PARAMETRIZING THE WORKERS COMPENSATION EXPERIENCE RATING PLAN

WILLIAM R. GILLAM

Abstract

This paper describes the development of the revised Workers Compensation Experience Rating Plan. The plan is based on sound statistical theory, certain modeling assumptions, and thorough empirical testing. It is heartening that the empirically derived parametrization is consistent with most of the assumptions needed to simplify the algebraic foundation.

The paper begins with an heuristic derivation of a general modification formula based on losses split into primary and excess portions. It delineates the assumptions about the components of loss ratio variance leading to the algebraic form of the formulae tested.

Iterative testing is used to parametrize those formulae. A simple preliminary test procedure is described to clarify the basic concepts. The operative test procedure is then specified, and results of iterative testing using that procedure are displayed for the selected formulae.

The parametrized formula finally approved by the National Council on Compensation Insurance (NCCI) was subject to certain adjustments to maintain continuity during the transition from the old to new plans. Credibilities have been scaled to account for differences in state benefit levels and the effect of inflation.

1. THEORETICAL JUSTIFICATION

A. Motivation

Compensation experience rating is a large-scale, ongoing application of credibility theory. The large volume of data supporting that application provides the raw material for the tests of that theory described below.

Researchers at the NCCI have used testing to create an improved experience rating plan. The power of modern electronic data processing has enabled them to reopen older experience rating files and recalculate experience modifications (mods) as if a hypothetical plan had been in place. The plans tested and the measurement of their performance are described in this paper.

The general strategy was to start with a formula based on sound theory, then use iterative testing to parametrize that formula. Least squares or Bayesian credibility was used to develop an algebraic form for the modification formula. Certain assumptions about loss ratio variance simplified the algebra. The parameters that worked best were consistent with a priori judgments about the components of loss variance.

B. Heuristic Derivation of Mod Formula

This section outlines the theoretical development of the split plan modification formula. Here, a *split plan* is one in which individual losses are split by formula into two components, primary and excess, and separate credibilities are assigned to the totals of the respective loss components.

The formula is based on a Bayesian view of the process of individual risk rating. The reader may refer to papers by Hewitt [1], Meyers [2], Mahler [3], and Venter [4] for more general theoretical background.

The split plan modification formula can be derived with one major simplifying assumption: that unconditional expected primary and excess losses are uncorrelated. This simplification is defensible more on the basis of its usefulness than its veracity. The standard used to select the final plan is how well it works, not how well it satisfies the assumptions. In the course of evaluating plan parameters, NCCI researchers found that a change in the primary/excess split formula improved the performance of the plan. They believe this change places the data used for rating in a form that better fits the assumptions.

The underlying analysis is simplified by assuming that most of the administrative features of the current experience rating plan are fixed.

To begin the derivation, hypothesize a linear approximation to the posterior mean experience, $P_o + X_o$ (split primary + excess), given experience P_t and X_t:

$$P_o + X_o = Y + Z_p P_t + Z_x X_t + e, \tag{1.1}$$

where

P = primary loss,

X = excess loss,

Y = constant to be determined,

t = (past) time period,

o = (future) time period, and

e = error.

Z_p and Z_x will be called the respective primary and excess credibilities; they and Y are the coefficients to be evaluated.

Time periods are fixed in the experience rating plan, so that time period t is the three most recently completed one-year policy periods before the prospective single policy period, labeled o. For example, the experience of completed policies with inception dates in 1986, 1987, and 1988 will be used to rate a 1990 policy.

Solving this equation for the coefficients that minimize the expected value of e^2 (with the assumption mentioned above) yields the following expressions:

$$Z_p = \frac{\mathrm{Var}_S\,[\mathrm{E}\,[P_t \mid S]\,]}{\mathrm{Var}\,[P_t]}, \text{ and}$$

$$Z_x = \frac{\mathrm{Var}_s\,[\mathrm{E}\,[X_t \mid S]\,]}{\mathrm{Var}\,[X_t]}\,. \tag{1.2}$$

And

$$Y = (1 - Z_p)\,\mathrm{E}\,[P_t] + (1 - Z_x)\,\mathrm{E}\,[X_t]\,. \tag{1.3}$$

where the condition S denotes a particular element of the parameter space (a particular risk) and the subscript s denotes the prior structure (the distribution of risk parameters).

Equation 1.2 has also been written

$$Z_p = \frac{1}{1 + \dfrac{\mathrm{E}_s\,[\mathrm{Var}\,[P_t \mid S]\,]}{\mathrm{Var}_s\,[\mathrm{E}\,[P_t \mid S]\,]}}\,,$$

using $\mathrm{Var}\,[P_t] = \mathrm{Var}_s\,[\mathrm{E}[P_t \mid S]\,] + \mathrm{E}_s\,[\mathrm{Var}\,[P_t \mid S]\,]$. The first term is the *variance of the conditional means* or the *between-variance*. The second term is the *expected value of the conditional variance* or the *within-variance*.

Using these equations, the linear credibility estimate of the posterior mean becomes

$$P_o + X_o = \mathrm{E}[P_t] + \mathrm{E}[X_t] + Z_p\,(P_t - \mathrm{E}[P_t]\,) + Z_x\,(X_t - \mathrm{E}[X_t]\,). \tag{1.4}$$

In practice, the loss functions are ratios to the prior expected total loss, so $\mathrm{E}[P_t] + \mathrm{E}[X_t] = 1$. In this paper, P and X are referred to as *loss ratios*, but the denominator is expected loss, not premium.

The rate modification factor is

$$M = 1 + Z_p\,(P_t - \mathrm{E}[P_t]\,) + Z_x\,(X_t - \mathrm{E}[X_t]\,)\,. \tag{1.5}$$

C. Variance Assumptions

More assumptions are needed to derive the form of the components of variance in the formulae for Z_p and Z_x.

First-Level Variance Assumptions

In Equation 1.4, loss ratio functions P and X were introduced. Those ratios have a variance that decreases as the size of risk increases. The sample ratios P_t and X_t are the emerged primary and excess actual losses of the individual risk divided by the unconditional expected total losses. The denominator is the *exposure*. The simplest assumption is that the large risk is essentially a combination of a large number of independent homogeneous units. That assumption leads to a within-variance of the risk loss ratio inversely proportional to exposure. The increase in exposure from additional time periods can be thought of as adding more independent units of exposure. The process variance decreases proportionately. Also, it is usually assumed that the variance of the conditional means is independent of exposure (i.e., size of risk). With those assumptions,

$$\mathrm{E}_s\,[\mathrm{Var}[P_t \mid S]\,] = a/E\ ,$$

and

$$\mathrm{Var}_s\,[\mathrm{E}[P_t \mid S]\,] = b\ ,$$

where E with no subscript represents the total expected losses, or exposure, of the individual risk: $E = \mathrm{E}[P_o + X_o \mid S]$.

Here b, the variance of ratios less than one, is small relative to a, which is measured in dollars of expected loss.

Using equation 1.2,

$$Z_p = \frac{b}{b + a/E}$$

$$= \frac{1}{1 + a/bE}$$

$$= \frac{E}{E + a/b}$$

This is the familiar expression

$$Z_p = \frac{E}{E + K_p}, \tag{1.6}$$

where K_p is constant. Similarly,

$$Z_x = \frac{E}{E + K_x},$$

where K_x is the excess credibility constant. This *compound fraction* form, with E alone in the numerator and K a ratio of components of variance, helps to simplify the mod formula.

Second-Level Variance Assumptions

Several investigators have refuted this simple variance assumption. Meyers [2] and Mahler [3] show that within-variance does not decrease in inverse proportion to exposure. Assuming there is a small, non-diversifiable component of risk loss ratio variance, averaging $c > 0$,

$$\mathrm{E}_s[\mathrm{Var}[P_t \mid S]] = c + d/E.$$

Using b again as the between-variance,

$$Z_p = \frac{b}{b + c + d/E}$$

$$= \frac{1}{1 + c/b + d/Be}$$

$$= \frac{E}{E + \frac{cE + d}{b}}, \text{ so}$$

$$Z_p = \frac{E}{E + K_p'}, \text{ where}$$

$$K_p' = \frac{cE + d}{b} . \tag{1.7}$$

Now K_p' is a linear function of the exposure. Here, b and c are small relative to d. The limiting value of primary credibility for the largest risks is less than unity, or $b/(b + c)$.

This form for K_p' and a similar one for K_x' are among possible formulae tested as described elsewhere in the paper. Because K_* ($*$ either p or x) is a linear function of the exposure rather than a constant, it performs better than the constant coefficient K, and considerably better than the formula B value of the old plan. However, it is not as good as the third-level formulae described below. The data show that K should not be constant, nor even a linear function of E, but rather should be a curve, increasing rapidly at first but then decreasing in slope to a more linear form for large values of E.

Third-Level Variance Assumptions

The variance assumptions resulting in the formula for K at this level were suggested by Mahler [3]. Mahler, in turn, credits Hewitt [5] with observation of the underlying phenomenon.

For this level, it is assumed that the between-variance is not constant across all risk sizes but has a component inversely proportional to exposure. This would follow if each larger risk was, at least in part, a random combination of non-homogeneous components. The effect is to flatten the variance of the conditional means as risk size increases. In this case,

$$\mathrm{Var}_s [\mathrm{E}\,[P_t \mid S]\,] = e + f/E .$$

Retaining the second assumption about individual risk variance,

$$Z_p = \frac{e + f/E}{e + f/E + c + d/E}$$

$$= \frac{1}{1 + \dfrac{c + d/E}{e + f/E}} .$$

In compound fraction form, this is

$$Z_p = \frac{E}{E + \dfrac{cE + d}{e + f/E}}$$

$$= \frac{E}{E + E\left(\dfrac{cE + d}{eE + f}\right)}, \text{ so}$$

$$Z_p = \frac{E}{E + K_p''},$$

where

$$K_p'' = E\left(\frac{cE + d}{eE + f}\right). \tag{1.8}$$

A similar form follows for K_x''. Notice that d and f are quite large compared to c and e. Since c is a small component of within loss ratio variance and e is a large component of between loss ratio variance, it is also plausible that $c < e$.

Dividing through by e, we define $C = c/e$, $D = d/e$, and $F = f/e$, so that

$$K_* = E\left(\frac{CE + D}{E + F}\right). \tag{1.9}$$

This form is selected for parametrization, so the superscripts have been dropped.

In all the sample parameters that worked well (as described below), C was consistently between 0 and 1, which is reasonable if c is in fact smaller than e. D and F are large positive numbers, as expected.

In a sense, the final parametrization selected is more general than the underlying variance assumptions. This is because performance testing was used to derive parameters for the modification that worked best. This obviates the need to estimate components of variance and reduces reliance on the correctness of assumptions. Thus, the only constraint on plan performance was the algebraic form of Equation 1.9, not the ability to analyze variance. Statisticians use *robust* to refer to models such as these that can fit a variety of processes while not necessarily satisfying the assumptions underlying the model.

With the definition of K_p and a similar one for K_x underlying Z in the form $E/(E + K_*)$, the modification formula becomes,

$$M = 1 + \frac{A_p - E_p}{E + K_p} + \frac{A_x - E_x}{E + K_x}, \tag{1.10}$$

where A and E are the actual and expected losses from the experience period and p denotes primary and x excess. The algebraic form of the modification used for most of the testing was Equation 1.10, with K_p and K_x defined as in Equation 1.9. For each of K_p and K_x, there are coefficients to estimate.

2. ESTIMATION OF PARAMETERS

A. Initial Testing

The concept of evaluating workers compensation individual risk credibility by looking back at how well it worked was discussed by Dorweiler [6]. He did not use his method to establish credibilities, but to check them for reasonableness.

Bailey and Simon [7] described a variant of the procedure for automobile merit rating wherein they were able to estimate the implicit credibilities of one, two, or three car-years. They were not trying to parametrize a continuous formula for credibility depending on exposure, however, but were trying to estimate only three values for one, two, or three car-years' experience.

Today, we are able to take their ideas a step further, largely because of the power of the computer.

In this study of experience rating, the criterion of "working best" is first measured by the ability of a plan to satisfy Dorweiler's necessary criterion for correct credibilities: that credit risks and debit risks would be made equally desirable insureds in the prospective period. In workers compensation, credibility is a function of risk size, so this property should exist across all size categories. We use this criterion as a *naive test*, which belies its great value to our early investigations. It also serves to simplify the demonstration of the basic idea behind the testing.

An example of this test is included in Exhibit 1. Note that the plan proposed earlier is still a long way from the plan that was eventually selected. The test begins with experience rated risks for policies effective in 1981. Their modifications are computed according to the formula to be tested. The 1981 loss experience that actually emerged may be found in the 1983 rating year files, i.e., the data underlying mods effective 1983. The risks in each size group are stratified by their 1981 modifications, so that risks with mods in the lower 50^{th} percentile would be in one stratum and risks with mods in the upper 50^{th} percentile would be in the other. (It should be noted that for the smaller size groups, the majority of risks have credit mods, so the upper percentile includes a proportion of risks with small credits.)

A canonical comparison would be of the subsequent loss ratios of the two strata: actual losses to manual premium on one side, and actual losses to modified premium on the other. The first ratios, actual to manual, should follow the predicted quality of the stratum. Risks with credit mods should prove to have favorable loss ratios on average, and those with debits should average poor ratios, showing that the plan was indeed able to "separate the wheat from the chaff." In order to see if the differences in predicted quality were correctly offset by the mod, the ratio of losses to modified premiums for the two groups should equal each other. It would be too much to expect that premium rates be correct in aggregate and that the two subsequent loss-to-modified-premium ratios be equal to the permissible loss ratio.

Effective manual premiums for the three policy periods used for the modification are retained in the experience rating files. Unfortunately, since they are not used for either ratemaking or experience rating, the

EXHIBIT 1

1981
ACTUAL TO EXPECTED LOSS RATIOS BEFORE AND AFTER EXPERIENCE RATING — 7 STATES TOTAL

CURRENT EXPERIENCE RATING FORMULA

Risk Size	Quality Indication	Subsequent Period: Loss Ratios	Subsequent Period: Modified Loss Ratios
2,500-5,000	50% Best	0.75	0.80
	50% Worst	1.12	1.05
5,000-10,000	50% Best	0.71	0.80
	50% Worst	1.11	1.01
10,000-25,000	50% Best	0.79	0.92
	50% Worst	1.12	0.96
25,000-100,000	50% Best	0.75	0.89
	50% Worst	1.15	0.93
Over 100,000	50% Best	0.71	0.93
	50% Worst	1.00	0.79
Sum of Absolute Differences		1.79	0.68

EARLY PROPOSED EXPERIENCE RATING FORMULA

Risk Size	Quality Indication	Subsequent Period: Loss Ratios	Subsequent Period: Modified Loss Ratios
2,500-5,000	50% Best	0.76	0.91
	50% Worst	1.10	1.00
5,000-10,000	50% Best	0.71	0.89
	50% Worst	1.11	0.98
10,000-25,000	50% Best	0.79	0.98
	50% Worst	1.12	0.95
25,000-100,000	50% Best	0.75	0.91
	50% Worst	1.15	0.94
Over 100,000	50% Best	0.72	0.89
	50% Worst	0.99	0.83
Sum of Absolute Differences		1.74	0.30

numbers are seldom checked and are considered unreliable. The expected loss rates, or ELRs, by class in these files, however, are subject to review by insureds and insurers. These are not the true loss costs underlying the rates, but estimates of emerged loss for three policy years as of a certain evaluation date. The ELRs used to estimate rating year 1981 expected losses are used to compute the expected ratable losses for modifications effective during the 1983 policy year. The ELRs are meant to be correct on average for the losses on three policy years, including, in this case, 1979, 1980, and 1981. The three policy years are at respective third, second, and first reports. ELRs are probably not correct for any single policy year, but should bear some reasonable relation to the rates effective in the latest year. A key assumption is that the ELRs will be uniformly redundant or inadequate over all insureds with the same rate.

The comparison is then between the ratios of actual to manual expected loss for each stratum and ratios of actual to expected losses *adjusted by the modification*, or modified expected loss. Specifically, 1981 actual loss to 1981 expected loss, taken from the 1983 rating year files, should reflect the predicted quality difference. Application of the 1981 modification to 1981 expected losses should make the ratios converge.

Just as it was observed in the case of premiums, it is reasonable to hope that the subsequent ratios to modified expected loss would be close to each other, but it is unreasonable in this case to require values near unity.

The need for credibility in a format not unlike the one finally selected is evidenced by application of the naive test. The smallest ratable risks have non-zero excess as well as primary credibilities. Starting with the smallest risks, credibility increases rapidly with risk size, but then increases at a slower rate, and never reaches full credibility for even the largest risks.

B. The Quintiles Test

As the testing of the plans progresses and more sophisticated actuarial theory is applied to the algebraic form of the credibility constants, it becomes apparent that a more sophisticated test is needed to measure the quality of alternative formulae. Dorweiler's sufficient criterion for correctness of the modification is that any subdivision of risks based on prior experience should produce uniform subsequent loss ratios (to modified premium).

Instead of good versus bad as in the naive test, the risks are grouped into five equal-sized strata according to the value of their modifications. The lowest 20% of the values belong to risks in the first quintile; the next 20% to the second; and so on. This is the prior subdivision. The subsequent aggregate unmodified loss ratios of the strata should reflect the quality difference recognized by the mod. Application of the modifications should cause the ratios to flatten across the strata.

This leads to the ratio of two sums of squared differences: the five squared deviations from the mean of the modified loss ratios, divided by the sum of the squared deviations of the ratios before modification. Lower values indicate greater reduction of loss ratio variances. The statistic would pertain to the experience of each group, so for a particular parametrized mod formula, several values are available for comparison. In most of the NCCI testing, coincidentally, five groups were considered.

The quintiles test was developed without reference to risk theory, but it can be characterized as the ratio of posterior structure variance to prior structure variance. The sum of five squared deviations does not capture the entire structure variance, either prior or posterior; but the ratio is valid. Experience rating should reduce this component of variance. Meyers [2] uses the more theoretically grounded "efficiency" standard: the proportion by which the total variance is reduced. Either statistic is useful; the quintiles test is computationally simpler and has an indisputable best value of zero.

Section 3 outlines the variants of the basic plan for which minimal values of the statistics were sought and it discusses some of the rationale for each. The exhibits shown are the final product of a large number of trial-and-error evaluations.

One sidebar test deserves mention. In this test, primary and excess credibilities were evaluated separately. The mod as a function of either primary or excess losses alone had far less predictive accuracy. Credibilities were lower when losses were used separately as the sole basis for the mod than when they were used together.

This conclusion should be contrasted with that of Meyers [2]—that a best modification formula could be based on primary losses only. His conclusion may be correct in the special case of a uniform, well-behaved severity distribution for all risks, which was the model he tested. The NCCI tests of real-world data support the split formula with a two-part credibility.

The workers compensation severity distribution is composed of many types of losses. An essential component of workers compensation ratemaking and individual risk rating is that the distribution of losses by type varies from class to class and risk to risk.

3. PLANS TESTED

A. Basic Specifications of the Former Plan

Potential revised experience rating plans were tested in comparison to the then-current experience rating formula, herein referred to as the "former" plan.

The former formula was derived through practical simplifications that made sense at the time of its development. It was partly these simplifications, however, that moved the plan away from whatever underlying credibility theory it may have had. The former formula is written:

$$M = \frac{A_p + WA_x + (1 - W)\,E_x + B}{E + B}\,. \qquad (3.1)$$

It is one fraction, with weighting value W and credibility ballast B—both linear functions of total expected losses. A denotes actual and E represents expected loss for the experience period. Subscripts p and x denote primary and excess portions of loss, respectively; and E with no subscript denotes total expected losses.

$$W = \begin{cases} 0 & \text{for } E < 25{,}000 \\ \dfrac{E - 25{,}000}{\text{SRP} - 25{,}000} & \text{for } 25{,}000 \le E \le \text{SRP} \\ 1 & \text{for } E > \text{SRP} \end{cases}$$

$$B = 20{,}000\,(1 - W)$$

Here the SRP is the state *Self-Rating Point*, 25 times the state average serious cost per case. This approach provides a nominal indexing to plan credibilities and ratable loss limits that should vary by state and by year.

In the former multi-split formula, the primary portion of a loss L was L_p.

$$L_p = \begin{cases} L & \text{if } L \le 2{,}000 \\ \dfrac{10{,}000L}{8{,}000 + L} & \text{if } L > 2{,}000 \end{cases} \tag{3.2}$$

To calculate the excess portion, losses are limited on a per-claim basis to 10% of the SRP, and on a per-accident basis to 20% of the SRP. Denoting the loss so limited by L_r, the excess portion of a loss greater than \$2,000 would be

$$L_x = L_r - L_p \tag{3.3}$$

where L_p is calculated as noted above.

B. Basic Specifications of the Proposed Plan

Many of the elements of the former plan are retained, including ELRs and D-ratios by class, the primary-excess split formula, and state ratable loss limitations. Payroll (in hundreds) by class is extended by the respective class ELRs to produce the total expected loss. D-ratios, which also vary by class, measure the primary portion of expected loss.

Putting $B = K_p$ and $W = (E + K_p)/(E + K_x)$ into Equation 3.1 results in algebraic equivalence of the new modification formula, Equation 1.10, and the former formula, Equation 3.1. Throughout the testing used to evaluate parameters, the NCCI researchers used Equation 1.10 for the mod, and concentrated on finding best values of K_p and K_x. The values of

K_p and K_x that worked best in all the testing lead to values of W and B quite unlike the former plan's values.

It is highly desirable that differences in benefit levels by state be reflected in the credibility constants K_p and K_x. The former formula used the SRP to effect a nominal difference in the W and B tables by state, but only really affected the risks whose expected losses were near the SRP. We want to use an adjustment that results in a true scaling by state, which would be valid across all risk sizes. That objective is accomplished by inserting a value G, measuring relative benefit levels by state, into the formulae for K_p and K_x. Equation 1.9 is modified to make the following expression for K_p by state:

$$K_p = E\left(\frac{CE + GD}{E + GF}\right). \tag{3.4}$$

A similar change was made to the formula for K_x.

The G-value not only accounts for differences in benefit levels, but also indexes credibility constants for inflation in average claim costs. This property is seen in the following analysis. Assume inflation of $1 + i$ between times t and s. For example, let primary credibility at time t be given by

$$Z(t) = \frac{E}{E + K_p}$$

$$= \frac{E}{E + E\left(\frac{CE + GD}{E + GF}\right)}.$$

With inflation but no real growth, both E and G increase by the factor $1 + i$. This factor cancels everywhere in the formula for Z(s) so that

$$Z(s) = Z(t).$$

The formula for G is one of the parameters that can be varied to optimize the test statistic. In most of the initial NCCI testing, G was taken as a linear function of the existing SRP.

The SRP is retained, but only for use in limitation of ratable losses. There can be no self-rating under any analytic plan, so the SRP is renamed the *State Reference Point.*

C. Tested Plans

One assumption underlying Equation 1.9 for the credibility ballast values K_p and K_x is that both primary and excess credibilities depend on total expected losses, E. The same assumption underlies the former formula, which is Perryman's First Formula. We call the first alternate formula *Perryman I* because it borrows much from the original.

Ultimately, the NCCI researchers tested four alternative plans in addition to the former plan, herein called Current Multi-Split. For each alternative plan, optimal values of the credibility parameters were chosen based on results of the testing. The selection of a final plan from among the four optimized alternatives took into consideration not only the associated values of the test statistic, but also the ease of understanding and implementation.

The tested plans include :

1. Current Multi-Split;
2. Perryman I Multi-Split;
3. Perryman II Multi-Split;
4. Perryman I Single-Split; and
5. Perryman II Single-Split.

Their specifications follow.

1. Current Multi-Split

The basic specifications for this plan have been given. They include the formulae for B, W, the SRP, the primary/excess split of actual losses, and the modification formula itself. They also include calculation of the ELRs and D-ratios by class. The rating values of each insured are included in the experience rating files for each year. In particular, rating years 1981 through 1984 were used in the testing.

2. Perryman I Multi-Split

As described in the introduction, this is the first alternative to the former plan. It is Equation 1.10, with Equation 3.2 used to split actual losses into primary and excess components. Values such as ELRs and D-ratios can be carried over directly from the experience rating files, while K_p and K_x can be calculated easily from the elements of the files: namely, total expected losses of the risk, state identification of the risk (which would be used to fetch indexed SRP and G values), and three coefficients for each formula, selected by trial and error.

3. Perryman II Multi-Split

This formula results from a different assumption about loss variance than the one used in Perryman I. It is only nominally related to Perryman's Second Formula, as noted below.

In the version tested, it is hypothesized that conditional primary loss variance is a function of expected primary losses and that excess loss variance is a similar function of expected excess losses.

The formula for credibilities takes the following form:

$$Z_p = \frac{E_p}{E_p + \tilde{K}_p} \qquad (3.5)$$

where $\tilde{K}_p = \underset{\sim}{C}E + GH/(1 + GF/E)$, and E_p is expected primary losses. Notice that $\tilde{K}_p$ ought to be expressed in terms of E_p, not E. This, however, further complicates the formula. The selection of C, F, and H, as determined by performance, could incorporate average D-ratios, if appropriate, and $\tilde{K}_*$ could be a function of total expected losses. The resulting credibility parameters could be put in tabular form by state according to expected primary or excess losses.

Denoting the average D-ratio by risk as δ results in the following formulae:

$$Z_p = \frac{E_p}{E_p + \tilde{K}_p}$$

$$= \frac{\delta E}{\delta E + \tilde{K}_p}$$

$$= \frac{E}{E + \tilde{K}_p/\delta} \quad . \tag{3.6}$$

Similarly,

$$Z_x = \frac{E_x}{E_x + \tilde{K}_x} \, ,$$

which yields

$$Z_x = \frac{E}{E + \tilde{K}_x/(1 - \delta)} \, . \tag{3.7}$$

Testing of this plan was accomplished using values available from the experience rating files.

For the sake of historical accuracy, the true Perryman's Second Formula actually resulted from the unusual expressions for credibilities

$$Z_p = \frac{E}{E_p + K_p} \, ;$$

$$Z_x = WZ_p \, .$$

Perryman does not derive these expressions and attempts (somewhat less than successfully) to rationalize their contradiction of credibility principles [8].

4. Perryman I Single-Split

One of the key assumptions of the tested formulae is the non-correlation of primary and excess loss components. As long as the primary losses had a severity component, the NCCI researchers were not fully satisfied with a credibility-based plan that uses the former primary-excess split.

It is classically assumed that frequency and severity are independent, hence uncorrelated. This is probably not a valid assumption, but it is reasonable. It is less reasonable to assume that primary and excess losses defined by the multi-split formula are uncorrelated. Thus, the NCCI researchers considered using a modification formula based strictly on frequency and severity. One problem with this idea would be the difficulty of obtaining a valid claim count. (For example, are small medical-only claims recorded on a consistent basis by all carriers for all risks?) Because this change would require the cooperation of so many different interests, it was not pursued.

A compromise is to use a single split (into primary and excess categories) of losses. The portion of a loss below the single threshold value would be primary; and the portion of a loss in excess of that value, if any, would be excess. Using $2,000 as the single-split point is a relatively easy choice: it is the smallest size for which individual claims data is reported, so it is the closest to a frequency/severity dichotomy we can obtain using available data.

To test a single-split plan against actual risk experience, expected losses can be taken directly from experience rating files. New D-ratios corresponding to the new split formula are needed. They are developed by adjusting the multi-split D-ratios in the files to maintain the aggregate adequacy of the D-ratios. For example, if the aggregate emerged actual-primary to ratable-total losses under the former formula had been 0.40, and under the new split it is 0.38, the D-ratios would all be adjusted (downward) by the factor (0.38)/(0.40). With D-ratios so adjusted, the formula is tested with K_p and K_e in the established Equation 1.10, until optimal values for the six coefficients are obtained.

5. Perryman II Single-Split

The last plan to test was the one that utilizes two major variations from Plan 2, Perryman I Multi-Split. Plan 5 uses a single primary-excess split of losses, with the credibility formulae from Plan 3. This is the "fully equipped" model as compared to the other "economy" versions. The question is whether there is enough improvement in performance to justify the additional cost and more difficult handling.

D. Summary

The NCCI researchers tested experience ratings effective on 1980 and 1981 policies. Best parametrizations for each of the plans tested may be seen in Exhibit 2. Of course, "best" is subjective in that no single set of coefficients in any plan produced a lowest value for all 10 evaluations (five size groups and two years). Still, the pattern that emerged for all evaluations was that the smaller sizes deserved more credibility and the larger sizes deserved much less credibility than under the current plan.

Exhibit 3 shows summary statistics and a sample calculation of the test statistic for Size Group Two in 1980.

Several credibilities are displayed in Figures 1, 2, and 3. The consistent pattern for the four optimized plans can be seen. The plans also bear a fairly logical relation to each other. In particular, credibilities seem to increase substantially in the passage from a multi-split to a single-split formula. This may be due to better satisfaction of the assumption that primary and excess losses are uncorrelated.

By contrast, the use of the Perryman II equation in place of Perryman I does not seem to increase average credibilities much. There is, of course, a slight improvement in the distribution of the credibility assigned to the individual risk, as reflected in the test statistic. As described in Section 4, the evaluation of all the plans included weighing the benefit of increased accuracy against the cost of increased complexity in application.

EXHIBIT 2

1. Perryman I Multi-Split

$$K_p = E\left[\frac{0.067\,E + 17{,}200G}{E + 3{,}100\,G}\right] \qquad K_x = E\left[\frac{0.60\,E + 563{,}000\,G}{E + 5{,}000\,G}\right]$$

$$G = \text{SRP}/570{,}000$$

2. Perryman II Multi-Split

$$K_p = E\left[\frac{0.068\,E + 7{,}000\,G}{E + 1{,}600\,G}\right] \qquad K_x = E\left[\frac{0.67\,E + 263{,}700\,G}{E + 5{,}500\,G}\right]$$

$$G = 1$$

3. Perryman I Single-Split *

$$K_p = E\left[\frac{0.10\,E + 2{,}570\,G}{E + 700\,G}\right] \qquad K_x = E\left[\frac{0.75\,E + 203{,}825\,G}{E + 5{,}100\,G}\right]$$

$$G = 0.85 + \text{SRP}/2{,}700{,}000$$

4. Perryman II Single-Split

$$K_p = 0.04\,E + 850\,G \qquad K_x = E\left[\frac{0.60\,E + 98{,}500\,G}{E + 2{,}500\,G}\right]$$

$$G = \text{SRP}/570{,}000$$

* As developed in the text, this is the form of the credibility constants used in the final plan, except that minimum values were established: min $K_p = 7{,}500$, min $K_x = 150{,}000$.

Also, G was changed so that $G = \frac{\text{SACC}}{1{,}000}$, where SACC is the average cost per case by state. At the same time, the SRP was defined as the State Reference Point, SRP = 250 × SACC, so $G = \text{SRP}/250{,}000$.

EXHIBIT 3
PART 1
SAMPLE STATISTICS

	1980					1981				
Formula	Size Group One	Size Group Two	Size Group Three	Size Group Four	Size Group Five	Size Group One	Size Group Two	Size Group Three	Size Group Four	Size Group Five
Former Plan	0.3277	0.2236	0.0918	0.0228	0.0293	0.3230	0.2361	0.1116	0.0453	0.2187
Perryman I Multi-Split	0.1978	0.1248	0.0994	0.0148	0.0012	0.2664	0.1674	0.0930	0.0380	0.0831
Perryman II Multi-Split	0.1632	0.1058	0.0976	0.0112	0.0033	0.1809	0.1333	0.0985	0.0414	0.0980
Perryman I Single-Split	0.0852	0.0519	0.0459	0.0169	0.0042	0.1140	0.0838	0.0688	0.0331	0.0782
Perryman II Single-Split	0.0803	0.0366	0.0380	0.0091	0.0075	0.0785	0.0735	0.0583	0.0312	0.0187

Smaller statistics are more desirable. There were many different samples tested. This was one of the series of tests used to select among the choices.

EXHIBIT 3
Part 2

1980
Actual to Expected Loss Ratios Before and After Experience Rating

15 States Total
Risk Size: $5,000-$10,000

Perryman I Single-Split

Quintile	Before	Squared Deviation From Mean	After	Squared Deviation From Mean
1	0.62	972	0.83	106
2	0.76	295	0.95	2
3	0.86	58	0.95	4
4	1.05	148	1.00	42
5	1.32	1,532	0.92	2
Mean Total:	0.93	3,005	0.93	156

Test Statistic

$$\frac{156}{3{,}005} = 0.0519$$

Current Multi-Split

Quintile	Before	Squared Deviation From Mean	After	Squared Deviation From Mean
1	0.68	623	0.79	192
2	0.70	532	0.78	214
3	0.87	37	0.93	0
4	1.06	173	1.04	112
5	1.30	1,372	1.03	94
Mean Total:	0.93	2,737	0.93	612

Test Statistic

$$\frac{612}{2{,}737} = 0.2236$$

FIGURE 1

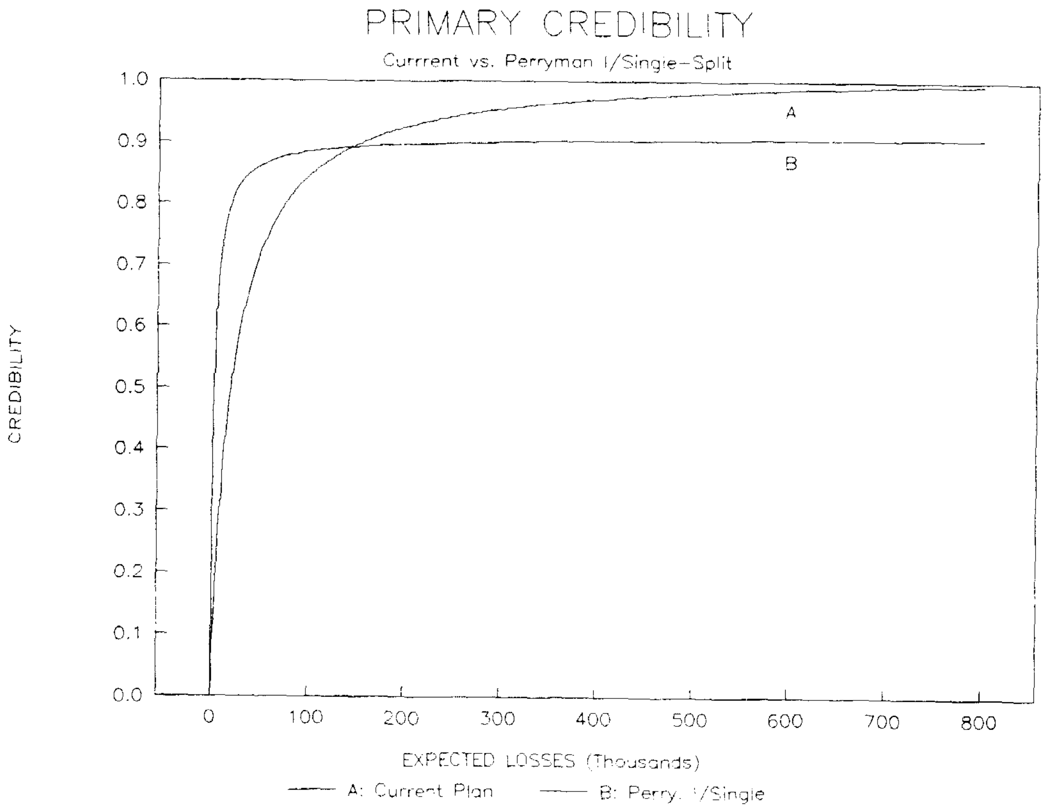

PRIMARY CREDIBILITY
Currrent vs. Perryman I/Single-Split
CREDIBILITY
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
A
B
0
100
200
300
400
500
600
700
800
EXPECTED LOSSES (Thousands)
A: Current Plan
B: Perry. I/Single

FIGURE 2

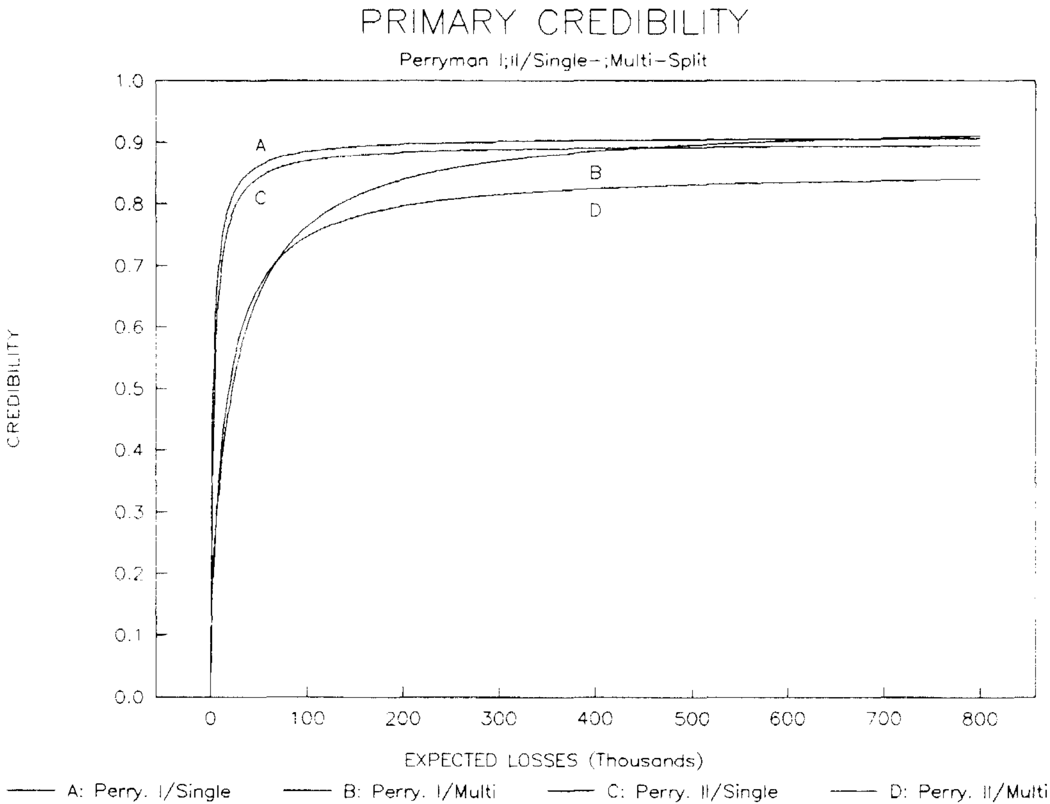
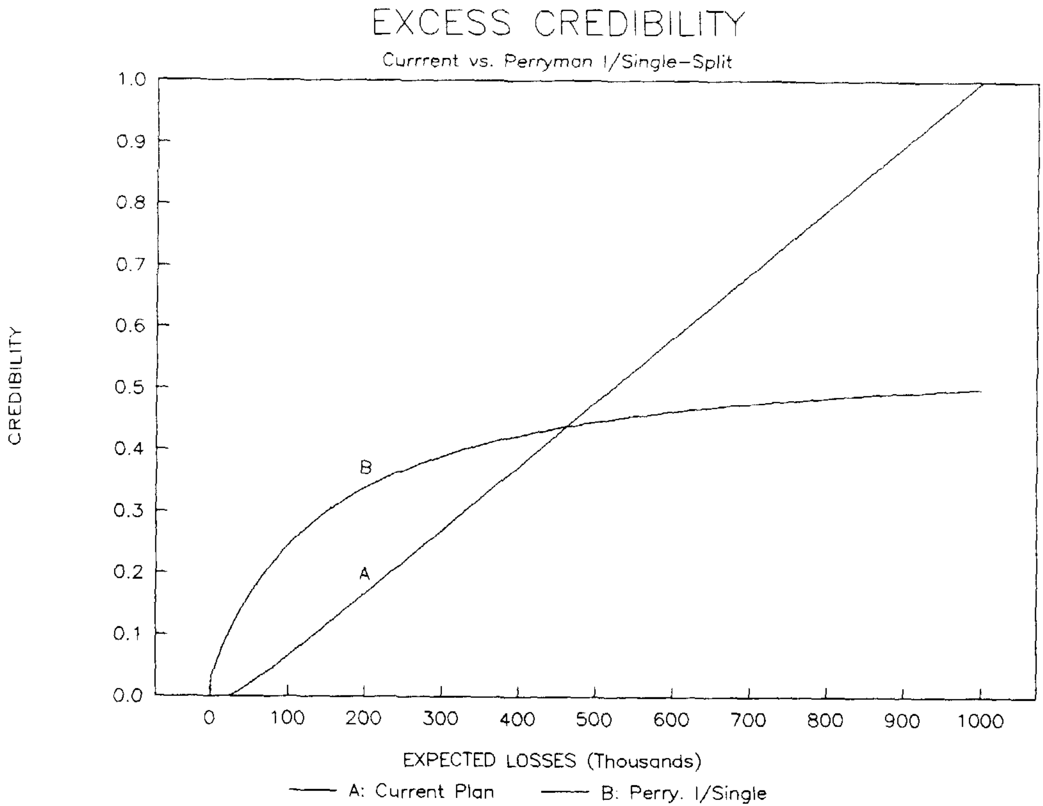

PRIMARY CREDIBILITY
Perryman I;II/Single-;Multi-Split
CREDIBILITY
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
A
B
C
D
0
100
200
300
400
500
600
700
800
EXPECTED LOSSES (Thousands)
A: Perry. I/Single
B: Perry. I/Multi
C: Perry. II/Single
D: Perry. II/Multi

FIGURE 3



EXCESS CREDIBILITY
Currrent vs. Perryman I/Single-Split
CREDIBILITY
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
B
A
0
100
200
300
400
500
600
700
800
900
1000
EXPECTED LOSSES (Thousands)
A: Current Plan
B: Perry. I/Single

4. PARAMETRIZING THE PLAN

A. Selection of Perryman I Single-Split

Selection of an experience rating formula was made primarily on the basis of performance, of course, but also on the basis of practical considerations. Ease of acceptance and implementation were among the considerations. Fortunately, this did not lead to any great compromise of actuarial principles.

The Individual Risk Rating Plans (IRRP) Subcommittee of the NCCI Actuarial Committee approved the Perryman I Single-Split plan as parametrized in Exhibit 2. Its performance was nearly as good as the Perryman II Single-Split plan, but the slight improvement offered by the latter did not appear to outweigh the effort necessary to make the more complex changes. The improvement offered by a single-split over a multi-split plan was significant and the transition would not be difficult.

B. Decreasing Swing

Consider Exhibit 4. It shows the average change in modification, plan to plan, for small risks grouped by value of the former modification. These risks constituted the smallest size group used in testing the original formula, as well as in testing this particular plan. Small risks whose 1985 mod exceeded 1.20 could expect an average increase, or "swing," of 62 points in their mods! Even if this selected plan reflected correctly calibrated credibility, several of the subcommittee representatives thought it would be unacceptable in the market. Some even doubted that it was correct at all, despite the evidence that credibilities were optimal for this size group. Of course, the tests worked on averages and these were extreme cases. Thus, it was possible to believe the tests, yet still believe there was a problem to be fixed.

To address this problem, two changes were made to the plan: the SRP was decreased, as it affected limitations on ratable losses; and minimum values for credibility constants K_p and K_x were established.

It was decided to make the SRP a multiple of average cost per case by state (SACC), rather than a multiple of the average cost per serious case.

EXHIBIT 4

1985

PERRYMAN I SINGLE-SPLIT

CHANGES IN AVERAGE MODIFICATIONS FOR RISKS GROUPED BY VALUE OF CURRENT MOD

$2,500-$5,000 EXPECTED LOSSES DURING EXPERIENCE PERIOD

Range of Current Mods	Number of Risks	Average Current Mod	Average Proposed Mod	Change
0.80-0.84	1	0.83	0.66	-0.17
0.85-0.89	32	0.89	0.70	-0.19
0.90-0.94	9,062	0.93	0.79	-0.14
0.95-0.99	9,600	0.96	0.87	-0.09
1.00	484	1.00	1.02	0.02
1.01-1.05	1,637	1.03	1.13	0.10
1.06-1.10	1,238	1.08	1.31	0.23
1.11-1.15	944	1.13	1.42	0.29
1.16-1.20	742	1.18	1.48	0.30
Over 1.20	2,578	1.35	1.97	0.62
Totals	26,318	1.01	1.03	0.02

Since the average cost per case was between \$750 and \$3,500 for most states in 1981, a multiple of $250 \times$ SACC generally led to smaller SRPs than the typical \$1 million SRPs effective at the time.

In order to test this plan, it was necessary to adjust the ELRs to compensate for the new limits on ratable losses. This was accomplished just like the adjustment of the D-ratios as described in Section 3(C.4).

Several minimum values of K_p and K_x were tested also, but the analysis quickly led to min K_x = 150,000 and min K_p = 7,500, which worked well in conjunction with the new loss limitations in the range above.

Exhibit 5 shows comparisons of the swing in mods for groups of risks by size with a 1986 mod greater than 1.2. (Computed for SRP = 250 or $300 \times$ SACC, and with min K_p = 7,500 and min K_x = 150,000.) In all cases, the swing was less than 25 points.

Changing the SRP formula also led to a re-examination of the calculation of the state scale factor G. The older G formula may be seen in Exhibit 2 as $G = 0.85 + \text{SRP}/2{,}700{,}000$, where the SRP was the value from the former plan. The new formula, resulting from some trial and error, was $G = \text{SACC}/1{,}000$, which worked well with the modified plan.

C. Caps on Modifications

Independent tests of the new plan still showed the potential for large swings in the values of the mod for risks in the smaller size categories. After considerable discussion, the IRRP subcommittee recommended one more change to the rating plan. Rather than tamper with credibility constants, loss limitations, or split points, the subcommittee decided to put absolute caps on the mods of smaller size risks. In this way a debit under the current formula could increase only a limited amount with the change to the new formula.

EXHIBIT 5

1986

PERRYMAN I SINGLE-SPLIT, AS MODIFIED*

CHANGE IN CURRENT MOD TO PROPOSED MOD

RISK HAVING CURRENT MOD GREATER THAN 1.2

SRP	Overall	Size Group One	Size Group Two	Size Group Three	Size Group Four	Size Group Five
$250 \times$ SACC	0.06	0.22	0.21	0.16	0.09	-0.05
$300 \times$ SACC	0.06	0.22	0.21	0.16	0.09	-0.05

* SRP = 250 or $300 \times$ SACC
Minimum $K_p = 7{,}500$ and $K_x = 150{,}000$

EXHIBIT 6

1986

Changes in Average Modifications For Risks Grouped by Value of Current Mod

\$2,500-\$5,000 Expected Losses During Experience Period

Range of Current Mod	Number of Risks	Average Current Mod	Average Proposed Mod	Change
0.80-0.84	18	0.83	0.79	-0.04
0.85-0.89	620	0.88	0.84	-0.04
0.90-0.94	10,951	0.93	0.88	-0.05
0.95-0.99	13,522	0.96	0.93	-0.03
1.00	682	1.00	1.02	0.02
1.01-1.05	2,307	1.03	1.08	0.05
1.06-1.10	1,675	1.08	1.20	0.12
1.11-1.15	1,278	1.13	1.26	0.13
1.16-1.20	1,066	1.18	1.31	0.13
Over 1.20	3,576	1.35	1.49	0.14
Totals	35,695	1.01	1.02	0.01

With minimum $K_p = 7{,}500$ and minimum $K_x = 150{,}000$

SRP = 250 × SACC

G = SACC ÷ 1,000

Mods Limited

The following table lists the limits by size.

Expected Loss Size	Maximum Modification
$0 < E < 5{,}000$	1.6
$5{,}000 \leq E < 10{,}000$	1.8
$10{,}000 \leq E < 15{,}000$	2.0

Exhibit 6 is similar to Exhibit 4 and shows that the swing problem is greatly reduced by this action.

No transition program was designed to phase the maximums out, except the impact of inflation. As experience rating eligibility increased, fewer risks would enjoy a potential 1.60 cap.

Although this action was largely pragmatic, it was not without some actuarial justification. Mods higher than the stated limits are probably not deserved, statistical arguments notwithstanding. The test results only showed that mods were correct for the worst 20% of risks *on average*. Other testing (not shown), using higher percentiles than the 80th, showed the new formula could result in unreasonably high mods, at least for the smallest risks. In addition, risks just below eligibility have a maximum modification of unity. There should be some continuity at the point of eligibility.

Exhibit 7 shows the quintiles test statistics for the finalized plan. These statistics compare reasonably well with the statistics shown in Exhibit 3—Part I.

D. Trending the Split Point

In the matter of a split point, the subcommittee also recommended that appropriate trending be applied to the single-split point used in the testing so that it would have the same relativity to the loss size distribution when actually applied. Trending was based on several years of change in the average cost per case, which led to the single-split point of $5,000 used in the filing. This is a reasonable value, given that it will be well after 1990 before the revised plan is widely accepted, and a few more years after that before any study can be done to revise the point.

EXHIBIT 7

1981

SUMMARY—QUINTILES TEST STATISTICS

SRP	Size Group One	Size Group Two	Size Group Three	Size Group Four	Size Group Five
250 × SACC	0.1800	0.1183	0.0546	0.0127	0.1588
300 × SACC	0.1704	0.1150	0.0645	0.0524	0.3138

SRP = 250 or 300 × SACC

Minimum $K_p = 7{,}500$ and $K_x = 150{,}000$

Still, several researchers thought $5,000 might be too high. After all, the single-split worked well because of its resemblance to a frequency/severity model, and $5,000 may be too high to resemble just claim frequency.

The rationale for such a high selected value was twofold. First, there would be no automatic adjustment, and the $5,000 would be retained until a study could be made to determine the optimal value. Second, the resultant D-ratios for the single-split point would not have to decrease dramatically from the ones in the former plan. Loss size trend would make primary losses increasingly resemble claim count.

It should be noted that a conscious decision was made not to index the split point. As a consequence, it can be expected that average D-ratios will decrease over time. Primary credibilities should be monitored. An initiative to study indexing the split point can be expected in the future.

REFERENCES

[1] Hewitt, Charles C., Jr., "Credibility for Severity," *PCAS* LVII, 1970, p. 148.

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

[3] Mahler, Howard C., Discussion of "An Analysis of Experience Rating," *PCAS* LXXIV, 1987, p. 119.

[4] Venter, Gary G., "Credibility," *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, 1990, Chapter 7.

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

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

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

[8] Perryman, Francis S., "Experience Rating Plan Credibilities," *PCAS* XXIV, 1937, p. 60.