BU_{1,s}}$ was used for the first step. For convenience, the remaining credibility after considering all five years was assigned to the first year of data. In practice, that would be assigned to the rate in effect when the period began, adjusted for trend and to the present rate level.

or exposure weighted average $\sum_s W_{n,s}(P_{n,s} - L_{n,s})^2$ could be computed in the spreadsheet. The resulting value could be called the "Target" and the solution routine or feature could be used to vary K and B until the lowest value of the "Target" is found.

A sample spreadsheet illustrating this approach with 12 data segments and common trend, process, and parameter variance constants, but different samples from those constants among the segments, is shown in Table 3. The expected loss ratios for each segment were simulated using a geometric Brownian

Table 3. Sample calculation of *K* and *B* from data for twelve separate segments subject to a common *K* and *B* (alternate annotation style for matrix data)

Part 1: Distribution Properties

Trend Variance	0.0016	(δ^2)	Values to Vary in Solver			True Values		
Process Variance	0.048	(τ^2—to be divided by premium)	$K =$	**9.2477**	(solver)	30.0	($K = \tau^2/\delta^2$)
Parameter Variance	.0009	(λ^2)	$B =$	**1.4732**	(solver)	.5625	($B = \lambda^2/\delta^2$)

Part 2: Premiums $\{U_{year,class} = U_{i,s}$(Data)$\}$

Year	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12
1	20.00	24.00	28.80	34.56	41.47	49.77	59.72	71.66	86.00	103.20	123.83	148.60
2	22.00	26.40	31.68	38.02	45.62	54.74	65.69	78.83	94.60	113.52	136.22	163.46
3	18.00	21.60	25.92	31.10	37.32	44.79	53.75	64.50	77.40	92.88	111.45	133.74
4	19.00	22.80	27.36	32.83	39.40	47.28	56.73	68.08	81.70	98.04	117.64	141.17
5	21.00	25.20	30.24	36.29	43.55	52.25	62.71	75.25	90.30	108.36	130.03	156.03
Target 6	22.00	26.40	31.68	38.02	45.62	54.74	65.69	78.83	94.60	113.52	136.22	163.46

Part 3: Loss Ratios $\{L_{i,s}$ (Data); in this example, generated using original means (.6,.65,.55 for each group of three) and unity mean/lognormal drift and observation error with drift variance δ^2, observation variance $\frac{\tau^2}{U_{i,s}} + \lambda^2\}$

1	65.9%	65.2%	48.2%	60.2%	62.3%	55.0%	60.1%	66.1%	53.9%	61.8%	67.2%	50.1%
2	62.3%	61.0%	58.3%	57.5%	66.0%	56.1%	57.7%	67.9%	54.3%	63.6%	66.7%	51.0%
3	59.0%	64.8%	53.8%	59.1%	61.8%	59.6%	55.1%	65.8%	52.3%	61.5%	63.6%	49.4%
4	64.3%	74.3%	54.6%	56.3%	64.8%	55.3%	54.9%	61.0%	55.0%	61.3%	66.1%	50.7%
5	68.0%	80.8%	58.3%	50.3%	61.7%	52.7%	51.7%	59.0%	55.9%	60.2%	68.9%	45.3%
Target 6	68.3%	73.6%	61.3%	53.2%	67.1%	51.0%	54.1%	65.8%	58.8%	65.4%	63.5%	43.6%

Part 4: Credibilities $\left\{Z_{i,s} = \frac{[U_{i,s} + Z_{i-1,s}(K + BU_{i,s})]}{[U_{i,s} + (1 + Z_{i-1,s})](K + BU_{i,s})}; \text{except } Z_{1,s} = \frac{U_{1,s}}{U_{1,s} + K + BU_{1,s}}\right\}$

1	0.34	0.35	0.36	0.36	0.37	0.38	0.38	0.38	0.39	0.39	0.39	0.39
2	0.46	0.47	0.48	0.49	0.49	0.50	0.50	0.50	0.51	0.51	0.51	0.51
3	0.49	0.50	0.51	0.51	0.52	0.52	0.53	0.53	0.53	0.53	0.54	0.54
4	0.50	0.51	0.51	0.52	0.52	0.53	0.53	0.53	0.54	0.54	0.54	0.54
5	0.51	0.51	0.52	0.52	0.53	0.53	0.53	0.54	0.54	0.54	0.54	0.54

Part 5: Weights Computed Using Credibilities $\{W_{5,s} = Z_{5,s}; W_{i,s} = Z_{i,s}(1 - Z_{i+1,s}) \times \ldots \times (1 - Z_{5,s}); W_{prior,s} = 100\% - (Z_{1,s} + Z_{2,s} + Z_{3,s} + Z_{4,s} + Z_{5,s})\}$

Prior	0.044	0.041	0.039	0.036	0.035	0.033	0.032	0.031	0.030	0.030	0.029	0.029
1	0.023	0.022	0.021	0.021	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019
2	0.058	0.057	0.055	0.054	0.053	0.052	0.052	0.051	0.050	0.050	0.050	0.049
3	0.121	0.120	0.119	0.117	0.116	0.115	0.114	0.114	0.113	0.113	0.112	0.112
4	0.247	0.248	0.248	0.248	0.248	0.248	0.248	0.247	0.247	0.247	0.247	0.247
5	0.506	0.513	0.518	0.523	0.528	0.532	0.535	0.537	0.540	0.542	0.543	0.544

(continued on next page)

Table 3. Sample calculation of *K* and *B* from data for twelve separate segments subject to a common *K* and *B* (alternate annotation style for matrix data) *(continued)*

Part 6: Projections $\{P_s = [\text{Sum of } W_{i,s} \times L_{i,s})$ for all years "i"$] + W_{prior,s} \times L_{1,s}\}$; and Squared Estimation Errors per Year 6 Observed Data $\{R_s = (P_s - \text{Target } L_{6,s})^2\}$												
Projection to year 6 $\{P_s\}$	65.5%	75.2%	56.3%	53.8%	62.8%	54.4%	53.6%	61.1%	55.1%	60.8%	67.4%	47.6%
Squared Error vs. Target 6 $\{R_s\}$	7.6E-04	2.4E-04	2.5E-03	3.1E-05	1.9E-03	1.2E-03	2.1E-05	2.2E-03	1.4E-03	2.1E-03	1.5E-03	1.6E-03
Straight Sum of Squared Errors $\{R_1 + \cdots + R_{12}\}$												1.54E-02
Weighted (with $U_{5,s}$ for each R_s) Sum of Squared Errors												
—Solution routine varied K and B in previous gray area to minimize this value												**1.46E-03**
Part 7: Validation with Actual Year 6 Underlying Expected Loss Ratios $\{E_s$ (Data)$\}$												
Expected Loss Ratio at Year 6	66.5%	73.8%	57.0%	49.6%	66.4%	50.5%	53.7%	66.4%	57.4%	62.5%	64.9%	43.8%
Sum of Errors Projecting Expected Loss with *K, B* in Gray $\{S = \text{sum of } (P_s - E_s)^2\}$												**1.06E-02**
Sum of Error w/True *K,B* {*T*; same as "*S*" only true *K, B* used throughout process}												**1.07E-02**
Ratio Error w/Est. *K, B* to True *K, B* {*S/T*}												**100%**

motions with the variance specified in Part 1. The actual loss ratios are also affected by the parameter variance and the process variance (a common factor, divided by the premium per the Law of Large Numbers) listed there. The actual values of K and B are on the very left of Part 1. Lastly, the K and B values that minimize the sum of premium-weighted sum of squared errors in projecting the sixth year's simulated value (using the credibility weights[27] defined by K, B, and the premium data) are highlighted in gray.

Note that the loss ratios for year 1 were deemed to have projection errors similar to the rate prior to the experience period, so they were used for the $S_{0,s}$'s.

What must be said. In testing this method, it appears that it may require a substantial number of data points to reliably estimate of K and B using this process. In particular, twelve classes do not appear to be sufficient for the test case above. However, the fact that K and B are combined as $K + BU$ in the equation means that they act together to impact the credibility. The only difference is that the "B" term reacts to exposure or premium volume, whereas "K" does not. In this case, at a premium of about 20 the estimated value of $K + BU$ is about equal to the true underlying value.

Next, the actual quality of the estimation, the errors in estimating the true (unaffected by process or parameter variance) expected loss ratios for year 6 (as shown at the top of Part 7) were computed. As one may see, the difference between the prediction error using the estimated K and B and the actual K and B is negligible. This suggests that, as long as the sample size (number of "s" values) is small and the difference in premiums, exposures, etc., is small, it may be more helpful to simply replace "$K + BU$" with "K" in the credibility formula.

4.3. Method 3: Estimating δ^2 and σ^2 from the historical data

This method involves using different linear combinations of squared differences between values. As such, it is oriented towards standard, linear, Brownian motion. However, note that the logs of values from a geometric Brownian motion form a linear Brownian motion. So, one may convert geometric Brownian motion data to linear data, estimate the values of δ^2 and σ^2 that work in the linear context, then convert those to comparable drift variance and process/

[27]For example, the most recent year has credibility Z_n, the former year has $(1 - Z_n)Z_{n-1}$, and then $(1 - Z_n)(1 - Z_{n-1})Z_{n-2}$, etc., as credibility has been applied at successive updates.

parameter variance values. For example, the geometric Brownian motion variance parameter would be $e^{\delta^2} - 1$ when δ^2 is the variance in the corresponding linear Brownian motion and the mean of the geometric Brownian motion steps is specified to be unity (no change in the multiplicative context).

So, the goal is to find functions of the S_i's that provide insight into the values of δ^2 and σ^2. For example, the squared difference between the beginning and ending values $(S_n - S_1)^2$ reflects two samples of parameter/process error at the two endpoints and $n - 1$ samples from the Brownian motion variance. So, if the two types of variance are similarly sized, the squared difference between the two endpoints should be dominated by a multiple of the Brownian motion variance δ^2. Similarly, if one adds the squared differences between adjacent points $\sum_{i=1}^{n-1}(S_{i+1} - S_i)^2$ one would expect the result to be dominated by a multiple of the process[28] variance σ^2. Further, one might expect that more precise approximations might be made by using linear combinations of those two values.

So, one might begin by computing the expected values of $(S_n - S_1)^2$ and $\sum_{i=1}^{n-1}(S_{i+1} - S_i)^2$. First, note that, since the mean expected change in values from the Brownian motion (after trend correction) is zero, and the expected process risk is zero.

$$E\left[(S_n - S_1)^2\right] = Var\left[S_n - S_1\right] \tag{4.4}$$

However, $S_n - S_1$ may be expressed as a sum of independent variables, each with mean zero, as $(S_n - L_n) + (L_n - L_1) + (L_1 - S_1)$. So, it is composed of a process error, a Brownian motion of length $n - 1$, and the negative of a process error. Therefore,

$$\begin{aligned} &E\left[(S_n - S_1)^2\right] \\ &= Var\left[S_n - L_n\right] + Var\left[L_n - L_1\right] + Var\left[L_1 - S_1\right] \\ &= \sigma^2 + (n-1)\delta^2 + \sigma^2 = (n-1)\delta^2 + 2\sigma^2. \end{aligned} \tag{4.5}$$

[28]To avoid the cumbersome phrase "parameter/process variance," the simpler phrase "process variance" should be understood to have the same meaning throughout this subsection.

Similarly,

$$\begin{aligned} &E\left[\sum_{i=1}^{n-1}(S_{i+1} - S_i)^2\right] \\ &= \sum_{i=1}^{n-1} Var\left[L_{i+1} - L_i\right] + 2\sum_{i=2}^{n-1} Var\left[S_i - L_i\right] \\ &\quad + Var\left[S_n - L_n\right] + Var\left[S_1 - L_1\right] \\ &= (n-1)\delta^2 + (n-2)\sigma^2 + \sigma^2 + \sigma^2 \\ &= (n-1)\delta^2 + 2(n-1)\sigma^2. \end{aligned} \tag{4.6}$$

Knowing those values, it is possible to construct estimators for δ^2 and σ^2. One may readily see that, by the linearity of expectations,

$$\begin{aligned} &\frac{E\left[\sum_{i=1}^{n-1}(S_{i+1} - S_i)^2 - (S_n - S_1)^2\right]}{2(n-2)} \\ &= \frac{(n-1)\delta^2 + 2(n-1)\sigma^2 - \left((n-1)\delta^2 + 2\sigma^2\right)}{2(n-2)} = \sigma^2, \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} &E\left[\frac{(n-1)(S_n - S_1)^2 - \sum_{i=1}^{n-1}(S_{i+1} - S_i)^2}{(n-1)(n-2)}\right] \\ &= \frac{(n-1)\left\{(n-1)\delta^2 + 2\sigma^2\right\} - (n-1)\delta^2 + 2(n-1)\sigma^2}{(n-1)(n-2)} \\ &= \delta^2. \end{aligned} \tag{4.8}$$

So, by creatively using the differences between the first and last point, and the differences between adjacent points, one may estimate the values of δ^2 and σ^2.

An example of the use of equations (4.7) and (4.8) is shown in Table 4. The actual observable data over 15 years in column 2 was generated randomly over 15 years, using the actual values $\delta = 3\%$ and $\sigma = 7\%$. The values of δ^2 and σ^2 were then estimated from the data. As one may see, the estimates are fairly close. But they nonetheless significantly overestimate the credibility.

Table 4. Sample estimation of δ^2 and σ^2 from historical data

Part 1: Data	
Brownian S.D.	3%
process S.D.	7%
Implied K	5.444

Part 2: Data and Analysis

(1) Year	(2) Brownian Expected Loss	(3) Process Error	(4) (2)+(3) Including Process Error (Observed Data)	(5) (4)–Previous (4) Annual Change	(6) (5)*(5) Squared Annual Changes	(7) (4)(end)–(4)(begin) Total Change from Beginning to End	(8) (7)*(7) Squared Total Change
1	0.650	0.032	0.682				
2	0.617	–0.051	0.566	–0.116	0.013		
3	0.621	0.116	0.738	0.172	0.029		
4	0.578	0.012	0.590	–0.148	0.022		
5	0.590	–0.033	0.557	–0.032	0.001		
6	0.593	–0.016	0.577	0.020	0.000		
7	0.603	0.082	0.685	0.108	0.012		
8	0.581	–0.032	0.549	–0.136	0.018		
9	0.613	–0.033	0.580	0.031	0.001		
10	0.585	0.004	0.589	0.008	0.000		
11	0.586	0.098	0.684	0.095	0.009		
12	0.618	–0.057	0.561	–0.123	0.015		
13	0.566	0.019	0.585	0.024	0.001		
14	0.557	–0.018	0.539	–0.046	0.002		
15	0.484	0.026	0.510	–0.029	0.001		
Total					0.1251 A.	–0.172	0.0297 B.

C. Estimate of process variance: [A.–B.]/[2(15–2)] Associated standard deviation	0.0037 6%
D. Estimate of variance parameter for Brownian motion: [(n–1)B.–A.]/[(15–1)(15–2)] Associated standard deviation	0.0016 4%
E. Value of K = C./D.	2.293
F. Estimated Steady-State Credibility (equation (2.11) formula using C. and D.)	48%
G. True Steady-State Credibility (equation (2.11) formula using values at top)	35%

A note about trend—The theory underlying this paper assumes that the expected loss, a priori, is the same for all years. That generally requires that historical losses have been trended (and premiums adjusted to the current rate and exposure level) before the calculations commence. Of course, if the trend is computed using the same data as the calculations, the calculated value of δ^2 may be suppressed. For example, if the random movement began with a large upward jump early in the period, and another jump later, because the value of δ^2 is high, the analysis of trend may incorrectly infer that it is high trend rather than a high Brownian motion variance. Of course, if the trend is clearly much larger than δ^2, it may well be less of an issue.

Further, as noted later, the problem of estimating δ^2 and σ^2 is relatively ill-conditioned.[29] So reducing

[29]For example, a big outlier could have arisen from either process variance or drift variance. Thus it would be difficult to infer which type of variance is "large" from the observed data points.

the degrees of freedom of the approximation by estimating trend simultaneously, given a small number of data points, may not be reliable. However, one might be advised to use some related data, such as calendar year reported loss frequency and calendar year closed claim severity, to estimate the trend. On the other hand, if there are a large number of data points relative[30] to the volatility in the data, then the impact of the random observation error in the initial and ending points on the trend estimate should be minimal.

A third aspect of trend deserves mention as well. Without a correction, the random lognormal aspect of geometric Brownian would produce a mean above one at all points after it begins. In effect, the randomness of the distribution combined with the skew of the lognormal tends to generate its own trend. So, the transformed (into a linear version) version of the data points, rather than having a normal-type[31] distribution with mean zero, must have a lognormal distribution with mean $-\frac{\delta^2}{2}$. That means that external trend must often be corrected, especially trend computed by averaging several year-to-year growth rates. To complicate matters, δ^2 is then unknown, so the value needed for the correction is unknown. However, some crude initial estimate of the value of δ^2 may be used when estimating trend, and then, once the trend is estimated, the δ^2 estimate may be refined, etc. The process may be continued iteratively until a consistent trend and δ^2 are computed. Consider that if the estimate is produced by loglinear regression of data with similar geometric Brownian motion variance, $-\frac{\delta^2}{2}$ should already be subsumed into the trend. Further, if quality surrogate data is available for trending, that option deserves serious consideration.

What must be said. There are some special considerations that should help explain why the approximations are not more precise. First, it may be difficult to distinguish say, whether a very high last point is due to a very high uptick in the Brownian motion because δ^2 is large, or a large process error because σ^2 is high. So, the basic problem of approximating δ^2 and σ^2 may often be ill-conditioned. Second, it is important to review Note 4 at the beginning of this section. At its core, Note 4 says that the error variance in computing the quantities above could be as much as the sum of the variances of the two items you are subtracting. While the error does not quite reach the sum of the variances (due to inter-correlation of the two quantities), one should still be extremely cautious if the difference (the estimate of δ^2 or σ^2) is much smaller than each of the values involved in the subtraction.

Nevertheless, even though the credibility determined using this method sometimes only has moderate precision, it is moderately close to the "best estimate" credibility. Therefore, it still has the potential to create more accurate estimates than the stability-centered classical credibility.

4.4. Method 4: Estimating σ^2 structurally from loss data and δ^2 by subtraction

Given the formulas in equations (4.5) and (4.6), it is clear that, once one of δ^2 and σ^2 is reliably estimated, the other may be estimated. It should also be clear that equation (4.5) has relatively more content in δ^2 than equation (4.6). So, if one has a quality estimate of σ^2, the formula

$$\delta^2 \cong \frac{(S_n - S_1)^2 - 2\sigma^2}{n-1} \tag{4.9}$$

may be used to estimate δ^2.

Some estimate of σ^2 is required to use that formula, though. One method for estimating σ^2 involves what may be described as a structural analysis. Such a process involves decomposing the process/parameter risk into its components and then estimating each component separately.

The process risk is some ways better represented in historical credibility formulas (such as the $P/(P+K)$, or $U/(U+K)$ in the notation of this paper), so it will be

[30]The author is not aware of any specific measure that would readily define this, so it would likely need to be assessed judgmentally.

[31]The author recognizes that use of the normal distribution is an implicit assumption, but Boor (2012) shows it is generally a reasonable approximation in the insurance line of business context.

analyzed first. Thankfully, as long as there are enough claims in the data to reliably estimate the upper end of the severity distribution, one may use the collective risk equation to calculate the process variance (which may be labeled "α^2"). Then,

$$\alpha^2 = E[\#claims] \times Var[severity] + Var[\#claims] \times E[severity], \quad (4.10)$$

or in the loss ratio or pure premium context,

$$\alpha^2 = \frac{E[\#claims] \times Var[severity] + Var[\#claims] \times E[severity]}{(premium \text{ or } exposures)^2}. \quad (4.11)$$

So, as long as the proper data is available,[32] the process variance is readily estimable.

The other portion that must be estimated is the parameter variance, which will similarly be denoted "β^2". Note that any year-to-year variations in the trend are subsumed into δ^2. So, in most cases the only parameter-type variance that need be considered is the uncertainty in loss development to ultimate. That variance has two parts: uncertainty about what the correct expected loss development factor is; and variance of the ultimate loss in each year, as estimated using loss development, around the actual ultimate loss.

It is not hard to see that the uncertainty about the expected loss development factor can be essentially ignored per Note 5 at the beginning of this section. The variance in future loss emergence[33] on the various years requires some analysis, though. Estimating the remaining random β^2, given appropriate volume in the triangle, can be done using some fairly well established procedures. For example, a paper by Hayne (1985) details one approach. The result of this approach would be a multiplicative distribution with a mean of unity and a variance of some β^2.

Of course, it is then necessary to combine α^2 and β^2. First, α^2 should be converted to a multiplicative distribution to use with the multiplicative loss development distribution. Such a distribution would represent the ratio $\frac{expected\ loss + process\ error}{expected\ loss}$, which has a mean of one and variance $\frac{\alpha^2}{(expected\ loss)^2}$. The multiplicative combination of these two clearly independent distributions gives

$$\left(\frac{Variance\ of\ process}{parameter\ variance\ in\ geometric\ Brownian\ motion\ space}\right) = \beta^2 + \frac{\alpha^2}{(expected\ loss)^2} + \frac{\alpha^2\beta^2}{(expected\ loss)^2}. \quad (4.12)$$

So, when that is converted to a parameter in the linear model[34], one may show that

$$\alpha^2 = \log\left(\beta^2 + \frac{\alpha^2}{(expected\ loss)^2} + \frac{\alpha^2\beta^2}{(expected\ loss)^2} + 1\right). \quad (4.13)$$

Then, that estimate may be combined with equation (4.9) to obtain an estimate of δ^2.

4.5. Method 5: Estimating δ^2 using a larger dataset and σ^2 by subtraction

Just as σ^2 may be estimated using alternate approaches, δ^2 may often be estimated in isolation as well. If a larger proxy dataset (for example, the

[32]The formulas are beyond the scope of this paper, but corrections for incomplete large loss samples and corrections to provide measures of the differing severities of developed losses may be done when they are needed.

[33]This approach assumes the key volatility in loss emerge lies in what has emerged to date, rather than that the future development may be heavily random due to fortuitous late, larger losses. If the latter is the key issue, it would be more accurate to view the process, given losses of maturity "*m*," as credibility-driven process of estimating the future reported losses at maturity *m,* then apply the loss development factor to the result. In such a case, the claim counts and severity distribution used in equations (4.10) and (4.12) should just use the loss data through *m* months.

[34]Consider the value of the variance of a lognormal distribution of mean one, compared to its normal mean, distribution variance parameter.

countrywide private passenger auto experience of a major carrier when rates are being made a low volume state) is available, and that dataset has very minimal process/parameter risk, then the formula (4.8) from subsection 4.3 should produce a very high quality estimate of δ^2. Then, using equation (4.6), σ^2 may be estimated via

$$\frac{\sum_{i=1}^{n-1}(S_{i+1}-S_i)^2}{2(n-1)}-\frac{\delta^2}{2}\cong\sigma^2 \tag{4.14}$$

4.6. All or many of the above

Several methods were presented above. They all have different strengths and weaknesses. Whenever possible, it may be helpful to review the results of more than one method. Note that that the credibility formula is not a formula in σ^2 and δ^2 per se, it is actually a formula in either the ratio $K=\dfrac{\delta^2}{\alpha^2}$ or in K and B. So, when different values for σ^2 and δ^2 result from different approaches, but the ratio K is similar, the methods fundamentally agree. Also, note that what may look like large changes in K may have a very minor effect on the credibility when K is very large. Lastly, should the methods disagree it creates an opportunity to evaluate the strengths and weaknesses of each one.

Summary

The "square root" or classical credibility process has been in use for many years. Nevertheless, that method has significant a flaw in that the statistical assumptions (confidence level and failure threshold) may be chosen arbitrarily. Further, it assumes that whatever data receives the complement of credibility is stable and reliable, even when that data is, say, four years of a 20% trend rate. It is hoped that this advancement, by providing a reliable credibility process that uses minimal assumptions, will restructure the credibility processes used by casualty actuaries. Then, the profession can be comfortable that rate indications that use the resulting credibility values are as accurate as possible.

References

Boor, J.A., "Credibility Based on Accuracy," *Proceedings of the Casualty Actuarial Society* 79, 1992, pp. 166–185.

Boor, J.A., "The Complement of Credibility," *Proceedings of the Casualty Actuarial Society* 83, 1996, pp. 1–40.

Boor, J.A., "An Analytic Approach to Estimating the Required Surplus, Benchmark Profit, and Optimum Reinsurance Retention for an Insurance Enterprise Using Moments of the Severity Distribution and Key Frequency Distribution Values," *Electronic Theses, Treatises and Dissertations, Paper 4726* (2012).

Bühlmann, H., "Experience Rating and Credibility," *ASTIN Bulletin* 4, 1967, pp. 199–207.

De Vlyder, F., "Practical Credibility Theory with Emphasis on Optimal Parameter Estimation," *ASTIN Bulletin* 12, 1981, pp. 115–131.

Jones, D.A. and H.U. Gerber, "Credibility Formulas of the Updating Type," *Transactions of the Society of Actuaries* 27, 1975, pp. 31–46.

Hayne, R.M., "An Estimate of Statistical Variation in Development Factor Methods," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 25–43.

Ledolter, J., S. Klugman, and C.-S. Lee, "Credibility Models with Time-Varying Trend Components," *ASTIN Bulletin* 21, 1991, pp. 73–91.

Mahler, H.C., "An Actuarial Note on Credibility Parameters," *Proceedings of the Casualty Actuarial Society* 73, 1986, pp. 1–26.

Mahler, H.C., "Credibility with Shifting Risk Parameters, Risk Heterogeneity, and Parameter Uncertainty," *Proceedings of the Casualty Actuarial Society* 85, 1998, pp. 455–653.

Appendix A

B, *K* and Updating Credibility Under Bühlmann-Type Assumptions and Modified Bühlmann-Type Assumptions

The assumptions used above are fairly broad. δ^2 is a fairly general trend volatility. More important, all the errors that come between each raw data point S_i and the true mean for that year L_i are assumed to have (log) normal distributions with mean 0 (or 1 for the lognormal case) and variance σ^2. So, each σ^2 would logically contain some process variance, PrV^2, and some parameter-type variance due to issues such as uncertainty of how the remaining losses will develop "γ^2". Bühlmann (1967) analyzed the class credibility, noting that the process variance decreased as the amount of exposures (in the case using premium as a proxy) increased. In effect, his formula arose from the Law of Large Numbers assertion that $PrV^2 = \tau^2/U$, where U represents units[35] of exposure (such as premium, payroll, house-years, expected losses, etc.), τ^2 is innate to the line of business, combined with an assumption that no parameter-type variance is present.

When one inserts the Bühlmann assumptions into the steady-state credibility formula from equations (2.11) and (2.18), the results are

$$Z = \frac{\delta^2\left(\sqrt{1+4\frac{\sigma^2}{\delta^2}}-1\right)}{2\sigma^2} = \frac{P\delta^2\left(\sqrt{1+4\frac{\tau^2}{P\delta^2}}-1\right)}{2\tau^2}. \tag{A.1}$$

Or, if one uses Bühlmann's assumption and the notation $K = \tau^2/\delta^2$, the result is

$$Z = \frac{U\left(\sqrt{1+\frac{4K}{U}}-1\right)}{2K} = \frac{\left(\sqrt{U^2+4KU}\right)-U}{2K}, \tag{A.2}$$

where the two formulas for K appear to be about equally useful.

Alternately, whereas Bühlmann assumed that all the observation error was due purely to inadequacy of the sample size, in practice, only part of the observation is usually due to sample size.[36] In most applications[37] part of the error in observing true expected losses in prior years is due to uncertainty about the ultimate losses that will emerge (which is more like a constant λ^2 when all the data is at a single maturity). So, we would say that $\sigma^2 = \lambda^2 + (\tau^2/U)$. That would mean

$$Z = \frac{\delta^2\left(\sqrt{1+4\frac{\sigma^2}{\delta^2}}-1\right)}{2\sigma^2} = \frac{U\delta^2\left(\sqrt{1+4\frac{\left(\tau^2+U\lambda^2\right)}{U\delta^2}}-1\right)}{2\left(\tau^2+U\lambda^2\right)}. \tag{A.3}$$

Substituting in $K = \tau^2/\sigma^2$, and adding in a new constant $B = \lambda^2/\sigma^2$, we get

$$Z = \frac{U\left(\sqrt{1+4\left(\frac{\tau^2}{U\delta^2}+\frac{\lambda^2}{\delta^2}\right)}-1\right)}{2\left(\frac{\tau^2}{\delta^2}+U\frac{\lambda^2}{\delta^2}\right)} = U\frac{\left(\sqrt{1+4\left(\frac{K}{U}+B\right)}-1\right)}{2(K+UB)}. \tag{A.4}$$

Carrying the algebra a step further gives

$$Z = \frac{\left(\sqrt{U^2+4\left(KU+BU^2\right)}\right)-U}{2(K+UB)}. \tag{A.5}$$

That appears to be the most refined formula reasonably possible for this scenario.

Similar calculations may be employed on the recursive/transitional credibility formula (equations (2.10) and (2.17)) as follows. First, under the process variance only assumptions of Bühlmann

$$Z_i \cong \frac{\delta^2+Z_{i-1}\sigma^2}{\delta^2+Z_{i-1}\sigma^2+\sigma^2} = \frac{1+Z_{i-1}\frac{\tau^2}{U\delta^2}}{1+(1+Z_{i-1})\frac{\tau^2}{U\delta^2}} = \frac{U+Z_{i-1}K}{U+(1+Z_{i-1})K}. \tag{A.6}$$

(with equality holding in the linear case). Then, under the combined process and loss development variance assumptions, the comparable formula in B, K, and U is

[35]In the classic Bühlmann paper, "P" is used rather than "U". The Jones and Gerber (1975) paper already reserved "P" to represent "projections" of expected loss costs, so an alternate variable name was required. Further, in this context the rate or rate level targeted is assumed to involve a division by something representing units of exposure.

[36]It is not difficult to show that, under a similar modification, Bühlmann's class credibility formula revises from $Z = \frac{U}{U+K}$ to $Z = \frac{U}{U(1+B)+K}$.

[37]Very short tail lines would be an exception to this.

$$Z_i \cong \frac{\delta^2 + Z_{i-1}\sigma^2}{\delta^2 + Z_{i-1}\sigma^2 + \sigma^2} = \frac{1 + Z_{i-1}\left(\frac{\tau^2}{U\delta^2} + \frac{\lambda^2}{\delta^2}\right)}{1 + (1 + Z_{i-1})\left(\frac{\tau^2}{U\delta^2} + \frac{\lambda^2}{\delta^2}\right)}$$

$$= \frac{U + Z_{i-1}(K + BU)}{U + (1 + Z_{i-1})(K + BU)}. \quad \text{(A.7)}$$

Appendix B

Experience Period Weights?

Two questions that might be asked are "What if I use, say, the last five years of data in each update, with some specified set of weights applied to the years? How can I apply the Gerber-Jones formula to that problem?" and "If I have some five years of data, what is the best set of weights to use?"

Consider first the scenario where a number of years of data are weighted together for each update. The data S_j for year "j" has some weight "U_j" in the combination of the most recent "n" years that is used to update the rate. Unfortunately, as briefly discussed in the body of this paper, such a scheme will not fulfill the assumptions of the Gerber-Jones formula. Equation (2.3) states that the scheme's weighted average for iteration "i" must have a constant covariance with all successive terms. It is hopefully clear that, up to differences in loss development estimation of ultimate losses, the covariance between the new data in the "$(i + 1)^{th}$" credibility iteration using the weighted average and the "i^{th}" would be

$$Z_i^2(1 - Z_{i+1})U_nU_{n-1}Var[S_i] + \\ ,\ldots, + \{(1 - Z_i)\ldots(1 - Z_{i-n+3})Z_{i-n+2}\}^2 \\ (1 - Z_{i+2})U_2U_1Var[S_{i-n+2}]. \quad \text{(B.1)}$$

But the covariance between the "$(i + n)^{th}$" weighted average and the "i^{th}", due to a lack of common terms, would be zero. Further, just to emphasize the point, the covariance between the "$(i + 2)^{th}$" weighted average and the "i^{th}", given no loss development uncertainty at any stage, would be

$$Z_i^2(1 - Z_{i+1})(1 - Z_{i+2})U_nU_{n-1}Var[S_i] + \\ ,\ldots, + \{(1 - Z_i)\ldots(1 - Z_{i-n+4})Z_{i-n+3}\}^2 \\ (1 - Z_{i+1})(1 - Z_{i+2})U_3U_1Var[S_{i-n+3}]. \quad \text{(B.2)}$$

So, the preconditions of the Gerber-Jones formula do not hold when experience periods that overlap from rate calculation to rate calculation are used. That does, of course, not imply that there is no optimal credibility to use in such a case. It merely means that a different process for determining the credibility must be followed.

Another approach that might be posited involves recognizing that if one may choose "n" weights U_j for n years, and one may choose either a single "credibility" "S" or a formula generating a set of credibility values "S_j", perhaps one might reproduce the target credibility formula generated using equation (2.16). Unfortunately, testing of this using a spreadsheet solution routine and a target steady-state credibility suggests that the solution is always $U_n = 100\%$, $U_{j\neq n} = 0\%$, and $S_j = Z_j$. Hence, each step just updates with the single year of data S_i and the single year's credibility Z_j. This suggests that, since the Gerber-Jones is the optimum combination of the years, the practice of using sets of years that overlap from rate calculation to rate calculation is suboptimal.

As a counterpoint to the difficulties with the overlapping experience periods, the case when multiple years of experience are used, but not updated, should be discussed. Say, for example, that an actuary making rates for some line "M", which has 30 classes, makes rates for the 30 classes using the last ten years of exposure and loss data for each class. Rather than use the present rate for the complement of credibility, the actuary uses the total population mean and some appropriately generated test correction process. Essentially, the problem for each class is that of how to weight the ten (or "n") years of data in each class to produce the optimum estimate of the costs in that class. This is not on its face an issue of an updating credibility formula. But, by viewing it as updating the oldest year with the second oldest year, and so on, it may be viewed as such a problem. Since the Gerber-Jones formula is designed to produce the optimum estimate of all combinations of the data points via an updating formula, it should also produce the optimum combination of the data points.

One may begin by postulating that the initial rate is simply the first data point $S_0 = \frac{\sum_{year\ 0} losses}{\# exposures_{year\ 0}}$. In other words, in the language of the Gerber-Jones method, $P_1 = S_0$. Then, to update with the second year of data, one might compute Z_1 using formula (2.5), yielding $P_2 = Z_1 X_1 + (1 - Z_1)P_1$. However, consider that formula (2.5) is

$$Z_1 = \frac{W_1}{W_1 + V_1}. \qquad \text{(B.3)}$$

One may assume that W_1 represents the drift variance between time 0 and time 1 (hence a Markov property would assure it fulfills the Gerber-Jones criteria), and V_1 the observation error at time 1 (completing the variance structure at time 1). Under the formula from Boor (1992), the credibility with such a variance structure is the average squared prediction of the complement of credibility, divided by the sum of the prediction errors of the new data and the complement. That is a roundabout way of showing that formula (B.3) expects the error of S_0 in predicting the value underlying S_1 to simply be a matter of drift variance. That shows that the Gerber-Jones formula implicitly assumes that one begins with a "perfect" estimate of the initial costs. That is unrealistic in this scenario, so one must replace equation (2.5) with

$$Z_1 = \frac{W_1 + V_0}{W_1 + V_1 + V_0}, \qquad \text{(B.4)}$$

where V_0 is the observation error inherent in S_0.

Whether the remaining credibilities Z_i can be perfected beyond those of equation (2.4) is beyond the scope of this short analysis, but consider the general utility of Gerber-Jones formula (2.4) when beginning with equation (2.5)/(B.3). The recursive nature of the credibilities and the utility of the Gerber-Jones formula suggest that if subsequent credibilities using (2.4), are determined for $S_2, S_3, \ldots, S_9$ the resulting credibility weighted sum

$$\sum_{i=0}^{9} (1 - Z_9)(1 - Z_8)\ldots(1 - Z_{i+1}) Z_i S_i \text{ (noting } Z_0 = 1) \qquad \text{(B.5)}$$

should have properties similar to those of the Gerber-Jones method. Therefore, it should be at least a nearly optimal projection of the costs for the subject class at time $n + 1$ or $n + 2$.