AN EXAMPLE OF CREDIBILITY AND SHIFTING RISK PARAMETERS

HOWARD C. MAHLER

Abstract

In this paper, the won-lost record of baseball teams will be used to examine and illustrate credibility concepts. This illustrative example is analogous to the use of experience rating in insurance. It provides supplementary reading material for students who are studying credibility theory.

This example illustrates a situation where the phenomenon of shifting parameters over time has a very significant impact. The effects of this phenomenon are examined.

Three different criteria that can be used to select the optimal credibility are examined: least squares, limited fluctuation and Meyers/Dorweiler. In applications, one or more of these three criteria should be useful.

It is shown that the mean squared error can be written as a second order polynomial in the credibilities with the coefficients of this polynomial written in terms of the covariance structure of the data. It is then shown that linear equation(s) can be solved for the least squares credibilities in terms of the covariance structure.

The author wishes to thank Julie Jannuzzi and Gina Brewer for typing this paper.

1. INTRODUCTION

In this paper, the won-lost record of baseball teams will be used to examine and illustrate credibility concepts. This illustrative example is analogous to the use of experience rating in insurance. The mathematical details are contained in the appendices.

One purpose of this paper is to provide supplementary reading material for students who are studying credibility theory. However, this paper also contains a number of points which should prove of interest to those who are already familiar with credibility theory.

Of particular interest is the effect of shifting risk parameters over time on credibilities and experience rating. This example illustrates a situation where the phenomenon of shifting parameters over time has a very significant impact.

The general structure of the paper is to go from the simplest case to the more general. The mathematical derivations are confined to the appendices.

Section 2 briefly reviews the use of credibility in experience rating.

Section 3 describes the data sets from baseball that are used in this paper in order to illustrate the concepts of the use of credibility in experience rating.

Section 4 is an analysis of the general structure of the data. It is demonstrated that the different insureds (baseball teams) have significantly different underlying loss potentials. It is also shown that for this example a given insured's relative loss potential does shift significantly over time.

Section 5 states the problem whose solution will be illustrated. One wishes to estimate the future loss potential using a linear combination of different estimates.

Section 6 discusses simple solutions to the problem presented in Section 5.

Section 7 discusses three criteria that can be used to distinguish between solutions to the problem in Section 5.

Section 8 applies the three criteria of Section 7 to the forms of solution presented in Section 6. The results of applying the three different criteria are compared. The reduction in squared error and the impact of the delay in receiving data are both discussed.

Section 9 discusses more general solutions to the problem than those presented in Section 6.

Section 10 applies the three criteria of Section 7 to the forms of the solution presented in Section 9.

Section 11 shows equations for Least Squares Credibility that result from the covariance structure assumed.

Section 12 discusses miscellaneous subjects.

Section 13 states the author's conclusions.

2. CREDIBILITY AND EXPERIENCE RATING

Experience rating and merit rating modify an individual insured's rate above or below average. From an actuarial standpoint, the experience rating plan is using the observed loss experience of an individual insured in order to help predict the future loss experience of that insured. Usually this can be written in the form:

$$\text{New Estimate} = (\text{Data}) \times (\text{Credibility}) + (\text{Prior Estimate}) \times (\text{Complement of Credibility})$$

For most experience rating plans, the prior estimate is the class average. However, in theory the prior estimate could be a previous estimate of the loss potential of this insured relative to the class average. This paper will treat both possibilities.

2.1 Shifting Parameters Over Time

There are many features of experience rating plans that are worthy of study by actuaries. Meyers [1], Venter [2], Gillam [3], and Mahler [4] present examples of recent work. The example in this paper will deal with only one aspect, that is, how to best combine the different years of past data.

The author, in a previous paper [5], came to the following conclusion concerning this point:

"When there are shifting parameters over time, older years of data should be given substantially less credibility than more recent years of data. There may be only a minimal gain in efficiency[1] from using additional years of data."

3. THE DATA SETS

This paper will examine two very similar sets of data in order to illustrate certain features of credibility. Each set of data is the won-lost record for a league of baseball teams.[2] One set is for the so-called National League while the other is for the American League.[3] Each set of data covers the sixty years from 1901 to 1960. During this period of time each league had eight teams.

For each year, called a season in baseball, for each team, we have the losing percentage, i.e., the percentage of its games that the individual team lost.

3.1 Advantages of this Data

This example has a number of advantages not to be found using actual insurance data. First, over a very extended period of time there is a constant set of risks (teams). In insurance there are generally insureds who leave the data base and new ones that enter.

Second, the loss data over this extended period of time are readily available, accurate and final. In insurance the loss data are sometimes hard to compile or obtain and are subject to possible reporting errors and loss development.

Third, each of the teams in each year plays roughly the same number of games.[4] Thus the loss experience is generated by risks of roughly equal "size." Thus, in this example, one need not consider the dependence of credibility on size of risk.

---

[1] Meyers [11] defines the efficiency of an experience rating plan as the reduction in expected squared error accomplished by the use of the plan. The higher the efficiency the smaller the expected squared error.

[2] Appendix A gives some relevant features of the sport of baseball.

[3] These two leagues are referred to as the major leagues. They generally contain the best players in North America. The data for the two leagues are independent of each other, since no inter-league games are included in the data.

[4] Over the 60 years in question, teams usually played about 150 games per year.

4. ANALYSIS OF THE GENERAL STRUCTURE OF DATA

The loss experience[5] by risk (team) by year are given in Table 1 for the National League and Table 2 for the American League.[6]

4.1 Is There an Inherent Difference Between Teams?

The first question to be answered is whether there is any real difference between the experience of the different teams, or is the apparent difference just due to random fluctuations. This is the fundamental question when considering the application of experience rating.

It requires only an elementary analysis in order to show that there is a non-random difference between the teams. The average experience for each team over the whole period of time differs significantly from that of the other teams. If the experience for each team were drawn from the same probability distribution, the results for each team would be much more similar. The standard deviation in losing percentage over a sample of about 9000 games[7] would be .5%.[8] Thus if all the teams' results were drawn from the same distribution, approximately 95% of the teams would have an average losing percentage between 49% and 51%.[9]

The actual results are shown on Table 3. In fact, only 3 of 16 teams have losing percentages in that range. The largest deviation from the grand mean is 15 times the expected standard deviation if the teams all had the same underlying probability distribution.

---

[5] For each of 60 years, the percentage of games lost is given for each team. The data are from *The Sports Encyclopedia* [6].

[6] For the National League the teams are in order: Brooklyn, Boston, Chicago, Cincinnati, New York, Philadelphia, Pittsburgh and St. Louis. For the American League the teams are in order: Boston, Chicago, Cleveland, Detroit, New York, Philadelphia, St. Louis and Washington. In both cases, the city given is that in which the team spent the majority of the data period.

[7] About 150 games for a team each year times 60 years.

[8] A binomial distribution with a 50% chance of losing, for 9000 games, has a variance of $9000(1/2)(1 - 1/2) = 2250$. This is a standard deviation of 47 games lost, or $47 \div 9000 = .5\%$ in losing percentage.

[9] Using the standard normal approximation, 95% of the probability is within two standard deviatioins of the mean which in this case is 50%.

TABLE 1

National League Losing Percentages

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1901	.500	.419	.619	.626	.620	.407	.353	.457
1902	.467	.457	.504	.500	.647	.591	.259	.582
1903	.580	.485	.406	.468	.396	.637	.350	.686
1904	.641	.634	.392	.425	.307	.658	.431	.513
1905	.669	.684	.399	.484	.314	.454	.373	.623
1906	.675	.566	.237	.576	.368	.536	.392	.653
1907	.608	.561	.296	.569	.464	.435	.409	.660
1908	.591	.656	.357	.526	.364	.461	.364	.682
1909	.706	.641	.320	.497	.399	.516	.276	.645
1910	.654	.584	.325	.513	.409	.490	.438	.588
1911	.709	.573	.403	.542	.353	.480	.448	.497
1912	.660	.621	.393	.510	.318	.520	.384	.588
1913	.543	.564	.425	.582	.336	.417	.477	.660
1914	.386	.513	.494	.610	.455	.519	.552	.471
1915	.454	.474	.523	.539	.546	.408	.526	.529
1916	.414	.390	.562	.608	.434	.405	.578	.608
1917	.529	.536	.519	.494	.364	.428	.669	.461
1918	.573	.548	.349	.469	.427	.553	.480	.605
1919	.590	.507	.464	.314	.379	.657	.489	.606
1920	.592	.396	.513	.464	.442	.595	.487	.513
1921	.484	.493	.582	.542	.386	.669	.412	.431
1922	.654	.506	.481	.442	.396	.627	.448	.448
1923	.649	.506	.461	.409	.379	.675	.435	.484
1924	.654	.403	.471	.458	.392	.636	.412	.578
1925	.542	.556	.558	.477	.434	.556	.379	.497
1926	.566	.536	.468	.435	.510	.616	.451	.422
1927	.610	.575	.444	.510	.403	.669	.390	.399
1928	.673	.497	.409	.487	.396	.717	.441	.383
1929	.636	.542	.355	.571	.444	.536	.425	.487
1930	.545	.442	.416	.617	.435	.662	.481	.403

TABLE 1

(CONTINUED)

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1931	.584	.480	.455	.623	.428	.571	.513	.344
1932	.500	.474	.416	.610	.532	.494	.442	.532
1933	.461	.575	.442	.618	.401	.605	.435	.464
1934	.483	.533	.430	.656	.392	.624	.507	.379
1935	.752	.542	.351	.556	.405	.582	.438	.377
1936	.539	.565	.435	.519	.403	.649	.455	.435
1937	.480	.595	.396	.636	.375	.601	.442	.474
1938	.493	.537	.414	.453	.447	.700	.427	.530
1939	.583	.451	.455	.370	.490	.702	.556	.399
1940	.572	.425	.513	.346	.526	.673	.494	.451
1941	.597	.351	.545	.429	.516	.721	.474	.366
1942	.601	.325	.558	.500	.441	.722	.551	.312
1943	.556	.471	.516	.435	.641	.584	.481	.318
1944	.578	.591	.513	.422	.565	.601	.412	.318
1945	.559	.435	.364	.604	.487	.701	.468	.383
1946	.385	.471	.464	.565	.604	.552	.591	.372
1947	.390	.442	.552	.526	.474	.597	.597	.422
1948	.455	.405	.584	.582	.494	.571	.461	.448
1949	.370	.513	.604	.597	.526	.474	.539	.377
1950	.422	.461	.582	.569	.442	.409	.627	.490
1951	.382	.506	.597	.558	.376	.526	.584	.474
1952	.373	.582	.500	.552	.403	.435	.727	.429
1953	.318	.403	.578	.558	.545	.461	.675	.461
1954	.403	.422	.584	.519	.370	.513	.656	.532
1955	.359	.448	.529	.513	.481	.500	.610	.558
1956	.396	.403	.610	.409	.565	.539	.571	.506
1957	.455	.383	.597	.481	.552	.500	.597	.435
1958	.539	.403	.532	.506	.481	.552	.455	.532
1959	.436	.449	.519	.519	.461	.584	.494	.539
1960	.468	.429	.610	.565	.487	.617	.383	.442

TABLE 2

American League Losing Percentages

	AL1	AL2	AL3	AL4	AL5	AL6	AL7	AL8
1901	.419	.390	.599	.452	.489	.456	.650	.545
1902	.438	.448	.493	.615	.638	.390	.426	.551
1903	.341	.562	.450	.522	.463	.444	.532	.686
1904	.383	.422	.430	.592	.391	.464	.572	.748
1905	.487	.395	.506	.484	.523	.378	.647	.576
1906	.682	.384	.418	.523	.404	.462	.490	.633
1907	.604	.424	.441	.387	.527	.393	.546	.675
1908	.513	.421	.416	.412	.669	.556	.454	.559
1909	.417	.487	.536	.355	.510	.379	.593	.724
1910	.471	.556	.533	.442	.417	.320	.695	.563
1911	.490	.490	.477	.422	.500	.331	.704	.584
1912	.309	.494	.510	.549	.671	.408	.656	.401
1913	.473	.487	.434	.569	.623	.373	.627	.416
1914	.405	.545	.667	.477	.545	.349	.536	.474
1915	.331	.396	.625	.351	.546	.717	.591	.444
1916	.409	.422	.500	.435	.481	.765	.487	.503
1917	.408	.351	.429	.490	.536	.641	.630	.516
1918	.405	.540	.425	.563	.512	.594	.525	.437
1919	.518	.371	.396	.429	.424	.743	.518	.600
1920	.529	.377	.364	.604	.383	.688	.503	.553
1921	.513	.597	.390	.536	.359	.654	.474	.477
1922	.604	.500	.494	.487	.390	.578	.396	.552
1923	.599	.552	.464	.461	.355	.546	.513	.510
1924	.565	.569	.562	.442	.414	.533	.513	.403
1925	.691	.487	.545	.474	.552	.421	.464	.364
1926	.699	.471	.429	.487	.409	.447	.597	.460
1927	.669	.542	.569	.464	.286	.409	.614	.448
1928	.627	.532	.597	.558	.344	.359	.468	.513
1929	.623	.612	.467	.545	.429	.307	.480	.533
1930	.662	.597	.474	.513	.442	.338	.584	.390

TABLE 2

(CONTINUED)

	AL1	AL2	AL3	AL4	AL5	AL6	AL7	AL8
1931	.592	.634	.494	.604	.386	.296	.591	.403
1932	.721	.675	.428	.497	.305	.390	.591	.396
1933	.577	.553	.503	.513	.393	.477	.636	.349
1934	.500	.651	.448	.344	.390	.547	.559	.566
1935	.490	.513	.464	.384	.403	.611	.572	.562
1936	.519	.464	.481	.461	.333	.654	.625	.464
1937	.474	.442	.461	.422	.338	.642	.701	.523
1938	.409	.561	.434	.455	.349	.651	.638	.503
1939	.411	.448	.435	.474	.298	.638	.721	.572
1940	.468	.468	.422	.416	.429	.649	.565	.584
1941	.455	.500	.513	.513	.344	.584	.545	.545
1942	.388	.554	.513	.526	.331	.643	.457	.589
1943	.553	.468	.464	.494	.364	.682	.526	.451
1944	.500	.539	.532	.429	.461	.532	.422	.584
1945	.539	.523	.497	.425	.467	.653	.464	.435
1946	.325	.519	.558	.403	.435	.682	.571	.506
1947	.461	.545	.481	.448	.370	.494	.617	.584
1948	.381	.664	.374	.494	.390	.455	.614	.634
1949	.377	.591	.422	.435	.370	.474	.656	.675
1950	.390	.610	.403	.383	.364	.662	.623	.565
1951	.435	.474	.396	.526	.364	.545	.662	.597
1952	.506	.474	.396	.675	.383	.487	.584	.494
1953	.451	.422	.403	.610	.344	.617	.649	.500
1954	.552	.390	.279	.558	.331	.669	.649	.571
1955	.455	.409	.396	.487	.377	.591	.630	.656
1956	.455	.448	.429	.468	.370	.662	.552	.617
1957	.468	.416	.503	.494	.364	.614	.500	.643
1958	.487	.468	.497	.500	.403	.526	.516	.604
1959	.513	.390	.422	.506	.487	.571	.519	.591
1960	.578	.435	.506	.539	.370	.623	.422	.526

TABLE 3

AVERAGE LOSING PERCENTAGES (1901-1960)

Risk (Team)	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
National League	53.4	49.9	47.3	51.8	44.7	56.5	47.8	48.8

Risk (Team)	AL1	AL2	AL3	AL4	AL5	AL6	AL7	AL8
American League	49.5	49.4	47.0	48.5	42.6	52.9	56.4	53.5

Thus there can be no doubt that the teams actually differ.[10] It is therefore a meaningful question to ask whether a given team is better or worse than average.

A team that has been worse than average over one period of time is more likely to be worse than average over another period of time. If this were not true, we would not have found the statistically significant difference in the means of the teams.

Thus if we wish to predict the future experience of a team, there is useful information contained in the past experience of that team. In other words, there is an advantage to experience rating.

4.2 Shifting Parameters Over Time

A similar, but somewhat different question of interest is whether for a given team the results for the different years are from the same distribution (or nearly the same distribution). In other words, are the observed different results over time due to more than random fluctuation? The answer is yes. This is a situation where the underlying parameters of the risk process shift over time.

[10] The situation here is somewhat complicated by the fact that one team's loss is another team's win. Thus the won-loss records of seven teams determine that of the remaining team. However, the author confirmed with a straightforward simulation that in this case this phenomenon would not affect the conclusion. For 8 teams each with the 50% loss rate playing 9000 games each, in 32 out of 600 cases (5%) a team had a winning percentage lower than 49% or more than 51%. In none of the 600 cases did a team have a winning percentage as low as 48% or as high as 52%.

As discussed in Section 2.1, the extent to which risk parameters shift over time has an important impact on the use of past insurance data to predict the future.

Whether the risk parameters shift over time can be tested in many ways. Two methods will be demonstrated here. These methods can be applied to insurance data as well as the data presented here.

The first method of testing whether parameters shift over time uses the standard chi-squared test. For each risk, one averages the results over separate 5 year periods.[11] Then one compares the number of games lost during the various 5 year periods. One can then determine by applying the chi-squared test that the risk process could not have the same mean over this entire period of time. The results shown in Table 4 are conclusive for every single risk. Even the most consistent risk had significant shifts over time.

In the second method of testing whether parameters shift over time, one computes the correlation between the results for all of the risks for pairs of years. Then one computes the average correlation for those pairs of years with a given difference in time. Finally, one examines how the average correlation depends on this difference. The results in our case are displayed in Table 5.

Observed values of the correlation different from zero are not necessarily statistically significant. For this example, a 95% confidence interval around zero for the correlation is approximately plus or minus .10.[12] Thus, for this example, the correlation decreases as the difference in time increases until about ten years when there is no longer a significant correlation between results.[13]

[11] The data were grouped in five year intervals for convenience. Other intervals could also have been used.

[12] For larger distances between the years, we have fewer observations to average, so the confidence interval expands to approximately plus or minus .12. The confidence intervals were determined via repeated simulation in which the actual data for each year were separately assigned to the individual risks at random; thus for the simulated data any observed correlation is illusory.

[13] For a difference of between 15 and 20 years there is again a small but significant positive correlation. The author has no explanation for this long term cycle.

TABLE 4

RESULTS OF CHI-SQUARED TEST OF SHIFTING PARAMETERS OVER TIME

For each risk (team) its experience over the 60 year period was averaged into 12 five-year segments. (The simplifying assumption was made of 150 games each year; this did not affect the results.) Then for each risk separately, the chi-square statistic was computed in order to test the hypothesis that each of the five year segments was drawn from a distribution with the same mean. The resulting chi-square values are:

NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
107	45	98	35	39	73	114	119
AL1	**AL2**	**AL3**	**AL4**	**AL5**	**AL6**	**AL7**	**AL8**
114	69	34	30	97	162	53	65

For example, for the risk (team) NL2 the data by five-year segments are as follows:

	'01 – '05	'06 – '10	'11 – '15	'16 – '20	'21 – '25	'26 – '30	'31 – '35	'36 – '40	'41 – '45	'46 – '50	'51 – '55	'56 – '60
(1) Games Lost*	402	451	412	357	370	389	391	386	326	344	354	310
(2) Expected Games Lost**	374	374	374	374	374	374	374	374	374	374	374	374
(3)=[(1)−(2)]^2/(2)	2.1	15.9	3.9	.8	0	.6	.8	.4	6.2	2.4	1.1	11.0

The sum of row (3) is 45, which is the chi-square value for this risk.

For each risk there is less than a .2% chance that the different five-year segments were drawn from distributions with the same mean.*** Thus we reject the hypothesis that the means are the same over time; we accept the hypothesis of shifting risk parameters over time.

*Assuming 150 games per year, and the observed losing percentage for the five year segment.

**Assuming 150 games per year, and the observed losing percentage for the whole 60 years.

***For 11 degrees of freedom, there is a .16% chance of having a chi-square value of 30 or more. There is a .004% chance of having a chi-square value of 40 or more.

TABLE 5

AVERAGE CORRELATIONS OF RISKS EXPERIENCE OVER TIME (1901-1960)

Difference Between Pairs of Years of Experience	Correlation	
	NL	AL
1	.651	.633
2	.498	.513
3	.448	.438
4	.386	.360
5	.312	.265
6	.269	.228
7	.221	.157
8	.190	.124
9	.135	.078
10	.100	.090
11	.083	.058
12	.103	.063
13	.154	.101
14	.176	.104
15	.180	.141
16	.246	.178
17	.278	.166
18	.219	.198
19	.176	.219
20	.136	.225
21	.090	.159
22	.065	.125
23	.055	.093
24	.004	.048
25	−.024	.006

TABLE 5

(CONTINUED)

Difference Between Pairs of Years of Experience	Correlation	
	NL	AL
26	−.028	.010
27	−.095	−.002
28	−.128	−.013
29	−.107	−.032
30	−.062	.006
31	−.061	−.019
32	−.028	.027
33	−.015	.002
34	.017	.088
35	.038	.143
36	−.014	.156
37	−.024	.214
38	−.012	.238
39	−.017	.138
40	−.095	.093
41	−.174	.055
42	−.216	.028
43	−.332	−.043
44	−.423	−.018
45	−.363	−.035
46	−.332	.066
47	−.324	.069
48	−.373	.136
49	−.423	.075
50	−.475	.145

The correlation between years that are close together is significantly greater than those further apart. This implies that the parameters of the risk process are shifting significantly over time. If the parameters were reasonably constant over time, the correlations would not depend on the length of time between the pair of years.

On the other hand, there is a significant correlation between the results of years close in time. Thus recent years can be usefully employed to predict the future.

5. STATEMENT OF THE PROBLEM

Let X be the quantity we wish to estimate. In this case, X is the expected losing percentage for a risk.

Let Y_1, Y_2, Y_3, etc., be various estimates for X. Then one might estimate X by taking a weighted average of the different estimates Y_i.

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

where X = quantity to be estimated,
Y_i = one of the estimates of X,
Z_i = weight assigned to estimate Y_i of X.

Here only linear combinations of estimators are being considered. In addition, the estimators themselves will be restricted to a single year of past data for the given risk or to the grand mean (which is 50% in this case).[14] No subjective information or additional data beyond the past annual losing percentages will be used.[15] In other words, this is a situation analogous to (prospective) experience rating. This is not a situation analogous to schedule rating.

[14] In other words, in this case, Y either equals the observed losing percentage for the risk in one year or equals the grand mean of 50%. Credibility methods can be applied to more general estimators.

[15] The use of information on the retirement of players or acquisition of new players might enable a significant increase in the accuracy of the estimate. The breakdown of the data into smaller units than an entire year might enable a significant increase in the accuracy of the estimate.

The problem to be considered here is what weights Z_i produce the "best" estimate of future losing percentage. In order to answer that question, criteria will have to be developed that allow one to compare the performance of the different methods to determine which is better. In the example being dealt with in this paper, it is easy to get unbiased estimators. Since all of the estimators being compared will be unbiased, the question of which method is better will focus on other features of the estimators.

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

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

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

6. SIMPLE SOLUTIONS TO THE PROBLEM

In this section, various relatively simple solutions to the problem will be presented.

6.1 Every Risk is Average

The first method is to predict that the future losing percentage for each risk will be equal to the overall mean of 50%. This method ignores all the past data; i.e., the past data are given zero credibility. While this is not a serious candidate for an estimation method in the particular example examined in this paper, it is a useful base case in general.

6.2 The Most Recent Year Repeats

The second method is to predict that the most recent past year's losing percentage for each risk will repeat. This is what is meant by giving the most recent year of data 100% credibility.

6.3 Credibility Weight the Most Recent Year and the Grand Mean

In the third method, one gives the most recent year of data for each risk weight Z, and gives the grand mean, which in this case is 50%, weight $1 - Z$.

When $Z = 0$, one gets the first method; when $Z = 1$, one gets the second method. Since each of these is a special case of this more general method, by the proper choice of Z one can do better than or equal to either of the two previous methods. This is an important and completely general result. It does not depend on either the criterion that is used to compare methods or the means of deciding which value of Z to use.

6.4 Determining the Credibility

When employing the third method, the obvious question is how does one determine the value of credibility to use. Ideally one would desire a theory or method that would be generally applicable, rather than one that only worked for a single example. There have been many fine papers on this subject in the actuarial literature.

Generally, the credibility considered "best" is determined by some objective criterion. This will be discussed later.

Using either Bühlmann/Bayesian credibility methods or classical/limited fluctuation credibility methods, one determines which credibility will be expected to optimize the selected criterion in the future. One can also empirically investigate which credibility would have optimized the selected criterion if it had been used in the past; i.e., one can perform retrospective tests. This will be discussed in more detail later.

6.5 Equal Weight to the N *Most Recent Years of Data*

In the fourth method, one gives equal weight to the N most recent years of data for each risk, and gives the grand mean, which in this case is 50%, weight $1 - Z$. This method gives each of the N most recent

years weight of Z/N.[16] When $N = 1$ this reduces to the previous method. Thus this method will perform at least as well as the previous method, with the proper choices of N and Z.

7. CRITERIA TO DECIDE BETWEEN SOLUTIONS

In this section, we will discuss *three* criteria that can be used to distinguish between solutions. These criteria can be applied in general and not just to this example.

7.1 Least Squared Error

The first criterion involves calculating the mean squared error of the prediction produced by a given solution compared to the actual observed result. The smaller the mean squared error, the better the solution.

The Bühlmann/Bayesian credibility methods attempt to minimize the squared error; i.e., they are least squares methods. Minimizing the squared error is the same as minimizing the mean squared error.

7.2 Small Chance of Large Errors

The second criterion deals with the probability that the observed result will be more than a certain percent different than the predicted result. The less this probability, the better the solution.

This is related to the basic concept behind "classical" credibility which has also been called "limited fluctuation" credibility [7]. In classical credibility, the full credibility criterion is chosen so that there is a probability, P, of meeting the test that the maximum departure from expected is no more than k percent.

The reason the criterion is stated in this way rather than the way it is in classical credibility is that, unlike the actual observations, one cannot observe directly the inherent loss potential.[17] However, the two concepts are closely related, as discussed in Appendix G.

---

[16] In later methods, the weights given to the different years of data will be allowed to differ from each other.

[17] It has been shown that the loss potential varies for a risk over time. Thus it cannot be estimated as the average of many observations over time.

7.3 Meyers/Dorweiler

The third criterion has been taken from Glenn Meyers' paper [1]. Meyers in turn based his criterion upon the ideas of Paul Dorweiler [8].

This criterion involves calculating the correlation between two quantities. The first quantity is the ratio of actual losing percentage to the predicted losing percentage. The second quantity is the ratio of the predicted losing percentage to the overall average losing percentage. The smaller the correlation between these two quantities, i.e., the closer the correlation is to zero, the better the solution.

To compute the correlation, the Kendall τ statistic is used.[18] This is explained in detail in Appendix B. The relation of this criterion as used here and as it is used by Meyers to examine experience rating is also discussed in that appendix.

8. THE CRITERIA APPLIED TO THE SIMPLE SOLUTIONS

In this section the three criteria in Section 7 will be applied to the simple solutions given in Section 6. More knowledgeable readers may wish to skip to Section 8.4 which compares the results of applying the three different criteria. Section 8.5 discusses the reduction in squared error. Section 8.6 examines the impact of a delay in receiving data.

8.1 The Two Base Cases

The two simplest solutions either always use as the estimate the overall mean ($Z = 0$), or always use as the estimate the most recent observation ($Z = 1$). While neither of these solutions is expected to be chosen, they serve as the base cases for testing the other solutions.

[18] Meyers in [1] used the Kendall τ statistic. In the example here, any other reasonable measure of the correlation could be substituted.

The first criterion is the smallest mean squared error. For the two data sets the results are:

	Mean Squared Error	
	NL	AL
$Z = 0$	.0091	.0095
$Z = 1$	.0059	.0068

The second criterion is to produce a small probability of being wrong by more than k percent. For the two data sets the results are as follows:

	Percent of time that the estimate is in error by more than 5%		Percent of time that the estimate is in error by more than 10%		Percent of time that the estimate is in error by more than 20%	
	NL	AL	NL	AL	NL	AL
$Z = 0$	82.2%	80.3%	64.8%	63.8%	29.0%	31.4%
$Z = 1$	75.8%	72.9%	52.3%	55.7%	19.1%	22.0%

The third criterion is to have a correlation as close to zero as possible between the ratio of the actual to estimated and the ratio of estimated to the overall mean. For the two data sets the results are as follows:

	Correlation (Kendall τ)	
	NL	AL
$Z = 0$*	.48	.46
$Z = 1$	−.24	−.27

* Limit as Z approaches zero.

8.2 Applying Credibility to the Latest Year of Data

The third prediction method, explained in Section 6.3, uses credibility to combine the latest year of data and the grand mean. The mean squared error depends on the credibility. As shown in Table 6, the mean squared error is a minimum for Z between 60% and 70%.[19] The probability of having errors of 20% or more is displayed in Table 7. Based on this second criterion, the optimal Z is between 50% and 80%.[20] This criterion does not distinguish very sharply between the different values of credibility.

The correlations used in the third criterion are displayed in Table 8. Based on the third criterion the optimal Z is approximately 70%.[21]

8.3 Applying Credibility to the Latest N *Years of Data*

The fourth method, explained in Section 6.4, uses credibility to combine the grand mean with the latest N years of data (giving each year of data the same weight.)

The results of applying the first criterion are shown in Table 6. Based on most actuarial uses of credibility, an actuary would expect the optimal credibilities to increase as more years of data are used. In this example they do not. In fact, using more than one or two years of data does an inferior job according to this criterion.

This result is to be expected, since the parameters shift substantially over time. Thus the use of older data (with equal weight) eventually leads to a worse estimate.[22]

---

[19] For the NL data set, the minimum occurs when $Z = 68\%$. For the AL data set, the minimum occurs when $Z = 66\%$. Also, it should be noted that the squared errors for $Z = 0$ vary somewhat with the number of years of data used, solely due to the differing periods of time over which the test can be performed.

[20] For the NL data set, the optimal Z is 75%. For the AL data set, the optimal Z is 55%. It should be noted that, given the limited number of observations, two values of Z can produce identical results for this criterion.

[21] For the NL data set, the correlation is closest to zero for $Z = 71\%$. For the AL data set, the correlation is closest to zero for $Z = 66\%$.

[22] The number of years of data to use to get the best estimate will depend on the particular example. This general subject was explored in Mahler [5].

TABLE 6

MEAN SQUARED ERROR (.0001)

NL

Z	N = 1	N = 2	N = 3	N = 4	N = 5	N = 7	N = 10	N = 15	N = 20	N = 25
0	91	90	90	89	87	84	80	80	80	80
.10	80	80	80	80	79	77	74	75	76	77
.20	70	72	72	72	71	71	70	72	72	74
.30	62	65	65	65	65	66	67	69	69	71
.40	56	59	60	60	60	62	64	67	67	69
.50	52	55	56	56	57	59	63	66	65	68
.60	50	53	53	53	55	57	62	65	64	68
.70	49	52	52	52	53	56	63	65	63	68
.80	51	53	52	52	54	57	64	66	64	69
.90	54	55	53	53	55	58	66	68	65	70
1.00	59	59	56	56	57	61	70	70	66	72

AL

0	95	96	96	95	95	95	95	92	91	95
.10	84	85	86	86	88	89	90	88	87	91
.20	75	77	78	79	81	83	86	85	83	87
.30	67	69	71	73	75	79	82	82	80	83
.40	61	64	66	68	71	75	80	81	78	80
.50	58	60	62	64	68	73	78	79	76	78
.60	56	57	60	62	66	71	78	79	74	76
.70	56	56	59	61	66	71	78	79	73	74
.80	58	57	59	62	66	72	79	79	73	73
.90	62	59	61	64	68	74	81	81	73	73
1.00	68	63	64	67	71	77	84	83	74	73

TABLE 7

PERCENT OF TIME THAT THE ESTIMATE IS IN ERROR BY MORE THAN 20%

Z	N = 1	N = 2	N = 3	N = 4	N = 5	N = 7	N = 10	N = 15	N = 20	N = 25
					NL					
0	29	29	29	28	28	27	25	25	25	26
.10	25	26	25	25	25	25	25	24	24	25
.20	23	23	23	23	24	24	23	24	23	26
.30	19	21	22	22	21	22	23	23	23	24
.40	18	19	20	20	22	23	22	22	23	25
.50	17	19	18	19	21	21	21	22	24	25
.60	17	18	18	18	18	21	21	22	24	25
.70	17	18	18	19	19	21	21	22	26	26
.80	17	17	18	19	20	22	23	24	25	27
.90	18	19	17	18	21	22	24	25	25	27
1.00	19	20	18	20	21	23	25	26	25	28
					AL					
0	31	31	32	32	32	32	33	32	32	34
.10	27	27	27	27	27	28	28	28	29	30
.20	23	24	25	25	25	27	27	27	28	30
.30	21	21	22	23	24	25	27	27	27	29
.40	19	19	21	23	24	26	27	26	26	27
.50	18	19	21	22	23	24	28	26	26	27
.60	18	18	21	21	22	25	27	25	26	26
.70	20	18	19	20	22	25	26	25	25	27
.80	19	19	18	20	21	26	27	26	25	26
.90	20	21	19	22	23	27	26	27	25	27
1.00	22	23	22	24	26	28	29	28	27	26

TABLE 8

CORRELATION (KENDALL τ)

NL

Z	N = 1	N = 2	N = 3	N = 4	N = 5	N = 7	N = 10	N = 15	N = 20	N = 25
0*	.48	.45	.46	.45	.43	.39	.31	.28	.29	.25
.10	.44	.41	.42	.41	.40	.35	.27	.24	.26	.22
.20	.38	.36	.37	.36	.35	.30	.23	.20	.22	.18
.30	.32	.30	.31	.31	.30	.25	.18	.16	.18	.14
.40	.25	.24	.25	.25	.24	.20	.12	.11	.14	.10
.50	.17	.17	.19	.19	.18	.14	.07	.06	.10	.06
.60	.09	.09	.12	.12	.12	.08	.02	.02	.05	.02
.70	.01	.02	.04	.05	.05	.02	−.04	−.03	.01	−.02
.80	−.08	−.06	−.03	−.02	−.02	−.04	−.09	−.07	−.03	−.06
.90	−.16	−.13	−.01	−.09	−.08	−.10	−.14	−.12	−.08	−.10
1.00	−.24	−.21	−.17	−.16	−.15	−.16	−.19	−.16	−.12	−.14

*Limit as Z approaches zero

CORRELATION (KENDALL τ)

AL

Z	N = 1	N = 2	N = 3	N = 4	N = 5	N = 7	N = 10	N = 15	N = 20	N = 25
0*	.46	.45	.44	.41	.38	.34	.28	.24	.27	.30
.10	.42	.41	.40	.38	.35	.30	.25	.21	.25	.28
.20	.36	.36	.35	.33	.30	.26	.21	.17	.21	.25
.30	.29	.30	.30	.27	.25	.21	.16	.13	.18	.22
.40	.22	.24	.23	.22	.19	.16	.12	.09	.15	.19
.50	.14	.16	.17	.15	.13	.11	.07	.05	.11	.16
.60	.05	.08	.10	.08	.07	.05	.02	.01	.07	.12
.70	−.03	.00	.02	.02	.00	−.01	−.03	−.03	.03	.09
.80	−.11	−.07	−.05	−.05	−.06	−.07	−.07	−.07	−.01	.05
.90	−.19	−.15	−.12	−.12	−.12	−.12	−.12	−.11	−.05	.01
1.00	−.27	−.22	−.19	−.18	−.18	−.17	−.16	−.15	−.09	−.02

*Limit as Z approaches zero

The results of applying the second criterion are displayed in Table 7. This criterion does not sharply distinguish between the different values of credibility. There is a broad range of credibilities all of which do reasonably well.[23] This is particularly true for larger values of N. Again the use of more years of data eventually leads to an inferior estimate.

The results of applying the third criterion are displayed in Table 8. Again the optimal credibility does not increase as N increases. Unlike the other criteria, the third criterion cannot be used to distinguish between values of N. For each N, there is a Z, such that the correlation is zero. Thus each value of N performs as well as all the others.

Meyers points out that the distribution of Kendall's τ can be used to obtain a confidence interval for the credibility. As explained in Appendix B, for this example a 95% confidence interval for τ around zero has a radius of about .07.

For example, using 10 years of data, the optimal credibility using the Meyers/Dorweiler criterion for the NL set of data is 63%. However, this point estimate for the credibility is actually an estimate of an interval of credibilities that correspond to τ between plus and minus .07. The optimal credibility is $63\% \pm 13\%$.[24]

8.4 Comparison of the Results of the Three Criteria

In Table 9 the optimal credibilities are displayed as determined by the three criteria for various values of N. Note that the listed values of credibility are those that happened to work best over the period of time observed. Values close to these values would also work well over this period of time.

One should think of the point estimates listed in Table 9 as the centers of interval estimates. This is illustrated when one compares the different estimates obtained by analyzing the NL and AL data sets. There is no inherent difference in the two data sets. Thus one would expect the credibilities from the two analyses to be the same. They are similar,

[23] This is true to a lesser extent for the first criterion. This subject is explored in Mahler [9].

[24] For $Z = 63.3\%$, $\tau = 0$. For $Z = 50.1\%$, $\tau = .07$. For $Z = 76.4\%$, $\tau = -.07$.

but far from identical. This indicates that the peculiarities of the specific observed values are sufficient to affect the answers somewhat. There is some lack of precision in the estimates in Table 9.

TABLE 9

OPTIMAL CREDIBILITY

Number of Years of Data Used	NL			AL		
	Criterion #1	Criterion #2	Criterion #3	Criterion #1	Criterion #2	Criterion #3
1	68%	75%	71%	65%	55%	66%
2	71	80	72	70	56	70
3	74	87	76	72	77	73
4	76	57	77	72	69	72
5	74	61	77	70	70	71
7	71	64	73	67	51	68
10	60	49	63	62	70	64
15	63	43	64	65	69	62
20	71	40	73	81	82	77
25	64	30	64	97	61	94

Criterion #1: Least Squares (Section 7.1)
Criterion #2: Small Chance of Large Errors (Section 7.2)
Criterion #3: Meyers/Dorweiler (Section 7.3)

This can be illustrated further by reversing the time arrow and analyzing the data sets going backwards in time rather than forwards. For example, one could use data from years 1902 to 1911 to "predict" 1901. This analysis is equally valid for determining optimal credibilities in this example as was the original anlysis.

For $N = 10$, one gets the following optimal credibilities for the different data sets, where NLR and ALR represent respectively the NL and AL data sets reversed in time.

	Optimal Credibilities (N = 10)				
	NLR	NL	ALR	AL	Average
Criterion #1	72%	60%	57%	62%	63%
Criterion #2	58	49	42	70	55
Criterion #3	77	63	58	64	65

The optimal credibilities differ between the four data sets. The amount of variation provides some idea of the imprecision of the different estimates. While the optimal credibilities differ between the three criteria, the differences do not appear to be sufficiently large to allow one to draw any definitive conclusions.

In this case, the use of any value of credibility between 50% and 70% would perform reasonably well according to all three criteria for all four data sets. As a practical matter, the difference in the predictions will not vary that much depending on which value of credibility is chosen in that range.[25]

In most applications of credibility, values for the credibility that differ somewhat from optimal perform reasonably well and the choice between these values has a relatively small practical impact.

8.5 Putting the Reduction in Squared Error in Context

The first criterion used to determine the optimal credibility is to minimize the squared error. Using the optimal credibility based on this criterion will reduce the squared error between the observed and predicted result. What should be considered a significant reduction in squared error?

[25] The maximum difference in any prediction for N = 10 between using 50% and 70% credibility is 3.3% in the losing percentage. In most cases it is much smaller. On average it would make about a 1% difference.

Let us examine an example. For the NL data, using one year of data, the optimal credibility is 68% as shown in Table 9. As shown in Table 6 the mean squared errors are:

Z	Mean Squared Error
0	.0091
68%	.0049
100%	.0059

In this case, by the use of credibility, the squared error has been reduced from .0059 if the data were relied upon totally, or .0091 if the data were totally ignored, to .0049. In this case, the squared error has been reduced to 83% (.0049/.0059) of its previous value.[26]

As discussed in Appendix E, in the current case, the best that can be done using credibility to combine two estimates is to reduce the mean squared error between the estimated and observed values to 75% of the minimum of the squared errors from either relying solely on the data or ignoring the data.[27]

The reduction of the squared error to 83% of its previous value appears significant in light of the maximum possible reduction to 75%.[28]

8.6 Effect of Delay in Receiving Data

It has been shown previously for the data set examined in this paper that the further apart in time two years are, the lower the correlation between them. Thus if there is a delay before the data are available for use in experience rating, the resulting estimate of the future will be less accurate.

---

[26] The "previous" value of the squared error is considered to be the minimum of the squared errors that result from either ignoring the data entirely or relying on the data entirely.

[27] When using more than two or more years of data, the reduction in squared error depends on the impact of shifting parameters over time. However, in the absence of shifting parameters over time, for N years with the same weight applied to each year, the maximum possible reduction is $1 \div (2(N + 1))$.

[28] The maximum reduction is possible when the squared errors for $Z = 0$ and $Z = 1$ are equal.

As is shown in Table 10, as the delay increases, the squared error increases significantly. The increase in squared error is particularly significant as one goes from a situation of having the data from the most recent year available to predict the coming year to a situation of having

TABLE 10

MINIMUM SQUARED ERROR (.0001)

Time Between Latest Data Point and Future Prediction	NL				
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
1	49	52	51	51	53
2	66	62	60	60	60
3	69	66	65	64	65
4	73	71	69	69	70
5	77	73	73	72	72
6	76	75	75	73	74
7	78	77	75	75	75
8	79	77	77	76	75
9	78	78	77	76	75
10	78	78	76	75	75

Time Between Latest Data Point and Future Prediction	AL				
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
1	56	56	59	61	66
2	71	70	71	74	76
3	78	77	80	81	83
4	83	85	85	87	88
5	89	89	90	91	91
6	91	91	92	93	93
7	93	93	94	93	94
8	95	94	94	94	93
9	95	94	94	93	93
10	94	94	93	93	94

only the next most recent year available. Unfortunately, the latter situation is more common in insurance than is the former.

As is shown in Table 11, the optimal credibility (as determined using the least squares criterion) decreases as the delay increases. Less current information is less valuable for estimating the future.

TABLE 11

OPTIMAL CREDIBILITY (CRITERION #1, LEAST SQUARES)

Time Between Latest Data Point and Future Prediction	NL				
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
1	68	71	74	76	74
2	51	59	64	64	63
3	47	53	55	56	55
4	40	45	47	47	45
5	33	38	40	39	36
6	30	33	34	32	30
7	24	26	26	25	24
8	19	20	21	21	20
9	14	16	17	18	20
10	11	13	15	18	21

Time Between Latest Data Point and Future Prediction	AL				
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
1	65	70	72	72	70
2	51	57	58	57	56
3	42	47	46	46	45
4	35	36	37	36	36
5	25	28	28	28	25
6	21	22	22	19	18
7	15	16	14	14	13
8	11	9	10	10	9
9	6	7	8	7	9
10	7	7	7	9	10

9. MORE GENERAL SOLUTIONS

In Section 6, four relatively simple forms of a solution were given. In this section, more general forms of a solution will be given.

9.1 Combine Previous Estimate and Most Recent Data

In the fifth method, one gives the latest year of *data* weight Z, and gives the previous *estimate* weight $1 - Z$. Of course, one has to choose an initial estimate.[29] In this case, for each risk the initial estimate will be taken as the grand mean of 50%.[30] Once this estimation method has been used for several years, the initial estimate has very little weight.

For example, let us assume $Z = 60\%$. Then the weights assigned to the given years of data used in estimating the result for the year 1911 would be as follows:

Year of Data	Weight in Estimate of 1911
1910	$Z = 60\%$
1909	$Z(1 - Z) = 40\% \times 60\% = 24\%$
1908	$Z(1 - Z)^2 = 40\% \times 40\% \times 60\% = 9.6\%$
1907	$Z(1 - Z)^3 = 9.6\% \times 40\% = 3.84\%$
1906	$Z(1 - Z)^4 = 3.84\% \times 40\% = 1.54\%$
1905	$Z(1 - Z)^5 = 1.54\% \times 40\% = .61\%$
1904 and Prior	$(1 - Z)^6 = .41\%$

The above assumes that the latest year of data is always given 60% weight, while the current estimate is given 40% weight.

Thus in this case, one gets a geometrically decreasing weight. This procedure is called (single) exponential smoothing [10]. It is an example of what mathematicians call a "filter."[31] Once the process of exponential

[29] This is precisely analogous to choosing a "seed" value in exponential smoothing.

[30] One could use subjective judgement to choose the initial estimate for each risk. Also one could use data from the period prior to that displayed in this paper; this has been avoided for the sake of simplicity.

[31] Morrison [11] gives this as an example of a "fading-memory polynomial filter."

smoothing gets "up to speed," it is equivalent to a weighted least squares regression, where the fitted line is horizontal,[32] and where the weights are geometrically decreasing as the data get less recent.

9.2 More General Varying Weights

In Section 9.1, one gave geometrically decreasing weight to years of data further in the past. More generally one can make the estimate:

$$F = \Sigma Z_i X_i + (1 - \Sigma Z_i)M$$

where the weights Z_i depend on how far in the past are the data X_i. For years for which data are not available (presumably because they are too far in the past) one uses the grand mean M instead of the data. This method is a generalization of the methods in Sections 6 and 9.1.

Unfortunately, calculating or empirically determining the optimal values of the weights Z_i becomes difficult as more years of data are used. Also, there are many vectors of Z_i that are very close to optimal; i.e., the n-dimensional volume of values $Z_1, \ldots, Z_n$ that produce close to optimal results is relatively large.

10. THE CRITERIA APPLIED TO THE MORE GENERAL SOLUTIONS

In this section the three criteria in Section 7 will be applied to the more general solutions to the problem given in Section 9. For simplicity, the results will be shown for the situation where there is no delay in obtaining the data for use in making the next estimate. In Section 8.6, an example was given of the results of such a delay in receiving data. The same general pattern would apply here.

10.1 Geometrically Decreasing Weights

In Section 9.1, weight Z is applied to the latest available year of data, while weight $1 - Z$ is applied to the previous estimate.

---

[32] Double exponential smoothing, sometimes called linear exponential smoothing, would be equivalent to a weighted linear least squares regression, with geometrically decreasing weights as the data got less recent.

Table 12 gives the mean squared errors for various values of Z. The optimal values of Z, using criterion #1 (least squares), are all close to 55%.[33] This results in weights to the various years of data very similar to those in the example in Section 9.1.

TABLE 12

MEAN SQUARED ERRORS* (.0001) THAT RESULT FROM APPLYING Z TO LATEST YEAR OF DATA AND $1 - Z$ TO PREVIOUS ESTIMATE

Z	NL	NLR**	AL	ALR**
0	79	97	95	96
.1	61	70	72	78
.2	56	63	65	71
.3	52	60	60	67
.4	50	57	57	64
.5	49	56	55	63
.6	50	56	55	63
.7	50	57	55	64
.8	52	58	56	66
.9	54	59	58	69
1.0	57	62	61	73

* First 10 years are not included in the computation of the squared errors in order to eliminate the calibration period.

** Data reversed in time.

In this case there is no significant reduction in squared error beyond what was previously obtained by applying credibility to the latest available year.[34]

Table 13 displays the results of applying criterion #2, limited fluctuation. Values of the credibility between 40% and 80% generally perform well.

[33] For the NL data set the optimal credibility is 53%. For the NLR data set, it is 58%. For the AL data set it is 60%. For the ALR data set it is 54%.

[34] Compare the results in Table 6 for $N = 1$ with those in Table 12.

Table 14 displays the results of applying criterion #3, Meyers/Dorweiler.[35] Unlike the previous two cases, the optimal credibilities are close to zero; 5% to 10% credibility produces correlations close to zero. The use of such small credibilities is approximately the same as using 10 to 20 years of data as the basis for the estimate, since the geometrically decreasing weights decline only slowly.

TABLE 13

PERCENT OF TIME* THAT THE ESTIMATE IS IN ERROR BY MORE THAN 20%
APPLYING Z TO LATEST YEAR OF DATA AND $1 - Z$ TO PREVIOUS ESTIMATE

Z	NL	NLR**	AL	ALR**
0	25	31	33	31
.1	23	23	25	26
.2	21	22	24	25
.3	18	21	21	23
.4	16	21	20	21
.5	16	20	19	22
.6	16	20	19	21
.7	17	19	20	22
.8	18	19	19	22
.9	18	20	18	23
1.0	19	21	19	26

* First 10 years are not included in the computation in order to eliminate the calibration period.

** Data reversed in time.

[35] In this case, the results of the first 20 years were excluded from the computation, in order to eliminate the calibration period. Twenty years were used, rather than ten years as in the previous two tables, since in this case smaller credibilities are optimal and smaller credibilities require a longer calibration period.

TABLE 14

CORRELATIONS* (KENDALL TAU) THAT RESULT FROM APPLYING Z TO LATEST YEAR OF DATA AND $1 - Z$ TO PREVIOUS ESTIMATE

Z	NL	NLR**	AL	ALR**
0***	.11	.16	.28	.14
.1	−.03	.01	.00	−.09
.2	−.05	−.05	−.04	−.10
.3	−.08	−.09	−.07	−.12
.4	−.10	−.12	−.10	−.13
.5	−.13	−.15	−.12	−.15
.6	−.15	−.18	−.14	−.17
.7	−.18	−.20	−.16	−.20
.8	−.20	−.23	−.18	−.22
.9	−.23	−.25	−.21	−.24
1.0	−.26	−.27	−.23	−.28

* First 20 years are not included in the computation of the correlations in order to eliminate the calibration period.

** Data reversed in time.

*** Limit as Z approaches zero.

10.2 More General Varying Weights

In Section 9.2, varying weights Z_i are applied to the most recent N years, while the remaining weight is given to the grand mean. This method will only be examined using criterion #1, least squares. One can solve numerically for the set of weights which produce the least squared error, using a given number of years of data.[36] The results are as follows:

[36] Unfortunately, as the number of years increases, the amount of computer time required also increases.

Using Most Recent Two Years of Data ($N = 2$, $\Delta = 1$)

	Credibility		
	Second Most Recent Year	Most Recent Year	Mean Squared Error (.0001)
NL	9.6%	61.1%	48
AL	13.1%	56.9%	54

Using Most Recent Three Years of Data ($N = 3$, $\Delta = 1$)

	Credibility			
	Third Most Recent Year	Second Most Recent Year	Most Recent Year	Mean Squared Error (.0001)
NL	16.4%	1.1%	59.0%	45
AL	8.1%	9.1%	55.7%	53

Most of the credibility is assigned to the most recent year. The complement of credibility, which is assigned to the grand mean, is about 25 to 35 percent, decreasing as N increases.

Complement of Credibility

	$N = 1$*	$N = 2$	$N = 3$
NL	32%	29%	24%
AL	35%	30%	27%

* One minus the optimal credibility from Table 9.

The mean squared error is reduced from that using only the latest year of data.[37]

[37] Since the use of fewer years of data is just a special case, the least squared error using more years of data must be less than or equal that using fewer years of data.

	Mean Squared Error (.0001)		
	$N = 1$*	$N = 2$	$N = 3$
NL	49	48	45
AL	56	54	53

* From Table 6.

11. EQUATIONS FOR LEAST SQUARES CREDIBILITIES

In Section 11.2 are equations to solve for the least squares credibility. These equations follow from the assumed covariance structure discussed in Section 11.1. In Section 11.3 the equations in Section 11.2 are modified to constrain them to place no weight on the grand mean. Section 11.4 compares the mean squared errors that result from different credibilities. Section 11.5 briefly discusses the validity of the results derived in this paper.

11.1 The Covariance Structure

By analyzing the covariance structure, one can set up matrix equations to solve for the credibilities that minimize the squared error. These matrix equations are discussed in the next section.

As shown in Appendix D, the variance of the data can be broken down into two pieces. There is the variance between the risks.[38] There is also the variance within the risks.[39] These two variances add up to the total variance.

	Between Variance	Within Variance	Total Variance[40]
NL	.001230	.007892	.009121
AL	.001619	.007875	.009494

[38] This has been denoted as τ^2.

[39] This has been denoted as $\delta^2 + \zeta^2$. δ^2 is what is usually termed process variance, while ζ^2 is the variance due to shifting parameters over time.

[40] May differ slightly from the sum of the other two variances due to rounding.

Also of interest is the covariance between the years of data. It is assumed that this is a function of the number of years separating the data. The observed values are given in Table 15. As was seen in Table 5, the covariance decreases as the years of data are further apart. After about 6 years the covariances are relatively close to zero.

TABLE 15

COVARIANCE (.0001)

Years Separating Data	NL	AL
0*	7892	7875
1	4919	4527
2	3416	3175
3	3128	2411
4	2541	1766
5	1810	780
6	1566	383
7	955	−99
8	387	−561
9	−74	−1068
10	−394	−878
11	−558	−980
12	−389	−1092
13	3	−737
14	59	−814
15	212	−453
16	603	−39
17	786	−139
18	302	214
19	47	279
20	−268	415

*Equal by definition to the within variance.

It is possible to divide the within variance into two parts. The first part is the process variance excluding the effect of shifting parameters over time.[41] The second part is that portion of the within variance due to shifting parameters over time.[42] While this division may aid our understanding, it is not necessary for the calculation of the least squares credibilities. Not coincidentally, this division cannot be performed based solely on the reported data contained in Tables 1 and 2. This subject is discussed in more detail in Appendix D.

11.2 Matrix Equations for Least Squares Credibilities

Using the estimation method described in Section 9.2:

$$F = \sum_{i=1}^{N} Z_i X_i + (1 - \sum_{i=1}^{N} Z_i)M \tag{11.1}$$

As derived in Appendix C, one gets the following expression for the expected squared error between the observation and prediction:

$$\begin{aligned} V(Z) = & \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j \, (\tau^2 + C(|i - j|)) \\ & - 2 \sum_{i=1}^{N} Z_i \, (\tau^2 + C(N + \Delta - i)) \\ & + \tau^2 + C(0) \end{aligned} \tag{11.2}$$

In equation (11.2) we have used the following quantities defined in Appendix D.

τ^2 = between variance
$C(k)$ = covariance for data for the same risk, k years apart
= "within covariance"
$C(0)$ = within variance
Δ = the length of time between the latest year of data used and the year being estimated

[41] This has been denoted as δ^2.
[42] This has been denoted as ζ^2.

Equation 11.2 shows that the squared error is a second order polynomial in the Z_i.[43] This equation is the fundamental result for analyzing least squares credibility.

One can differentiate equation 11.2 in order to get N linear equations in N unknowns, which can be solved for the optimal credibilities.

$$\sum_{j=1}^{N} Z_j(\tau^2 + C(|i - j|)) = \tau^2 + C(N + \Delta - i) \quad i = 1, 2, \ldots N \qquad (11.3)$$

The set of equations 11.3 can be solved on a computer relatively easily using the usual methods from matrix theory. The results of doing so for $\Delta = 1$, using the average of the variances and covariance determined from the NL and AL data separately,[44] are shown in Table 16.

TABLE 16

LEAST SQUARES CREDIBILITIES, SOLUTIONS OF MATRIX EQUATIONS 11.3 ($\Delta = 1$)

Number of Years of Data Used (N)	Years Between Data and Estimate									
	1	2	3	4	5	6	7	8	9	10
1	66.0%	—	—	—	—	—	—	—	—	—
2	57.7	12.6	—	—	—	—	—	—	—	—
3	56.1	4.8	13.5	—	—	—	—	—	—	—
4	55.6	4.6	11.5	3.5	—	—	—	—	—	—
5	55.7	5.1	11.7	6.0	−4.4	—	—	—	—	—
6	55.9	4.9	11.3	5.8	−6.6	3.9	—	—	—	—
7	56.0	4.7	11.5	6.2	−6.5	5.9	−3.5	—	—	—
8	56.0	4.7	11.4	6.2	−6.3	5.9	−2.8	−1.2	—	—
9	56.1	4.9	11.0	6.6	−6.7	5.3	−3.1	−4.3	5.6	—
10	55.9	5.0	11.2	6.4	−6.4	5.1	−3.4	−4.5	3.6	3.5

The complement of credibility is applied to the grand mean.

First column is the credibility applied to the most recent year, second column is the credibility applied to the next most recent year, etc.

Note: Based on the average of the variances and covariances determined from the NL and AL data separately; however, assumes that for a separation of eight years or more, the covariance is zero.

[43] When $N = 1$, the squared error is a parabola as a function of the credibility. This has been noted before, for example in Appendix B of Meyers [12].

[44] It is assumed that for a separation of eight years or more, the covariance is zero.

The results conform reasonably well to those determined in Section 10.2.

The credibilities applied to the most recent year quickly converge to about 56%. The credibilities for the less recent years are much smaller. However, these credibilities do not monotonically decline as the years get less recent. There is a complicated pattern of weights determined by the covariance matrix. Some of the weights are even less than zero.[45]

The optimal credibilities are uniquely determined given the covariance structure. However, there are many other sets of credibilities which produce expected squared errors very close to minimal. The precise values of the credibilities are not particularly important, although the general range of credibilities that perform well might be instructive.

One can apply equation (11.2) to the method discussed in Sections 6.5 and 8.3 of applying equal weight Z_i to the latest N years of data, where

$Z_i = Z/N$ for $i = 1, \ldots, N$

As shown in Appendix C, the least squares credibility in this case is given by:

$$Z = N \frac{N\tau^2 + \sum_{i=1}^{N} C(N + \Delta - i)}{N^2\tau^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} C(|i - j|)} \tag{11.4}$$

The results of applying this equation for $\Delta = 1$, using the average of the variances and covariances determined from the NL and AL data separately,[46] are shown in Table 17.

Table 17 can be compared to Table 11.

The results in Table 11 conform reasonably well to those determined empirically for each data set (for $\Delta = 1$).

[45] Giving negative weight to some years allows a larger weight to be given to other years. The net effect is to reduce the expected squared error.

[46] It is assumed that for a separation of eight years or more, the covariance is zero.

TABLE 17

LEAST SQUARES CREDIBILITY, SOLUTION TO EQUATION 11.4 ($\Delta = 1$)

Number of Years of Data Used (N)	Z	$Z \div N$
1	66.0%	66.0%
2	70.3	35.2
3	72.9	24.3
4	73.6	18.4
5	72.2	14.4
6	71.3	11.9
7	69.9	10.0
8	68.2	8.5
9	67.3	7.5
10	66.9	6.7

Equal weight Z/N is applied to each of the N most recent years of data. The complement of credibility, $1 - Z$, is applied to the grand mean.

Note: Based on the average of the variances and covariances determined from the NL and AL data separately; however, assumes that for a separation of eight years or more, the covariance is zero.

11.3 Placing No Weight on the Grand Mean

Once the estimation method described in Sections 9.1 and 10.1 gets "up to speed," the initial estimate, which was taken as the grand mean, has very little weight. For all intents and purposes each risk is estimated based on its own past data, without relying on the data of other risks, in particular the grand mean.[47]

[47] The covariance structure is herein estimated using the data for all risks. This in turn is used to estimate the optimal credibilities. However, the credibilities are applied to the data for the particular risk we are estimating.

One can constrain the credibilities used in equation 11.1, so that they add to unity, thus giving no weight to the grand mean. Equation 11.1 then becomes

$$F = \sum_{i=1}^{N} Z_i X_i \tag{11.5}$$

with the constraint

$$\sum_{i=1}^{N} Z_i = 1. \tag{11.6}$$

The least squres credibilities for equations 11.5 and 11.6 are derived in Appendix C using the method of Lagrange Multipliers. The result is a set of $N + 1$ linear equations in $N + 1$ unknowns, the Z_i for $i = 1, \ldots, N$, and λ, the Lagrange Multiplier. There is the single constraint equation 11.6, plus the N equations 11.7.

$$\sum_{j=1}^{N} Z_j \, C(|i - j|) = C(N + \Delta - i) + \frac{\lambda}{2}, \quad i = 1, 2, \ldots, N \tag{11.7}$$

The set of equations 11.6 and 11.7 can be solved on a computer relatively easily using the usual methods from matrix theory. The results of doing so for $\Delta = 1$, using the average of the variances and covariances determined from the NL and AL data separately,[48] are shown in Table 18.

11.4 Mean Squared Errors

The mean squared errors that result from using the credibilities in Tables 16, 17, and 18 are displayed in Table 19.

When applying general weights to the latest N years of data, giving the most remote year of data no weight is equivalent to the case of using the latest $N - 1$ years of data. Since using the latest $N - 1$ years of data is a special case of using the latest N years of data, we expect the squared errors to decline, or remain constant.

This is what we observe for the credibilities from Table 16. They decline until $N = 6$, where the point of diminishing returns is reached.

[48] It is assumed that for a separation of eight years or more, the covariance is zero.

TABLE 18

LEAST SQUARES CREDIBILITIES, SOLUTIONS OF EQUATIONS 11.6 AND 11.7 ($\Delta = 1$)

Number of Years of Data Used (N)	Years Between Data and Estimate									
	1	2	3	4	5	6	7	8	9	10
1	100.0%	—	—	—	—	—	—	—	—	—
2	72.6	27.4	—	—	—	—	—	—	—	—
3	66.1	10.3	23.6	—	—	—	—	—	—	—
4	63.5	9.1	16.0	11.4	—	—	—	—	—	—
5	63.1	8.7	15.8	9.5	2.9	—	—	—	—	—
6	62.8	7.6	14.1	8.6	−3.9	10.8	—	—	—	—
7	62.5	7.7	13.8	8.2	−4.1	9.0	2.9	—	—	—
8	62.3	7.3	14.0	7.7	−4.8	8.6	−0.2	5.1	—	—
9	61.8	7.3	13.0	8.3	−5.7	7.0	−1.1	−1.9	11.2	—
10	60.8	7.5	13.1	7.7	−5.2	6.3	−2.2	−2.5	6.1	8.4

The credibilities are constrained to sum to unity.

First column is the credibility applied to the most recent year, second column is the credibility applied to the next most recent year, etc.

Note: Based on the average of the variances and covariances determined from the NL and AL data separately; however, assumes that for a separation of eight years or more, the covariance is zero.

Applying the same weight to each year is a special case of using the general weights. Thus the squared errors that result from using the credibilities from Table 17 should be greater than or equal to those that result from the credibilities from Table 16. This is the case, as shown in Table 19. Also, as was observed in Section 8.3, using more years of data leads in this case to larger squared errors.

Applying no weight to the grand mean is a special case of using the general weights. Thus the squared errors that result from using the credibilities from Table 18 should be greater than or equal to those that result from the credibilities from Table 16. As is shown in Table 19, the squared errors are substantially greater, with the gap narrowing as the number of years increases.

TABLE 19

Number of Years of Data Used (N)	Mean Squared Errors (.0001)*		
	Using the Credibilities From Table 16**	Using the Credibilities From Table 17***	Using the Credibilities From Table 18****
1	52	52	63
2	51	54	58
3	49	55	54
4	48	57	52
5	48	60	52
6	47	61	51
7	47	64	51
8	47	66	51
9	47	68	51
10	47	70	50

* Mean squared error using the stated credibilities to predict for the NL and AL data sets.

** The complement of credibility is given to the grand mean.

*** Equal weight to N years, with the complement of credibility given to the grand mean.

**** The credibilities add up to one, and thus no weight is given to the grand mean.

11.5 Validity of Results

The credibilities determined in Sections 10 and prior are all determined empirically by directly working with the data. In this section equations for the least squares credibilities have been introduced along with an assumed covariance structure.

The credibilities resulting from the use of the equations in this section are comparable to those determined in the previous sections empirically. As is shown in Appendix F, the observed pattern of squared errors is comparable to that derived from the assumed covariance structure.

Therefore, the results of this section are an appropriate means of estimating least squares credibilities for this example. How well these results would apply to another situation would depend on the covariance structure that underlies the particular data set.

12. MISCELLANEOUS

Section 12.1 contrasts the Meyers/Dorweiler Criterion vs. the other criteria. Section 12.2 discusses a somewhat artificial ratemaking example. It is intended to point the way towards applying these or similar methods to practical situations. Section 12.3 compares the baseball example to typical insurance applications. Section 12.4 shows that the estimates that result herein from the use of the credibilities are in balance. Section 12.5 discusses the question of what estimation method to select for predicting the future loss record of baseball teams. It is included in order to complete the illustrative example used throughout this paper.

12.1 Contrasting the Meyers/Dorweiler Criterion vs. the Other Criteria

Section 10.1 provides a good example of how criterion #3, Meyers/Dorweiler, differs on a basic conceptual level from the first two criteria. Both of the other criteria are concerned with eliminating large errors.[49] Criterion #1, least squares, does this since even a few large errors will

[49] Mahler [7] compares the credibilities that result from the application of the Bühlmann/least squares criterion and the credibilities that result from the application of the classical/limited fluctuation criterion.

greatly increase the sum of squared errors. Criterion #2, limited fluctuation, does this directly by minimizing the number of errors larger than the selected size.

In contrast, criterion #3, Meyers/Dorweiler, is concerned with the pattern of the errors. Large errors are not a problem, as long as there is no pattern relating the errors to the experience rating modifications. For example, consider the following two situations. In each case, for simplicity, only four risks are assumed.

Situation #1	
Modification	Error
1.20	+30%
1.20	−30%
.80	+40%
.80	−40%

Situation #2	
Modification	Error
1.30	+2%
1.10	+1%
.90	−1%
.70	−2%

Situation #2 with its small errors is preferable under the first two criteria. Situation #1 with its lack of a pattern of errors is preferable under the Meyers/Dorweiler criterion. Most actuaries would prefer Situation #2.

This example is not meant to discourage use of the Meyers/Dorweiler criterion. Rather it is meant to point out the potential hazards of relying solely on any single criterion, as well as the importance of understanding exactly what is being tested by any criterion that is being used.

12.2 A Ratemaking Example

Assume for a given line of insurance that the five most recent annual loss ratios are being combined to calculate a rate level indication.[50] Assume that it is three years from the latest year of data to the average date of loss under the proposed new rates.[51] A weighted average of the annual loss ratios will be used to estimate the future loss ratio.

If we assume a given covariance structure, equations 11.6 and 11.7 can be used to calculate the optimal least squares set of weights, Z_i, such that

$$\sum_{i=1}^{5} Z_i = 1.$$

Assume the covariance of the loss ratios separated by a given number of years is as follows:[52]

Separation in Years	Covariance in Loss Ratios (.00001)
0	130
1	60
2	55
3	50
4	45
5	40
6	35
7	30

Then the optimal weights are: 11.6%, 13.4%, 17.3%, 23.8%, 33.9%, with the more recent data receiving more weight. It is interesting to note that these weights can be reasonably approximated by the weights used in Walters [13], i.e., 10%, 15%, 20%, 25%, and 30%.

This example is for illustrative purposes only. It should not be taken as a derivation of the correct weights to use in any real world application. Unfortunately, in order to apply this idea to real world applications one

[50] The loss ratios for the separate years are presumed to have been adjusted for trend, development, and any other factors such as law changes.

[51] This period will vary, but $\Delta = 3$ is not uncommon.

[52] This would be produced by $\delta^2 = .0004$, $\zeta^2 = .0009$, $\ell(1) = .667$, $\ell(2) = .611$, $\ell(3) = .556$, $\ell(4) = .500$, $\ell(5) = .494$, $\ell(6) = .389$, $\ell(7) = .333$, where the quantities are defined as in Appendix D.

has to estimate the covariance matrix. This will be affected by shifting parameters over time. It will also be affected by the varying quantity of data available in each year.[53] It will be affected by the uncertainty in the trend estimates and loss development estimates applied to each year. These complications are beyond the scope of this paper.

12.3 Baseball Example vs. Typical Insurance Applications

In many typical insurance applications, credibility is used in the process of determining relativities. For example, credibility is used to determine the rate for a class or territory relative to the overall rate level. Credibility is also used in experience rating, where the rate for an individual risk is adjusted relative to an average.

In these situations, where a class, territory, or individual risk is compared relative to an average, the result depends on the other classes, territories, or risks which make up the average. An automobile territory with a low relativity in Massachusetts could have a higher loss potential than a automobile territory with a high relativity in Vermont. A workers compensation insured could have a credit experience modification simply because of the bad loss experience of several other employers in the same business in the same state. An insured with a .9 experience modification could have a higher loss potential than another risk with 1.1 modification in a different business or in a different state. The baseball example has this same feature. A team is being compared relative to the average in the league. The losing percentage only has relevance to rank teams in a single league relative to the average for that league. The difference in this example is that the average is a known constant. The grand mean is always .500.

In baseball if somebody loses, then somebody else wins. Thus the win-loss records of seven teams determine that of the remaining team in an eight team league.[54]

---

[53] In this paper, each year of baseball data represented a comparable number of games, so this aspect was not important.

[54] The win-loss record of teams in the same league should be negatively correlated by an amount proportional to the number of games the two teams have played.

This could have had a major impact on the analysis of this example. However, each team played each other team in the league approximately the same number of times each year and each team played approximately the same number of games in total.[55] Thus no team had its results distorted by playing a weaker or stronger schedule.

12.4 Estimates in Balance

The estimation methods used herein are always in balance.

The most general estimation method considered herein was given by equation 9.1, where the subscript j has been added to identify team j:

$$F_j = \sum_{i=1}^{N} Z_i X_{ij} + \left(1 - \sum_{i=1}^{N} Z_i\right) M$$

Then the average of the estimates F_j for all the teams in the league is given by:

$$\frac{1}{8}\sum_{j=1}^{8} F_j = \sum_{i=1}^{N} Z_i \left(\frac{1}{8}\sum_{j=1}^{8} X_{ij}\right) + \left(1 - \sum_{i=1}^{N} Z_i\right) M$$

$$= \left(\sum_{i=1}^{N} Z_i\right) M + \left(1 - \sum_{i=1}^{N} Z_i\right) M$$

$$= M$$

Note that for a given year i, the credibility Z_i assigned to each team's experience X_{ij} for that year is the same for all teams. Also note the fact that the grand mean is the same for all years.

That the estimates are in balance can be verified directly for the example given in Table 20. The predicted losing percentages for each year average to .500, subject to rounding.

12.5 Choice of a Prediction Method

The example in this paper is for illustrative purposes only; the purpose of this paper was not to predict baseball teams' win-loss records. Nevertheless, it may be of interest to choose a reasonable prediction method

[55] The schedule was exactly balanced, but a few scheduled games are sometimes not played.

TABLE 20

National League, Predictions of Losing Percentages

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1904	.541	.479	.461	.495	.469	.575	.379	.606
1905	.582	.568	.432	.456	.398	.610	.423	.534
1906	.615	.613	.424	.480	.368	.504	.408	.588
1907	.627	.568	.334	.533	.389	.531	.421	.598
1908	.594	.559	.351	.544	.448	.463	.426	.616
1909	.578	.598	.375	.529	.408	.476	.405	.631
1910	.633	.599	.366	.508	.427	.498	.354	.614
1911	.614	.576	.371	.509	.426	.492	.430	.581
1912	.651	.563	.411	.524	.400	.490	.443	.522
1913	.624	.582	.414	.511	.376	.508	.425	.557
1914	.561	.555	.438	.550	.377	.454	.471	.596
1915	.458	.526	.478	.570	.441	.504	.515	.509
1916	.468	.493	.505	.541	.504	.443	.517	.529
1917	.437	.438	.536	.574	.464	.440	.551	.559
1918	.503	.506	.519	.511	.423	.442	.603	.492
1919	.534	.519	.425	.493	.440	.512	.514	.565
1920	.560	.512	.467	.394	.413	.584	.509	.565
1921	.567	.448	.488	.458	.449	.573	.490	.528
1922	.509	.486	.543	.501	.419	.618	.449	.474
1923	.592	.492	.499	.469	.426	.596	.461	.466
1924	.596	.503	.485	.448	.412	.626	.450	.479
1925	.615	.448	.478	.462	.418	.605	.440	.536
1926	.553	.522	.525	.474	.441	.562	.418	.505
1927	.556	.516	.485	.458	.488	.583	.452	.465
1928	.571	.550	.472	.497	.441	.610	.422	.436
1929	.613	.509	.441	.487	.434	.648	.452	.418
1930	.603	.530	.406	.539	.449	.558	.442	.471
1931	.556	.472	.430	.570	.448	.614	.476	.434
1932	.564	.487	.452	.586	.448	.559	.498	.403
1933	.513	.478	.441	.584	.504	.520	.467	.492

TABLE 20

(CONTINUED)

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1934	.487	.537	.455	.588	.442	.564	.460	.468
1935	.487	.523	.447	.609	.434	.578	.492	.433
1936	.633	.534	.405	.558	.427	.568	.460	.417
1937	.545	.543	.442	.532	.426	.603	.470	.440
1938	.518	.563	.421	.582	.412	.579	.457	.467
1939	.498	.536	.436	.490	.449	.635	.450	.507
1940	.543	.486	.456	.437	.477	.641	.518	.445
1941	.547	.458	.494	.398	.508	.635	.495	.466
1942	.569	.406	.522	.433	.510	.659	.491	.411
1943	.572	.381	.538	.477	.472	.661	.525	.378
1944	.551	.452	.519	.457	.573	.590	.492	.368
1945	.559	.530	.515	.451	.544	.586	.455	.363
1946	.546	.470	.428	.543	.513	.629	.472	.399
1947	.450	.487	.468	.538	.562	.559	.538	.400
1948	.434	.459	.511	.531	.495	.579	.559	.433
1949	.453	.439	.548	.554	.504	.554	.497	.451
1950	.413	.492	.571	.564	.511	.502	.527	.419
1951	.440	.470	.564	.556	.470	.454	.570	.477
1952	.414	.501	.572	.548	.429	.503	.563	.472
1953	.411	.542	.518	.541	.428	.458	.646	.457
1954	.375	.455	.553	.543	.503	.475	.627	.469
1955	.416	.456	.554	.521	.423	.497	.626	.507
1956	.395	.454	.532	.515	.481	.497	.594	.531
1957	.419	.434	.572	.453	.521	.523	.566	.512
1958	.451	.421	.567	.482	.533	.504	.571	.471
1959	.507	.425	.538	.492	.501	.532	.492	.512
1960	.464	.451	.523	.509	.482	.551	.502	.518

Note: Using latest three years of data, with weights of 10%, 10%, 55% (55% weight to the most recent year; 25% weight to grand mean).

for this particular problem. Assume that $\Delta = 1$, i.e., 1910 data are available to predict 1911, etc.

Based on Table 19, the credibilities in Table 16 work well.

The author would recommend avoiding using many years of data unless it substantially improved the accuracy. It is better to keep things simple. For this particular problem, based on Table 19, there seems little advantage to using more than 3 years of data. For example, giving 55% weight to the most recent year, 10% weight to the next most recent year, 10% weight to the third most recent year, and the remaining 25% weight to the grand mean works reasonably well.[56]

The predictions that result from this method of estimation applied to the National League data are shown in Table 20.[57] The errors are shown in Table 21.

The mean squared error is .0046.[58] There is a 14% chance of an error of more than 20%. The correlation used in the Meyers/Dorweiler criterion is .02, not significantly different from zero. Thus according to all three criteria this prediction method works well.

13. CONCLUSIONS

The data from baseball used in this paper provide a useful way to examine and illustrate credibility concepts.

The methods and concepts illustrated here can be applied to problems actuaries deal with in insurance. However, this paper is only a first step; there is further work required to apply these general concepts to any specific practical situation.

---

[56] Many other choices would also work reasonably well. This illustrates the typical situation where once the general form of the weights is determined, there is a range of weights that work well. Usually, the specific choice of weights within that range has relatively little impact on the final result.

[57] For example, the 1904 entry under NL2: .479 = (.10)(.419) + (.10)(.457) + (.55)(.485) + (.25)(.500), where the first three values are from Table 1, and .500 is the grand mean.

[58] The mean squared error is .0049 when the method is applied to both the AL and NL data sets. This is a standard deviation of 10½ losses out of a season of 150 games; the process standard deviation is about 6 losses out of a season of 150 games.

TABLE 21

National League, Errors of Predictions in Table 20

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1904	−.100	−.155	.069	.070	.162	−.083	−.052	.093
1905	−.087	−.116	.033	−.028	.084	.156	.050	−.089
1906	−.060	.047	.187	−.096	.000	−.032	.016	−.065
1907	.019	.007	.038	−.036	−.075	.096	.012	−.062
1908	.003	−.097	−.006	.018	.084	.002	.062	−.066
1909	−.128	−.043	.055	.032	.009	−.040	.129	−.014
1910	−.021	.015	.041	−.005	.018	.008	−.084	.026
1911	−.095	.003	−.032	−.033	.073	.012	−.018	.084
1912	−.009	−.058	.018	.014	.082	−.030	.059	−.066
1913	.081	.018	−.011	−.071	.040	.091	−.052	−.103
1914	.175	.042	−.056	−.060	−.078	−.065	−.081	.125
1915	.004	.052	−.045	.031	−.105	.096	−.011	−.020
1916	.054	.103	−.057	−.067	.070	.038	−.061	−.079
1917	−.092	−.098	.017	.080	.100	.012	−.118	.098
1918	−.070	−.042	.170	.042	−.004	−.111	.123	−.113
1919	−.056	.012	−.039	.179	.061	−.145	.025	−.041
1920	−.032	.116	−.046	−.070	−.029	−.011	.022	.052
1921	.083	−.045	−.094	−.084	.063	−.096	.078	.097
1922	−.145	−.020	.062	.059	.023	−.009	.001	.026
1923	−.057	−.014	.038	.060	.047	−.079	.026	−.018
1924	−.058	.100	.014	−.010	.020	−.010	.038	−.099
1925	.073	−.108	−.080	−.015	−.016	.049	.061	.039
1926	−.013	−.014	.057	.039	−.069	−.054	−.033	.083
1927	−.054	−.059	.041	.052	.085	.086	.062	.066
1928	−.102	.053	.063	.010	.045	−.107	−.019	.053
1929	−.023	−.033	.086	−.084	−.010	.112	.027	−.069
1930	.058	.088	−.010	−.078	.014	−.104	−.039	.068
1931	−.028	−.008	−.025	−.053	.020	.043	−.037	.090
1932	.064	.013	.036	−.024	−.084	.065	.056	−.129
1933	.052	−.097	−.001	−.034	.103	−.085	.032	.028

TABLE 21

(CONTINUED)

	NL1	NL2	NL3	NL4	NL5	NL6	NL7	NL8
1934	.004	.004	.025	−.068	.050	−.060	−.047	.089
1935	−.265	−.019	.096	.053	.029	−.004	.054	.056
1936	.094	−.031	−.030	.039	.024	−.081	.005	−.018
1937	.065	−.052	.046	−.104	.051	.002	.028	−.034
1938	.025	.026	.007	.129	−.035	−.121	.030	−.063
1939	−.085	.085	−.019	.120	−.041	−.067	−.106	.108
1940	−.029	.061	−.057	.091	−.049	−.032	.024	−.006
1941	−.050	.107	−.051	−.031	−.008	−.086	.021	.100
1942	−.032	.081	−.036	−.067	.069	−.063	−.060	.099
1943	.016	−.090	.022	.042	−.169	.077	.044	.060
1944	−.027	−.139	.006	.035	.008	−.011	.080	.050
1945	.000	.095	.151	−.153	.057	−.115	−.013	−.020
1946	.161	−.001	−.036	−.022	−.091	.077	−.119	.027
1947	.060	.045	−.084	.012	.088	−.038	−.059	−.022
1948	−.021	.054	−.073	−.051	.001	.008	.098	−.015
1949	.083	−.074	−.056	−.043	−.022	.080	−.042	.074
1950	−.009	.031	−.011	−.005	.069	.093	−.100	−.071
1951	.058	−.036	−.033	−.002	.094	−.072	−.014	.003
1952	.041	−.081	.072	−.004	.026	.068	−.164	.043
1953	.093	.139	−.060	−.017	−.117	−.003	−.029	−.004
1954	−.028	.033	−.031	.024	.133	−.038	−.029	−.063
1955	.057	.008	.025	.008	−.058	−.003	.016	−.051
1956	−.001	.051	−.078	.106	−.084	−.042	.023	.025
1957	−.036	.051	−.025	−.028	−.031	.023	−.031	.077
1958	−.088	.018	.035	−.024	.052	−.048	.116	−.061
1959	.071	−.024	.019	−.027	.040	−.052	−.002	−.027
1960	−.004	.022	−.087	−.056	−.005	−.066	.119	.076

Note: Predicted Losing Percentage minus Actual Losing Percentage

When shifting parameters over time is an important phenomenon, older years of data should be given substantially less credibility than more recent years of data. The more significant this phenomenon, the more important it is to minimize the delay in receiving the data that is to be used to make the prediction.

In this paper three different criteria were examined that can be used to select the optimal credibility: least squares, limited fluctuation, and Meyers/Dorweiler. In applications, one or more of these three criteria should be useful. While the first two criteria are closely related, the third criterion can give substantially different results than the others.

Generally the mean squared error can be written as a second order polynomial in the credibilities. The coefficients of this polynomial can be written in terms of the covariance structure of the data. This in turn allows one to obtain linear equation(s) which can be solved for the least squares credibilities in terms of the covariance structure.

REFERENCES

[1] G. G. Meyers, "An Analysis of Experience Rating," *PCAS* LXXII, 1985, p. 278.

[2] G. G. Venter, "Experience Rating–Equity and Predictive Accuracy," *NCCI Digest,* Volume II, Issue I, April 1987, p. 27.

[3] W. R. Gillam, "Parameterizing the Workers' Compensation Experience Rating Plan," Casualty Actuarial Society Discussion Paper Program, May 1990, p. 857.

[4] H. C. Mahler, "An Actuarial Analysis of the NCCI Revised Experience Rating Plan," *Casualty Actuarial Society Forum,* Winter 1991, p. 35.

[5] H. C. Mahler, Discussion of [1], *PCAS* LXXIV, 1987, p. 119.

[6] D. S. Neft and R. M. Cohn, *The Sports Encyclopedia: Baseball,* St. Martin's Press, 1987.

[7] G. G. Venter, "Classical Partial Credibility with Application to Trend," *PCAS* LXXIII, 1986, p. 27.

[8] P. Dorweiler, "A Survey of Risk Credibility in Experience Rating," *PCAS* XXI, 1934, p. 1.

[9] H. C. Mahler, "An Actuarial Note on Credibility Parameters," *PCAS* LXXIII, 1986, p. 1.

[10] S. C. Wheelwright and S. Makridakis, *Forecasting Methods for Management,* John Wiley and Sons.

[11] N. Morrison, *Introduction to Sequential Smoothing and Prediction,* McGraw-Hill, 1969.

[12] G. G. Meyers, "Empirical Bayesian Credibility for Workers' Compensation Classification Ratemaking," *PCAS* LXXI, 1984, p. 96.

[13] M. A. Walters, "Homeowners Insurance Ratemaking," *PCAS* LXI, 1974, p. 15.

[14] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Volume 2, Macmillan, 1979.

[15] F. De Vylder, *Introduction to the Actuarial Theory of Credibility*. English translation by Charles A. Hachmeister, 1980.

[16] H. U. Gerber and D. A. Jones, "Credibility Formulas of the Updating Type," *Credibility: Theory and Applictions*, edited by P. M. Kahn, Academic Press, 1975, pp. 89–105.

APPENDIX A

SOME RELEVANT FEATURES OF BASEBALL

Baseball is a competitive sport involving a combination of luck and skill. Two teams play against each other in a game; the team that scores the most "runs" wins the game, the other team loses.[1]

Each team has nine players in the game at a time.[2] Players may be substituted for, but once they leave the game they cannot return. Over this period of time each team had 20 to 25 players on its roster.[3] The individual skills of the players, as well as how their skills complement each other, has a direct impact on the quality of the team.

In addition to the players, a team has coaches and a field manager. By supervising the players' training and conditioning, providing advice, deciding who plays, and by various decisions throughout the game, these people have some effect on the percentage of games lost or won by the team.

Each team has an owner(s) and other office personnel.[4] By developing new players, trading for players with other teams, etc., management has some effect on the percentage of games lost or won by the team.

All of these elements that affect the quality of the team shift over time. A team's roster of players typically changes a little during the course of a single year; over the course of several years the changes are substantial. It is unusual for a player to be with a single team for more than 10 years, although on very rare occasions a player has played for a single team for 20 years.

Even if the identity of the players were to stay the same, the skill level of individual players changes over time. The most important effect is aging; as a player gets older he generally improves until he reaches a

[1] While it is possible for a baseball game to end in a tie, such games are ignored in major league standings.

[2] Currently the American League has added a tenth player, the designated hitter.

[3] Of the players on the roster, about half get most of the playing time, while the remainder see much less playing time.

[4] During the latter half of this period a team had a general manager.

peak and then declines. Injuries can have a profound impact on a player's skill; sometimes that impact is temporary while sometimes it is permanent.

The field managers and coaching staff also change over time.[5] In addition, the owner(s) and upper management change, but much less frequently.

Finally, a team occasionally relocates to another city.

It might be useful to think of the following analogy to a workers compensation risk. The baseball players correspond to the workers in the factory. The field manager corresponds to the plant manager. The baseball upper management corresponds to the corporation's upper management.

[5] Quite often the departure of the field manager will be related to the poor record of the team.

APPENDIX B

MEYERS/DORWEILER CRITERION AND KENDALL'S TAU

If an experience rating plan works properly, then after the application of experience rating, an insurer should be equally willing to write debit and credit risks. In other words, the modified loss ratio of expected losses to modified premiums should be the same for debit and credit risks.

Mathematically, we desire that the correlation between the experience modification and the modified loss ratio be zero.[1]

In the example in this paper, the experience modification corresponds to the ratio of predicted losing percentage to the grand mean losing percentage.[2] For example, a predicted losing percentage of 60% is equivalent to an experience modification of 60% ÷ 50% = 1.2. The modified loss ratio corresponds to the ratio of the actual losing percentage and the predicted losing percentage.[3] The third criterion used in this paper is that the correlation between these two ratios be zero. This corresponds to the criterion used by Meyers.

Meyers [1] uses the Kendall τ to measure correlation.

Let X and Y be two vectors of length n.[4] Kendall's τ can be calculated as follows [14]. Suppose Y is arranged in its natural order. Assume that the corresponding ranks of X are $X_1, X_2, \ldots X_n$, a permutation of 1, 2, ... n. Let Q be the number of inversions in $X_1, X_2, \ldots X_n$.[5] Then let

$$\tau = 1 - \frac{4Q}{n(n-1)}$$

[1] If the correlation is positive, then insurers would prefer to write credit risks. The credits and debits given are on average too small, i.e., the credibility assigned to the experience is too small. The situation is reversed for a negative correlation.

[2] The predicted losses are equal to the experience modification times the expected losses for an average risk in the class. In a more general situation one would have classifications of risks; in this example we have only one such classification and thus use the grand mean rather than the class mean.

[3] In general the modified loss ratio is equal to the expected loss ratio times the actual losses over the predicted losses. In this example, the expected loss ratio can be thought of as unity.

[4] In our case, X would be the experience modifications and Y would be the corresponding modified loss ratios.

[5] For example, in the X-ranking 3214 for $n = 4$, there are 3 inversions of order.

τ is symmetrically distributed on the range $[-1, +1]$. As is usual for measures of correlation, $+1$ signifies complete agreement and -1 signifies complete disagreement.

As shown in Kendall and Stuart [14],

$$\text{Var } \tau = \frac{2(2n + 5)}{9n(n - 1)}$$

As n approaches infinity the distribution of τ approaches the normal distribution.

In the examples in this paper, the variance of τ varies from .0009 to .0016.[6] The standard deviation of τ goes from .031 to .040. Thus an approximate 95% confidence interval around zero for τ has a radius of approximately .07, about two standard deviations.

[6] n = 8 teams times 59 years = 472 to n = 8 times 35 = 280.

APPENDIX C

MATRIX EQUATIONS FOR LEAST SQUARES CREDIBILITY

In this appendix, equations 11.2, 11.3, 11.4, and 11.7 in the main text are derived. The squared error is written as a second order polynomial in the credibilities, with the coefficients depending on the covariance structure discussed in Appendix D. This squared error is minimized by setting the partial derivative(s) with respect to the credibilities equal to zero.

Assume an estimate for year $N + \Delta$, using N years of data, is given by:

$$F = \sum_{i=1}^{N} Z_i X_i + (1 - \Sigma Z_i)M$$

where X_i is the data for year i, and M is the grand mean.[1] Let $Z_0 = 1 - \Sigma Z_i$. Write Z for the vector $Z_0, Z_1, \ldots, Z_N$.

Then the mean squared error between the prediction and the observation is given by the expected value of the squared difference between F and $X_{N+\Delta}$.

$$\begin{aligned} V(Z) &= \mathrm{E}[(F - X_{N+\Delta})^2] \\ &= \mathrm{E}\left[\left\{ \sum_{i=1}^{N} Z_i \left(X_i - X_{N+\Delta}\right) + Z_0 \left(M - X_{N+\Delta}\right)\right\}^2\right] \\ &= \sum_{i=1}^{N} Z_i^2 \mathrm{E}[(X_i - X_{N+\Delta})^2] \\ &\quad + \sum_{i=1}^{N} \sum_{j \neq i} Z_i Z_j \, \mathrm{E}[(X_i - X_{N+\Delta})(X_j - X_{N+\Delta})] \\ &\quad + 2 \sum_{i=1}^{N} Z_0 Z_i \, \mathrm{E}[(X_i - X_{N+\Delta})(M - X_{N+\Delta})] \\ &\quad + Z_0^2 \, \mathrm{E}[(M - X_{N+\Delta})^2] \end{aligned}$$

[1] It is assumed that the grand mean is known. This is the case in this paper. It is the case whenever one is only concerned with relativities compared to the overall average.

From Appendix D we have,[2]

$$\mathrm{E}[(X_i - X_{N+\Delta})^2] = 2\delta^2 + 2\zeta^2(1 - \ell(N + \Delta - i))$$

$$\mathrm{E}[(X_i - X_{N+\Delta})(X_j - X_{N+\Delta})] = \delta^2 + \zeta^2(1 + \ell(|i - j|) - \ell(N + \Delta - i) - \ell(N + \Delta - j))$$

$$\mathrm{E}[(X_i - X_{N+\Delta})(M - X_{N+\Delta})] = \delta^2 + \zeta^2(1 - \ell(N + \Delta - i))$$

$$\mathrm{E}[(M - X_{N+\Delta})^2] = \delta^2 + \zeta^2 + \tau^2$$

Therefore

$$V(Z) = \sum_{i=1}^{N} Z_i^2(2\delta^2 + 2\zeta^2(1 - \ell(N + \Delta - i)))$$

$$+ \sum_{i=1}^{N} \sum_{j \neq i} Z_i Z_j \left[\delta^2 + \zeta^2(1 + \ell(|i - j|) - \ell(N + \Delta - i) - \ell(N + \Delta - j))\right]$$

$$+ 2Z_0 \sum_{i=1}^{N} Z_i(\delta^2 + \zeta^2(1 - \ell(N + \Delta - i))$$

$$+ Z_0^2(\delta^2 + \tau^2 + \zeta^2)$$

$$V(Z) = \delta^2 \left[\sum_{i=0}^{N} \sum_{j=0}^{N} Z_i Z_j + \sum_{i=1}^{N} Z_i^2\right] + Z_0^2 \tau^2$$

$$+ \zeta^2 \left[\sum_{i=0}^{N} \sum_{j=0}^{N} Z_i Z_j + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j(\ell(|i - j|) - \ell(N + \Delta - i) - \ell(N + \Delta - j)) - 2Z_0 \sum_{i=1}^{N} Z_i \ell(N + \Delta - i)\right]$$

but

$$\sum_{i=0}^{N} Z_i = Z_0 + \sum_{i=1}^{N} Z_i = (1 - \sum_{i=1}^{N} Z_i) + \sum_{i=1}^{N} Z_i = 1$$

[2] In Appendix D, $X(\theta,t)$ = the observation for risk θ at time t. Since in this appendix none of the calculations are performed for individual risks, the θ has been suppressed in order to simplify the notation.

Therefore

$$V(Z) = \delta^2 + Z_0^2\tau^2 + \zeta^2 + \delta^2 \sum_{i=1}^{N} Z_i^2$$

$$+ \zeta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j (\ell(|i-j|) - \ell(N+\Delta-i) - \ell(N+\Delta-j))$$

$$- \zeta^2 Z_0 2 \sum_{i=1}^{N} Z_i \ell(N+\Delta-i)$$

$$V(Z) = \delta^2 + \zeta^2 + \tau^2 + \tau^2 \left[\sum_{i=1}^{N} Z_i \right]^2 - 2\tau^2 \sum_{i=1}^{N} Z_i$$

$$+ \delta^2 \sum_{i=1}^{N} Z_i^2 + \zeta^2 \sum_{i=1}^{N} \sum_{j=i}^{N} Z_i Z_j (\ell(|i-j|) - \ell(N+\Delta-i)$$

$$- \ell(N+\Delta-j)) - 2\zeta^2 \sum_{i=1}^{N} Z_i \ell(N+\Delta-i)$$

$$+ 2\zeta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j \ell(N+\Delta-i)$$

$$V(Z) = \delta^2 + \zeta^2 + \tau^2 + \tau^2 \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j - 2\tau^2 \sum_{i=1}^{N} Z_i$$

$$+ \delta^2 \sum_{i=1}^{N} Z_i^2 + \zeta^2 \sum_{i=1}^{N} \sum_{j=i}^{N} Z_i Z_j (\ell(|i-j|) - \ell(N+\Delta-j))$$

$$- 2\zeta^2 \sum_{i=1}^{N} Z_i \ell(N+\Delta-i)$$

$$V(Z) = \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j (\delta^2 \delta_{ij} + \tau^2 + \zeta^2 \ell(|i-j|))$$

$$- 2 \sum_{i=1}^{N} Z_i (\tau^2 + \zeta^2 \ell(N+\Delta-i)) + \delta^2 + \zeta^2 + \tau^2$$

This is equation 11.2 in the main text, with $\delta^2\delta_{ij} + \zeta^2\ell(|i-j|) = C(|i-j|)$ the covariance between data for a given risk $|i-j|$ years apart. It is left as an exercise to the reader to verify that the formula for the mean squared error compared to the underlying mean rather than the observed value would be exactly δ^2 less.

In order to minimize this squared error, one sets the partial derivatives with respect to Z_i equal to zero. This yields the following set of N equations.

$$2Z_i(\delta^2 + \tau^2 + \zeta^2) + \sum_{j \neq i} 2Z_j(\tau^2 + \zeta^2\ell(|i-j|)) - 2(\tau^2 + \zeta^2\ell(N + \Delta - i)) = 0, \quad i = 1, \ldots, N$$

$$\sum_{j=1}^{N} Z_j(\delta^2\delta_{ij} + \tau^2 + \zeta^2\ell(|i-j|)) = \tau^2 + \zeta^2\ell(N + \Delta - i), \quad i = 1, \ldots, N$$

This is equation 11.3 in the main text, again with

$$\delta^2\delta_{ij} + \zeta^2\ell(|i-j|) = C(|i-j|).$$

It is worth noting that equation 11.3 is very similar to the usual general matrix equation for optimal least squares credibilities:

$$\vec{Z} = \frac{\text{COV}[\vec{X}, Y]}{\text{COV}[\vec{X}, \vec{X}]}$$

where $\vec{X}$ is the vector of observations, and Y is the quantity to be estimated.[3] Here in equation 11.3, there is an additional term of τ^2, the between variance, added to the covariances. This is due to the application of the complement of credibility to the grand mean.

In the absence of shifting parameters over time ($\zeta^2 = 0$), the squared error is given by:

$$V(Z) = \delta^2 \left(1 + \sum_{i=1}^{N} Z_i^2\right) + \tau^2 \left(1 - \sum_{i=1}^{N} Z_i\right)^2$$

[3] See, for example, Theorem 3.3 in Chapter III of De Vylder [15].

The optimal credibilities are given by the solution to the equations:

$$\sum_{j=1}^{N} Z_j(\delta^2\delta_{ij} + \tau^2) = \tau^2, \quad i = 1, \ldots, N$$

The solution has all the credibilities equal:

$$Z_i = \frac{\tau^2}{N\tau^2 + \delta^2}, \quad i = 1, \ldots, N$$

$$\sum_{i=1}^{N} Z_i = \frac{N\tau^2}{N\tau^2 + \delta^2} = \frac{N}{N + \delta^2/\tau^2}$$

This is the familiar expression for the least squares credibility in the absence of shifting parameters over time.

If we set $Z_i = Z/N$ for $i = 1, \ldots, N$ then equation 11.2 becomes:

$$V(Z) = \frac{Z^2}{N^2}\left\{N\delta^2 + N^2\tau^2 + \zeta^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \ell(|i - j|)\right\}$$

$$- 2\frac{Z}{N}\left\{N\tau^2 + \zeta^2 \sum_{i=1}^{N} \ell(N + \Delta - i)\right\} + \delta^2 + \zeta^2 + \tau^2$$

Setting the derivative of $V(Z)$ equal to zero gives the least squares credibility:

$$Z = N \frac{N\tau^2 + \zeta^2 \sum_{i=1}^{N} \ell(N + \Delta - i)}{N^2\tau^2 + N\delta^2 + \zeta^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \ell(|i - j|)}$$

This is equation 11.4 in the main text, with $C(|i - j|) = \delta^2\delta_{ij} + \zeta^2\ell(|i - j|)$.

We can minimize $V(Z)$ in equation 11.2, given the constraint $\Sigma_{i=1}^{N} Z_i = 1$, by using Lagrange Multipliers.

We set the partial derivatives with respect to Z_i of

$V(Z) - \lambda\left(\sum_{i=1}^{N} Z_i - 1\right)$ equal to zero.

This produces the following N equations:

$$\sum_{j=1}^{N} Z_j(\delta^2\delta_{ij} + \zeta^2\ell(|i-j|)) = \zeta^2\ell(N + \Delta - 1) + \frac{\lambda}{2} \quad i = 1, \ldots, N$$

This is equation 11.7 in the main text. It is worth noting the absence from the above equation of τ^2, the between variance. This follows logically from the fact that the grand mean is given no weight and each risk is estimated solely from its own data.

APPENDIX D

COVARIANCE STRUCTURE

In this appendix, the covariance structure for the data sets in Tables 1 and 2 will be analyzed. As discussed in Section 11.1, the variance is the sum of three pieces, the between variance, the variance due to shifting parameters over time, and the process variance excluding the effect of shifting parameters over time. The analysis herein will define these three pieces.

Let $X(\theta,t)$ be the observation for risk θ at time t.

Let $\mu(\theta,t)$ be the expected value for risk θ at time t.

$\mu(\theta,t) = E[X(\theta,t)]$.

Let $\mu(\theta) = E_t[X(\theta,t)]$.

Let M be the grand mean.

$M = E[\mu(\theta,t)] = E_\theta[\mu(\theta)]$.

In our case, θ and t are both discrete rather than continuous variables. We can observe X. M is known since we are dealing with relativities compared to the overall average. On the other hand $\mu(\theta,t)$ is unknown and can never be observed directly.

We can observe the squared error that results from using different estimations. This squared error can be usefully expressed in another form. To do so, we split the variance of X into various pieces. Define

$$\delta^2 = E_\theta[E_t[E[(X(\theta,t) - \mu(\theta,t))^2|\theta,t]]]$$

$$\zeta^2 = E_\theta[E_t[(\mu(\theta,t) - \mu(\theta))^2|\theta]]$$

$$\zeta^2\ell(s) = E_\theta[E_t[COV[X(\theta,t),X(\theta,t+s)]|\theta]]$$

$$= E_\theta[E_t[COV[\mu(\theta,t),\mu(\theta,t+s)]|\theta]]$$

$$\tau^2 = VAR_\theta[E_t[\mu(\theta,t)]] = VAR_\theta[\mu(\theta)]$$

Then δ^2 is the process variance excluding any impact of shifting risk parameters over time. ζ^2 is the variance due to shifting parameters over time. $\ell(s)$ is a correlation measuring how much the risk parameters shift over time. $\ell(0) = 1$. $\ell(s) \leq 1$ for $s > 0$.[1] τ^2 is the parameter variance, the variance between the different risks.

For later convenience of notation define

$$\delta^2(\theta,t) = E[(X(\theta,t) - \mu(\theta,t))^2|\theta,t]$$
$$\delta^2(\theta) = E_t[\delta^2(\theta,t)]$$
$$\zeta^2(\theta) = E_t[(\mu(\theta,t) - \mu(\theta))^2|\theta]$$
$$\ell(s,\theta) = E_t[COV[\mu(\theta,t), \mu(\theta,t+s)]] + \zeta^2(\theta)$$

then

$$\delta^2 = E_\theta[\delta^2(\theta)] = E_{\theta,t}[\delta^2(\theta,t)]$$
$$\zeta^2 = E_\theta[\zeta^2(\theta)]$$
$$\ell(s)\zeta^2 = E_\theta[\ell(s,\theta)\zeta^2(\theta)]$$

It is useful to rearrange the definitions of the variances in the usual manner so as to express the expected value of a quantity squared as the sum of a squared mean and a variance.

$$E[X^2(t,\theta)|t,\theta] = \mu^2(t,\theta) + \delta^2(\theta,t)$$
$$E_t[\mu^2(t,\theta)] = \mu^2(\theta) + \zeta^2(\theta)$$
$$E_\theta[\mu^2(\theta)] = M^2 + \tau^2$$

A similar expression can be derived from the definition of the covariance.

$$E_t[\mu(t,\theta)\mu(t+s,\theta)] = \mu^2(\theta) + \ell(s,\theta)\zeta^2(\theta)$$

For the formula for the expected value of the squared error of the estimate from the observation, one needs to express various expected values in terms of the variances and correlations defined above.

[1] One should note that it is an assumption that this correlation depends only upon the separation of the two years in question. Whether or not this is a reasonable approximation to reality is an empirical question which depends on the particular application.

$$\begin{aligned}
E_{t,\theta}[X^2(t,\theta)] &= E_\theta[E_t[E[X^2(t,\theta)|t,\theta]]] \\
&= E_\theta[E_t[\mu^2(t,\theta) + \delta^2(\theta,t)]] \\
&= E_\theta[\mu^2(\theta) + \zeta^2(\theta) + \delta^2(\theta)] \\
&= M^2 + \tau^2 + \zeta^2 + \delta^2
\end{aligned}$$

$$\begin{aligned}
E_{t,\theta}[X(t,\theta)X(t+s,\theta)] &= E_\theta[E_t[E[X(t,\theta)X(t+s,\theta)|t,\theta]]] \\
&= E_\theta[E_t[\mu(t,\theta)\mu(t+s,\theta)]] \\
&= E_\theta[\mu^2(\theta) + \ell(s,\theta)\zeta^2(\theta)] \\
&= M^2 + \tau^2 + \ell(s)\zeta^2
\end{aligned}$$

$$E_{t,\theta}[MX(t,\theta)] = ME[X(t,\theta)] = M^2$$

Then it follows that:

$$\begin{aligned}
E_{t,\theta}[(X(t,\theta) - X(t+s,\theta))^2] &= E_{t,\theta}[X^2(t,\theta)] + E_{t,\theta}[X^2(t+s,\theta)] \\
&\quad - 2E_{t,\theta}[X(t,\theta)X(t+s,\theta)] \\
&= M^2 + \tau^2 + \zeta^2 + \delta^2 + M^2 + \tau^2 \\
&\quad + \zeta^2 + \delta^2 - 2(M^2 + \tau^2 + \ell(s)\zeta^2) \\
&= 2\delta^2 + 2\zeta^2(1 - \ell(s))
\end{aligned}$$

$$\begin{aligned}
&E_{t,\theta}[(X(t,\theta) - X(t+s,\theta))(X(t+u,\theta) - X(t+s,\theta))] \\
&\quad = E_{t,\theta}[X(t,\theta)X(t+u,\theta)] + E_{t,\theta}[X^2(t+s,\theta)] \\
&\qquad - E_{t,\theta}[X(t+s,\theta)X(t+u,\theta)] - E_{t,\theta}[X(t,\theta)X(t+s,\theta)] \\
&\quad = M^2 + \tau^2 + \ell(u)\zeta^2 + M^2 + \tau^2 + \zeta^2 + \delta^2 - (M^2 + \tau^2 \\
&\qquad + \ell(s-u)\zeta^2) - (M^2 + \tau^2 + \ell(s)\zeta^2) \\
&\quad = \delta^2 + \zeta^2(1 + \ell(u) - \ell(s-u) - \ell(s))
\end{aligned}$$

$$\begin{aligned}
&E_{t,\theta}[(X(t,\theta) - X(t+s,\theta))(M - X(t+s,\theta))] \\
&\quad = M^2 - M^2 + (M^2 + \tau^2 + \zeta^2 + \delta^2) \\
&\qquad - (M^2 + \tau^2 + \ell(s)\zeta^2) \\
&\quad = \delta^2 + \zeta^2(1 - \ell(s))
\end{aligned}$$

$$E_{t,\theta}[(M - X(t,\theta))^2] = M^2 - 2M^2 + (M^2 + \tau^2 + \zeta^2 + \delta^2)$$
$$= \delta^2 + \zeta^2 + \tau^2$$

These results are used in Appendix C.

It is of interest to note that variance of $X = E_{t,\theta}[(M - X(t,\theta))^2] = \delta^2 + \zeta^2 + \tau^2$. This is the split of the variance of X into three pieces that was discussed above.

Let $C(s)$ = Covariance for data for the same risk, Δ years apart. Then for $s > 0$

$$C(s) = E[(X(t,\theta) - \mu(\theta))(X(t+s,\theta) - \mu(\theta))]$$

$$C(s) = E[X(t,\theta)X(t+s,\theta)] - E[X(t,\theta)\mu(\theta)] - E[X(t+s)\mu(\theta)] + E[\mu^2(\theta)]$$
$$= M^2 + \tau^2 + \ell(s)\zeta^2 - (M^2 + \tau^2) - (M^2 + \tau^2) + M^2 + \tau^2$$
$$= \ell(s)\zeta^2$$

$$C(0) = E[(X(t,\theta) - \mu(\theta))^2] = E(X^2(t,\theta)] - 2E[X(t,\theta)\mu(\theta)] + E[\mu^2(\theta)]$$
$$= M^2 + \tau^2 + \zeta^2 + \delta^2 - 2(M^2 + \tau^2) + M^2 + \tau^2$$
$$= \zeta^2 + \delta^2$$

It is worth noting that the covariance structure assumed herein differs from that in Gerber and Jones [16]. The covariance structure which in Gerber and Jones is shown to give credibility formulas of the updating variety[2] can be written as:

$$COV[X_i,X_j] = \begin{cases} W_i & i < j \\ W_i + V_i & i = j \end{cases}$$

That covariance structure would assume for example that the covariance of the 1940 data with the data for each of the years earlier than

[2] Credibilities of the updating variety are such that new estimate = (prior estimate × complement of credibility) + (new data × credibility). This is the form of the estimate discussed in Section 9.1.

1940 is the same. In fact we observe that the distance between the years has an extremely significant impact on the covariance between the years.

The covariance structure assumed here can be written as:

$$\mathrm{COV}[X_i,X_j] = \begin{cases} \ell(j-i)\zeta^2 & i < j \\ \zeta^2 + \delta^2 & i = j \end{cases}$$

Thus the optimal least squares credibilities that result from the matrix equations that are given in Appendix C will generally not be of the updating variety.[3]

We can directly estimate only the following quantities from the data: τ^2, C(0), C(1), C(2), etc. Not coincidentally, these are the quantities that enter into the formula in Appendix C for the squared error. Thus, these are also the quantities that enter into the calculation of the optimal credibilities.

Thus, it is not necessary to estimate δ^2 by itself. However, if one does so, the values for ζ^2 and $\ell(i)$ follow. We will estimate δ^2 here solely in order to aid our understanding; it does not affect any of the calculated values of the credibilities.[4]

For a binomial process, with a success rate of .4 or .6, the variance is $.24n$.[5] This is approximately the variance for the average risk in this example, with $n = 150$.[6] The resulting variance of games lost is (150)(.24). The variance in losing percentage is $(150)(.24)/(150)^2 = .0016$.

Thus a reasonable approximate value for δ^2 is .0016. The values for the variances and correlations are shown in Table D1. It should be noted that as the difference in years increases, the correlations get close to zero.

For example, the observed value for the NL data for $\delta^2 + \zeta^2 = .007892$. Thus since we assume $\delta^2 = .001600$, we estimate $\zeta^2 = .006292$. The observed value of $\zeta^2\ell(1) = .004919$. Thus we estimate $\ell(1) = .004919/.006292 = .782$. For this example, the observed value of $\tau^2 = .001230$.

[3] They will be of the updating variety when $\ell(s) = 1$ for all s.

[4] In general if something cannot be observed in the squared errors, then it is not needed to calculate the optimal least squares credibilities.

[5] The variance is $p(1-p)n$.

[6] Teams played about 150 games per year over this period of time.

It is important to note that the total variance of the observations is equal to $\delta^2 + \zeta^2 + \tau^2 = .009122$. Thus, what has been done here is just an analysis of variance, breaking the variance into its various sources. For this example, about 13.5% of the variance of the observation is due to the differences between the risks, about 17.5% is due to the process variance, and about 69.0% is due to shifting parameters over time.

One can verify that the observed pattern in the covariance structure in Table D1 is not due solely to random chance. One can rearrange the data in random fashion, and observe the covariances.

TABLE D1

COVARIANCE STRUCTURE

	NL	AL
τ^2	.001230	.001619
δ^2*	.001600	.001600
ζ^2**	.006292	.006275
$\ell(0)$***	1.000	1.000
$\ell(1)$	.782	.721
$\ell(2)$	.543	.506
$\ell(3)$	.497	.384
$\ell(4)$	.404	.283
$\ell(5)$	.288	.124
$\ell(6)$	.249	.061
$\ell(7)$	.158	−.016
$\ell(8)$	.062	−.089
$\ell(9)$	−.012	−.170
$\ell(10)$	−.063	−.140

* δ^2 estimated as .001600 based on an assumed binomial process.

** ζ^2 is based on the assumed value of δ^2 and the observed value of $\zeta^2 + \delta^2$.

*** $\ell(0)$ is unity by definition.

First one can rearrange the entries in each row of Table 1; for each row separately, assign each entry in that row to a randomly selected column. Similarly one can rearrange the entries in each column of Table 1; for each column separately, assign each entry in that column to a randomly selected row. The resulting covariances that are computed for these two "scrambled" data sets are shown in Table D2. All of the covariances $\ell(i)$, $i > 0$ are close to zero. Therefore, one can conclude that there is a significant pattern being displayed in Table D1.

TABLE D2

COVARIANCE STRUCTURE, SCRAMBLED DATA

	NL Entries in Each Row Rearranged	NL Entries in Each Column Rearranged
τ^2	.000191	.001230
δ^2*	.001600	.001600
ζ^2**	.007330	.006292
$\ell(0)$***	1.000	1.000
$\ell(1)$	.010	−.117
$\ell(2)$	−.009	.021
$\ell(3)$	.008	−.070
$\ell(4)$	−.084	−.035
$\ell(5)$	−.025	−.039
$\ell(6)$	−.020	−.006
$\ell(7)$	−.030	−.053
$\ell(8)$	−.058	.082
$\ell(9)$	.049	.091
$\ell(10)$	.042	−.019

* δ^2 estimated at .001600 based on an assumed binomial process.
** ζ^2 is based on the assumed value of δ^2 and the observed value of $\zeta^2 + \delta^2$.
*** $\ell(0)$ is unity by definition.

APPENDIX E

PUTTING THE REDUCTION IN SQUARED ERROR IN CONTEXT

The first criterion used to determine the optimal credibility is to minimize the squared error. Using the optimal credibility based on this criterion will reduce the squared error between the observed and predicted result. What should be considered a significant reduction in squared error?

Let us examine an example. For the NL data set, using one year of data, the optimal credibility is 68% as shown in Table 9. As shown in Table 6 the mean squared errors are:

Z	Mean Squared Error
0	.0091
68%	.0049
100%	.0059

In this case, by the use of credibility, the squared error has been reduced from .0059 if the data were relied upon totally, or .0091 if the data were totally ignored, to .0049. In this case, the squared error has been reduced to 83% (.0049/.0059) of its previous value.[1]

All of these squared errors include the variation of the observed results around the expected value.[2] The use of credibility does not affect this source of variation. Thus credibility methods cannot reduce the squared error between the observed value and the estimated/predicted value to as great an extent as they reduce the squared error between the true mean and the estimated/predicted mean.[3]

It is shown in Mahler [9] that the best that can be done using credibility to combine two estimates is to halve the mean squared error between the estimated and theoretical true underlying mean. However,

[1] The "previous" value of the squared error is considered to be the minimum of the squared errors that result from either ignoring the data entirely or relying on the data entirely.

[2] This random variation is usually referred to as process risk.

[3] It should be noted that the former squared error is concrete and easily observed, while the latter squared error is theoretical and difficult if not impossible to observe.

in this paper the squared error being examined is between the estimated/predicted and the observed result, rather than the true underlying mean. This squared error is inherently larger due to the random variation in the observed result. Also the result derived in Mahler [9] was derived in the absence of shifting parameters over time.

It turns out that, in the current case, the best that can be done using credibility to combine two estimates is to reduce the mean squared error between the estimated and observed values to 75% of the minimum of the squared errors from either relying solely on the data or ignoring the data.[4] One can think of half[5] of the squared error as being due to two sources: the inherent process variance associated with comparing to observed results, and the presence of shifting parameters over time. This portion of the squared error is independent of the value chosen for the credibility. The remainder of the squared error can be thought of as that which is affected by the choice of the value of credibility; as stated above this can be at most cut in half by the use of credibility methods. If half of the squared error is cut in half, this reduces the total squared error to 75% of what it was.

Assume one is estimating the future by credibility weighting together a single year of data and the grand mean.[6] Let $V(0)$ be the squared error between the predicted and observed results for $Z = 0$. Let $V(1)$ be the squared error between the predicted and observed results for $Z = 1$. Then as is shown in Appendix F:

Z	Squared Error Between Predicted and Observed
0	$V(0)$
Optimal	$V(1)\left(1 - \frac{V(1)}{4V(0)}\right)$
100%	$V(1)$

with the optimal credibility given by: Z optimal $= 1 - V(1)/2V(0)$.

[4] When using more than two or more years of data, the reduction in squared error depends on the impact of shifting parameters over time. However, in the absence of shifting parameters over time, for N years with the same weight applied to each year, the maximum possible reduction is $1/(2(N + 1))$.

[5] This is only a half for the case when the squared errors for $Z = 0$ and $Z = 1$ are equal. However, this is the case when one gets the maximum reduction in squared error.

[6] The formula given below does not hold when using several years of data.

In the example above, we had $V(0) = .0091$, $V(1) = .0059$. Using these values in the above formula gives Z optimal = 68%, equal to the empirically determined 68%. The formula for the minimum squared error gives a value of .0049, which is equal to the empirical minimum squared error. The reduction of the squared error to 83% of its previous value appears significant in light of the maximum possible reduction to 75%.[7]

[7] The maximum reduction is possible when the squared errors for $Z = 0$ and $Z = 1$ are equal.

APPENDIX F

SQUARED ERRORS

In Appendix C, the fundamental formula for the squared error was derived:

$$V(Z) = \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j (\delta^2 \delta_{ij} + \tau^2 + \zeta^2 \ell(|i - j|)) - 2 \sum_{i=1}^{N} Z_i (\tau^2 + \zeta^2 \ell(N + \Delta - i)) + \delta^2 + \zeta^2 + \tau^2.$$

One can actually check this result against the observed squared errors.[1] For example, let $N = 2$ and $\Delta = 3$. Then

$$V(Z_1, Z_2) = Z_1^2(\delta^2 + \tau^2 + \zeta^2) + 2Z_1Z_2(\tau^2 + \zeta^2\ell(1)) + Z_2^2(\delta^2 + \tau^2 + \zeta^2) - 2Z_1(\tau^2 + \zeta^2\ell(4)) - 2Z_2(\tau^2 + \zeta^2\ell(3)) + \delta^2 + \zeta^2 + \tau^2$$

Using the average of the NL and AL values in Table D1 for the covariance structure:

$$\tau^2 = .001425 \qquad \delta^2 + \zeta^2 = .007884$$

$$\zeta^2\ell(1) = .004723 \qquad \zeta^2\ell(3) = .002770 \qquad \zeta^2\ell(4) = .002158$$

$$V(Z_1, Z_2) = Z_1^2(.009309) + Z_1Z_2(.012296) + Z_2^2(.009309) - Z_1(.007166) - Z_2(.008390) + .009309$$

Table F1 contains the results of the test for various values of Z_1 and Z_2. (Z_1 is the credibility applied to the less recent year of the two.) The mean squared errors are a close match to those given by the equation.[2]

[1] The covariances were estimated from the same data as is being used to test the equation for the squared error. Thus, the magnitude of the covariances is not being tested. However, the validity of the assumed form of the covariance structure as well as the validity of the derivation of the equation for $V(Z)$ are being tested.

[2] The differences are largely due to the fact that at the two ends of the data period there are either no predictions or no actual observation to enter into the computation of an error.

When $N = 1$, one gets the following parabola for $V(Z)$:

$$V(Z) = Z^2(\delta^2 + \tau^2 + \zeta^2) - 2Z(\tau^2 + \zeta^2\ell(\Delta)) + \delta^2 + \zeta^2 + \tau^2$$

$V(0) = \delta^2 + \tau^2 + \zeta^2 =$ squared error ignoring the data

$V(1) = 2\delta^2 + 2\zeta^2(1 - \ell(\Delta)) =$ squared error relying solely on the data

$$Z \text{ optimal} = \frac{\tau^2 + \zeta^2\ell(\Delta)}{\tau^2 + \delta^2 + \zeta^2} = \frac{V(0) - V(1)/2}{V(0)} = 1 - \frac{V(1)}{2V(0)}$$

$$V(Z \text{ optimal}) = -\frac{(\tau^2 + \zeta^2\ell(\Delta))^2}{\tau^2 + \delta^2 + \zeta^2} + \delta^2 + \zeta^2 + \tau^2$$

$$= -\frac{(V(0) - V(1)2)^2}{V(0)} + V(0)$$

$$= -V(0) + V(1) - \frac{V(1)^2}{4V(0)} + V(0)$$

$$= V(1)\left(1 - \frac{V(1)}{4V(0)}\right)$$

This is the result referred to in Appendix E. The reduction in mean squared error is greatest when $V(1) = V(0)$; then the squared error is reduced to 75% of the minimum of the squared errors that result from relying solely on the data or ignoring the data.

In the absence of shifting parameters over time,[3] the estimate improves as one uses more and more years of data. For large N, relying solely on the data produces a very good estimate; this is reflected in the fact that the optimal credibility approaches 1 as N gets large. Thus for large N, one cannot reduce the squared error significantly by using credibility.

[3] In the presence of shifting parameters over time the situation is much more complicated.

TABLE F1

MEAN SQUARED ERRORS (.0001)

Z_1	Z_2	Observed	Estimated by 2nd Order Polynomial
0	0	9,182	9,309
0	.25	7,592	7,793
.25	0	7,963	8,099
0	.5	7,202	7,441
.15	.35	7,087	7,293
.25	.25	7,172	7,352
.5	0	7,949	8,053
0	.75	8,011	8,253
.25	.5	7,581	7,769
.5	.25	7,957	8,057
.75	0	9,140	9,171
0	1	10,020	10,228
.25	.75	9,189	9,349
.5	.5	9,165	9,260
.75	.25	9,947	9,961
1	0	11,536	11,452
.75	.75	15,162	15,031
1	1	25,162	24,667

Note: Mean Squared Errors in estimating NL and AL data. $N = 2$, $\Delta = 3$. Estimate uses data from the fourth and third years prior to the estimation period with weights Z_1 and Z_2, respectively, and the complement of credibility applied to the grand mean. $Z_1 = 15\%$ and $Z_2 = 35\%$ is the solution to equation 11.3 for the least squares credibility.

The exact behavior can be derived using the results of Appendix C. In the absence of shifting parameters over time ($\zeta^2 = 0$), and applying equal weight Z/N to each of N years, based on the result in Appendix C, the squared error is given by:

$$V(Z) = Z^2\left(\frac{\delta^2}{N} + \tau^2\right) - 2Z\tau^2 + \delta^2 + \tau^2$$

$$V(0) = \delta^2 + \tau^2$$

$$V(1) = \delta^2\left(\frac{N+1}{N}\right)$$

$$Z \text{ optimal} = \frac{N\tau^2}{N\tau^2 + \delta^2} = \frac{(N+1)V(0) - NV(1)}{(N+1)V(0) - (N-1)V(1)}$$

$$V(Z \text{ optimal}) = \delta^2 + \tau^2 - \frac{\tau^4 N}{N\tau^2 + \delta^2}$$

$$= V(1)\left(1 - \frac{V(1)}{(N+1)^2V(0) - (N^2-1)V(1)}\right)$$

The maximum reduction in squared error compared to the minimum of $V(0)$ and $V(1)$ occurs when $V(0) = V(1)$. For this case

$$Z \text{ optimal} = 1/2$$

$$V(Z \text{ optimal}) = V(1)\left(1 - \frac{1}{2(N+1)}\right)$$

As N gets large, there is no significant reduction in squared error due to using credibility (in the absence of shifting parameters over time).

APPENDIX G

THE SECOND CRITERION AND LIMITED FLUCTUATION CREDIBILITY

The second criterion in Section 7 deals with the probability that the observed result will be more than a certain percent different than the predicted result. The less this probability, the better the solution.

This is related to the basic concept behind "classical" credibility which has also been called "limited fluctuation" credibility [7]. In classical credibilty, the full credibility criterion is chosen so that there is a probability, P, of meeting the test, that the maximum departure from expected is no more than k percent.

The reason the criterion is stated in this way rather than the way it is in classical credibility is that, unlike the actual observations, one cannot observe directly the inherent loss potential.[1]

However, the two concepts are closely related. If there is a small chance of the estimate differing by a large amount from the true value of the inherent loss potential, then, since the observed values are distributed about the true value, the chance of the estimate differing by a large amount from the observed value will be smaller than it would otherwise be.

For example, assume the inherent loss potential is .550 and that the observed values are distributed approximately normally with a standard deviation of .050. Then there is approximately a 95% probability that the observed value will be between .452 and .648.[2]

Assume the estimated values are also approximately normally distributed about the inherent loss potential.[3] Assume a standard deviation of .028. Then there is a 95% chance that the estimate will be between .495 and .605, i.e., within 10% of the true inherent loss potential.

---

[1] It has been shown that the loss potential varies for a risk over time. Thus, it cannot be estimated as the average of many observations over time.

[2] The mean plus or minus 1.96 standard deviations.

[3] An unbiased estimator has the same expected value as the inherent loss potential.

The difference between the estimated value and the observed value will also be approximately normally distributed about zero.[4] The standard deviation is .057.[5] Thus, there would be a 95% chance that the absolute difference between the estimated and observed value will be less than .112. This corresponds to about a 95% chance that the estimated value will be within ±20% of the observed value.[6]

In a particular example, the result would depend on the relative size of the variances of the observations and the estimates. However, the smaller the variance in the estimates, the smaller the variance in the difference between the estimates and the observations. Thus the smaller the probability that the estimate and the true mean differ by a large amount, the smaller the probability that the estimate and the observation differ by a large amount.

[4] The sum or difference of two normal distributions is also a normal distribution. The new mean is the difference of the two means.

[5] The new variance is the sum of the two variances.

[6] .112 ÷ .550 = .204.