EXPERIENCE RATING REASSESSED

BY

ROBERT A. BAILEY

Introduction

The heterogeneity of risks and the need for experience rating is a widespread problem and is not confined only to insurance or to casualty insurance. An illustration is the familiar passage: "Beware of false prophets, which come to you in sheep's clothing, but inwardly they are ravening wolves. Ye shall know them by their fruits." (Matthew 7:15)

The development of commercial package policies with their combination of a broad spectrum of property and casualty coverages has brought about the need for reassessing the different experience rating plans which we presently use in the various separate lines of insurance. When one policy embraces several lines of insurance the question naturally occurs as to which of the present experience rating plans, if any, is appropriate for the package. These new packages probably need experience rating more than the separate coverages where the rates, classifications and coverages have been seasoned by many years of experience. Just as the experience incurred under the homeowners policies led to a number of changes in the coverage and rating of those policies, so also the experience under these new commercial package policies will undoubtedly lead to modifications and changes in the original programs. In such a transitional period the experience incurred by an individual risk is of particular value in adjusting the rate closer to the inherent hazard of that risk.

There is quite a variety of experience rating plans to choose from, ranging from the multiple location experience rating plan for fire insurance on contents to an interesting one which is used for Bankers and Brokers Blanket Bonds which sets the modification equal to .500 plus ½ the loss ratio plus the square of the loss ratio, subject to certain limitations. The rationale of the latter plan, while based on sound principles, must certainly seem elusive to some of the policyholders and agents. Nevertheless the various experience rating plans have several things in common although in varying degrees. Every plan limits the effect of a single large loss. This is accomplished in many ways—such as by credibility factors and/or limitations on the largest loss or on all large losses. In addition the compromise is often evident between the desire to give the best risks as large a credit as possible and the desire to prevent large fluctuations in the rating.

Fundamental Criterion for Experience Rating

It is not accidental that the various experience rating plans have these common features. These common features all represent attempts to satisfy the fundamental criterion for experience rating, which is:

I. Each dollar of loss, or absence thereof, should contribute to the risk's adjusted rate an amount equivalent to the amount of information it provides regarding the future losses of the same risk for the same amount of exposure.

A number of other criteria are imposed which are in the nature of limitations on this fundamental criterion. They are:

II. The risk's premium should not fluctuate widely from year to year. If it fluctuates too widely, the purpose of insurance is defeated.

III. One dollar of actual loss should not increase the adjusted losses by more than one dollar. Otherwise the insured might find it to his advantage to pay his own losses. (The term "adjusted losses" means the weighted average of the actual and the expected losses which is used to determine the adjusted rate for the risk.)

IV. The experience rating plan should not be too expensive to administer.

Basic Formula of Experience Rating

Letting f_t represent the frequency of losses of t dollars or more (which is the same as the frequency of the t-th dollar of each loss), E the expected losses contemplated by the tariff or standard rates, M the experience rating modification, K a constant, ω the size of the largest possible loss, $\underline{E}$ ( ) the expected value of whatever is inside the parentheses, and Z_t the multiple regression coefficient between f_t and ME, the basic formula would be:

$$ME = K + \sum_{t=1}^{\omega} Z_t f_t$$

or
$$ME = E + \sum_{t=1}^{\omega} Z_t \left[f_t - \underline{E}(f_t)\right] \tag{1}$$

since $K = E - \sum_{t=1}^{\omega} Z_t \underline{E}(f_t)$, if $\underline{E}(ME) = E$; that is, the plan should balance.

(Ideally, the experience period should be subdivided into several time intervals with different Z_t for each interval.) The fundamental criterion would be satisfied if we had sufficient data available to calculate these multiple regression coefficients. The difficulty is that we will probably never have sufficient data available to calculate all or even many of these regression coefficients. It is rare that we get enough data to calculate even one coefficient.

If we do get enough data to calculate one coefficient it is usually Z_1 which corresponds to the claim frequency. Automobile merit rating statistics have been one such source where risks have been classified according to their claim frequency and where we can obtain Z_1. For example, using the data on page 163 of [5] for class 1 private passenger cars in Canada, we find

M = .945 for risks which had no losses of 1 dollar or more during an experience period of one year.

$E = 25.44$ (That is, \$84,607,000/3,325,714)

$\underline{E}(f_1) = .087$ (Equivalent to the average claim frequency)

Our formula is $ME = E + Z_1 [f_1 - \underline{E}(f_1)]$. Putting in the known values we obtain $.945 \times 25.44 = 25.44 + \overline{Z}_1 (0 - .087)$

$Z_1 = 16.08$

This illustrates that when we calculate only a single regression coefficient for f_1, which is the claim frequency, we can usually expect a value for Z_1 which far exceeds 1.000. The other Z_t exist, of course, but when we classify risks according to their claim frequency and disregard the size of each loss, our implicit assumption is that all Z_t except Z_1 are equal to zero, and hence we throw all the weight on Z_1. For the approximately 90% of all Canadian automobile liability insurance claims which exceed \$16.08, this presents no problem. But for any loss less than \$16.08, we are adding more than one dollar to the adjusted losses for each dollar of actual loss. As a natural consequence, some of these small losses are not reported to the insurers by the insureds.

If we had sufficient data to calculate more than one Z_t, the value of Z_1 would undoubtedly be less. But if we had sufficient data to calculate many or all Z_t, we would have so much data that in all likelihood our classification plans would be so thoroughly refined and the rates so accurate that the need for experience rating would be considerably reduced. This could be termed the Actuarial Theory of Indeterminancy which would state that when we get sufficiently refined statistics in sufficient volume to be able to determine the correct values for an experience rating plan, we won't use the information that way because we can then determine a far better class plan instead. It is when the data is limited and hence the rates less accurate that the need for experience rating is greater. And the need for experience rating is greatest when we have no data at all, such as the case with new commercial multiple line packages. So it appears that in practice we will have to rely heavily on judgment to establish our Z_t.

The True Values of Z_t

If the inherent severity of claims is the same for every risk, and the only difference among risks is in their inherent frequency, general reasoning tells us that f_1 would include all the information contained in the experience, and hence Z_1 would be a large positive number, its size dependent on the amount of dispersion in the inherent frequencies, and all other Z_t would be zero. But we know that risks differ in their inherent severity of claims.

If the inherent severities of claims vary by risk but are independent of the inherent frequencies, we can conclude that each f_t provides additional information and that each f_t is positively correlated with the total inherent hazard of the risk, hence all Z_t would be greater than zero and Z_1 would be much less than under the previous assumption. The values of Z_t would depend on the dispersion of the inherent frequencies and severities.

The assumption of independence between frequency and severity has been customarily made by authors who have discussed the mathematical distributions of actual losses, that is, the mathematical theory of risk. See [8], Sections 3.1 and 6.1. See also [4], p. 22, "The Unsolved Problem". The assumption of independence greatly simplifies the mathematics. While it is not an inappropriate assumption in collective risk theory, it is an inaccurate assumption for the experience rating of individual risks. This paper by no means solves "the unsolved problem" but just because we cannot solve the mathematical theory behind a problem does not mean that we are free to ignore the problem.

If the inherent frequencies and inherent severities vary by risk and if they are correlated either positively or negatively, the values of some Z_t can easily be less than or equal to zero. Some can also be greater than 1.000. This can be verified by the reader by setting up some simple models and calculating the values of Z_t.

It can be seen that the true values of Z_t for a class of risks may have a considerable range, are not restricted to $0 \leqq Z_t \leqq 1$, and that they would not necessarily be constantly increasing or decreasing, all depending on the nature of the variation in the inherent hazards of the risks. However our knowledge of the variation in the inherent frequencies and severities and the correlation between them is incomplete, to say the least. In such a situation we must use our best judgment to estimate the values of Z_t. While our estimates will probably be incorrect to some extent in every case, if our estimates produce a rate for each risk which is sufficiently more accurate than the tariff or standard rates to justify the expense of experience rating we will have accomplished our purpose. And strange as it may seem, our chances of accomplishing this are greatest when the least data is available, that is, when the tariff or standard rates are themselves based largely on judgment.

Estimates of Z_t

One possible method of estimating the values of Z_t is to proceed as follows. Let us ask ourselves what is the indicative value of the t-th dollar of each loss. For a single risk, the actual number of such losses will follow a Poisson distribution. In the case of a Poisson distribution it has been shown, [4] pp. 14 & 15, that the best unbiased linear estimate of the true expected number of losses, T, (the inherent hazard of the risk) per unit of exposure when we have observed n losses in N exposures is

$$\underline{E}\left(\frac{T}{N}\middle|\frac{n}{N}\right) = Z\frac{n}{N} + (1\text{-}Z)m \tag{2}$$

and that

$$Z = \frac{Nm}{Nm + \frac{m^2}{\sigma^2}} \tag{3}$$

where: m and σ^2 are the mean and variance of T per unit of exposure for all risks in the same rating class, and Nm is the expected number of losses for a risk with N exposures. (For many lines of insurance, premium could be used as the measure of exposure.)

Returning to the t-th dollar of each loss, we find that it should be given a weight of

$$Z_t = \frac{Nm_t}{Nm_t + \frac{m^2_t}{\sigma^2_t}} = \frac{\underline{E}\ (f_t)}{\underline{E}\ (f_t) + \frac{m^2_t}{\sigma^2_t}} \tag{4}$$

where m_t and σ^2_t are the mean and variance of the *inherent* number of losses of t dollars or more per unit of exposure. σ^2_t can be estimated from an analysis of variance by subtracting m_t from the variance of the *actual* number of losses of t dollars or more per unit of exposure, since m_t equals the part due to chance of the variance of the actual number of losses. σ^2_t/m^2_t can also be estimated by using the technique used in [6].

We will assume that m^2_t/σ^2_t is constant for all t. We make this assumption because it produces credibilities which meet Mr. Perryman's axioms (See below) which are an expression of our intuitive sense of credibility. If we were able to calculate Z_t or m^2_t/σ^2_t from actual experience we would modify this assumption to fit our data, but in the absence of any data, this seems to be a reasonable assumption, and it produces reasonable results.

Mr. Perryman's Axioms

The weight given to a loss of C dollars would be $\sum_{t=1}^{C} Z_t$. Expressed as an average credibility factor, Z, it becomes: $Z = \frac{\sum_{t=1}^{C} Z_t}{C}$ Formula (4) for Z_t is such that Z meets all three of Mr. Perryman's axioms. [10], p. 63.

"(i) the credibility should be not less than zero and not greater than unity.

(ii) the credibility should increase (or more strictly speaking not decrease) as the size of the risk increases.

(iii) As the size of the risk increases, the percentage charge for any loss of given size should decrease."

Somewhat as an extension of Mr. Perryman's axioms, we should observe that formula (4) for Z_t also satisfies the following conditions for an individual risk.

(iv) A loss of t dollars has more value than a loss of t-1 dollars.

(v) A loss of 2t dollars has less than twice the value of a loss of t dollars, and far less when t is large in proportion to the size of the risk's expected losses, and almost the same value as a loss of t when t is very large.

(vi) Two losses amounting to a total of t dollars have more value than one loss of t dollars, and similarly three losses totalling t dollars have more value than two losses totalling t dollars.

All this does not necessarily mean that these estimates of Z_t are the best estimates, or even good estimates. All it means is that they are not unreasonable estimates. As mentioned above, it is possible that the actual data may not conform to Mr. Perryman's axioms and the three extensions, but when we do not know what the actual data is, we feel inclined to make our estimates conform to these "axioms". Any measurement of how good these estimates are would require an analysis of the actual experience of experience rated risks. For some studies of the experience of experience rated risks, see [5], [6], [9], [11] and [12].

Primary Losses

If we give a weight of Z_t to the t-th dollar of each loss, the experience rating formula becomes:

$$M = \frac{\sum_{t=1}^{\omega} Z_t A_t + \sum_{t=1}^{\omega} (1-Z_t)\, E_t}{E} = \frac{A_p + E_e}{E} \tag{5}$$

where

$$A_p = \sum_{t=1}^{\omega} Z_t A_t \tag{6}$$

A = actual losses
E = expected losses
t subscript refers to the t-th dollar of each loss
p subscript = primary
e subscript = excess
ω is the maximum size of loss.

It would be interesting to determine the values of A_p produced by $Z_t = \dfrac{\underline{E}(f_t)}{\underline{E}(f_t) + \dfrac{m^2_t}{\sigma^2_t}}$ using some actual data. The primary losses shown below were calculated on the basis of the actual distribution of 139,458 Workmen's Compensation losses during the first half of 1956 in Michigan, and assuming $m^2_t/\sigma^2_t = 1$ for all t. The distribution and some examples of the calculations are shown in the Appendix.

Primary Loss

Actual Loss	*E = 107* *E(f_t) = 1*	*E = 1,070* *E(f_t) = 10*	*E = 10,700* *E(f_t) = 100*	*E = 107,000* *E(f_t) = 1000*	$f_t \div f_t$
10	4	9	10	10	.450
100	17	64	94	99	.100
500	38	204	430	492	.034
1,000	50	300	779	971	.018
5,000	81	587	2,466	4,458	.0039
10,000	92	693	3,347	7,811	.0014
50,000	99	768	4,050	12,545	.00001
110,000	100	774	4,110	13,139	0
average	7.4	29.8	66.1	94.1	

It would usually be considered impractical to have a different table of primary losses for each size of expected losses, particularly if the tables extended down to the most frequent sizes of loss. So let us consider the various possible approximations for primary losses.

The experience rating plan used in most states for Workmen's Compensation insurance probably is the best multi-split plan that can be devised on the basis of judgment and with the restriction that there can be only one table of primary losses. This plan has one table of primary losses for all sizes of risk, and introduces variations by size of risk through a multiplier (called a credibility factor) which varies by size of risk. The combined operation of the table of primary losses and the primary and excess credibility factors adds the following amounts to the adjusted losses for each actual loss. The primary losses shown in the previous table are comparable to the following amounts, since the previous table was developed on the basis that the credibility factors were 1 for the primary losses and 0 for the excess losses.

	Addition to Adjusted Losses (WC Plan — 1961 Revision — Mich.)			
Actual Loss	*E=107**	*E=1,070*	*E=10,700*	*E=107,000*
10	0	1	6	10
100	1	12	59	96
500	7	62	294	478
1,000	13	117	552	917
5,000	33	293	1,383	3,121
10,000	41	360	1,708	5,115
50,000	48	424	2,037	11,730
110,000	48	424	2,037	11,730
average	1.0	8.8	41.3	79.3

*This size not eligible for experience rating.

These additions to the adjusted losses used in WC fulfill all of Mr. Perryman's axioms and the first extension. And they fulfill the second and third extensions for t greater than $750. But they have what appears to be one serious defect. They give insufficient weight to small losses. While a $1000 loss may deserve to add only 117 dollars or 552 dollars respectively to the adjusted losses for risks of size E = $1070 and E = $10,700, certainly a $10 loss should add more than $1.25 and $5.88 respectively to the adjusted losses for risks of these sizes. In other words, a $10 loss on risks of these sizes should be treated as fully credible. Half of all WC losses in Michigan in 1956 were $10 or less. In fact, if it were not for criterion III, we might even be tempted to add more than $10 to the adjusted losses for a loss of $10.

In other casualty lines of insurance the actual losses are limited to an amount that varies by size of risk and then multiplied by a credibility factor which also varies by size of risk. This suffers from the same defect as mentioned above for WC.

Comparisons of the Additions to the Adjusted Losses

Shown below are some comparisons of the amounts added to the adjusted losses. The primary losses calculated from $\sum_{t=1}^{C} Z_t$ are used as the standard for comparison. Compared with this are the amounts added to the adjusted losses by (1) the experience rating plan used in most states for WC, a multi-split plan, (2) the experience rating plan used in Pennsylvania for WC, taken as an example of a single-split plan and (3) a modified single-split plan using 100% of the first I dollars.

Size of Loss	Standard $\underline{E}(f_i) = 1$ $\overline{E} = 107$	Equal Average: WC $E = 890$	Equal Average: 100% of first 11	Equal Average: WC-Pa $E = 1600$ 8½% of first 4674	Minimum Error: WC $E = 440$	Minimum Error: 100% of first 8	Minimum Error: WC-Pa $E = 946$ 5% of first 4500
10	4	1	10	1	1	8	1
100	17	11	11	9	6	8	5
500	38	53	11	43	28	8	25
1,000	50	99	11	85	52	8	50
5,000	81	248	11	397	129	8	225
10,000	92	306	11	397	159	8	225
50,000	99	360	11	397	187	8	225
110,000	100	360	11	397	187	8	225
Average	7.4	7.4	7.3	7.4	3.9	5.9	4.3
Average Error	0	5.4	5.2	6.4	4.6	5.0	5.4

Size of Loss	Standard $\underline{E}(f_i) = 10$ $\overline{E} = 1070$	Equal Average: WC $E = 5540$	Equal Average: 100% of first 188	Equal Average: WC-Pa $E = 8200$ 32½% of first 6334	Minimum Error: WC $E = 4040$	Minimum Error: 100% of first 210	Minimum Error: WC-Pa $E = 4800$ 22% of first 5481
10	9	4	10	3	4	10	2
100	64	43	100	33	35	100	22
500	204	213	188	163	175	210	110
1,000	300	399	188	325	328	210	220
5,000	587	996	188	1,625	820	210	1,100
10,000	693	1,226	188	2,059	1,010	210	1,206
50,000	768	1,444	188	2,059	1,189	210	1,206
110,000	774	1,444	188	2,059	1,189	210	1,206
Average	29.8	29.8	29.8	30.0	24.5	31.3	19.8
Average Error	0	11.5	11.7	18.0	10.7	11.6	16.7

Size of Loss	Standard $E(f_1) = 100$ $\bar{E} = 10{,}700$	Equal Average: WC $E = 47{,}700$	Equal Average: 100% of first 1680	Equal Average: WC-Pa $E = 32{,}000$ 65% of first 12,215	Minimum Error: WC $E = 47{,}700$	Minimum Error: 100% of first 2320	Minimum Error: WC-Pa $E = 24{,}500$ 59% of first 10,428
10	10	9	10	7	9	10	6
100	94	88	100	65	88	100	59
500	430	440	500	325	440	500	295
1,000	779	833	1,000	650	833	1,000	590
5,000	2,466	2,378	1,680	3,250	2,378	2,320	2,950
10,000	3,347	3,383	1,680	6,500	3,383	2,320	5,900
50,000	4,050	6,025	1,680	7,940	6,025	2,320	6,153
110,000	4,110	6,025	1,680	7,940	6,025	2,320	6,153
Average	66.1	66.1	66.1	66.0	66.1	72.4	58.8
Average Error	0	3.7	13.9	21.2	3.7	12.6	20.5

Size of Loss	Standard $E(f_1) = 1{,}000$ $\bar{E} = 107{,}000$	Equal Average: WC $E = 199{,}400$	Equal Average: 100% of first 7,000	Equal Average: WC-Pa $E = 83{,}000$ 89% of first 23,402	Minimum Error: WC $E = 211{,}000$	Minimum Error: 100% of first 10,000	Minimum Error: WC-Pa $E = 95{,}000$ 92½% of first 25,658
10	10	10	10	9	10	10	9
100	99	99	100	89	99	100	93
500	492	494	500	445	495	500	463
1,000	971	968	1,000	890	973	1,000	925
5,000	4,458	4,091	5,000	4,450	4,208	5,000	4,625
10,000	7,811	7,605	7,000	8,900	7,913	10,000	9,250
50,000	12,545	20,467	7,000	20,828	21,560	10,000	23,734
110,000	13,139	20,467	7,000	20,828	21,560	10,000	23,734
Average	94.1	94.1	94.1	94.2	95.8	99.1	98.0
Average Error	0	5.2	7.9	10.5	5.0	6.4	9.9

It can be seen from these comparisons that the single-split plan uniformly produces the poorest fit, as might be expected. The multi-split plan produces an excellent fit in the central range of sizes but in the remainder of the comparisons it is about equalled by the modified single-split. Moreover, all three plans produce poor fits at the smallest sizes, indicating the need for special techniques for small risks. (Small risks present other problems as well, because the assumption of a linear regression formula becomes inaccurate for small risks, as indicated in [2] p. 18 and [4] p. 19.)

The size of E for the "standard" would be changed if we changed our assumption regarding σ^2_t/m^2_t since E is inversely proportional to σ^2_t/m^2_t. For example, if we assumed that $\sigma^2_t/m^2_t = ½$ for all t, then the same "standard" primary losses would be shown for twice the size of E. That is, $E = 107$ would become $E = 214$. From this we can determine the approximate assumptions in the two WC plans regarding σ^2_t/m^2_t. For the multi-split plan, σ^2_t/m^2_t is assumed to be about ⅛ for the smallest sizes of E, increasing to ½ for the largest sizes. For the single-split plan it increases from $\frac{1}{16}$ for the smallest sizes to more than 1 for the largest sizes. It seems unreasonable that σ^2_t/m^2_t should increase with the size of the risk, but this result was probably produced by the attempt to limit the maximum effect of a single loss. There is some available evidence that σ^2_t/m^2_t is larger than what is implicitly assumed by these two plans for the smaller sizes of E. For example, see [6]. But to assume larger values would increase the maximum effect of a single loss and would cause the risk's premium to fluctuate too widely, contrary to criterion II. If we had reason to believe that σ^2_t/m^2_t had larger values and we wanted to recognize them but we still wanted to limit the maximum fluctuation caused by a single loss to the present amounts, we would approach rather closely to the modified single-split which assumes high values of σ^2_t/m^2_t for small and medium sized losses but limits the effect of a single loss to a fixed amount. The modified single-split in effect would ignore the indications of large losses in excess of a certain amount, in order to limit the effect of a single loss. The two WC plans used for comparison have reduced the credibility of all losses, large and small, in order to limit the maximum effect of a single loss, rather than reducing the credibility of only the large losses. This is what has caused the insufficient weights to be given to the small losses.

The consequences of giving a loss less credibility than it deserves were mentioned in [2] where it is stated that "if an arbitrarily chosen credibility . . . is less than:"

$$\frac{\underline{E}(f_t) - \dfrac{m^2_t}{\sigma^2_t}}{\underline{E}(f_t) + \dfrac{m^2_t}{\sigma^2_t}} \qquad \text{(in my symbols)}$$

"it can be shown . . . that the use of the arbitrary credibility has produced a greater error-variance than would have resulted from giving each observation 100% credibility." This situation frequently occurs in WC and elsewhere for the smaller sizes of loss. For example, if the credibility of a $100 loss should be 94% as developed for $E = 10{,}700$, then any credibility less than 88% will produce a greater error-variance than would have resulted from using a credibility of 100%. For a \$100 loss when $E = 10{,}700$, the WC plan gives 59% credibility and the WC—Pa. plan gives 38.5% credibility.

The difference between reducing the credibility of all losses and reducing the credibility of only the large losses can be illustrated from Workmen's Compensation. The average D ratio ($D = E_p \div E$) under the WC experience rating formula (1961 revision) is about .600. For a risk with $Z_p = .25$, the proportion of losses which affect the rating is $.600 \times .25 = .15$, and the maximum effect of a single loss is about $\$3400 \times .25 = \$850$. A modified single split using 100% of the first \$850 would permit about .500 of the losses to affect the rating instead of only .15, and even 100% of the first \$500 would permit about .450 of the losses to affect the rating. This may not be much of a problem in Workmen's Compensation insurance where the D ratios are high. But a commercial package policy has considerably lower D ratios because of the greater catastrophe hazards. Hence the maximum possible effectiveness of experience rating measured by the portion of losses which affect the rating is correspondingly less. For commercial package policies, therefore, we need to conserve all the effectiveness we can, and any substantial reduction such as would be caused by an arbitrary reduction in the credibility of small claims as well as large claims could easily prove fatal to the whole experience rating plan for a multiple line policy.

Basing Experience Rating on Experience

The previous comparisons have been made with the assumption that the "standard" formula is correct. While we have good reason to believe that that "standard" is more appropriate than any of the other formulas, there is little reason to believe that it is anywhere near correct. The only way to know would be to analyze the actual experience of experience-rated risks, which unfortunately is either unavailable or difficult to obtain. While we should do the best we can under the circumstances, we should recognize that the extensive use of highly refined and technical judgment can be like straining at gnats, and if we don't use some actual experience to modify our judgment, we may swallow a camel unawares.

An experience rating formula which is not based on experience is somewhat of an anomaly. The merit rating plans in use in private passenger automobile insurance may appear crude in comparison to a highly refined multi-split experience rating formula, but at least they are or will be based on actual experience. And an experience rating formula which is based on experience has a substantial advantage over any experience rating formula based entirely on judgment no matter how carefully refined that judgment may be.

The extensive use of judgment in the design of the experience rating plans in WC insurance where the size of the credits and debits to be given for various specified losses or lack of losses has been based almost entirely on judgment, is comparable to the extensive use of judgment in establishing the size of the credits and debits given in fire insurance for various safety or hazardous features of the risk. In fact a general comparison can be made between

WC and fire insurance on their entire rate making methods. In both lines the statewide rate level and some statewide class relativities are based on experience. From these class rates in both lines, credits and debits are given to recognize the peculiarities of individual risks, and the size of the credits and debits are based in both lines almost entirely on judgment. In WC insurance the credits and debits are for the presence or absence of certain previous losses and in fire insurance they are for the presence or absence of certain safety or hazardous characteristics of the risk. While the details are different of how the experience and judgment are used in the two lines, the basic role of judgment is the same. In both cases the judgment used to determine the size of the credits and debits and the relationships among the various credits and debits has been very carefully refined. Both systems are probably equally as sound and both probably would benefit equally as much from the use of more experience, which unfortunately is equally difficult to obtain in both lines of insurance.

When the Tariff Rate is Not Based Entirely on Experience

Another assumption made in the previous developments is that the manual rate is equal to the average true rate for all risks with the same manual rate. In the formula

$$\underline{E}\left(\frac{T}{N}\,\middle|\,\frac{n}{N}\right) = Z\frac{n}{N} + (1 - Z)\,m \qquad (2)$$

m was assumed to be

$$m = \underline{E}\left(\frac{T}{N}\right)$$

This is a good assumption when the tariff rate is based on experience. But it is a questionable assumption in the particular case of a new commercial package policy where modifications in rates and coverages have been based on judgment, and it is questionable also in the case of many long-standing property insurance rates where the relativities for many important elements in the rates, such as for watchman service, non-standard floor openings, size of building, and protection are based largely on judgment.

When the m in formula (2) is not equal to the mean for the class or at least has limited credibility, what kind of experience rating formula should we use? (Enter Judgment again.)

Criterion I places considerable reliance on the tariff rate in keeping with the assumption that the tariff rate is a reliable average of the risks in the class. When we cannot make such an assumption it seems that the best course of action would be to place less reliance on the tariff rate, in fact, as little as possible. To do this we should base as much of the rate as possible on the experience, consistent with the credibility of the experience, and use judgment to estimate the remainder of the rate. This is equivalent to revising the fundamental criterion for experience rating shown at the beginning of this paper to read as follows:

Ia. The proportion of total losses which influences the rating, $\frac{E_pZ_p + E_eZ_e}{E}$, should be as large as possible. This is the same as saying that the average credibility, $DZ_p + (1\text{-}D)\ Z_e$, should be as large as possible.

Criterion Ia alone would make $Z_p = Z_e = 1$ which obviously is too high. So we must define what we mean by "consistent with the credibility of the experience". Let us define this as follows:

IIa. A maximum single loss should not increase the adjusted losses by more than a predetermined percentage, h, of the expected losses, E.

It will be noted that IIa is an approximation to criterion II, an approximation which is more definite and somewhat narrower in scope, and one which is used in many experience rating plans.

Now let us state criterion III mathematically as follows:

IIIa. $Z \leqq 1$

Criterion IV is not capable of precise mathematical expression so let us leave that one to judgment.

Derivation of the Plan From the Criteria

The three criteria expressed mathematically are as follows:

$DZ_p + (1\text{-}D)\ Z_e$ is a maximum

$IZ_p + (C\text{-}I)\ Z_e \leqq Eh,\ C > I$

$Z \leqq I$

where I = the loss limitation which defines primary losses

$$D = E_p \div E = \frac{\int_0^I Cf_C\, dC + I \int_I^\infty f_C\, dC}{\int_0^\infty Cf_C\, dC}$$

f_C = number of claims of size C

In [7] Mr. Borch shows that if we are presented with the problem of reducing the variation in the expected losses as much as possible with the transfer to a reinsurer of a minimum amount of expected losses, we should buy a 100% excess of loss contract (assuming that the expense and contingency loadings would be the same percentage for any type of re-insurance contract). The point at which the reinsurance would attach would be selected so that both the variation of the retained losses and the expected amount of the ceded losses would be within acceptable bounds. In other words, for a selected level of stability and assuming the same percentage for expense and contingency loadings, a 100% excess of loss contract will require the smallest transfer of premium to the reinsurer. Or, for a selected amount of reinsurance premium and assuming the same expense and contingency loadings, a 100% excess of loss contract will produce the greatest reduction in the variation of the retained expected losses. While Mr. Borch's paper deals with stop loss reinsurance (yearly aggregate) the reasoning is equally applicable for our purposes for excess of loss (each loss) reinsurance. Mr. Borch's conclusion is the same as saying in our criterion Ia that Z_e should equal zero. It means that in order to retain in the portfolio (or in the rating) as large a portion of the total losses as possible and at the same time to make that portion meet a selected level of stability, we should include 100% of all losses within an appropriately selected limitation and exclude 100% of the excess losses. This is the modified single split discussed earlier. It should be obvious, just from general reasoning, that 100% of the first $100 of each loss represents a larger portion of total losses and has a smaller coefficient of variation, $\frac{\sigma}{m}$, than for example, 10% of the first $1,000 of each loss, although both would produce the same h.

For the moment, let us use a single split, at I, of losses and set $Z_e = 0$ in criterion Ia but defer consideration of the fact that the conclusions of Mr. Borch's paper seem also to specify that $Z_p = 1$.

We may now express the three criteria as follows:

$$\text{DZ is a maximum} \tag{7}$$

$$IZ = Eh \tag{8}$$

$$Z \leqq 1 \tag{9}$$

We seek the best simultaneous solution of these three criteria. From (8) we obtain $Z = \frac{Eh}{I}$. Substituting in (7) we obtain $\frac{D\,Eh}{I}$ is a maximum, or $\frac{D}{I}$ is a maximum since Eh is a constant. From inspection of the definition of D it is evident that D is a function of I and that

$\frac{D}{I}$ is continuous,

$$\frac{d}{dI}\frac{D}{I} \leqq 0,$$

$$\lim_{I\to 0}\frac{D}{I} = \lim_{I\to 0}\frac{\int_0^I \frac{C}{I} f_C\, dC + \int_I^\infty f_C\, dC}{\int_0^\infty C\, f_C\, dC} = 1 \div \text{average claim cost}$$

and $\lim_{I\to\infty}\frac{D}{I} = 0$

Therefore in order to make $\frac{D}{I}$ as large as possible we should make I as small as possible. From (8) we see that this is done by making Z as large as possible consistent with (9). Hence we obtain Z = 1 which is in agreement with the result derived by Mr. Borch and discussed above. We also obtain $I = Eh$. Thus it appears that the experience rating formula which represents the best simultaneous solution of the three criteria listed above is

$$M = \frac{E + A_p - E_p}{E} = \frac{A_p + E_e}{E} \tag{10}$$

Where $I = Eh$

and $A_p = 100\%$ of the first Eh dollars of each loss.

The part of the rate based on actual experience is $\frac{A_p}{E}$ and the part based on judgment is $\frac{E_e}{E}$. $\frac{A_p + E_e}{E}$ will be a better estimate of the true rate for the risk than E will be if A_p is correct and if we are able to estimate E_e with less absolute error than E. A_p will not be precisely correct, but it has a high probability of being closer to the true value than E_p. Moreover, A_p is unbiased over the long run, unlike E_p. A_p is subject to some chance variation, but with a proper choice of h, this variation will be within acceptable limits. E_p has no variation unless we consider the variation between the values of E_p estimated by different ratemakers for the same risk. If it were not for the restrictions imposed by rate regulation, this latter variation in E_p could easily be greater than the chance variation in A_p. Finally, it seems reasonable that we should be able to estimate part of the rate, E_e, with less absolute error than we can estimate the whole rate, E.

Criterion I puts less weight on the small losses because Criterion I assumes that the present rate is reasonably accurate and puts more reliance on it. Criterion Ia puts as little weight as possible on the present rate in keeping with the assumption that the present rate may not be very accurate at all.

Rationale

The formula $\frac{A_p + E_e}{E}$ is similar to an excess of loss contract or a deductible plan or a retrospective rating plan without a minimum where the insured pays the full cost of losses below his retention and buys insurance at a fixed cost above his retention. It also is similar to the Comprehensive Medical insurance plans which have become widespread in recent years as a replacement for the conventional hospital and surgical plans which provide first dollar coverage and limit the benefits per day and per procedure.

The formula $\frac{A_p + E_e}{E}$ is also a very simple formula. Oddly enough, its simplicity may be a drawback, because this plan is just a small step away from self-insurance. The small step is the expense loading that the company applies to the losses which the insured will weigh against the value he receives for the services rendered. The complexity of most other plans, along with their credibility weighting, obscures the expense loading and confuses everyone alike, including the insured.

Alternate Derivations

The same result as formula (10) can also be derived as follows. When the m in formula (2) is not equal to the mean for the class or at least has limited credibility, that is, when the tariff or standard rates are not based on a reliable volume of data for that class, σ^2_t/m^2_t will be increased if σ^2_t is measured from the class rate, rather than from m_t, the true class average. Hence Z_t will be correspondingly increased. This is in keeping with the concept that the less reliable the tariff rate is, the more weight we should put on the actual experience for the risk. However, even though the credibility for experience rating would justify large weights to be put on the risk's experience, we should not permit the weights to be so large that they violate criterion II. In effect, we are seeking the best compromise between the "Greatest Accuracy" credibility and the "Limited Fluctuation" credibility discussed in [1] pp. 63-65 in the chapter on "Two Kinds of Credibility". If we apply a limitation on the effect of each loss of Eh as in criterion IIa, but use the full weight justified by the experience rating credibility for smaller losses, we obtain something very close to the modified single-split of 100% of the first Eh dollars of each loss. How close it is can be seen by truncating the theoretical primary losses shown above for E = 107, E = 1,070, E = 10,700 and E = 107,000 at selected values of Eh, and considering the effect of increased values of σ^2_t/m^2_t.

Another derivation of formula (10) can be based on [4] pp. 21 & 22, "Primary and Excess Values" where it was shown that the first J dollars of each loss should be given 100% credibility and that the excess portions should be given a lesser weight. If we limit the maximum effect of a single loss to Eh in order to meet criterion IIa, we obtain formula (10) since h is usually less than 1, and J as defined in [4] is close to E.

Comparison With Other Experience Rating Formulas

A number of comparisons have already been made, but a comparison with the plans which have widespread use in Workmen's Compensation and the liability lines would be of interest. These are formula plans and permit a ready comparison. Many of the other lines of insurance use tabular plans which are more difficult to compare exactly, although the tabular plans generally are based on similar underlying formulas.

The formula developed above is:

$$M = \frac{E + (A_p - E_p)}{E} = \frac{A_p + E_e}{E} \qquad (10)$$

This compares with (when $Z_e = 0$):

$$M = \frac{E + (A_p - E_p)\, Z_p}{E} \qquad \text{(WC)}$$

$$M = \frac{E + (A_p - E)\, Z_p}{E} \qquad \text{(WC-Pa)}$$

$$M = \frac{E + (A_{pb} - E_b)\, Z_b + E_i \dfrac{A_{pb} - E_b}{E_b} Z_b}{E} \qquad \text{(Liability)}$$

and (when $Z_e > 0$): (Z_e never > 0 for WC-Pa and for liability in many states)

$$M = \frac{E + (A_p - E_p)\ Z_p + (A_e - E_e)\ Z_e}{E} \qquad \text{(WC)}$$

$$M = \frac{E + (A_{pb} - E_b)\ Z_b + (A_{pi} - E_i)\ Z_i + E_i \dfrac{A_{pb} - E_b}{E_b} Z_b\ (1 - Z_i)}{E}$$

(Liability, in some states)

where the subscripts mean:

- p primary
- e excess
- pb primary basic limits
- b basic limits
- pi primary increased limits
- i increased limits

The loss limitation, I, is constant in the WC plan but varies by size of premium in the other plans. All these plans have a built-in limitation on the effect of a single large loss (usually about 25%).

D Ratios

Any experience rating plan which uses a loss limitation must cope with D ratios. This is a vexing problem but an unavoidable consequence of loss limitations. Some plans, such as the plans used in Workmen's Compensation in Pennsylvania and in other casualty lines, do their best to ignore this complication by assuming that the D ratios equal 1.000, that is, that $E_p = E$, or $E_{pb} = E_b$. Probably this is because D ratios would increase the complexity of these plans to an intolerable level. Not much harm is done anyway if the D ratios are close to 1.000. But in the plan (10) developed above, D ratios are important because of the low loss limitations. In addition, D ratios are doubly important for any policy which includes fire insurance because of the large portion of premium devoted to excess losses.

For a new commercial package policy, judgment must play a significant role in establishing proper D ratios just as it has in establishing the rates, at least until a large volume of experience has been accumulated under these new package policies. Claim distributions for fire insurance on commercial properties are difficult to obtain because of the practice in conventional fire insurance of insuring the same building pro-rata in several different policies. The limited data available on the value of deductibles, large and small, is useful. Claim distributions are more readily available for casualty lines and can be used in proportion to their share of the package premium.

For both property and casualty insurance, D ratios will vary by rate class. But for property insurance, D ratios will also vary by size of building (or by size of the probable maximum loss). To some extent, this is true also for casualty insurance as is illustrated in Homeowners insurance where a large portion of policyholders with high valued homes take increased limits for comprehensive personal liability, but where practically none of the policyholders with low valued homes take increased CPL limits. The policyholders with high valued homes evidently believe they have a greater probability of having a large CPL claim, which is equivalent to believing they have lower D ratios, and they are probably correct. However, the problem in property insurance is more serious than in casualty insurance because the variations in the D ratios by size of building for property insurance are more direct.

The claim distributions of many casualty lines can be closely approximated by a log-normal curve. Some available data indicates that this is true also for fire insurance. Because of this, the log-normal curve can be used as an additional guide for establishing the D ratios and also as a graduating device. Methods for fitting the log-normal curve to actual data, and calculating primary and excess ratios from the fitted curve are discussed in [1], p. 58 ff and [4], p. 20 ff. Some other techniques of calculating D ratios are presented in [3].

Summary

The changes and developments which have taken place in the insurance business in recent years have created the need for reassessing our procedures for the experience rating of individual risks, particularly in reference to multiple line policies which include both property and casualty coverages. Are we to cease experience rating the casualty portion of a package policy or are we to begin experience rating the property portion? It seems unreasonable to experience rate only half of a package.

If a package included only casualty coverages it would be easy to find an appropriate experience rating plan. But when it includes both property and casualty coverages, it is a different matter, because property coverages have not usually been experience rated.

When we think of experience rating, most of us think of the type of experience rating used in casualty insurance. Casualty experience rating plans, however, do not work well for property insurance, simply because property insurance is different from casualty. Property insurance has lower claim frequencies and higher catastrophe hazards. So it is not surprising that the casualty experience rating plans do not work well for property insurance.

The same thing is true in other lines of insurance when an experience rating plan is designed especially for a certain type of policy, and such a plan often is unsuitable for other types of policies. For example, take the experience rating plan used in individual life insurance. We don't usually think of the rating plan used in ordinary life insurance as experience rating, but actually it is. The rates for ordinary life insurance are based almost entirely on the length of the insured's own claim-free experience period. The only difference is that the longer the claim-free experience period, the higher the rate. We could never apply an experience rating plan like that to casualty insurance.

So if the experience rating plan used in ordinary life insurance does not fit casualty insurance, it is not surprising that the experience rating plans designed for casualty insurance do not fit property insurance. It's the same old problem of not being able to put new wine into old bottles. But we should not let that prevent us from designing a new bottle.

In this paper, the attempt has been made to go back to the fundamental principles of experience rating and to develop from them the basis for an experience rating plan which will cope with the problems of low claim frequencies and high catastrophe hazards, and which therefore will work well for property insurance, and for combinations of property and casualty.

Experience rating is widely accepted as a sound rating tool. Its soundness can be demonstrated both from the actual experience of experience rated risks and also from actuarial and statistical theory. But we can't just blindly use any experience rating plan. We have to use one which is suited to the type of risk to be experience rated. Some plans are better than others. So we aim for the best plan possible. But we will never have a perfect plan because of the necessity to compromise between actuarial precision and the practical need for simplicity. In the mathematical-actuarial parts of the paper it is shown that one of the best compromises for a commercial multiple-line package policy from a theoretical standpoint and also from a practical standpoint is a type of plan which works very much like a deductible.

A loss limitation per occurrence is established for each risk. The size of the loss limitation is related to the size of the premium for the risk. The risk's actual losses during an experience period of, say, three years are given full credibility up to the loss limitation, and the losses, if any, in excess of the limitation are given no credibility. In effect the premium for the risk is self-rated for coverage up to the limitation, and the portion of the premium for coverage in excess of the limitation is unaffected by the risk's loss experience.

If the risk over the past three years, say, has incurred an average amount of losses within its loss limitation, it gets regular manual rates. If it has had less losses than average within its limitation, or more than average, its rate is correspondingly adjusted. If it has had no losses, it gets credit for the full value of the corresponding deductible.

Multiple-line policies, which are now becoming an important factor in the non-personal lines, present an unusual opportunity for a carefully designed experience rating plan to perform a valuable and much needed function.

Acknowledgment

I wish to acknowledge the assistance of Ruth E. Salzmann and LeRoy J. Simon who read early versions of this paper and made many helpful comments.

APPENDIX

Cumulative Claim Frequency

Michigan — Workmen's Compensation — First Half of 1956

t	$f_t \div f_1$	t	$f_t \div f_1$	t	$f_t \div f_1$	t	$f_t \div f_1$	t	$f_t \div f_1$	t	$f_t \div f_1$
1	1.000	41	.180	160	.076	620	.029	3,100	.0069	26,000	.00006
2	1.000	42	.177	170	.073	640	.028	3,200	.0067	27,000	.00006
3	.850	43	.174	180	.070	660	.027	3,300	.0065	28,000	.00005
4	.750	44	.172	190	.068	680	.026	3,400	.0063	29,000	.00005
5	.660	45	.170	200	.067	700	.025	3,500	.0061	30,000	.00005
6	.600	46	.167	210	.066	720	.024	3,600	.0059	40,000	.00003
7	.550	47	.164	220	.064	740	.023	3,700	.0057	50,000	.00001
8	.510	48	.162	230	.063	760	.022	3,800	.0055	80,000	.00001
9	.470	49	.160	240	.061	780	.022	3,900	.0053	110,000	.00001
10	.450	50	.158	250	.059	800	.021	4,000	.0052	110,001	.00000
11	.430	52	.154	260	.058	820	.021	4,100	.0050		
12	.410	54	.150	270	.056	840	.020	4,200	.0049		
13	.390	56	.146	280	.054	860	.020	4,300	.0047		
14	.370	58	.143	290	.053	880	.019	4,400	.0046		
15	.355	60	.140	300	.052	900	.019	4,500	.0044		
16	.344	62	.137	310	.051	920	.019	4,600	.0043		
17	.333	64	.134	320	.050	940	.019	4,700	.0042		
18	.322	66	.131	330	.049	960	.018	4,800	.0041		
19	.311	68	.129	340	.048	980	.018	4,900	.0040		
20	.300	70	.127	350	.047	1,000	.018	5,000	.0039		
21	.291	72	.124	360	.046	1,100	.016	6,000	.0034		
22	.282	74	.122	370	.045	1,200	.015	7,000	.0025		
23	.273	76	.120	380	.044	1,300	.014	8,000	.0020		
24	.264	78	.118	390	.043	1,400	.013	9,000	.0016		
25	.255	80	.116	400	.043	1,500	.013	10,000	.0014		
26	.249	82	.114	410	.042	1,600	.012	11,000	.0012		
27	.243	84	.112	420	.041	1,700	.011	12,000	.0010		
28	.237	86	.110	430	.040	1,800	.011	13,000	.0009		
29	.232	88	.108	440	.039	1,900	.010	14,000	.0008		
30	.227	90	.107	450	.039	2,000	.010	15,000	.0007		
31	.222	92	.105	460	.038	2,100	.0097	16,000	.00055		
32	.217	94	.103	470	.037	2,200	.0094	17,000	.00039		
33	.212	96	.102	480	.036	2,300	.0091	18,000	.00032		
34	.207	98	.101	490	.035	2,400	.0088	19,000	.00024		
35	.202	100	.100	500	.034	2,500	.0085	20,000	.00020		
36	.198	110	.095	520	.033	2,600	.0082	21,000	.00017		
37	.194	120	.091	540	.033	2,700	.0079	22,000	.00012		
38	.190	130	.087	560	.032	2,800	.0076	23,000	.00009		
39	.186	140	.083	580	.031	2,900	.0073	24,000	.00008		
40	.183	150	.079	600	.030	3,000	.0071	25,000	.00007		

This table is based on the actual distribution of 139,458 claims compiled by the National Council on Compensation Insurance. The actual distribution was grouped into various size intervals, for example, 0-$499, $500-$599. These intervals were subdivided graphically using log-normal graph paper in such a manner as to reproduce the same number and amount of claims in each interval. For simplicity it was assumed that f_t between any two intervals shown in the table above was the same as the f_t shown for the larger end of the interval. That is, $f_t \div f_1 = .016$ for $1001 \leqq t \leqq 1100$. Hence some $f_t \div f_1$ shown in the table for the end of each interval are slightly higher than the values calculated from the actual claim distribution for t equal to the end of the corresponding interval. The average claim produced by this table is 107.2 compared to 107.4 for the actual distribution.

The primary losses were calculated as follows:

	$f_1 = 1$			$f_1 = 10$		
t	f_t	$Z_t = f_t \div (f_t + 1)$	Primary Loss $= \sum_{t=1} Z_t$	f_t	$Z_t = f_t \div (f_t + 1)$	Primary Loss $= \sum_{t=1} Z_t$
1	1.000	.500	.500	10.00	.909	.909
2	1.000	.500	1.000	10.00	.909	1.818
3	.850	.459	1.459	8.50	.895	2.713
4	.750	.429	1.888	7.50	.882	3.595
5	.660	.398	2.286	6.60	.868	4.463
6	.600	.375	2.661	6.00	.857	5.320
7	.550	.355	3.016	5.50	.846	6.166
8	.510	.338	3.354	5.10	.836	7.002
9	.470	.320	3.674	4.70	.825	7.827
10	.450	.310	3.984	4.50	.818	8.645
etc.						

The primary loss, rounded to the nearest dollar, for an actual loss of $10 is $4 for $f_1 = 1$ and $9 for $f_1 = 10$. These are the values shown in the table of primary losses included in the body of the paper in the section, "Primary Losses".

REFERENCES

[1] Bailey, A. L., "Sampling Theory in Casualty Insurance," *PCAS* 1943, p. 31

[2] ———— "A Generalized Theory of Credibility," *PCAS* 1945, p. 20

[3] ———— "Workmen's Compensation D-Ratio Revisions," *PCAS* 1948, p. 26

[4] ———— "Credibility Procedures—La Place's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data," *PCAS* 1950, p. 7

[5] Bailey, R. A. & Simon, L. J., "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car," *PCAS* 1959, p. 159

[6] Bailey, R. A., "Any Room Left for Skimming the Cream?," *PCAS* 1960, p. 30

[7] Borch, K., "An Attempt to Determine the Optimum Amount of Stop Loss Reinsurance," *XVI International Congress of Actuaries*, 1960, Vol. I, p. 597

[8] Cramér, H., *Collective Risk Theory*, Scandia Insurance Company, Stockholm, 1955, Sections 3.1 and 6.1

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

[10] Perryman, F. S., "Experience Rating Plan Credibilities," *PCAS* XXIV (1937), p. 63

[11] Stevens, W. A., "A Review of the Experience of Massachusetts Workmen's Compensation Experience Rated Risks," *PCAS* 1959, p. 87

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