DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXII

AN ANALYSIS OF EXPERIENCE RATING

GLENN G. MEYERS
VOLUME LXXII

DISCUSSION BY HOWARD C. MAHLER

The author wishes to thank Robert Conger, Glenn Meyers, Stephen Philbrick, and Gary Venter for reading an earlier version of this discussion and providing helpful comments. I also wish to thank Lesley Phipps for typing this discussion.

1. INTRODUCTION

This paper is another valuable contribution by Glenn Meyers to the actuarial literature [1]. In it, the author analyzes many aspects of experience rating formulas. Mr. Meyers's paper contains a remarkable amount of material.

It can be divided into four parts, each of which would have been a useful paper on its own. His first two sections give an introduction to experience rating. His third section examines private passenger automobile merit rating data, illustrating a general result in credibility theory with important practical implications. Meyers's fourth, fifth, and sixth sections examine commercial lines experience rating in terms of a useful general concept which Meyers has called efficiency. His seventh section gives a generally applicable method of applying statistical tests to choose the most appropriate form of an experience rating plan.

Although I will concentrate my discussion on certain portions of Mr. Meyers's paper, this in no way reflects upon the importance of the other portions of this paper. Rather, it reflects the large amount of significant material Mr. Meyers has presented, and the inability of this author to analyze it all thoroughly in a single discussion of tractable length.

Section 2 of this discussion concerns Meyers's discussion of the Bailey-Simon results [2]. Meyers proposes an explanation for the observed credibilities based on parameter uncertainty. I also discuss two other similar phenomena—risk heterogeneity and shifting parameters over time. Section 3 of this discussion presents simple examples of the phenomena discussed in Section 2.

Section 4 of this discussion summarizes Meyers's fourth section on the efficiency criterion.

Section 5 of this discussion presents the Bühlmann credibility result for a split experience rating plan.[1] It gives the general formulas to use to assign credibility to the primary and excess losses so as to maximize efficiency.

In Section 6 of this discussion, the formulas derived in Section 5 are used to analyze Meyers's General Liability example. Among the important points discussed is the use of credibilities other than the optimal credibilities from Section 5.

Section 7 of this discussion continues that analysis in more detail. The loss in efficiency due to the use of other than the optimal credibilities is shown to be small for this example. Also, the effects of the choices of different loss limits is explored.

Section 8 of this discussion points out that under certain circumstances it is theoretically valid to have a self-rating point.

Section 9 of this discussion contains the conclusions I draw from my analysis of Meyers's General Liability example.

Section 10 of this discussion analyzes Meyers's Workers' Compensation example. The analysis of the multi-split plan parallels that of the General Liability single split plan. In addition, the multi-split plan is compared to a single split plan, and is found not to perform significantly better for this example.

Section 11 of this discussion summarizes Meyers's seventh section, which gives a generally applicable method of testing experience rating plans.

Section 12 of this discussion gives my conclusions. I believe that some of Mr. Meyers's conclusions do not follow from the work he presents even though they may well turn out to apply in many specific cases encountered in real world applications.

Following the main body of this discussion, there are ten appendices which provide the mathematical support. Appendix A contains two results for covariances which should be more widely known among actuaries. Appendix J contains an interesting example with continuous mixing functions. The other appendices should be of interest to those with a serious interest in credibility theory.

[1] The mathematical derivation is in Appendix F.

2. FORMULAS FOR THE CREDIBILITY

In the third section of his paper, Mr. Meyers discusses two formulas for the credibility. The first, Meyers's formula 3.2, is the usual Bayesian credibility formula

$$Z = \frac{N}{N + K}, \qquad K \geqq 0. \tag{2.1}$$

The second is Meyers's formula 3.3

$$Z = \frac{N}{JN + K}, \qquad J \geqq 1, K \geqq 0. \tag{2.2}$$

which the author derives assuming parameter uncertainty. (See Appendix B for a further discussion of this formula.)

2.1 Parameter Uncertainty and the Bailey-Simon Data

He goes on to see how well the two formulas fit data from the classic paper by Bailey and Simon on the credibility of a single private passenger car [2].[2] He estimates values of J and K from the credibility for one and two years of data. He finds that formula 2.2 does a better job of fitting the credibility observed for three years. In itself, this should not be surprising since formula 2.1 is a special case of formula 2.2, and the extra choice of parameter available should allow a better fit for formula 2.2. Nevertheless, the resulting fit for Classes 1 and 2 is quite impressive.[3] Even for the other classes the fit is a substantial improvement over that for formula 2.1. It should be noted that Class 1, with 3 million car years, has over ten times the data in any of the other classes.

There is an explanation for the poor fit of formula 2.2 to the Bailey-Simon data for Class 4; this same explanation applies, to a lesser extent, to Class 5. The key point is that one cannot have three clean years of experience unless one has been licensed for at least three years. Class 4 includes many drivers who have less than three years of driving experience. Those risks with one year of experience go into Merit Rating Class Y (clean for one year) if they are clean, and Merit Rating Class B (clean for less than one year) if they are not,

[2] This section of Mr. Meyers's paper constitutes a discussion of this remarkable paper by Bailey and Simon written over a quarter of a century ago.

[3] The definitions of the classes are given in the Bailey-Simon paper. Class 1 is Pleasure—No Male Operator under 25. Class 2 is Pleasure—Non-principal Male Operator under 25. Class 3 is Business Use. Class 4 is Unmarried Owner or Principal Operator under 25. Class 5 is Married Owner or Principal Operator under 25.

as explained in Wittick [3]. Both Merit Rating Class A (clean for three years) and Merit Rating Class X (clean for two years) contain no risks with only one year of experience. We expect drivers with only one year of experience to be worse than the average for Class 4. Thus Merit Rating Class A (clean for three years) for driving Class 4, will have a lower frequency than the average for driving Class 4, merely because all of its drivers have at least three years of experience. Thus when we compare it to the remainder of driving Class 4, the resulting Bailey-Simon credibility for three years of data is overstated. The same is true to a lesser extent for the Bailey-Simon credibility for two years of data.[4]

2.2 Practical Implications of Parameter Uncertainty

As noted by the author, formula 2.2 has a maximum credibility of $1/J$. Based on the fit to the Bailey-Simon data, this implies maximum credibilities between 7% and 13%.[5] This implies that no private passenger automobile merit rating scheme can ever attain extremely large credits regardless of how many years of data are used. More generally, when parameter uncertainty is present ($J>1$), then the maximum credibility is less than 100%.

If formula 2.2 holds, the law of diminishing returns sets in very quickly. Using Mr. Meyers's parameters, roughly two-thirds[6] of the theoretical maximum credibility has been achieved using three years of data.

2.3 Shifting Parameters Over Time

An important conceptual distinction should be made between adding up separate units during the same time period (e.g., a large commercial risk) and adding up different years of experience (e.g., a private passenger automobile merit rating plan). While similar formulas might fit the observations in both cases, they do not have exactly the same meaning.

---

[4] This problem, which applies to an analysis of many merit rating plans, could have been avoided if it were possible to remove from the data all risks for which the insured and/or principal operator has been licensed for less than three years.

[5] If a more refined class plan were used, the credibilities would be lower. If the number of accidents rather than just the number of years since the last accident were taken into account, the credibilities would differ. If severity were taken into account, the credibilities would differ. The credibilities will differ depending on whether just accidents or accidents and convictions are taken into account.

[6] The value differs by class. It is 62% for Class 1, and 75% or greater for the other classes.

There are other similar phenomenon which, when important, cause formula 2.1 to no longer apply. One phenomenon is the shifting of parameters over time, which is discussed briefly by both Bailey-Simon and Meyers. Bailey and Simon put this forward as one possible explanation for the observation that extra years of data add relatively little credibility. "It can be fully accounted for only if an individual insured's chance for an accident changes from time to time within a year and from one year to the next, or if the risk distribution of individual insureds has a marked skewness reflecting varying degrees of accident proneness."[7,8]

In Appendix C, a formula is derived for the credibility when the parameters shift over time.[9] The exact solution is complicated for $N \geqq 3$. However, the following formula is approximate for $N = 3$, and exact for $N = 1$ or $N = 2$. (For $N \geqq 3$ this formula produces credibilities slightly too high.)

$$Z = \frac{\rho^{\Delta} \sum_{i=1}^{N} \rho^{i-1}}{\left(\sum_{i=1}^{N} \rho^{i-1} \right) + K} \tag{2.3}$$

where $\rho \leqq 1$ is the covariance between the risk processes one year apart[10] and Δ is the time between the mid-point of the last year of experience used in the rating and the mid-point of the policy year to which the rating will be applied.[11]

[7] Bailey and Simon explain in their subsequent paper [4] that what they meant by "marked skewness" leads to formula 2.1.

[8] Bailey and Simon also put forward as a partial explanation the fact that risks enter and leave the various classes. In addition, their use of a premium basis for frequency does not completely eliminate the maldistribution that would result from the use of an imperfect exposure base, as pointed out in the discussion by Hazam [5]. Finally, the Bailey-Simon credibilities are estimated by only looking at the indicated claims-free discounts. In contrast, the optimal credibility is a least squares fit to the Bayesian result for all the observed levels of claims.

[9] A simple assumption is made to quantify the impact of the shift. Other assumptions could be made which lead to other formulas. However, the basic idea remains, if the parameters shift over time, then data from far in the past can be of minimal value in predicting the future.

[10] ρ would capture some aspects that might be considered to be due to parameter uncertainty.

[11] Typically, $\Delta = 2$ for workers' compensation, and $\Delta = 1$ for private passenger automobile merit rating.

We see that the $\sum_{i=1}^{N} \rho^{i-1} \leqq N$ has replaced N in formula 2.1. Also, there is a maximum credibility of

$$\frac{\rho^{\Delta}}{1 + K(1 - \rho)} . \tag{2.4}$$

For ρ considerably less than one, adding more years of data quickly reaches the point of having no practical advantage.

Using the Bailey-Simon credibilities for one and two years of data, we can solve for the parameters in formula 2.3 ($\Delta = 1$). The results are

TABLE 2.1

			Three Year Credibility	
Class	K	ρ	Predicted	Observed
1	10.9	.55	8.0%	8.0%
2	7.7	.39	6.5%	6.8%
3	6.9	.40	7.4%	8.0%
4	3.0	.28	8.7%	9.9%
5	8.6	.37	5.5%	5.9%

Formula 2.3 produces a good fit for Class 1, a fair fit to Classes 2, 3, and 5, but it is unacceptable for Class 4. Formula 2.2 does considerably better for Classes 2 and 3. As has already been explained, we do not expect a good fit for Classes 4 and 5. The maximum credibilities indicated range from 8% to 13%, roughly the same range as indicated by formula 2.2.

One can use all three years of data in an attempt to estimate the parameters in either formula 2.2 or formula 2.3. Using a least squares fit, the results for formula 2.2 are given in Table 2.2 and for formula 2.3 in Table 2.3. We note that overall the fits of the formulas to the Bailey-Simon data, which is reproduced for convenience in Table 2.4, are as good as can be expected given the nature of the data. While the assumptions behind formula 2.3 seem more applicable to the situation here, formula 2.2 does at least as good a job of fitting the observed data. We note that the indicated maximum credibilities ($N = \infty$) are consistently lower for formula 2.3 than for formula 2.2.

TABLE 2.2

FORMULA 2.2 FIT TO THE DATA IN TABLE 2.4

Class	J	K	$N = 1$	$N = 2$	$N = 3$	$N = \infty$
1	7.88	13.78	4.62%	6.77%	8.02%	12.7%
2	10.96	11.30	4.49	6.02	6.79	9.1
3	9.00	10.85	5.04	6.93	7.93	11.1
4	8.30	6.06	6.96	8.83	9.69	12.0
5	12.37	14.33	3.75	5.12	5.83	8.1

TABLE 2.3

FORMULA 2.3 ($\Delta = 1$) FIT TO THE DATA IN TABLE 2.4

Class	J	ρ	$N = 1$	$N = 2$	$N = 3$	$N = \infty$
1	11.14	.557	4.59%	6.83%	8.00%	9.4%
2	8.61	.428	4.45	6.09	6.75	7.2
3	8.47	.473	4.99	7.01	7.89	8.7
4	4.51	.381	6.91	8.93	9.63	10.0
5	11.09	.448	3.71	5.17	5.80	6.3

TABLE 2.4

EMPIRICAL CREDIBILITIES FROM BAILEY-SIMON PAPER

Class	$N = 1$	$N = 2$	$N = 3$
1	4.6%	6.8%	8.0%
2	4.5	6.0	6.8
3	5.1	6.8	8.0
4	7.1	8.5	9.9
5	3.8	5.0	5.9

The extent to which parameters actually shift over time for any given line of insurance is an important empirical question worthy of further investigation. One would examine the correlations between years of data separated from each other by different time spans. One would also examine the empirical credibility of one year of data being used to predict some later year of data, for different separations between the two years.

The results of such an investigation should be quite useful in the design of experience rating plans. It would help to decide how many years of data should go into the plan. Also, it would help decide whether it is worthwhile, i.e. produces a significant increase in efficiency (as defined by Meyers), to give more weight to the more recent years of data.[12] It would also help in deciding what those relative weights should be.

2.4 Risk Heterogeneity

Another phenomenon is risk heterogeneity. In other words, a large risk may be made up of smaller risks. If we treat the smaller risks within a single large risk as independent observations from the same distribution we get the usual Bayesian formula 2.1. However, if smaller risks were grouped together in a totally random fashion to give larger risks, then there would be no increase in credibility between a small risk and a large risk. The actual situation is generally somewhere between those two extremes.

As shown in Appendix D, this would lead to a formula for credibility of the form

$$Z = \frac{N + I}{N + K}, \qquad 0 \leqq I \leqq K. \tag{2.5}$$

It should be noted that the value of K in formula 2.5 differs from that in formula 2.1. Formula 2.5 does not fit the Bailey-Simon data.

Formula 2.5 was derived for large risks. It would not apply for small risks, i.e., those too small to have separate and distinct subunits.[13] Specifically, no conclusion should be drawn from the fact that formula 2.5 has a minimum credibility of I/K.

[12] For example, the Massachusetts Private Passenger Automobile Safe Driver Insurance Plan currently gives less weight to older incidents via a so-called aging process.

[13] It should be noted that generally experience rating plans have an eligibility requirement which excludes very small risks.

If both parameter uncertainty and risk heterogeneity are important, as shown in Appendix D, the formula for credibility takes the form

$$Z = \frac{N + I}{JN + K}, \qquad 0 \leqq I \leqq K, J \geqq 1. \tag{2.6}$$

Formula 2.6 can be fit to the Bailey-Simon data. However, using three years of data to fit three parameters I, J, and K leaves no way to test the predictions.

Let M be a measure of the (average) size of the risk in each year. Let N be the number of years of data used for experience rating. Then if all three phenomena are taking place, we get the following formula in Appendix E:[14]

$$Z = \frac{\rho^{\Delta} \left(\sum_{i=1}^{N} \rho^{i-1} \right) (M + I)}{\left(\sum_{i=2}^{N} \rho^{i-1} \right) (M + I) + JM + K}. \tag{2.7}$$

For $\rho = 1$, formula 2.7 reduces to

$$Z = \frac{N(M + I)}{(N-1)(M + I) + JM + K}. \tag{2.8}$$

For $N = 1$, formula 2.8 reduces to formula 2.6, as it should.

2.5 *Conclusions*

I suspect that each of the three phenomena discussed is taking place to some extent. It would be worthwhile to obtain a more current set of private passenger automobile data that followed a risk for more than three years. Then one could determine the relative importance of the three phenomena. It would also be worthwhile to investigate the effects of these phenomena on other lines of insurance. For example, parameter uncertainty and risk heterogeneity would be expected to be particularly important for large commercial risks.

More generally, it would be worthwhile to determine empirically the credibility associated with each size of risk.[15] For such an investigation, identifying

[14] As stated previously, ρ captures certain aspects that might be labeled parameter uncertainty. Here J captures only those aspects of parameter uncertainty that relate to adding up subunits at the same point in time.

[15] The National Council on Compensation Insurance is currently doing so for workers' compensation.

the underlying causes would be helpful but not necessary. However, the above reasoning leads to useful candidates to check against the observed behavior with size of risk.[16]

3. AN EXAMPLE ILLUSTRATING PARAMETER UNCERTAINTY, SHIFTING PARAMETERS OVER TIME, AND RISK HETEROGENEITY

This example will try to illustrate what is meant by the three related but somewhat different concepts of parameter uncertainty, shifting parameters over time, and risk heterogeneity. The mathematics are developed and discussed in Appendices B, C, D, and E.

Assume we have the legendary little old lady from Pasadena who only uses her car to drive back and forth to church on Sundays. Let us ignore any seasonal variations in driving conditions. Further assume she always travels at the same time of day and always uses the same route.

One year actually consists of 52 observations of her claims process. If we treat these as 52 independent observations from the same distribution then we would get formula 2.1 for the credibility.

3.1 Parameter Uncertainty

However, we note that there are factors outside of her risk process that will vary her loss potential randomly, i.e., change the parameters of her risk process on any given day. To take one example, whether or not it is raining would affect her risk process.

Assume that there is a higher chance of an accident when it rains. Further, assume whether or not it rains during her trip is a random variable. Then we have a risk process such as described by Mr. Meyers's algorithm A.1 (ignoring, for simplicity, claim severity). This is an example of parameter uncertainty.

There is a kernel of uncertainty in the number of accidents she has in a year due to the variability caused by the different possible states of the universe. Extra observations will not reduce the effect of this kernel of uncertainty. This is why we get the lower credibilities indicated by formula 2.2. This is also why the maximum credibility is less than one.

3.2 Shifting Parameters Over Time

Let us change the example to illustrate risk parameters shifting over time.

[16] The table at the end of Appendix G gives a good list of such candidates.

(They shift in a definite direction, but we cannot predict beforehand in which direction.) Assume that the little old lady changes churches, and that her trip is significantly different (e.g., longer or shorter). Then her accident potential is different. Experience based on old data when she was driving to her old church is less useful for predicting her expected future claims experience driving to her new church. Thus the credibility assigned to it would be lower. The credibility would be given by formula 2.3. Once again, the maximum credibility is less than one. Many observations in a single year will not take this shift into account, and observations over a longer period of time include, of necessity, "stale" data.

3.3 Risk Heterogeneity

Finally, let us change the example to illustrate risk heterogeneity. Let us assume we have a classification of risks which is made up solely of cars which are only driven by little old ladies, who only use them to drive to church on Sundays.[17] Further, let us assume that in each case the car is jointly owned by two little old ladies. Finally, assume that they take turns driving, each one driving every other Sunday. In this case, the average process variance per car for the class remains the same as the case where each car was driven by one little old lady.[18]

However, since the distribution of the loss potential of cars has become more concentrated toward the mean, the variance between the cars making up the class would be less than in the case where each car was driven by one little old lady.[19] Thus, the claims data for a car is less credible than the similar situation where we had only one driver of each car. It would be even less credible if they alternated churches each Sunday, as well as drivers.

The formula for credibility when there were heterogeneous risks was given by formula 2.5. Another simple but illustrative example is given in Appendix D of this discussion.

[17] While this is clearly unrealistic, many class plans for private passenger automobiles do have a senior citizens class.

[18] This would not be the case if they flipped a coin in order to decide who did the driving each Sunday. In that case, the process variance would be greater. This can be usefully thought of as a case of parameter uncertainty.

[19] This would not be the case if for each pair of little old ladies who jointly own a car, the two drivers in each pair have the exact same loss potential as each other.

3.4 *Conclusions*

It should be noted that part of the difficulty in assigning credibilities to one car year of exposure is that a car year can mean considerably different things depending on how the car is used, how far it is driven, and how many drivers it has. Part of the purpose of experience rating is to make up for any such inadequacies in the exposure base or risk classification system.

The interested reader would probably find it useful to construct a similar example of his own for a large commercial risk. One could take a workers' compensation insured consisting of ten separate locations of equal size, and give examples of each of the three phenomena.

4. THE EFFICIENCY CRITERION

In Section 4 of his paper, Mr. Meyers defines the efficiency of an experience rating plan as the reduction in the expected squared error. (See formula 5.2 below.) The higher the efficiency, the more accurate the experience rating plan.

The author defines the efficiency so that it is never more than 100%.[20] The efficiency can only reach 100% if all the risks in the class have the same mean. Since classes are usually not perfectly homogeneous, the efficiency obtainable by any estimator is usually less than 100%. The author shows that the maximum efficiency using credibility is achieved when the credibility is equal to the Bühlmann (i.e., Bayesian credibility) result. For this case, the efficiency equals the credibility.

The author also shows that the efficiency as a function of the credibility is a parabola. Thus, even if the credibility used is not quite the Bühlmann result, there is still a substantial improvement in accuracy due to the use of credibility.[21] In the next section of his paper, the author shows how this general principle applies to the use of a self-rating point.

5. MAXIMIZING EFFICIENCY, PRIMARY AND EXCESS LOSSES

It is possible to generalize the Bühlmann result to the cases Meyers examines. Assume we have an experience rating plan, and our estimate of the mean

[20] The efficiency can be negative for a particularly poor choice of estimator.

[21] This pleasant and useful property of credibility estimates is explored in more detail in Mahler [6].

F is given by

$$F = (1 - Z_p)\,E_p + Z_pA_p + (1 - Z_e)\,E_e + Z_eA_e\,. \tag{5.1}$$

The subscripts p and e will stand for primary and excess; however, for now they can be treated as any two well defined portions of the total losses. E_p and E_e are the expected losses of each type. A_p and A_e are the actual losses of each type. Z_p and Z_e will be thought of as the credibilities assigned to each portion of the losses; however, for now they can be treated as just numbers to be determined.

In accordance with Meyers, define the efficiency of F by the expression

$$1 - \frac{E[(F - \mu)^2]}{E[(M - \mu)^2]}\,, \tag{5.2}$$

Where M is the grand mean, and μ is the mean for individual risks. In this case, $M = E_p + E_e$.

In order to maximize the efficiency, one must minimize $E[(F - \mu)^2]$. In Appendix F, Z_p and Z_e are determined so as to maximize the efficiency. Let:

$a =$ total variance of the primary losses
$b =$ total variance of the excess losses
$c =$ variance of the hypothetical means of the primary losses
$d =$ variance of the hypothetical means of the excess losses
$r =$ total covariance of the primary and excess losses
$s =$ covariance of hypothetical means of the primary and excess losses.

Then the optimum Z_p and Z_e are

$$Z_p = \frac{(c + s)b - (d + s)r}{ab - r^2}\,, \text{ and} \tag{5.3}$$

$$Z_e = \frac{(d + s)a - (c + s)r}{ab - r^2}\,. \tag{5.4}$$

It is interesting to note that if we set the primary losses equal to the total losses and thus the excess losses equal to zero, then the solution to the equations becomes

$$Z_p = \frac{c}{a}\,,$$

which is the usual expression for credibility, as in Meyers's equation 3.1.

However, unlike the usual case for credibilities, formulas 5.3 and 5.4 do not have the property of restricting Z_p or Z_e to the closed interval between zero and one. Thus, although it may merely be a matter of semantics, some caution is required before labelling Z_p and Z_e as credibilities. For simplicity of exposition, I will refer to them as credibilities, but perhaps a more precise term to apply would be weights.

The maximum efficiency that results from the optimal values of Z_p and Z_e given by formulas 5.3 and 5.4 is

$$\text{Maximum Efficiency} = \frac{Z_p(c + s) + Z_e(d + s)}{c + d + 2s} . \tag{5.5}$$

Thus, the maximum efficiency is a weighted average of the two credibilities that produce this maximum.[22]

In Appendix G, the dependence of the credibilities and efficiency on the size of risk is explored. One does not get the familiar formula 2.1 that we had for the no-split situation.[23]

6. SINGLE SPLIT PLANS

In Section 5 of his paper, Mr. Meyers illustrates the advantage of having a loss limit as per the General Liability single split plan. He does this by means of an example in which he assumes four types of risks.[24] It is useful to think of these risks as excellent, good, bad, and terrible. While this choice simplifies the computations, it still captures the essence of experience rating, which is to distinguish between risks to the extent that they are otherwise not distinguished by the class plan.

The claim count distribution is chosen as a binomial with N trials. N is used as a measure of the size of the insured. Once again this simplifies the computations, but captures the essential features. There are high and low frequency risks and the process variance increases linearly with N.

[22] The denominator is the variance of the hypothetical means of the total losses (primary plus excess).

[23] Of course as discussed above, one can make alternative assumptions, and get alternative formulas for the no-split situation, as for example formulas 2.2, 2.3, and 2.5.

[24] In Appendix J, an example is given of a continuous distribution of types of risks.

The severity distribution is chosen as a discrete version of the Shifted Pareto.[25] The use of the discrete version again simplifies the computations while maintaining the essential features. There are high and low severity risks. Most of the claims are small; however, the large claims contribute a large part of the mean and most of the variance.[26]

Mr. Meyers employs the Panjer algorithm to derive the aggregate loss distribution from the assumed frequency and severity distributions.[27] The reader should note that, for these simple examples, it is relatively simple to calculate the aggregate distributions directly via convolutions. Also, all of the calculations necessary to explore the behavior of the credibility results can be done from the separate frequency and severity distribution without first obtaining the aggregate loss distribution.[28]

6.1 Frequency versus Severity

Mr. Meyers looks at three different examples. In his first example, which is displayed in his Table 5.2, only the frequency distributions vary between the risks. In his second example, which is displayed in his Table 5.3, only the severity distributions vary between risks. In his third example, which is displayed in his Table 5.4, both the frequency and severity distributions vary between risks. In actual applications, which example is a better approximation to reality will depend on the relative importance of the variance between risks of the frequency and the variance between risks of the severity. For each example, Mr. Meyers displays the results of using Bayes Theorem as well as credibility. I will only discuss the results of using credibility.

Mr. Meyers points out the conflicting roles of the frequency and severity variances between risks in the choice of a loss limit. If only the frequency distributions vary, then the loss limit should be low. The assumption in Meyers's

---

[25] While the Pareto has been extensively used as a size of loss distribution for a group of risks, it is unclear whether or not the Pareto is an appropriate size of loss distribution for individual risks.

[26] In fact, for Mr. Meyers's choice of parameters, $q = 1.25$, the unlimited Pareto has an infinite variance.

[27] The Panjer algorithm is explained in Venter [7]. It is simpler than the Heckman-Meyers algorithm and designed to handle the case where one has a discrete severity distribution. The Heckman-Meyers algorithm is explained in Heckman and Meyers [8].

[28] In fact, they need only be done completely for $N = 1$, with the results for other values of N following from the results in Appendix G. Mr. Meyers's results for the use of Bayes Theorem do require that the aggregate loss distributions be calculated.

first example is that the size of a claim is completely random and does nothing to distinguish good and bad risks. On the other hand, if only the severity distributions vary, we want a higher loss limit. In Meyers's second example, we want to capture as much of the valuable information contained in the size of claim as is useful. When both the severity and frequency distributions vary, the optimal loss limit is somewhere between the results for the first two cases.

In his third example, Mr. Meyers takes the frequency and severity as highly correlated.[29] Those risks with a high mean frequency also have a high mean severity. Thus, although frequency and severity are assumed to be independent for a given risk, they cannot be treated as independent when looking at all risks combined.

This high correlation chosen by Mr. Meyers, as well as the particular choice of parameters, affects the particular results obtained. Thus, when attempting to apply Mr. Meyers's method of analysis to a particular real world situation, it is important to carefully choose those assumptions which most closely match that situation. With this caveat, the method of analysis should be widely applicable. I will analyze Meyers's third example extensively below in Section 7.

6.2 Basic Limits versus Total Limits

Sometimes the question that is asked is as important as the answer. Mr. Meyers poses the question in Section 5 of his paper as trying to maximize the efficiency, where only the error in predicting basic limits losses is considered in the efficiency. A useful extension might be to consider the error in predicting total limits losses,[30] which would produce different results. If you assume that each of these risks would receive the same increased limits factor, then one could explore the behavior of the efficiency for various total limits using the methods discussed below.

6.3 Primary and Excess Credibilities

Formulas 5.3 and 5.4 give the primary and excess credibilities which will maximize the efficiency. (However, these "credibilities" do not necessarily lie between zero and one.) It is possible to use other values for the credibility, but, of course, the efficiencies will be lower.

[29] In Meyers's Section 6, when examining the Workers' Compensation Experience Rating Plan, frequency and severity are taken as independent of each other.

[30] In the next section, Mr. Meyers deals with unlimited losses while exploring the features of the Workers' Compensation Experience Rating Plan.

As shown in Appendix F, if the excess variance is relatively large, as it is for this example, then it is a good approximation to the optimal credibilities to take

$$Z_p = \frac{c+s}{a}\text{ , and} \tag{6.1}$$

$$Z_e = 0\,.$$

In fact, the General Liability Plan sets $Z_e = 0$. Subject to that constraint, formula 6.1 gives the maximum efficiency. As shown in Appendix F, this is equivalent in the General Liability Plan to taking

$$Z = \left(\frac{c+s}{a}\right)\left(\frac{E_p}{E_p + E_e}\right)\,. \tag{6.2}$$

In actual application, formula 6.2 can lead to values greater than one. Thus, it might be more practical to employ

$$Z = \text{MIN}\left[1, \left(\frac{c+s}{a}\right)\left(\frac{E_p}{E_p + E_e}\right)\right]\,. \tag{6.3}$$

In his tests, Mr. Meyers uses

$$Z = \frac{c}{a}\,. \tag{6.4}$$

The ratio of the credibility given by formula 6.2 to that given by formula 6.4 is

$$\left(1 + \frac{s}{c}\right) \div \left(1 + \frac{E_p}{E_e}\right)\,.$$

For the assumptions used here, this is independent of N. For the Meyers example, this expression is greater than one, so that the credibilities given by formula 6.2 are larger than those given by formula 6.4.

In Appendix F, a formula is given for the loss in efficiency due to using formula 6.4 rather than formula 6.2.[31] As will be seen below for the cases explored by Mr. Meyers, the loss in efficiency is relatively small. Nevertheless, for certain applications, it may be significant.

[31] The relative loss in efficiency turns out to be independent of N.

7. MEYERS'S GENERAL LIABILITY EXAMPLE IN MORE DETAIL

Meyers displayed in more detail in Exhibit 5.1 the case of varying frequencies and severities when $N = 4$ with a loss limit of 4.[32] For this case we have $a = 4.744$, $b = 29.740$, $c = 1.026$, $d = .753$, $r = 4.752$, and $s = .874$.

Using formulas 5.3 and 5.4 gives $Z_p = 41.2\%$, and $Z_e = -1.1\%$, with a resulting efficiency of 21.7%.

Using formula 6.2 results in $Z = 26.3\%$ and an efficiency of 21.6%.[33] In this case $Z \leqq 1$, so formula 6.3 gives the same result as formula 6.2.

Formula 6.4 (used by Meyers) gives $Z = 21.6\%$ and an efficiency of 20.9%, which matches Meyers's result.[34]

The behavior observed by Mr. Meyers for credibilities in Table 5.4 can be explained in terms of formula 6.4 being an approximation to formula 6.3, which is in turn an approximation to formula 6.2, which in turn is an approximation to formulas 5.3 and 5.4, the true optimal credibility result.[35]

7.1 Results of the Various Formulas for Credibility

In Tables 7.1 to 7.4, I have calculated the equivalent of Meyers's Table 5.4 (frequency and severity both vary) for these various different credibility formulas. I have extended the tables to cover more values of N and loss limits.

[32] There are four prior distributions given equal weight. They have binomial parameters p of .2, .3, .4, and .5 respectively. They have first parameters of the Pareto, which Meyers calls b, of .25, .50, .75, and 1.00 respectively. They all have the second parameter of the Pareto, which Meyers calls q, equal to 1.25.

[33] Note that this is approximately equal to the Bühlmann credibility from formula 6.4 given below. Why this is the case is explained in Appendix H.

[34] The efficiency is not equal to the credibility as one might expect from Meyers's Appendix B, since we are measuring the error in predicting the basic limits losses rather than just the primary portion of the basic limits losses.

[35] Which is in turn a linear approximation to the optimal Bayesian result.

TABLE 7.1

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
PRIMARY AND EXCESS CREDIBILITIES AS PER FORMULAS 5.3 AND 5.4

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	19.9%	33.1%	49.6%	66.2%	79.4%	88.2%	93.4%	100.0%
2	22.7	37.0	53.9	69.9	82.1	90.0	94.5	100.0
2.5	22.8	37.1	54.1	70.1	82.2	90.1	94.6	100.0
3	22.5	36.7	53.6	69.7	82.0	90.0	94.6	100.0
4	21.7	35.6	52.5	68.8	81.4	89.6	94.4	100.0
6	20.0	33.3	49.9	66.5	79.9	88.7	93.9	100.0
8	18.5	31.3	47.6	64.5	78.4	87.8	93.5	100.0
12	16.4	28.2	44.0	61.1	75.9	86.2	92.6	100.0
16	15.0	26.1	41.4	58.5	73.8	84.9	91.8	100.0

TABLE 7.2

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.2

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	18.6%	31.4%	47.6%	64.3%	77.9%	87.1%	92.6%	98.8%
2	22.6	36.9	53.8	69.8	82.0	89.9	94.4	99.4
2.5	22.8	37.1	54.0	70.0	82.2	90.0	94.5	99.5
3	22.5	36.7	53.6	69.7	82.0	90.0	94.5	99.6
4	21.6	35.5	52.3	68.7	81.3	89.6	94.4	99.7
6	19.6	32.8	49.4	66.1	79.5	88.5	93.9	99.9
8	18.0	30.5	46.7	63.7	77.8	87.5	93.3	99.9
12	15.5	26.9	42.4	59.5	74.6	85.4	92.1	100.0
16	13.8	24.3	39.1	56.2	71.9	83.7	91.1	100.0

TABLE 7.3

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.3

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	18.6%	31.4%	47.6%	63.8%	73.2%	77.9%	80.3%	82.7%
2	22.6	36.9	53.8	69.8	80.8	86.3	89.1	91.9
2.5	22.8	37.1	54.0	70.0	81.6	87.7	90.7	93.7
3	22.5	36.7	53.6	69.7	81.8	88.9	91.7	95.0
4	21.6	35.5	52.3	68.7	81.3	89.0	92.8	96.6
6	19.6	32.8	49.4	66.1	79.5	88.5	93.3	98.1
8	18.0	30.5	46.7	63.7	77.8	87.5	93.1	98.8
12	15.5	26.9	42.4	59.5	74.6	85.4	92.1	99.4
16	13.8	24.3	39.1	56.2	71.9	83.7	91.1	99.7

TABLE 7.4

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.4, I.E. SHOULD MATCH MEYERS'S TABLE 5.4

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	15.6%	26.2%	39.8%	53.8%	65.2%	72.9%	77.5%	82.7%
2	20.9	34.1	49.7	64.5	75.8	83.1	87.2	91.9
2.5	21.5	34.9	50.9	65.9	77.4	84.7	89.0	93.7
3	21.4	35.0	51.1	66.5	78.2	85.8	90.2	95.0
4	20.9	34.4	50.7	66.5	78.8	86.8	91.4	96.6
6	19.3	32.2	48.5	64.9	78.1	87.0	92.2	98.1
8	17.8	30.1	46.2	62.9	76.9	86.5	92.2	98.8
12	15.4	26.7	42.1	59.2	74.2	85.0	91.6	99.4
16	13.8	24.2	38.9	56.0	71.7	83.4	90.8	99.7

7.2 Efficiency as a Function of Loss Limit

Table 7.4 here should match Meyers's Table 5.4.[36] We see the same behavior noted by Mr. Meyers. Taking larger loss limits gives higher efficiency only up to a point; then it decreases. The optimal loss limit increases with size of risk, as was noted by Mr. Meyers. Table 7.1, using primary and excess credibilities, shares the former feature, but not the latter feature.

The fact that, for fixed N, the efficiencies in Table 7.1 have a maximum somewhere in between $L = 0$ and $L = 50$ (the limit for basic limit losses in Meyers's example) is not surprising. If $L = 50$, then all of the losses are primary. If $L = 0$, then all of the losses are excess. In each case, the solution reduces to that of the no-split plan. Thus, the two endpoints have the same efficiency. For $0 < L < 50$, the special case where we restrain $Z_p = Z_e$ reduces to that of the no-split plan. Thus we know that the split plan allowing Z_p and Z_e to vary independently so as to get the maximum efficiency does at least as well as the no-split plan, which is just a special case.

Thus $0 < L < 50$ does at least as well as $L = 0$ or $L = 50$. In fact, it does better. The efficiency peaks somewhere between the endpoints and decreases as L approaches the two extremes. This behavior carries over to Tables 7.2, 7.3, and 7.4 which can be thought of as successive approximations getting further and further from the optimal results in Table 7.1

In Table 7.1, the optimal loss limit is about 2.5, independent of N.[37] This behavior carries over to Table 7.2. In Table 7.3, the optimal loss limit increases slowly with N, after $N = 32$. The reason is that we have restricted $Z \leqq 1$;[38] if formula 6.2 would indicate $Z > 1$, we set $Z = 1$ instead. Thus we are giving

[36] In fact, it does not for a loss limit of \$1000. Apparently, Mr. Meyers mistakenly set these efficiencies equal to those from his Table 5.2 where only the claim count distributions vary. The efficiencies here should be lower, even though the two cases are very similar. It is true that since the severity distributions are discrete in units of \$1000, choosing a loss limit of \$1000 means that all limited claims are of size \$1000. In other words, the experience rating plan ignores the size of claim in both cases for a loss limit of \$1000. In each case, the experience rating plan is explaining the same amount of variation, based solely on the observed difference in claim counts. However, in the case here, the total observed variation in losses is greater than when only the claim count distributions vary, since here the severity distributions also vary. Thus, a smaller proportion of the total observed variation is explained here. Therefore, the efficiency is lower here.

[37] The reason this is so for Table 7.2 is explained in Appendix H.

[38] This becomes applicable in the upper righthand corner of the table. For example, for $N = 64$ and a loss limit of 2.5, the credibility indicated by formula 6.2 is 109.1%.

less than the optimum weight to the primary losses. We can afford to raise the loss limit somewhat (staying within the range where formula 6.2 would indicate $Z \geqq 1$, thus formula 6.3 indicates $Z = 1$) so as to make the mean primary loss larger, in order to make up for the lessened weight that is being applied to the primary losses.

Finally, Table 7.4 has the optimum loss limit increase faster with N than does Table 7.3. This is so since formula 6.4 is close to formula 6.3, but in Meyers's example yields lower credibilities. Thus, once again, we raise the loss limit to make up for a too low weight applied to the primary losses.

In Table 7.3, we notice that, for a given loss limitation, after a certain point there is little increase in efficiency with increasing size of risk. This is explained in Appendix F.

8. SELF-RATING POINT

Formulas 6.2 and 6.3 for credibility illustrate the theoretical validity of a self-rating point. One can have $Z \geqq 1$. In fact, this is the case for Meyers's third example.

It makes sense to define the self-rating point as the smallest size risk such that $Z = 1$ at the optimum loss limit. Using formula 6.2 this means that

$$\frac{N^2\,\hat{c} + N^2\,\hat{s}}{N^2\,\hat{c} + N\,\hat{t}} \left(\frac{E_p}{E_p + E_e}\right) = 1, \tag{8.1}$$

whereas in Appendix G we let:

$$\hat{c} = \frac{c}{N^2},$$

$$\hat{s} = \frac{s}{N^2}, \text{ and}$$

$$\hat{t} = t/N = (a - c)/N.$$

Formula 8.1 can be solved for

$$N = \frac{\hat{t}\,(E_p + E_e)}{\hat{s}\,E_p - \hat{c}\,E_e} \tag{8.2}$$

In the Meyers example, we saw in Table 7.2 that the optimum loss limit is about 2.5. For this loss limit we have $\hat{c} = .041$, $\hat{s} = .054$, $\hat{t} = .548$, $E_p = .495$, and $E_e = .375$. Thus, formula 8.2 gives $N \cong 42$.

Thus, in this example, the self-rating point would be 42.[39] When using formula 6.3, we expect the optimum loss limit to increase above this self-rating point. This is confirmed by Table 7.3 where the optimum loss limit began to increase for $N \geqq 64$ after remaining constant for $N \leqq 32$.

9. CONCLUSIONS FROM MEYERS'S GENERAL LIABILITY EXAMPLE

Those features and assumptions of Meyers's third example that are probably true for most applications are:

(1) Both the frequency and severity distributions vary between risks (although perhaps not in the same relative importance as in this example).
(2) Frequency and severity are somewhat correlated (although probably not to the extent they are in the example).
(3) The excess losses have a much higher coefficient of variation than do the primary losses.
(4) The primary severity and excess severity are highly correlated.

Based on the above analysis of Meyers's example, when the general assumptions of his example hold, the behavior we expect to see for a single split experience rating plan is as follows:

(1) The optimum loss limit increases slowly as the size of risk gets larger up to the self-rating point.[40] (It shall be shown when examining Meyers's next example that the optimum loss limit remains virtually constant here because of the particular choice of parameters for this example.)
(2) The optimum loss limit increases more rapidly as the size of risk increases beyond the self-rating point.
(3) The more important the differences in severity, the higher the optimum loss limit. The more important the differences in frequency, the lower the optimum loss limit.
(4) The efficiency is very close to optimal for loss limits close to optimal.
(5) The optimal credibilities will not be of the form $N/(N + K)$, although such a formula will give efficiencies close to optimal. (This will not be true if any of the phenomena discussed in Sections 2 and 3 of this discussion are significant.)

---

[39] This corresponds in this example to expected basic limits losses of about \$37,000, or about 26 claims on average.

[40] In certain cases it may not be appropriate to have a self-rating point. For example, this will be the case if the phenomena discussed in Sections 2 and 3 of this discussion significantly reduce the credibilities.

10. MULTI-SPLIT PLAN

In Section 6 of his paper, Mr. Meyers constructs an example to illustrate the behavior of a multi-split plan such as that currently used for workers' compensation. As in his Section 5, the risks are divided into a small number of possible types.[41]

He uses a Poisson distribution to model the claim counts. He uses a (continuous) Weibull distribution to model claim severity. The frequency and severity are treated as independent.[42]

No overall limit is applied by the author to the losses. In actual application of the Experience Rating Plan, a per claim accident limit is applied.[43] The value differs considerably by state, but I will use \$100,000 here for illustrative purposes. If the severity distribution has a large tail, this accident limitation can add to the efficiency of the plan.

10.1 The Current Workers' Compensation Experience Rating Plan

Mr. Meyers examines whether the current Workers' Compensation Experience Rating Plan or his formula 6.1 works better, i.e., which produces higher efficiency. Mr. Meyers concludes that his formula 6.1, which gives no credibility to the excess losses, outperforms the current Workers' Compensation plan.

The first thing to note is that the current workers' compensation formula can be written as the new estimate of expected losses equals the old estimate of expected losses times the experience modification,

$$F = \frac{A_p + W A_e + (1 - W) E_e + (1 - W) K}{E + (1 - W) K} E . \tag{10.1}$$

Then, following Snader [9], let

$$Z_p = \frac{E}{E + K_p} , \tag{10.2}$$

$$Z_e = \frac{E}{E + K_e} = W Z_p , \tag{10.3}$$

$$K_p = (1 - W)K = B, \text{ and} \tag{10.4}$$

[41] In Appendix J, an example is given of a continuous distribution of risks.

[42] In Section 5 of his paper, Mr. Meyers's example had the frequency and severity highly correlated.

[43] For accidents involving multiple claims, the limitation is twice that for single claim accidents.

$$K_e = \frac{(1 - W)(K + E)}{W} . \tag{10.5}$$

Then formula 10.1 can be written as

$$F = E_p (1 - Z_p) + A_pZ_p + E_e (1 - Z_e) + A_eZ_e . \tag{10.6}$$

Formula 10.6 is the form of F that has been previously discussed. As was shown previously, the solution for the optimal Z_p and Z_e given in formulas 5.3 and 5.4 do not have the form of formulas 10.2 and 10.3, even if K_p and K_e were constants with size of risk. In the current workers' compensation plan, $1 - W$ and thus K_p decreases with increasing size of risk until it is zero for self-rated risks. Below a certain value, $W = 0$ and thus $Z_e = 0$. Above that value, W increases to 1 with increasing size of risk, and K_e decreases with increasing size of risk until it reaches zero for self-rated risks. Values of K_p, K_e, Z_p, and Z_e are displayed in Table 10.1 for a typical choice of parameters.[44]

Mr. Meyers makes the excellent point that there is no theoretical framework in which the standard workers' compensation formula (with this particular variation of K_p and K_e with size of risk) is optimal.[45]

10.2 Meyers's Alternative, Zero Excess Credibility

Meyers's formula 6.1 can be written as

$$F = E \left(\frac{A_p + K}{E_p + K}\right) , \text{ or}$$

$$F = E_p \left\{1 - \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right)\right\}$$

$$+ A_p \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right) + E_e . \tag{10.7}$$

which is a special case of formula 5.1 with

$$Z_p = \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right) , \text{ and} \tag{10.8}$$

$$Z_e = 0$$

[44] These are the values currently in use in Massachusetts.

[45] It is also of interest to note that Z_e increases very quickly with the size of risk.

TABLE 10.1

WORKERS' COMPENSATION EXPERIENCE RATING
CURRENT PLAN
EXAMPLE FOR TYPICAL VALUES*

Expected Losses	Credibility Parameters K (00)		Credibilities (Percent)	
(000)	Primary	Excess	Primary	Excess
5	200	Infinite	20	0
10	200	Infinite	33	0
15	200	Infinite	43	0
20	200	Infinite	50	0
25	200	Infinite	56	0
30	198	49500	60	1
35	196	26950	64	1
40	194	19400	67	2
45	194	21017	70	2
50	192	16800	72	3
60	188	12533	76	5
70	184	10350	79	6
80	182	10111	81	7
90	178	8900	83	9
100	174	8031	85	11
120	168	7350	88	14
140	162	6821	90	17
160	154	6026	91	21
180	148	5692	92	24
200	140	5133	93	28
225	132	4756	94	32
250	124	4405	95	36
275	116	4074	96	40
300	106	3609	97	45
325	98	3315	97	50
350	90	3027	97	54
375	82	2745	98	58
400	72	2363	98	63
425	64	2094	99	67
450	56	1828	99	71
475	48	1563	99	75
500	38	1220	99	80
525	30	962	99	85
550	22	704	100	89
575	14	448	100	93
600	6	192	100	97
625	0	0	100	100

* Self-Rating Point varies by state. Self-Rating Point taken as S = \$615,000. Q = \$25,000. K = \$20,000. See Snader [9].

As we saw in the discussion of Meyers's Section 5, the use of the credibilities in formula 10.8 results in less efficiency than the use of the theoretically optimal credibilities, but for many applications the loss in efficiency may be acceptable.

10.3 A Modification of Meyers's Example

The loss in efficiency will be examined for the example in Meyers's section 6. In order to simplify the calculations, a discrete version of the Weibull will be used.[46] In order to better match the current plan, an accident limitation will be used. The accident limitation will be chosen at \$100,000. The probability, $F(x)$, that a claim will be less than or equal to x is given by:[47]

$$F(x) = 1 - e^{-(x/b)^c} \qquad x = \$100, \$200, \ldots \$100{,}000 \tag{10.9}$$

The remaining probability will be at the accident limitation \$100,000.

The primary portion of each loss will be determined by the multi-split formula

$$x_p = \begin{cases} x, & x \leqq B \\ \dfrac{(B + C)\,x}{x + C}, & x \geqq B\,. \end{cases} \tag{10.10}$$

The current plan has $C = 4B$ and $B = \$2{,}000$.

The frequency will be Poisson as per Mr. Meyers; however, we will take the parameter λ equal to N times 4, 7, 10, 13, and 16. Thus N represents the size of risk, with $N = 10$ in Meyers's example.[48]

Appendix I shows how to calculate the quantities that enter formulas 5.3 and 5.4, when the frequency and severity are independent.

The results for this example with $N = 10$ are $a = 86961$, $b = 349814$, $c = 65952$, $d = 101857$, $r = 123788$, and $s = 77843$, $Z_p = 185.8\%$, $Z_e = -14.4\%$, and Efficiency $= 74.6\%$.

---

[46] The use of the discrete version changes some of the actual values, but the essence of the example is preserved. It should be noted that the continuous Weibull usually does not fit the observed size of loss distribution for small claims. For example, with $b = 50$ and $c = .25$, 31% of the losses will be one dollar!

[47] Meyers takes $c = .25$. He lets $b = 30$, 40, 50, 60, and 70 with equal probability.

[48] It should be noted that really small risks are currently not eligible for experience rating. The eligibility level for a risk with three years of experience eligible for experience rating is generally set so that for each state it approximates the average premium of a risk with 10 full time employees. (For example, in one state, this is currently \$3500 in premium per year.)

Using formula 6.2 would result in $Z = 81.4\%$ and an efficiency of 73.5%. This is quite close to optimal.

Using formula 6.4 would result in $Z = c/a = 75.8\%$ and an efficiency of 73.2%. This is still quite close to optimal. Thus, as we saw in Section 7 of this discussion, the use of formula 6.4, as suggested by Mr. Meyers, results in a relatively small loss in efficiency compared to optimal.

10.4 Results of Using the Various Formulas for Credibilities

Table 10.2 gives, for various sizes of risks and various choices of B (as per formula 10.10), the efficiencies for this example using formulas 5.3 and 5.4 for credibility. We notice that for smaller risks, the optimal choice of B is much lower than the current value of \$2000.[49] The optimal B rises with size of risk. For the largest risks it reaches \$2000.

Table 10.3 is similar to Table 10.2, except that the credibilities are calculated using formula 6.2. There is a similar pattern to Table 10.2. The loss in efficiency is relatively small compared to Table 10.2. Table 10.4 uses formula 6.3 in order to calculate the credibilities. It is virtually identical to Table 10.3.

Table 10.5 is similar to the preceding tables, except that the credibilities are calculated using formula 6.4, as recommended by Meyers. The pattern is again very similar, and the losses in efficiency are probably sufficiently small to be acceptable for most practical applications. Thus, at least for this example, there is little disadvantage to setting the excess credibility equal to zero.

Table 10.6 is similar to Table 10.2, except that a single split plan has been used instead of a multi-split plan. The loss in efficiency is relatively small. Thus, at least for this example, there is little disadvantage to the use of the simpler single split plan.

Table 10.7 is similar to Table 10.2, except that an accident limitation of \$200,000 rather than \$100,000 has been used. The pattern is similar to that in Table 10.2. As expected, the optimal value of B is (slightly) higher, since the tail of the severity distribution is relatively more important. The efficiencies in Table 10.2 are lower than those in Table 10.7, but it is inappropriate to compare

[49] This result depends on Meyers's choice in this example of the relative importance of variation of frequency and variation of severity as well as the use of the Weibull distribution even for small sizes of claims.

them directly. Table 10.7 is the result of attempting to estimate losses capped at $200,000. This is a more difficult task than trying to estimate losses capped at $100,000 as in Table 10.2.[50]

TABLE 10.2

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULAS 5.3 AND 5.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	78.4%	85.9%	92.7%
$200	41.4	63.6	79.1	86.9	93.1
$500	34.6	58.9	78.5	87.8	93.7
$1,000	29.6	54.3	77.0	88.0	94.0
$2,000	25.2	49.4	74.6	87.7	94.3
$5,000	20.3	42.9	70.5	86.5	94.3
$10,000	17.2	38.3	66.8	85.0	94.1

TABLE 10.3

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.2

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	77.1%	81.2%	82.7%
$200	41.0	63.5	78.7	84.4	86.6
$500	33.6	58.3	78.5	87.1	90.6
$1,000	28.2	53.0	76.7	87.9	92.7
$2,000	23.2	47.0	73.5	87.6	93.9
$5,000	17.4	38.7	67.3	85.4	94.3
$10,000	13.9	32.7	61.7	82.6	93.7

[50] One could put the two tables on a comparable basis by adjusting the efficiencies in Table 10.2 to what they would have been if one measures efficiency in terms of the variation of the losses capped at $200,000. However, this is beyond the scope of this discussion.

TABLE 10.4

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.3

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	77.1%	81.1%	82.6%
$200	41.0	63.5	78.7	84.3	86.3
$500	33.6	58.3	78.5	87.1	90.3
$1,000	28.2	53.0	76.7	87.9	92.4
$2,000	23.2	47.0	73.5	87.6	93.8
$5,000	17.4	38.7	67.3	85.4	94.3
$10,000	13.9	32.7	61.7	82.6	93.7

TABLE 10.5

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.6%	65.3%	76.8%	80.9%	82.5%
$200	40.8	63.2	78.3	84.0	86.2
$500	33.4	58.0	78.1	86.6	90.1
$1,000	28.0	52.7	76.3	87.4	92.2
$2,000	23.0	46.8	73.2	87.2	93.5
$5,000	17.4	38.6	67.1	85.2	94.1
$10,000	13.9	32.6	61.6	82.5	93.6

TABLE 10.6

EFFICIENCIES
SINGLE SPLIT PLAN, FORMULAS 5.3 AND 5.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
\$100	47.8%	63.8%	74.8%	83.4%	91.9%
\$200	48.2	65.4	76.8	84.6	92.3
\$500	44.1	64.7	78.6	86.2	92.8
\$1,000	38.8	61.9	78.9	87.2	93.3
\$2,000	33.0	57.5	78.1	87.9	93.8
\$5,000	25.6	49.9	74.9	87.7	94.3
\$10,000	20.6	43.5	70.9	86.6	94.4
\$20,000	16.4	36.9	65.6	84.4	94.0

TABLE 10.7

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULAS 5.3 AND 5.4
\$200,000 ACCIDENT LIMITATION

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
\$100	44.7%	64.0%	76.6%	83.9%	91.1%
\$200	40.4	62.2	77.5	85.3	91.7
\$500	33.7	57.7	77.3	86.6	92.5
\$1,000	28.7	53.1	75.8	87.0	93.1
\$2,000	24.2	48.0	73.4	86.7	93.5
\$5,000	19.1	41.1	68.8	85.4	93.7
\$10,000	16.0	36.1	64.7	83.7	93.5

11. TESTING AN EXPERIENCE RATING PLAN ON ACTUAL DATA

In Section 7 of his paper, Mr. Meyers gives a generally applicable method of testing experience rating plans. It is a more modern and statistically sophisticated version of the method presented by Dorweiler. The author uses this method on actual data to test which formula for credibilities performed best, as well as to test which values of parameters worked best given a particular formula.

Meyers looked at formulas 2.1 and 2.2, and found that formula 2.2, which assumes parameter uncertainty, performed better. It would be interesting to perform the same test on other candidates, such as formula 2.5. In fact, given the other features of the plan, one can find by trial and error a relation of credibility with size of risk that works well.

Given the appropriate data, the general method presented by the author should be able to answer the following questions which were not tested in the paper.

1. What is the best loss limit to use?
2. Does a loss limit which increases with the size of risk significantly improve the performance of the plan?
3. Does a multi-split plan perform significantly better than a single split plan?
4. Does assigning non-zero credibility to the excess losses perform significantly better than assigning zero credibility to the excess losses?

Mr. Meyers is able to use the method in this section not only to get a point estimate of the credibility parameter K, but also to get a confidence interval for K. It is interesting to note that the best estimate of K is not at the center of the confidence interval. Rather, the best estimate is nearer the low end of the confidence interval. This is at least partially explained by the fact that it is the ratio of different estimates of K rather than their difference which is important.[51]

---

[51] This feature of estimates of K is discussed in Mahler [6]. For example, if one either doubles or halves K, the resulting maximum changes in credibility are the same. The connection to the result here was pointed out to this author by Mr. Meyers.

12. CONCLUSIONS AND SUMMARY

While Mr. Meyers's paper is an excellent contribution to the actuarial literature which opens up many areas for further investigation, I think the author goes a little too far in drawing conclusions from his work. I will arrange my conclusions in a manner parallel to the author's, in order to allow ready comparison.

MEYERS	MAHLER
1. A loss limit can be an effective tool for increasing the accuracy of an experience rating formula. Loss limits are particularly helpful when there are differences in claim frequency. Even if the only differences among the insureds are in claim severity, little accuracy will be lost with a loss limit.	1a. A loss limit can be an effective tool for increasing the accuracy of an experience rating formula.[52] 1b. If the differences between risks within rating classes are mainly due to differences in frequency, then a lower loss limit is optimal. If they are mainly due to severity, then a higher loss limit is optimal. 1c. The efficiency is relatively insensitive to the choice of the loss limit. If your chosen loss limit is close to optimal, and assuming you choose credibilities close to optimal, then your efficiency will be very close to optimal. 1d. The larger the maximum loss that can occur,[53] and the thicker the tail of the size of loss distribution,[54] the more important it is to have a loss limit.

[52] Appendix J gives an example where there is no practical advantage to the use of a loss limit.

[53] Here we mean the largest loss considered by the plan, that is $50,000 in the Meyers's general liability example and $100,000 in my version of Meyers's workers' compensation example.

[54] The Pareto has a very thick tail. The Weibull has a thick tail for Meyers's $c < 1$, but not quite as thick as the Pareto. See Hogg and Klugman [10].

MEYERS	MAHLER
	1e. The optimal loss limit increases slowly or remains virtually constant up to a large size of risk. For very large risks, the optimal loss limit increases more quickly.
	1f. The optimal credibilities depend on the loss limit chosen.
	1g. Under certain conditions, there is a theoretical as well as a practical justification for having a self-rating point.
2. The current formula in the Workers' Compensation Experience Rating Plan, which has a separate treatment of primary and excess losses, is less accurate than a formula which uses only primary losses.	2a. The current formula used in the Workers' Compensation Experience Rating Plan can be improved.
	2b. The current manner in which the credibilities in the Workers' Compensation Plan vary with size of risk has no theoretical justification. Empirical studies should be done to come up with more appropriate relationships.
	2c. The gain in efficiency from the use of a multi-split rather than a single split plan may not be large enough to justify the use of the more complicated multi-split plan.
	2d. The excess credibilities are expected to be relatively small. The gain in efficiency may not be large enough to justify the use of the excess losses.
	2e. The per claim limitation in the Workers' Compensation Plan serves a useful purpose.

MEYERS

3. There are some very plausible situations when the standard credibility formula $Z = E/(E + K)$ is not appropriate. These include parameter uncertainty over time and loss limit which increases with the size of the insured. Failure to recognize this will result in overstating credibilities for larger insureds.

MAHLER

3a. The traditional formula for credibility, formula 2.1, applies in only limited special situations.
3b. Three specific phenomena are examined: parameter uncertainty, shifting parameters over time, and risk heterogeneity. Each of these will tend to lower credibilities for large risks compared to those from the traditional formula. (Formulas for each are presented.) One or more of them are expected to be important in many situations.
3c. Under certain conditions, the optimal credibility will remain substantially less than one, regardless of how large the risk gets or how many years of data are used.
3d. Under certain circumstances, 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 from using additional years of data.
3e. If the loss limit changes, the optimal credibilities also change. This is another reason why formula 2.1 may not apply.
3f. The efficiencies are relatively insensitive to the choice of the credibilities. The credibilities, in turn, are relatively insensitive to the choice of parameters entering the formula.

MEYERS

4. The author would recommend an experience rating formula based on the credibility formula 2.2

$$Z = E/(JE + K).$$

A loss limit that does not vary by size of insured should be a part of the plan. Excess losses should not be a part of the plan. This formula is less complicated than current formulas and should be easier to administer.

MAHLER

4. Any reasonable experience rating plan is expected to achieve a substantial increase in efficiency. However, theoretical and empirical studies should allow a significant improvement in the efficiency of most plans.

Mr. Meyers's paper has already stimulated work on experience rating plans which should lead to substantial improvements in the design of these plans in the near future. The paper also examines some interesting features of credibility which should have implications outside the area of experience rating plans.

REFERENCES

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

[2] Robert A. Bailey and LeRoy J. Simon, "An Actuarial Note on the Credibility of a Single Private Passenger Car," *PCAS* XLVI 1959, p.159.

[3] Herbert E. Wittick, "The Canadian Merit Rating Plan for Individual Automobile Risks," *PCAS* XLV 1958, p.214.

[4] Robert A. Bailey and LeRoy J. Simon, "Two Studies in Automobile Insurance Ratemaking," *PCAS* XLVII 1960, p.1.

[5] William J. Hazam, Discussion of [2], *PCAS* XLVII 1960, p.150.

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

[7] Gary G. Venter, Discussion of [8], *PCAS* LXX 1983, p.62.

[8] Philip E. Heckman and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX 1983, p.22.

[9] Richard H. Snader, "Fundamentals of Individual Risk Rating and Related Topics," Part I of Study Note, Casualty Actuarial Society, 1980.

[10] Robert V. Hogg and Stuart A. Klugman, *Loss Distributions,* John Wiley & Sons, 1984.

[11] George W. Snedecor and William G. Cochran, *Statistical Methods,* The Iowa State University Press, 1980.

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

[13] Richard G. Woll, "A Study of Risk Assessment," *PCAS* LXVI 1979, p.84.

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

[15] Charles C. Hewitt, "Loss Ratio Distributions—A Model," *PCAS* LIV, 1967, p.70.

APPENDIX A

TWO RESULTS FOR COVARIANCES

In this appendix will be established two useful results involving covariances. The first result is:

Total covariance = expected value of the process covariance
+ parameter covariance

where the parameter covariance is another term for the covariance of the hypothetical means. It should be noted that the similar result involving variances is just a special case of this result.[55]

If θ represents the set of parameters, this result can be written as follows.

Theorem: $\mathrm{COV}[X,Y] = E_\theta\,[\mathrm{COV}[X,Y|\theta]] + \mathrm{COV}_\theta\,[E[X|\theta], E[Y|\theta]]$

Proof:
$$\begin{aligned}
\mathrm{COV}[X,Y] &= E[XY] - E[X]\,E[Y] \\
&= E_\theta\,[E[XY|\theta]] - E_\theta[X|\theta]\,E_\theta[Y|\theta] \\
&= E_\theta\,[E[XY|\theta]] - E_\theta\,[E[X|\theta]\,E[Y|\theta]] \\
&\quad + E_\theta\,[E[X|\theta]\,E[Y|\theta]] - E_\theta\,[X|\theta]\,E_\theta\,[Y|\theta] \\
&= E_\theta\,[E[XY|\theta] - E[X|\theta]\,E[Y|\theta]] \\
&\quad + E_\theta\,[E[X|\theta]\,E[Y|\theta]] - E_\theta[X|\theta]\,E_\theta[Y|\theta] \\
&= E_\theta\,[\mathrm{COV}[X,Y|\theta]] + \mathrm{COV}_\theta\,[E[X|\theta], E[Y|\theta]]
\end{aligned}$$

The second result puts the covariance of the primary losses and excess losses in terms of the means, variances, and covariance of the primary and excess severity and the frequency. The familiar result for the variance of the losses in terms of frequency and severity is a special case of this result.[56]

Theorem: Assume a claims process in which frequency and severity are independent of each other, and the claim sizes are mutually independent random variables with a common distribution. Then let each claim be divided into two pieces in a well-defined manner not dependent on the number of claims. For convenience, we refer to these two pieces as primary and excess.

[55] Both results are familiar to statisticians. See, for example, Snedecor and Cochran [11]. However, the result for variances seems more familiar to actuaries.

[56] The proof given here parallels that for the more familiar result given in Appendix 2 of Venter [12].

Let:

T_p = Primary Losses
T_e = Excess Losses
X_p = Primary Severity
X_e = Excess Severity
N = Frequency

Then:

$\text{COV}\ [T_p, T_e] = E[N]\ \text{COV}\ [X_p, X_e] + \text{VAR}[N]\ E[X_p]\ E[X_e]$

Proof: T_p is the sum of the individual primary portions of claims $X_p(i)$, where i runs from 1 to N, the number of claims. Similarly, T_e is a sum of $X_e(i)$. Since N is a random variable, both frequency and severity contribute to the covariance of T_p and T_e.

To compute the covariance of T_p and T_e, begin by calculating

$E\ [T_p\ T_e \mid N = n]$,

fix the number of claims n and find

$$E\left[\left(\sum_{i=1}^{n} X_p(i)\right)\left(\sum_{i=1}^{n} X_e(i)\right)\right].$$

Expanding the product yields n^2 terms of the form $X_p(i)\ X_e(j)$. When $i = j$ the expected value of the term is

$E[X_p(i)\ X_e(i)] = \text{COV}\ [X_p, X_e] + E[X_p]\ E[X_e]$

from the definition of covariance. Otherwise it is $E[X_p]\ E[X_e]$, since then $X_p(i)$ and $X_e(j)$ are independent. Thus

$$E\left[\left(\sum_{i=1}^{n} X_p(i)\right)\left(\sum_{i=1}^{n} X_e(i)\right)\right] = n\ \text{COV}\ [X_p, X_e] + n^2\ E[X_p]\ E[X_e].$$

Now, by general considerations of conditional expectations,

$E\ [T_pT_e] = E_n\ [E[T_pT_e \mid N = n]]$.

Thus, taking the expected value of the above equation with respect to N gives

$$\begin{aligned} E\ [T_pT_e] &= E[N]\ \text{COV}\ [X_p, X_e] + E[N^2]\ E[X_p]\ E[X_e], \text{ and} \\ \text{COV}\ [T_p, T_e] &= E[T_pT_e] - E[T_p]\ E[T_e] \\ &= E[N]\ \text{COV}[X_p, X_e] + (\text{VAR}[N] + E^2[N])\ E[X_p]\ E[X_e] \\ &\quad - E[N]\ E[X_p]\ E[N]\ E[X_e] \\ &= E[N]\ \text{COV}[X_p, X_e] + \text{VAR}[N]\ E[X_p]\ E[X_e]. \end{aligned}$$

APPENDIX B

PARAMETER UNCERTAINTY

In this appendix, the effect of parameter uncertainty on the formula for credibility is discussed. This discussion is intended to aid in the understanding of both Meyers's result for this phenomenon and the results obtained here in the later appendices for the two other similar phenomena, risk heterogeneity and shifting parameters over time.

As explained in Mr. Meyers's Appendix A, when there is parameter uncertainty, the credibility as a size of risk no longer follows formula 2.1, but rather follows formula 2.2.

The important point is the behavior of the expected variance within classes, which Meyers labels δ^2.

$$\delta^2 = E\,[(A - \mu)^2]$$

This is normally thought of as the expected value of the process variance. If for each risk the parameters of the risk process themselves vary randomly,[57] then δ^2 really is made up of two pieces.[58] The first piece of δ^2 is due to the variance of the parameters due to different states of the universe.[59] The second piece of δ^2 is due to the process variance, given a specific state of the universe. The first piece is expected to be proportional to N^2, just as was the variance between risks due to different parameters.[60] The second piece is expected to be proportional to N as usual.

In other words, we can write δ^2 as

$$\delta^2 = N^2\,\alpha^2 + N\,\chi^2.$$

[57] An example is given in Section 3 of this discussion.

[58] This is a special case of the first result in Appendix A.

[59] In the example in Section 3, there were for simplicity two states, determined by whether it was raining or not.

[60] Thus somewhat paradoxically, δ^2, which is usually thought of as "process variance" actually includes a piece of "parameter variance," albeit of a very special variety.

The "good" piece of δ^2 goes up only as N, while the "bad" piece goes up as N^2. This bad piece of δ^2 was introduced due to the assumed different possible states of the universe. Unlike the good piece of δ^2, this piece of δ^2 increases as quickly as the variation between risks (Meyers's τ^2), which also increases as N^2.

Thus taking more observations will not get rid of the effect due to the variation inherent in the universe.[61]

If

$$\tau^2 = N^2 \beta^2.$$

then

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N}{N(1 + \alpha^2/\beta^2) + \chi^2/\beta^2},$$

which is of the form

$$Z = \frac{N}{NJ + K}, \qquad J \geqq 1,$$

which is equation 2.2.

[61] The same idea has applications to risk assessment efficiency and class homogeneity. See Woll [13].

APPENDIX C

RISK CHARACTERISTICS CHANGING OVER TIME

In this appendix, the effect of shifting parameters over time on the formula for credibility is discussed. A general formula is derived and the results are applied for a reasonable special case. It is this special case that results in formula 2.3 in the main text.

Assume that the parameters that describe the loss process of a risk are not constant over time.[62] Then the theoretical true mean for each risk is a function of time. For example, the shift might be due to a change in the attitude of management with regard to safety or due to a change in the upkeep of the roads on which the insured usually drives his car. We are not including shifts that are expected to affect all risks in the same manner, for example, claim cost trend.

Assume we have an experience rating plan, and we use N years of data, such that our estimate of the mean F we expect in year $N + \Delta$ is given by:

$$F = \left(1 - \sum_{i=1}^{N} Z_i\right) E_{N+\Delta} + \sum_{i=1}^{N} Z_i A_i$$

where A_i is the actual losses observed for year i, (brought up to the expected level of year $N + \Delta$),[63] Z_i is the credibility assigned to that year losses, and $E_{N+\Delta}$ is the expected losses for year $N + \Delta$.[64]

We assume that the individual years of data are generated by the same size of risk (or same number of risks). Thus except for the assumed shifting parameters over time, we would assign each of the years equal weight.

General Case

We wish to find Z_i for $i = 1$ to N, such that the efficiency is maximized. In order to proceed, we will assume a covariance structure. We will assume that the correlation expected between two years of data separated in time by i

[62] We assume that while individual risk parameters shift, the overall distribution of risks remains the same. Also, we assume there is no way to predict the shift of an individual risk, and that the class plan doesn't pick up the shift.

[63] Thus we assume $E_i = E_j = E$.

[64] Typically $N = 3$. For workers' compensation usually $\Delta = 2$. For private passenger auto usually $\Delta = 1$.

years, is a function of i, $l(i)$. It is assumed that the expected correlation decreases (or stays the same) as the years get further apart. Specifically, we assume:

$$l(0) = 1$$

$$l(i) \geqq l(i + 1).$$

The covariance structure assumed is:

$$E[(A_i - E)(A_j - E)] = \beta^2\, l(|i - j|) + \delta_{ij}\chi^2$$

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

where for different years, $i \neq j$, we get the variation of the hypothetical means, β^2, times a factor equal to a correlation $l(|i - j|) \leqq 1$, dependent on the number of years of difference. For $i = j$, we get the variation of the hypothetical means β^2, plus the expected value of the process variance χ^2. The closer $l(|i - j|)$ is to 1, the less shifting of parameters there is over time.

$$\text{Efficiency} = 1 - \frac{E\,[(F - \mu_{N+\Delta})^2]}{E\,[(E - \mu_{N+\Delta})^2]}$$

Substituting our expression for F, letting $K = \chi^2/\beta^2$ and simplifying we get:

$$\text{Efficiency} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} Z_i\, Z_j\, \{l(|i - j|) + K\delta_{ij}\} - 2\sum_{i=1}^{N} Z_i\, l(|N + \Delta - i|)}{E[(E - \mu_{N+\Delta})^2]}$$

We get the maximum efficiency by setting each of the partial derivatives with respect to the different credibilities equal to zero. This gives the following set of N linear equations in N unknowns.

$$\sum_{i=1}^{N} Z_i\, \{l(|i - j|) + \delta_{ij}K\} = l(|N + \Delta - j|) \qquad j = 1, 2, \ldots, N$$

These equations can be solved simply by using the usual matrix methods. However, for $N > 2$ the expressions for the solutions are complicated to actually write out. For $N = 2$ remembering that $l(0) = 1$, we get:

$$Z_1 = \frac{(1 + K)\, l(\Delta + 1) - l(1)\, l(\Delta)}{(1 + K)^2 - l(1)^2}$$

$$Z_2 = \frac{(1 + K)\, l(\Delta) - l(1)\, l(\Delta + 1)}{(1 + K)^2 - l(1)^2}$$

The ratio of Z_1 to Z_2 is given by:

$$\frac{Z_1}{Z_2} = \frac{(1 + K)\, l(\Delta + 1) - l(1)\, l(\Delta)}{(1 + K)\, l(\Delta) - l(1)(\Delta + 1)}$$

For the usual case where $l(\Delta + 1) < l(\Delta)$, this ratio is less than one, and thus $Z_1 < Z_2$. As expected the more recent data (year 2) is given more credibility than the less recent data (year 1).

$$Z_1 + Z_2 = \frac{(1 + K) - l(1))(l(\Delta + 1) + l(\Delta))}{(1 + K)^2 - l(1)^2}$$

$$= \frac{l(\Delta + 1) + l(\Delta)}{1 + l(1) + K}$$

For $l(i) = 1$ for all i, this reduces to the familiar $2/(2 + K)$.

For $N > 2$, one can approximate the exact solution as follows.

Adding up the N linear equations gives after rearranging the order of summation:

$$\sum_{i=1}^{N} Z_i \left\{ \left[\sum_{j=1}^{N} l(|i - j|) \right] + K \right\} = \sum_{j=1}^{N} l(|N + \Delta - j|)$$

The term $\sum_{j=1}^{N} l(|i - j|)$ depends on the value of i. For $i = 1$ and $i = N$ it is equal to $\sum_{j=0}^{N-1} l(j)$. For values of i between 1 and N, smaller values of $|i - j|$ are duplicated in the sum, while larger values of $|i - j|$ no longer enter it.[65] Thus since we have assumed that l is a decreasing function, we have

$$\sum_{j=1}^{N} l(|i - j|) \geqq \sum_{j=0}^{N-1} l(j)$$

Thus if we substitute $\sum_{j=0}^{N-1} l(j)$ into our previous equation wherever $\sum_{j=1}^{N} l(|i - j|)$ appeared we get:

$$\left(\sum_{i=1}^{N} Z_i \right) \left\{ \left(\sum_{j=0}^{N-1} l(j) \right) + K \right\} \leqq \sum_{j=1}^{N} l(|N + \Delta - j|)$$

This suggests the following approximation, which gives lower credibilities than the exact solution.

---

[65] For example, if $N = 5$ and $i = 2$, then the sum is $l(1) + l(0) + l(1) + l(2) + l(3)$.

$$\sum_{i=1}^{N} Z_i \cong \frac{\sum_{i=0}^{N-1} l(\Delta + i)}{\left(\sum_{i=0}^{N-1} l(i)\right) + K}$$

The sum of the Z_i is the credibility assigned to the data for all the separate years combined. In the main text this is called Z.

For $l(i) = 1$ for all i, this reduces to the familiar $Z = N/(N + K)$.

If the covariance structure has the basic property assumed here, that the correlation between individual years of data is smaller the longer the time span between the years, then the credibilities will have the following properties. The credibility assigned to a more recent year of data should be higher than that assigned to a more distant year of data.[66] If the correlation between distant years of data is significantly lower, i.e., l (i) gets small relatively quickly, then beyond a certain point using more recent years of data will lead to little improvement in the estimate of the mean. If the individual years of data are given the same weight, then using more years of data will eventually lead to a *worse* estimate of the mean, since the very old data provides a poor estimate of the future mean. Finally, the smaller Δ is, i.e., the less the delay is in getting and using the data, the higher the credibilities and the better the resulting estimate of the future mean.[67]

Special Case

Let us take a special case of the general covariance structure that has been assumed. Let $l(i) = \rho^i$, $\rho \leqq 1$. Then:

$$E[(A_i - E)(A_j - \mathrm{E})] = \beta^2 \rho^{|i-j|} + \delta_{ij} X^2$$

Thus for two different years of data, their covariance is proportional to a constant ρ taken to the power equal to the number of years separating them. For $\rho < 1$, the covariance decreases the larger the separation, and goes to zero rather quickly.

[66] This concept of giving more recent data more weight than less recent data is a familiar one to actuaries. See for example, "Homeowners Insurance Ratemaking" by Walters [14]. However, when estimating values at ultimate, it might be appropriate in certain circumstances to assign less weight to more recent but immature data.

[67] For experience rating generally $\Delta \geqq 1$.

While for illustrative purposes this is not an unreasonable assumption for the structure of covariances, it is far from the only assumption that could be made. The actual covariance structure for a particular real world application would have to be determined empirically.

For this special case, we have for $N = 2$:

$$Z_1 = \frac{\rho^{\Delta}\rho K}{(1 + K)^2 - \rho^2}$$

$$Z_2 = \frac{\rho^{\Delta}}{(1 + K)^2 - \rho^2}(1 + K - \rho^2)$$

$$\frac{Z_2}{Z_1} = \frac{1 + K - \rho^2}{\rho K} = \frac{1}{\rho} + \frac{1}{K}\left(\frac{1}{\rho} - \rho\right) > 1$$

$$Z_1 + Z_2 = \frac{\rho^{\Delta}(1 + \rho)}{(1 + \rho) + K}$$

For $\rho = 1$, $Z_1 + Z_2$ reduces to the familiar $2/(2 + K)$.

As discussed above, one can approximate the exact solution by[68]:

$$Z = \sum_{i=1}^{N} Z_i \cong \frac{\rho^{\Delta}\left(\sum_{i=1}^{N} \rho^{i-1}\right)}{\left(\sum_{i=1}^{N} \rho^{i-1}\right) + K}$$

This is formula 2.3. For $\rho = 1$, this reduces to the familiar $N/(N + K)$.

[68] The exact solution gives lower credibilities.

APPENDIX D

RISK HETEROGENEITY

In this appendix, the effect of risk heterogeneity on the formula for credibility for large risks is discussed. In addition, the combined effect of risk heterogeneity and parameter uncertainty is discussed.

In general, a large risk is made up of smaller risks. For example, a large commercial risk might consist of a grouping of separate factories. Assume our different risks consist in each case of grouping together N factories of the same size.[69] Then how does the variance between different risks, $E[(E - \mu)^2]$, which Meyers calls τ^2, depend on the size of risk N?[70]

A Simple Example

To illustrate the point, let us examine a very simple example. Assume that half the factories are "good" and half are "bad." The good factories have an expected mean of one, while the bad factories each have an expected mean of two. Depending on how the factories are grouped together to form risks, τ^2 has a different dependence on N.

Case 1

Assume that the risks consist solely of good factories or bad factories, but never a mixture. Then the risks of size N have an expected mean of either N or $2N$, with equal frequency. Thus τ^2 is $N^2/4$.

Case 2

Assume that the risks consist of good and bad factories grouped together totally at random. A risk of size N, is merely a random sample of size N from the set of all factories. Thus in this case, τ^2 is N times the variance between individual factories. $\tau^2 = N/4$.

Case 3

Assume that half the risks are "superior" and half "inferior". Each factory in a superior risk has a 2/3 chance to be good and a 1/3 chance to be bad. The situation is reversed for inferior risks. Then the expected means of the superior

[69] In certain cases, a factory could be usefully broken up into smaller subunits. We are merely presenting a simple example here.

[70] τ^2 measures how homogeneous the classes are. The smaller τ^2, the less the separation between the risks and the more homogeneous the class.

risks of size N extend from N to $2N$, with the probabilities given by the binomial distribution with $p = 1/3$. The inferior risks also have expected means from N to $2N$, but with the probabilities given by the binomial distribution with $p = 2/3$.[71]

One can compute the variance τ^2 for specific values of N in the usual straightforward manner.[72] However, τ^2 can be broken up into two pieces.[73] The first piece is the variation among different superior risks or the variation among different inferior risks. This is just $2N/9$, since the variance of the binomial distribution is just $Np(1 - p)$. The second piece of τ^2 is the variance between the grand mean of superior risks and the grand mean of the inferior risks. Since these grand means are $4N/3$ and $5N/3$ respectively, this piece of τ^2 is just $N^2/36$. Adding the two pieces together, one gets:

$$\tau^2 = 2N/9 + N^2/36.$$

Generalizing the Simple Example

In Case 1, it was assumed that the risks are homogeneous. Each of the factories making up a risk has the same expected mean. In this case, τ^2 is proportional to N^2, since the expected mean for each risk just gets multiplied by N. This special case is the one that is usually dealt with.

On the other hand, in Case 2, we have assumed the other extreme, that the factories are grouped together totally at random.[74] Each risk is merely a sample of size N from the overall set of factories, and thus τ^2 is N times the variance between the individual factories. In this special case τ^2 is proportional to N.

Thus in the two extreme cases, we have either τ^2 proportional to N or τ^2 proportional to N^2. We expect most real world situations to be in the intermediate situation, such as Case 3, where bad factories are more likely to be grouped

[71] The use of the terms superior and inferior could be thought of in terms of some underwriting criterion. While the average superior risk has a lower expected mean than the average inferior risk, there are inferior risks with low means and vice versa.

[72] For example, for $N = 3$, the risks will have expected means of 3, 4, 5, and 6 with probabilities of 1/6, 1/3, 1/3, 1/6. Thus $\tau^2 = 11/12$. The reader can verify that this matches the formula for τ^2 given below.

[73] This is a special case of the first result given in Appendix A.

[74] As demonstrated in Hewitt [15], for loss ratio distribution purposes, the sum of two $50,000 risks doesn't act the same as a single $100,000 risk. Thus, Case 2 is not a good model of the reality; it is an extreme case chosen for illustrative purposes.

together with bad factories, but a single risk can be made up of both good and bad factories. For this intermediate case τ^2 had the form

$$\tau^2 = N\pi^2 + N^2\beta^2$$

This form for τ^2 follows from breaking the variance into two pieces. The first piece is the variation among risks of similar type.[75] This piece of τ^2 is proportional to N. The second piece is the variance between the grand means of different types of risks. This piece of τ^2 is proportional to N^2.

In order to get the credibility we must combine τ^2 with δ^2. As explained in Appendix B, without parameter uncertainty it makes sense to assume

$$\delta^2 = N\chi^2,$$

then

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N + \pi^2/\beta^2}{N + \dfrac{\chi^2 + \pi^2}{\beta^2}}\,.$$

which can be written in the form

$$Z = \frac{N + I}{N + K}\,, \qquad 0 \leqq I \leqq K$$

which is formula 2.5 in the main text.

While it at first appears that $I > 0$ (i.e. risk heterogeneity) leads to higher credibilities than formula 2.1 ($I = 0$), that is not the case. One must remember that the K in formula 2.5 is not equal to the K in formula 2.1. The K here has an additional term of π^2/β^2 compared to the K in formula 2.1. Thus since risk heterogeneity affects both I and K, a more careful analysis is required.

Formula 2.5 can be rewritten in the form

$$Z = 1 - \frac{\chi^2}{\chi^2 + N(\beta^2 + \pi^2) - (N - 1)\pi^2}\,.$$

The variance between single units (for example, between individual factories) is $\beta^2 + \pi^2$. Keeping $\beta^2 + \pi^2$ constant, we can see from the above equation, that as π^2 increases, Z decreases (for $N > 1$). In other words, the

[75] Thus somewhat paradoxically τ^2, which can be thought of as the "between variance," actually includes a piece of "within variance," albeit of a very special variety. In Appendix B, a similar but reversed situation was explained for δ^2, when there is parameter uncertainty.

greater the risk heterogeneity that is present, the lower the credibility, all other things being equal. Looked at another way, the more risk heterogeneity, the smaller τ^2 is, all other things being equal (for $N > 1$). Therefore, the more risk heterogeneity, the smaller the credibility.

Risk Heterogeneity and Parametric Uncertainty

If both risk heterogeneity and parameter uncertainty (see Appendix B) are important then we have:

$$\delta^2 = N^2\alpha^2 + N\chi^2$$
$$\tau^2 = N\pi^2 + N^2\beta^2$$

Thus:

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N + \pi^2/\beta^2}{N\left(1 + \frac{\alpha^2}{\beta^2}\right) + \frac{\chi^2 + \pi^2}{\beta^2}}$$

This can be written in the form

$$Z = \frac{N + I}{NJ + K}, \qquad 0 \leqq I \leqq K, J \geqq 1$$

which is equation 2.6.

APPENDIX E

PARAMETER UNCERTAINTY, SHIFTING RISK PARAMETERS AND RISK HETEROGENEITY

In this appendix, all three phenomena discussed in Appendices B, C, and D will be assumed to be of importance. The resulting formula for the sum of the credibilities will be a combination of the features of those in the prior appendices.[76] The notation from the previous appendices will be used.

Let M be a measure of the (average) size of the risk in each year. Let N be the number of years of data used for experience rating.

Assume the following covariance structure:

$$E\,[(A_i - E)(A_j - E)] = \rho^{|i-j|}\,(M^2\beta^2 + M\pi^2) + \delta_{ij}\,(M^2\alpha^2 + M\chi^2)$$

$$E\,[(\mu_i - E)(\mu_j - E)] = \rho^{|i-j|}\,(M^2\beta^2 + M\pi^2)$$

Then proceeding as in the previous appendices we get the following set of linear equations for the optimal credibilities Z_m, $m = 1, \ldots, N$.

$$\sum_{i=1}^{N} Z_i\,\{(M^2\beta^2 + M\pi^2)\,\rho^{|i-m|} + \delta_{im}\,(M^2\alpha^2 + M\chi^2)\} = \rho^{\Delta+N-m}$$

This set of equations can be solved by matrix methods. As in Appendix C, we can get an approximate solution that is exact for $N < 3$. (I, J, and K are defined in the previous appendices.)

$$Z = \sum_{i=1}^{N} Z_i \cong \frac{\rho^{\Delta}\left(\sum_{i=1}^{N}\rho^{i-1}\right)(M + I)}{\left(\sum_{i=2}^{N}\rho^{i-1}\right)(M + I) + JM + K}$$

This is formula 2.7.

[76] The formulas from the prior appendices are special cases of the formula presented here.

APPENDIX F

EFFICIENCY AND CREDIBILITIES FOR SPLIT EXPERIENCE RATING PLANS

In this appendix, the optimal primary and excess credibilities for a split experience rating plan are derived. The solution, equations 5.3 and 5.4 in the main text, is a generalization of the familiar Bühlmann result for the no-split situation. The second part of this appendix explores the results of using credibilities other than the optimal ones.

Assume we have an experience rating plan, and our estimate of the mean F is given by

$$F = (1 - Z_p)\, E_p + Z_pA_p + (1 - Z_e)\, E_e + Z_eA_e\,.$$

The subscripts p and e will stand for primary and excess; however, for now they can be treated as any two well-defined portions of the total losses. E_p and E_e are the expected losses of each type. A_p and A_e are the actual losses of each type. Z_p and Z_e will be thought of as the credibilities assigned to each portion of the losses; however, for now they can be treated as just numbers to be determined.

Efficiency

In accordance with Meyers, define the efficiency of F by

$$\text{Efficiency} = 1 - \frac{E[(F - \mu)^2]}{E[(E_p + E_e - \mu)^2]}\,,$$

where μ is the theoretical true mean for each risk. Let μ_p and μ_e be the excess and primary pieces of μ. $\mu = \mu_p + \mu_e$. μ is a function of the parameters that describe each risk.

$$\begin{aligned}(F - \mu)^2 = {} & Z_p^2\,(A_p - E_p)^2 + Z_e^2\,(A_e - E_e)^2 + (E_p + E_e - \mu)^2 \\ & + 2Z_pZ_e\,(A_p - E_p)\,(A_e - E_e) + 2Z_p\,(A_p - E_p)\,(E_p - \mu_p) \\ & + 2Z_p\,(A_p - E_p)\,(E_e - \mu_e) + 2Z_e\,(A_e - E_e)\,(E_e - \mu_e) \\ & + 2Z_e\,(A_e - E_e)\,(E_p - \mu_p)\end{aligned}$$

$$(E_p + E_e - \mu)^2 = (E_p - \mu_p)^2 + (E_e - \mu_e)^2 + 2(E_p - \mu_p)\,(E_e - \mu_e)$$

Let:

a = total variance of the primary losses;
b = total variance of the excess losses;
c = variance of the hypothetical means of the primary losses;
d = variance of the hypothetical means of the excess losses;
r = total covariance of the primary and excess losses; and,
s = covariance of hypothetical means of the primary and excess losses.

Remembering that μ is only a function of the set of parameters ϕ, μ is subject to parameter variance but not process variance. The actual observed losses A_p and A_e are subject to both parameter and process variance. E_p and E_e are the overall grand means and are subject to neither kind of variance.

Then we have

$E[(A_p - E_p)^2] = a;$
$E[(A_e - E_e)^2] = b;$
$E[(A_p - E_p)(\mu_p - E_p)] = E[(\mu_p - E_p)(\mu_p - E_p)] = c;$
$E[(A_e - E_e)(\mu_e - E_e)] = d;$
$E[(A_e - E_p)(A_e - E_e)] = r;$
$E[(A_p - E_p)(\mu_e - E_e)] = E[(\mu_p - E_p)(\mu_e - E_e)] = s;$ and,
$E[(A_e - E_e)(\mu_p - E_p)] = E[(\mu_e - E_e)(\mu_p - E_p)] = s.$

Substituting these values back in the definition of efficiency, we get

$$\text{Efficiency} = \frac{2Z_p(c + s) + 2Z_e(d + s) - Z_p^2 a - Z_e^2 b - 2Z_pZ_e r}{c + d + 2s}$$

Optimal Credibilities

The optimal credibilities are given by the least squares solution, which results in the maximum efficiency. In order to maximize the efficiency, we set the partial derivatives with respect to Z_p and Z_e equal to zero. This gives

$a\,Z_p + r\,Z_e = c + s$, and
$r\,Z_p + b\,Z_e = d + s.$

The solution of this simple set of two equations in two unknowns is

$$Z_p = \frac{(c + s)b - (d + s)r}{ab - r^2}, \text{ and}$$

$$Z_e = \frac{(d + s)a - (c + s)r}{ab - r^2}.$$

This is the desired result, which is equations 5.3 and 5.4 in the main text.

If we substitute these values of $Z_p = Z_{p,m}$ and $Z_e = Z_{e,m}$ into the formula for efficiency and simplify we get

$$\text{Maximum Efficiency} = \frac{Z_{p,m}(c + s) + Z_{e,m}(d + s)}{(c + d + 2s)} .$$

Thus, the maximum efficiency is a weighted average of the two credibilities that produce this maximum. This is similar to Meyers's result in his Appendix B, where the maximum efficiency was equal to the credibility that produces the maximum.

Zero Excess Credibility

Now we will explore the results of using credibilities other than the optimal ones. Let us take a special case when the variance of the excess losses is very large; in other words, $b \to \infty$. Then

$$Z_p \cong \frac{c + s}{a}, \text{ and}$$

$$Z_e \cong 0$$

In fact, if we set $Z_e = 0$, then the formula for efficiency becomes

$$\text{Efficiency} = \frac{2Z_p(c + s) - Z_p^2 a}{c + d + 2s} .$$

For this case, the value of Z_p such that the efficiency is maximized is given by

$$Z_p = \frac{c + s}{a} .$$

This is equation 6.1 in the main text.

The credibility obtained by applying the Bühlmann method to the primary losses alone is c/a. The credibility here is larger (for $s > 0$) since we have taken into account the positive correlation between the primary and excess severities.

When $Z_p = \dfrac{c + s}{a}$ and $Z_e = 0$ we get

$$\text{Efficiency} = \frac{(c + s)^2}{(c + d + 2s)a}.$$

If we let our estimate of the losses be given as in the General Liability Experience Rating Plan, using credibility Z as described by Meyers then

$$F = ZA_p \left(1 + \frac{E_e}{E_p}\right) + (1 - Z)E_p + (1 - Z)E_e.$$

This is a special case of the previous case, with

$$Z_p = Z\left(1 + \frac{E_e}{E_p}\right).$$

Thus, we know the maximum efficiency occurs when

$$Z\left(1 + \frac{E_e}{E_p}\right) = Z_p = \frac{c + s}{a}, \text{ or}$$

$$Z = \left(\frac{c + s}{a}\right)\left(\frac{E_p}{E_p + E_e}\right).$$

This is equation 6.2 in the main text.

Meyers's Case

Meyers instead uses $Z = c/a$. Putting the corresponding value of Z_p into the formula for efficiency gives, after simplifying terms

$$\text{Meyers's Efficiency} = \frac{(s + c)^2 - \left(s - c\,\dfrac{E_e}{E_p}\right)^2}{(c + d + 2s)\,a}.$$

Comparing this efficiency to that obtained when $Z = \left(\dfrac{c + s}{a}\right)\left(\dfrac{E_p}{E_p + E_e}\right)$,

we get:

$$\frac{\text{Meyers's Efficiency}}{\text{Maximum Efficiency}} = 1 - \frac{\left(s - c\dfrac{E_e}{E_p}\right)^2}{(s + c)^2}.$$

So, using the usual value of Z in accordance with Meyers leads to a loss of efficiency. However, the loss will be small whenever the second term is small. This is the case for all the examples tested in Meyers's Tables 5.2, 5.3, and 5.4.

Credibility Equal to One

Another useful special case is when in formula 6.3 in the main text for large N we take $Z = 1$ and thus

$$Z_p = 1 + \frac{E_e}{E_p}, \text{ and } Z_e = 0.$$

Then the efficiency is given by

$$\text{Efficiency} = \frac{2\left(1 + \frac{E_e}{E_p}\right)(c + s) - \left(1 + \frac{E_e}{E_p}\right)^2 a}{c + d + 2s}.$$

Using the notation of Appendix G, this becomes

$$\text{Efficiency} = 1 - \frac{\hat{c}\,\frac{E_e^2}{E_p^2} + \hat{d} + \frac{\hat{t}}{N}\left(1 + \frac{E_e}{E_p}\right)^2 - 2\,\frac{E_e}{E_p}\,\hat{s}}{\hat{c} + \hat{d} + 2\hat{s}},$$

which has a limit as N gets large of:

$$\text{Efficiency} = 1 - \frac{\hat{c}\,\frac{E_e^2}{E_p^2} + \hat{d} - 2\,\frac{E_e}{E_p}\,\hat{s}}{\hat{c} + \hat{d} + 2\hat{s}}.$$

Thus using formula 6.3 for a fixed loss limit, the maximum efficiency is less than 100%, and we expect to get relatively little improvement in efficiency beyond the point where $Z = 1$. This is the behavior observed in Table 7.3.

APPENDIX G

DEPENDENCE ON SIZE OF RISK

In this appendix, the variation of credibility with size of risk is explored for the cases examined in Meyers's Sections 5 and 6. Also tables of the specific values of the parameters entering the credibility formulas are given for two specific cases from Meyers's Sections 5 and 6. Finally, the general behavior of $N(1 - Z)/Z$ with size of risk is examined. In the examples in Meyers's Sections 5 and 6 there is no parameter uncertainty, the individual risks are homogeneous, and the parameters for an individual do not change over time. Then, as discussed by Meyers in his Section 3, the factors that go into formulas 5.3 and 5.4 for Z_p and Z_e vary as follows with the size of risk N.

Increase as N	Increase as N^2
$a - c = t =$ process variance of primary losses	c
$b - d = u =$ process variance of excess losses	d
$r - s = v =$ process covariance of primary and excess losses	s

$$\text{Let } \hat{t} = \frac{a - c}{N} = \frac{t}{N}$$

$$\hat{c} = \frac{c}{N^2}$$

with similar definitions, for the other quantities, such that we obtain new quantities which are independent of N.

Then substituting into formula 5.3 we get:

$$Z_p = N \frac{N(\hat{c}\hat{d} - \hat{s}^2) + \hat{c}\hat{u} + \hat{s}\hat{u} - \hat{d}\hat{v} - \hat{s}\hat{v}}{N^2(\hat{c}\hat{d} - \hat{s}^2) + N(\hat{c}\hat{u} + \hat{d}\hat{t} - 2\hat{s}\hat{v}) + \hat{u}\hat{t} - \hat{v}^2}$$

Substituting into formula 5.4 would give a similar complicated formula for Z_e. If we set the primary losses equal to the total losses, and set the excess losses equal to zero, then

$$Z_p = \frac{c}{a} = \frac{N}{N + \hat{t}/\hat{c}},$$

which is the familiar expression for credibility, given in Meyers's formula 3.2. However, we notice that in the more general expression, we do not have the familiar simple function of N. Instead we have:

$$Z_p = N \frac{N + K_3}{N^2 + NK_1 + K_2}$$

$$\text{where } K_1 = \frac{\hat{c}\hat{u} + \hat{d}\hat{t} - 2\hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

$$K_2 = \frac{\hat{t}\hat{u} - \hat{v}^2}{\hat{c}\hat{d} - \hat{s}^2}$$

$$K_3 = \frac{\hat{c}\hat{u} + \hat{s}\hat{u} - \hat{d}\hat{v} - \hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

Similarly

$$Z_e = N \frac{N + K_4}{N^2 + NK_1 + K_2}$$

where

$$K_4 = \frac{\hat{d}\hat{t} + \hat{s}\hat{t} - \hat{c}\hat{v} - \hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

For large N, if $K_1 > K_3$ we have $Z_p < 1$, but if $K_3 > K_1$, we have $Z_p > 1$.[77] In the latter case, it makes sense to refer to risks such that $Z_p \geqq 1$ as self-rated. Notice, the difference from the usual result, formula 2.1 in the main text, where Z remains strictly less than one, but gets so close to one so as to make no practical difference in the resulting efficiencies. As $N \rightarrow \infty$ we do have $Z_p \rightarrow 1$ and $Z_e \rightarrow 1$.[78]

We can write formula 5.5 for the maximum efficiency as:

$$\text{Maximum Efficiency} = \frac{Z_p(\hat{c} + \hat{s}) + Z_e(\hat{d} + \hat{s})}{\hat{c} + \hat{d} + 2\hat{s}}$$

[77] For the example as per Meyers's Table 5.4, the table below uses a loss limit of 2500, $K_1 = 8{,}308$, $K_2 = 110{,}847$, $K_3 = 19{,}119$ and $K_4 = 76$. For the example as per Meyers's Section 6, the table below uses B = 2000, $K_1 = 169$, $K_2 = 461$, $K_3 = 407$, and $K_4 = -42$.

[78] Thus, this is a different phenomenon than discussed in Meyers's Section 3, where $N \rightarrow \infty$, $Z \rightarrow 1/J \leqq 1$.

The dependence of the maximum efficiency on N is solely contained in Z_p and Z_e themselves. The weights are independent of N. As $N \to \infty$, we have $Z_p \to 1$, $Z_e \to 1$, Maximum Efficiency $\to 1$.

EXAMPLE AS PER MEYERS'S TABLE 5.4

Loss Limit ($000's)	$\hat{c}$	$\hat{d}$	$\hat{r}$	$\hat{s}$	$\hat{t}$	$\hat{u}$	$\hat{v}$	$N = 1$ E_p	$N = 1$ E_e
1	.013	.129	.338	.040	.215	9.303	.299	.350	.520
2	.033	.084	.626	.052	.441	8.525	.574	.463	.407
2.5	.041	.072	.752	.054	.548	8.171	.698	.495	.375
3	.050	.061	.848	.055	.684	7.846	.793	.528	.343
4	.064	.047	1.024	.055	.930	7.247	.970	.572	.299
6	.087	.031	1.287	.051	1.418	6.225	1.236	.631	.240
8	.104	.022	1.470	.047	1.897	5.374	1.422	.670	.200
12	.130	.012	1.678	.039	2.817	4.020	1.639	.722	.149
16	.148	.007	1.749	.033	3.694	2.987	1.717	.756	.115

EXAMPLE AS PER MEYERS'S SECTION 6[79]

B ($000's)	$\hat{c}$	$\hat{d}$	$\hat{r}$	$\hat{s}$	$\hat{t}$	$\hat{u}$	$\hat{v}$	$N = 1$ E_p	$N = 1$ E_e
.1	50	2550	784	317	41	35110	467	1.65	9.38
.2	92	2305	1278	419	105	34263	859	2.23	9.26
.5	209	1864	2394	581	365	32094	1813	3.30	8.19
1	380	1454	3698	700	901	29190	2998	4.37	7.12
2	660	1019	5373	778	2101	24796	4595	5.65	5.84
5	1234	499	7580	751	5751	16676	6829	7.53	3.95
10	1803	221	8218	605	11057	9803	7612	8.93	2.56

[79] Meyers's example corresponds to $N = 10$, B = \$2000, except that here a claim limitation of \$100,000 has been used, and the Weibull distribution has been approximated by a discrete analog. See Section 10 of this discussion.

When doing empirical studies it is sometimes useful to focus on the following quantity as a function of size of risk, rather than the credibilities themselves. Define $\kappa = N(1 - Z)/Z$. Then for the usual simple case, formula 2.1 in the main text, $\kappa = K$. With parameter uncertainty, formula 2.2 in the main text, $\kappa = (J - 1)N + K$. For the split plan case handled previously in this appendix, $\kappa_p = [N(K_1 - K_3) + K_2]/(N + K_3)$. One can take other cases and derive the expected behavior of κ as a function of N, the size of risk. For some cases, this requires combining the results of this appendix with those in Appendices B, D, and F. Note that for the case of a split plan, the same general functional form applies to κ_p and κ_e, even though the specific coefficients may make the specific curves look very different. The following table presents the results for the different cases.

$$\kappa = \frac{N(1 - Z)}{Z} \text{ as a function of } N\text{, the size of risk}$$

	No-Split Plan	Split Plan (Primary vs. Excess)
No Parameter Uncertainty No Risk Heterogeneity	Constant	$\frac{\text{Linear}}{\text{Linear}}$
With Parameter Uncertainty No Risk Heterogeneity	Linear	$\frac{\text{Quadratic}}{\text{Linear}}$
No Parameter Uncertainty With Risk Heterogeneity	$\frac{N}{\text{Linear}}$	$N\frac{\text{Linear}}{\text{Quadratic}}$
With Parameter Uncertainty With Risk Heterogeneity	$N\frac{\text{Linear}}{\text{Linear}}$	$N\frac{\text{Quadratic}}{\text{Quadratic}}$

APPENDIX H

MAXIMUM EFFICIENCY AS A FUNCTION OF LOSS LIMIT AND SIZE OF RISK

In this appendix, the behavior of the maximum efficiency with changing loss limit and size of risk is explored. The observed behavior for the examples in Meyers's Sections 5 and 6 is explained in terms of the underlying mathematics and the specific choices of parameters for those examples. Which example is a better approximation to a real world application of experience rating will determine whether the loss limit used should increase significantly for large risks.

Using the credibilities from formula 6.2 in the main text, and the notation of Appendix G, the maximum efficiency is given by

$$\text{Maximum Efficiency} = \frac{(c + s)^2}{(c + d + 2s)a} = \frac{(\hat{c} + \hat{s})^2 N^4}{(\hat{c} + \hat{d} + 2\hat{s})\, N^2\, (\hat{t}N + \hat{c}N^2)}$$

$$= \left[\frac{(\hat{c} + \hat{s})^2}{\hat{c}\,(\hat{c} + \hat{d} + 2\hat{s})}\right] \frac{N}{N + \dfrac{\hat{t}}{\hat{c}}}\,.$$

The first term is independent of N.[80] It can be rewritten as

$$\left[\frac{(\hat{c} + \hat{s})^2}{\hat{c}(\hat{c} + \hat{d} + 2\hat{s})}\right] = 1 - \frac{\hat{c}\hat{d} - \hat{s}^2}{\hat{c}(\hat{c} + \hat{d} + 2\hat{s})}$$

We expect $\hat{s}^2$ to be close to $\hat{c}\hat{d}$, since we expect the hypothetical mean primary losses to be highly correlated with the hypothetical mean excess losses. (If they were perfectly correlated then $\hat{s}^2 = \hat{c}\hat{d}$.) Thus we expect the first term in the expression for the maximum efficiency to be close to one. In fact, for the examples here, that is the case, as shown in the table below.

The second term in the expression for the maximum efficiency is the Bühlmann credibility used by Meyers. It varies with the loss limit only in so far as $K = \hat{t}/\hat{c}$ does. The smaller K, the larger the second term. However, in general, we do not expect the second term to be extremely sensitive to K.

Thus neither term is expected to be extremely sensitive to the loss limit chosen. In fact, the observed optimal efficiencies for fixed N are relatively insensitive to the loss limit.

---

[80] In fact, $\hat{c} + \hat{d} + 2\hat{s}$, which is part of the denominator, is the variation of the hypothetical means of the total losses, and is thus independent of the loss limit chosen.

For the examples in Meyers's Table 5.4, the variation of the second term with loss limit is more important than that of the first term. Thus, selecting the smallest K will produce the largest efficiency regardless of N. As shown in the table below, a loss limit of 2.5 gives the smallest K. As was seen in Table 7.2 the largest efficiency was indeed obtained by taking a loss limit of 2.5 regardless of N, the size of the risk.

In the example in Meyers's Section 6, the first term increases significantly with the loss limit. Thus in order to get the largest efficiency, we have a conflict between choosing a smaller value of K (low loss limit) and a larger value of the first term (high loss limit). For larger N, the second term depends less on K, thus the first term is relatively more important. Thus we expect the optimal loss limit to increase significantly with size of risk. This was indeed the behavior observed in Table 10.1.

EXAMPLE AS PER MEYERS'S TABLE 5.4

Limit ($000's)	$\frac{(\hat{c} + \hat{s})^2}{\hat{c}\,(\hat{c} + \hat{a} + 2\hat{s})}$	$K = \frac{\hat{t}}{\hat{c}}$
1	.988	17.2
2	.994	13.6
2.5	.995	13.5
3	.996	13.7
4	.997	14.5
6	.999	16.3
8	.999	18.2
12	1.000	21.7
16	1.000	24.9

EXAMPLE AS PER MEYERS'S SECTION 6[81]

B ($000's)	$\dfrac{(\hat{c} + \hat{s})^2}{\hat{c}\,(\hat{c} + \hat{d} + 2\hat{s})}$	$K = \dfrac{\hat{t}}{\hat{c}}$
.1	.834	.82
.2	.876	1.14
.5	.922	1.74
1	.949	2.37
2	.969	3.19
5	.987	4.66
10	.995	6.13

[81] Meyers's example corresponds to $N = 10$, B = $2000, except that here a claim limitation of $100,000 has been used, and the Weibull distribution has been approximated by a discrete analog. See Section 10 of this discussion.

APPENDIX I

CALCULATION OF THE QUANTITIES ENTERING THE CREDIBILITY FORMULA

In this appendix, expressions for the parameters entering the formulas for credibility will be derived. The results are summarized at the end of the appendix. We will assume that the frequency and severity are independent and that the individual claims are taken independently from a size of loss distribution.[82]

Let ϕ be a set of parameter(s) that describes the claims process.[83]

We assume that ϕ may take on different values, with probability density function $f(\phi)$; f is commonly referred to as the mixing distribution.

Let θ be the parameter(s) which specify the severity distribution. Assume θ takes on different values, with probability density function $g(\theta)$. Similarly, let ψ be the parameter(s) which specify the frequency distribution. Assume ψ takes on different values, with probability density function $h(\psi)$.

Since frequency and severity have been assumed to be independent we have:

$$f(\phi) = g(\theta)\, h(\psi)$$

Let

E_p = expected value of the primary losses taken over all values of ϕ.
$E_p(\phi)$ = expected value of the primary losses given a specific set of parameters ϕ.

Use a similar definition of the corresponding symbols for excess losses.

Definitions

Define the following quantities, which will be useful:

$\bar{n}$ = average frequency
m_p = average primary portion of a claim
$= E_p \div \bar{n}$ = primary severity

[82] This assumption is made by Meyers in his Algorithm 6.1. For real risks, this may not be true. In Meyers's Section 5, although for a given risk the frequency and severity are independent, the frequency and severity between risks are not independent, thus the formulas in this appendix do not all apply to that situation.

[83] Following Meyers's Algorithm 3.2, ϕ is assumed constant over time for an individual risk, but may be different for different risks. In fact, for real risks ϕ varies over time, as noted by Meyers in his discussion of the Bailey and Simon results on the credibility of a single private passenger car.

m_e = average excess portion of a claim
= $E_e \div \bar{n}$ = excess severity
α_f = parameter variance of the frequency
α_p = parameter variance of the primary severity
= variance of the hypothetical mean primary severities
α_e = parameter variance of the excess severity
β_f = expected value of the process variance of the frequency
β_p = expected value of the process variance of the primary severity
β_e = expected value of the process variance of the excess severity
γ = covariance of hypothetical mean primary severity and excess severity
ζ = expected value of the process covariance of the primary severity and excess severity

Also let

$m_p(\theta)$ = expected value of primary severity for a specific set of parameters θ

$$m_p(\theta) = E_p(\theta) \div \bar{n}$$
$$m_e(\theta) = E_e(\theta) \div \bar{n}$$

Derivation of Results

Let c = variance of the hypothetical means of the primary losses.

$$\begin{aligned} &= \int E_p^2(\phi)\, f(\phi)\, d\phi - E_p^2 \\ &= \int \bar{n}^2\, (\psi)\, h(\psi)\, d\psi \int m_p^2(\theta)\, g(\theta)\, d\theta - \bar{n}^2 m_p^2 \\ &= (\alpha_f + \bar{n}^2)\, (\alpha_p + m_p^2) - \bar{n}^2 m_p^2 \\ &= \alpha_p\alpha_f + \alpha_p\bar{n}^2 + \alpha_f m_p^2 \end{aligned}$$

Similarly, let d = variance of the hypothetical means of the excess losses.

$$d = \alpha_e\alpha_f + \alpha_e\bar{n}^2 + \alpha_f m_e^2$$

Let a = total variance of the primary losses. By a well-known result used by Meyers,[84] total variance equals parameter variance plus expected value of the process variance.

Thus

$a = c$ + expected value of the process variance of the primary losses.

[84] This result is a special case of a result derived for covariances in Appendix A of this discussion.

The process variance of the primary losses can be put in terms of the frequency and severity in the usual manner.[85]

$$a = c + E_\theta[m_p^2(\theta) \text{ (process variance of frequency) } + \bar{n}(\theta) \text{ (process variance of primary severity)}]$$

$$= c + E_\theta[m_p^2(\theta)\, \overline{\beta}_f + n\beta_p]$$
$$= c + (\alpha_p + m_p^2)\, \beta_f + \bar{n}\beta_p$$
$$= (\alpha_p + m_p^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_p + \bar{n}\beta_p$$

Similarly, let b = total variance of the excess losses.

$$b = (\alpha_e + m_e^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_e + \bar{n}\beta_e$$

Let s = covariance of the hypothetical means of the primary and excess losses

$$s = \int E_p(\phi)\, E_f(\phi)\, f\phi\, d\phi - E_pE_e$$
$$= \int m_p(\theta)\, m_e(\theta)\, g(\phi)\, d\phi \int \bar{n}^2(\psi)\, h(\psi)\, d\psi - \bar{n}^2 m_p \mathrm{m}_e$$
$$= (\gamma + m_pm_e)(\alpha_f + \bar{n}^2) - \bar{n}^2 m_pm_e$$
$$= \gamma\alpha_f + \gamma\bar{n}^2 + \alpha_f m_pm_e$$

Let r = total covariance of the primary and excess losses

The total covariance can be split into two pieces in a manner similar to that for the variance.[86]

Total covariance = parameter covariance +
expected value of the process covariance.

Thus, $r = s$ + expected value of the process covariance.

The process covariance can be written in terms of the frequency and severity in a manner similar to the usual formula for the process variance.[86] Given a set of parameters ϕ:

process covariance of the primary and excess losses = (process covariance of the frequency) (mean primary severity) (mean excess severity) + (mean frequency) (process covariance of the primary and excess severity)

$$= \mathrm{VAR}[n|\psi]\; m_p(\theta)\; m_e(\theta) + \bar{n}(\psi)\; \zeta(\theta)$$

---

[85] This result is also a special case of a result derived for covariances in Appendix A.

[86] This result is derived in Appendix A.

Taking the expected value over all values of the parameters gives the expected value of the process covariance of the primary and excess losses equal to

$$\beta_f(\gamma + m_p m_e) + \bar{n}\zeta.$$

Thus

$$r = s + \beta_f(\gamma + m_p m_e) + \bar{n}\zeta$$
$$r = (\alpha_f + \beta_f)(\gamma + m_p m_e) + \bar{n}^2\gamma + \bar{n}\zeta$$

For the special case of a single split plan[87] ζ has a relatively simple form.

Let the probability density function of the severity be $\pi(x;\theta)$. For a fixed value of θ, the process covariance is

$$\begin{aligned} \zeta(\theta) &= \textstyle\int_0^\infty \mathrm{MIN}[x,L]\ \mathrm{MAX}[0,x-L]\ \pi(x;\theta)\ dx - m_p(\theta)m_e(\theta) \\ &= \textstyle\int_0^L 0\ dx + \int_L^\infty L(x-L)\ \pi(x;\theta)\ dx - m_p(\theta)m_e(\theta) \\ &= Lm_e(\theta) - m_p(\theta)m_e(\theta) \end{aligned}$$

Thus, the expected value of the process covariance is:

$$\begin{aligned} \zeta &= Lm_e - (\gamma + m_p m_e) \\ &= m_e(L - m_p) - \gamma. \end{aligned}$$

Thus, for a single split plan we have

$$r = (\alpha_f + \beta_f)(\gamma + m_p m_e) + \bar{n}^2\delta + \bar{n}(m_e(L - m_p) - \gamma).$$

Summary of Results

$$\begin{aligned} a &= (\alpha_p + m_p^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_p + \bar{n}\beta_p \\ b &= (\alpha_e + m_e^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_e + \bar{n}\beta_e \\ c &= \alpha_p\alpha_f + \alpha_p\bar{n}^2 + \alpha_f m_p^2 \\ d &= \alpha_e\alpha_f + \alpha_e\bar{n}^2 + \alpha_f m_e^2 \\ r &= (\alpha_f + \beta_f)(\alpha + m_p m_e) + \bar{n}^2\gamma + \bar{n}\zeta \\ s &= \gamma\alpha_f + \gamma\bar{n}^2 + \alpha_f m_p m_e \end{aligned}$$

[87] Those dollars below L are primary; those above are excess. The current Workers' Compensation Experience Rating Plan is a multi-split plan. The difference is discussed in Snader [9].

APPENDIX J

GAMMA-POISSON, GAMMA-EXPONENTIAL PROCESS

In this appendix, the results of Appendix I are carried forward for a specific choice of distributions. In addition, the resulting efficiencies are shown for a specific choice of parameters.

We assume a single split rating plan with a loss limit of L. We assume no overall limitation on claims.[88] The frequency is given by a Gamma-Poisson process.[89] The frequency for an individual risk is Poisson, with parameter ψ:

$$\omega(n;\psi) = e^{-\psi}\,\frac{\psi^n}{n!}\,; \qquad \text{mean } \psi, \text{ variance } \psi.$$

In turn, the parameter ψ has a mixing distribution which is Gamma, with parameters of η and ϵ:

$$h(\psi) = \frac{\epsilon^{\eta}\psi^{\eta-1}\,e^{-\epsilon\psi}}{\Gamma(\eta)}\,; \qquad \text{mean } \frac{\eta}{\epsilon}\,,\ \text{variance } \frac{\eta}{\epsilon^2}\,.$$

The severity is given by a Gamma-Exponential process.[90] The severity for an individual risk is exponential, with parameter θ:

$$\pi(x;\theta) = \theta e^{-\theta x}\,; \qquad \text{mean } 1/\theta,\ \text{variance } 1/\theta^2.$$

In turn, the parameter θ has a mixing distribution which is Gamma, with parameters ξ and ν:

$$g(\theta) = \frac{\nu^{\xi}}{\Gamma(\xi)}\,e^{-\nu\theta}\theta^{\xi-1}\,; \qquad \text{mean } \frac{\xi}{\nu}\,,\ \text{variance } \frac{\xi}{\nu^2}\,.$$

We assume $\xi > 2$ so that the resulting Pareto has finite variance. If $2 > \xi > 1$ the means are finite, but the excess variances are infinite, and the following formulas are not valid.[91]

---

[88] The mathematics are only slightly more complicated for an overall limitation. However, an overall limitation applied in both Sections 7 and 10 of this discussion.

[89] The resulting overall frequency distribution is a negative binomial.

[90] The resulting overall severity distribution is a Pareto. The exponential distribution is a special case of the Gamma. The more general Gamma-Gamma process results in a Generalized Pareto Distribution; see Hogg and Klugman [10]. While the Gamma-Gamma is probably a better model of reality, the mathematics here would be much more complicated.

[91] With an overall limitation on losses, the variances would be finite.

Quantities Entering the Credibility Formulas

$\bar{n} = \eta/\epsilon$

β_f = expected value of the process variance of the frequency $= \bar{n} = \eta/\epsilon$

α_f = variance of the hypothetical mean frequencies $= \eta/\epsilon^2$

$$m_p(\theta) = \frac{1}{\theta} - \frac{e^{-\theta L}}{\theta}$$

$$m_e(\theta) = \frac{e^{-\theta L}}{\theta}$$

$$m_p = \frac{\nu}{\xi - 1} - \frac{\nu}{\xi - 1}\left(1 + \frac{L}{\nu}\right)^{1-\xi}$$

$$m_e = \frac{\nu}{(\xi - 1)}\left(1 + \frac{L}{\nu}\right)^{1-\xi}$$

α_p = variance of the hypothetical mean primary severity

$$= \frac{\nu^2}{(\xi - 1)(\xi - 2)} - \frac{2\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} + \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_p^2$$

α_e = variance of the hypothetical mean excess severity

$$= \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_e^2$$

γ = covariance of the hypothetical mean primary and excess severity

$$= \frac{\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_p m_e$$

Process Variance of the Primary Severity (θ) =

$$\frac{1}{\theta^2} - \frac{2Le^{-\theta L}}{\theta} - \frac{e^{-\theta L}}{\theta^2}$$

Process Variance of the Excess Severity (θ) =

$$\frac{2e^{-\theta L}}{\theta^2} - \frac{e^{-2\theta L}}{\theta^2}$$

Process Covariance of the Primary and Excess Severity (θ) =

$$\frac{Le^{-\theta L}}{\theta} + \frac{e^{-2\theta L}}{\theta^2} - \frac{e^{-\theta L}}{\theta^2}$$

β_p = Expected Value of the Process Variance of the Primary Severity

$$= \frac{\nu^2}{(\xi - 1)(\xi - 2)} - \frac{2L\nu(1 + L/\nu)^{1-\xi}}{(\xi - 1)} - \frac{\nu^2(1 + 2L)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

β_e = Expected Value of the Process Variance of the Excess Severity

$$= \frac{2\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

ζ = Expected Value of the Process Variance of the Excess Severity

$$= \frac{L\nu(1 + L/\nu)^{1-\xi}}{(\xi - 1)} + \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

A Specific Example

In the Gamma distribution the first parameter controls the shape[92], while the second parameter basically determines the scale once the first parameter is chosen.

The mixing distribution of the frequency is Gamma, with parameters η and ϵ. Let $\eta = 4$ and $\epsilon = 4/N$. Thus since $\bar{n} = \eta/\epsilon$, N is the mean number of claims.

The mixing distribution of the severity is Gamma, with parameters ξ and ν. Let $\xi = 2.5$ and $\nu = 4500$. Thus, since $m = m_p + m_e = \nu/(\xi - 1)$, the average size of claim is $4500/1.5 = 3000$.

Then the resulting efficiencies are as follows:

[92] For parameters η and ϵ, the skewness of the Gamma is two over the square root of η. For η = 1, the Gamma is the exponential distribution. For η large, the Gamma approaches the normal distribution.

EFFICIENCIES

L	$N = 1$	$N = 10$	$N = 100$
0	31.43%	82.09%	97.86%
100	32.11	82.25	97.87
1000	32.01	82.23	97.87
10000	31.67	82.14	97.87
∞	31.43	82.09	97.86

We note that for this example, the efficiencies are almost independent of the loss limit L. In fact, the single split plan has no practical advantage over the no split plan ($L = 0$ or $L = \infty$).