SAMPLING THEORY IN CASUALTY INSURANCE

PARTS III THROUGH VII

BY

ARTHUR L. BAILEY

Introduction

It has been the intent of the writer to develop a fairly complete mathematical theory of the variations in casualty insurance statistics as well as to develop such mathematical aids as are necessary to the computations involved in the use of the theory; but to leave to others any interpretation of the results of application of the theory. In accordance with this intent, only light and superficial treatment by way of illustration has been given in Parts III and VII to the application of the theory to underwriting and to the description of two kinds of credibility. The discussion in Part VI of the use and computation of excess pure premium ratios covers considerable detail felt to be advisable at this time because of the recent extension of rating procedures based on such ratios. The reading of this part at least should bring about a realization that the figures in a table of excess pure premium ratios are by no means exact and are at best only rough approximations.

It will be noted, in Part IV covering the modification of the formulae of Part I to recognize various types of fluctuation other than chance, that the individual observations are in each case weighted to obtain the various averages. This weighting process will appear to the reader either as obviously necessary or as a completely unnecessary and arbitrary complexity. The writer can only state that some very erroneous results were at first obtained when the weighting procedure was omitted.

It will be recognized by some that one of the most important types of variation to be found in casualty insurance statistics has not yet been covered. This is the variation in the accuracy of the data or in the underlying conditions with the passage of time. The effects of such variation will have to be investigated prior to the application of the theories to rate making and experience rating. It is hoped that this can be presented in a subsequent part, together with the applications to rate making, experience rating, and the problems of excess and deductible coverages.

In view of the contemporary work of Mr. Satterthwaite, it seems advisable for the writer to say a few words in defense of having taken a very circuitous route to reach results which to many will appear to be the same as those reached immediately by Mr. Satterthwaite. From the earliest days of statistical theory, there have been two schools of mathematical statistics. One of these is broadly spoken of as the Pearsonian school, being identified by its development of concise algebraic formulae by means of highly advanced and very elegant mathematical processes and by its insistence, in the application of these formulae, that the data fit the formula rather than that the formula fit the data. The other school is known as the Scandinavian or "sledge hammer" school. This latter term describes rather well the processes used by it in the development of formulae. The essential difference, however, is that the entire effort of this latter school is aimed at obtaining formulae which will describe the actual data, with the description being made in terms of symbols having specific interpretations. It will be obvious to anyone who has glanced through either Parts I or IV of this paper that the writer most certainly has used the sledge hammer method as contrasted to the neat development of the "Generalized Poisson Distribution" and the "Hyper-geometric Distribution" by Mr. Satterthwaite. This procedure has been necessary, however, in order to obtain, instead of algebraic formulae with indefinite parameters, a mathematical description of the moments of the various casualty insurance statistics in terms of fundamental statistics subject to exact or approximate determination from actual data.

Attention is called to the proofreading error on page 73 of Part I, where, in the fourth line

$$U_{2:R'} = \frac{H}{E'} U_{2:m} \text{ should read } U_{2:R'} = \frac{H}{E'} + U_{2:m}.$$

Thanks to the assistance of several individuals, Miss Eva Dorenstreich in particular, the following parts are presented with somewhat more confidence as to their algebraic accuracy than were the first two.

III.

Use of Sampling Theory in Individual Risk Underwriting

The tables of the normal sampling range due to chance fluctuations only which were developed in Part II are designed to be used in the evaluation of past individual risk experience in the determination of the future desirability of the risk. The use of these tables can best be explained by their application to individual risks as examples. The examples do not attempt to cover all possible cases but are given only to illustrate that definite answers to specific questions can be provided from the tables of sampling distributions due to chance fluctuations only. Actual problems will frequently require the testing of the risk experience for individual years to point out any trends and will usually involve the separate analysis of the experience of more than one line or type of insurance. Although large risks are used in the examples, the tables are equally applicable to small risks. Likewise, the tables are equally applicable to the combined experience of all risks in a territory, class, or production office; and it is in this application that much of their value can be realized by a carrier.

The A Laundry Company

Let us consider the A Laundry Company, for which the automobile property damage premium for the exposure of the past three years at present manual rates is \$4,531. The permissible loss ratio for this premium is .517, and the average claim cost for the classification in the entire state experience of all companies is \$32. To compare with this, we have from the experience of the risk during these three years 8 claims incurred, with a total loss of \$356 and an average claim cost of \$44.50. Testing this average claim cost first, we refer to Table 11, to find that a ratio of $\frac{\$44.50}{\$32.00} = 1.39$ would not be an unusual ratio for a risk having 8 claims. The table shows us that a ratio of 1.991 would even be quite normal for a risk having 10 claims. We thus find that any unusual element of the risk must lie in the claim frequency of the risk. To test this we refer to Table 5 and enter it with the expected number of claims of $\frac{\$4{,}351 \times .517}{\$32.00} = 73$, to find that a ratio of $\frac{8}{73} = .11$ is entirely below the normal range. This indicates that at manual rates the A Laundry Company is a very desirable risk to put on the books.

In an extreme case of this kind, this same conclusion would undoubtedly be reached by any underwriter without reference to tables of any kind; and a certain amount of competitive rating would probably be encountered on such a risk. We thus have the problem of determining just how much rate recognition can safely be given to the experience of such a risk. Before doing this, let us examine the results produced under the New York State Automobile Experience Rating Plan, which in this case would produce a credit of 46%, or a rate modification of .54. With such a modification, the expected number of claims would be $73 \times .54 = 39$, and the ratio of actual to expected claims would be $\frac{8}{39} = .21$. Referring again to Table 5, we find that the risk is still far below the level of claims expected, even with a 46% credit.

In an open state, where risks are "equity" rated, competition might well be offering such a risk more than a 46% credit; and we must decide for our company how great a credit we are willing to offer such a risk with a reasonable assurance that the risk will continue to be a good risk and not immediately deteriorate. The first decision must be as to the level of significance that our company will adopt as its standard of excellence, the $P = .005$, $P = .025$, $P = .050$, or some other level. Having adopted a particular significance level as our standard, we would find a chart prepared from Table 5 to be of considerable assistance. Figure 2 is illustrative of such a chart and shows for the $P = .050$ and $P = .950$ levels the relationship between actual and expected numbers of claims. Referring to Figure 2, we find for the A Laundry Company that the 8 observed claims would represent a $P = .050$

FIGURE 2

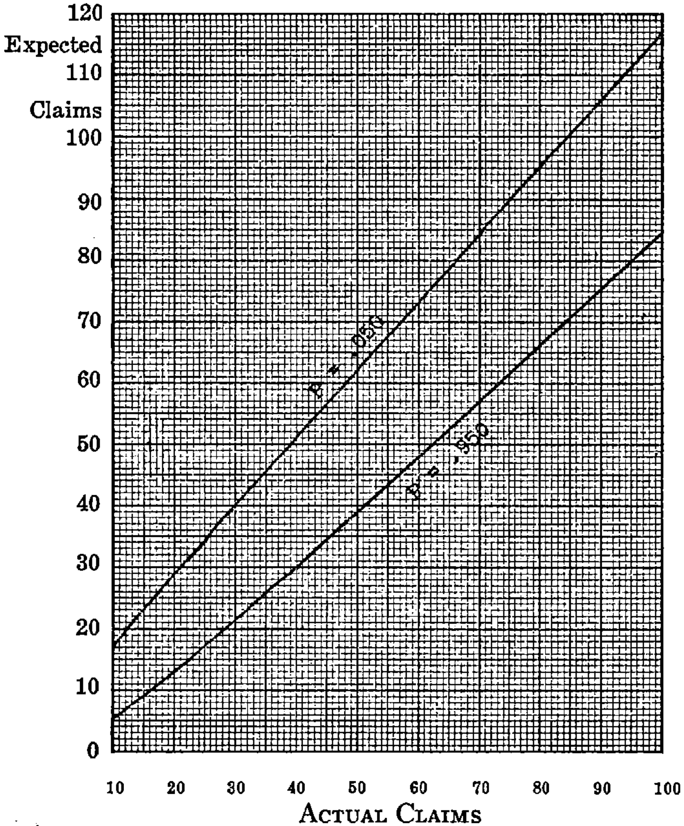
NORMAL SAMPLING RANGE OF EXPECTED NUMBER OF CLAIMS CORRESPONDING TO THE ACTUAL NUMBER OF CLAIMS

PROBABILITY LEVELS OF .050 AND .950

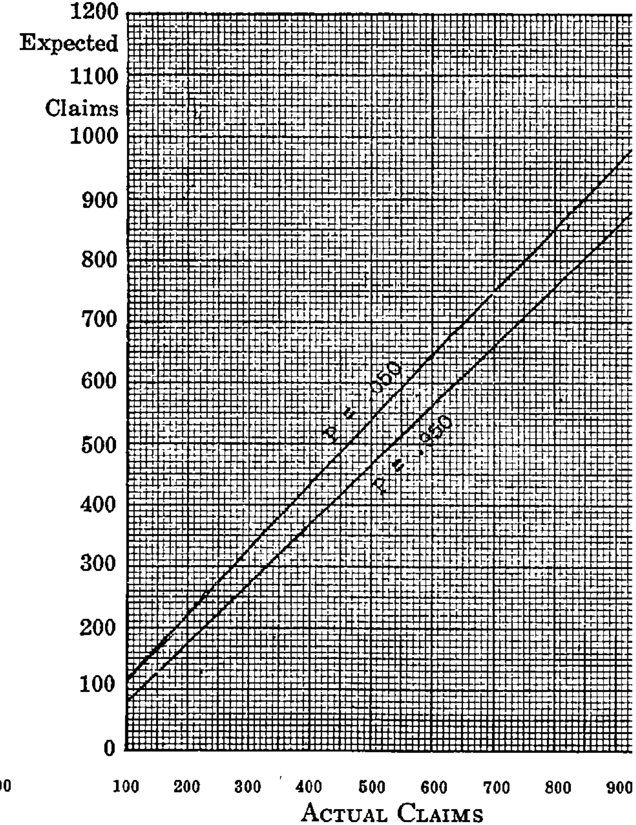
0-10 ACTUAL CLAIMS

Actual Claims	Expected Claims .050 Level	Expected Claims .950 Level
0	2.6	—
1	4.9	—
2	6.4	.3
3	7.8	.7
4	9.2	1.3
5	10.6	2.0
6	11.9	2.6
7	13.2	3.2
8	14.5	4.0
9	15.8	4.7
10	17.0	5.4

10-100 ACTUAL CLAIMS



100-1000 ACTUAL CLAIMS



level when the expected number of claims was 14.5. Thus we can give this risk a modification of $\frac{14.5}{73} = .20$, or an 80% credit. Obviously, a discount of more than 80% would make such a risk just an ordinary risk instead of a good one.

The B Brewery

Our next risk is the B Brewing Company, with a premium at present rates for the past three years of \$16,996, a permissible loss ratio of .517, and an expected average claim cost of \$27, from which we calculate the expected total loss as $\$16{,}996 \times .517 = \$8{,}787$ and the expected number of claims as $\frac{\$8{,}787}{\$27} = 325$. The experience of this risk for this three-year period included 544 claims totaling \$13,389 and averaging \$24.61 per claim. Testing the average claim cost in Table 11, we find that a ratio of $\frac{\$24.61}{\$27.00} = .91$ is a normal occurrence when 544 claims actually occurred. For the claim frequency, however, we find from Table 5 that the ratio of $\frac{544}{325} = 1.67$ is definitely above the normal range. For this risk the Experience Rating Plan would produce a 30% debit, and with such a debit the expected number of claims would be $325 \times 1.30 = 423$. Re-entering Table 5 with 423 expected claims and a ratio of $\frac{544}{423} = 1.29$, we still find the risk to be definitely above the normal claim frequency range.

Before considering how much greater debit than 30% our company would require before feeling safe to write this risk, let us assume that we knew nothing of the number of claims actually occurring and only knew the total losses of the risk during the past three years. To test these total losses, we enter Table 10 with 325 expected claims and a ratio of actual to expected total losses of $\frac{\$13{,}389}{\$8{,}787} = 1.52$, reaching exactly the same conclusion as before, that the risk's loss level is considerably above the range to be normally expected. Considering the 30% debit of the Experience Rating Plan, however, we would enter Table 10 with 423 expected claims and a ratio of $\frac{\$13{,}389}{\$8{,}787 \times 1.30} = 1.17$, to find the risk to be just about on the $P = .950$ level, indicating that the risk is probably, but not definitely, bad. This exercise illustrates only that if we want to obtain all of the information from the available data, we must use all of it and not take the easiest way, thereby getting only part of the answer from part of the available data.

The $P = .950$ line on Figure 2 will assist us in determining the minimum debit modification which a carrier with the $P = .950$ level of deficiency for bad risks would require in order to write the risk. For the B Brewing Company we find, for a deficiency level of $P = .950$, that the 544 actual claims correspond to expected claims of 505. Thus the modification would be $\frac{505}{325} = 1.55$, and we would require a 55% debit or more applicable to the present rates in order to make worth while the chances involved in insuring this risk. Even under these conditions it would be indicated that the B Brewing Company would be a fertile field for some effective safety engineering service.

The C Bus Line

The C Bus Line has in the past been self-insured and is now making application for full coverage insurance. It has provided our carrier under affidavit with lists of equipment used during the past three years and summaries of the losses which it has incurred under its self-insurance. Applying the present manual rates to this risk, we find that it would have developed \$14,832 of premium during the past three years in a classification having a permissible loss ratio of .607 and an average claim cost of \$43. Thus total expected losses of \$9,003 and 209 expected claims would be indicated. Their statement of loss experience shows that 179 claims, totaling \$4,395 and averaging \$24.55 per claim, were incurred by them. Testing the claim frequency from Table 5, we enter it with 209 expected claims and a ratio of $\frac{179}{209} = .86$, to find the risk at just about the $P = .025$ level. The Experience Rating Plan applicable to this risk produces a 26% credit, so that we return to Table 5 with expected claims of $209 \times .74 = 155$ and a ratio of $\frac{179}{155} = 1.15$, finding that the application of the 26% credit has shifted the risk from the $P = .025$ level to the $P = .975$ level and that, from a loss frequency point of view, the risk may no longer be desirable.

Entering Table 11 to test the average claim cost with 179 actual claims and a ratio of $\frac{\$24.55}{\$43.00} = .57$, we find that the average claim cost is far below the level to be normally expected. For this risk we have the average claim cost in one direction and the frequency in the other direction, and it thus behooves us to review the total losses. We enter Table 10 with the expected number of claims of 155; and with the ratio of actual to expected losses after experience rating of $\frac{\$4,395}{\$9,003 \times .74} = .66$, to find that the risk is apparently a desirable one, being just below the $P = .005$ level of significance. Reviewing our findings, we conclude that the desirability of this risk hinges entirely on its low average claim cost; and that because this low average claim cost may have been the result of a fictitious deflation accomplished by the elimination from the report of experience of a few large losses, and because there is a certain doubt in our minds that this low average claim cost can be continued with the settlement of claims transferred to our company as a third party, a review of the risk's claim folders in considerable detail is advisable and a continuous check on the average claim cost of our own losses for the risk should be maintained.

The D Distributing Company

The D Distributing Company is a combined local and long haul truckman written on a gross receipts basis. On the basis of its currently developed gross receipts rate, including a 57% experience rating credit, the premium for the past three years would be \$34,587 in a classification having a permissible loss ratio of .617 and an average claim cost of \$41, indicating expected total losses of $\$34{,}587 \times .617 = \$21{,}340$ and an expected number of claims of $\frac{\$21{,}340}{\$41.00} = 520$. The experience of these years shows 441 claims incurred, totaling \$15,791 and averaging \$35.81. Reference to Table 11 for 441 actual claims and a ratio of $\frac{\$35.81}{\$41.00} = .87$ shows the average claim cost to be on about the $P = .025$ level. Reference to Table 5 for 520 expected claims and a ratio of $\frac{441}{520} = .85$ shows the claim frequency below the $P = .005$ level. Apparently this risk could be afforded a greater credit than the 57% provided by the Experience Rating Plan.

A chart similar to Figure 2 but prepared for New York commercial automobile—property damage coverage—total losses from Table 10, would show us that, for actual losses of \$15,791 equivalent to $\frac{\$15{,}791}{\$41} = 385$ claims of the expected average amount, the expected number of such normal sized claims would be 450 for a level of significance of $P = .050$. Thus a further modification of $\frac{450 \times \$41.00}{\$21{,}340} = .86$, or a total modification of $.86 \times .43 = .37$ (a credit from manual rates of 63%), could be safely afforded the risk.

IV.

Modifications of the Basic Formulae for the Distribution of Casualty Insurance Statistics to Recognize Diversity of Risks, Hazard Levels of Classifications and Ratemaking Errors

A. Diversity of Risks in the Same Classification

The formulae of Part I were developed for application to individual risks and involved statistics of the expected number of claims and the distribution of losses by size of loss for the individual risk. It is necessary to modify these formulae for application to all risks of a classification when only the average statistics for the classification as a whole are known. It will be recognized that, although the individual risks grouped into a classification may be similar, only in very rare instances are they identical in all respects. The differences between the individual risk and the average of the classification, on a percentage basis, will be spoken of as the risk diversities.

Because of the necessity of weighting the statistics of individual risks to obtain averages for the classification, we shall find it necessary to introduce the symbols e for exposure and f for claim frequency. Thus $C = ef$. Statistics of the classification as a whole will be denoted by a prime ($'$). The diversities of claim frequencies, average claim costs, and pure premiums will be denoted by p, q, and m respectively, being defined by:

$$f = f'(1+p), \quad V_{1:x} = V_{1:x'}(1+q) \quad \text{and} \quad (1+m) = (1+p)(1+q)$$

The moments of p, q, and m will be defined by:

$$V_{n:p} = \frac{\Sigma\, e \cdot p^n}{\Sigma\, e}, \quad V_{n:q} = \frac{\Sigma\, e \cdot f \cdot q^n}{\Sigma\, e \cdot f}, \quad \text{and} \quad V_{n:m} = \frac{\Sigma\, e \cdot m^n}{\Sigma\, e}$$

where it will be noted from the following identities that:

$$V_{1:p} = V_{1:q} = V_{1:m} = 0.$$

$$f' = \frac{\Sigma\, e \cdot f}{\Sigma\, e} = \frac{\Sigma\, e \cdot f' \cdot (1+p)}{\Sigma\, e} = f'\,\frac{\Sigma\, e\,(1+p)}{\Sigma\, e} = f'\left(1 + \frac{\Sigma\, e \cdot p}{\Sigma\, e}\right)$$

$$V_{1:x'} = \frac{\Sigma\, e \cdot f \cdot V_{1:x}}{\Sigma\, e \cdot f} = \frac{\Sigma\, e \cdot f \cdot V_{1:x'}\,(1+q)}{\Sigma\, e \cdot f} = V_{1:x'}\left(1 + \frac{\Sigma\, e \cdot f \cdot q}{\Sigma\, e \cdot f}\right)$$

$$f' \cdot V_{1:x'} = \frac{\Sigma\, e \cdot f \cdot V_{1:x}}{\Sigma\, e} = \frac{\Sigma\, e \cdot f'\, V_{1:x'}\,(1+m)}{\Sigma\, e} = f'\, V_{1:x'}\left(1 + \frac{\Sigma\, e \cdot m}{\Sigma\, e}\right)$$

1. *Number of Claims.*

We shall concern ourselves with the development of formulae applicable to risks for which C claims are expected on the basis of the classification claim frequency. Such risks will have an exposure of C/f' and, as they have a claim frequency of $f'(1+p)$, the true expected number of claims for such a risk will be $C(1+p)$. From the formulae of section B of Part I we can write the first three moments, about the origin, of the actual number of claims occurring for such a risk as:

$$V_{1:n} = C(1+p)$$
$$V_{2:n} = C(1+p) + C^2(1+p)^2$$
$$V_{3:n} = C(1+p) + 3C^2(1+p)^2 + C^3(1+p)^3$$

These moments for individual risks, when weighted by the exposures of the risks and averaged for all risks, will give us the moments of the actual number of claims occurring for all risks in the classification when C claims are expected on the basis of the classification claim frequency. These are:

$$V_{1:n'} = C$$
$$V_{2:n'} = C + C^2(1+V_{2:p})$$
$$V_{3:n'} = C + 3C^2(1+V_{2:p}) + C^3(1+3V_{2:p} + V_{3:p}), \text{ and}$$
$$U_{2:n'} = C + C^2 V_{2:p}$$
$$U_{3:n'} = C + 3C^2 V_{2:p} + C^3 V_{3:p}$$

The moments of r', the ratio of actual to expected number of claims, can then be obtained by dividing by the powers of C as:

$$V_{1:r'} = 1,\ U_{2:r'} = 1/C + V_{2:p}, \text{ and } U_{3:r'} = 1/C^2 + 3V_{2:p}/C + V_{3:p}$$

2. *Total Cost of a Fixed Number of Claims.*

We must now deal with a group of risks whose average claim costs are admittedly different but for which we have available only the distribution of losses by size of loss for all risks combined. No progress can be made without some assumption as to the form of the distributions for individual risks. A minimum assumption is that the distributions for all risks in the class have the same coefficient of variation and, for the purpose of developing third moments, also have the same skewness. Indicating the coefficient of variation by (CV) we have:

$$(CV)^2 = \frac{U_{2:x}}{V^2_{1:x}} = \frac{V_{2:x} - V^2_{1:x}}{V^2_{1:x}} = \frac{V_{2:x}}{V^2_{1:x}} - 1$$

$$V_{2:x} = V^2_{1:x}[(CV)^2+1] = V^2_{1:x'}[(CV)^2+1]\,(1+q)^2$$

$$\text{but } V_{2:x'} = \frac{\Sigma\, e\cdot f\cdot V_{2:x}}{\Sigma\, e\cdot f} = \frac{\Sigma\, V^2_{1:x'}\,[(CV)^2+1]\, e\cdot f\,(1+q)^2}{\Sigma\, e\cdot f}$$

$$= V^2_{1:x'}\,[(CV)^2+1]\,(1+V_{2:q})$$

so that $V^2_{1:x'}[(CV)^2+1] = \dfrac{V_{2:x'}}{1+V_{2:q}}$ and by substituting this we have:

$$V_{2:x} = \frac{V_{2:x'}\,(1+q)^2}{1+V_{2:q}} \text{ and } V_{2:x}\cdot V_{1:x} = \frac{V_{2:x'}\cdot V_{1:x'}\,(1+q)^3}{1+V_{2:q}}$$

A similar procedure involving the skewness, $\alpha_{3:x}$, gives us that:

$$V_{3:x} = \frac{V_{3:x'}\,(1+q)^3}{1+3\,V_{2:q}+V_{3:q}}$$

Using these values of $V_{2:x}$, $V_{2:x}\cdot V_{1:x}$, and $V_{3:x}$ in the formulae of section C of Part I we have the first three moments of the actual costs of n claims, for an individual risk, as:

$$V_{1:t} = n\cdot V_{1:x'}\,(1+q)$$

$$V_{2:t} = \left[\frac{n\cdot V_{2:x'}}{1+V_{2:q}} + n\,(n-1)\,V^2_{1:x'}\right](1+q)^2$$

$$V_{3:t} = \left[\frac{n\cdot V_{3:x'}}{1+3\,V_{2:q}+V_{3:q}} + \frac{3\,n(n-1)\,V_{2:x'}\,V_{1:x'}}{1+V_{2:q}} + n(n-1)(n-2)\,V^3_{1:x'}\right](1+q)^3$$

These moments for individual risks, when weighted by the expected number of claims of the risks, $e\cdot f$, to measure the relative number of times such a number of n claims will occur, will give us the moments of the actual amount of losses from n claims of all risks in the classification as:

$$V_{1:t'} = n\cdot V_{1:x'}$$

$$V_{2:t'} = n\cdot V_{2:x'} + n(n-1)\,V^2_{1:x'}\,(1+V_{2:q})$$

$$V_{3:t'} = n\cdot V_{3:x'} + 3\,n(n-1)\,V_{2:x'}\cdot V_{1:x'}\left[\frac{1+3\cdot V_{2:q}+V_{3:q}}{1+V_{2:q}}\right] + n(n-1)(n-2)\,V^3_{1:x'}(1+3\cdot V_{2:q}+V_{3:q})$$

and $U_{2:t'} = n\cdot U_{2:x'} + n(n-1)\,V_{2:q}\cdot V^2_{1:x'}$

$$U_{3:t'} = n\cdot U_{3:x'} + 3\,n(n-1)\,V_{2:x'}\cdot V_{1:x'}\left[\frac{2\cdot V_{2:q}+V_{3:q}}{1+V_{2:q}}\right] + n(n-1)(n-2)\,V^3_{1:x'}\left[V_{3:q} - \frac{6\cdot V_{2:q}}{n-2}\right]$$

The moments of a', the average claim cost of a fixed number of claims, can then be obtained by dividing by the powers of n and the moments of s', the ratio of the actual to expected total losses or average claim costs for a fixed number of claims, can be obtained by dividing by the powers of $n\cdot V_{1:x'}$.

3. *Total Losses When C Claims are Expected.*

Concerning ourselves again with risks for which C claims are expected on the basis of the classification claim frequency, we can write, from the formulae of section D of Part I, the moments of the total actual losses, T, for an individual risk as:

$$V_{1:T} = C \cdot V_{1:x'} (1 + m)$$
$$V_{2:T} = C \cdot V_{2:x} (1 + p) + C^2 V^2_{1:x'} (1 + m)^2$$
$$V_{3:T} = C \cdot V_{3:x} (1 + p) + 3 C^2 V_{2:x} \cdot V_{1:x} (1 + p)^2 + C^3 V^3_{1:x'} (1 + m)^3$$

The averaging of these moments, when weighted by the risk exposures, is a straightforward process for the right hand terms of the above values for $V_{1:T}$, $V_{2:T}$, and $V_{3:T}$ involving only powers of m and the constants C and $V_{1:x'}$. The averaging of the left hand terms of the values for $V_{2:T}$ and $V_{3:T}$ involve only the substitutions:

$$\frac{\Sigma\, e\, (1 + p)\, V_{2:x}}{\Sigma\, e} = \frac{\Sigma\, e \cdot f'\, (1 + p)\, V_{2:x}}{\Sigma\, e \cdot f} = \frac{\Sigma\, e \cdot f \cdot V_{2:x}}{\Sigma\, e \cdot f} = V_{2:x'}$$

and
$$\frac{\Sigma\, e (1 + p)\, V_{3:x}}{\Sigma\, e} = \frac{\Sigma\, e \cdot f'\, (1 + p)\, V_{3:x}}{\Sigma\, e \cdot f} = \frac{\Sigma\, e \cdot f \cdot V_{3:x}}{\Sigma\, e \cdot f} = V_{3:x'}$$

No exact symbolic evaluation of the middle term of the value for $V_{3:T}$ can be made except one involving the correlations between powers of p and q. However, under the assumptions of the previous section regarding the individual risk distributions of losses by size of loss, such correlations can be treated as a group and it can be shown that:

$$\frac{\Sigma\, e\, (1+p)^2\, V_{2:x} \cdot V_{1:x}}{\Sigma\, e} = \frac{V_{2:x'} \cdot V_{1:x'}\, (1+V_{2:p})\, (1+3\, V_{2:q}+V_{3:q})}{1 + V_{2:q}} \pm \frac{G}{3\, C^2}$$

where G jointly represents all such correlations and is zero when all are zero. The algebraic derivation of this identity becomes very involved and will not be shown. Its accuracy can readily be verified by the erection of a numerical problem that fulfills the conditions of no correlation between any power of p and q.

We thus have for the moments of the actual total losses, for all risks in a classification for which C claims are expected on the basis of the classification claim frequency:

$$V_{1:T'} = C \cdot V_{1:x'}$$
$$V_{2:T'} = C \cdot V_{2:x'} + C^2\, V^2_{1:x'}\, (1 + V_{2:m})$$
$$V_{3:T'} = C \cdot V_{3:x'} + 3\, C^2\, V_{2:x'} \cdot V_{1:x'} \left[\frac{(1+V_{2:p})\, (1+3\, V_{2:q}+V_{3:q})}{1 + V_{2:q}}\right] + C^3\, V^3_{1:x'}\, (1 + 3\, V_{2:m} + V_{3:m}) \pm G$$

and
$$U_{2:T'} = C \cdot V_{2:x'} + C^2\, V^2_{1:x'} \cdot V_{2:m}$$
$$U_{3:T'} = C \cdot V_{3:x'} + 3\, C^2\, V_{2:x'} \cdot V_{1:x'} \left[\frac{(1+V_{2:p})\, (1+3\, V_{2:q}+V_{3:q})}{1 + V_{2:q}} - 1\right] + C^3\, V^3_{1:x'} \cdot V_{3:m} \pm G$$

The moments of R', the ratio of actual to expected losses, can be obtained from these by dividing by the powers of the expected total loss, $C \cdot V_{1:x'}$.

B. Hazard Levels of Classifications to Be Combined

Classifications and territories are made in order to discern and recognize any demonstrable differences in loss costs per unit of exposure. Such of these differences as are due to differences in claim frequencies have no effect on the sampling distributions with which we are concerned. Only such differences as are due to differences in the distributions of losses by size of loss need be considered.

It frequently occurs that the only available distribution of losses by size of loss includes risks from a wide group of classes. Although it might be reasonable to assume that the coefficients of variation, and even the skewnesses, of the distributions for individual classifications are constant, it would not be reasonable to assume that the average claim costs for all classes are likewise constant. Moreover classification average claim costs are usually available and actually can be taken into consideration. Thus for the purpose of sampling theory the "hazard" of a class or territory will be represented by its average claim cost.

Statistics for a group of classes will be denoted by a double prime ($''$). The symbol B will be used to represent the relative hazard of a classification and will be defined as:

$$V_{1:x'} = V_{1:x''}(1+B) \quad \text{with } V_{n:B} = \frac{\Sigma\, e' f' B^n}{\Sigma\, e' f'}$$

It will be noted that $V_{1:B} = 0$ as:

$$V_{1:x''} = \frac{\Sigma\, e' f' V_{1:x'}}{\Sigma\, e' f'} = V_{1:x''} \frac{\Sigma\, e' f' (1+B)}{\Sigma\, e' f'} = V_{1:x''}(1 + V_{1:B})$$

The other difference between classes will be in the diversity of risks within the classes. It will be unreasonable in most cases to assume that the moments of p, q, and m are constant for all of a group of classes. The moment formulae for risks in a group of classes will, therefore involve moment functions of p, q, and m defined as:

$$V_{n:p'} = \frac{\Sigma\, e' V_{n:p}}{\Sigma\, e'}, \quad V_{n:q'} = \frac{\Sigma\, e' f' V_{n:q}}{\Sigma\, e' f'} \quad \text{and } V_{n:m'} = \frac{\Sigma\, e' V_{n:m}}{\Sigma\, e'}$$

It will be important to note that all powers of B are independent of any moment of p, q, or m as this independence is utilized in evaluating the summations and averages in the following paragraphs.

1. *Number of Claims*

The moments of the actual number of claims occurring when C are expected on the basis of the classification claim frequency of section A-1 of this Part may be averaged for a group of classes, using weights equal to the class exposure, e', to obtain:

$$V_{1:n''} = C$$
$$V_{2:n''} = C + C^2 (1 + V_{2:p'})$$
$$V_{3:n''} = C + 3\,C^2 (1 + V_{2:p'}) + C^3 (1 + 3\,V_{2:p'} + V_{3:p'})$$

2. *Total Cost of an Actual Number of Claims Whose Expected Total is E*

If E losses are expected for a risk, in a class having an average claim cost of $V_{1:x'}$, because of the actual occurrence of n claims, then n must be $E/V_{1:x'}$. The moments of the actual total losses, of all risks in a classification for which $E/V_{1:x'}$ claims have occurred, may be written from the formulae of section A-2 of this Part as:

$$V_{1:t'} = E$$
$$V_{2:t'} = E\frac{V_{2:x'}}{V_{1:x'}} + E\,(E - V_{1:x'})(1 + V_{2:q})$$
$$V_{3:t'} = E\frac{V_{3:x'}}{V_{1:x'}} + 3\,E\,(E - V_{1:x'})\cdot\frac{V_{2:x'}}{V_{1:x'}}\cdot\frac{1 + 3\,V_{2:q} + V_{3:q}}{1 + V_{2:q}}$$
$$+ E\,(E - V_{1:x'})(E - 2\,V_{1:x'})(1 + 3\,V_{2:q} + V_{3:q})$$

These moments must then be weighted by $e'\,f'\,V_{1:x'}$ or by $e'\,f'\,(1 + B)$ to obtain the corresponding moments for risks of all classifications. First it will be necessary to make the substitutions:

$$\frac{V_{2:x'}}{V_{1:x'}} = \frac{V_{2:x''}}{V_{1:x''}}\cdot\frac{1 + B}{1 + V_{2:B}} \text{ and } \frac{V_{3:x'}}{V_{1:x'}} = \frac{V_{3:x''}}{V_{1:x''}}\cdot\frac{(1 + B)^2}{(1 + 3\,V_{2:B} + V_{3:B})}$$

This produces:

$$V_{1:t''} = E$$
$$V_{2:t''} = E\frac{V_{2:x''}}{V_{1:x''}} + E^2 (1 + V_{2:q'}) - E\cdot V_{1:x''}\,(1 + V_{2:q'})(1 + V_{2:B})$$
$$V_{3:t''} = E\,\frac{V_{3:x''}}{V_{1:x''}} + 3\,\frac{V_{2:x''}}{V_{1:x''}}\cdot\frac{1 + 3\,V_{2:q'} + V_{3:q'}}{1 + V_{2:q'}}\left[E^2 - E\,V_{1:x''}\,\frac{1 + 3\,V_{2:B} + V_{3:B}}{1 + V_{2:B}}\right]$$
$$+ (1 + 3\,V_{2:q'} + V_{3:q'})\,[E^3 - 3\,E^2\,V_{1:x''}\,(1 + V_{2:B})$$
$$+ 2\,E\,V^2_{1:x''}\,(1 + 3\,V_{2:B} + V_{3:B})]$$

3. *Total Losses When Losses of E Are Expected*

If E losses are expected for a risk in a class having an average claim cost of $V_{1:x'}$, then C, the expected number of claims on the basis of the classification claim frequency, will be $E/V_{1:x'}$. The moments of the actual total losses, of all risks in a classification for which $E/V_{1:x'}$ claims are expected, may be written from the formulae of section A-3 of this Part as:

$$V_{1:T'} = E$$

$$V_{2:T'} = E \frac{V_{2:x'}}{V_{1:x'}} + E^2 (1 + V_{2:m})$$

$$V_{3:T'} = E \frac{V_{3:x'}}{V_{1:x'}} + 3 E^2 \frac{V_{2:x'}}{V_{1:x'}} \cdot \frac{(1 + V_{2:p})(1 + 3 V_{2:q} + V_{3:q})}{1 + V_{2:q}}$$
$$+ E^3 (1 + 3 V_{2:m} + V_{3:m}) \pm G$$

These moments must then be weighted by $e' f' V_{1:x'}$ or by $e' f' (1 + B)$ and the substitutions of the previous paragraph made to obtain the corresponding moments for all classifications as:

$$V_{1:T''} = E$$

$$V_{2:T''} = E \frac{V_{2:x''}}{V_{1:x''}} + E^2 (1 + V_{2:m'})$$

$$V_{3:T''} = E \frac{V_{3:x''}}{V_{1:x''}} + 3 E^2 \frac{V_{2:x''}}{V_{1:x''}} \cdot \frac{(1 + V_{2:p'})(1 + 3 V_{2:q'} + V_{3:q'})}{1 + V_{2:q'}}$$
$$+ E^3 (1 + 3 V_{2:m'} + V_{3:m'}) \pm G'$$

C. *Ratemaking Errors, or Errors in the Available Estimates of Classification Averages*

Up to this point the formulae have been developed on the premise that actual averages of classification data were available. In hindsight analysis this is usually the case; while in prospective application it is not. The next step, then, will be to recognize the possibility of errors in the available estimates of classification averages. To do this we shall represent the estimated classification claim frequency by F where $f' = F(1 + P)$, the estimated average claim cost by A where $V_{1:x'} = A(1 + Q)$, and the estimated average pure premium by FA where $f' V_{1:x'} = FA(1 + M)$. Thus P, Q, and M represent percentage errors in classification estimates corresponding to the percentage diversities, p, q, and m, of individual risks.

The moments of P, Q, and M will be defined by:

$$V_{n:P} = \frac{\Sigma e' P^n}{\Sigma e'}, \quad V_{n:Q} = \frac{\Sigma e' F Q^n}{\Sigma e' F}, \text{ and } V_{n:M} = \frac{\Sigma e' M^n}{\Sigma e'}$$

so that $V_{1:P}$, $V_{1:Q}$, and $V_{1:M}$ represent the component parts of or the entire error in rate level for the group of classes as a whole.

It will be important to note here that P, Q, and M are independent of the moments of p, q, and m; as this independence will be utilized in evaluating the averages of the following paragraphs.

1. *Number of Claims.*

The value of C in section A-1 of this Part must be replaced by $C(1 + P)$ before the averaging process of section B-1 is performed, to obtain:

$$V_{1:n''} = C\,(1 + V_{1:P})$$
$$V_{2:n''} = C\,(1 + V_{1:P}) \;+ C^2\,(1+V_{2:p'})\,(1+2\,V_{1:P}+V_{2:P})$$
$$V_{3:n''} = C\,(1 + V_{1:P}) \;+ 3\,C^2\,(1+V_{2:p'})\,(1+2\,V_{1:P}+V_{2:P}) + C^3\,(1+3\,V_{2:p'}+V_{3:p'})\,(1+3\,V_{1:P}+3\,V_{2:P}+V_{3:P}).$$

2. *Total Cost of an Actual Number of Claims Whose Expected Total is E.*

The value of E in the first paragraph of section B-2 of this Part must be replaced by $E\,(1 + Q)$ with the result that the following replacements must be made in the moments of the second paragraph:

$$E\,(1 + V_{1:Q}) \text{ for } E, \quad E^2\,(1 + 2\,V_{1:Q} + V_{2:Q}) \text{ for } E^2, \text{ and}$$
$$E^3\,(1 + 3\,V_{1:Q} + 3\,V_{2:Q} + V_{3:Q}) \text{ for } E^3.$$

3. *Total Losses When Losses of E Are Expected.*

The value of E in the first paragraph of section B-3 of this Part must be replaced by $E\,(1 + M)$ with the result that the following replacements must be made in the moments of the second paragraph:

$$E\,(1 + V_{1:M}) \text{ for } E, \quad E^2\,(1 + 2\,V_{1:M} + V_{2:M}) \text{ for } E^2,$$
$$E^3\,(1 + 3\,V_{1:M} + 3\,V_{2:M} + V_{3:M}) \text{ for } E^3.$$

V.

Calculating Methods to Obtain Estimates of $U_{2:R''}$

In Part II all of the calculation of the sampling moments based on the formulae developed in Part I involved only the moments of the distribution of claims by size of claim. The formulae developed in Part IV involve these moments together with the moments of p, q, m, P, Q, and M. For these variables we can not determine the moments by any direct means but must always obtain them from data in which these variables are in combination with variations due to chance. Most frequently these moments can be calculated from the moments of R'' and while at times we will use $U_{3:R''}$ we will usually need only $U_{2:R''}$ and the following discussion will be limited to this second moment in order to reduce its length. The methods of approach can be extended to the third moment by the reader as necessary.

Theoretically we can only estimate the value of $U_{2:R''}$ if we have a sufficiently large number of risks or classes with exactly the same expected losses. This is a condition so rarely met in practice that we must investigate the possibilities of obtaining estimates from groups of risks or classes that differ as to their expected losses although being contained within a limited range.

A. The True Value of $U_{2:R''}$

What we want to calculate from the data of individual risks or classes is the value of $U_{2:R''}$ corresponding to the average value of E for a group of risks grouped by size of E. From section C-3 of Part IV we have for

$$E = V_{1:E}$$

$$V_{1:T''} = V_{1:E}\,(1 + V_{1:M})$$

$$V_{2:T''} = V_{1:E}\,(1 + V_{1:M})\,\frac{V_{2:x''}}{V_{1:x''}} + V^2_{1:E}\,(1 + V_{2:m'})\,(1 + 2\cdot V_{1:M} + V_{2:M})$$

$$U_{2:T''} = V_{1:E}\,(1 + V_{1:M})\frac{V_{2:x''}}{V_{1:x''}} + V^2_{1:E}\,[U_{2:M}\,(1 + V_{2:m'}) + V_{2:m'}\,(1 + V_{1:M})^2]$$

from which

$$V_{1:R''} = 1 + V_{1:M} \text{ and}$$

$$U_{2:R''} = \frac{1 + V_{1:M}}{V_{1:E}}\cdot\frac{V_{2:x''}}{V_{1:x''}} + U_{2:M}\,(1 + V_{2:m'}) + V_{2:m'}\,(1 + V_{1:M})^2$$

There will also be times when we shall want these moments of R'' after they have been corrected for the error in the rate level of $(1 + V_{1:M})$. The corrected moments would then be:

Corrected $V_{1:R''} = 1$, and

$$\text{Corrected } U_{2:R''} = \frac{1}{V_{1:E}\,(1 + V_{1:M})}\cdot\frac{V_{2:x''}}{V_{1:x''}} + V_{2:m'} + (1 + V_{2:m'})\cdot\frac{U_{2:M}}{(1 + V_{1:M})^2}$$

B. Calculations from Individual Values of R''

The most obvious procedure would be to calculate the value of R'' for each risk in the group and to proceed from these to calculate $V_{1:R''}$ and $U_{2:R''}$. Let us then see what the results of this calculation will produce. For a particular value of E we would have:

$$V_{1:R''} = 1 + V_{1:M}$$

$$V_{2:R''} = \frac{1 + V_{1:M}}{E}\cdot\frac{V_{2:x''}}{V_{1:x''}} + (1 + V_{2:m'})\,(1 + 2\cdot V_{1:M} + V_{2:M})$$

and as averages for all values of E we would have:

$$V_{1:R''} = 1 + V_{1:M}$$

$$V_{2:R''} = (1+V_{1:M})\;\frac{V_{2:x''}}{V_{1:x''}}\left(\text{Average }\frac{1}{E}\right) + (1 + V_{2:m'})\,(1+2\cdot V_{1:M}+V_{2:M})$$

$$U_{2:R''} = (1+V_{1:M})\frac{V_{2:x''}}{V_{1:x''}}\left(\text{Average }\frac{1}{E}\right) + U_{2:M}\,(1+V_{2:m'}) + V_{2:m'}\,(1+V_{1:M})^2$$

or, if each value of E had been multiplied by $(1+V_{1:M})$ before the individual values of R'' were calculated, we would have:

$$V_{1:\text{Adjusted } R''} = 1 \text{ and}$$

$$U_{2:\text{Adjusted } R''} = \frac{1}{(1+V_{1:M})} \cdot \frac{V_{2:x''}}{V_{1:x''}} \cdot \left(\text{Avg.} \frac{1}{E}\right) + V_{2:m'} + (1+V_{2:m'}) \cdot \frac{U_{2:M}}{(1+V_{1:M})^2}$$

For a range of E in which the greatest value of E is r times the least value of E we have approximately that:

$$\frac{1}{V_{1:E}} = \left(\text{Average } \frac{1}{E}\right) \cdot \frac{.8686\,(r-1)}{(r+1)\log_{10} r}$$

For example if $r=2$ then $\frac{1}{V_{1:E}} = .962\left(\text{Average } \frac{1}{E}\right)$

Thus this method of calculation of $U_{2:R''}$ overestimates its value by an error of approximately:

$$\text{Calculated } U_{2:R''} - \text{True } U_{2:R''} = (1+V_{1:M}) \frac{V_{2:x''}}{V_{1:x''}} \frac{1}{V_{1:E}} \left[\frac{(r+1)\log_{10} r}{.8686\,(r-1)} 1 - 1\right]$$

where the $(1+V_{1:M})$ term becomes $\frac{1}{1+V_{1:M}}$ if the values of R'' are corrected for the error in rate level.

C. Calculations Based on the Z-Function

A second method of calculation would be based on the z-function described in section F of Part I. The value of Z would be calculated for each risk or class, the moments of these observed Z's calculated and the value of $U_{2:R''}$ calculated from these. For a particular value of E we would have:

$$V_{1:Z''} = \sqrt{E}\,(V_{1:R''} - 1) = V_{1:M}\sqrt{E}$$

$$V_{2:Z''} = E(V_{2:R''} - 2 \cdot V_{1:R''} + 1)$$

$$= E\left[\frac{1+V_{1:M}}{E} \cdot \frac{V_{2:x''}}{V_{1:x''}} + (1+V_{2:m'})(1+2\,V_{1:M}+V_{2:M}) - 2\,(1+V_{1:M}) + 1\right]$$

$$\doteq (1+V_{1:M}) \frac{V_{2:x''}}{V_{1:x''}} + E\,[V_{2:M} + V_{2:m'}\,(1+2\,V_{1:M}+V_{2:M})]$$

and for all values of E we would have:

$$V_{1:Z''} = V_{1:M}\,(\text{Average } \sqrt{E})$$

$$V_{2:Z''} = (1+V_{1:M}) \frac{V_{2:x''}}{V_{1:x''}} + V_{1:E}\,[V_{2:M} + V_{2:m'}\,(1+2\,V_{1:M}+V_{2:M})]$$

$$U_{2:Z''} = (1+V_{1:M})\frac{V_{2:x''}}{V_{1:x''}} + V_{1:E}\,[U_{2:M}\,(1+V_{2:m'})+V_{2:m'}\,(1+V_{1:M})^2]$$
$$+ [V_{1:E}-(\text{Average } \sqrt{E})^2]\, V^2_{1:M}$$

$$U_{2:R''} = \frac{U_{2:Z''}}{V_{1:E}} = \frac{1+V_{1:M}}{V_{1:E}} \cdot \frac{V_{2:x''}}{V_{1:x''}} + U_{2:M}\,(1+V_{2:m'}) + V_{2:m'}\,(1+V_{1:M})^2$$
$$+ V^2_{1:M}\left[1 - \frac{(\text{Average } \sqrt{E})^2}{V_{1:E}}\right]$$

As we have approximately that:

$$\frac{(\text{Average } \sqrt{E})^2}{V_{1:E}} = \frac{8\,(r^{3/2}-1)^2}{9\,(r-1)^2\,(r+1)}$$

we have that $U_{2:R''}$ is overestimated by this method of calculation by:

$$\text{Calculated } U_{2:R''} - \text{True } U_{2:R''} = V^2_{1:M}\left[1 - \frac{8\,(r^{3/2}-1)^2}{9\,(r-1)^2\,(r+1)}\right]$$

Obviously the Z-method gives exactly the right answer if the expected losses have been balanced to the actual losses prior to the calculation of the individual values of Z, as the value of $V_{1:M}$ then becomes zero.

D. Calculations based on the z-Function.

The most readily obtained exact value of $U_{2:R''}$ (and the closest approximation to $U_{3:R''}$) is calculated from the average value of $z^2 = \frac{(T'')^2}{E}$. This, however, requires the separate calculation of z^2 for each risk.

For a particular value of E we would have:

$$V_{1:T''} = E\,(1+V_{1:M})$$
$$V_{2:z''} = (1+V_{1:M})\cdot\frac{V_{2:x''}}{V_{1:x''}} + E(1+V_{2:m'})(1+2\,V_{1:M}+V_{2:M})$$

and for all values of E:

$$V_{1:T''} = V_{1:E}\,(1+V_{1:M})$$
$$V_{2:z''} = (1+V_{1:M})\frac{V_{2:x''}}{V_{1:x''}} + V_{1:E}\,(1+V_{2:m'})\left[U_{2:M} + (1+V_{1:M})^2\right]$$

so that:

$$\frac{V_{2:z''}}{V_{1:E}} = \frac{(1+V_{1:M})}{V_{1:E}}\cdot\frac{V_{2:x''}}{V_{1:x''}} + U_{2:M}\,(1+V_{2:m'}) + V_{2:m'}\,(1+V_{1:M})^2 + (1+V_{1:M})^2$$

and the true value of $U_{2:R''}$ is seen to be:

$$U_{2:R''} = \frac{V_{2:z''}}{V_{1:E}} - \left(\frac{V_{1:T''}}{V_{1:E}}\right)^2$$

and the true value of $U_{2:R''}$ corrected for the error in rate level:

$$\text{Corrected } U_{2:R''} = \frac{V_{1:E}\,V_{2:z''}}{V^2_{1:T''}} - 1$$

Note: If $V_{3:z}$ is the average value of $\frac{(T'')^3}{E}$, a very close approximation to the true value of the corrected $U_{3:R''}$ is given by:

$$\text{Corrected } U_{3:R''} = \frac{V_{1:E}\,V_{3:z''}}{V^3_{1:T''}} - 3\frac{V_{1:E}\,V_{2:z''}}{V^2_{1:T''}} + 2$$

E. Calculations Based on the W-Function

The calculation of individual values of R'' for a large number of risks becomes quite laborious and the calculation of the individual values of Z'' becomes prohibitive. The data is usually available on punch cards in the form of actual losses and either premiums at manual rates or expected losses. The following method of computation assumes that expected losses are cut on the cards. The necessary adjustments of the formulae to use premiums in place of expected losses will be left to the reader. It will be assumed that the reader is familiar with the methods of obtaining sums of squares and of cross products of data cut on punch cards and has obtained the values:

$\Sigma\,(T''-E) = \Sigma\,T'' - \Sigma\,E$ and $\Sigma\,(T''-E)^2 = \Sigma\,(T'')^2 + \Sigma\,E^2 - 2\,\Sigma\,T''E$

from which values of $V_{1:(T''-E)}$ and $V_{2:(T''-E)})$ have been calculated.

For a particular value of E we would have:

$$\begin{aligned}
V_{1:T''} &= E\,(1+V_{1:M})\\
V_{1:(T''-E)} &= V_{1:T''} - E = E\cdot V_{1:M}\\
V_{2:(T''-E)} &= V_{2:T''} - 2\,E\cdot V_{1:T''} + E^2\\
&= E\,(1+V_{1:M})\,\frac{V_{2:x''}}{V_{1:x''}} + E^2\,[V_{2:M}+U_{2:M}\cdot V_{2:m'}+V_{2:m'}\,(1+V_{1:M})^2]
\end{aligned}$$

and for all values of E:

$$\begin{aligned}
V_{1:T''} &= V_{1:E}\,(1+V_{1:M})\\
V_{1:(T''-E)} &= V_{1:E}\cdot V_{1:M}\\
V_{2:(T''-E)} &= V_{1:E}\,(1+V_{1:M})\frac{V_{2:x''}}{V_{1:x''}} + V_{2:E}\,[V_{2:M}+U_{2:M}\cdot V_{2:m'}+\\
&\qquad V_{2:m'}\,(1+V_{1:M})^2]
\end{aligned}$$

Now if we calculate:

$$V_{1:W} = \frac{V_{1:(T''-E)}}{V_{1:E}} = V_{1:M} \text{ and } V_{2:W} = \frac{V_{2:(T''-E)}}{V^2_{1:E}} \text{ we have:}$$

$$V_{2:W} = \frac{1+V_{1:M}}{V_{1:E}}\cdot\frac{V_{2:x''}}{V_{1:x''}} + \frac{V_{2:E}}{V^2_{1:E}}\,[V_{2:M}+U_{2:M}\cdot V_{2:m'}+V_{2:m'}\,(1+V_{1:M})^2]$$

$$\begin{aligned}
U_{2:W} &= \frac{1+V_{1:M}}{V_{1:E}}\cdot\frac{V_{2:x''}}{V_{1:x''}} + U_{2:M}\,(1+V_{2:m'})+V_{2:m'}\,(1+V_{1:M})^2\\
&\quad+\left[\frac{V_{2:E}}{V^2_{1:E}}-1\right]\cdot[V_{2:M}+U_{2:M}\cdot V_{2:m'}+V_{2:m'}\,(1+V_{1:M})^2]
\end{aligned}$$

and make use of the following approximation:

$$\frac{V_{2:E}}{V^2_{1:E}} - 1 = \frac{(r-1)^2}{3\,(r+1)^2}$$

we have that $U_{2:W}$ is an overestimate of $U_{2:R''}$ having an error of:

True $U_{2:R''}$ — $U_{2:R''}$ calculated from $U_{2:W}$ =

$$-\frac{(r-1)^2}{3\,(r+1)^2}\left[V_{2:M} + U_{2:M}\cdot V_{2:m'} + V_{2:m'}\,(1+V_{1:M})^2\right]$$

Similarly, if we calculate:

$$V_{1:W'} = \frac{V_{1:(T''-E)}}{V_{1:T''}} \quad \text{and} \quad V_{2:W'} = \frac{V_{2:(T''-E)}}{V^2_{1:T''}}$$

then $U_{2:W'}$ is an overestimate of the Corrected $U_{2:R''}$ having an error of:

True Corrected $U_{2:R''}$ — $U_{2:R''}$ calculated from $U_{2:W'}$ =

$$-\frac{(r-1)^2}{3\,(r+1)^2}\left[V_{2:m'} + \frac{V_{2:M} + U_{2:M}\cdot V_{2:m'}}{(1+V_{1:M})^2}\right]$$

In most cases this estimate of $U_{2:R''}$ is closer to the true value than that obtained from the calculation of R'' for each risk. Although not as exact as the estimate obtained from the z-function, it is so much more easily obtained as to make its use mandatory in all but the most exact studies.

VI.

Excess Pure Premium Ratios

A. The Use of Excess Pure Premium Ratios

Tables of excess pure premium ratios are made available only to serve as the means of calculating the "insurance charge" to be included in the basic premium of a retrospective rating plan. All other parts of the final restrospective premium are on an actual cost-plus basis and as such do not represent insurance. Because of the complete reliance on tables of excess pure premium ratios for the determination of the entire insurance portion of retrospective premiums, it is necessary to analyse very carefully all of the conditions under which the tabular values of these excess pure premium ratios may be in error.

The "insurance charge" is made up of an expected amount of losses, or loss portion, loaded for expenses, such as claim adjustment expenses, that are assumed to vary directly with losses and, in some cases, for taxes or other expenses which are to vary with the final premium. The loss portion of the "insurance charge" is the net difference between the loss portion of the "charge for losses in excess of those contemplated by the maximum premium" and the loss portion of the "saving on minimum premium risks." It is customary to express all of these as ratios to the standard premium, P.

The loss portion of the charge for losses in excess of those contemplated by the maximum premium is equal to the product of the average loss ratio, A, and the excess pure premium ratio corresponding to the loss ratio necessary to reach the maximum premium, $B_{(max)}$. Symbolically this is expressed as:

Loss on Maximum Premium Risks $= A$ (x-ratio for $B_{(max)}$, P, A)

The loss portion of the savings on minimum premium risks is equal to the loss ratio contemplated by the minimum premium, minus the average loss ratio, plus the product of the average loss ratio and the excess pure premium ratio corresponding to the loss ratio necessary to reach the minimum premium, $B_{(min)}$, *i.e.*:

Saving on Min. Prem. Risks $= B_{(min)} - A + A$ (x-ratio for $B_{(min)}$, P, A)

In deriving the insurance charges for a retrospective plan, the average loss ratio, A, is assumed to be the permissible loss ratio, L. In actual application, however, the actual loss level varies considerably above and below the permissible creating considerable differences between the true insurance cost and that obtained from the use of the permissible loss ratio and the tabular values of excess pure premium ratios.

When the actual loss level, A, is equal to aL, the true value of the excess pure premium ratio is equal to the tabular excess pure premium ratio corresponding to a loss ratio of B/a, a premium size of aP, and the permissible loss ratio L. The loss portion of the insurance charge thus becomes:

$$aL - B_{(min)} - aL\left[(x\text{-ratio for } \frac{B_{(min)}}{a}, aP, L) - (x\text{-ratio for } \frac{B_{(max)}}{a}, aP, L)\right]$$

The effect of departures from the expected loss level can best be seen from a consideration of specific examples. Let us take a hypothetical plan in which the minimum and maximum premiums for a \$10,000 standard premium risk contemplate loss ratios of .400 and .800 respectively. The loss portion of the insurance charge included in such a plan, if based on a permissible loss ratio of .598, would be:

$$.598 - .400 - .598\ [(x\text{-ratio for } .400,\ \$10{,}000,\ .598) - (x\text{-ratio for } .800,\ \$10{,}000,\ .598)] = .198 - .598\ [.467 - .213] = .046$$

If manual rates were redundant to an extent that the average loss ratio was .498, the actual cost of insured losses would be:

$$.498 - .400 - .498\ [(x\text{-ratio for } .480,\ \$12{,}000,\ .598) - (x\text{-ratio for } .961,\ \$12{,}000,\ .598)] = .098 - .498\ [.382 - .144] = -.021$$

If, however, the manual rates were inadequate and the average loss ratio was .698, the actual cost would be:

$$.698 - .400 - .698\ [(x\text{-ratio for } .343,\ \$8{,}570,\ .598) - (x\text{-ratio for } .685,\ \$8{,}570,\ .598)] = .298 - .698\ [.532 - .281] = .123$$

This example illustrates three important points. First, the actual insured losses under a retrospective plan may even be negative when the rate level is redundant. Second, the actual insured losses may be several times greater than provided for, by the insurance charges in the plan, when the rate level is inadequate. Third, the actual insured losses under a retrospective plan always average to an amount greater than provided for in a retrospective plan over any period of years in which the actual loss levels varied although averaging out to the permissible level. A comparison of the loss portion of the insurance charge with the permissible loss ratio also indicates the extent to which the hazards of the insurance have been transferred to the risk by the carrier with, in most cases, but little change being made by the carrier in the charges for the expense of providing it.

In addition to the variations in the insurance costs due to departures of actual average loss ratios from the permissible, there are many other causes of variation in these costs due to departures of actual excess pure premium ratios from their tabular values. It will be necessary to study in detail the composition of an excess pure premium ratio in order to analyse such variation and the conditions under which it will occur.

The excess pure premium ratio corresponding to a loss ratio of B, a standard premium risk size of P, and a permissible loss ratio of L is defined as the ratio, to the total of all losses, of losses in excess of BP per risk. When the average loss ratio of all risks is equal to the permissible loss ratio the excess pure premium ratio has the form:

$$x\text{-ratio for } B,P,L = 1 - \frac{B}{L} + \int_0^{B/L}\int_0^{B/L} F_{(R)} \cdot dR \cdot dR \,.$$

where $F_{(R)}$ is the probability that a risk of standard premium size P will have actual losses of $P \times L \times R$ when the expected losses are $P \times R$.

It is obvious from the above form of the excess pure premium ratio that, for fixed values of B and L, any variation to occur must result from variation in the value of $F_{(R)}$. It has been shown that $F_{(R)}$ takes the form of a frequency distribution with a mean of unity and a variance of:*

$$U_{2:R} = \frac{1}{P \times L}\,\frac{V_{2:x}}{V_{1:x}} + U_{2:m} + U_{2:M} + U_{2:m}\,U_{2:M}$$

where:

$V_{1:x}$ and $V_{2:x}$ are the first and second moments, about the origin, of the distribution of amounts of individual claims by size of claim.

$U_{2:m}$ is the variance of the inherent hazards of risks assigned to a classification about the average hazard of the classification, (on a percentage basis).

* The (′) and (″) notation has been omitted throughout this Part and must be inferred as necessary.

$U_{2:M}$ is the variance of the errors in ratemaking and experience rating (on a percentage basis).

The importance of the variance of $F_{(R)}$, as a measure of the value of the double definite integral in the formula for the excess pure premium ratio, can be visualized by recognizing

$$\int_0^{B/L} F_{(R)} \cdot dR \text{ as the ogive of the frequency distribution of } F_{(R)}.$$

This ogive is a continually ascending curve with zero as its minimum and unity as its maximum. The greater the variance of $F_{(R)}$, the less steep will be the slope of this ogive. The double definite integral is the area under this ogive up to the abscissa B/L and will obviously be greater for an ogive with a more moderate slope (corresponding to a larger value of $U_{2:R}$) than for one with a steep slope (corresponding to a small value of $U_{2:R}$). Thus the following information as to variation in the excess pure premium ratio can be obtained directly from the above equation for $U_{2:R}$.

1. The value of the excess pure premium becomes less as the size of the risk increases; but will never reach the lowest possible value (zero, or $1 - B/L$ if $1 - B/L$ is greater than zero) except for completely self-rated risks whose hazards never change from one year to another.

2. The excess pure premium ratios will be higher during periods of rapidly changing conditions than during periods of comparatively stable conditions, due to the greater error in manual rates and experience modifications at such times.

3. The excess pure premium ratios will vary by state, as a result of the material effect of differences in law levels on the value of $\frac{V_{2:x}}{V_{1:x}}$ They will be higher for states (such as Pennsylvania) that have fewer and broader classifications than for states having many special classifications, because broadening of classifications increases $U_{2:m}$ by increasing the differences between risks in the same classification. They will be lower for states with a large volume of business, because the rates in such states will be more accurate and the value of $U_{2:M}$ will be smaller.

4. The excess pure premium ratios will vary considerably by classification, because of variation in the values of $\frac{V_{2:x}}{V_{1:x}}$ by classification arising from differences in the expected frequencies of large losses. For homogeneous classifications in which the inherent hazard of all risks is very nearly the same, they will be lower than for heterogeneous or N.O.C. classifications, because of lower values of $U_{2:m}$ in homogeneous classes. They will be lower for large classifications than for small classes, due to the greater accuracy of the manual rates producing lower values of $U_{2:M}$.

5. The excess pure premium ratios will vary by date of valuation of the experience; being lower for first reports and higher successively for second, third and later reports. This is the result of the use of averages as estimated values of unsettled cases in early reports having the effect of depressing the value of $\frac{V_{2:x}}{V_{1:x}}$ below its ultimate and correct value.

B. *Computation of Excess Pure Premium Ratios from Actual Data for Individual Risks*

The usual procedure for the calculation of excess pure premium ratios (hereafter called x-ratios) from actual risk data is shown on Exhibit A on the basis that the individual risk data consists of punch cards containing the expected and actual losses and the ratio of actual to expected losses. The data used as an example in Exhibits A, B and C is that for 173 New York State workmen's compensation risks having a standard premium between $4,000 and $6,687 and having a governing classification contained in a particular group of classifications indicated as Hazard Group 1 by the writer and characterized by low average loss cost per claim. Such risks have, as would be expected, x-ratios quite different from the average of similar risks but of all classifications.

If the risk data consists of premium, actual losses, and the loss ratio, exhibit A would have the following columns:

Col. (1) Desired Loss Ratio
Col. (1a) Actual Loss Ratio $=$ (2a) $\div F$
Col. (2) Desired Upper Sorting Limit of Loss Ratios $=$ (1) $\times F$
Col. (2a) Actual Upper Sorting Limit of Loss Ratios
Col. (3) Tabulated Data — Number of Risks
Col. (4) Tabulated Data — Premium
Col. (5) Tabulated Data — Actual Losses
Col. (6) Premium Cumulated Up $=$ Col. (4) Cumulated Up
Col. (7) Actual Losses Cumulated Up $=$ Col. (5) Cumulated Up
Col. (8) Adjusted Premium Cumulated Up $=$ (6) $\times F$
Col. (9) Excess Losses $=$ (7) $-$ (8) (1a) or $=$ (7) $-$ (6) (2a)
Col. (10) Excess Pure Premium Ratio $=$ (9) $\div$ Total (7)

$$F = \frac{\text{Total (5)}}{\text{Total (4)} \times \text{P.L.R.}}, \quad \text{Average Adjusted Premium} = \frac{\text{Total (5)}}{\text{Total (3)} \times \text{P.L.R.}}$$

It should be noted that unless the ratios of actual to expected losses or the loss ratios for individual risks have been calculated to more decimal places than is usually the case, the resulting x-ratios will not correspond to exactly the same loss ratios (column 1a) for all risk size groups. Because some variation is bound to occur, it is frequently advantageous to use the same set of sorting limits for all risk size groups in order to avoid the need for a hand

EXHIBIT A

Calculation of Excess Pure Premium Ratios

(New York State Workmen's Compensation Risks — Premium Size Group $4,000 to $6,687 — Hazard Group 1)

(1)	(2)	(2a)	(3)	(3a)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(11a)
					Tabulated Data								
Desired Loss Ratio	Desired Upper Limit of R $\frac{(1)}{PLR}$ 581	Actual Upper of R $\frac{(3a)}{F}$	Desired Upper Sorting Limit (2) × F	Actual Upper Sorting Limit	No. of Risks	Expected Losses	Actual Losses	Expected Losses Cumulated Up Col. (5) Cumul. Up	Actual Losses Cumulated Up Col. (6) Cumul. Up	Adjusted Expected Losses Cumulated Up (7) × F	Excess Losses (8) − (9) (2a)	Excess Pure Premium Ratio $\frac{(10)}{\text{Total (6)}}$	Actual Loss Ratio PLR × (2a)
.000	.0000	.00000	.0000	.000	—	——	——	534,403	518,947	518,960	518,947	1.00000	.00000
.200	.3442	.34497	.3343	.335	34	106,787	22,865	427,616	496,082	415,259	352,830	.67990	.20043
.400	.6884	.69509	.6685	.675	47	142,351	73,362	285,265	422,720	277,022	230,165	.44352	.40385
.500	.8606	.87015	.8357	.845	15	47,685	35,218	237,580	387,502	230,715	186,745	.35985	.50556
.600	1.0327	1.03491	1.0029	1.005	19	59,321	54,363	178,259	373,139	173,108	153,988	.29673	.60128
.800	1.3769	1.38503	1.3371	1.345	20	62,280	73,393	115,979	259,746	112,628	103,753	.19993	.80470
1.000	1.7211	1.72485	1.6714	1.675	13	39,768	57,162	76,211	202,584	74,009	74,930	.14439	1.00214
1.500	2.5817	2.57955	2.5071	2.505	17	50,727	104,945	25,484	97,639	24,748	33,800	.06513	1.49872
2.000	3.4422	3.44455	3.3427	3.345	4	12,599	35,545	12,885	62,094	12,513	18,992	.03660	2.00128
		6.18371	Over	6.005	4	12,885	62,094	——	——	——	——	——	3.60132
			Total	Total	173	534,403	518,947						

$F = \frac{518947}{534403} = .9711$ Average Adjusted Expected Loss $= \frac{518047}{173} = 3{,}000$ Avg. Adj. Prem. $= \frac{3000}{PLR} = 5{,}164$ 581

Note: Except in very accurate work columns (1a), (2a) and (3a) are omitted and columns (1), (2) and (3) are used in their place. Although column (9) is usually calculated and used to obtain column (10), it should be noted that column (9) is not necessary as column (10) can be calculated as equal to (8) — (7) (3a).

sort of the cards and the calculation of the desired sorting limit. Resulting x-ratios are then plotted against the actual loss ratio to which they correspond and interpolated values read, from such charts of all risk size groups, for the desired loss ratios.

Whenever it is necessary to make a wide grouping of risk sizes in order to obtain a sufficient number of risks, the problem arises as to the effect of this variation by size of risk. The above procedure produces x-ratios for the observed distribution of risks. If the risks of any one size group were divided into two sub-groups in some random fashion, two sets of x-ratios would result, which frequently would be widely different. This only illustrates the difference between the calculation of the actual x-ratios for a set of observed risks and the estimation of the most probable set of x-ratios to use in the rating of other risks or of the same risks in future years. To obtain the most probable set of x-ratios for future use, some type of smoothing procedure should be applied to the observed data.

One smoothing procedure of considerable value consists of the reference to already smoothed values obtained from a much larger group of risks of various risk sizes. Such values are available, for example, in the tables of x-ratios for New York workmen's compensation risks prepared by the New York Compensation Insurance Rating Board. Such a procedure will, of course, include in the smoothing process any imperfections that were incorporated in the table used for reference by the methods of its development.

The most important single statistic characterizing a set of x-ratios is the second moment of R, $U_{2:R}$, of the distribution of risks which it represents. Just as the x-ratios can be calculated from the distribution of risks by R, so can the value of $U_{2:R}$ be calculated from the x-ratios. The formula for doing this is based on the procedures of determining moments by successive summations and is:

$$U_{2:R} = 2\,B \sum_{0}^{\infty} (x\text{-ratios}) + B - 1$$

$$\text{where } B = \frac{\text{class interval of the loss ratios in } x\text{-ratio table}}{\text{permissible loss ratio of the } x\text{-ratio table}}$$

For the New York workmen's compensation table of x-ratios

$$B = \frac{.010}{.598} = .016722$$

and the formula becomes:

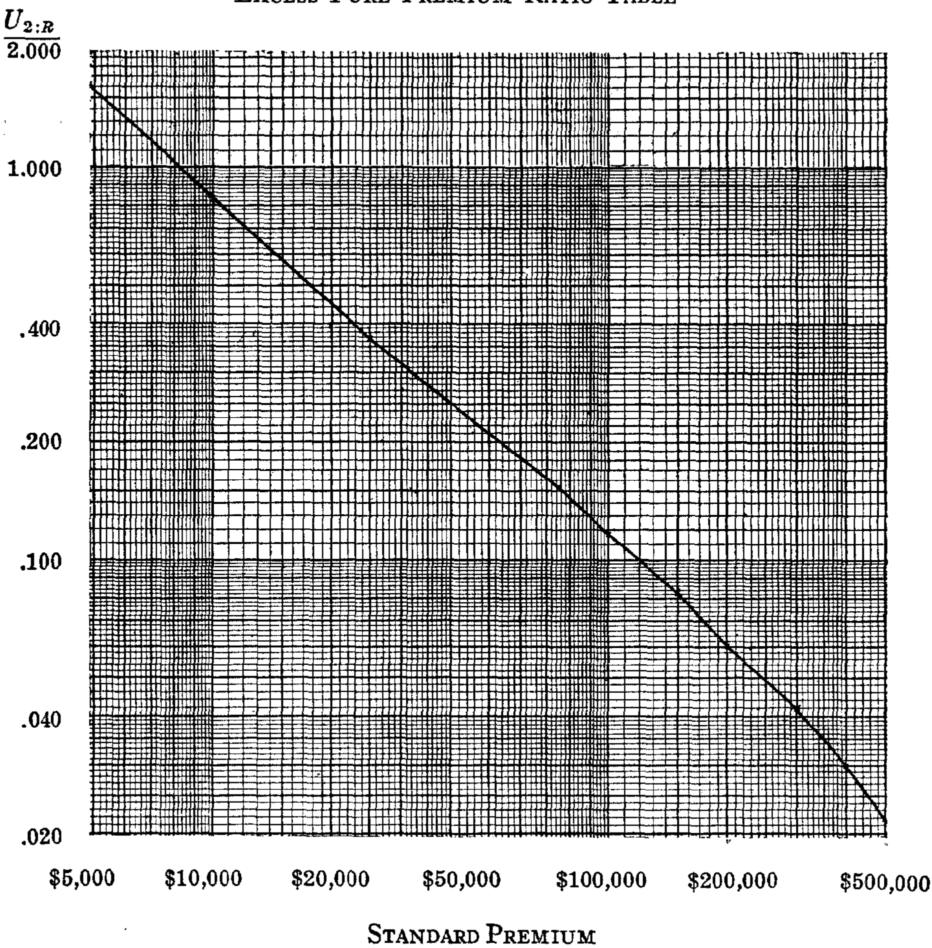
$$U_{2:R} = .033444 \sum_{0}^{\infty} (x\text{ ratios}) - .983278$$

The values of $U_{2:R}$ calculated from this formula for the New York x-ratio table are shown on Figure 3 plotted against the standard premium amount. Thus the standard premium amount in the New York Table can be read for any desired value of $U_{2:R}$.

FIGURE 3

STANDARD PREMIUM AMOUNTS FOR VALUES OF $U_{2:R}$

NEW YORK WORKMEN'S COMPENSATION INSURANCE EXCESS PURE PREMIUM RATIO TABLE



In the example we are using the value of $U_{2:R}$ was calculated from the individual risk values of R as .794. The equivalent standard premium amount from Figure 3 is $10,500. The x-ratios obtained from the New York table for this amount are compared below with those calculated from Exhibit A.

		Excess Pure Premium Ratios		
R	Corresponding Loss Ratio $= R \times .598$	From Exhibit A	From N. Y. x-Ratio Table Prem. = $10,500	Normal Logarithmic Freq. Distribution (See Following Discussion)
.34497	.2063	.680	.679	.687
.69509	.4157	.444	.448	.453
.87015	.5203	.360	.361	.366
1.03491	.6189	.297	.297	.301
1.38503	.8282	.200	.198	.199
1.72485	1.0315	.144	.141	.136
2.57955	1.5426	.065	.060	.055
3.44455	2.0598	.037		.024
6.18371	3.6979	.000		.003

The x-ratios calculated on Exhibit A or obtained on the basis of the moments of R do not correspond to any particular premium size, being only averages applicable to the range of risk sizes from which they were obtained. The importance of this is seen from the fact that such an average is always less than the true value of the x-ratio corresponding to the average risk size of the size group. An example using x-ratios from the New York table will serve to demonstrate this:

(1) Risk	(2) Premium	(3) Expected Losses .598 (2)	(4) x-Ratio for 100% Loss Ratio	(5) Expected Excess Losses (3) × (4)
A	5,000	2,990	.237	709
B	5,000	2,990	.237	709
C	10,000	5,980	.153	915
D	20,000	11,960	.109	1,017
Average	10,000	5,980	.140	837
True Value	10,000		.153	

There is no way of directly calculating the true x-ratios following any procedure such as that of Exhibit A. The proper correction can, however, be made in the determination of $U_{2:R}$ by obtaining its value from $U_{2:\,(\text{Corrected } x)}$ or from $U_{2:W'}$, as these functions are designed to provide the function of the average expected loss rather than an average function for all expected loss sizes in the size group. The value of $U_{2:R}$ determined from the W' moments is .883 in our example and corresponds to a standard premium of $9,300 in the table of New York x-ratios. Estimates of the true x-ratios on this basis compare with the calculated values from Exhibit A as follows:

R	Corresponding Loss Ratio $= R \times .598$	Calculated x-Ratios From Exh. A	Estimates of True x-Ratios: From N. Y. x = Ratio Table Prem. = $9,300	Estimates of True x-Ratios: Normal Logarithmic Freq. Distribution (See following discussion)
.34497	.2063	.680	.684	.688
.69509	.4157	.444	.458	.459
.87015	.5203	.360	.373	.375
1.03491	.6189	.297	.309	.311
1.38503	.8282	.200	.212	.212
1.72485	1.0315	.144	.155	.149
2.57955	1.5426	.065	.070	.066
3.44455	2.0598	.037		.031
6.18371	3.6979	.000		.004

The above smoothing procedure cannot be applied when it is known that the distribution of risks by R has a skewness widely different from that underlying the reference table x-ratios of the standard premium amount having the same $U_{2:R}$. It obviously should not be used to develop a new table of x-ratio values. An independent procedure making use of all known facts should be applied in the latter case.

There is one peculiar characteristic of an observed frequency distribution which should be recognized here. This is, that no matter how many observations have been made, there is always the possibility that the next observation will have a value greater than any yet observed. Similarly, for the loss ratio distribution, if no zero loss ratios have been observed, there is always the possibility that one will occur. The result is that any smoothing process applied directly to a frequency distribution will include the probabilities of more extreme cases occurring than any of those observed.

This same condition when followed through into the x-ratios will produce, from a smoothed frequency distribution, higher x-ratios than those observed for high loss ratios and usually higher values for the low loss ratios. Intermediate values of the x-ratios will be uniformly lower than the observed values. Such a consistent departure of the smoothed x-ratios from those calculated from observations should not be viewed with alarm but should be recognized as the provision against the eventual occurrence of extreme cases.

The normal logarithmic frequency distribution has been found to satisfactorily fit many distributions of R. In the particular case in hand this distribution fits very well as shown by the following:

Range of R	Actual No. of Risks	Expected No. Based on Normal Log. Freq. Distribution	$\frac{(A-E)^2}{E}$
0- .19	16	18	.22
.20- .39	27	25	.16
.40- .59	22	26	.62
.60- .79	29	23	1.57
.80- .89	19	18	.06
1.00-1.49	32	30	.13
1.50-1.99	10	15	1.67
2.00-2.99	13	12	.08
3.00 & Over	5	6	.17
	173	173	$\chi^2 = 4.68$

The value of 4.68 for Chi-Square, with $n = 6$ (obtained by subtracting the 3 parameters used in fitting the distribution from the 9 groups), represents a probability of between .5 and .7 that, purely as a result of chance variation, a divergence from the theoretical distribution as great or greater than that actually observed would occur.

This distribution, or any other providing a reasonable fit to the observed data, may be used to smooth the observed data prior to the calculation of x-ratios from the formula:

$$\chi_{(R',E)} = 1 - \frac{R - {}^{R}\!\int_0 {}^{t}\!\int_{-\infty} \phi_t \cdot dt \cdot dR}{V_{1:R}}$$

which gives the x-ratio, to be applied when $V_{1:R} = 1$, for values of R' equal to $\frac{R}{V_{1:R}}$. When the normal logarithmic distribution is used, the value of ϕ_t and ${}^{t}\!\int_{-\infty} \phi_t \cdot dt$ can be read from tables of the normal distribution directly, and the value of ${}^{R}\!\int_0 {}^{t}\!\int_{-\infty} \phi_t \cdot dt \cdot dR$ can be calculated from these by the application of the Euler-Maclaurin formula which produces:

$${}^{R}\!\int_0 {}^{t}\!\int_{-\infty} \phi_t \cdot dt \cdot dR = \frac{w}{2} \, {}^{t(n)}\!\int_{-\infty} \phi_t \cdot dt + w \sum_0^{n-1} {}^{t(n)}\!\int_{-\infty} \phi_t \cdot dt$$

$$- \frac{.03619117\, w^2}{\sigma_e\,(R-a)} \phi_{t(n)} - G$$

where $G = \frac{w}{2} \, {}^{t(0)}\!\int_{-\infty} \phi_t \cdot dt = \frac{.03619117\, w^2}{\sigma_e\,(R-a)} \phi_{t(0)}$ and

where $t = \frac{\log_{10}(R-a) - l_0}{\sigma_e}$ and, a, l_0 and σ_e are the parameters of the normal logarithmic distribution, and where ${}^{t}\!\int_{-\infty} \phi_t \, dt$ and ϕ_t are obtained for values of R starting with 0 and increasing by intervals equal to w. Thus $R_{(n)} = 0 + n\,w$ and $t_{(n)}$ is the value of t corresponding to $R_{(n)}$.

The application of this procedure to the previously used data to obtain x-ratios is shown in detail in Exhibit B. The results for specific values of B, obtained by second difference interpolation from the values in Exhibit B, have been previously shown for comparative purposes.

In the above described processes the x-ratios are obtained on the basis that the difference between $V_{1:R}$ and unity, its expected value, is a significant difference that should be carried through as an adjustment to the entire distribution of risks according to R. That is, the factor $F = V_{1:R}$ has been applied to the unit of measurement and, in effect, enters as F^2 into the correction of $V_{2:R}$ and as F^3 for $V_{3:R}$. The actual facts of the case may be, however,

EXHIBIT B

Calculation of Excess Pure Premium Ratios Based on Best Fitting Normal Logarithmic Frequency Distribution

(New York State Workmen's Compensation Risks — Premium Size Group \$4,000 to \$6,687 — Hazard Group 1)

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
n	R	$R\text{-}a$	$log_{10}(R\text{-}a)$	t	ϕt	$\dfrac{\phi t}{R\text{-}a}$	$\int_{-\infty}^{t} \phi_t \cdot dt$	$\sum_{0}^{n\text{-}t} \int_{-\infty}^{t} \phi_t \cdot dt$	$\int_{0}^{R} \int_{-\infty}^{t(n)} \phi_t \cdot dt dR$	$\dfrac{(9)}{V_{1:R}}$	Adjusted R $= R'$	Excess Pure Premium Ratios
	$= nw$	(1) + .247609	$log_{10}(2)$	.012562 +3.605605(3)	From Normal Freq. Dist. Table	(5) ÷ (2)	From Normal Frequency Dist. Table	$\sum_{0}^{n-1}$ (7)	—.001305(6) +.05(7)+.1(8) —.000546	1.03222(9)	1.03222(1)	1+(10)—(11)
0	.00	.247609	—.6062336	—2.173	.03763	.15197	.01489	——	.00000	.00000	.00000	1.00000
1	.10	.347609	—.4589089	—1.642	.10362	.29809	.05029	.01489	.00307	.00317	.10322	.89995
2	.20	.447609	—.3491012	—1.246	.18357	.41011	.10639	.06518	.01076	.01111	.20644	.80467
3	.30	.547609	—.2615294	— .930	.25888	.47275	.17619	.17157	.02480	.02560	.30967	.71593
4	.40	.647609	—.1886872	— .668	.31916	.49283	.25207	.34776	.04619	.04768	.41289	.63479
5	.50	.747609	—.1262255	— .443	.36165	.48374	.32889	.59983	.07525	.07767	.51611	.56156
6	.60	.847609	—.0718044	— .246	.38705	.45664	.40284	.92872	.11187	.11547	.51933	.49614
7	.70	.947609	—.0233708	— .072	.39791	.41991	.47130	1.33156	.15563	.16064	.72255	.43809
8	.80	1.047609	.0201992	.085	.39750	.37944	.53387	1.80286	.20594	.21258	.82578	.38680
9	.90	1.147609	.0597939	228	.38870	.33870	.59017	2.33673	.26219	.27064	.92900	.34164
10	1.00	1.247609	.0960785	359	.37404	.29981	.64021	2.92690	.32376	.33419	1.03222	.30197
etc.												

$V_{1:R} = .968786$ $\quad a = -.247609$

$U_{2:R} = .745064$ $\quad l_0 = -.003484$

$\alpha_{3:R} = 1.486148$ $\quad \sigma_e = +.277346$

$$t = \frac{log_{10}(R\text{-}a) - l_0}{\sigma_e}$$

$$t = -.012562 + 3.605605(3)$$

$$w = .1$$

$$\int_{0}^{R} \int_{-\infty}^{t(n)} \phi_t \, dt dR = \frac{-.03619117 w^2}{\sigma_e} \times \frac{\phi_t}{R\text{-}a} + \frac{w}{2} \int_{-\infty}^{t} \phi_t \, dt + w \sum_{0}^{n-1} \int_{-\infty}^{t} \phi_t \cdot dt - G$$

$$= -.001305(6) + .05(7) + .1(8) - G$$

$$G = -.001305(.15197) + .05(.01489) = .000546$$

that the observed difference between unity and $V_{1:R}$ is entirely the result of chance variations. Similarly the values of $V_{2:R}$ and $V_{3:R}$ used to fit the theoretical distribution are likewise subject to chance fluctuations. The results of these chance fluctuations are readily seen when the values of the x-ratios, for a particular loss ratio, are plotted against the average adjusted size of risk for several risk size groups. Whatever method of calculation of the x-ratios has been used, such a chart will show an appreciable variation about a trend. Rather than use some artificial method of smoothing these x-ratios by size of risk, it would seem more reliable to use the available knowledge of the functional form of the variation in the mutually independent statistics: $V_{1:R''}$, $(CV)^2$, and $\alpha_{3:R''}$. From the results of C-3 of Part IV we have that these statistics are functions of E of the following form:

$$E\cdot V_{1:R''} = A\cdot E$$

$$E\cdot (CV)^2 = B\cdot E + C$$

$$E\cdot \alpha_{3:R''} = D\cdot E + F\sqrt{E} + G + \frac{H}{\sqrt{E}}$$

where E is the expected loss and H and G are of least importance and should be disregarded unless a very large amount of data is available. In determining the parameters in these equations the observed values for the various size groups should be weighted by the number of risks in the groups.

When $V_{1:R}$ shows a consistent downward (or upward) trend as E increases, as often happens, the further problem arises as to whether such a trend should be recognized in the preparation of a table of x-ratios. If, for example, the table were to be used to experience rate all risks, it would be possible to have manual rates keyed to the level of losses of small risks and to include the credit (or debit) for the larger risks in the x-ratio table. The x-ratios in such a table would start at $V_{1:R}$ for a loss ratio of 0 rather than at unity. The usual procedure is for the rate structure to correct for any loss level differences by size of risk that are definite enough to be recognized. In that case, or if the table is not to be used for rating all risks, it is necessary to base the table of x-ratios on an adjusted value of $V_{1:R}$ of unity for all risk sizes and to carry this adjustment through into the values of E.

The procedure developed by Mr. Dorweiler consisted of smoothing the calculated x-ratios for different size groups by the use of an empirical formula having no a priori relationship to the expected form of curve, and then smoothing such values for a single risk size by the judicial use of a french curve. Such a procedure obviously cannot be used if the data is available for only a single size group. Such a method would also produce consistently too low x-ratios due to the use of uncorrected size group averages as pointed out above. Within the range of the majority of observations, the results of different types of smoothing processes would differ only slightly. Beyond the range of the majority of observations, all methods are subject to the dangers inherent in any attempt to extrapolate. Thus, the values in any x-ratio table for loss ratios above which only a few risks have actually been observed should be used only in conjunction with an appreciable "balance for contingencies."

VII.

Two Kinds of Credibility

When two sources of experience indicate different values of a statistic (claim frequency, average claim cost, pure premium, or loss ratio), it is customary to use a weighted average of the two using "credibilities" as weights. Such credibilities are, or may be, designed to accomplish one but not both of two separate results. One type provides an average value which will not fall outside of a specified range of accuracy in more than a specified proportion of all cases as a result of fluctuations in one value due to chance only. This type might be termed the limited fluctuation credibility and is the type generally used in developing manual rates. The other type is that designed to provide the most accurate average of the two values irrespective of how much variation will result in the average and recognizing all types of variation in both of the individual values and not just the chance variation in one of them. This is the type of credibility that most actuaries have in mind in dealing with experience rating plans and which would be most effective if applied in retrospective rating plans.

A. The "Limited Fluctuation" Credibility

If the two sets of values to be averaged have the same average value, and if Z is the credibility given to the value which is subject to chance variation, then the variation due to chance in the weighted average will be Z times the variation due to chance in the value having such variation. If it is specified that $B\%$ of all cases may fall outside of the range to be specified, then the difference between the values in the $\left(\frac{B}{2}\right)$ and $\left(1-\frac{B}{2}\right)$ columns of Tables 5, 10, and 11 of Part II give the range of variation of the statistic. Furthermore, if the specified range is to be $A\ \%$ of the average, then

$$A = Z\left[\left(1-\frac{B}{2}\right)\text{Value} - \left(\frac{B}{2}\right)\ \text{Value}\right]$$

$$\text{and } Z = \frac{A}{\left(1-\frac{B}{2}\right)\text{Value} - \left(\frac{B}{2}\right)\ \text{Value}}$$

where Z varies from 0 to 1.00.

It has frequently been prescribed in developing credibility formulae that $A = B = 10\%$. The following table shows the values of Z calculated from Tables 5, 11, and 10, respectively, with such specifications for each of claim frequencies, average claim costs, and for pure premiums or loss ratios. For comparison with these, credibility values are shown calculated from the formula:

$$Z = \sqrt{\frac{C}{C \text{ for } 100\% \text{ Credibility}}}$$

It should be noted that these credibilities for average claim costs, total losses, pure premiums, and loss ratios are applicable only to New York commercial automobiles—property damage coverage.

COMPARISON OF "LIMITED FLUCTUATION" CREDIBILITIES

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number of Claims or Claim Frequencies		Average Claim Costs of a Fixed Number of Claims		Total Losses, Pure Premiums, or Loss Ratios	
Number of Claims*	From Table 5	$\sqrt{\frac{(1)}{1{,}084}}$	From Table 11	$\sqrt{\frac{(1)}{3{,}520}}$	From Table 10	$\sqrt{\frac{(1)}{4{,}590}}$
1	.033	.030	.031	.017	.026	.015
4	.057	.061	.050	.034	.040	.030
10	.100	.096	.070	.053	.057	.047
40	.190	.192	.120	.107	.101	.093
90	.286	.288	.171	.160	.146	.140
160	.385	.384	.223	.213	.192	.187
250	.481	.480	.276	.267	.238	.233
360	.575	.576	.328	.320	.284	.280
490	.671	.672	.382	.373	.331	.327
640	.763	.768	.437	.426	.379	.373
810	.862	.864	.488	.480	.426	.420
1,000	.962	.960	.538	.533	.472	.467
1,440	1.000	1.000		.640	.565	.560
1,960	1.000	1.000		.746	.658	.653
2,560	1.000	1.000		.853	.752	.747
3,240	1.000	1.000		.959	.847	.840
4,000	1.000	1.000		1.000	.935	.934
4,840 & Over	1.000	1.000		1.000	1.000	1.000

* For columns (2), (3), (6), and (7) these are the expected number of claims; while for columns (4) and (5) they are the actual number of claims.

B. The "Greatest Accuracy" Credibility

In section *I* of Part I the least squares solution for Z in the formula:

True Inherent Hazard $= Z$ (Actual Losses) $+ (1 - Z)$ (Expected Losses)

was found to be $Z = \frac{U_{2:m}}{U_{2:R'}}$ where $U_{2:m}$ represented the second moment of all types of variation of the actual losses from the expected losses except that due to chance, and where $U_{2:R'}$ represented the second moment of the ratio of actual to expected losses. In the notation of Part IV we would have:

$$Z = \frac{U_{2:R''} - U_{2:R}}{U_{2:R''}}$$

where $U_{2:R''}$ is to be *calculated from the data to which the credibilities are to be applied* as outlined in Part V and where $U_{2:R}$ is to be calculated from the moments of the distribution of claims by size of claim for the same data from the formula:

$$U_{2:R} = \frac{1}{E} \cdot \frac{V_{2:x''}}{V_{1:x''}}$$

One very important point in regard to this type of credibility is that it does not always or even usually have the range of possible values from 0 to 1.00. One extreme case is where the "expected" losses are those based on one year of actual experience and the "actual" losses are those of another year in which all elements of exposure and hazard remain unchanged. Under these conditions Z would be .50 irrespective of the value of E. In an application to New York commercial automobile property damage experience of individual risks, a minimum credibility of approximately .30 and a maximum credibility of approximately .85 have been developed. This only serves to remind us that, no matter how little experience is available for the most recent period, it is worth looking at and considering and, conversely, no matter how much recent experience is available, it is worth giving some consideration to the past experience.

From section C-3 of Part IV we would evaluate Z as:

$$Z = \frac{\frac{V_{1:M}}{V_{1:E}} \cdot \frac{V_{2:x''}}{V_{1:x''}} + U_{2:M}\,(1+V_{2:m'}) + V_{2:m'}\,(1+V_{1:M})^2}{U_{2:R''}}$$

or, if adjustments have been made to eliminate rate level differences,

$$Z = \frac{U_{2:M} + V_{2:m'} + U_{2:M} \cdot V_{2:m'}}{U_{2:R''}}$$

for individual risk experience or

$$Z = \frac{U_{2:M}}{U_{2:R''}}$$

for classification experience. In reviewing these formulae, it should be borne in mind that $U_{2:M}$ may vary by size of classification and $V_{2:m'}$ may vary by size of risk.

The most important feature, however, is that this credibility will be greatest when the new experience differs most from that underlying manual rates: that is, when there is the greatest need for a revision of rates, and will approach zero when the new experience evidences only chance variation from that underlying the existing rates.