CREDIBILITY WITH SHIFTING RISK PARAMETERS, RISK HETEROGENEITY, AND PARAMETER UNCERTAINTY

HOWARD C. MAHLER

Abstract

This paper explores the important effects on credibility of three phenomena: shifting risk parameters, risk heterogeneity, and parameter uncertainty. When any of these phenomena are significant, the Bühlmann credibility formula no longer applies.

Covariance structures corresponding to these phenomena both separately and in combination are shown. Linear equations for the corresponding credibilities are derived.

Possible applications to classification ratemaking, overall rate indication calculation, and experience rating are illustrated in detail. The procedure for estimating the parameters of the covariance structure is discussed for each situation. Illustrative credibilities are then calculated for each situation.

ACKNOWLEDGEMENT

The author would like to thank Curtis Gary Dean for his many extremely helpful suggestions and corrections. I also wish to thank Dotty Culleton and Silvia Fitcher for typing this paper.

1. INTRODUCTION

In Mahler [1] Markov chains were used to model shifting risk parameters. This model was applied to calculate credibilities in four situations. This paper will expand on that work in a number of important areas.

The phenomena of parameter uncertainty and risk heterogeneity will be incorporated. The behavior of credibility as the size of risk changes will be explored. Possible implications for ratemaking, classification pricing, and experience rating will be discussed.

The three phenomena examined in this paper can be defined as follows:

Shifting Risk Parameters: The parameters defining the risk process for an individual insured are not constant over time. There are (a series of perhaps small) permanent changes to the individual insured's risk process as one looks over several years.

Risk Heterogeneity: An insured is a sum of subunits, and not all of the subunits have the same risk process.

Parameter Uncertainty: There are random fluctuations from year to year in the risk processes of insureds. Parameter uncertainty involves fluctuations that affect most or all insureds somewhat similarly, regardless of size.

Each phenomena can be understood and distinguished in the context of the dice examples to be presented.[1] Insurance examples of each phenomena include:

Shifting Risk Parameters: An automobile insured's risk parameters might shift if a major new road were opened in his locality or if he changed the location to which he commutes to work. Similarly, the automobile experience of a town relative to the rest of the state could shift as that town becomes more densely populated.

Risk Heterogeneity: A workers compensation insured may own several factories that have somewhat different risk characteristics.

---

[1] See Table 3 for a summary of the dice examples.

Parameter Uncertainty: Automobile insureds' risk processes might vary depending on the severity of the winter weather in each year.

The so-called Bühlmann credibility formula is:

$$Z = E/(E+K), \tag{1.1}$$

where E is a measure of size of risk and K is the Bühlmann credibility parameter.

As will be shown, these three phenomena have different effects on the covariance structure between years of data and the resulting credibilities. In the presence of any or all of these three phenomena, the credibility formula in Equation 1.1 does not hold.

Section 2 reviews the results of Mahler [1] relating to shifting risk parameters over time. Section 3 extends the simple dice example from Mahler [1] in order to incorporate parameter uncertainty. Then parameter uncertainty and shifting risk parameters are combined in one model. Section 4 extends the dice example to include risk heterogeneity. Then the model is expanded to include both risk heterogeneity and parameter uncertainty or risk heterogeneity and shifting risk parameters.

In Section 5 the model is expanded to include all three phenomena. The general form of the covariances is given. Section 6 illustrates the calculations of credibilities for various situations. The credibilities for very small risks are discussed. The effect of varying volumes of data by year is discussed. Finally, the case in which no weight is given to the grand mean is discussed.

Section 7 shows how the techniques developed in the prior sections might be applied to the calculation of classification rate relativities. Section 8 extends the results in Section 7 to the use of data from outside the state. Section 9 shows how these techniques might be applied to the calculation of an overall rate indication. Section 10 shows how these techniques might be applied to experience rating. Section 11 covers miscellaneous subjects. Section 12 contains conclusions and a summary.

In order to calculate credibilities there are three steps necessary. First, we must specify the covariance structure between years of data. This structure will vary depending on the phenomena that are important as well as the particular situation.[2] The different covariance structures are listed in Table 1. The general form of the covariance structure is given by Equations 5.10 and 5.11. Second, we must estimate and/or select the parameters appearing in the covariance structure. Finally, we must solve the appropriate set of linear equations for the credibilities. Table 2 lists the different sets of linear equations for the credibilities.

1.1. Bühlmann Credibility[3]

The Bühlmann credibility formula, Equation 1.1, is the least squares credibility corresponding to the following covariance structure between years of data:

$$\mathrm{Cov}[X_i, X_j] = \tau^2 + (\eta^2/E)\delta_{ij}, \tag{1.2}$$

where η^2 is the Expected Value of the Process Variance (for a risk of size 1),

τ^2 is the Variance of the Hypothetical Means,

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

and E is some measure of size of risk. If the Bühlmann credibility parameter is defined as $K = \eta^2/\tau^2$, then Equation 1.2 can be rewritten as:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{1 + (K/E)\delta_{ij}\}. \tag{1.3}$$

---

[2] For example, are we dealing with a single split experience rating plan?

[3] Bühlmann credibility is discussed, for example, in Mayerson [2], Hewitt [3], Hewitt [4], Philbrick [5], Herzog [6], Venter [7], Klugman, Panjer and Willmot [8], and Mahler [9].

TABLE 1

EQUATIONS FOR DIFFERENT COVARIANCE STRUCTURES

Shifting Risk Parameters	Risk Heterogeneity	Parameter Uncertainty	Varying Sizes of Risk	$\mathrm{Cov}[X_i, X_j]$		Credibility Assigned to Y Years of Data, Each of Size E[1]
No	No	No	No	$\tau^2\{1 + K\delta_{ij}\}$	Bühlmann	$Y/(Y+K)$
No	No	No	Yes	$\tau^2\{1 + (K/E)\delta_{ij}\}$	(1.2)	$EY/(EY+K)$
Yes	No	No	No	$\tau^2\{\lambda^{\lvert i-j\rvert} + K\delta_{ij}\}$	(3.15)	Not Applicable[2]
Yes	No	No	Yes	$\tau^2\{\lambda^{\lvert i-j\rvert} + (K/E)\delta_{ij}\}$	(3.16)	Not Applicable[2]
No	No	Yes	Yes	$\tau^2\{1 + ((K/E)+J)\delta_{ij}\}$	(3.5)	$EY/(E(Y+J)+K)$
Yes	No	Yes	Yes	$\tau^2\{\lambda^{\lvert i-j\rvert} + ((K/E)+J)\delta_{ij}\}$	(3.20)	Not Applicable[2]
No	Yes	No	Yes	$r^2\{1 + (I/E) + (K/E)\delta_{ij}\}$	(4.3)	$(E+I)Y/(EY+YI+K)$
No	Yes	Yes	Yes	$r^2\{1 + (I/E) + ((K/E)+J)\delta_{ij}\}$	(4.13)	$(E+I)Y/(E(Y+J)+YI+K)$
Yes	Yes	No	Yes	$r^2\{\rho^{\lvert i-j\rvert} + (I/E)\gamma^{\lvert i-j\rvert} + (K/E)\delta_{ij}\}$	(4.34)	Not Applicable[2]
Yes	Yes	Yes	Yes	$r^2\{\rho^{\lvert i-j\rvert} + (I/E)\gamma^{\lvert i-j\rvert} + ((K/E)+J)\delta_{ij}\}$	(5.5)	Not Applicable[2]

[1] In those cases where there are not different sizes of risk, take $E = 1$.
[2] Credibilities are the result of solving a set of linear equations, as listed in Table 2. They do not have a simple algebraic form.

In the presence of shifting risk parameters, risk heterogeneity, parameter uncertainty, and varying sizes of risk, the general covariance structure is given by Equations 5.10 and 5.11:

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{\lvert i-j\rvert} + \gamma^{\lvert i-j\rvert} I/\sqrt{E_i E_j} + \delta_{ij}\left(K/\sqrt{E_i E_j} + J\right)\right\}, \qquad \sqrt{E_i E_j} \geq \Omega \tag{5.10}$$

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{\lvert i-j\rvert} + \gamma^{\lvert i-j\rvert} I/\Omega + \delta_{ij}\left(K/\sqrt{E_i E_j} + J\right)\right\}, \qquad \sqrt{E_i E_j} \leq \Omega. \tag{5.11}$$

TABLE 2

LINEAR EQUATIONS TO SOLVE FOR CREDIBILITIES

Situation	
Y years of data X_i being used to predict Year $Y+\Delta$. Weight to the overall mean.	$\sum_{i=1}^{Y} \mathrm{Cov}[X_i, X_k] Z_i = \mathrm{Cov}[X_k, X_{Y+\Delta}]$, $k = 1,2,\ldots Y$ (2.4)
Y years of data X_i being used to predict Year $Y+\Delta$. No weight to the overall mean.	$\sum_{i=1}^{Y} \mathrm{Cov}[X_i, X_k] Z_i = \mathrm{Cov}[X_k, X_{Y+\Delta}] + \lambda/2$, $k = 1,2,\ldots Y$; $\sum_{i=1}^{Y} Z_i = 1$ (6.7)
Y years of classification data, both from within and outside the state, being used to predict classification relativities for Year $Y+\Delta$. No weight to the overall mean. S_{ij} = covariances within the state. T_{ij} = covariances outside the state. U_{ij} = covariances between state and outside the state.	$\sum_j Z_j S_{ij} + \sum_j W_j U_{ij} = \frac{\lambda}{2} + S_{i,Y+\Delta}$, $i = 1,2,\ldots Y$; $\sum_i Z_i U_{ij} + \sum_i W_i T_{ij} = \frac{\lambda}{2} + U_{Y+\Delta,j}$, $j = 1,2,\ldots Y$; $\sum_i Z_i + \sum_j W_j = 1$ (8.1)
Y years of experience rating data, primary and excess, being used to predict Year $Y+\Delta$. S_{ij} = covariances of primary losses. T_{ij} = covariances of excess losses. U_{ij} = covariances between primary and excess losses.	$\sum_{i=1}^{Y} (Z_{Pi} S_{ik} + Z_{Xi} U_{ki}) = S_{k,Y+\Delta} + U_{k,Y+\Delta}$, $k = 1,2,\ldots Y$ (10.12); $\sum_{i=1}^{Y} (Z_{Pi} U_{ik} + Z_{Xi} T_{ki}) = U_{Y+\Delta,k} + T_{k,Y+\Delta}$, $k = 1,2,\ldots Y$ (10.13)

In those situations where size of risk is not important, Equation 1.3 could be rewritten by setting $E = 1$:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{1 + K\,\delta_{ij}\}. \tag{1.4}$$

For Y years of data each of size E, the covariance structure given by Equation 1.3 corresponds to a Bühlmann/least squares credibility assigned to these Y years of data of:[4]

$$Z = \frac{EY}{EY + K}. \tag{1.5}$$

As displayed in Table 1, in the presence of any or all of the three phenomena discussed above, the simple covariance structure of Equation 1.3 and the simple credibility formula of Equation 1.5 no longer apply.

2. SHIFTING RISK PARAMETERS

The parameters defining the risk process for an individual insured are not constant over time. For example, for automobile insurance the expected claims frequency of an insured compared to the average changes over time. Mahler [1] presents a Markov chain model of shifting risk parameters which quantifies the effects of shifts over time in the risk process of an insured via the covariances between years of data.

2.1. Covariances, Shifting Risk Parameters

For this Markov chain model, in most cases the covariances can be approximated by:[5]

$$\mathrm{Cov}[X_i, X_j] = \tau^2 \lambda^{|i-j|} + \delta_{ij}\eta^2, \tag{2.1}$$

where

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

η^2 is the Expected Value of the Process Variance,

τ^2 is the Variance of the Hypothetical Means,

---

[4]See Section 3.1 for an example of the calculation of Bühlmann credibility. The Bühlmann credibility is calculated as $\mathrm{Cov}[X_i, X_j]/\mathrm{Cov}[X_i, X_j]$. This is the ratio of the variance of the hypothetical means to the expected value of the process variance (each for Y years of data each of size E).

[5]This is Equation 7.1 in Mahler [1].

and λ is the dominant eigenvalue (other than unity) of the transpose of the transition matrix of the Markov chain.

X has different meanings depending on the application. X_i could be the claim frequency for an insured in year i, the loss ratio for a state in year i, the relativity for a class in year i, the die roll in trial i, etc.

From Equation 2.1,

$$\mathrm{Var}[X] = \mathrm{Cov}[X,X] = \tau^2 + \eta^2, \qquad \text{and}$$

$$\text{Total Variance} = \text{VHM} + \text{EPV},$$

the usual relationship that the total variance can be split into the Variance of the Hypothetical Means and the Expected Value of the Process Variance.

As the separation between years of data increases, the (expected) covariance and correlation between years declines.

For example, if $\tau^2 = \text{VHM} = 1{,}000$, $\eta^2 = \text{EPV} = 5{,}000$, and $\lambda = .9$, then the variance-covariance matrix given by Equation 2.1 for four consecutive years of data would be:

6,000	900	810	729
900	6,000	900	810
810	900	6,000	900
729	810	900	6,000

This contrasts with the situation in the absence of shifting risk parameters; if $\lambda = 1$, then the variance-covariance matrix has entries of 6,000 along the diagonal and 1,000 off the diagonal. With no shifting risk parameters, Equation 2.1 reduces to the usual Bühlmann covariance structure $\mathrm{Cov}[X_i, X_j] = \tau^2 + \delta_{ij}\eta^2$.

2.2. *Rate of Shifting Risk Parameters*

It is not vital to understand the precise derivation of λ; rather it is important to understand that λ quantifies the rate at which the parameters shift. The smaller λ is, the faster the parameters shift. The closer λ is to unity, the slower the parameters shift. In the limit for $\lambda = 1$, there is no shifting of parameters.

The "half-life" is a useful way to quantify the rate of shifting parameters. The half-life is defined as the length of time necessary for the correlations between years to have declined by a factor of one-half:

$$\begin{aligned} \lambda^{\text{half-life}} &= .5, \\ \text{half-life} &= \frac{\ln .5}{\ln \lambda} = \frac{-.693}{\ln \lambda}. \end{aligned} \tag{2.2}$$

The longer the half-life, the slower the rate of shifting parameters over time.

2.3. Correlations Between Years of Data, Shifting Risk Parameters

If the Markov chain model holds, the correlations between different years of data should decline approximately exponentially. For $i \neq j$, Equation 2.1 gives $\mathrm{Cov}[X_i, X_j] = \tau^2 \lambda^{|i-j|}$.

Thus, as the distance between years grows, the expected covariance between the data from those years declines. Another feature of the Markov chain model is that even though the risk parameters of individuals vary over time, the overall portfolio of insureds looks (relatively) stable from year to year. Specifically, Equation 2.1 gives the same variance for each year of data, $\mathrm{Var}[X_i] = \mathrm{Var}[X_j] = \tau^2 + \eta^2$.

Therefore, the correlations between different years of data are:

$$\begin{aligned} \mathrm{Corr}[X_i, X_j] &= \left(\frac{\tau^2}{\tau^2 + \eta^2}\right) \lambda^{|i-j|}, \qquad \text{and} \\ \ln \mathrm{Corr}[X_i, X_j] &= \ln\left(\frac{\tau^2}{\tau^2 + \eta^2}\right) + |i - j| \ln \lambda, \qquad i \neq j. \end{aligned} \tag{2.3}$$

Therefore, if the Markov chain model holds, the log-correlations for years separated by a given amount should decline approximately linearly. The slope of this line is (approximately) $\ln\lambda$. The intercept is approximately

$$\ln\left(\frac{\tau^2}{\tau^2+\eta^2}\right).$$

Note that $\tau^2/(\tau^2+\eta^2)$ = VHM/Total Variance = credibility in the absence of shifting risk parameters.

Thus given a data set, we can determine whether this (simple) Markov chain model might be appropriate. We determine whether the log-correlations as a function of the separation between years (not including zero separation) can be approximated by a straight line.[6] Then we can estimate the parameter λ and the ratio $\tau^2/(\tau^2+\eta^2)$ from the slope and intercept of the fitted straight line.

2.4. *Credibilities, Shifting Risk Parameters*

These estimates can be used in turn to estimate credibilities. If we have data X_i from years $1,2,\ldots Y$ and are estimating year $Y+\Delta$, then the least squares credibilities Z_i to be assigned to individual years of data are found by solving the Y linear equations in Y unknowns:[7]

$$\sum_{i=1}^{Y}\mathrm{Cov}[X_i,X_k]Z_i = \mathrm{Cov}[X_k,X_{Y+\Delta}], \qquad k=1,2,\ldots Y. \tag{2.4}$$

3. PARAMETER UNCERTAINTY

Parameter uncertainty and its effect on credibilities is discussed in Meyers [10], Mahler [11] and Mahler [12]. Random

[6]In many cases there is a large amount of random fluctuation so even if the expected log-correlations are precisely along a straight line, the log-correlations estimated from the data will vary widely around a straight line. See Figure 10 in Mahler [1].

[7]See Equations 2.8 in Mahler [1].

fluctuations occur from year to year in the risk processes of insureds. Parameter uncertainty involves fluctuations that affect most or all insureds somewhat similarly, regardless of size.

While the distinction between parameter uncertainty and shifting risk parameters is not always clear-cut, parameter uncertainty tends to involve fluctuations not related to the insured while shifting risk parameters tend to involve (a series of perhaps small) permanent changes to the individual insured's risk process. For example, shifting risk parameters would occur if a workers compensation insured implemented a new safety program.

An example of parameter uncertainty occurs in workers compensation insurance, where the level of losses is affected by economic events that affect even very large employers. This creates a potential random fluctuation in the loss potential above and beyond what we normally think of as the process variance. The important feature is that while the large size of an employer reduces the impact of the random fluctuations inherent in observed accidents per year, it either does not reduce or only partially reduces the impact of (seemingly) random changes in the overall economy.

There is a kernel of uncertainty in the frequency of workers compensation claims that will not be reduced by observing more workers during a single year. In these circumstances, the credibilities as a function of the size of risk E will not be of the Bühlmann form $E/(E+K)$.

The covariance structure in the presence of parameter uncertainty is somewhat more complicated, as shown in Equations 3.4 and 3.5. When both parameter uncertainty and shifting risk parameters are present, the covariance structure, as shown in Equations 3.19 and 3.20, contains a combination of the features of each phenomenon separately. These covariance structures will be developed in the context of the simple dice example from

TABLE 3

VARIOUS DICE EXAMPLES

Shifting Risk Parameters	Risk Heterogeneity	Parameter Uncertainty	Section(s)	People	Different Colors of Dice
No	No	No	3.1	Joe	No
No	No	Yes	3.2, 3.4	Joe, Mary	No
Yes	No	No	3.5	Joe, Beth	No
Yes	No	Yes	3.6	Joe, Mary, Beth	No
No	Yes	No	4.1	Joe	Yes
No	Yes	Yes	4.4	Joe, Mary	Yes
Yes	Yes	No	4.9	Joe, Rose, Gwen	Yes
Yes	Yes	Yes	5.1	Joe, Mary, Rose, Gwen	Yes

Joe initially selects either N identical dice in the cases without different colors of dice, or N identical red dice and N possibly different green dice.

Mary flips a coin prior to each trial (year).

Beth, prior to each trial, may alter all the dice from one type to another. (For example, 6-sided dice could be switched to 4-sided dice.)

Rose, prior to each trial, may alter the type of all the red dice.

Gwen, prior to each trial, may alter the type of one or more of the green dice; Gwen acts independently on each green die.

Mahler [1]. Table 3 summarizes the various examples that will be presented.

Section 3.1 will present this simple dice example. Section 3.2 will expand on the dice example in order to incorporate parameter uncertainty. Section 3.3 will discuss how this example relates to parameter uncertainty in general. Section 3.4 will expand the example to observing several years of data. Section 3.5 will introduce shifting risk parameters into the example, in the absence of parameter uncertainty. Section 3.6 extends the example to include both parameter uncertainty and shifting risk parameters. Section 3.7 compares the credibilities corresponding to the various covariance structures discussed. Many readers may find it helpful to go directly to this graphical comparison of results.

3.1. Simple Dice Example, No Shifting Risk Parameters, No Parameter Uncertainty

Assume Joe selects N dice of the same type and rolls them. Assume Joe selected either four-sided,[8] six-sided[9] or eight-sided dice[10] with a priori probabilities of 25%, 50%, and 25%, respectively. Joe tells you how many dice he rolled and the resulting sum, but you do not know which type of dice Joe selected. Joe will roll the same dice again.

The process variances for 4, 6, and 8-sided dice are respectively 1.25, 2.92, and 5.25. Therefore, the expected value of the process variance (for one die) is (25%)(1.25) + (50%)(2.92) + (25%)(5.25) = 3.08. The means for 4, 6, and 8-sided dice are respectively 2.5, 3.5, and 4.5. Therefore, the a priori overall mean is (25%)(2.5) + (50%)(3.5) + (25%)(4.5) = 3.5. The variance of the hypothetical means is .500.

In this case, the Bühlmann credibility for estimating the sum of the next roll of the dice can be written as:

$$Z = \frac{N}{N+K} \tag{3.1}$$

where

$$K = \frac{\text{Expected Value of the Process Variance (for } N = 1)}{\text{Variance of the Hypothetical Means (for } N = 1)}$$

$$= \frac{\eta^2}{\tau^2} = \frac{3.08}{.5} = 6.16.$$

The credibility Z is to be applied to the data (the sum of Joe's dice), while the complement of credibility $1 - Z$ is to be applied to the a priori grand mean of 3.5.

[8]With numbers 1, 2, 3, and 4 on the faces.
[9]With numbers 1 through 6 on the faces.
[10]With numbers 1 through 8 on the faces.

3.2. Parameter Uncertainty, Dice Example

Assume that in a modification of the previous example, Mary flips a single coin[11] and adds the result to each of Joe's N die rolls.[12] Each head adds $\frac{1}{2}$ to the result of a die, while each tail subtracts $\frac{1}{2}$ from the result of a die. You are again told the result of the combination of Joe's and Mary's actions but see neither the coin nor the dice.

While the addition of a coin flip does not change any of the means, the overall risk process has changed. The amount of credibility we would assign to a single observation has also changed. As will be shown, there is a fundamental change in the behavior of the credibility as a function of N, the number of dice per roll.

The expected value of the process variance is the sum of the expected value of the process variances from Joe's and Mary's actions, since these processes are independent. The expected value of the process variance of Mary's actions is $.25N^2$ since we multiply the result of a single coin flip by N and since $\mathrm{Var}[NX] = N^2\,\mathrm{Var}[X]$. Thus, since the EPV for Joe's action is $3.08N$, the overall expected value of the process variance is $3.08N + .25N^2$.

The hypothetical means have not been changed by the introduction of the coin flips. Therefore, the variance of the hypothetical means remains $.5N^2$.

This covariance structure can be written as:

$$\mathrm{Cov}[X_i, X_j] = .5N^2 + (3.08N + .25N^2)\,\delta_{ij}. \tag{3.2}$$

Equation 3.2 can be rewritten for more general situations than this specific dice example. It will be useful to substitute E, representing some measure of size of risk such as expected losses, for N, the number of dice that Joe rolls in this specific example.

---

[11] For simplicity, we assume the coins are fair, with equal probability of heads or tails.

[12] Equivalently, one could add N times the result of the single coin flip to the sum of the die rolls.

If η^2 = expected value of the process variance = 3.08, u^2 = variance due to parameter uncertainty = 0.25, τ^2 = variance of the hypothetical means = 0.5, and E is a measure of size of risk, then Equation 3.2 can be rewritten as

$$\mathrm{Cov}[X_i, X_j] = \tau^2 E^2 + (\eta^2 E + u^2 E^2)\delta_{ij}. \qquad (3.3)$$

Suppose that, instead of the sum of the dice, one were estimating the average per die rolled, in a manner analogous to claim frequency, claim severity or pure premium. Then, since the quantity of interest is divided by E, all the variances and covariances in Equation 3.3 are divided by E^2:

$$\mathrm{Cov}[X_i, X_j] = \tau^2 + (\eta^2/E + u^2)\delta_{ij}. \qquad (3.4)$$

Letting $J = u^2/\tau^2$ and $K = \eta^2/\tau^2$, Equation 3.4 can be rewritten as:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{1 + ((K/E) + J)\delta_{ij}\}. \qquad (3.5)$$

Equations 3.4 and 3.5 are the covariances in the presence of parameter uncertainty. A new parameter J has been introduced in addition to Bühlmann's K.

The credibility is the variance of the hypothetical means for N dice divided by the sum of the variance of the hypothetical means for N dice and the expected value of the process variance for N dice:

$$Z = \frac{.5N^2}{3.08N + .25N^2 + .5N^2} = \frac{N}{1.5N + 6.16}. \qquad (3.6)$$

With $J = u^2/\tau^2 = .25/.5 = .5$ and $K = \eta^2/\tau^2 = 3.08/.5 = 6.16$, this is of the form:[13]

$$Z = \frac{N}{(1+J)N + K}, \qquad J > 0 \quad \text{and} \quad K > 0. \qquad (3.7)$$

The form of the credibility as a function of size is fundamentally different. As $N \to \infty$, $Z \to 1/(1+J) < 1$. Therefore, no

[13]The notation in Meyers [10], Mahler [11] and Mahler [12] has been changed so that J there is called $1 + J$ here. As will be seen, this cosmetic difference makes it easier to write the formulas involving more than one year of data.

matter how many dice Joe rolls, the credibility assigned to the observation stays less than $1/(1+J)$ or $1/1.5 = 67\%$ in this case. The fact that Joe is rolling more and more dice cannot eliminate the noise added by Mary's single coin flip, which is added to each and every die, and thus cannot increase the credibility beyond 67%.[14]

This is an example of the phenomenon of parameter uncertainty. We can think of this risk process as Joe selects (at random) which type of dice to roll and then Mary's coin flip alters the parameters of the risk process. If for example, Joe selects 6-sided dice, then prior to Mary's coin flip we are uncertain whether this time the expected value of Joe's roll is $3N$ or $4N$. Once Mary flips her coin, if it is tails, the expected value of Joe's roll is $3N$ (after subtracting $.5N$) and if it is heads, the expected value of Joe's roll is $4N$ (after adding $.5N$). The variance of this parameter uncertainty is $.25N^2$.

The value of J which quantifies the impact of parameter uncertainty in the credibility formula was:

$$J = .25/.5 = .5 = \frac{\text{variance due to parameter uncertainty}}{\text{variance of the hypothetical means}} = \frac{u^2}{\tau^2}. \tag{3.8}$$

The larger the J, the greater the impact of parameter uncertainty.

3.3. *Parameter Uncertainty in General*[15]

When parameter uncertainty is important, the within class variance will have two pieces. The "good" piece increases as N and is the expected value of the process variance in the absence of parameter uncertainty. The "bad" piece increases as N^2 and is the variance introduced by parameter uncertainty. Unlike

---

[14] If instead Mary had flipped N coins, one for each die rolled by Joe, then the credibilities would not have behaved in this manner. Instead they would have followed the usual Bühlmann formula, in this case, $Z = N/(N + 6.66)$. The Bühlmann credibility parameter would have been $6.16 + .5 = 6.66$.

[15] See Meyers [10] and Mahler [11].

the good piece, the bad piece increases as quickly as the variance between classes, which also increases as N^2. Thus taking many observations (in a single year) will not get rid of the effect of parameter uncertainty.

This effect is assumed to be due to the different possible states of the universe. Taking more observations will not get rid of the variation inherent in the universe.

In the simple example, Mary's single coin flip represented this random variation in the universe from year to year. In the case of workers compensation insurance, changes in the economy affect the relative costs of claims. These changes can affect firms with 1,000 workers as much as those with 100 workers. Such changes are therefore expected to affect the risk process in a manner similar to Mary's single coin flip (although there is a continuous spectrum of possible states of the economy).

If parameter uncertainty has an important impact on workers compensation insurance, one would expect the credibility to be of the form of Equation 3.7:

$$\frac{E}{(1+J)E+K}, \qquad J>0, \quad K>0,$$

where E represents the size of risk. This is one of the refinements introduced in the NCCI's Revised Experience Rating Plan.[16]

3.4. *Dice Example, Several Years of Data*

The dice example with parameter uncertainty will be extended to the situation in which more than one year of data is observed.

Assume Joe selects N dice of a given type and rolls them in each of Y years, while Mary flips a separate coin each year. Then the expected value of the process variance is Y times what it was for a single year: $3.08NY + .25N^2Y$. The variance of the

[16] See Gillam [13], and Mahler [12].

hypothetical means is Y^2 times what it was for a single year: $.5N^2Y^2$. Therefore, the credibility is:

$$
\begin{aligned}
Z &= \frac{.5N^2Y^2}{.5N^2Y^2 + 3.08NY + .25N^2Y} \\
Z &= \frac{NY}{NY + .5N + 6.16} \\
Z &= \frac{NY}{N(Y + .5) + 6.16}.
\end{aligned}
\tag{3.9}
$$

In general, for size of risk E and number of years Y Equation 3.9 can be written as:

$$Z = \frac{EY}{E(Y+J)+K}, \qquad J > 0, \quad K > 0. \tag{3.10}$$

The credibility has a different form. In the presence of parameter uncertainty, the accumulation of Y separate years does not enter into the formula in the same way as size of risk E. There is the "extra" term involving E, where E is multiplied by J, which is the ratio of the variance due to parameter uncertainty divided by the variance of the hypothetical mean. For one year of data Equation 3.10 reduces to Equation 3.7, the previous result for parameter uncertainty in a single year, which for this example is $Z = E/(1.5E + 6.16)$.

For any fixed number of years, Z has the form $E/$(Linear in E), although the values of the coefficients depend on Y. For fixed size of the insured E, the formula reduces to the usual Bühlmann formula in terms of Y, the number of years. For fixed E as $Y \rightarrow \infty$, $Z \rightarrow 1$. Increasing the number of years of observations overcomes the impact of parameter uncertainty. We can in fact average over the different assumed random states of the universe in each year by averaging over time.

Observing a fleet of 100 cars for 10 years is *not* the same as observing a similar fleet of 1,000 cars for a single year. In the latter case, we cannot average out those aspects peculiar to that one individual year. For example, a gasoline shortage due to an oil embargo or a severe winter might produce unusual results in an individual year regardless of the size of the fleet.

In summary, in the presence of parameter uncertainty, one must carefully distinguish between size of risk and number of years of data.

3.5. *Shifting Parameters Over Time, Dice Example*

Shifting risk parameters over time were discussed in Section 2. In Mahler [1], shifting risk parameters were introduced into the simple dice example in Section 3.1 by altering the risk process as follows:

Joe selects a die and rolls it. Then prior to the next trial, Beth may at random replace that die with another die. Assume Beth's replacement process works such that:

1. A four-sided die will be replaced 20% of the time by a six-sided die.[17]
2. A six-sided die will be replaced 10% of the time by a four-sided die and 15% of the time by an eight-sided die.
3. An eight-sided die will be replaced 30% of the time by a six-sided die.

Then the process repeats: Joe rolls a die and Beth (possibly) replaces the die.

Beth's risk process is just a simple example of a Markov chain. See Appendix A for a discussion of Markov chains. There are three "states": 4-sided die, 6-sided die, and 8-sided die. For each trial there is a new, possibly different, state. The probability of being in a state depends only on the state for the previous trial.

[17]The remaining 80% of the time the die is left alone.

Beth's Markov chain is completely described by the "transition probabilities" between the states.

Generally, the transition probabilities for a Markov chain are arranged in a matrix P. For Beth's "risk process," the matrix of transition probabilities is:

	Four	Six	Eight
Four	.80	.20	0
Six	.10	.75	.15
Eight	0	.30	.70

As shown in Mahler [1], in this case, the covariances between years of data are given by:

$$\mathrm{Cov}[X_i, X_j] = (.468)(.769)^{|i-j|} + (.032)(.481)^{|i-j|} + \delta_{ij}(3.08). \tag{3.11}$$

In general, for years of data X_i and X_j:

$$\mathrm{Cov}[X_i, X_j] = \sum_{k>1} \zeta_k \lambda_k^{|i-j|} + \delta_{ij}\eta^2, \tag{3.12}$$

where η^2 is the Expected Value of the Process Variance, $\delta_{ij} = 0$ for $i \neq j$ and 1 for $i = j$, λ_k are the eigenvalues of the transpose of the transition matrix and the ζ_k are a function of the transition matrix P and the means of the states.[18] In general,

$$\sum_{k>1} \zeta_k = \tau^2 = \text{variance of the hypothetical means.}$$

Equation 3.11 can be approximated by:

$$\mathrm{Cov}[X_i, X_j] \approx (.5)(.769)^{|i-j|} + \delta_{ij}(3.08). \tag{3.13}$$

Equation 3.13 can be written in general as:

$$\mathrm{Cov}[X_i, X_j] \approx \tau^2 \lambda^{|i-j|} + \delta_{ij}\eta^2. \tag{3.14}$$

[18] See Mahler [1].

where λ is the dominant eigenvalue of the transpose of the transition matrix (other than unity), τ^2 is the variance of the hypothetical means, and η^2 is the expected value of the process variance. Taking as before $K = \eta^2/\tau^2$, Equation 3.14 could be rewritten as:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{\lambda^{|i-j|} + K\delta_{ij}\}. \tag{3.15}$$

For a size of risk E, Equation 3.15 becomes:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{\lambda^{|i-j|} + (K/E)\delta_{ij}\}. \tag{3.16}$$

One could then use Equation 2.4 to solve linear equations for the credibilities.

In the absence of shifting risk parameters $\lambda = 1$ and Equation 3.16 becomes the usual Bühlmann covariance structure, Equation 1.3:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{1 + (K/E)\delta_{ij}\}.$$

3.6. Combining Parameter Uncertainty and Shifting Risk Parameters

Let us now combine the models of parameter uncertainty and shifting risk parameters. Assume that Joe selects N dice (of the same kind) and rolls them. Mary then flips a coin and adds the result ($+1/2$ if heads and $-1/2$ if tails) to the result of each die. The sum is the result of one trial or year. After each trial, Beth (possibly) changes the type of all N dice, with transition matrix P.

Beth does not affect the variance of a single year. As discussed previously in the example involving just Joe and Mary, the total variance of a year of data for this example is $(3.08N + .25N^2) + .5N^2 = 3.08N + .75N^2$.

The covariances between different years are what they were in the absence of Mary, because Mary's action in one year is independent of her action in another year.

Therefore, the covariances of the years of data are for this example:[19]

$$\mathrm{Cov}[X_i, X_j] = \{(.468)(.769^{|i-j|}) + (.032)(.481^{|i-j|})\}N^2 + \{.25N^2 + 3.08N\}\delta_{ij}. \tag{3.17}$$

In general, where E is a measure of the size of the insured:

$$\mathrm{Cov}[X_i, X_j] = \left\{\sum_{k>1} \zeta_k \lambda_k^{|i-j|}\right\} E^2 + \delta_{ij}\{\eta^2 E + u^2 E^2\}. \tag{3.18}$$

where η^2 = Expected Value of the Process Variance and u^2 = variance due to parameter uncertainty. Given Y years of data, we can solve Y linear equations in Y unknowns, Equations 2.4, for the credibilities to be assigned to each year of data. Note that the solution is the same if we divide all of the variances and covariances by E^2:

$$\begin{aligned}\mathrm{Cov}[X_i, X_j]/E^2 = {} & \sum_{k>1} \zeta_k \lambda_k^{|i-j|} \\ & + \delta_{ij}\{(\text{Variance Due to Parameter Uncertainty}) \\ & \quad + ((\text{Expected Value of the Process Variance})/E)\}.\end{aligned}$$

This isolates the effect of the size of risk E. As will be discussed subsequently, this is the form that will apply in insurance applications where one is estimating claim frequency rather than total number of claims, pure premiums rather than total losses, etc.

As was done previously, the covariances can be approximated in terms of λ, the dominant eigenvalue of the transpose of the transition matrix (other than unity). For claim frequency, pure premiums, etc., the covariances in the presence of parameter un-

---

[19] Note that for $i = j$, $\mathrm{Cov}[X_i, X_j] = \mathrm{Var}[X_i] = (.468 + .032)N^2 + .25N^2 + 3.08N = 3.08N + .75N^2$, as stated above.

certainty and shifting risk parameters are then approximately

$$\mathrm{Cov}[X_i, X_j] \approx \tau^2 \lambda^{|i-j|} + (\eta^2/E + u^2)\delta_{ij}. \qquad (3.19)$$

Taking as before $J = u^2/\tau^2$ and $K = \eta^2/\tau^2$, Equation 3.19 can be rewritten as:

$$\mathrm{Cov}[X_i, X_j] = \tau^2\{\lambda^{|i-j|} + (K/E + J)\delta_{ij}\}. \qquad (3.20)$$

In the absence of shifting risk parameters over time, $\lambda = 1$ and Equation 3.20 would reduce to Equation 3.5. In the absence of parameter uncertainty, $J = 0$ and Equation 3.20 would reduce to Equation 3.16. In the absence of both phenomena Equation 3.20 would reduce to the usual Bühlmann covariance structure. These covariance structures are compared in Table 1.

3.7. Graphical Comparison of Results

Assuming the covariances given by Equation 3.20, we can solve Equation 2.4 for the corresponding credibilities. This has been done for the dice example, which had parameters $J = .5$, $K = 6.16$, and $\lambda = .769$.

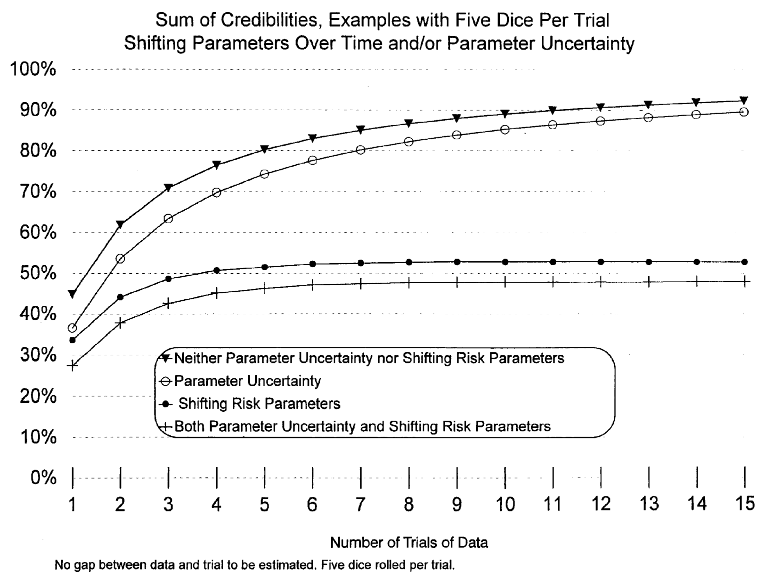
Figure 1 compares the behavior of the credibilities with and without parameter uncertainty as well as with and without shifting risk parameters over time, for five dice per year.[20] In general, both phenomena reduce the credibility assigned to the data by introducing additional noise to the results.

In this particular case with five dice, it so happens that each phenomenon individually results in roughly the same credibility being assigned to a single year of data.[21] Yet we see a radically different behavior as the number of years increases. With just parameter uncertainty, in the limit the effect of parameter uncertainty vanishes; the sum of the credibilities approaches unity.

[20] In Equation 3.20, $E = 5$.

[21] The relative importance of parameter uncertainty increases as the number of dice increases. In this case $Z = Y/(Y + .5 + 6.16/N)$ for Y years and N dice with parameter uncertainty but no shifting risk parameters.

FIGURE 1



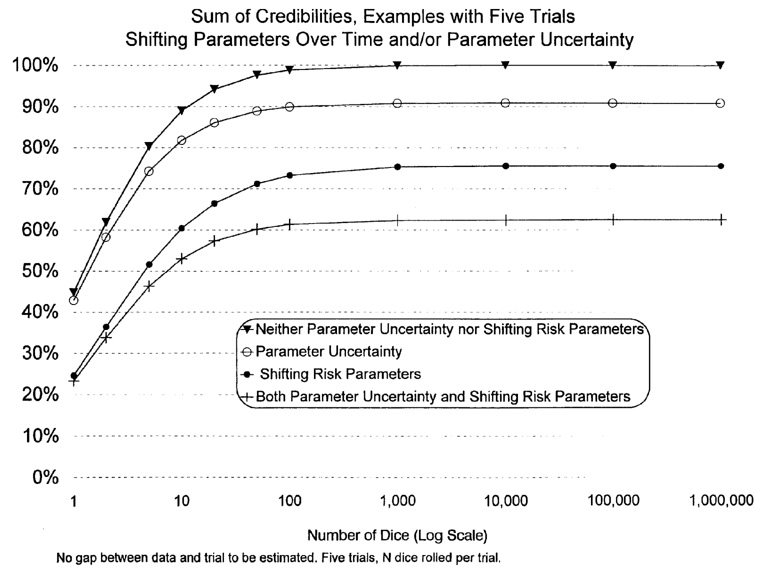
With just shifting risk parameters over time, the sum of the credibilities approaches a limit strictly less than unity.[22]

The credibilities in the presence of both phenomena are lower than those with only one of the phenomena. These credibilities approach an even lower limit as the number of years approaches infinity than when we had solely shifting risk parameters.[23] While similar behavior would be expected in general, the details will depend on the amount of parameter uncertainty and the speed at which the parameters shift.

Figure 2 compares for 5 years of data the dependence of the sum of the credibilities on the number of dice per year with the presence or absence of the two phenomena. As expected, with no shifting or parameter uncertainty, we get the usual Bühlmann

[22]In this example, the sum of the credibilities approaches .528.
[23]In this case, with both phenomena present, this limit is .480.

FIGURE 2



credibility, which goes to unity as the number of dice approaches infinity.[24] With parameter uncertainty, the credibilities are somewhat less. Also, as the number of dice approaches infinity, the credibility approaches a limit less than unity.[25]

With shifting risk parameters over time, the credibilities are less than in the absence of shifting risk parameters. As seen in Figure 2, as the number of dice approaches infinity, the credibilities approach a value less than unity.[26] With both phenomena present, the credibilities are lower.[27]

---

[24]In this case, $Z = 5/(5 + 6.16/N)$ for the sum of the credibilities for 5 years.

[25]In this case, $Z = 5/(5.5 + 6.16/N)$ which approaches $1/1.1 = 90.9\%$ as N approaches infinity. Using 5 years of data, one cannot get rid of the effects of parameter uncertainty (although it has less effect than if one relied on fewer than 5 years).

[26]In this case, the limit is .755.

[27]As the number of dice $N \to \infty$, the sum of the credibilities in this case approaches the limit .625.

4. RISK HETEROGENEITY

The phenomenon of risk heterogeneity and its effect on credibilities was discussed in Mahler [11] and Mahler [12]. As stated in Hewitt [14], "For loss ratio distribution purposes—two \$50,000 risks don't make a \$100,000 risk. Nor is a \$100,000 risk for one year the same as a \$50,000 risk for two years." Risk heterogeneity involves an insured which is a sum of individual subunits, where not all the subunits have the same risk process.

Assume we have a large workers' compensation insured. It might consist of several locations or several factories. It is reasonable to assume that the factories making up this insured will be affected by some of the same efforts of management. Therefore, if one factory has better than average expected losses for its mix of classifications, it is likely that another factory that is part of the same insured will have better than average expected losses.

Thus, the combined experience of the different factories has higher credibility for experience rating than the experience of a single factory. However, since the factories also differ in some ways, the larger risk is to some extent heterogeneous. The credibility will not increase as quickly as if the factories were identical; the credibilities are not of the form: $Z = E/(E + K)$.

In general, subunits are combined into one overall insured.[28] If the subunits of the overall insured have the same risk process,[29] then we have the familiar Bühlmann assumptions as in the simple dice example. If on the other hand the subunits of the overall insured are selected at random from the total available population, then there is no increase in the experience rating

[28]The term "subunit" is intended to be vague. It is intended to convey the general concept rather than a particular situation.

[29]"Risk process" refers to the random process that generates the observed quantity of interest. So in the dice example, it would be determined by the number of sides of the dice being rolled. In a Poisson frequency example it would be determined by the average frequency.

credibility of the overall risk compared to its subunits. If the subunits are more similar to each other on average than the total available population, then there is some increase in the experience rating credibility as risk size increases, but not as quickly as in the Bühlmann case where $Z = E/(E + K)$.

As with the prior phenomena, the behavior in the presence of risk heterogeneity will be demonstrated via the simple dice example from Mahler [1]; the example in Section 3.1 will be expanded upon in Section 4.1 in order to incorporate risk heterogeneity.

Section 4.2 discusses risk heterogeneity in general. Equation 4.3 is the corresponding covariance structure. Section 4.3 discusses a refinement to this covariance structure for very small risks.

In Section 4.4, the phenomena of parameter uncertainty and risk heterogeneity are combined in the dice example. Equation 4.13 is the corresponding covariance structure for insurance applications. Section 4.5 gives formulas for credibility in the absence of shifting risk parameters. Section 4.6 discusses a refinement for very small risks to the covariance structure with risk heterogeneity and parameter uncertainty. Sections 4.7 and 4.8 illustrate how this refinement might be applied to workers compensation experience rating.

In Section 4.9, the phenomena of risk heterogeneity and shifting risk parameters are combined in the dice example. Equation 4.34 is the corresponding covariance structure for insurance applications. Section 4.10 discusses the behavior with size of risk for this covariance structure for risk heterogeneity and shifting risk parameters.

4.1. Risk Heterogeneity, Dice Example

As before Joe selects dice, either four-sided, six-sided or eight-sided dice. However, he selects N red dice all of one type and N green dice of possibly different types. Then Joe rolls the dice and tells you the result:

$(1-h)$ (the sum of N red dice) + (h) (the sum of N green dice), where h is a known parameter $0 \le h \le 1$.

Assume Joe selected the type of red dice as either four-sided, six-sided, or eight-sided, with a priori probabilities of 25%, 50%, and 25%, respectively. All N of the red dice are of the same type.

Joe independently selected the type of each green die as either four-sided, six-sided, or eight-sided, with a priori probabilities[30] of 25%, 50%, and 25%. The N green dice will usually be a mixture of the three types.

The important feature that distinguishes this example from the prior examples is the different manner in which the green dice are selected compared to the red dice. The N red dice are identical, while the N green dice are a random mixture.

Thus, the green and red dice contribute differently to the variance of the hypothetical means. For a single die with means of 2.5, 3.5 or 4.5 selected with probabilities 25%, 50% and 25%, the variance of the hypothetical means is 0.5. For N identical dice each hypothetical mean is multiplied by N, so the variance of the hypothetical means for the sum of the N red dice is $.5N^2$. For N randomly selected dice the variances add. For the sum of the N green dice the variance of the hypothetical means is $.5N$.

Since the green and red dice are chosen independently of each other, the variance of the hypothetical means for $(1-h)$ (N red dice) + h (N green dice) is:

$$(1-h)^2(.5N^2) + h^2(.5N) = .5N^2(1-h)^2 + .5Nh^2.$$

This is the key effect of risk heterogeneity: the variance of the hypothetical means increases more slowly than the square of the risk size.

[30]The a priori probabilities for the green dice and red dice were selected to be equal solely for simplicity of illustration. This is not an essential feature.

Therefore, one feature of risk heterogeneity is that there is less variation between larger risks than between smaller risks. Specifically, for the dice example, the coefficient of variation[31] of the hypothetical means is $(\sqrt{.5}/3.5)\sqrt{(1-h)^2 + h^2/N}$, which decreases to a positive constant as N increases.

Here, as in Mahler [11] and Mahler [12], the variance of the hypothetical means increases in a combination of a linear and a quadratic term. The question of how quickly the variance of the hypothetical means increases goes back to the origins of workers compensation experience rating.[32] While a rate of increase between linear and quadratic was indicated, the assumption of a quadratic increase was used for practical reasons. This led to the now famous formula for credibility, $Z = E/(E+K)$, which was used for experience rating workers compensation, as discussed in Whitney [15] and Michelbacher [16].

The expected value of the process variance for a single die is 3.08. For the sum of N green dice or N red dice, the expected value of the process variance is $N(3.08)$, since the die rolls are independent. The expected value of the process variance for $(1-h)$ (N red dice) + h (N green dice) is: $(1-h)^2N(3.08) + h^2N(3.08)$.

This model might have some applicability to large commercial insureds. For example, assume a commercial automobile fleet involves N drivers. There are many features such as driver selection, driver training, vehicle maintenance, use of vehicle, etc., that are likely to cause the N drivers' risk processes to be more similar than those of the general population of drivers for similarly classified fleets. On the other hand, the N drivers are unlikely to each have the exact same risk process.

In the dice example, each driver's result could be taken as $(1-h)$ (roll of a red die) + h (roll of a green die). Then the red

[31] The coefficient of variation is the standard deviation divided by the mean. The overall mean in the dice example is 3.5.

[32] Whitney [15, p. 287] states that the variance of the hypothetical means seemed to increase as $P^{5/4}$, where P was the loss pure premium.

die captures that part of the risk process that is similar across the particular fleet[33] while the green die captures those aspects that mirror the variation across the total classification to which this fleet belongs. The smaller h, the more similar the drivers' risk processes across the fleet, and the smaller the impact of risk heterogeneity.

The credibility is:

$$\begin{aligned} Z &= \mathrm{VHM}/(\mathrm{VHM}+\mathrm{EPV}) \\ &= \frac{.5N^2(1-h)^2 + .5Nh^2}{.5N^2(1-h)^2 + .5Nh^2 + 3.08N(1-h)^2 + 3.08Nh^2} \\ &= \frac{N + \left(\dfrac{h}{1-h}\right)^2}{N + \left(\dfrac{h}{1-h}\right)^2 + 6.16 + 6.16\left(\dfrac{h}{1-h}\right)^2}. \end{aligned} \tag{4.1}$$

If $h = 0$, then $Z = N/(N + 6.16)$, the familiar Bühlmann result in the absence of risk heterogeneity, as in Equation 3.1.

If $h = 1$, then $Z = 1/(1 + 6.16) = 14\%$, the Bühlmann credibility for a single die. If the subunits are chosen totally at random, ($h = 1$), then there is no increase in credibility with size of risk.

Let $I = h^2/(1-h)^2$ while $K = 6.16$, the usual Bühlmann credibility parameter in this case. Then we can rewrite Equation 4.1 as:

$$Z = \frac{N + I}{N + I + K + IK}. \tag{4.2}$$

Equation 4.2 is of the same general form as given in Mahler [11] and Mahler [12].[34] The additional parameter I is zero in the absence of risk heterogeneity. In the presence of risk heterogeneity $I > 0$, and the credibility is of the form: (size + constant)/(size + different constant).

---

[33] While the red dice are identical, the outcomes of the rolls are independent. They represent the same risk process, *not* the same outcome of that risk process.

[34] However, the definition of the parameters is not precisely the same.

While there are some specific assumptions that could be altered,[35] this is one reasonable model which captures the key effect of risk heterogeneity; the Variance of Hypothetical Means has a piece which increases more slowly than N^2 does.

4.2. Risk Heterogeneity in General

The key impact of risk heterogeneity in general is that the covariance between years of claim counts, losses, etc. increases more slowly than the square of the size of risk. Put another way, the covariance between years of claim frequency, pure premiums, etc. decreases with the size of risk. Here, as in Mahler [11] and Mahler [12], the assumption will be made of a covariance structure in the presence of risk heterogeneity of:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + I/E + (K/E)\delta_{ij}\}, \qquad I, K \geq 0. \quad (4.3)$$

Between different years, Equation 4.3 gives a covariance of $r^2\{1 + I/E\}$, which has one term independent of size of risk and one term that declines as one over the size of risk. If $I = 0$, there is no risk heterogeneity, and Equation 4.3 reduces to the usual Bühlmann covariance structure.

In Equation 4.3, the Variance of the Hypothetical Mean frequencies, pure premiums, etc. is $r^2\{1 + I/E\}$. Assuming the mean claim frequency, pure premium, etc. is (largely) independent of the risk size E, then the coefficient of variation of the hypothetical means declines as E increases. As measured by the coefficient of variation of the hypothetical means, larger insureds are more similar to each other than smaller insureds are to each other. Larger insureds are likely to be a sum of somewhat dissimilar subunits; if we added up enough randomly selected subunits, then we would approach the overall average. Thus with risk heterogeneity, in some sense insureds get closer to average as they get very large.

[35]For example, h, the parameter that quantifies the heterogeneity, was assumed to not depend on N.

For one year of data, substituting the covariance structure given by Equation 4.3 into Equation 2.4 gives the following equation for the credibility:[36]

$$(1 + I/E + K/E)Z = 1 + I/E.$$

Thus, the credibility is of the form:

$$(\text{size} + \text{constant})/(\text{size} + \text{different constant}),$$

$$Z = \frac{E + I}{E + I + K}. \tag{4.4}$$

With three years of data, all of size E, Equations 2.4 become the following 3 linear equations in three unknowns:

$$(1 + I/E + K/E)Z_1 + (1 + I/E)Z_2 + (1 + I/E)Z_3 = 1 + I/E,$$

$$(1 + I/E)Z_1 + (1 + I/E + K/E)Z_2 + (1 + I/E)Z_3 = 1 + I/E, \quad \text{and}$$

$$(1 + I/E)Z_1 + (1 + I/E)Z_2 + (1 + I/E + K/E)Z_3 = 1 + I/E.$$

This has solution:

$$Z_1 = Z_2 = Z_3 = \frac{E + I}{3E + 3I + K}.$$

If we let Z be the sum of these three credibilities,

$$Z = Z_1 + Z_2 + Z_3 = \frac{(E + I)3}{E3 + 3I + K}.$$

If instead of 3 years of data we have Y years of data, all of size E, then the sum of the credibilities obtained by solving Equations 2.4 is:

$$Z = \frac{(E + I)Y}{EY + YI + K}. \tag{4.5}$$

Equation 4.3 for the covariance structure and Equation 4.5 for the credibility have the same general behavior as in the dice

[36]The factors of r^2 on each side of the equation cancel out, and have no effect on the credibility.

example with risk heterogeneity, although the parameters are somewhat different. Equations 4.3 and 4.5 are in the form that will later be applied to insurance examples. Also, the covariance structure in Equation 4.3 will form the basis for the covariance structure when other phenomena besides risk heterogeneity are present.

4.3. Very Small Risks and Risk Heterogeneity

For the phenomena of risk heterogeneity we will now introduce a refinement for very small sizes of risk. In the dice example in Section 4.1, risk heterogeneity only applies for risks above a certain size, those with more than one die.

Similarly, in insurance examples we might expect that the effects of risk heterogeneity will apply only above a certain size. For commercial automobile insurance, this might be when there is more than one vehicle or more than five vehicles. For workers compensation insurance, this minimum size might be more than one worker, more than a dozen workers, or more than one location. In general, below a certain size, we might expect that there are no subunits which are being grouped and, therefore, no risk heterogeneity. In any case, we will assume there is some minimum size, Ω, which depends on the particular application, below which the phenomena of risk heterogeneity does not apply.

Then for sizes of risk less than Ω, Equation 4.5 will not give the appropriate credibility. It will give too much credibility to the very smallest risks; as $E \rightarrow 0$ in Equation 4.5, $Z \rightarrow (I/(I + K/Y)) > 0$.

In practical applications we can apply special caps to the effect of credibility for small risks.[37] In the NCCI Revised Experience Rating Plan for workers compensation insurance, there are caps

[37] In general one should cap the effects of credibility. See for example Mahler [17] and Mahler [18].

on the maximum debit for small risks.[38] In addition, below a certain size risks are not eligible for experience rating.[39]

It is worthwhile to explore the expected behavior of the credibilities for very small risks. For experience rating one may devise a simplified merit rating plan to apply to smaller risks. For classification rating one must assign the data of every class a credibility, no matter how small the volume of data.

We will assume a covariance structure and derive a formula for the credibilities that apply for risks of the smallest sizes. Equation 4.3 is assumed to be valid for risks of size $\geq \Omega$

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + I/E + (K/E)\delta_{ij}\}, \qquad E \geq \Omega. \quad (4.6)$$

For $E = \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + I/\Omega + (K/\Omega)\delta_{ij}\}.$$

We assume that for $E < \Omega$, the term related to risk heterogeneity, I/Ω, does not decline as the risk size declines below Ω, and thus acts as if the risk was homogeneous.[40] In other words:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + I/\Omega + (K/E)\delta_{ij}\}, \qquad E \leq \Omega. \quad (4.7)$$

Thus, for risks of size less than Ω, the Variance of the Hypothetical Means is $r^2 + r^2 I/\Omega$, independent of size. This is the type of behavior we expect in the absence of risk heterogeneity.[41] While the dice example was useful for developing the ideas in

---

[38]See Mahler [12]. Recently the maximum debit has been revised. It is now given via a continuous formula for all sizes: $1 + (.00005)[E + 2E/g]$, where g is NCCI's state specific parameter.

[39]The minimum is based on premiums and varies by state. For example, for Massachusetts it is currently $5,500 in annual premium.

[40]The term related to risk homogeneity, r^2, is independent of the size of risk, and thus below Ω remains the same.

[41]Although r^2 can be thought of as the piece of the VHM which is related to risk homogeneity, the VHM for small risks is assumed to be $r^2 + r^2 I/\Omega$. If one desired, one could reparametrize the covariances setting $\tau^2 = r^2 + r^2 I/\Omega$ and then use τ^2 rather than r^2. However, such a reparametrization would not in and of itself alter the credibilities.

this paper, it has its limitations. In the dice example N can never be less than one.

Using the covariance structure given by Equation 4.6, for $E \geq \Omega$ the credibilities are given by Equation 4.5

$$Z = \frac{Y(E+I)}{YE+YI+K}, \qquad E \geq \Omega. \tag{4.8}$$

However, for $E \leq \Omega$, the covariances are given by Equation 4.7, and the solution to Equations 2.4 is, in the absence of shifting risk parameters and parameter uncertainty:

$$\begin{aligned} Z &= \frac{Y(1+I/\Omega)}{Y\{1+(I/\Omega)\}+(K/E)} \\ Z &= \frac{YE}{YE+K'}, \qquad E \leq \Omega \end{aligned} \tag{4.9}$$

where $K' = K(\Omega/I+\Omega)$.

Equation 4.9 is of the same form as the Bühlmann credibility formula, but with the parameter K adjusted by a factor of $\Omega/(I+\Omega)$.

The credibilities given by Equation 4.9 approach zero as the risk size approaches zero. As expected, for $E = \Omega$, Equations 4.8 and 4.9 give the same credibility:

$$\begin{aligned} Z &= \frac{Y(\Omega+I)}{(Y)\Omega+YI+K} = \frac{Y(\Omega+I)}{Y(\Omega+I)+K} \\ &= \frac{Y}{Y+K\left(\dfrac{\Omega}{\Omega+I}\right)\left(\dfrac{1}{\Omega}\right)} \\ &= \frac{Y}{Y+K'/\Omega} = \frac{Y\Omega}{Y\Omega+K'}. \end{aligned}$$

Equations 4.8 and 4.9 together combine the usual Bühlmann credibility formula for small risks with that applicable in the presence of risk heterogeneity for large risks.

4.4. Risk Heterogeneity and Parameter Uncertainty, Dice Example

The dice models of risk heterogeneity and parameter uncertainty can be easily combined. Joe picks N identical red dice and N randomly selected green dice as in Section 4.1, and Mary flips a coin as in Section 3.2. Then the result is:

$$(1-h)(\text{Sum of } N \text{ Red Dice}) + h(\text{Sum of } N \text{ Green Dice}) + N(\text{Coin Flip}),$$

where the coin flip is counted as $-\frac{1}{2}$ if tails and $+\frac{1}{2}$ if heads.

Then, per Sections 3.2 and 4.1, the Expected Value of the Process Variance is the sum of Joe and Mary's individual process variances:

$$(3.08)(1-h)^2N + (3.08)h^2N + .25N^2.$$

The presence of the coin flips has not altered the hypothetical means. Therefore, according to Section 4.1, the variance of the hypothetical means is:

$$.5N^2(1-h)^2 + .5Nh^2.$$

The EPV and VHM can be combined into the covariance structure:

$$\text{Cov}[X_i, X_j] = .5N^2(1-h)^2 + .5Nh^2 + \{(3.08)(1-h)^2N + (3.08)h^2N + .25N^2\}\delta_{ij}. \tag{4.10}$$

The credibility is:

$$\begin{aligned} Z &= \text{VHM}/(\text{VHM} + \text{EPV}) \\ &= \frac{.5N^2(1-h)^2 + .5Nh^2}{.5N^2(1-h)^2 + .5Nh^2 + 3.08N(1-h)^2 + 3.08Nh^2 + .25N^2} \\ &= \frac{N + \left(\dfrac{h}{1-h}\right)^2}{N + \left(\dfrac{h}{1-h}\right)^2 + 6.16 + 6.16\left(\dfrac{h}{1-h}\right)^2 + \dfrac{.5N}{(1-h)^2}}. \end{aligned} \tag{4.11}$$

As before let $I = h^2/(1-h)^2$ while $K = 6.16$, the usual Bühlmann credibility parameter. Let $J = .5/(1-h)^2$, which for $h = 0$ reduces to the situation in Section 3.2 where J was .5. Then Equation 4.11 can be rewritten as:

$$Z = \frac{N + I}{N(1+J) + I + K + IK}. \tag{4.12}$$

4.5. *Credibilities, No Shifting Risk Parameters*

For insurance applications to frequency, pure premiums, etc., it will be useful to rewrite the covariance structure in Equation 4.10 with a somewhat different parametrization than in the dice example. Combining the features of Equations 3.5 and 4.3, the covariance structure with risk heterogeneity and parameter uncertainty is:[42]

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + (I/E) + ((K/E) + J)\delta_{ij}\}, \qquad I, J, K \geq 0. \tag{4.13}$$

When one uses Y years of data to predict a future year, Equations 2.4 become with the covariances from Equation 4.13:

$$(K/E + J)Z_i + \sum_{j=1}^{Y}(1 + I/E)Z_j = 1 + I/E, \qquad i = 1, 2, \ldots, Y.$$

By symmetry the credibilities for the individual years, Z_i, are all equal.

Let $Z_i = Z/Y$, where $Z = \sum Z_i$, the total credibility applied to the data.[43] Then the sum of the credibilities for Y years of data

[42]Where as before E is the size of risk, I quantifies risk heterogeneity, J quantifies parameter uncertainty, and K is the Bühlmann credibility parameter. The size of risk enters as $1/E$ since we are estimating quantities such as frequency or pure premiums rather than the sum of die rolls, the total number of claims, the total losses.

[43]The credibility applied to each year is in this case the total credibility divided by the number of years.

each of size E is:

$$Z = \frac{(E+I)Y}{E(Y+J)+YI+K}; \qquad I,J,K \geq 0. \tag{4.14}$$

For one year of data, $Y = 1$, Equation 4.14 becomes:[44]

$$Z = \frac{E+I}{E(1+J)+I+K}. \tag{4.15}$$

While Equation 4.14 with $Y = 1$ differs slightly from Equation 4.12, they have the same essential form as a function of size of risk.

Equation 4.14 is of the same general form as given in Mahler [11] and Mahler [12].[45] In the absence of parameter uncertainty, J=0 and the credibility is given by Equation 4.5. In the presence of parameter uncertainty, $J > 0$. In the absence of risk heterogeneity $I = 0$ and the credibility is given by Equation 3.10. In the presence of risk heterogeneity, $I > 0$.

The parameter I largely affects the credibilities for smaller risks. The parameter J largely affects the credibilities for larger risks. The maximum credibility as the size of risk approaches infinity is $Y/(Y+J) < 1$. The credibility is of the form: (linear function of size)/(linear function of size).

Equation 4.14 for the credibility in the presence of risk heterogeneity and parameter uncertainty is the form used in the NCCI Revised Experience Rating Plan for workers compensation. The primary and excess credibilities depend on a state specific parameter g as follows:[46]

$$\begin{aligned} Z_p &= (E' + 700g)/(1.1E' + 3270g), \qquad \text{and} \\ Z_x &= (E' + 5{,}100g)/(1.75E' + 208{,}925g), \end{aligned} \tag{4.16}$$

---

[44]This is the same general form of the credibilities in the presence of risk heterogeneity and parameter uncertainty, shown in Mahler [12]. This is the same basic form as Equation 4.4 of Mahler [12], with a slightly different treatment of the parameters I and K.

[45]However, the definition of the parameters is not precisely the same.

[46]See Mahler [12]. The parameter g is the average cost per case divided by 1,000; g is rounded to the nearest 0.05. Recently the NCCI has revised the excess parameters

where E' is the expected losses for the sum of 3 years of data. E' is the equivalent of $3E = YE$ in Equation 5.9. Equations 4.16 are the same as Equation 4.14 with $Y = 3$ and the parameters:[47]

	Primary	Excess
I	$700g/3$	$1{,}700g$
J	.3	2.25
K	$2{,}570g$	$203{,}825g$

(4.17)

Note that as $E \to 0$ in Equation 4.14, $Z \to YI/(YI + K)$. Thus the minimum credibility is $1/(1 + (K/IY))$. This is greater than zero for $I > 0$. For the NCCI Revised Experience Rating Plan the minimum primary credibility is $1/(1 + (2{,}570/700)) = 21.4\%$. The minimum excess credibility is $1/(1 + (203{,}825/5{,}100)) = 2.4\%$.

As $E \to \infty$ in Equation 4.14, $Z \to (YE)/(Y + J)E = Y/(Y + J) = 1/(1 + J/Y)$. This is less than 1 for $J > 0$. For the NCCI Revised Experience Plan, the maximum primary credibility is $1/(1 + .3/3) = 1/1.1 = 90.9\%$. The maximum excess credibility is $1/1.75 = 57.1\%$.

Without parameter uncertainty, $J = 0$ and Equation 4.14 becomes:

$$Z = \frac{(E + I)Y}{EY + YI + K} = \frac{E + I}{E + I + K/Y}. \tag{4.18}$$

somewhat to take effect during 1998 and later. $J_x = 1.125$ rather than 2.25. $K_x = 150{,}000g$ rather than $203{,}825g$. In addition, only 30% of Medical Only losses will be included in experience rating.

[47]This differs from the values shown in Mahler [12] due to the somewhat different treatment of the parameters here. The important point is that the credibilities are of the form Linear/Linear. The Revised Experience Rating Plan was developed under the direction of Gary Venter while he was at the National Council on Compensation Insurance. As described in Gillam [13], this was the form of credibilities that worked well in the tests performed by the NCCI. Note that while in Section 10 of the current paper explicit recognition of the impact of the covariance of primary and excess losses is taken, this was not the case in the derivation of the credibilities in the NCCI Revised Experience Rating Plan.

For one year of data (and only risk heterogeneity) Equation 4.18 becomes:

$$Z = \frac{E+I}{E+I+K}. \tag{4.19}$$

Without risk heterogeneity $I = 0$, and Equation 4.14 becomes:

$$Z = \frac{YE}{E(Y+J)+K}. \tag{4.20}$$

For one year of data (and only parameter uncertainty) Equation 4.19 becomes:

$$Z = \frac{E}{E(1+J)+K}. \tag{4.21}$$

4.6. Very Small Risks, Risk Heterogeneity and Parameter Uncertainty

As in Section 4.3, we will introduce a refinement for very small sizes of risk. In the dice example, risk heterogeneity applies only for risks above a certain size, those with more than one die. Similarly, in insurance examples we might expect that the effects of risk heterogeneity will apply only above a certain size.

We will assume a covariance structure and derive a formula for the credibilities that apply for risks of the smallest sizes. Equation 4.13 is assumed to be valid for risks of size $\geq \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + (I/E) + ((K/E) + J)\delta_{ij}\}, \qquad E \geq \Omega. \tag{4.22}$$

For $E = \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + (I/\Omega) + ((K/\Omega) + J)\delta_{ij}\}.$$

We assume that for $E < \Omega$, the term related to risk heterogeneity, I/Ω, does not decline as the risk size declines below Ω, and thus acts as if the risk were homogeneous.[48] In other

[48]The term related to risk homogeneity, r^2, is independent of the size of risk, and thus below Ω remains the same.

words:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + (I/\Omega) + ((K/E) + J)\delta_{ij}\}, \qquad E \leq \Omega. \tag{4.23}$$

Using the covariance structure given by Equation 4.22, for $E \geq \Omega$ the credibilities are given by Equation 4.14:

$$Z = \frac{Y(E+I)}{(Y+J)E + YI + K}, \qquad E \geq \Omega. \tag{4.24}$$

However, for $E \leq \Omega$, the covariances are given by Equation 4.23, and the solution to Equations 2.4 is, in the absence of shifting risk parameters:

$$\begin{aligned} Z &= \frac{Y(1 + (I/\Omega))}{Y(1 + (I/\Omega)) + (K/E) + J} \\ &= \frac{YE((I+\Omega)/\Omega)}{YE((I+\Omega)/\Omega) + JE + K} \\ &= \frac{YE}{(Y+J')E + K'}, \qquad E \leq \Omega \end{aligned} \tag{4.25}$$

where

$$J' = J\left(\frac{\Omega}{I+\Omega}\right) \qquad \text{and} \qquad K' = K\left(\frac{\Omega}{I+\Omega}\right).$$

Equation 4.25 is of the same form as Equation 4.20, but with the parameters J and K each adjusted by a factor of $\Omega/(I+\Omega)$. This is the Bühlmann credibility formula with an additional parameter J' to account for parameter uncertainty. For very small risks, the parameter J' has very little effect; thus Equation 4.25 gives approximately the same result as the usual Bühlmann credibility formula.

The credibilities given by Equation 4.25 approach zero as the risk size approaches zero. As expected, for $E = \Omega$, Equations 4.24 and 4.25 give the same credibility:

$$
\begin{aligned}
Z &= \frac{Y(\Omega + I)}{(Y+J)\Omega + YI + K} = \frac{Y(\Omega + I)}{Y(\Omega + I) + J\Omega + K} \\
&= \frac{Y}{Y + J\left(\dfrac{\Omega}{\Omega + I}\right) + K\left(\dfrac{\Omega}{\Omega + I}\right)\left(\dfrac{1}{\Omega}\right)} \\
&= \frac{Y}{Y + J' + K'/\Omega} = \frac{Y\Omega}{(Y+J')\Omega + K'}.
\end{aligned}
$$

4.7. Very Small Risks, Workers Compensation Experience Rating

For example, consider the NCCI Revised Experience Rating Plan with parameters given in Equations 4.17 and $Y = 3$. Take solely for illustrative purposes $\Omega = \$1{,}000g$. If $g = 2$, corresponding to an average claim size of \$2,000, then $\Omega = \$2{,}000$. This would correspond to \$6,000 in expected losses[49] over 3 years. Assuming the expected loss rate is about 40% of the manual rate, then \$6,000 in expected losses corresponds to about \$15,000 in premium over 3 years.

This would be among the smallest risks eligible for experience rating. Nevertheless, let us ignore the eligibility criterion, and compare the primary credibilities given by Equations 4.14 and 4.25 for risks with expected annual losses less than $\Omega = 1{,}000g$. For $g = 2$, we get parameters in Equation 4.17 of:

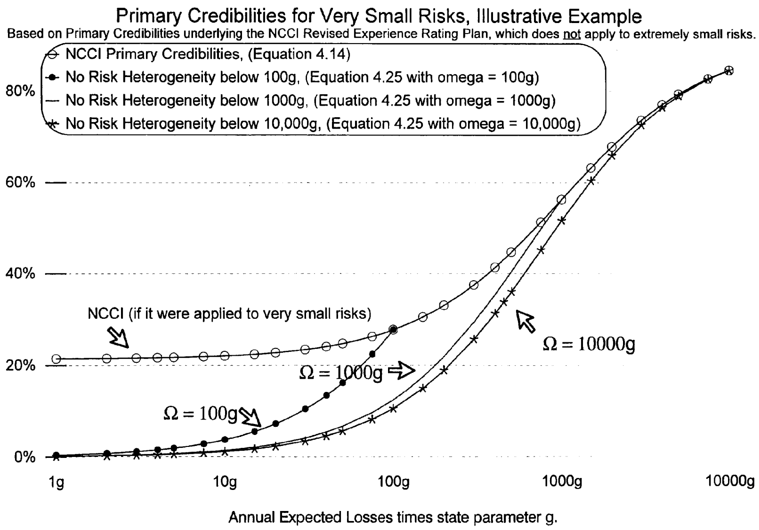
	Primary	Excess
I	466.67	3,400
J	0.3	2.25
K	5,140	407,650

Using Equation 4.14 with $Y = 3$, we get primary credibilities of:

$$Z_p = \frac{3E + 1{,}400}{3.3E + 6{,}540}.$$

[49] At first, second, and third reports as limited for experience rating.

FIGURE 3



Assumes three years of data. Risk Heterogeneity and Parameter Uncertainty, but no shifting risk parameters. g = state average claim cost / 1000. Equation 4.14 corresponds to the NCCI primary credibilities, if they were applied to very small risks. Equation 4.25 assumes that for annual expected losses below Ω, risk heterogeneity no longer applies.

For example, for $E = 100g = 200$, the primary credibility would be $2{,}000/7{,}200 = 27.8\%$. In contrast, using Equation 4.25 with $Y = 3$, $\Omega = 2{,}000$, $J' = J\ (\Omega/(I+\Omega)) = .243$, and $K' = K(\Omega/(I+\Omega)) = 4167.6$, we get primary credibilities of:

$$Z_p = \frac{3E}{3.243E + 4167.6}, \qquad E \le 2{,}000.$$

For example, for $E = 100g = 200$, the primary credibility would be $600/4816.2 = 12.5\%$.

As shown in Figure 3, the credibilities given by Equation 4.25 decline quickly to zero, while those from Equation 4.14 have a minimum value of $YI/(YI+K) = 1{,}400/(1{,}400 + 5{,}140) = .214$.

For example, for expected annual losses of $100g$, the primary credibilities are 27.8% from Equation 4.14 and 12.5% from Equation 4.25. For $1{,}000g$ the credibilities from the two equations are equal. Similarly for $E = 100g = 200$, the excess credibilities are 2.6% from Equation 4.14 and .4% from Equation 4.25.

For $E = 200$, weighting together the primary and excess credibilities, assuming a D-ratio[50] of roughly 30%, produces credibilities of 10% from Equation 4.14 and 4% from Equation 4.25. The contrast is even greater for much smaller risks.

Expected Annual	NCCI Formulas[51]		Alternate Formula 4.25 with $\Omega = 1{,}000g$	
Losses	Z_p	Z_x	Z_p	Z_x
$10g$	22.1%	2.5%	1.4%	0.04%
$100g$	27.8%	2.6%	12.5%	0.4%
$1{,}000g$	56.6%	3.8%	56.6%	3.8%

The lower credibilities from Equation 4.25 make much more sense for very small risks. For $g = 2$, $10g = \$20$ in expected annual losses.[52] The alternative formula corresponding to Equation 4.25 gives a credibility of .4% (assuming a D-ratio of roughly 30%),[53] which at least has a possibility of being reasonable. The NCCI formulas corresponding to Equation 4.14 are not applied to such small risks, nor could they be. The resulting credibility of 8.4% (assuming a D-ratio of roughly 30%)[54] is way too high. Thus, the refinement to the covariance structure for very small risks, as in Equation 4.23, is at least a step in the right direction towards obtaining reasonable experience rating credibilities for very small risks.

[50] The D-ratio is the ratio of primary losses to primary plus excess losses.

[51] Equation 4.16 with $g = 2$ and E' equal to three times expected annual losses.

[52] A single full-time clerical employee might have \$20 or more in expected annual losses for workers compensation. This is very far below the size of risk that is experience rated.

[53] $(1.4\%)(30\%) + (.04\%)(1 - 30\%) = .448\%$, where from the table $Z_p = 1.4\%$ and $Z_x = .04\%$.

[54] $(22.1\%)(30\%) + (2.5\%)(1 - 30\%) = 8.38\%$, where from the table $Z_p = 22.1\%$ and $Z_x = 2.5\%$.

Figure 3 also displays the result of choosing $\Omega = 10{,}000g$ rather than $\Omega = 1{,}000g$. The credibilities are relatively insensitive to the choice between these two values of Ω. Either value of Ω used with Equation 4.25 allows a smooth transition down to zero from the NCCI credibilities for very small risks. The transition at $E = \Omega$ between Formulas 4.24 and 4.25 will be smoothest when the slopes at $E = \Omega$ are similar.

If the credibility Z is given by Formula 4.24, then

$$\frac{dZ}{dE} = \frac{Y(K - IJ)}{((Y+J)E + YI + K)^2} = \frac{Y(K - IJ)}{(YE(1 + I/E) + JE + K)^2}.$$

If instead the credibility is given by Formula 4.25, then

$$\frac{dZ}{dE} = \frac{YK'}{((Y+J')E + K')^2} = \frac{YK(1 + I/\Omega)}{(YE(1 + I/\Omega) + JE + K)^2}.$$

At $E = \Omega$, the denominators of the derivatives of the two formulas for Z are equal.

Thus, it follows that at $E = \Omega$, the ratio of the derivative with respect to E of Z given by Formula 4.25 to the derivative with respect to E of Z given by Formula 4.24 is: $(1 + I/\Omega)/(1 - IJ/K)$. The transition will be smoothest when the slopes of the curves are close, which occurs when this ratio of derivatives is close to unity.[55] In most applications IJ/K will be small, and thus $1/(1 - IJ/K)$ will be close to unity.[56] Thus, if Ω is at least $5I$, the ratio of derivatives at Ω will be close to unity, producing a smooth transition between the two credibility formulas.

Figure 3 also displays the result of choosing $\Omega = 100g$. This value would not allow a smooth transition between the two credibility formulas. The credibilities using $\Omega = 100g$ differ significantly from those obtained from using $\Omega = 1{,}000g$. Which value of Ω is most appropriate is an empirical question whose answer

---

[55] The ratio is greater than unity since $1 + I/\Omega > 1$ and $1 - IJ/K < 1$.

[56] For the NCCI Revised Experience Rating Plan, IJ/K is .027 for primary losses and .019 for excess losses.

depends on obtaining as much information as possible about the covariance structure in the particular situation.

4.8. W and B Values, Workers Compensation Experience Rating

In workers compensation experience rating it is common to display tables of W (weighting) and B (ballast) values rather than primary and excess credibilities.[57] The primary and excess credibilities are then given in terms of W, B and expected losses as:[58]

$$Z_p = \text{Expected Losses}/(\text{Expected Losses} + B), \qquad \text{and}$$

$$Z_x = WZ_p.$$

Thus, B acts like a Bühlmann credibility parameter, except that B varies by size of risk. W quantifies for a given size of risk how much smaller the excess credibility is than the primary credibility. For three years of data, each with expected annual losses of E, $Z_p = 3E/(3E + B)$.

We can calculate the ballast value B that corresponds to the primary credibilities calculated in the prior section. Prior to the imposition of a minimum value, $B = 3E(1/Z_p - 1)$, where E is the expected annual losses and Z_p is the primary credibility.[59]

Using Equation 4.25, which assumes risk homogeneity below risk size Ω, with parameters $I_p = 466.67$, $J_p = 0.3$, $K_p = 5{,}140$ and $\Omega_p = 2{,}000$ from the prior section, we can calculate the primary credibility and corresponding value of B. For example, for expected annual losses of $E = 200$, $Z_p = 12.5\%$ and thus $B = 600((1/.125) - 1) = 4{,}200$. Keeping the other parameters fixed, we can alter Ω_p, resulting in different graphs of B versus E, as shown in Figure 4.[60]

---

[57] See Gillam and Snader [19], Gillam [13] or Mahler [12].

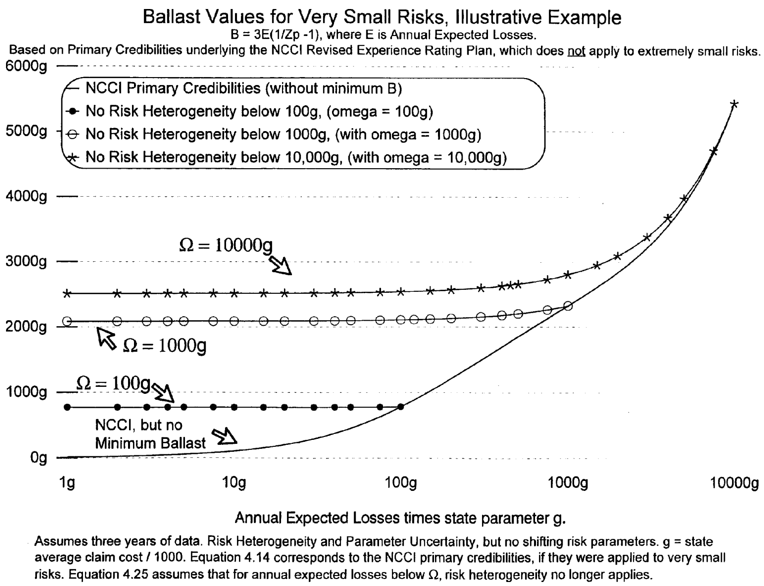
[58] Mahler [12] relates these equations to Equations 4.16.

[59] Thus, for 3 years of data, with expected losses $3E$,

$$3E/(3E + B) = 3E/(3E + 3E \cdot (1/Z_p - 1)) = Z_p.$$

[60] For example, for $\Omega_p = 1{,}000g = 2{,}000$ and $E = 100g = 200$, $B = 2{,}100g = 4{,}200$.

FIGURE 4



As seen in Figure 4, the assumption of no risk heterogeneity below Ω, with respect to primary credibilities, corresponds approximately to the imposition of a minimum ballast value. As the value of Ω varies from $100g$ to $10{,}000g$, the minimum B varies from around $800g$ to $2{,}500g$.

For $E \leq \Omega_p$ from Equation 4.25 we have

$$Z_p = \frac{YE}{(Y + J_p')E + K_p'} \approx \frac{YE}{YE + K_p'} \qquad \text{where} \quad K_p' = K_p\left(\frac{\Omega_p}{I_p + \Omega_p}\right)$$

and J_p' is small.

For $E \leq \Omega$, the credibilities approximately follow the usual Bühlmann formula, thus the minimum ballast value should be approximately

$$K_p\left(\frac{\Omega_p}{I_p + \Omega_p}\right).$$

For $I_p = 700g/3$, $K_p = 2{,}570g$, and illustrative values of Ω we get:

Ω_p	$\Omega_p/(I_p + \Omega_p)$	K_p	K'_p
$100g$	30.0%	$2{,}570g$	$771g$
$300g$	56.3%	$2{,}570g$	$1{,}447g$
$1{,}000g$	81.1%	$2{,}570g$	$2{,}084g$
$3{,}000g$	92.8%	$2{,}570g$	$2{,}385g$
$10{,}000g$	97.7%	$2{,}570g$	$2{,}511g$

Thus, for this range of values for Ω, the range of minimum ballast values K'_p is from about $800g$ to $2{,}500g$.[61] In any case, some minimum ballast value is appropriate regardless of the value of Ω. The minimum B should be a function of the state specific parameter g, and must be less than K_p.

Similarly the weighting value W is equal to $W = Z_x/Z_p$. For $E \leq \Omega_p$ and $E \leq \Omega_x$, using Equation 4.25, $W \approx (YE/(YE + K'_x))/(YE/(YE + K'_p)) = (YE + K'_p)/(YE + K'_x)$. As the size of risk goes to zero, $E \to 0$,

$$W \to K'_p/K'_x = (K_p/K_x)(\Omega_p/\Omega_x)\frac{(I_x + \Omega_x)}{(I_p + \Omega_p)}.$$

If, for example, we were to take $\Omega_x = \Omega_p = 10{,}000g$, then using the current NCCI values $I_p = 700g/3$, $K_p = 2{,}570g$, $I_x = 1{,}700g$, and $K_x = 203{,}825g$, the minimum W value would be .014; this compares to a current minimum W of .07.

4.9. Risk Heterogeneity and Shifting Risk Parameters, Dice Example

In this section, the phenomenon of shifting risk parameters will be added to the model in Section 4.1.

[61] The NCCI has introduced a minimum B of $2{,}500g$, which as seen here corresponds to $\Omega \approx 10{,}000g$.

Joe initially selects N identical red dice and N possibly different green dice.

Prior to each trial, Rose may alter the type of all the red dice. Prior to each trial, Gwen may alter the type of one or more of the green dice; Gwen acts independently on each green die. Then Joe rolls all the dice and the result is taken as: $(1-h)$ (the sum of the N red dice) + h (the sum of the N green dice.)

Assume that Rose's replacement of red dice follows the transition matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{matrix} .96 & .04 & 0 \\ .02 & .95 & .03 \\ 0 & .06 & .94 \end{matrix}$$

Thus, if the red dice are 6-sided, there is a 2% chance Rose will change them to 4-sided, a 3% chance Rose will change them to 8-sided, and a 95% chance Rose will leave them alone.

Similarly, assume that Gwen's replacement of individual green dice follows the transition matrix $\boldsymbol{G}$:

$$\boldsymbol{G} = \begin{matrix} .60 & .40 & 0 \\ .20 & .50 & .30 \\ 0 & .60 & .40 \end{matrix}$$

Gwen is ten times as likely to switch dice as is Rose.[62] Thus, the parameters of the green dice shift more swiftly than those of the red dice.[63] The dominant eigenvalue[64] (other than unity) of the transpose of $\boldsymbol{R}$ is $\rho = .954$, with a half-life of 15 trials. The dominant eigenvalue[65] (other than unity) of the transpose of $\boldsymbol{G}$ is $\gamma = .537$ with a half-life of 1.1 trials. The transition matrices $\boldsymbol{G}$ and $\boldsymbol{R}$ have been chosen such that they each have the same stationary distribution:[66] .25, .50, and .25.

---

[62] We have chosen this simple relation for illustrative purposes. Gwen could switch dice at any rate relative to Rose.

[63] One could just as easily model the reverse situation.

[64] The three eigenvalues of $\boldsymbol{R}$ are 1, .954 and .896.

[65] The three eigenvalues of $\boldsymbol{G}$ are 1, .537 and $-.037$.

[66] One could model a somewhat more complicated situation where the green and red dice had different stationary distributions.

For now take the simplest case in which Joe rolls a single die of each color, $N = 1$. (The next section will deal with the more general case of $N \geq 1$.)

As shown in Mahler [1], the covariance of trials X_i and X_j for either a single red or green die is given by Equation 3.12:

$$\mathrm{Cov}[X_i, X_j] = \sum_{k>1} \zeta_k \lambda_k^{|i-j|} + \delta_{ij}\eta^2$$

where η^2 is the Expected Value of the Process Variance, $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ij} = 1$ for $i = j$, λ_k are the eigenvalues of the transition matrix and the ζ_k are a function of the transition matrix and the means of the different dice.[67]

For transition matrix $\boldsymbol{R}$:

k	λ_k	ζ_k
1	1	12.25
2	.954	.4676
3	.896	.0324

For transition matrix $\boldsymbol{G}$:

k	λ_k	ζ_k
1	1	12.25
2	.537	.4676
3	−.037	.0324

Note that since we have chosen the same basic structure for the shifting of the green and red dice the ζ values are the same. Also note that $\sum_{k>1} \zeta_k = .5$ = Variance of the Hypothetical Means in the absence of shifting risk parameters. The eigenvalues are different, reflecting the different rates of shifting parameters.

In this case the expected value of the process variance = η^2 = 3.08. Thus, for the red dice the covariance between trials of data

[67] The dice in this case are the different states of the Markov chain. See Mahler [1].

is:

$$\mathrm{Cov}[Y_i, Y_j] = (.4676)(.954^{|i-j|}) + (.0324)(.896^{|i-j|}) + 3.08\,\delta_{ij}. \tag{4.26}$$

For the green dice the covariance between trials of data is:

$$\mathrm{Cov}[W_i, W_j] = (.4676)(.537^{|i-j|}) + (.0324)((-.037)^{|i-j|}) + 3.08\,\delta_{ij}. \tag{4.27}$$

Equation 4.27 can be approximated as:

$$\mathrm{Cov}[W_i, W_j] \approx (.5)(.537)^{|i-j|} + 3.08\,\delta_{ij}. \tag{4.28}$$

Similarly, Equation 4.26 can be approximated as:[68]

$$\mathrm{Cov}[Y_i, Y_j] \approx (.5)(.954)^{|i-j|} + 3.08\,\delta_{ij}. \tag{4.29}$$

Equations 4.28 and 4.29 are each of the form given by Equation 3.14:

$$\mathrm{Cov}[X_i, X_j] \approx \tau^2 \lambda^{|i-j|} + \eta^2\,\delta_{ij}. \tag{4.30}$$

In both cases the Variance of the Hypothetical Means[69] = τ^2 = .5 while the Expected Value of the Process Variance = η^2 = 3.08.

Let Y_i = result of a red die, W_i = result of a green die, and $X_i = (1-h)Y_i + hW_i$ = result of a trial (for one die of each kind). Then

$$\begin{aligned}\mathrm{Cov}[X_i, X_j] &= \mathrm{Cov}[(1-h)Y_i + hW_i, (1-h)Y_j + hW_j]\\ &= (1-h)^2\,\mathrm{Cov}[Y_i, Y_j] + (1-h)h\,\mathrm{Cov}[Y_i, W_j]\\ &\quad + (1-h)h\,\mathrm{Cov}[W_i, Y_j] + h^2\,\mathrm{Cov}[W_i, W_j].\end{aligned}$$

[68]Depending on the particular example, putting the covariance in terms of the principal eigenvalue other than unity will represent more or less of an approximation. For example, for the green dice, the approximate covariances from Equation 4.28 for separations of 1, 2, and 3 trials are .2685, .1442, and .0774. These compare to the exact covariances from Equation 4.27 of .2499, .1349 and .0724. On the other hand, the approximation of Equation 4.26 by Equation 4.29 is an example where the approximate covariances are close to the exact covariances.

[69]In the absence of shifting risk parameters.

However, the green and red die are independent of each other, so that

$$\mathrm{Cov}[Y_i, W_j] = \mathrm{Cov}[W_i, Y_j] = 0.$$

Thus, $\mathrm{Cov}[X_i, X_j] = (1-h)^2\,\mathrm{Cov}[Y_i, Y_j] + h^2\,\mathrm{Cov}[W_i, W_j]$.

$$\mathrm{Cov}[X_i, X_j] \approx (1-h)^2(.5)(.954^{|i-j|}) + h^2(.5)(.537^{|i-j|}) + 3.08(1-h)^2\,\delta_{ij} + 3.08h^2\,\delta_{ij}. \tag{4.31}$$

In general Equation 4.31 can be written as:

$$\mathrm{Cov}[X_i, X_j] \approx (1-h)^2\tau_1^2\rho^{|i-j|} + h^2\tau_2^2\gamma^{|i-j|} + (1-h)^2\eta_1^2\,\delta_{ij+h^2}\eta_2^2\,\delta_{ij} \tag{4.32}$$

where we have allowed for possibly different values of the variance of the hypothetical means[70] τ_1^2 and τ_2^2, as well as possibly different values of the expected value of the process variance η_1^2 and η_2^2, for the "red" and "green" risk processes.

4.10. Behavior by Size of Risk with Risk Heterogeneity and Shifting Risk Parameters

According to Section 4.1, the green and red dice contribute differently to the Variance of the Hypothetical Means and to the covariances as the number of dice N increases. For the sum of N identical red dice, the VHM is $.5N^2 = N^2\tau_1^2$. For the sum of N possibly different green dice, the VHM is $.5N = N\tau_2^2$. In both case the EPV $= N\eta^2 = 3.08N$.

Thus, for N dice, Equation 4.32 becomes:

$$\mathrm{Cov}[X_i, X_j] = (1-h)^2N^2\tau_1^2\rho^{|i-j|} + h^2N\tau_2^2\gamma^{|i-j|} + (1-h)^2N\eta_1^2\,\delta_{ij} + h^2N\eta_2^2\,\delta_{ij}. \tag{4.33}$$

---

[70] In the absence of shifting risk parameters.

For insurance applications to frequency, pure premiums, etc., it will be useful to rewrite Equation 4.33 as:[71]

$$\mathrm{Cov}[X_i, X_j] = r^2\{\rho^{|i-j|} + (I/E)\gamma^{|i-j|} + (K/E)\delta_{ij}\},$$
$$1 \geq \rho, \gamma \geq 0 \quad I, K \geq 0. \qquad (4.34)$$

Equation 4.34 for the covariances in the presence of shifting risk parameters and risk heterogeneity combines the features of Equation 3.16 with shifting risk parameters and Equation 4.3 with risk heterogeneity.

Equation 4.33 displays the typical behavior in the presence of risk heterogeneity ($h > 0$); there is a piece of the variance of hypothetical means that increases as N^2 and a piece that increases only as N, the size of risk. Therefore, the relative importance of the two dominant eigenvalues ρ and γ varies by size of risk N. For N large, ρ is relatively more important than for N small. Thus for large size risks the log-correlations decline at a rate of approximately ρ. For medium size risks, the decline rate will be between ρ and γ. For very small risks, the decline rate should be approximately γ. This same behavior also holds for Equation 4.34.

For the dice example, $\rho = .954$ and $\gamma = .537$, thus larger risks should have their log-correlations decline approximately with a slope of $\ln .954$,[72] while smaller risks would see their log-correlations decline more quickly. For $h = .8$, Figure 5 shows the behavior for various sizes of risk. The correlations are both smaller for fewer numbers of dice and decline more quickly as the separation of years increases.

For this example, plugging into Equation 4.33, the values $h = .8$, $\tau_1^2 = \tau_2^2 = .5$, $\eta_1^2 = \eta_2^2 = 3.08$, we obtain:

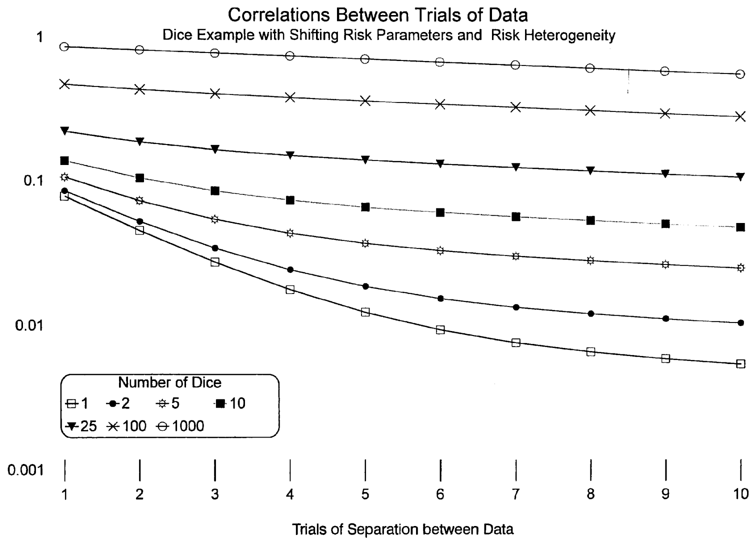
$$\mathrm{Cov}[X_i, X_j] = .02N^2 .954^{|i-j|} + .32N .537^{|i-j|} + 2.0944N\,\delta_{ij}.$$

---

[71] Where as before E is the size of risk, I quantifies risk heterogeneity and K is the Bühlmann credibility parameter. ρ and γ quantify the rate(s) of shifting of risk parameters.

[72] The correlation declines approximately as $.954^{|i-j|}$, thus, its log declines approximately as $|i-j|(\ln .954)$.

FIGURE 5



Note h=.8, indicating a large impact from risk heterogeneity. The two dominant eigenvalues are .954 and .537.

Thus, $\mathrm{Var}[X] = \mathrm{Cov}[X,X] = .02N^2 + 2.4144N.$

Thus, for this example the correlations between years are given by:

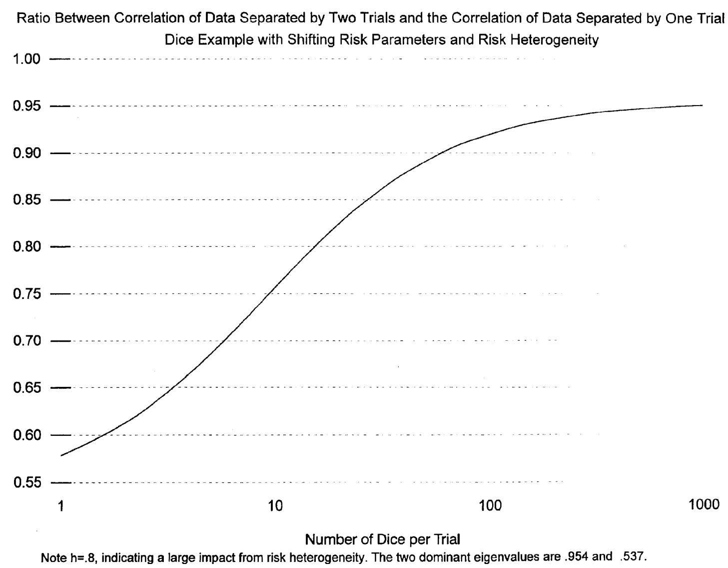
$$\mathrm{Corr}[X_i,X_j] = \frac{.02N^2.954^{|i-j|} + .32N.537^{|i-j|}}{.02N^2 + 2.414N}, \qquad i \neq j. \tag{4.35}$$

Figure 6 shows the ratio of $\mathrm{Corr}[X_i,X_i+2]$ to $\mathrm{Corr}[X_i,X_i+1]$.[73] As the number of dice increases this ratio gets closer to $\rho = .954$. In this example, larger risks have less quickly shifting risk parameters over time.[74]

[73]Figure 6 shows the approximation given by Equation 4.35. The more exact results that would be obtained starting with Equations 4.26 and 4.27 including terms for all the eigenvalues, would display the same behavior.

[74]If the transition matrices for Gwen and Rose had been reversed, then the larger risks would have had more quickly shifting risk parameters than smaller risks.

FIGURE 6



For general insurance applications, Equation 4.35 would become for the covariances written as in Equation 4.34:

$$\mathrm{Corr}[X_i, X_j] = \frac{E\rho^{|i-j|} + I\gamma^{|i-j|}}{E + I + K}, \qquad i \neq j. \tag{4.36}$$

In Equation 4.36, as $E \to \infty$, $\mathrm{Corr}[X_i, X_j] \to \rho^{|i-j|}$, while as $E \to 0$, $\mathrm{Corr}[X_i, X_j] \to \gamma^{|i-j|} I/(I+K)$. As will be discussed subsequently, examining the behavior of the correlations between years of data as the separation between years and the size of risk vary will allow one to estimate the parameters of the covariance structure which are needed to calculate credibilities.

5. SHIFTING RISK PARAMETERS, RISK HETEROGENEITY, AND PARAMETER UNCERTAINTY

In this section, the effects of shifting risk parameters, risk heterogeneity and parameter uncertainty will be combined. In

Section 5.1, the three phenomena will be combined for the dice example. The model will be put into a form useful for insurance applications in Section 5.2. Section 5.3 will incorporate the previously discussed refinement to the covariance structure for very small risks. Section 5.4 will discuss all three phenomena in the context of Philbrick's target shooting example.

5.1. All Three Phenomena, Dice Example

Combining the examples in Sections 3.6, 4.4, and 4.9 we can incorporate shifting risk parameters, risk heterogeneity, and parameter uncertainty.

Joe initially selects N identical red dice and N possibly different green dice. Prior to each trial, Rose may alter the type of all the red dice. Prior to each trial, Gwen may alter the type of one or more of the green dice; Gwen acts independently on each green die.

For each trial Joe rolls all the dice and Mary flips a coin. The result of a trial is:

$$(1-h)(\text{Sum of } N \text{ Red Dice}) + h(\text{Sum of } N \text{ Green Dice}) \\ + N(\text{Result of Coin Flip})$$

where the coin flip is counted as $-\frac{1}{2}$ for tails and $\frac{1}{2}$ for heads.

The presence of the coin flip does not alter the hypothetical means. However, as in Section 4.4, the Expected Value of the Process Variance is $(3.08)(1-h)^2N + (3.08)h^2N + .25N^2$. Combining this with the Variance of the Hypothetical Means from Section 4.10, the covariance between the results of trials i and j is:

$$\mathrm{Cov}[X_i, X_j] = (1-h)^2N^2(.5).954^{|i-j|} + h^2N(.5).537^{|i-j|} \\ + (1-h)^2N(3.08)\,\delta_{ij} + h^2N(3.08)\,\delta_{ij} + (.25)N^2\,\delta_{ij}. \tag{5.1}$$

Equation 5.1 can be written more generally as:

$$\begin{aligned}\mathrm{Cov}[X_i, X_j] = {}& (1-h)^2 N^2 \tau_1^2 \rho^{|i-j|} + h^2 N \tau_2^2 \gamma^{|i-j|} \\ & + (1-h)^2 N \eta_1^2 \delta_{ij} + h^2 N \eta_2^2 \delta_{ij} + u^2 N^2 \delta_{ij}.\end{aligned} \tag{5.2}$$

In insurance we normally are interested in quantities such as claim frequency[75] or pure premium,[76] which have the volume of data in the denominator. This introduces a factor of $1/\text{volume}^2$ into the variances and covariances.

In the dice example, this would be the equivalent of the result of a trial being the previously defined "result of a trial" divided by N:

$$\frac{1}{N}\left\{\begin{array}{l}(1-h)(\text{Sum of } N \text{ Red Dice}) + h(\text{Sum of } N \text{ Green Dice}) \\ +N(\text{Result of Coin Flip})\end{array}\right\}.$$

In that case, Equation (5.2) is modified to:

$$\begin{aligned}\mathrm{Cov}[X_i, X_j] = {}& (1-h)^2 \tau_1^2 \rho^{|i-j|} + h^2 \tau_2^2 \gamma^{|i-j|}/N \\ & + (1-h)^2 \eta_1^2 \delta_{ij} + h^2 \eta_2^2 \delta_{ij} + u^2 \delta_{ij}.\end{aligned} \tag{5.3}$$

There are those portions of the covariance that are independent of size of risk and those portions such as the process variance which decline with size of risk, when dealing with claim frequencies, pure premiums, etc.

5.2. General Form of Covariances, All Three Phenomena

Equation 5.3 contains four different types of terms. There are those that decrease as the inverse of the size of risk N and those that do not depend on N. There are those involving δ_{ij} that are related to the process variance and are not present in the covariance between different years. On the other hand, there

[75] Frequency = claims/exposures.
[76] Pure Premiums = losses/exposures.

are those involving $\lambda^{|i-j|}$ that are related to the variance of the hypothetical means.

In specific examples, the key elements will be the speed with which parameters shift and thus the half-lives of ρ and γ, and the relative weights of each of the four types of terms. With this in mind it will be worthwhile to rewrite Equation 5.3. Let $r^2 = (1-h)^2\tau_1^2$, $g^2 = h^2\tau_2^2$, $e^2 = (1-h)^2\eta_1^2 + h^2\eta_2^2$, and rather than N use E as some appropriate measure of size of risk.[77]

Then Equation 5.3 becomes:

$$\begin{aligned} \mathrm{Cov}[X_i, X_j] &= r^2\rho^{|i-j|} + g^2\gamma^{|i-j|}/E + \delta_{ij}(e^2/E + u^2) \\ \mathrm{Var}[X] = \mathrm{Cov}[X, X] &= r^2 + g^2/E + e^2/E + u^2. \end{aligned} \tag{5.4}$$

As before letting $I = g^2/r^2$, $J = u^2/r^2$ and $K = e^2/r^2$, then

$$\mathrm{Cov}[X_i, X_j] = r^2\{\rho^{|i-j|} + \gamma^{|i-j|}(I/E) + (J + K/E)\delta_{ij}\}. \tag{5.5}$$

Thus, the correlations are:

$$\mathrm{Corr}[X_i, X_j] = \frac{E\rho^{|i-j|} + I\gamma^{|i-j|}}{E(1+J) + K + I}. \tag{5.6}$$

For large risks the term with $\rho^{|i-j|}$ will dominate, while for small risks the term with $\gamma^{|i-j|}$ will dominate. For large risks the log-correlations will decline as ρ, while for small risks the log-correlations will decline as γ. For risks of medium size the decline will be between ρ and γ.

Thus, this model will be particularly useful when and if there are different decline rates in correlations by size of risk.[78] ρ can be estimated from the slopes for large risks of the log-correlations versus separations. γ can be estimated from the slopes for small risks of the log-correlations. The size of I can be

---

[77] For example, E could be expected losses in workers compensation experience rating.

[78] Where the rate of decline in the correlations is not dependent on size of risk, one can set $\rho = \gamma$.

estimated by what constitutes a "medium-size risk," where the decline rate of the log covariances are about halfway between ρ and γ. At that size $E \approx I$.

As we take larger and larger risks, Equation 5.6 for the correlations approaches

$$\lim_{E\to\infty} \mathrm{Corr}[X_i, X_j] = \frac{\rho^{|i-j|}}{1+J}.$$

Thus, we can estimate J, quantifying the impact of parameter uncertainty, by examining for large risks the correlations between years. For example, if we fit an exponential regression to such correlations versus the separations, then the intercept can be used to estimate J. For large risks:

$$\ln \mathrm{Corr}[X_i, X_j] \approx -\ln(1+J) + |i-j| \ln \rho.$$

For any size:

$$\ln \mathrm{Corr}[X_i, X_j] = \ln(E\rho^{|i-j|} + I\gamma^{|i-j|}) - \ln(E(1+J) + K + I).$$

Assuming a fixed set of parameters I, J, K, ρ and γ, then for a fixed size of risk E, the second term is constant, while the first term depends on the separation between years $|i-j|$. We expect the decline rate to be some rate between ρ and γ, depending on the relative sizes of E and I. Very approximately:[79]

$$\ln(E\rho^{|i-j|} + I\gamma^{|i-j|}) \approx |i-j| \ln\left(\frac{E\rho + I\gamma}{E+I}\right) + \ln(E+I)$$

Thus,

$$\ln \mathrm{Corr}[X_i, X_j] \approx |i-j| \ln\left(\frac{E\rho + I\gamma}{E+I}\right) + \ln\left(\frac{E+I}{E(1+J)+K+I}\right).$$

Thus, if we fit an exponential least squares regression to the correlations by separations ≥ 1, we would expect to have a slope between ρ and γ and an "intercept" of $(E+I)/(E(1+J)+K+I)$.

[79]For $E = 0$, $\ln(I\gamma^{|i-j|}) = |i-j| \ln\gamma + \ln I$. For $I = 0$, $\ln(E\rho^{|i-j|}) = |i-j| \ln\rho + \ln E$. For $|i-j| = 1$, the approximation is exact. The approximation is poor when $|i-j|$ is large, ρ and γ differ substantially, and E is approximately the same as I.

This intercept[80] is equal to the credibility for a single year of data in the absence of shifting risk parameters, as in Equation 4.15.

We can therefore approximate some of the necessary parameters from the behavior of the observed correlations as the size of risk and number of years of separation vary.

For each of various sizes of risk we can fit exponential least squares regressions to the correlations for years separated by one year or more. The intercept for each size category is an estimate of the credibility of one year of data in the absence of shifting risk parameters over time. These credibilities by size of risk can be used to estimate the parameters I, J and K. The slope (exponential rate of decline) of the correlations varies between γ and ρ as the size of risk increases. At an intermediate size of about I, the slope will be about halfway between γ and ρ.

In the situation where the years X_i and X_j have different expected volumes of data E_i and E_j, Equation 5.5 can be generalized to:

$$\mathrm{Cov}[X_i, X_j] = r^2 \left\{\rho^{|i-j|} + \gamma^{|i-j|} I/\sqrt{E_i E_j} + \left(J + K/\sqrt{E_i E_j}\right)\delta_{ij}\right\} \tag{5.7}$$

In the covariance, those terms that were divided by E in Equation 5.5 are now in Equation 5.7 divided by the geometric average of the sizes of risk, $\sqrt{E_i E_j}$. If $E_i = E_j = E$, then $\sqrt{E_i E_j} = E$, so that Equation 5.7 would reduce to Equation 5.5. The use of the square root function in the generalization was motivated by the $\sqrt{\mathrm{VAR}[X_1]\mathrm{VAR}[X_2]}$ that appears in the denominator of the correlation of X_1 and X_2.

Equations 5.5 or 5.7 can be used to calculate all of the covariances necessary to solve Equations 2.4 for the credibilities.

An example of how to calculate the credibilities in general will be given in Section 6. However, prior to that it is worthwhile to

[80] For convenience in this paper, $(E + I)/(E(I + J) + K + J)$, rather than the natural log of that quantity, will be referred to as the intercept.

generalize Equation 5.7 for the covariance in order to take into account the different behavior of very small risks.

5.3. Very Small Risks, General Covariance Structure

In Sections 4.3 and 4.6 a refinement for very small sizes of risk was introduced. In this section, this refinement will be introduced into the general covariance structure.

The same logic concerning risk heterogeneity and very small risks applies as well when both parameter uncertainty and shifting risk parameters are considered. If we assume risk heterogeneity does not apply for $E \le \Omega$, then Equation 5.4 for the covariances is split into two separate equations, per Equations 4.22 and 4.23.

For $E \ge \Omega$, Equation 5.5 holds:

$$\mathrm{Cov}[X_i, X_j] = r^2\{\rho^{|i-j|} + \gamma^{|i-j|}(I/E) + \delta_{ij}((K/E) + J)\}, \qquad E \ge \Omega. \tag{5.8}$$

For $E \le \Omega$, the term involving I takes on its value at $E = \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\{\rho^{|i-j|} + \gamma^{|i-j|}(I/\Omega) + \delta_{ij}((K/E) + J)\}, \qquad E \le \Omega. \tag{5.9}$$

In the situation where the years X_i and X_j have different expected volumes E_i and E_j, Equations 5.8 and 5.9 can be generalized to:[81]

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{|i-j|} + \gamma^{|i-j|} I/\sqrt{E_i E_j} + \delta_{ij}\left(K/\sqrt{E_i E_j} + J\right)\right\}, \qquad \sqrt{E_i E_j} \ge \Omega \tag{5.10}$$

[81] It should be noted that in Equations 5.10 to 5.11 the expression $\sqrt{E_i E_j}$ only enters due to the presence of risk heterogeneity. This results in terms such as $I/\sqrt{E_i E_j}$. In contrast, where $\sqrt{E_i E_j}$ divides K it is multiplied by δ_{ij}. These terms are zero unless $i = j$, so $\sqrt{E_i E_j}$ could be replaced in these terms by either E_i or E_j. This simplification in notation is conventional in the absence of risk heterogeneity.

$$\mathrm{Cov}[X_i, X_j] = r^2 \left\{ \rho^{|i-j|} + \gamma^{|i-j|} I/\Omega + \delta_{ij} \left(K/\sqrt{E_i E_j} + J \right) \right\},$$
$$\sqrt{E_i E_j} \leq \Omega. \qquad (5.11)$$

5.4. Philbrick's Target Shooting Example

Philbrick [5] explains credibility concepts by using a target shooting example. There are four marksmen each shooting at his own target. Each marksman's shots are assumed to be distributed around his target, with expected mean equal to his target. Once we observe a shot or shots from a single *unknown* marksman, we could use Bühlmann credibility to estimate the location of the next shot from the *same* marksman.

The key features of Bühlmann credibility are explained by Philbrick as follows by altering the initial conditions of the target shooting example:

Feature of Target Shooting Example	Mathematical Quantification	Bühlmann Credibility
Better Marksmen	Smaller EPV	Larger
Targets Further Apart	Larger VHM	Larger
More Shots	Larger N	Larger

These mathematical relationships also follow from Bühlmann's credibility formula, Equation 1.1.

We can modify the example in Philbrick to include each of the three phenomena discussed in this paper.

In Philbrick, it is assumed that each marksman continues to shoot at his target.[82] Within a single example in Philbrick, the risk parameters do not shift over time. If instead there were a small random chance that between each shot a marksman would

[82] It is also assumed within an example that the targets are stationary, the marksmen remain the same and do not get better or worse, nor do the marksmen move closer to or further from the targets.

switch targets, then one would have shifting risk parameters over time.[83] In this case, the credibility assigned to a given shot would be less than if each marksman always shot at the same target. The informational content of a shot for purposes of predicting the next shot from the same marksman has been reduced by the presence of shifting parameters over time.

The Philbrick example can also be altered in order to incorporate risk heterogeneity. Assume we have teams of marksmen. Assume each marksman on a team shoots at his own target. Assume that while members of a team each shoot at a possibly different target, the members of a team are more likely to shoot at the same target than are marksmen who are not members of the same team. For example, the six members of Team 1 might shoot at targets A, A, A, B, C, and D respectively. For the purpose of predicting the next shot, the informational content of a given number of shots from Team 1 is less than if all the members of the team always had the same target. Risk heterogeneity has reduced the credibility assigned to a given number of shots.

Assume, for example, as the teams got bigger each additional marksman in Team 1 was assigned target A half the time and targets B, C, and D one-sixth of the time. Then as the teams got bigger, the credibility assigned to a set of shots, one per team member, would not be the same as the Bühlmann case in which each team member shot at the same target. With risk heterogeneity the credibility would increase more slowly as the teams increase in size; the incremental informational content of another team member is less when they do not all shoot at the same target.

As discussed previously, in the presence of risk heterogeneity, the credibilities are given by Equation 4.4:

$$Z = \frac{E+I}{E+I+K}.$$

[83]This is analogous in the dice example to Beth possibly replacing dice between the rolls.

The derivative of Z with respect to the size of risk E is $K/(E+I+K)^2$. This derivative decreases as I increases; the greater the impact of risk heterogeneity, the more slowly the credibility increases with size of risk.

Finally, the Philbrick example can be altered to incorporate parameter uncertainty. Again assume that there are teams of marksmen, but each marksman shoots at the same target. Assume that for each round of shots, one per team member, every team member uses the same rifle. However, between rounds the rifle is replaced by another. Further assume the rifles look alike but some shoot high, some shoot low and to the left, etc. Also assume the marksmen on a team do not communicate with each other, nor adjust their aim based on their teammate's shots, so that all team members are equally affected by the peculiarities of the given rifle. The errors introduced by the switching rifles reduce the informational content of the shots; in the presence of parameter uncertainty less credibility is assigned to the data, holding all else equal.

In addition, adding more team members can never eliminate the effect of an individual, randomly chosen rifle. In the presence of parameter uncertainty the credibility of a single year of data does not approach unity as the risk size increases; rather in Equation 3.7 the credibility goes to $1/(1+J)$ as the risk size approaches infinity.

However, by observing many rounds of shots, assuming the errors of the rifles average to zero, one can eliminate their impact. In the presence of parameter uncertainty (and no shifting risk parameters over time), the credibility of a given size of risk goes to unity as the number of years goes to infinity; the credibilities in Equation 3.10 go to unity as the number of years increases.

Clearly, we could modify the Philbrick target shooting example to incorporate two or all three of the phenomena discussed in this paper.

6. ILLUSTRATIVE EXAMPLES OF CALCULATING CREDIBILITIES

This section will present illustrative examples of calculating credibilities based on the general covariance structure presented in Section 5. Section 6.1 deals with large risks, while Section 6.2 includes the refinement to the covariance structure for very small risks. Section 6.3 shows how differing volumes of data by year would affect the credibilities. Section 6.4 shows an example in which no weight is given to the overall mean.

6.1. An Example of Calculating Credibilities, Large Risks

As an example, take the following illustrative values in Equations 5.5 or 5.7 for the covariances in the presence of all three phenomena:[84]

$\rho = .9$ (rate of shifting parameters related to risk homogeneity),

$\gamma = .7$ (rate of shifting parameters related to risk heterogeneity),

$e^2 = 9{,}000$ (expected value of process variance without parameter uncertainty),

$u^2 = 2$ (variance related to parameter uncertainty),

$r^2 = 3$ (portion of variance of hypothetical means related to risk homogeneity),

$g^2 = 4{,}000$ (portion of variance of hypothetical means related to risk heterogeneity),

$$I = g^2/r^2 = 1{,}333,$$

$$J = u^2/r^2 = .6667, \qquad \text{and}$$

$$K = e^2/r^2 = 3{,}000.$$

[84]These values were chosen solely to present an example. Note that if one multiplies e^2, u^2, r^2 and g^2 all by the same constant, then all the covariances are multiplied by that same constant, but the credibilities are unchanged.

Assuming each year of data has equal volume E, Equation 5.5 becomes:

$$\mathrm{Cov}[X_i, X_j] = (3).9^{|i-j|} + (4{,}000/E).7^{|i-j|} + \delta_{ij}(9{,}000/E + 2). \tag{6.1}$$

Thus, the variance is:

$$\mathrm{Var}[X_i] = \mathrm{Cov}[X_i, X_i] = (13{,}000/E) + 5.$$

The covariance between years of data separated by two years is:

$$\mathrm{Cov}[X_1, X_3] = (1{,}960/E) + 2.43.$$

For 4 years of data each of volume E, the variance-covariance matrix is:

$$\begin{matrix} (13{,}000/E)+5 & (2{,}800/E)+2.7 & (1{,}960/E)+2.43 & (1{,}372/E)+2.187 \\ (2{,}800/E)+2.7 & (13{,}000/E)+5 & (2{,}800/E)+2.7 & (1{,}960/E)+2.43 \\ (1{,}960/E)+2.43 & (2{,}800/E)+2.7 & (13{,}000/E)+5 & (2{,}800/E)+2.7 \\ (1{,}372/E)+2.187 & (1{,}960/E)+2.43 & (2{,}800/E)+2.7 & (13{,}000/E)+5. \end{matrix}$$

For example, if $E = 1000$ then the variance-covariance matrix is:

$$\begin{matrix} 18 & 5.5 & 4.39 & 3.559 \\ 5.5 & 18 & 5.5 & 4.39 \\ 4.39 & 5.5 & 18 & 5.5 \\ 3.559 & 4.39 & 5.5 & 18. \end{matrix}$$

Assume we are using three years of data to estimate the fourth year directly following them. Then Equations 2.4 for the credibilities to assign to each of the three years of data are:

$$\begin{aligned} 18Z_1 + 5.5Z_2 + 4.39Z_3 &= 3.559, \\ 5.5Z_1 + 18Z_2 + 5.5Z_3 &= 4.39, \qquad \text{and} \\ 4.39Z_1 + 5.5Z_2 + 18Z_3 &= 5.5. \end{aligned} \tag{6.2}$$

Equations 6.2 are three linear equations in three unknowns, with solution:

$$\begin{aligned} Z_1 &= 9.62\%, \\ Z_2 &= 14.15\%, \qquad \text{and} \\ Z_3 &= 23.88\%, \end{aligned}$$

where Z_1 is the credibility assigned to the oldest year of data and Z_3 is the credibility assigned to the most recent year of data. Note that $Z_1 + Z_2 + Z_3 = 47.65\% < 100\%$. The remaining weight of 52.35% is given to the grand mean.[85]

It should be noted that Equations 2.4 for the credibilities[86] were derived so as to minimize the expected squared error of the estimate. As derived in Mahler [1][87] the expected squared difference between the estimate and observation as a function of the variance-covariance matrix and the credibilities is:

$$V(Z) = \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j C_{ij} - 2\sum_{i=1}^{Y} C_{i,Y+\Delta} Z_i + C_{Y+\Delta,Y+\Delta}. \quad (6.3)$$

In this particular case for $E = 1{,}000$, we get for various selected values of the credibilities the following expected squared errors:

Z_1	Z_2	Z_3	$V(Z)$
0	0	0	18
1/3	1/3	1/3	18.454
1/2	0	0	18.941
0	1/2	0	18.110
0	0	1/2	17.000
9.62%	14.15%	23.88%	15.722

Thus, the use of (the optimal least squares) credibilities of 9.62%, 14.15%, 23.88% does indeed seem to have reduced the expected squared errors.[88]

Figure 7 shows how the sum of the credibilities for three years of data varies with size of risk. In addition to the case where all

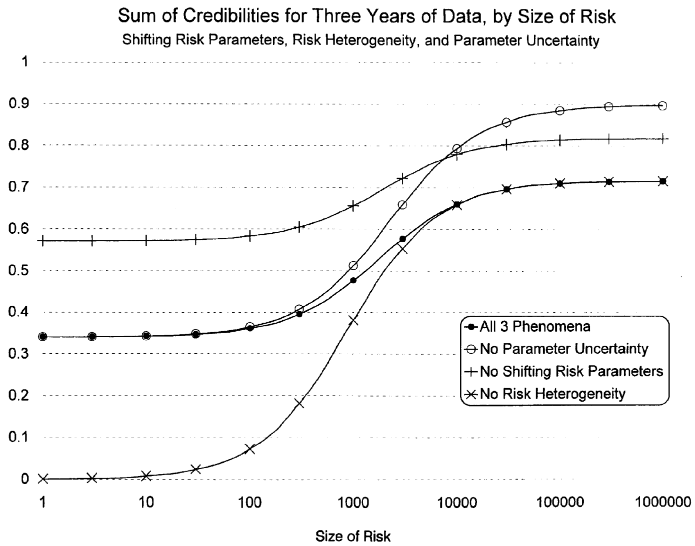
[85] The situation in which no weight is given to the grand mean is discussed below.

[86] Which are Equations 6.2 for this specific example with $E = 1{,}000$.

[87] See Appendix C in Mahler [1]. The derivation parallels that in Appendix B of the current paper.

[88] In this case, the expected squared error is about $15.722 \div 18 = 87\%$ of what one would obtain by ignoring the observations (assigning the observations zero credibility).

FIGURE 7



No gap between data and year to be estimated. See the example in Section 6.1.

three phenomena are present, cases are shown in which only two of the phenomena are present.

For no parameter uncertainty, J is set equal to zero rather than .6667. For large risks the credibility is higher than in the presence of parameter uncertainty. Nevertheless, the maximum credibility is less than 100%, due to the impact of shifting risk parameters over time.

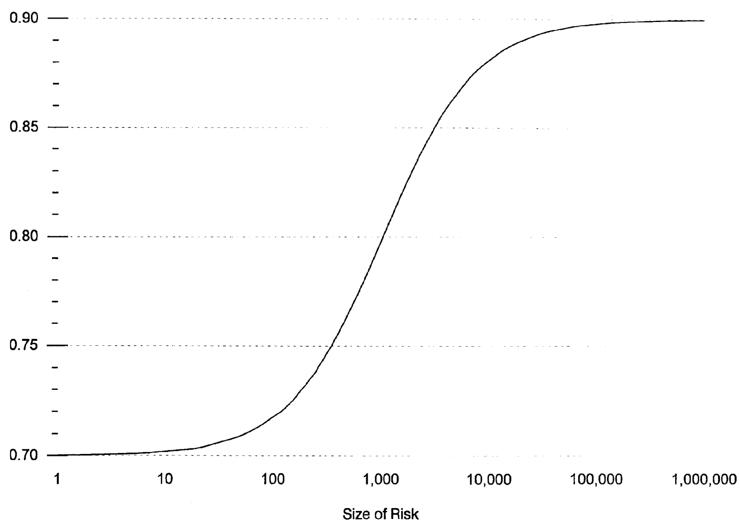
For no shifting risk parameters, $\rho = \gamma = 1$ rather than $\rho = .9$ and $\gamma = .7$. Credibilities are higher. The credibilities are given by Equation 4.14.

For no risk heterogeneity, I is set equal to zero rather than 1333. With risk homogeneity the credibilities go to zero as the risk size declines.[89]

[89]As discussed in Section 5.3, Equation 5.5 and the resulting credibilities are not appropriate for very small risks.

FIGURE 8

Ratio Between Correlation of Data Separated by Two Years and the Correlation of Data Separated by One Year



The two rates of decline of the log correlations are: $\rho = .9$ and $\gamma = .7$. See the example in Section 6.1.

The decline rate of the correlations is close to $\rho = .9$ for large risks and close to $\gamma = .7$ for small risks. Specifically, the ratio of the correlation between years separated by two years to the correlation between years separated by one year is:

$$\mathrm{Corr}[X_1, X_3]/\mathrm{Corr}[X_1, X_2] = (2.43E + 1{,}960)/(2.7E + 2{,}800). \tag{6.4}$$

Figure 8 shows how this decline rate given by Equation 6.4 varies by size of risk.

In general if the covariances are given by Equation 5.5, we expect this decline rate to be given by:

$$\mathrm{Corr}[X_1, X_3]/\mathrm{Corr}[X_1, X_2] = (\rho^2 E + I\gamma^2)/(\rho E + I\gamma). \tag{6.5}$$

If $\rho > \gamma$, then we expect to see something like Figure 8. If instead $\rho < \gamma$, we expect the curve to decrease from γ to ρ as the size increases.

The intermediate size at which the decline rate is about equally distant between ρ and γ is approximately I. This could be used to estimate I from data. In the example, $I = 1{,}333$. In Figure 8 for this size the decline rate is .81, roughly halfway between $\rho = .9$ and $\gamma = .7$.

6.2. Credibilities, Small Risks

In the example in Section 6.1, let us assume there is no risk heterogeneity for $E \leq \Omega = 100$. Then the covariances and credibilities are different for $E < 100$ than they were in Section 6.1.

For $E \leq 100$, the covariances are given by Equation 5.9:

$$\mathrm{Cov}[X_i, X_j] = (3)(.9^{|i-j|}) + (40)(.7^{|i-j|}) + \delta_{ij}(9{,}000/E + 2).$$

For $E \geq 100$, the covariances are given by Equation 5.8:

$$\mathrm{Cov}[X_i, X_j] = (3)(.9^{|i-j|}) + (4{,}000/E)(.7^{|i-j|}) + \delta_{ij}(9{,}000/E + 2).$$

For example, for $E = 10$, the variance-covariance matrix is:

$$\begin{matrix} 945 & 30.7 & 22.03 & 15.907 \\ 30.7 & 945 & 30.7 & 22.03 \\ 22.03 & 30.7 & 945 & 30.7 \\ 15.907 & 22.03 & 30.7 & 945. \end{matrix}$$

Assume we are using three years of data (each with $E = 10$), in order to estimate the fourth year directly following them. Then Equations 2.4 for the credibilities to assign to each of the three years of data are:

$$\begin{aligned} 945Z_1 + 30.7Z_2 + 22.03Z_3 &= 15.907, \\ 30.7Z_1 + 945Z_2 + 30.7Z_3 &= 22.03, \qquad \text{and} \\ 22.03Z_1 + 30.7Z_2 + 945Z_3 &= 30.7. \end{aligned} \tag{6.6}$$

Equations 6.6 are three linear equations in three unknowns, with solutions:

$$\begin{aligned} Z_1 &= 1.5\%, \\ Z_2 &= 2.2\%, \qquad \text{and} \\ Z_3 &= 3.1\% \end{aligned}$$

where Z_1 is the credibility assigned to the oldest year of data. The remaining weight not given to any of the years of data is given to the grand mean.

These credibilities assuming no risk heterogeneity below E = 100 are significantly smaller than those derived from Equation 5.5, which assumes risk heterogeneity for all sizes of risk. For $E = 10$, using Equation 5.5 to calculate the covariances rather than Equation 5.9 would result in credibilities of:

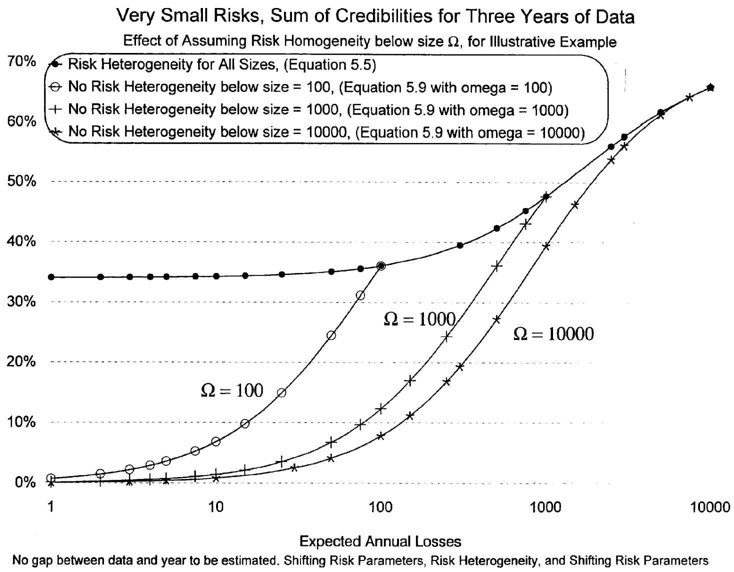
$$Z_1 = 5.7\%,$$
$$Z_2 = 9.9\%, \qquad \text{and}$$
$$Z_3 = 18.6\%.$$

Equation 5.9 produces credibilities that decline to zero as the risk size decreases in a manner similar to the usual Bühlmann formula, in contrast to Equation 4.14. Figure 9 contrasts this behavior for very small sizes, assuming $\Omega = 100$. Shown are the sum of the credibilities for three years of data as calculated above. For example, for $E = 10$, the credibilities for three years of data with risk heterogeneity sum to 34.2%, while those without risk heterogeneity (below $E = \Omega = 100$) sum to 6.8%. As E gets even smaller, in the presence of risk heterogeneity, the sum of the credibilities remains about 34%, while in the absence of risk heterogeneity it goes to zero.

Intuitively the credibility should approach zero as the size of risk approaches zero. Without the refinement discussed in Sections 4.3, 4.6 and 5.3, the covariance structure incorporating risk heterogeneity would produce credibilities that make no sense to actuaries. Credibility formulas such as Equation 4.14 or covariance structures such as Equation 5.5 should not be applied to very small risks.

Also shown in Figure 9 are the results of using Equation 5.9 with the alternate values $\Omega = 1{,}000$ or $\Omega = 10{,}000$ rather than $\Omega = 100$. In this case, the credibilities using the latter value

FIGURE 9



are significantly different than using either of the two former values.

In this example $I = 1{,}333$. This is the parameter related to risk heterogeneity, and it controls the behavior of the credibilities that result from Equation 5.5 for small risks. For $E < I$ these credibilities start leveling off significantly. Taking Ω significantly less than I, as for example 100 compared to 1,333, starts the steep descent to zero of the credibilities resulting from Equation 5.9 from an otherwise very small slope. In contrast, taking Ω either roughly equal to or greater than I, starts the descent in a much smoother manner, as is the case for $\Omega = 1{,}000$ or 10,000.

6.3. *Credibilities for Years with Differing Volumes of Data*

Returning to the example in Section 6.1, assume that the three years have differing volumes of data. Assume $E_1 = 600$, $E_2 = 1{,}600$, and $E_3 = 800$, where E_1 is the most distant of the three years. Using the inputs from before, Equation 5.7 becomes:

$$\mathrm{Cov}[X_i, X_j] = (3)(.9^{|i-j|}) + (4{,}000)(.7^{|i-j|})/\sqrt{E_i E_j} + \delta_{ij}\left(\left(9{,}000/\sqrt{E_i E_j}\right) + 2\right).$$

Assume that the year to be estimated will have a volume of data $E_4 = 1{,}000$, the average of the observed years.[90] Then the variance-covariance matrix is:

26.667	5.558	5.259	3.958
5.558	13.125	5.175	3.98
5.259	5.175	21.25	5.83
3.958	3.98	5.83	18

The credibilities are given by the solution to Equations 2.4:

$$Z_1 = 6.68\%,$$

$$Z_2 = 19.16\%, \qquad \text{and}$$

$$Z_3 = 21.12\%.$$

Thus, as expected, years 1 and 3 with their smaller volumes are given less credibility than in Section 6.1, while year 2 with its larger volume of data is given more credibility than before.

It is interesting to note that in the presence of risk heterogeneity[91] the credibilities depend on the assumed volume of data for the year being estimated, year 4.

	$E_4 = 100$	$E_4 = 1{,}000$	$E_4 = 10{,}000$
Z_1	13.15%	6.68%	4.64%
Z_2	31.18%	19.16%	15.36%
Z_3	48.44%	21.12%	12.47%

[90]While $\mathrm{Var}[X_4]$ will not enter into the equations for the credibility, $\mathrm{Cov}[X_1, X_4]$ and similar terms will. $\mathrm{Cov}[X_1, X_4]$ depends on E_4, due to the presence of risk heterogeneity. In the absence of risk heterogeneity, one need not assume a value for E_4.

[91]Whether or not there are shifting risk parameters over time.

When E_4 is large, the covariances of the data years with the year to be estimated are smaller, and therefore we assign less credibility.[92]

As E_4 gets larger, we are assuming the insured will be larger in year 4, the year to be predicted. As discussed previously, one implication of risk heterogeneity is that larger insureds are in some sense more similar to average than are smaller insureds. The less distinct insureds are from average, the less credibility we give to the data from individual insureds and the more weight we give to the overall average.[93] Thus, if E_4 is larger, we give less credibility to this insured's data and more weight to the overall average.

For mechanical applications of the methodology,[94] we would probably just assume that the volume of data in the future would be some average of that observed in the recent past for that insured. In this example, we might assume as above that:

$$E_4 = (E_1 + E_2 + E_3)/3 = 1{,}000.$$

6.4. *Credibilities, No Weight Given to the Grand Mean*

So far we have assumed that the complement of credibility is given to the grand mean. In some cases the grand mean either does not exist or is not used. In those situations, we can have the credibilities be constrained to add to 100%.

Assume that we are using three years of data to estimate the fourth year directly after them, but that no weight is given to the grand mean. Then Equations 2.4 no longer apply.

---

[92]This differs from the Bühlmann case in which the covariances between the claim frequencies of different years are assumed to be independent of the size of risk.

[93]In the target shooting example in Philbrick [5], as the targets get closer together less credibility is given to each observed shot.

[94]For example, if one were performing many thousands of experience ratings by computer.

As shown in Appendix B, the general equations for credibility when no weight is applied to the grand mean are:[95]

$$\sum_{i=1}^{Y} \mathrm{Cov}[X_i, X_k] Z_i = \mathrm{Cov}[X_k, X_{Y+\Delta}] + \frac{\lambda}{2}, \qquad k = 1, \ldots, Y$$
$$\sum_{i=1}^{Y} Z_i = 1, \tag{6.7}$$

where λ is a Lagrange Multiplier.[96]

For the covariances used in the previous Section 6.1, with $E = 1{,}000$, Equations 6.7 become:

$$18Z_1 + 5.5Z_2 + 4.39Z_3 = 3.559 + \lambda/2,$$
$$5.5Z_1 + 18Z_2 + 5.5Z_3 = 4.39 + \lambda/2,$$
$$4.39Z_1 + 5.5Z_2 + 18Z_3 = 5.5 + \lambda/2, \qquad \text{and}$$
$$Z_1 + Z_2 + Z_3 = 1.$$

These are four linear equations in four unknowns.[97] The solution is:[98]

$$Z_1 = 27.60\%,$$
$$Z_2 = 30.53\%, \qquad \text{and}$$
$$Z_3 = 41.86\%.$$

We note that $Z_1 + Z_2 + Z_3 = 1$ as desired. The most recent year is given weight $41.86\% > 27.60\%$, the weight given to the most distant year.

[95] See Equation 11.7 in Mahler [20].

[96] The Lagrange Multiplier is introduced due to the constraint equation $\Sigma Z_i = 1$. Note that λ is used to denote the Lagrange Multiplier here and was used to denote the dominant eigenvalue in prior sections. λ is commonly used in both these roles, but the reader should not be confused. There is no connection between these two separate uses of the same Greek letter.

[97] Although we are really not particularly interested in the value of the Lagrange Multiplier.

[98] The Lagrange Multiplier $\lambda = 9.853$.

Usually, as the size of risk increases, the need for stability in the estimation procedure declines, so that we give more weight to recent years of data. However, in this case that is counteracted to some extent by the assumption that large risks have more stable risk parameters over time.[99] Thus the estimation procedure can afford to be less responsive.

In this example, this leads to the credibilities being relatively insensitive to the size of risk:

	Size of Risk 1	1,000	1 Million
Z_1	28.23%	27.60%	24.93%
Z_2	30.60%	30.53%	30.21%
Z_3	41.17%	41.86%	44.86%

If we switch the rates of shifting parameters and instead takes $\rho = .7$ and $\gamma = .9$, we get a significantly different behavior by size of risk:

	Size of Risk 1	1,000	1 Million
Z_1	30.32%	27.96%	21.96%
Z_2	32.34%	30.87%	25.81%
Z_3	37.34%	41.17%	52.23%

As risk size increases, the weight given to the recent year increases more substantially than before. In general, the dependence of credibility on size of risk will depend significantly on the relative magnitudes of ρ and γ.

[99]Larger risks correspond to a decline rate in the log-correlations of $\rho = .90$ rather than $\gamma = .70$.

7. CLASSIFICATION RATE RELATIVITIES

In this section, the ideas developed so far will be applied to a simplified version of the estimation of classification rate relativities.[100] While the example draws from workers compensation, it is intended to illustrate the general applicable concepts rather than the details of workers compensation insurance.

Section 7.1 defines rate relativities. Section 7.2 describes the classification data examined. Section 7.3 describes the covariance structure and explains how correlations were estimated. Section 7.4 describes how regressions were fit to the correlations in order to estimate the parameters γ, ρ, I and J. Section 7.5 describes how the parameter K was estimated. Section 7.6 describes how the parameter Ω was selected.

Section 7.7 calculates credibilities with no weight given to the overall mean. Section 7.8 calculates credibilities with weight given to the overall mean. Section 7.9 discusses using prior estimates of the class relativities.

Section 7.10 discusses the impact of maturity of data in general. Section 7.11 gives an example of the impact of maturity on correlations while Section 7.12 gives the corresponding credibilities.

7.1. Rate Relativities

Assume that we are trying to estimate for a number of individual classes the expected pure premiums relative to the average for that group of classes. Further, assume we will do so by weighting together the observed relativities for that class over several recent years.[101] If R_{ic} is the relativity for year i, for class

---

[100]For an introduction to classification ratemaking see, for example, the Risk Classification chapter of *Foundations of Casualty Actuarial Science* [21].

[101]This is a simplification of how we might get indicated pure premiums by classification for workers compensation insurance. In that case, the relative pure premiums by class would be compared to those for an industry group. Also, the “serious,” “non-serious”

c, then the estimate of the relativity for that class for year $N+\Delta$ is: $\sum_{i=1}^{N} Z_{ic}R_{ic}$, where $\sum_{i=1}^{N} Z_{ic} = 1$. This is the situation covered by Equations 6.7.

If instead we gave the complement of credibility to the grand mean, which in this example is a relativity of unity, then Equations 2.4 would apply instead of Equations 6.7. In either case, in order to estimate credibilities the key step will be the estimation of the (expected) covariances between years of data.

7.2. Classification Data

The data to be examined is 13 (consecutive) years of classification experience in one state for workers compensation insurance.[102] For each class we will use its payroll and losses to compute its pure premium relative to its industry group for that year. If L_{ic} is the loss[103] and P_{ic} the payroll,[104] then the relative pure premium in year i for class c is:[105]

$$R_{ic} = (L_{ic}/P_{ic}) \Big/ \left(\sum_c L_{ic} \Big/ \sum_c P_{ic}\right). \tag{7.1}$$

In order to estimate the behavior of the covariances by size of class, the data for the Manufacturing and Goods and Services industry groups will be examined.[106] The Manufacturing industry

---

and "medical" pure premiums might be treated separately. See Kallop [22] and Feldblum [23]. In addition, we might rely on "National" as well as state pure premiums by class. See Harwayne [24].

[102] See Appendix C for details on the data set examined.

[103] In this illustration, the losses are paid losses plus case reserves, at latest report, for medical plus indemnity, without any limitation by claim size.

[104] Payroll is in units of \$100.

[105] Note that the relativity of an individual class within an industry group depends both on the experience of that class, the experience of the other classes, as well as the exposures by class within the industry group. Thus, a given class relativity may change over time for a number of different reasons, some of which may have little to do with the individual class.

[106] Currently five industry groups are most commonly used for workers compensation ratemaking: Manufacturing, Construction, Office and Clerical, Goods and Services, and Miscellaneous.

group will be particularly useful since it has about 270 separate classes of various sizes. The Goods and Services industry group, with only about 100 separate classes, will not allow as detailed a breakdown by size of class.[107]

7.3. Covariance Structure

The covariance structure will be assumed to be that given by Equations 5.10 and 5.11. However, for estimation purposes we will use the simpler Equations 5.8 and 5.9, which ignore the varying volume of data by year for a class.[108]

For an industry group we compute the relative pure premiums for each class for each year. Then we can compute the covariances and correlations between the different years. By examining the behavior of these covariances and correlations as the size of class and the number of years of separation vary, we can roughly estimate the parameters that appear in the covariance Formulas 5.8 and 5.9.

For this purpose, we will restrict our attention to one size category of class at a time.[109] There are a number of ways to categorize the volume of data. This example uses an estimate of the average annual expected losses for a class based on its reported payroll.[110] Other reasonable measures of volume should produce roughly similar results.

For each such size category, we estimate the covariance between any two years of observed relative pure premiums R_{ic} and R_{jc} for $c = 1,\ldots,k$ where there are k classes in the size

---

[107]The Office and Clerical industry group has only around 14 classes. The Construction industry group has about 71 classes. The Miscellaneous industry group has about 49 classes.

[108]As will be seen, the estimation process is sufficiently imprecise that this simplification is appropriate.

[109]Nevertheless, the pure premiums are relative to the *entire* industry group, regardless of size of class.

[110]The details of how the expected losses were estimated for each class for each year are described in Appendix C.

FIGURE 10

category:[111]

$$\mathrm{Cov}[R_{ic}, R_{jc}] \approx \frac{\sum_{c=1}^{k} \sqrt{P_{ic}P_{jc}} R_{ic} R_{jc}}{\sum_{c=1}^{k} \sqrt{P_{ic}P_{jc}}} - \frac{\sum_{c=1}^{k} P_{ic} R_{ic}}{\sum_{c=1}^{k} P_{ic}} \frac{\sum_{c=1}^{k} P_{jc} R_{jc}}{\sum_{c=1}^{k} P_{jc}}. \tag{7.2}$$

The payrolls P_{ic} have been used as weights, in order to take into account the fact that for some classes the volume of data may be radically different by year. The variances are estimated in the same manner. Then as usual the estimated correlations are:

$$\mathrm{Corr}[R_{ic}, R_{jc}] = \mathrm{Cov}[R_{ic}, R_{jc}] / \sqrt{\mathrm{Var}[R_{ic}]\mathrm{Var}[R_{jc}]}. \tag{7.3}$$

For example, Figure 10 shows the observed correlations for the Manufacturing classes with expected annual losses between

[111]Recall that $\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$.

\$300,000 and \$1 million. There are a total of 61 such classes. With 13 separate years of data, we can estimate $(13)(12)/2 = 78$ correlations. These correlations correspond to a separation of between one year and twelve years. We note considerable random fluctuation. Nevertheless, as the separation grows the correlations tend to decline.

7.4. Fitting Regressions to Correlations, Estimating γ, ρ, I, and J

Figure 10 shows the results of fitting a linear regression to the logs of these correlations. The fitted curve is (approximately) $y = (.46)(.94^x)$. The y-intercept is .46, and the decline rate or slope is .94.

Thus, we might estimate for this size of class the decline rate is about .94.[112] In the assumed covariance model this corresponds to some sort of weighted average of γ and ρ, with the weights depending on the size of risk E and the variances g^2 and r^2.

On the other hand the intercept of .46 represents an estimate of the credibility (of a single year of data) in the absence of shifting risk parameters. That is, using Equation 4.15,

$$\frac{E + I}{E(1 + J) + I + K} \approx .46 \qquad \text{for} \quad E \approx \$650{,}000.$$

Similar regressions were fit to the correlations for various size categories. However, in order to improve stability, the correlations for the same separations were first averaged.[113] So for example, the 12 correlations for one year of separation in Figure 10 average to .498. Then a weighted regression was fit to

---

[112] The slope of the log-correlations is about $\ln .94$.

[113] The averaging of the correlations prior to the regression versus time lag is not necessarily the best procedure to employ in this particular application, let alone in general. Ideally one would identify the variables causing the wide dispersion in observed correlations between individual years of data, as seen, for example, in Figure 10. However, I was unable to do so, beyond convincing myself that some substantial portion of this dispersion was a result of the process variance inherent to a data set of this size. While for the illustrative example here the technique used seems sufficient, it would be preferable to find a technique that directly makes use of all the available data. This is a potential area for future research, which could lead to a sharper estimate of the time dependence.

the logs of these average correlations,[114] with weights equal to the number of observed correlations of that separation. For data shown in Figure 10, this would result in a very similar fitted curve:[115] $y = (.46)(.94^x)$.

The results for the Manufacturing industry group, for six size categories with substantial number of classes are:

Expected Annual Losses ($000)	Number of Classes	"Slope"	Intercept
10 to 30	22	1.109	.075
30 to 100	40	.758	.329
100 to 300	37	.979	.375
300 to 1,000	61	.944	.469
1,000 to 3,000	50	.977	.744
3,000 to 10,000	13	.887	.911

The intercepts reflect a general pattern of increasing credibility with size of class, as expected. The smallest and largest size categories have too few classes to reliably estimate correlations.[116] Thus one should not rely on the estimated slopes; the estimated intercepts for these categories are less reliable than those for the other size categories.

For the four size categories with a large number of classes, there is some indication that the "slope" is closer to unity for large classes than for small classes. This data provides a weak indication that the risk parameters of larger classes shift more slowly than those of smaller classes.

The results of fitting regressions to the correlations of two size categories for Goods and Services classes are:

---

[114]If the average correlation was negative as occasionally happened, that separation was not included in the regression.

[115]The curve is the same in this case to the number of decimal places displayed.

[116]Also for the smallest size category, there is a lot of random fluctuation in the pure premiums of the classes.

Expected Annual Losses ($000)	Number of Classes	"Slope"	Intercept
100 to 1,000	38	.938	.605
1,000 to 10,000	38	.994	.837

The same general pattern applies, but with only two size categories we cannot infer much.

As discussed in Section 5.2, we expect the decline rate of the correlations to be approximately $(E\rho + I\gamma)/(E + I)$. Note that this actually applies only when $E \geq \Omega$. For $E \leq \Omega$ the decline rate of the correlations should be approximately $(\Omega\rho + I\gamma)/(\Omega + I)$.[117] In any case, the largest classes should have a decline rate near ρ, while the smaller classes have a decline rate closer to γ.

From the data for these two industry groups,[118] we might estimate that the largest classes have a decline rate for correlations of about .98; thus we might estimate $\rho \approx .98$. The smaller classes might have a decline rate below .90; thus we might estimate $\gamma \approx .85$. Note that ρ corresponds to a half-life of 34 years, while γ corresponds to a half-life of about 4 years. There is clearly a great deal of uncertainty in these estimates.[119]

The midway point at which the decline in the correlations is between ρ and γ is even harder to estimate. As discussed in Section 5.2, we expect this midway point to be at about I. For illustrative purposes estimate this as \$100,000, so that $I \approx \$100{,}000$.

As discussed previously in Section 4.5, the maximum credibility in the absence of shifting risks parameters for one year of data is $1/(1 + J)$. Thus, if J were .1, the intercepts would ap-

[117]For the parameters selected in this section $(\Omega\rho + I\gamma)/(\Omega + I) = ((50{,}000)(.98) + (100{,}000)(.85))/(50{,}000 + 100{,}000) = .89$.

[118]We ignore here the real possibility that the covariance structure might differ significantly among different industry groups, since this data is well short of being able to distinguish if that is the case.

[119]Better estimates would require looking at similar data from a large number of individual states, each of reasonable size.

proach $1/1.1 = .909$ for large risk sizes. While it is unclear from this limited data precisely what that maximum intercept is, it is almost certainly greater than .85. Thus, J is probably .15 or less. In any case, for illustrative purposes $J = .10$ will be used.

7.5. Estimating K

The estimates of J and I, together with the intercepts by size of risk, can be used to estimate the value of K. In the absence of shifting risk parameters, the credibility for a single year of data is given by Equation 4.15:

$$Z = \frac{E+I}{E(1+J)+I+K}, \qquad E \geq \Omega.$$

Thus,

$$K = \left(\frac{1}{Z} - 1\right)(E+I) - (JE). \tag{7.4}$$

Given an estimate of Z from the intercept, for a size E, and the previously estimated $I = \$100{,}000$ and $J = .10$, we can estimate K.

We get the following estimates:

Industry Group	Size[120] (000)	Intercept	Estimated K (\$000)
Manufacturing	20	.075	1,478
Manufacturing	65	.329	330
Manufacturing	200	.375	480
Manufacturing	650	.469	784
Manufacturing	2,000	.744	523
Manufacturing	6,500	.911	−5
Goods & Services	550	.605	369
Goods & Services	5,500	.837	541

Recall that for Manufacturing the smallest and largest size categories really do not contain enough classes to adequately

[120] Based on the midpoint of the size category.

quantify the intercept. In any case, for the largest size category, the estimate of K is extremely sensitive to the selection of J. For the smallest two size categories, Equation 5.9 rather than Equation 5.8 is likely to hold, since $E \leq \Omega$; thus the above estimate of K using the two smallest categories is likely to be invalid. Averaging the middle three size categories for Manufacturing plus the two size categories from Goods and Services, we get $K \approx \$500{,}000$. This value of K will be used for illustrative purposes.

7.6. Selecting Ω

Finally, we must select Ω, the value below which the classes are homogeneous; i.e., there is no significant impact from risk heterogeneity below size Ω. Conceptually, this is the size at which a class is likely to be made up of one significant sized employer.[121] On the other hand, it was seen before that choosing Ω somewhere close to I produces a smooth decline in credibilities.

For illustrative purposes choose $\Omega = \$50{,}000$. This corresponds for this data set to somewhere between 50 and 75 full-time employees.[122]

In the absence of shifting risk parameters over time, Equation 4.25 gives the credibility for one year of data as:

$$Z = \frac{E}{(1+J')E + K'}, \qquad E \leq \Omega = \$50{,}000$$

where

$$J' = J\left(\frac{\Omega}{I+\Omega}\right) = (.10)\left(\frac{50}{150}\right) = .033 \qquad \text{and}$$

[121] While situations where the data for a class comes from one significant employer are not common, they do occur.

[122] Assuming reported losses (at unit statistical plan level) of about 2.5% of payrolls and a State Average Weekly Wage of about \$600, 65 full-time employees have \$50,700 in expected annual losses.

$$K' = K\left(\frac{\Omega}{I+\Omega}\right) = 500{,}000\left(\frac{50}{150}\right) = \$166{,}667.$$

For $E = \$20{,}000$,

$$Z = \frac{20}{((1.033)(20) + 166.667)} = \frac{20}{187.3} = 10.7\%.$$

This compares to the estimated intercept of .075. Given the uncertainty of the estimated parameters, the uncertainty of the estimated intercept, and the approximate nature of the regression relation itself, these values of .107 and .075 are not inconsistent. Getting a somewhat more precise estimate of Ω would require analyzing data from many states over many years.

With all these caveats, we have estimated the essential features of the covariances. Equation 5.10 states for $\sqrt{E_iE_j} \geq \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{|i-j|} + I\gamma^{|i-j|}/\sqrt{E_iE_j} + \delta_{ij}\left(K/\sqrt{E_iE_j} + J\right)\right\}, \quad \sqrt{E_iE_j} \geq \Omega.$$

Similarly Equation 5.11 states that for $\sqrt{E_iE_j} \leq \Omega$:

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{|i-j|} + I\gamma^{|i-j|}/\Omega + \delta_{ij}\left(K/\sqrt{E_iE_j} + J\right)\right\}, \quad \sqrt{E_iE_j} \leq \Omega.$$

In both cases there is a factor of r^2 that multiplies the covariances that does not affect the credibilities.

7.7. Illustrative Credibilities, No Weight to Overall Mean

We can use Equations 5.10 and 5.11 together with the values of the parameters estimated in the previous section to estimate the covariances. These in turn can be used to estimate the credibilities using Equations 6.7 (for the case where no weight is being given to the mean).

The following illustrative values will be used to calculate credibilities:

$\rho = .98$	(rate of shifting parameters related to class homogeneity),
$\gamma = .85$	(rate of shifting parameters related to class heterogeneity),
$I = \$100{,}000$	(related to class heterogeneity),
$J = .10$	(related to parameter uncertainty),
$K = \$500{,}000$	(Bühlmann credibility parameter, related to process variance), and
$\Omega = \$50{,}000$	(size limit for class homogeneity).

For example, for years 1, 2, 3 and 4 being used to predict year 8, with each year of data having \$1 million in expected losses, Equations 6.7 become:

$$1.7Z_1 + 1.065Z_2 + 1.0327Z_3 + 1.0026Z_4 = .9002 + \lambda/2,$$
$$1.065Z_1 + 1.7Z_2 + 1.065Z_3 + 1.0327Z_4 = .9236 + \lambda/2,$$
$$1.0327Z_1 + 1.065Z_2 + 1.7Z_3 + 1.065Z_4 = .9483 + \lambda/2,$$
$$1.0026Z_1 + 1.0327Z_2 + 1.065Z_3 + 1.7Z_4 = .9746 + \lambda/2,$$
$$\text{and} \qquad Z_1 + Z_2 + Z_3 + Z_4 = 1,$$

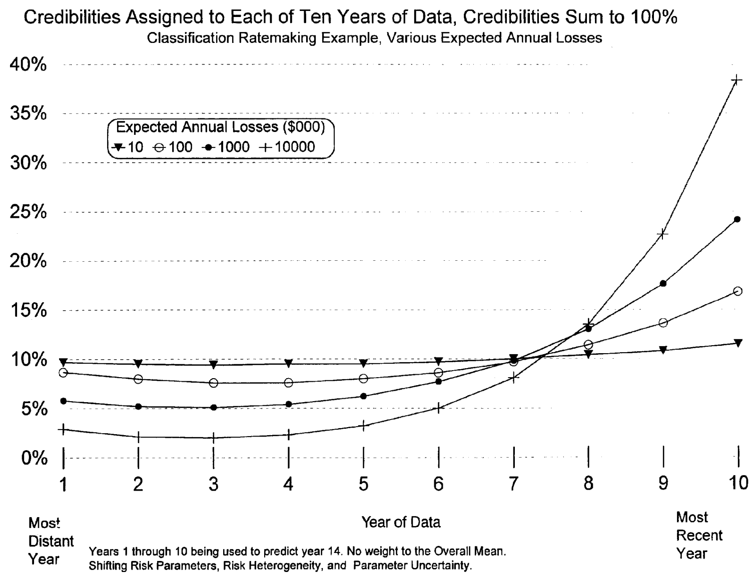
with solution:[123]

$$Z_1 = 21.08\%,$$
$$Z_2 = 21.98\%,$$
$$Z_3 = 25.34\%, \qquad \text{and}$$
$$Z_4 = 31.60\%.$$

Note that since no weight is given to the overall mean, the credibilities have been constrained to add up to 100%.

[123]The Lagrange Multiplier is .5416.

FIGURE 11



The credibilities assigned to ten individual years are shown in Figure 11 for various size classes, for years $1, 2, \ldots, 10$ being used to predict year 14. Note here it has been assumed that the credibilities are constrained to add to unity. Thus, the default weight is 10% to each of the ten years. However, as the classes get bigger and bigger we can make the estimation process more responsive and give more weight to the more recent data.[124] For \$10 million dollars in expected annual losses the most recent year gets about 38% of the weight. For small classes, we must use a

[124] The most distant year gets a slight amount of extra weight, due to the "edge effect." Year 1 contains valuable information about Year 0 due to the fact that they are correlated. Therefore, by giving a little more weight to Year 1, one gets some of the same benefit as if Year 0 were in the database. While, for example, Year 3 contains valuable information about Year 2 and Year 4, Years 2 and 4 are already in the database. In general, with shifting risk parameters over time, the most distant year(s) should receive somewhat more weight, due to this edge effect, than they would otherwise receive. In Figure 11 for \$10 million in Expected Annual Losses, as one goes to more distant years, at the edge the graph of credibilities bends slightly upwards rather than continuing to decline.

more stable method and give every available year of data significant weight. However, for small classes the parameters shift more quickly and thus there is a counter-balancing tendency to weight these older years less than more recent years. Nevertheless, for \$10,000 in annual expected losses the weights are all about 10%.

7.8. Illustrative Credibilities, Weight to Overall Mean

We can use Equations 5.10 and 5.11 together with the values of the parameters listed in Section 7.7 to estimate the covariances. These in turn can be used to estimate the credibilities using Equations 2.4, for the case where the complement of credibility is being given to the overall mean.

Assuming years $1,2,\ldots,10$ are being used to predict year 14, the credibilities assigned to the given years are shown in Figure 12. Larger sizes give more weight to recent years as well as more total credibility. Figure 13 shows the sums of the credibilities assigned to different classes. For ten years of data, the larger size classes are assigned up to 90% credibility.[125] The credibility goes to zero as the size of class goes to zero.[126] Also shown are the results for three years and one year of data.

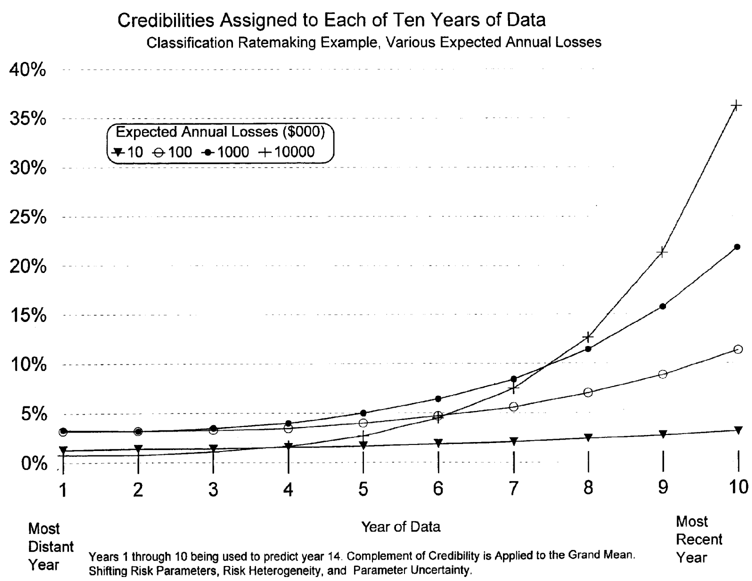
The class (expected) pure premiums within an industry group can easily vary by a factor of ten from lowest to highest. Thus, the average industry group pure premium, or equivalently a relativity of unity, is not a very good predictor for most classes. Therefore, the credibilities assigned to the classification data are relatively large. Assigning the complement of credibility to the average pure premium for the industry group, as illustrated here, is not generally done in practice.

---

[125]Without shifting risk parameters, the maximum credibility would be $Y/(Y+J) = 10/10.1 = 99\%$. With 10 years of data and $J = .1$, the effects of parameter uncertainty are not very significant.

[126]Since we've assumed no risk heterogeneity below size Ω.

FIGURE 12



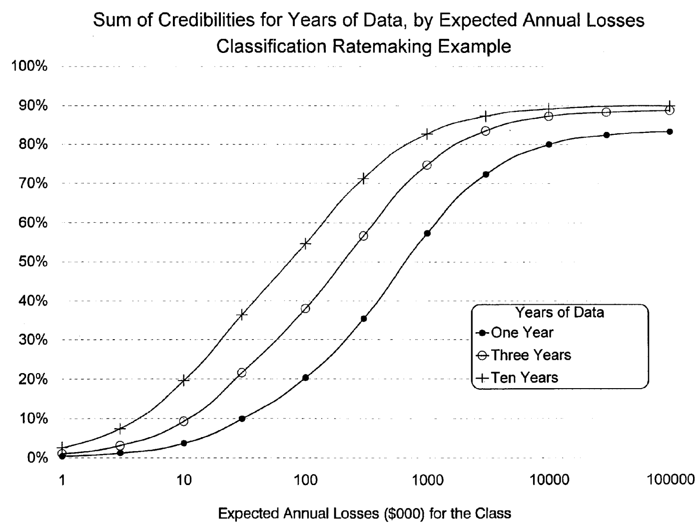
An alternative would be to work with loss ratios to premiums at current rates, as is done in Meyers [25]. Then the complement of credibility is given to the loss ratio for the industry group;[127] i.e., each class rate is changed by the industry group average rate change. This follows the general practice and is equivalent to giving the complement of credibility to the prior estimated relativity for each class.

7.9. Using Prior Estimates of Relativities

Assume that we have been estimating classification relativities for a long time. Then we might weight together the estimated relativity for each class based on the most recent data and the

[127]Meyers does not appear to divide the classes into industry groups. However, the technique presented could be applied equally well to industry groups. We would have to take a little more care in estimating the Bühlmann credibility parameters.

FIGURE 13



For example, Years 1 through 10 being used to predict year 14. Complement of Credibility is Applied to the Grand Mean. Shifting Risk Parameters, Risk Heterogeneity, and Parameter Uncertainty.

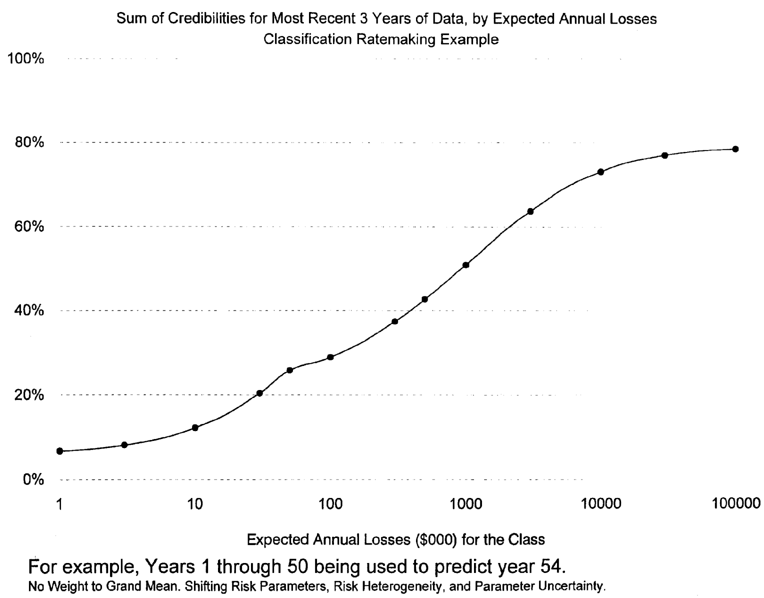
prior estimate of the relativity for that class. The issue is how much weight to apply to each of these estimates.

While there are other ways to think of this problem, we could fit it into the current framework by assuming some very long series of data, for example 50 years.[128] Then as in Section 7.7, we can compute the credibility to be assigned to each of these 50 years of data (with no weight to the overall mean). If three years of recent data are being used, then we can assign as the weight to the prior estimate the sum of the credibilities for the 47 less recent years.

For example, using the values from Section 7.7, for \$1 million in expected annual losses, for years $1, 2, \ldots, 50$ being used to

[128] In the case of a workers compensation rating bureau, classification relativities have been estimated for about 80 years.

FIGURE 14



predict year 54, years 48, 49, and 50 have credibilities of: 11.8%, 16.3% and 22.8%. The prior estimate would be assigned a weight of 100% – (11.8% + 16.3% + 22.8%) = 49.1%.

Figure 14 shows the weight assigned to the most recent three years of data as the expected annual losses vary. The recent data for large classes gets less than 100% credibility; both the prior estimate and that from the recent data are assumed to be good estimators for large classes. The recent data for small classes gets considerable credibility; the prior estimate as well as that from the recent data are assumed to be poor estimators for small classes.

Note that the credibility curve in Figure 14 has a discontinuous derivative at the point $\Omega = 50{,}000$. This will be typical as we switch from Equations 5.10 for the covariances to Equations

5.11, as we go from a size where risks are generally heterogeneous to one where risks are generally homogeneous.[129]

7.10. General Effect of Differences in the Maturity of the Data

Conceptually the goal has been to estimate the expected future class relativity at *ultimate* report. Assume, as in Figure 11, we were predicting year 14 using data from years 1 to 10. Then we expect that year 1 at 10th report would be a better predictor of year 14 at ultimate than would year 1 at 5th report. Year 1 at 5th report is in turn a better predictor than year 1 at 1st report. Generally, the more mature the data from a single given year the better predictor of the future ultimate losses we expect it to be.[130]

Thus, actuaries will usually rely upon the latest *available* report for each year of data. In the case of the workers compensation classification example, we would have years 1 to 6 at fifth report,[131] year 7 at fourth report, year 8 at third report, year 9 at second report, and year 10 at first report.

In the example in Section 7.7, there is no weight to the overall mean; the credibilities assigned to the data sum to 100%. Thus in that situation, the credibilities reflect how good an estimator each year is *relative* to the others. If each of the ten years of data were at the same report, their relative value as estimators would be unaffected by maturity.

However, year 10 is only at first report while years 1 through 6 are at fifth report. Therefore, the tenth year of data is a poorer estimator relative to the other years than if it were available at fifth report. Thus, we should give year 10 somewhat less cred-

[129]In the model this switch is abrupt, leading to the discontinuous derivative of the credibility. While we could refine the model to make this derivative continuous, this would seem to be unlikely to have any practical significance.

[130]Thus, there is a dilemma. We prefer more recent years of data in order to minimize the impact of shifting risk parameters, but we also prefer more mature data. Section 9 discusses this from the point of view of an overall rate indication.

[131]Usually workers compensation classification data is only collected up to fifth report.

TABLE 4

CORRELATIONS BETWEEN REPORTS OF CLASSIFICATION RELATIVITIES

Various Size Classes, Based on Annual Expected Losses ($000)

	10 to 30 (19 classes)				30 to 100 (28 classes)			
	2nd	3rd	4th	5th	2nd	3rd	4th	5th
1st	.783	.757	.914	.834	.850	.754	.656	.624
2nd		.978	.849	.805		.860	.745	.730
3rd			.835	.783			.839	.809
4th				.949				.898
	100 to 300 (39 classes)				300 to 1,000 (52 classes)			
	2nd	3rd	4th	5th	2nd	3rd	4th	5th
1st	.863	.814	.822	.800	.879	.830	.809	.799
2nd		.935	.945	.917		.968	.945	.927
3rd			.957	.929			.980	.964
4th				.975				.975
	1,000 to 3,000 (49 classes)				3,000 to 10,000 (22 classes)			
	2nd	3rd	4th	5th	2nd	3rd	4th	5th
1st	.955	.924	.902	.884	.970	.964	.947	.939
2nd		.962	.946	.932		.980	.971	.965
3rd			.977	.965			.988	.977
4th				.986				.992

For each of five composite policy years, 84/85, 85/86, 86/87, 87/88 and 88/89, class relativities were calculated for the Manufacturing industry group. Then for each year, for classes in a given size category, correlations were calculated between the relativities at two different reports. The correlation matrices displayed here are an average of the five separate correlation matrices, one from each year.

ibility than was calculated in Section 7.7, while other years are assigned somewhat more credibility.

7.11. Correlations Between Differing Maturities

This effect of the differing maturities of data will be estimated by examining the correlations between class relative pure premiums from the same year of data but at different maturities. These correlations are calculated using Equation 7.2, where a difference in maturity is substituted for a difference in year. Table 4 displays correlation matrices for various size categories of classes from the Manufacturing industry group.[132] So, for example, for classes with expected annual losses between $300,000 and $1 million, the correlation between class relativities calculated from the same year of data at second report and fourth report is .945. In contrast, that between first and fifth report is .799. As expected, since more development occurs between first and fifth reports than between second and fourth report, the classification relativities are less highly correlated.[133]

In general, it is expected that the more loss development between two reports, the smaller the correlation of the relativities. The observed loss development factors (LDFs) were:[134]

1st to 2nd	1.249
2nd to 3rd	1.123
3rd to 4th	1.059
4th to 5th	1.040

Also, we expect that the relativities for smaller classes will be more affected by the random fluctuations caused by loss development. Therefore, the smaller the size category, the smaller the correlation of the relativities for different reports.

The simplest type of model would be one in which the correlation was some linear function of the class size and the loss development factor between reports. Since I was unable to find a useful model of that type, instead I first took the log of both the loss development factor and the correlation. Then I examined linear models involving the ln (correlation), ln (LDF), and size of class.

[132]Each correlation matrix is the average of five correlation matrices calculated for composite policy years 84/85, 85/86, 86/87, 87/88 and 88/89. A composite policy year includes July 1 to June 30.

[133]First and fifth report are further apart so that their correlation has more opportunity to decline from unity.

[134]For the Manufacturing industry group, for composite policy years 84/85, 85/86, 86/87, 87/88 and 88/89. All data was included independent of the size of class. Recall that the losses are paid plus case reserves.

One model of the relation of the correlations to the development that will have the desired properties is:[135]

$$\begin{aligned}\ln(\text{Correlation}) &= -\ln(\text{LDF})/(\text{Linear Function of Size})\\ -\ln(\text{LDF})/\ln(\text{Correlation}) &\\ &= \text{Linear Function of Size.} \qquad (7.5)\end{aligned}$$

This model has the desired property that the correlation is 1 when the LDF is 1.[136] If the right hand side of Equation 7.5 is positive, then the correlation decreases as the amount of development increases. If the right hand side of the Equation 7.5 increases with size of class, then as desired the correlations will be closer to unity for larger classes.

For the data in Table 4, we can compute the ratio of the $-\ln(\text{LDF})/\ln(\text{correlation})$. For example, for the second to fourth report the LDF is $(1.123)(1.059) = 1.189$. For the size category \$300,000 to \$1 million in expected annual losses, the correlation between 2nd and 4th report is .945. Thus, $-\ln(\text{LDF})/\ln(\text{correlation}) = (-\ln(1.189)/\ln(.945)) = 3.06$.

Averaging over each correlation matrix we obtain by size of risk:[137]

Size (\$ million)	$-\ln(\text{LDF})/\ln(\text{correlation})$
.02	1.76
.065	.76
.2	1.88
.65	2.38
2.0	3.35
6.5	6.20

[135]For a given size of risk this model assumes the correlations decline as per a constant to the power of the "effective time" between reports. The "effective time" between reports is taken as the logarithm of the loss development factors.

[136]In lines of insurance where salvage and subrogation are significant, the loss development factor can be less than unity. Equation 7.5 would not apply.

[137]Expected Annual Losses for the midpoint of the size category.

A least squares linear regression would give

$$1.57 + .73\ (\text{Size}/1\ \text{million}).$$

Let,

$$c = -\ln(\text{LDF})/\ln(\text{correlation}).$$

Then,

$$\text{correlation} = \left(\frac{1}{\text{LDF}}\right)^{1/c}.$$

If we take for illustrative purposes:

$$c = 1.5 + .75\ (\text{Size}/1\ \text{million}),$$

then substituting into Equation 7.5 gives

$$\text{correlation} = 1/\text{LDF}^{1/(1.5+.75\ (\text{Size}/1\ \text{million}))}. \tag{7.6}$$

For example, for the size category \$300,000 to \$1 million if we take a size of \$650,000 equivalent to the midpoint, then Equation 7.6 gives an estimated correlation of $1/\text{LDF}^{.5}$. For example, for the 2nd to 4th report the LDF is 1.189. Thus, for this size category the model correlation is about .91. (The observed correlation is .945.)

Table 5 displays the model correlations between reports for classes of various sizes. While the particular model represented by Equation 7.6 should be taken as solely for illustrative purposes, the general pattern of correlations in Table 5 is what we would expect. For a given report interval, the larger the class the higher the correlation. For a given size of class, the more development in a report interval, the lower the correlation. This pattern of correlations can be incorporated into the calculation of credibilities.

7.12. Credibilities Taking Into Account Differing Maturities

Returning to the example in Section 7.7, we can incorporate the impact of the differences in maturity. Given years 1 through 6 at fifth report, year 7 at fourth report, year 8 at third report, year

TABLE 5

CLASSIFICATION RATE RELATIVITIES
MODEL CORRELATIONS BETWEEN REPORTS

	Expected Annual Losses ($000)					
Reports	20	65	200	650	2,000	6,500
1 vs. 2	.864	.866	.874	.894	.929	.966
1 vs. 3	.800	.804	.815	.843	.893	.948
1 vs. 4	.770	.775	.787	.819	.876	.940
1 vs. 5	.750	.755	.768	.803	.865	.934
2 vs. 3	.926	.928	.932	.943	.962	.982
2 vs. 4	.892	.894	.900	.916	.944	.973
2 vs. 5	.869	.872	.879	.899	.932	.967
3 vs. 4	.963	.964	.966	.972	.981	.991
3 vs. 5	.938	.940	.943	.953	.968	.985
4 vs. 5	.974	.975	.977	.980	.987	.994

9 at second report and year 10 at first report, we try to predict year 14 at fifth report.

For a class with expected annual losses of $1 million, Equation 7.6 estimates the correlation between classification relativities at different reports as $1/\mathrm{LDF}^{.444}$. For 2nd to 4th report, the estimated correlation is $(1.189)^{-.444} = .926$. Prior to taking into account the differences in maturity, the model covariance between year 7 and year 9 was 1.033.[138] It will be estimated that the covariance between year 7 at 4th report and year 9 at second report will be lower by a factor of the correlation[139] .926; $(.926)(1.033) = .957$.

The other model covariances involving at least one year of data at prior to 5th report are similarly adjusted.[140] (The vari-

[138]For $r^2 = 1$, $1 million in expected annual losses, and the parameters in Section 7.7.

[139]This is an approximation based on an assumption that the impact of maturity is largely independent of the other factors previously considered.

[140]For purposes of adjustment it was assumed Year 14, the year to be predicted, was at 5th report. If one assumed instead for example 20th report, all the covariances involving

ances along the diagonal of the variance-covariance matrix are unaffected.) The least squares credibilities differ from those obtained in Section 7.7:[141]

CREDIBILITY FOR CLASS WITH \$1 MILLION IN EXPECTED ANNUAL LOSSES

Year, Report	Section 7.7	Taking into Account Differences in Maturity
1@5th	5.8%	6.7%
2@5th	5.2	6.2
3@5th	5.1	6.4
4@5th	5.4	7.1
5@5th	6.2	8.6
6@5th	7.7	10.8
7@4th	9.8	11.5
8@3rd	13.0	12.7
9@2nd	17.6	13.9
10@1st	24.2	16.0

As expected, more mature years of data are given more credibility than previously while less mature years receive less. For example, the data from year 10 at first report gets 16.0% credibility compared to the 24.2% credibility calculated in Section 7.7.

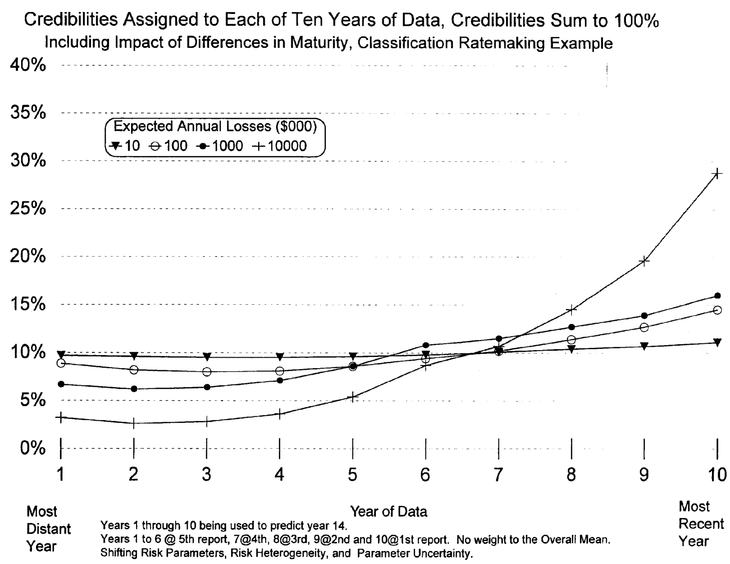
Figure 15 displays the credibilities for other size classes. The credibilities shown in Figure 15 that take into account differences in maturity can be compared to those in Figure 11, which ignore these differences. While the precise impact depends on the particular amount of loss development and the particular model used to estimate the correlations, the general pattern displayed here should occur in most situations.

The weights which would otherwise be given to immature years of data should decrease significantly for larger size classes.

Year 14 would be lower by the same factor but the resulting credibilities would all be the same, since they've been constrained to sum to 100%.

[141] As shown in Figure 11.

FIGURE 15



For smaller classes the weights assigned to recent years are already close to the default weight, in this case 10%, so taking into account their immaturity only produces a small decrease in the weight they would otherwise receive. In all cases, the more mature years of data receive more weight than when we ignored maturity. In this example, the largest increase in weight occurs for year 6, which is the most recent year which is available at "ultimate."[142]

So while taking into account shifting risk parameters over time tends to give more weight to recent years, taking into account the difference in maturity tends to counterbalance that tendency somewhat.[143] This will be true for overall ratemak-

---

[142]In this example, fifth report is the ultimate report actually received of Unit Statistical Plan data, even though there is loss development beyond fifth report.

[143]An example is given in Section 9 in which the most recent year of data is so immature it is given very little weight.

ing and experience rating as well as for classification ratemaking.

8. USE OF DATA FROM OTHER STATES

In estimating the classification relativities in a given state one may supplement the data from that state with data from other states, as in Harwayne [24].

8.1. Use of Data from One Other State

As a simple example, assume we are estimating Massachusetts relativities and will use New York experience in addition to that from Massachusetts. The key assumption is that the underlying expected class relativities in New York are similar to those in Massachusetts. Thus, observed relativities in New York are useful for predicting future relativities in Massachusetts. However, all other things being equal, a given volume of New York data is assumed to be less useful in predicting Massachusetts relativities than would be similar data from Massachusetts.[144] Thus, we expect that in this case the credibilities assigned to a given volume of data will be less for New York data than for Massachusetts data.

There are three steps to calculating the credibilities to assign to the years of data from Massachusetts and New York. First, we must model the covariance structures. Second, we must estimate the parameters in the covariance structures. Third, we must use these covariances together with the appropriate set of linear equations, in this case Equations 8.1, in order to solve for the credibilities. In this case, the first two steps will build on the results on classification relativities obtained in Section 7.

[144] Similarly, New York data would be more useful for predicting New York relativities than would data from Massachusetts.

8.2. Covariance Structure, Use of Data From One Other State

There are three types of variance-covariance matrices. The first type involves covariances between data from Massachusetts: S_{ij} = covariances within Massachusetts. The illustrative values from Section 7.7 will be used for these covariances. The second type involves covariances between data from New York: T_{ij} = covariances within New York. For a given volume of data, assume a similar covariance structure within New York to that estimated for Massachusetts; the illustrative values from Section 7.7 will therefore be used for the covariances within New York, T_{ij}.

The third type of covariance is that involving data from Massachusetts versus data from New York: U_{ij} = covariances between Massachusetts and New York. It is expected that for a given volume of data, the correlation of relativities between states is less than the correlation of relativities within states. This is what is observed.

8.3. Estimating Parameters, Between State Covariances

Classification data for Massachusetts, New York and several other large states was examined as discussed in Appendix F. Correlations of classification relativities between states were calculated for classes in various size categories for both the Manufacturing and the Goods and Services industry groups.

Based on the analysis discussed in Appendix F, with three exceptions the same parameters will be used for the interstate and intrastate covariances. The K parameter, related to the expected value of the process variance, will be zero for the interstate covariances. The J parameter, related to parameter uncertainty, will be selected for the interstate covariances as half of the intrastate J parameter.[145]

[145]The credibilities are relatively insensitive to this choice.

The r^2 parameter, setting the scale for the covariances, will be taken for the interstate covariances as 70% of its value for the intrastate covariances.[146] This will result in correlations of relativities between states that are lower than those within a state, all else being equal.

For the covariances the following inputs are used:[147]

	Intrastate	Interstate
r^2	1	.7
ρ	.98	.98
γ	.85	.85
I	100,000	100,000
J	.10	.05
K	500,000	0
Ω	50,000	50,000

8.4. *Equations for Credibilities, One Other State*

Assume we are trying to estimate class relativities in Massachusetts, without any weight to the overall mean. Let Z_i be the weight applied to the Massachusetts data and let W_i be the weight applied to the New York data. Then $\Sigma Z_i + \Sigma W_i = 1$, since there is no weight given to the overall mean. As shown in Appendix E, if we use Y years of data from each state, in order to predict year $Y + \Delta$, we obtain $2Y + 1$ equations in $2Y + 1$ unknowns:[148]

[146] The relative size of the interstate and intrastate covariances affects the calculation of credibilities. However, there is still an arbitrary choice of overall scale which does not affect the credibilities.

[147] The r^2 values contain an arbitrary scale factor. Since it is only their relationship that affects the credibilities, the actual r^2 values have not been estimated. Unlike Section 7.12, no adjustment is made for differing maturities here. Such an adjustment in the case of more than one state would parallel that for a single state as shown in Sections 7.11 and 7.12.

[148] There are YZ's, YW's, plus the Lagrange Multiplier λ. The equations would be somewhat different if the years for which we have Massachusetts data and New York data are not the same. Appendix E gives an example.

$$\sum_j Z_j S_{ij} + \sum_j W_j U_{ij} = \frac{\lambda}{2} + S_{i,Y+\Delta}, \qquad i = 1,2,\ldots Y,$$

$$\sum_i Z_i U_{ij} + \sum_i W_i T_{ij} = \frac{\lambda}{2} + U_{Y+\Delta,j}, \qquad j = 1,2,\ldots Y, \qquad (8.1)$$

$$\sum_i Z_i + \sum_j W_j = 1,$$

where the covariance matrices are:

S_{ij} = covariances within Massachusetts,[149]

T_{ij} = covariances within New York,[150]

U_{ij} = covariances between Massachusetts and New York.[151]

8.5. *Illustrative Credibilities, One Other State*

For example, assume we are estimating year 54 class relativities in Massachusetts using data from years 1 to 50, with \$1 million of expected annual losses in Massachusetts and \$5 million of expected annual losses in New York. Then using Equations 8.1 the most recent three years of Massachusetts data would be given credibilities of 9.7%, 13.3% and 18.6%, while the three most recent years of New York data would be given credibilities of 2.5%, 7.0% and 15.3%. We could give the prior estimate the remaining weight of 33.6%.

8.6. *Using Data From Several Other States*

This example where the data from one outside state is used can be extended to one where data is used from several other states. Assume for simplicity that "countrywide data" is from

---

[149]More generally within the state for which we are trying to estimate class relativities.
[150]More generally within the supplementary data from outside the state of interest.
[151]More generally between the data from the state of interest and the data from outside the state of interest.

ten states, other than the state for which we are estimating class relativities.

Let C be the covariance matrix within states, while D is the covariance matrix between states.[152] Assume for simplicity that for a given volume of data, C is the same in each state and D is the same for each pair of states. Assume that the "countrywide data" is the average of data from ten states, each with a volume of data $\tilde{E}/10$. Let the covariance matrix between two non-Massachusetts states be D' and the covariance within a non-Massachusetts state be C'. Then the covariance between the countrywide data is the sum of 100 terms, 90 of which are between states, D', and 10 of which are within states, C'.[153] The covariance between the countrywide data is therefore $(90D' + 10C')/100 = .9D' + .1C'$. In general, if we had data from n other states each of the same size, the covariance between the countrywide data would be $((n-1)D' + C')/n$.

We have assumed $D' < C'$, so that $.9D' + .1C' < C'$. Due to the lack of homogeneity of the countrywide data, its covariance is less than that for an equivalent volume of data all from a single state.

The covariance of the countrywide data[154] with Massachusetts is just the average of ten similar terms all involving the covariance between the states.[155] Thus, the covariance between Massachusetts and the countrywide data is D.

In summary, for C and D calculated for the appropriate volumes of data for the state(s) involved:

S_{ij} = covariances within Massachusetts = C,

[152]Both C and D are a function of the volume of data in the state(s).

[153]$\mathrm{Cov}[\frac{1}{10}\{Y_1 + Y_2 + \cdots + Y_{10}\}, \frac{1}{10}\{Y_1 + Y_2 + \cdots + Y_{10}\}] = \frac{1}{100}\{\mathrm{Cov}[Y_1,Y_1] + \mathrm{Cov}[Y_1,Y_2] + \mathrm{Cov}[Y_1,Y_3] + \cdots \mathrm{Cov}[Y_{10},Y_{10}]\}$.

[154]The state of interest, in this case Massachusetts, is assumed to be excluded from the countrywide data.

[155]$\mathrm{Cov}[X,(Y_1 + Y_2 + \cdots + Y_{10})/10] = \frac{1}{10}\{\mathrm{Cov}[X,Y_1] + \mathrm{Cov}[X,Y_2] + \cdots + \mathrm{Cov}[X,Y_{10}]\}$.

T_{ij} = covariances within Countrywide[156] $= .1C' + .9D'$,

U_{ij} = covariances between Massachusetts and Countrywide $= D$.

8.7. *Illustrative Credibilities, Data From Several Other States*

These covariances can then be used in Equations 8.1, in order to solve for the credibilities. For example, assume we are estimating year 54 class relativities in Massachusetts using data from years 1 to 50, with \$1 million of expected annual losses in Massachusetts and \$1 million of expected annual losses in each of ten other states. Then using Equations 8.1, the most recent three years of Massachusetts data would be given credibilities of 8.5%, 11.0% and 14.9%. The most recent three years of countrywide data would be given credibilities of 1.8%, 9.8% and 28.5%. The remaining weight of 25.5% could be given the prior estimate of the class relativity.[157]

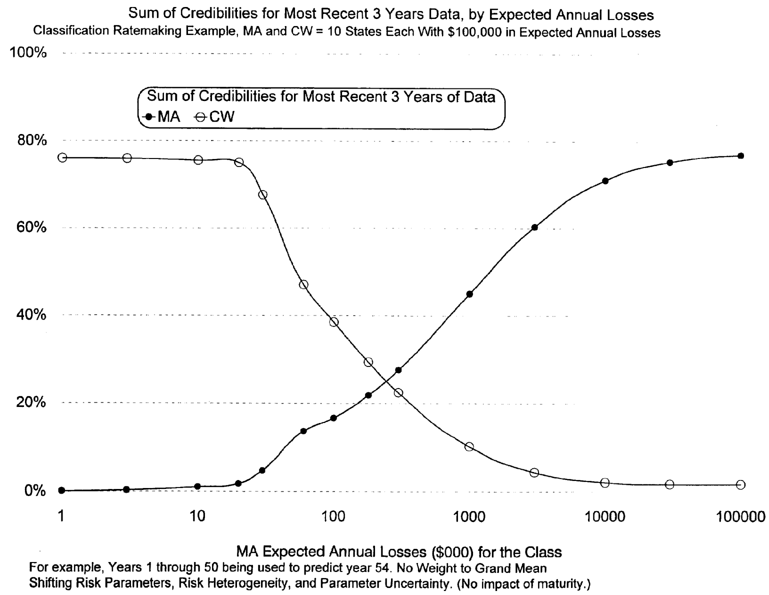
Figure 16 shows for a fixed amount of countrywide data, how the credibilities vary as the volume of data in Massachusetts changes. Since in Figure 16 there is assumed to be \$100,000 in expected annual losses in each of ten states other than Massachusetts, there is sufficient countrywide data to get a reasonable estimate of the class relativity. When there is very little Massachusetts data, for example \$3,000 in expected annual losses, then the most recent three years of Massachusetts data are given virtually no weight,[158] while the most recent three years of coun-

[156]This is for the case where "countrywide" data consists of 10 equal sized states. In general, the covariance of countrywide data will be some mixture of C and D covariance matrices.

[157]It should be noted that for this case, many older years of countrywide data are given negative weight. As a practical matter these weights could be set equal to zero and the weights given to more recent years of countrywide data could be reduced accordingly. This would increase the weight given to the prior estimate.

[158]This is in contrast to Figure 13 where, in the absence of the use of countrywide data, the Massachusetts data was given small but significant weight.

FIGURE 16

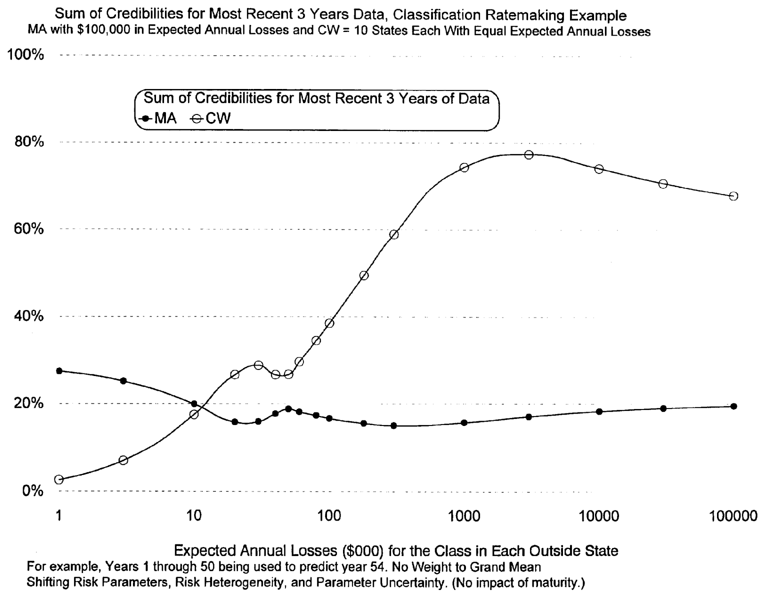


trywide data are given a weight of about 75%.[159] As the volume of Massachusetts data increases, while the volume of countrywide data remains the same, the weight assigned to the most recent three years of Massachusetts data increases up to about 75%, while that assigned to the countrywide data declines to zero.

Figure 17 displays the credibilities assigned to the most recent three years of data, for a fixed amount of Massachusetts data while the volume of countrywide data varies. As the volume of countrywide data increases, the credibility assigned to the most recent three years of countrywide data increases nonmonotonically to about 75%. The credibility assigned to the latest three years of Massachusetts data (with $100,000 in expected

[159] The remaining weight is given to the prior estimate.

FIGURE 17

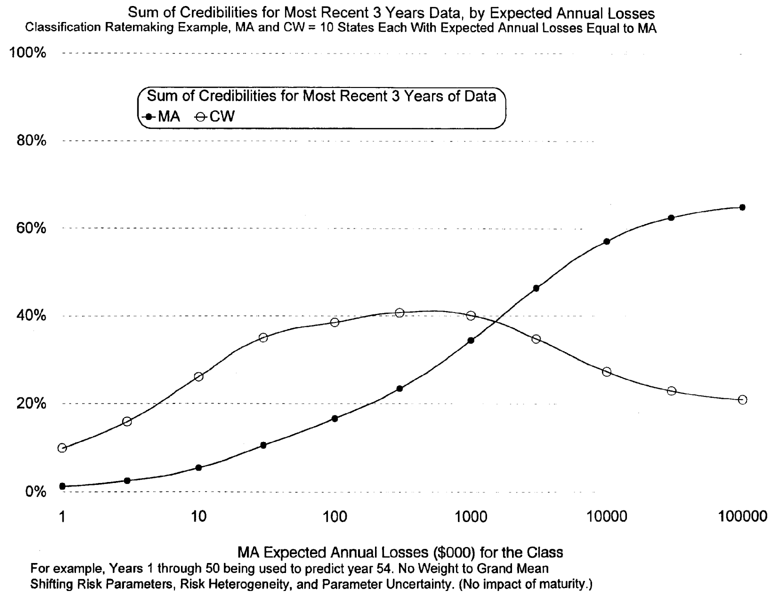


annual losses) varies between 27% and 15% as the volume of countrywide data varies.

Figure 18 displays the credibilities if Massachusetts and each of ten other states all have the same expected annual losses for a given class. As the size of class increases, the sum of the credibilities given to the most recent three years of Massachusetts data increases to about 65%. As the size of class increases, the sum of the credibilities given to the most recent three years of countrywide data increases and then decreases, as for very large classes the Massachusetts data is given more weight.

This behavior means that no simple formula for the amount of credibility given to the countrywide data will be appropriate. We must know how much data is available within the state of interest, before we know how much credibility to assign to the

FIGURE 18



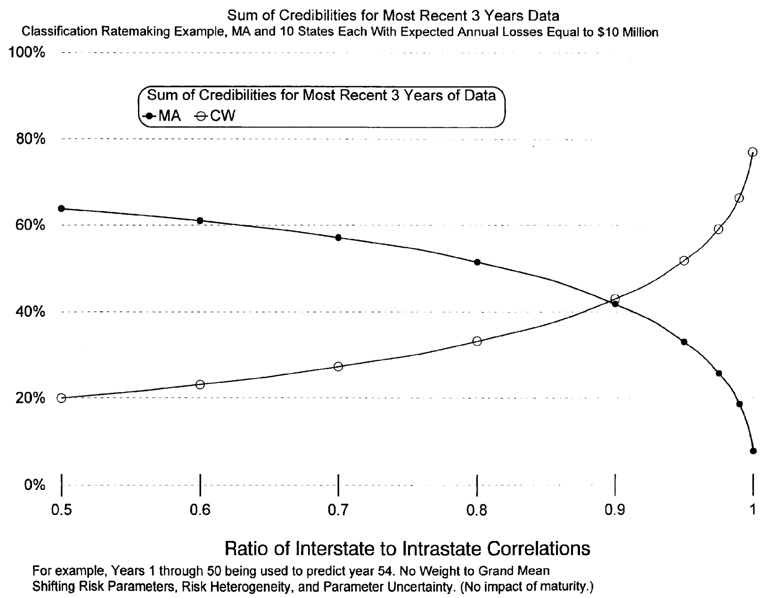
countrywide data.[160] If a simple formula such as the "square root rule" or the "Bühlmann credibility formula" were to be applied based solely on the volume of countrywide data, it would have to be supplemented by some other restriction on the credibility assigned to countrywide data. One commonly used rule of thumb is to restrict the credibility assigned to the countrywide data to be no more than:

$$(\tfrac{1}{2})(1 - \text{credibility assigned to the state data}).$$

Figure 19 displays the sensitivity of the credibilities to the selected ratio of the interstate correlations to the intrastate correlations. For values of this ratio close to the selected value of 70%, the credibilities are relatively insensitive. Note that if the

[160]The reverse is also true, but the credibility of the Massachusetts data is less sensitive to the amount of countrywide data, as seen in Figure 16.

FIGURE 19



intrastate and interstate correlations were equal, then each outside state would get the same 7.7% credibility[161] as would Massachusetts.

9. A RATEMAKING EXAMPLE

This section will illustrate how the ideas in this paper might be applied to the calculation of an overall rate indication.[162] The issue explored here is how much weight should be given to different years of data. This example will illustrate how adjustments to the data for trend, development, etc. will affect the optimal weights.

Assume that for a given line of insurance the six most recent years of data are being combined in order to calculate a rate

[161] For a sum of 77% for ten outside states.

[162] This is an expansion of an example in Mahler [20].

indication. Specifically, assume we have loss ratios[163] from policy years[164] 1991, 1992, 1993, 1994, 1995 and 1996, all as of 12/31/96, which will be used to get a rate indication for policy year 1998.

9.1. Estimation Errors Due to Adjustments to Data

It is assumed that appropriate adjustments have been made to each year's data for development, trend, law changes, changes in deductibles, etc.[165] The necessity of these adjustments introduces estimation error into the process. For example, if we had policy year 1995 at ultimate rather than at first report, we could make a more precise estimate of policy year 1998 at ultimate.

The important consideration for this illustrative example is the pattern of errors for the different types of adjustments for the different years. For purposes of simplicity only two types of adjustments will be assumed. Development will be assumed to have larger estimation errors for recent years. In particular the "incomplete" policy year 1996 as of 12/31/96 will have an extremely large amount of development to ultimate. Trend[166] will be assumed to have larger estimation errors for more distant years.

For example, assume the reported Policy Year 1993 losses at 12/31/96 were $90 million. Further, assume that the point estimate[167] of Policy Year 1993 losses at ultimate is $96 million. This corresponds to a point estimate of the age to ultimate loss development factor of approximately 1.067. However, there is an error associated with this point estimate.

---

[163]The general ideas explored in this example would apply equally well to pure premiums.

[164]The general ideas explored in this example would apply equally well to calendar years or accident years of data.

[165]We assume that each of the adjusted loss ratios is intended to be an unbiased estimate of the Policy Year 1998 loss ratio.

[166]For illustrative purposes this can be thought of as trend, law amendment and other adjustments.

[167]Using data evaluated as of 12/31/96.

For example, an interval estimate of these ultimate losses might be \$92 million to \$100 million. This would correspond to an interval estimate of the age to ultimate loss development factor of approximately $1.067 \pm .044$.

A 95% confidence interval corresponds to about plus or minus two standard deviations. Therefore, this interval estimate of the loss development factor could result from a standard deviation of .022 or a variance of $.022^2 \approx .0005$. Any estimate is subject to error and in general one can estimate the variance of any estimator.[168]

Generally, estimation errors are quantified via variance-covariance matrices.[169] The covariances are introduced in order to capture the fact that the estimation errors for the years are usually positively correlated. If the development estimated for 1995 is too high, then it is likely that the development estimated for 1994 is too high as well. Similarly, if the trend applied to 1993 is too high, that applied to 1992 is likely to be too high as well.

Let $\boldsymbol{D}$ be the variance-covariance matrix quantifying the estimation errors related to development. An illustrative example of such a matrix is:

$$\boldsymbol{D} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 50 & 45 & 70 & 180 \\ 0 & 0 & 45 & 100 & 125 & 300 \\ 0 & 0 & 70 & 125 & 350 & 600 \\ 0 & 0 & 180 & 300 & 600 & 5{,}000 \end{pmatrix} \times 10^{-5}$$

[168] See, for example, Klugman, Panjer and Willmot [8].

[169] The diagonal elements are the variances quantifying the estimation errors. In this case, the element in the first row and second column is the covariance between the 1991 and 1992 errors. Readers may be familiar with the use of the inverse of the information matrix as a variance-covariance matrix when estimating parameters of loss distributions via the method of maximum likelihood. See, for example, Klugman, Panjer and Willmot [8].

The rows and columns correspond to the six years of data. For example, the variance of the estimated age to ultimate loss development factor of Policy Year 1993 is 50×10^{-5}.[170] The covariance between the estimated age to ultimate loss development factors for Policy Years 1993 and 1994 is 45×10^{-5}.

The particular values are chosen for illustrative purposes.[171] While the values would vary considerably depending on the particular application, the general pattern is expected to apply. The estimation errors for recent years are large,[172] and there is a positive correlation between the estimation errors for the different years.

Similarly, let $\boldsymbol{T}$ be the variance-covariance matrix quantifying the estimation errors related to trend. An illustrate example of such a matrix is:

$$\boldsymbol{T} = \begin{pmatrix} 350 & 292 & 240 & 192 & 150 & 110 \\ 292 & 300 & 247 & 198 & 155 & 114 \\ 240 & 247 & 250 & 201 & 157 & 115 \\ 192 & 198 & 201 & 200 & 156 & 115 \\ 150 & 155 & 157 & 156 & 150 & 110 \\ 110 & 114 & 115 & 115 & 110 & 100 \end{pmatrix} \times 10^{-5}$$

For example, the variance of the estimated trend factor from Policy Year 1994 to 1998 is 200×10^{-5}.[173] The covariance be-

[170]Thus, the standard deviation is $\sqrt{50 \times 10^{-5}} = .022$. If the point estimate of this loss development factor were, for example, 1.067, then using two standard deviations would result in an interval estimate of $1.067 \pm .044$.

[171]In particular, for longer tailed lines of insurance there would still be considerable development left for Policy Year 1991. In actual applications the actuary may have a good idea of how accurate an estimate is likely to be and thus could judgementally select a variance-covariance matrix.

[172]The error in developing the incomplete Policy Year 1996 is potentially extremely large.

[173]Thus, the standard deviation is $\sqrt{.002} = .045$. If the point estimate of this trend factor were, for example, 1.148, then using two standard deviations would result in an interval estimate of $1.148 \pm .090$.

tween the estimated trend factors to 1998 from 1994 and 1995 is 156×10^{-5}.

Again the particular values are chosen for illustrative purposes. The pattern was chosen such that the estimation error from trend is larger for more distant years and such that there is a large positive correlation[174] between the estimation errors for different years.[175]

9.2. *Covariance Structure for Years of Data*

Next we need to assume a variance-covariance structure for the year's loss ratios in the absence of any estimation error. Let this matrix be C. Then following the development in Mahler [1] of shifting risk parameters, assume that C has the form:[176]

$$C_{ij} = \delta_{ij} e^2 / \sqrt{E_i E_j} + r^2 \rho^{|i-j|}, \qquad \text{where} \quad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}. \tag{9.1}$$

It is not necessary to know the source of e^2, r^2 and ρ in order to proceed. However, it may be helpful to think of ρ as the dominant eigenvalue (other than unity) of the transpose of a transition matrix of a Markov chain, r^2 as the variance of the hypothetical means, and $e^2/\sqrt{E_i E_j}$ as the expected value of the process variance.

In any case, ρ determines the rate of decline in the covariances as the separation between years increases.[177] So $\rho = .90$ would

[174]For example, the correlation between the estimated trend factors to 1998 from 1994 and 1995 is $156/\sqrt{(200)(150)} = .90$.

[175]A similar pattern would be expected for on-level factors to adjust for law amendments.

[176]This is the covariance structure in the presence of shifting risk parameters, equivalent to Equation 3.16. If appropriate we could instead use one of the more complicated covariance structures, for example Equations 5.10 and 5.11.

[177]For data from an individual insurer, one of the reasons that the covariances between years declines as the separation increases may be nonrenewals of insureds. The higher the lapse rate the faster the expected rate of decline in these covariances. As discussed in Busche [26], the higher the lapse rate, the lower the weight given to older years of data.

represent a more rapid decline than would $\rho = .99$; the former would correspond to more rapidly shifting parameters over time than the latter.

The relative magnitudes of r^2 and e^2 will control how much weight is given to distant versus recent years. The larger e^2, the more random noise there is in the data from any one year; when e^2 is large we must give each of the available years significant weight. The smaller e^2, the less random noise there is in the data and larger weight can be given to more recent years and insignificant weight to older years. When e^2 is small, we can use a more responsive method. When e^2 is large we have to use a more stable method.

If everything else is equal, the larger the volume[178] of data in a year, the smaller we expect the process variance of the loss ratios to be. We assume the process variance is inversely proportional to the volume of data.[179] Thus, how responsive our estimation method should be depends on the volume of data available per year. If more data is available per year, then the estimation method can be more responsive.

9.3. Credibilities

Assume we are estimating the year $Y+\Delta$ by weighting together years $1,2,\ldots,Y$. Then as shown in Appendix B, the least squares weights Z_i, $i = 1,2,\ldots,Y$, with $\sum_{i=1}^{N} Z_i = 1$, are the solution to the $Y+1$ Equations 6.7:

$$\begin{aligned} &\sum_{i=1}^{Y} Z_i V_{ik} = V_{k,Y+\Delta} + \lambda/2, \qquad k = 1,2,\ldots,Y \qquad \text{and} \\ &\sum_{i=1}^{Y} Z_i = 1. \end{aligned} \tag{9.2}$$

---

[178]The measurement of the volume of data would depend on the particular application. For example, it could be house-years, man-weeks, car-years, inflation adjusted sales, etc. See Bouska [27].

[179]For this example, it has been assumed e^2 is the same for each year.

where V is the variance-covariance matrix and λ is the Lagrange Multiplier.[180]

9.4. No Estimation Error

If we do not include any estimation error, then in our example $V = C$, $Y = 6$ and $\Delta = 2$. Thus, the Equations 9.2 become:

$$\sum_{i=1}^{6} Z_i C_{ik} = C_{k,8} + \lambda/2, \qquad k = 1,2,\ldots,6 \qquad \text{and}$$
$$\sum_{i=1}^{6} Z_i = 1. \tag{9.3}$$

Given values for E_i, e^2, r^2, and ρ we can use Equation 9.1 to calculate the matrix C and then solve these linear Equations 9.3 for the weights Z_i.

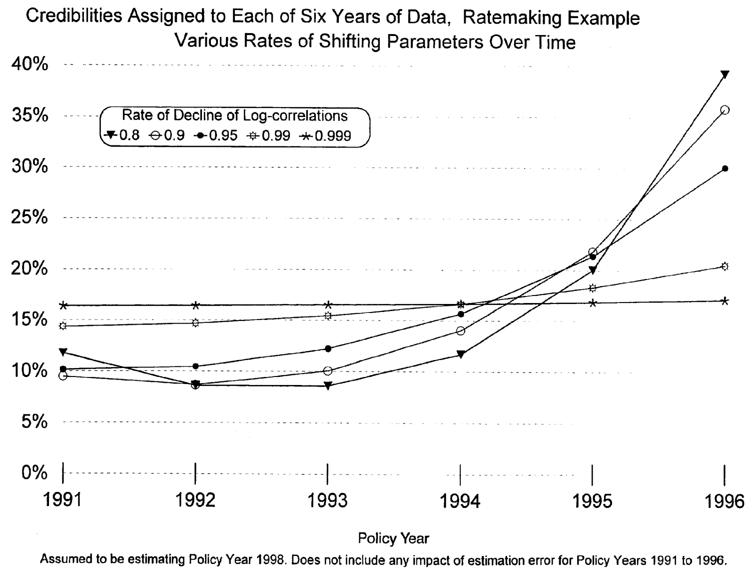
For example, with $E_i = 1$ for $i = 1$ to 8, $e^2 = .005$, $r^2 = .007$ and $\rho = .90$, we would get:

$$Z_1 = 9.5\%,$$
$$Z_2 = 8.7\%,$$
$$Z_3 = 10.1\%,$$
$$Z_4 = 14.0\%,$$
$$Z_5 = 21.8\%, \qquad \text{and}$$
$$Z_6 = 35.9\%.$$

Thus, as expected in the presence of shifting risk parameters, the more recent years 1996 and 1995 get more weight, while the earlier years 1991 and 1992 get less weight. Note that there is an "edge effect." The credibility assigned to 1991 is somewhat

[180] λ is an auxiliary variable, whose value will not be of particular interest.

FIGURE 20



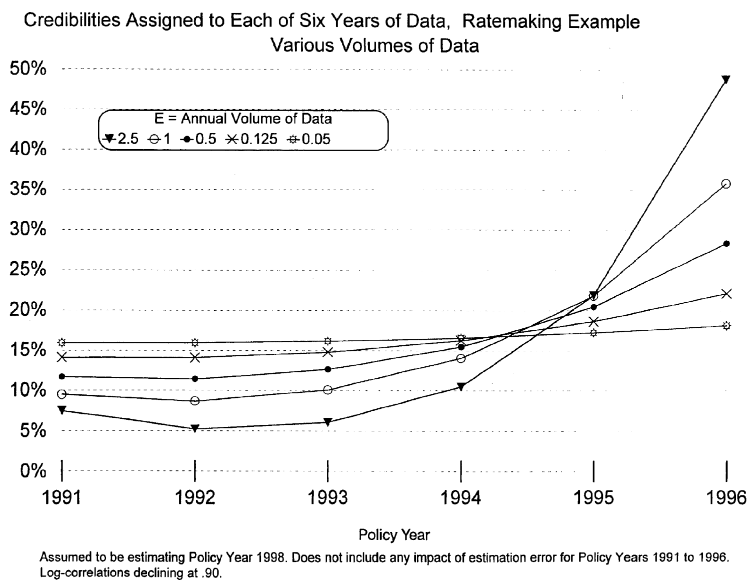
larger than it would otherwise be, since it is assumed to contain more unique information compared to 1992; the information content of 1992 is captured to some extent by the years 1991 and 1993 bracketing it on either side. The same "edge effect" applies to 1996, raising its credibility weight somewhat.

Figure 20 displays what happens as we vary ρ. As ρ approaches unity, parameters are shifting less rapidly, and therefore approximately equal weight is given to different years.[181] As ρ approaches zero, parameters are shifting more rapidly, and therefore less weight is given to the older years.

If we were to increase the expected value of the process variance, by taking $E = \frac{1}{2}$, keeping $e^2 = 0.005$, $r^2 = .007$ and $\rho = .90$,

[181] Recall that for this illustrative example the volume of data for each year is assumed to be the same.

FIGURE 21



then the weights are:

$$
\begin{aligned}
Z_1 &= 11.7\%, \\
Z_2 &= 11.4\%, \\
Z_3 &= 12.6\%, \\
Z_4 &= 15.5\%, \\
Z_5 &= 20.5\%, \qquad \text{and} \\
Z_6 &= 28.4\%.
\end{aligned}
$$

Compared to $E = 1$, with $E = \frac{1}{2}$ (a smaller volume of data) there is less weight given to more recent years and more weight given to more distant years. Figure 21 displays what happens as we vary E. As E (the volume of data) gets smaller, the weights become more equal. As E gets larger, more weight is given to recent years.

9.5. Taking Into Account Estimation Error

We can now introduce the impact of estimation error. First take the sum of the variance-covariance matrix discussed above for $E_i = 1$ for $i = 1$ to 8, $e^2 = .005$, $r^2 = .007$ and $\rho = .90$, and D, the assumed variance-covariance matrix for the estimation errors associated with development.[182]

$$V_{ij} = (.005)\,\delta_{ij} + (.007).90^{|i-j|} + D_{ij}, \qquad \text{for} \quad i,j \leq 6,$$

$$V_{ij} = (.005)\,\delta_{ij} + (.007).90^{|i-j|}, \qquad \text{for} \quad i \text{ or } j > 6.$$

For i or $j > 6$, the year is one whose losses we are trying to estimate. Since we are trying to estimate ultimate losses there is no additional development to be applied to those years. Thus, there is no D term, or alternately $D_{ij} = 0$ for i or $j > 6$.

Solving the Equations 9.2 for the weights we get:

$$Z_1 = 18.4\%,$$

$$Z_2 = 18.7\%,$$

$$Z_3 = 16.5\%,$$

$$Z_4 = 21.0\%,$$

$$Z_5 = 23.1\%, \qquad \text{and}$$

$$Z_6 = 2.3\%.$$

Taking into account the estimation errors due to development has decreased the weight given to recent years. In particular the weight given to incomplete policy year 1996 has declined very significantly. This is in line with the general practice of giving reduced or no weight to the incomplete policy year.

[182] We have assumed for simplicity that the estimation errors due to development are independent of the variance of the ultimate values for the years, so that the two variance-covariance matrices add. Also, we have for simplicity not had D depend on the volume of data E, even though in actual applications it is likely to be dependent.

Similarly we can include the impact of the estimation error due to trend using the previously selected variance-covariance matrix T

$$V_{ij} = (.005)\,\delta_{ij} + (.007).90^{|i-j|} + T_{ij}, \qquad \text{for} \quad i,j \le 6,$$

$$V_{ij} = (.005)\,\delta_{ij} + (.007).90^{|i-j|}, \qquad \text{for} \quad i \text{ or } j > 6.$$

Solving the Equations 9.2 for the weights one gets:

$$Z_1 = 7.8\%,$$
$$Z_2 = 6.7\%,$$
$$Z_3 = 8.5\%,$$
$$Z_4 = 12.1\%,$$
$$Z_5 = 23.3\%, \qquad \text{and}$$
$$Z_6 = 41.6\%.$$

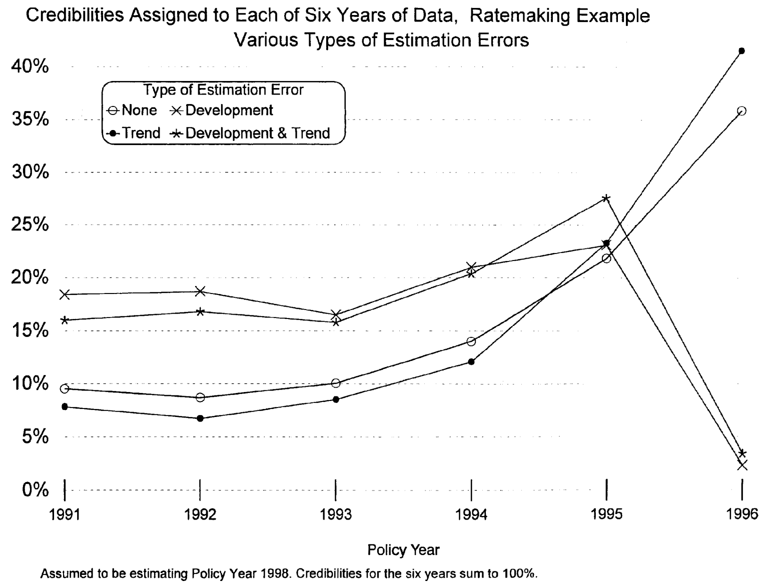
Taking into account the estimation errors due to trend has decreased the weight given to older years.

Finally, we can include the impact of both forms of estimation error by using the matrix $D + T$ in place of either D or T. (This assumes the estimation errors due to development and trend are independent.) The resulting weights are:

$$Z_1 = 16.0\%,$$
$$Z_2 = 16.8\%,$$
$$Z_3 = 15.8\%,$$
$$Z_4 = 20.4\%,$$
$$Z_5 = 27.6\%, \qquad \text{and}$$
$$Z_6 = 3.4\%.$$

Figure 22 compares the weights with and without the estimation errors.

FIGURE 22



9.6. *General Effects of Estimation Error*

The inputs used in the illustrative example can be varied. We can use more or less than six years of data. The gap between the latest year of data and the year to be estimated can differ. The volume of data and therefore the expected value of the process variance can vary by year. The relative size of the variance of the hypothetical means and the expected value of the process variance can differ. The rate at which parameters shift can be faster or slower. The pattern of estimation errors and their relative importance can differ.[183]

As any of these inputs vary, so do the calculated weights. Nevertheless, approximate values of the inputs can be used to estimate a pattern of weights that would be reasonable to use for a particular application.

[183] Also, in some cases the estimation errors would depend on the volume of data.

The general conclusions from analyzing this model all make sense. When we have a smaller volume of data per year we choose a more stable method.[184] Years with less data get less weight. When there is a lot of potential error from estimating loss development for a year, we give that year relatively less weight; this tends to affect more recent years. When there is a lot of potential error from estimating trend or on-level factors for a year, we give that year relatively less weight; this tends to affect more distant years. As there are more rapidly shifting parameters over time we choose a more responsive method.

Recall that in this illustrative example the weights always add to 100%. Thus the weight given to a particular year is a reflection of its value relative to the other years. Giving two years equal weight implies that they have the same value for purposes of estimation, but tells us nothing about what that value is in any absolute sense.

10. EXPERIENCE RATING

In this section, the previous results will be applied to a single split experience rating plan. While the values for the covariance structure used in this section were selected based on analyzing some workers compensation data from one state, they should be viewed as for illustrative purposes.

Section 10.1 describes the structure of a single-split experience rating plan. Section 10.2 describes the covariance structure. Section 10.3 displays the set of linear equations to be solved in order to get the credibilities. The parameters of the covariance structure are estimated and selected in Sections 10.4 to 10.8. Section 10.9 displays the credibilities that correspond to this covariance structure and parameters. Section 10.10 discusses the

[184]This customary practice is illustrated in Stern [28, p. 77]. The larger the premium volume, the more weight given to the latest year of data and the less weight given to the prior year of data. The smaller the premium volume, the more equal the weights given to the two years of data.

impact on the credibilities of taking into account the maturity of experience rating data.

10.1. Structure of the Experience Rating Plan

Assume we have split the losses into primary and excess portions, with the first \$5,000 of losses primary.[185] Assume only the first \$175,000 of any claim enters into experience rating.[186]

Assume we have Y years of data being used to predict year $Y + \Delta$.[187] We wish to determine credibilities to apply to the primary and excess data for each year.

Define the following quantities:

E_{Pi} = Expected Primary Losses for Year i,

E_{Xi} = Expected Excess Losses for Year i,

$E_i = E_{Pi} + E_{Xi}$ = Expected Losses for Year i,

A_{Pi} = Actual Primary Losses for Year i,

A_{Xi} = Actual Excess Losses for Year i,

$D_i = E_{Pi}/E_i$ = D-ratio for Year i,

$P_i = A_{Pi}/E_i$,

$X_i = A_{Xi}/E_i$,

$\pi_i = P_i - D_i = (A_{Pi} - E_{Pi})/E_i$
= Primary "Deviation Ratio" for Year i,

$\xi_i = X_i - (1 - D_i) = (A_{Xi} - E_{Xi})/E_i$
= Excess "Deviation Ratio" for Year i, and

M = Experience Modification.

[185]This is a single split experience rating plan. The \$5,000 split point is currently used for workers compensation. The general results illustrated here would be similar with a different split point.

[186]The \$175,000 limit is currently used in Massachusetts workers compensation. Other states use different limits.

[187]Typically $Y = 3$ and $\Delta = 2$. Years 1, 2 and 3 are predicting Year 5.

Then the experience modification will be of the form:

$$M = 1 + \sum_{i=1}^{Y} \pi_i Z_{Pi} + \sum_{i=1}^{Y} \xi_i Z_{Xi}. \tag{10.1}$$

The primary deviation ratio for year i, π_i, is given weight Z_{Pi}. The excess deviation ratio for year i, ξ_i, is given weight Z_{Xi}. The complement of credibility is given to unity, i.e., the average modification and the expected ratio of actual losses to expected losses.

If we were to introduce ballast and weighting values, as in the current experience rating plan,[188] then one could rewrite the credibilities as:

$$\begin{aligned} Z_{Pi} &= E_i/(E_i + B_i), \\ Z_{Xi} &= W_i Z_{Pi} = W_i E_i/(E_i + B_i). \end{aligned} \tag{10.2}$$

Note that there would be separate ballast and weighting values for each year in the treatment here. In the current experience rating plan there is a single B and W for a given insured.[189]

Then using the definitions of the deviation ratios:

$$\pi_i = (A_{Pi} - E_{Pi})/E_i \qquad \text{and} \qquad \xi_i = (A_{Xi} - E_{Xi})/E_i,$$

we can rewrite Equation 10.1 as:

$$M = 1 + \sum_{i=1}^{Y} \frac{A_{Pi} - E_{Pi} + W_i A_{Xi} - W_i E_{Xi}}{E_i + B_i}. \tag{10.3}$$

By giving each year its own weight, Equations 10.1 or 10.3 differ somewhat from the usual Equation 10.4.[190] If all the years of data were added together and assigned one combined primary credibility and one combined excess credibility, then Equation

[188] See Mahler [12] or Gillam and Snader [19].

[189] If an insured is interstate rated, the W and B values are a weighted average of those that would apply to that size risk if it were intrastate rated in each of the states involved.

[190] See for example Gillam and Snader [19] or Mahler [12].

10.1 would reduce to an equivalent of the usual equation for the experience modification for the Workers Compensation single split plan:

$$\begin{aligned} M &= 1 + Z_p(A_p/E - E_p/E) + Z_x(A_x/E - E_x/E) \\ &= 1 + \frac{A_p - E_p + WA_x - WE_x}{E + B} = \frac{A_p + WA_x + (1-W)E_x + B}{E + B} \end{aligned} \tag{10.4}$$

10.2. Variances and Covariances

The credibilities that appear in Equations 10.1 or 10.2 will be derived from the variance-covariance structure.[191]

There are three types of variances and covariances: those involving just primary deviation ratios, those involving just excess deviation ratios, and those involving both primary and excess deviation ratios. Each covariance will involve ratios from two (possibly different) years.

Define the relevant covariances as:

$$\begin{aligned} S_{ij} &= \mathrm{Cov}[\pi_i, \pi_j] = S_{ji}, \\ T_{ij} &= \mathrm{Cov}[\xi_i, \xi_j] = T_{ji}, \qquad \text{and} \\ U_{ij} &= \mathrm{Cov}[\pi_i, \xi_j]. \end{aligned} \tag{10.5}$$

Each of these three variance-covariance structures S, T and U would need to be modeled and/or estimated in a manner similar to that performed in previous sections of this paper. The covariances would differ by the amount of data and would be affected by risk heterogeneity, parameter uncertainty, and shifting risk parameters over time.

[191]The "best" credibilities will be taken as those that minimize the expected squared error. See Appendix D.

When the two years involved are the same, we obtain the total[192] variance or covariance:

$S_{1,1}$ = total variance of the primary deviation ratio,

$T_{1,1}$ = total variance of the excess deviation ratio, and

$U_{1,1}$ = total covariance of the primary and excess deviation ratios.

When the two years involved differ, we obtain in the *absence* of shifting risk parameters over time, the variance or covariance of the hypothetical means:

$S_{1,2}$ = variance of the hypothetical mean primary deviation ratios,

$T_{1,2}$ = variance of the hypothetical mean excess deviation ratios, and

$U_{1,2}$ = covariance of the hypothetical mean primary and excess deviation ratios.

In the presence of shifting risk parameters over time, it will be assumed that S, T and U each have a structure similar to that in Equations 5.10 and 5.11:

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{|i-j|} + I\gamma^{|i-j|}/\sqrt{E_iE_j} + \delta_{ij}\left(K/\sqrt{E_iE_j} + J\right)\right\}, \quad \sqrt{E_iE_j} \geq \Omega;$$

$$\mathrm{Cov}[X_i, X_j] = r^2\left\{\rho^{|i-j|} + I\gamma^{|i-j|}/\Omega + \delta_{ij}\left(K/\sqrt{E_iE_j} + J\right)\right\}, \quad \sqrt{E_iE_j} \leq \Omega.$$

The parameters r^2, I, J, K, ρ, γ and Ω in general may vary between the covariance structures for S, T and U. Thus, we will write each parameter with a subscript, p for primary, x for excess,

[192] "Total" means including both the variance (or covariance) of the hypothetical means and the expected value of the process variance (or covariance). See Mahler [11].

and m for mixed, resulting in the following equations:

$$S_{ij} = r_p^2 \left\{ \rho_p^{|i-j|} + I_p \gamma_p^{|i-j|} / \sqrt{E_i E_j} + \delta_{ij} \left(K_p / \sqrt{E_i E_j} + J_p \right) \right\}, \quad \sqrt{E_i E_j} \geq \Omega_p; \qquad (10.6)$$

$$S_{ij} = r_p^2 \left\{ \rho_p^{|i-j|} + I_p \gamma_p^{|i-j|} / \Omega_p + \delta_{ij} \left(K_p / \sqrt{E_i E_j} + J_p \right) \right\}, \quad \sqrt{E_i E_j} \leq \Omega_p; \qquad (10.7)$$

$$T_{ij} = r_x^2 \left\{ \rho_x^{|i-j|} + I_x \gamma_x^{|i-j|} / \sqrt{E_i E_j} + \delta_{ij} \left(K_x / \sqrt{E_i E_j} + J_x \right) \right\}, \quad \sqrt{E_i E_j} \geq \Omega_x; \qquad (10.8)$$

$$T_{ij} = r_x^2 \left\{ \rho_x^{|i-j|} + I_x \gamma_x^{|i-j|} / \Omega_x + \delta_{ij} \left(K_x / \sqrt{E_i E_j} + J_x \right) \right\}, \quad \sqrt{E_i E_j} \leq \Omega_x; \qquad (10.9)$$

$$U_{ij} = r_m^2 \left\{ \rho_m^{|i-j|} + I_m \gamma_m^{|i-j|} / \sqrt{E_i E_j} + \delta_{ij} \left(K_m / \sqrt{E_i E_j} + J_m \right) \right\}, \quad \sqrt{E_i E_j} \leq \Omega_m; \quad \text{and} \qquad (10.10)$$

$$U_{ij} = r_m^2 \left\{ \rho_m^{|i-j|} + I_m \gamma_m^{|i-j|} / \Omega_m + \delta_{ij} \left(K_m / \sqrt{E_i E_j} + J_m \right) \right\}, \quad \sqrt{E_i E_j} \leq \Omega_m. \qquad (10.11)$$

The covariance structure given by Equations 10.6 to 10.11 includes a total of 21 parameters. In theory, these parameters can be estimated using techniques similar to those used in the previous sections of this paper. As a practical matter, some of the parameters such as Ω, ρ and γ can be taken equal or approximately equal for S, T and U. So, for example, we could assume $\Omega_p = \Omega_x = \Omega_m$; in other words, we could assume that the transition from risk homogeneity to risk heterogeneity occurs at (approximately) the same size[193] for all three covariance structures.

[193] As applied here to experience rating, I have followed the current practice of using the total expected losses rather than the primary or excess losses to define the size of risk.

10.3. Equations for Credibilities

Set aside for now the difficult task of estimating the variance-covariance matrices: S, T and U. As shown in Appendix D, we can derive 2 Y linear equations for the 2 Y credibilities:

$$\sum_{i=1}^{Y}(Z_{Pi}S_{ik} + Z_{Xi}U_{ki}) = S_{k,Y+\Delta} + U_{k,Y+\Delta}, \qquad k = 1,2,\ldots Y, \quad \text{and} \tag{10.12}$$

$$\sum_{i=1}^{Y}(Z_{Pi}U_{ik} + Z_{Xi}T_{ki}) = U_{Y+\Delta,k} + T_{k,Y+\Delta}, \qquad k = 1,2,\ldots Y. \tag{10.13}$$

If the excess losses are set equal to zero; i.e., we have a no-split plan, then Equation 10.12 reduces to Equation 2.4. In the absence of shifting risk parameters over time, as shown in Appendix D, Equations 10.12 and 10.13 reduce to those derived in Mahler [11].[194]

10.4. Estimating the Parameters of the Covariance Structure

Prior sections have discussed how we might estimate some of the needed parameters. Also, the National Council on Compensation Insurance has estimated quantities which are similar to the I, J and K parameters here.[195] These NCCI estimates can aid in choosing the relative sizes of the I, J and K parameters.

The available data was insufficient to allow independent estimates of ρ_p, ρ_x and ρ_m, so it is assumed that $\rho_p \approx \rho_x \approx \rho_m$.

Thus, for a given insured, the size of risk to which we compare Ω_p, Ω_x or Ω_m is the same. In this case, I think it unlikely that Ω_p, Ω_x and Ω_m would differ. Nevertheless, for generality, I have labeled Ω with subscripts even though in the example $\Omega_p = \Omega_x = \Omega_m$.

[194]See Equations 5.3 and 5.4 in Mahler [11] for Z_p and Z_x for a split experience rating plan.

[195]See Gillam [13] and Mahler [12]. Note that the credibilities in the NCCI Revised Experience Rating Plan were derived without explicit recognition of the impact of what has been called herein U_{ij}, the covariance of the primary and excess losses.

Similarly, assume $\gamma_p = \gamma_x = \gamma_m$ and $\Omega_p = \Omega_x = \Omega_m$. So we have assumed that the rate of shifting parameters over time as it impacts S, T and U is similar and that risk homogeneity applies for risks of size less than Ω.

The primary losses are less subject to random fluctuations than the excess losses. Therefore, whenever possible the results of analyzing the primary deviation ratios will be relied upon.

The data analyzed was that used for intrastate experience rating in one state over a five year period.[196] The analysis was limited to risks that were experience rated over this whole period of time.[197] For each such risk, for each "rating year" the data consists of three separate years of actual primary losses, actual excess losses, expected primary losses, and expected excess losses, that were used to calculate the experience modification. The variance-covariance structure of this data was analyzed by size of risk.

For example, for risks with expected annual losses between \$10,000 and \$20,000 the correlations between primary deviation ratios, $(A_p - E_p)/(E_p + E_x)$, were computed for different separations and different reports.[198] For example, this primary correlation was .331 between the "rating year" 1991 data at first report[199] and the "rating year" 1992 data at first report. Table 6 displays the correlations.

There are 12 correlations corresponding to a separation of 1 year, 9 for 2 years, 6 for 3 years, and 3 for 4 years. Based

[196]For experience modifications applied to policies written during 1991, 1992, 1993, 1994 and 1995 in Massachusetts workers compensation.

[197]Employers who went out of business, left the state, became self-insured or became too small to be experience rated would therefore be excluded.

[198]This differs somewhat from Mahler [12] where correlations between $A_p/(E_p + E_x)$ were examined. The two sets of correlations are very similar.

[199]Generally data from a 1989 policy at first report, a 1988 policy at second report, and a 1987 policy at third report would be used to calculate the experience modification to apply to the 1991 policy. The data from the 1989 policy at first report is what is being referred to here.

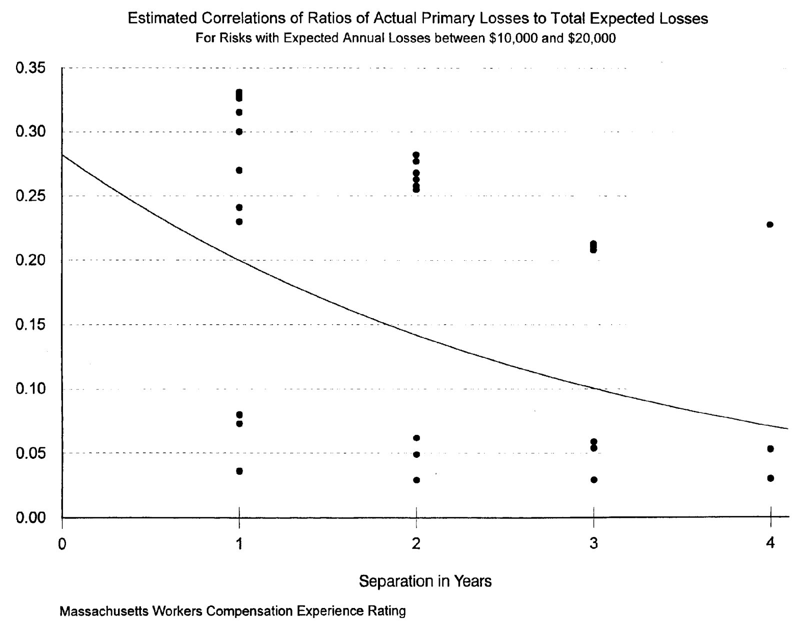
TABLE 6

CORRELATIONS OF RATIOS OF ACTUAL PRIMARY LOSSES TO TOTAL EXPECTED LOSSES
Expected Annual Losses[1] Between $10,000 and $20,000
Massachusetts Workers Compensation Experience Rating

Rating Years[2]	Report	Separation	Correlation
91 92	1	1	.331
91 92	2	1	.230
91 92	3	1	.270
92 93	1	1	.328
92 93	2	1	.326
92 93	3	1	.241
93 94	1	1	.080
93 94	2	1	.300
93 94	3	1	.330
94 95	1	1	.036
94 95	2	1	.073
94 95	3	1	.315
91 93	1	2	.282
91 93	2	2	.255
91 93	3	2	.258
92 94	1	2	.062
92 94	2	2	.268
92 94	3	2	.263
93 95	1	2	.049
93 95	2	2	.029
93 95	3	2	.277
91 94	1	3	.059
91 94	2	3	.211
91 94	3	3	.208
92 95	1	3	.054
92 95	2	3	.029
92 95	3	3	.213
91 95	1	4	.053
91 95	2	4	.030
91 95	3	4	.228

[1] If E_1 and E_2 are the expected losses (primary plus excess) for the given report which are used for experience rating the two rating years, then $\sqrt{E_1 E_2}$ is between $10,000 and $20,000. There were an average of 3,060 such risks.
[2] 91 refers to experience modifications applied to policies written in 1991.

FIGURE 23



on Mahler [1], it is expected that the logs of these correlations will decline linearly as the separation increases. A least squares regression was fit to these correlations, and the result was $c = (.282).709^s$, where c is the correlation and s is the separation. The value of .282 will be referred to as the "intercept" while the value of .709 will be referred to as the "slope" of this regression. This regression is illustrated in Figure 23.

Similar regressions were fit to the correlations for other size categories.[200] A similar analysis was performed for the correlations of excess deviation ratios and the correlations between primary and excess deviation ratios. The resulting slopes and intercepts are displayed in Table 7.

---

[200] A few estimated correlations were not positive and were excluded from the regressions.

TABLE 7

RESULTS OF EXPONENTIAL REGRESSIONS FIT TO CORRELATIONS OF RATIOS
Massachusetts Workers Compensation Experience Rating

Expected Annual Losses ($000)	Average Number of Risks	Primary		Excess		Mixed	
		Intercept	Slope	Intercept	Slope	Intercept	Slope
3 to 5	3,952	.224	.727	.034	.820	.089	.775
5 to 10	4,798	.169	.836	.086	.792	.102	.841
10 to 20	3,060	.282	.709	.098	.741	.126	.737
20 to 50	2,197	.380	.992	.146	.842	.167	.950
50 to 100	770	.579	.869	.260	.809	.272	.855
100 to 200	356	.717	.865	.442	.723	.408	.790
200 to 500	186	.661	.877	.471	.812	.355	.825
500 to 1,000	45	.869	.658	.693	.687	.397	.781
1,000 to 2,000	14	.882	.973	.776	.850	.583	.828

10.5. Estimating I_p, J_p, K_p, I_x, J_x and K_x

As discussed previously, the intercepts of the primary correlations are an estimate of the credibility to be assigned to a single year of data in the absence of shifting risk parameters.[201] Thus we expect a curve of the form:

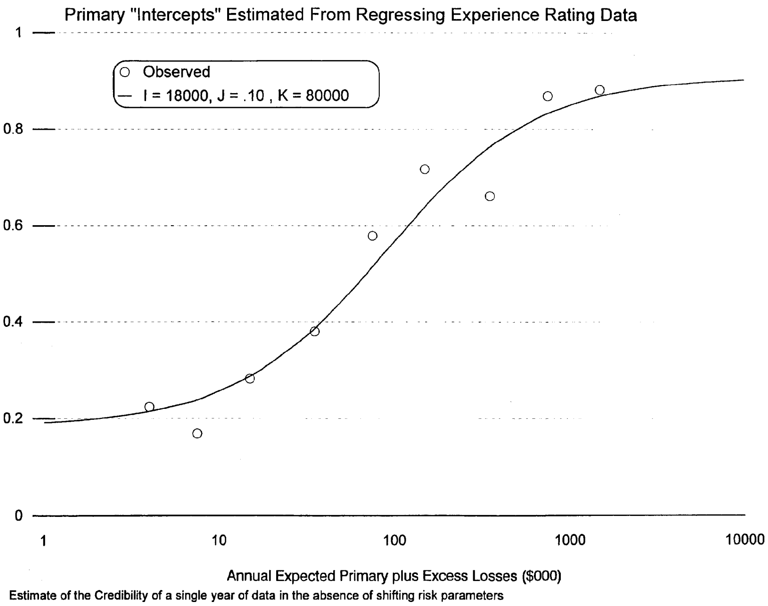
$$Z = (E + I_p)/\{(1 + J_p)E + I_p + K_p\}, \qquad \text{for} \quad E \geq \Omega_p.$$

As shown in Figure 24, the values $I_p = 18{,}000$, $J_p = .10$, and $K_p = 80{,}000$ do a reasonable job of approximating the estimated intercepts for the primary deviation ratios.[202]

Similarly, as seen in Figure 25, values of $I_x = 20{,}000$, $J_x = .15$, and $K_x = 315{,}000$ do a reasonable job of approximating the estimated intercepts for the excess deviation ratios.

[201] The correlation between primary and excess losses is also ignored.
[202] More data on extremely large risks would improve the estimate of J_p.

FIGURE 24



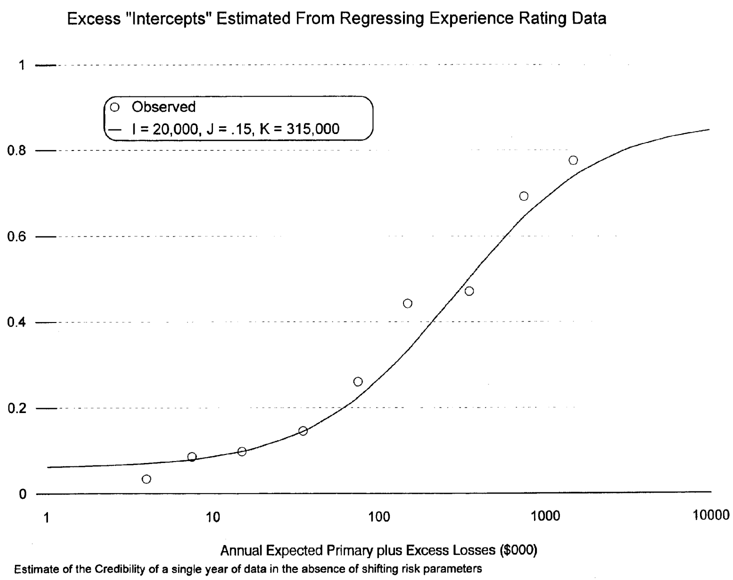
The intercepts of the regressions fit to the mixed correlations have a somewhat different interpretation. Using Equations 10.6, 10.8 and 10.10, for a primary deviation ratio π_i and excess deviation ratio ξ_j for different years $i \neq j$, we have for $\sqrt{E_i E_j} \geq \Omega_m$, $E_i \geq \Omega_p$ and $E_j \geq \Omega_x$:

$$\operatorname{Corr}[\pi_i, \xi_j] = \operatorname{Cov}[\pi_i, \xi_j] / \sqrt{\operatorname{Var}(\pi_i)\operatorname{Var}(\xi_j)} = U_{ij} / \sqrt{S_{ii} T_{jj}}, \qquad \text{and}$$

$$\operatorname{Corr}[\pi_i, \xi_j] = \frac{r_m^2 \left\{ \rho_m^{|i-j|} + I_m \gamma_m^{|i-j|} / \sqrt{E_i E_j} \right\}}{r_p r_x \sqrt{(1 + J_p + (I_p + K_p)/E_i)(1 + J_x + (I_x + K_x)/E_j)}}. \tag{10.14}$$

Note that the mixed correlation between different years does not involve J_m and K_m. Thus the regression fit to the mixed intercepts cannot be used to estimate these parameters. The intercept

FIGURE 25



of that regression should be:[203]

$$\frac{r_m^2\left\{1+I_m/\sqrt{E_iE_j}\right\}}{r_pr_x\sqrt{(1+J_p+(I_p+K_p)/E_i)(1+J_x+(I_x+K_x)/E_j)}}. \tag{10.15}$$

These intercepts by size of risk will be used subsequently to check the reasonableness of selected parameter values.

10.6. Estimating γ and ρ

The slopes of the regressions fit to the correlations are displayed in Table 7. There is considerable random fluctuation, but generally the slopes are somewhere in the 75% to 90% range. As discussed previously, the slope for smaller sizes should be

[203]The result of substituting unity for ρ_m and γ_m in Equation 10.14.

approximately equal to γ, while that for larger sizes should be approximately equal to ρ. There is some tendency for the primary ratios for the slopes to be closer to unity for large sizes. For illustrative purposes, $\gamma_p = .80$ and $\rho_p = .85$ will be selected.

There is less evidence of a dependence on size of risk for the excess ratios; $\gamma_x = \rho_x = .80$ will be selected. The rates of shifting for the mixed correlations are similar to those for the primary and excess correlations; $\gamma_m = .80$ and $\rho_m = .83$ will be selected.

10.7. Estimating r^2, I_m, J_m and K_m

Besides analyzing correlations between data from different years, we need to analyze the variance of data from a single year. The variance is S_{ii} for primary deviation ratios:

$$S_{ii} = r_p^2(1 + J_p + (I_p + K_p)/E_i), \qquad E_i \geq \Omega. \qquad (10.16)$$

Similarly, for the excess deviation ratios the variance is

$$T_{ii} = r_x^2(1 + J_x + (I_x + K_x)/E_i), \qquad E_i \geq \Omega. \qquad (10.17)$$

For a given year, the covariance between the primary and excess deviation ratios is U_{ii}:

$$U_{ii} = r_m^2(1 + J_m + (I_m + K_m)/E_i), \qquad E_i \geq \Omega. \qquad (10.18)$$

The estimated variances and covariances for various sizes of risk are shown in Table 8.[204] Using the estimated primary variances and the previously selected values I_p, J_p and K_p we can estimate r_p^2. Similarly, we can estimate r_x^2. The covariances can be used to estimate r_m^2, I_m, J_m and K_m.

Table 9 shows the estimates of r_p^2 and r_x^2 that result from the estimated variances for the different sizes of risk and Equations

[204]In each case, the value shown is an average of 15 values from 5 years and 3 reports.

TABLE 8

VARIANCES AND COVARIANCES FOR A SINGLE YEAR
Massachusetts Workers Compensation Experience Rating

Expected Annual Losses ($000)	Average Number of Risks	Primary Variance	Excess Variance	Mixed Covariance[1]
3 to 5	6,228	.271	15.444	1.221
5 to 10	6,619	.181	9.754	.791
10 to 20	4,081	.101	5.526	.442
20 to 30	1,569	.077	3.344	.301
30 to 50	1,318	.064	2.459	.222
50 to 100	1,034	.047	1.816	.177
100 to 200	506	.034	1.045	.112
200 to 500	262	.022	.637	.069
500 to 1,000	67	.019	.405	.055
1,000 to 2,000	23	.016	.213	.034

In each case the estimate shown is the average of 15 estimates from each of 5 years at 3 reports.
[1]Covariance of primary and excess deviation ratios for the same year.

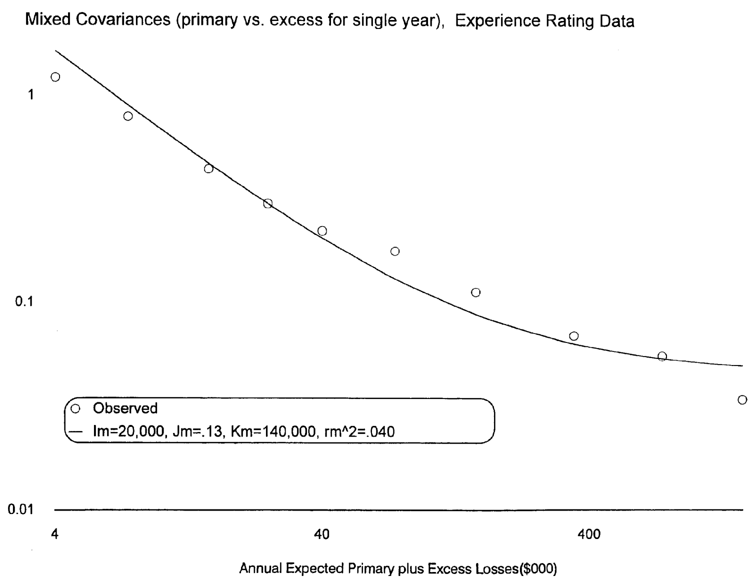
TABLE 9

ESTIMATES OF r^2 FROM OBSERVED VARIANCES

Expected Annual Losses ($000)	Primary Variance	r_p^2	Excess Variance	r_x^2
4	.271	.011	15.444	.182
7.5	.181	.013	9.754	.213
15	.101	.013	5.526	.235
25	.077	.015	3.344	.230
40	.064	.018	2.459	.258
75	.047	.020	1.816	.323
150	.034	.019	1.045	.309
350	.022	.016	.637	.302
750	.019	.015	.405	.254
1,500	.016	.014	.213	.155

$r^2 = \text{Variance}/(1 + J + (I + K)/E)$
$I_p = 18{,}000$, $J_p = .10$, $K_p = 80{,}000$
$I_x = 20{,}000$, $J_x = .15$, $K_x = 315{,}000$

FIGURE 26



10.16 and 10.17. The values of r_p^2 are all in the range of .015. The similarity of the estimates of r_p^2 that result from the different size categories tends to confirm the reasonableness of the previously selected values of I_p, J_p and K_p.[205]

Similarly, Table 9 displays estimates for r_x^2 from the different size categories using Equation 10.17. The values of r_x^2 vary considerably. A value $r_x^2 \approx .26$ will be selected.

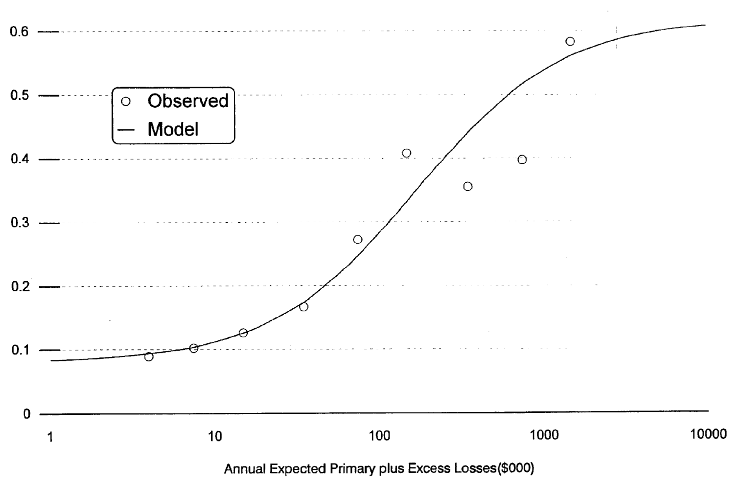
As seen in Figure 26, using Equation 10.18, the set of parameters: $I_m = 20{,}000$, $J_m = .13$, $K_m = 140{,}000$, and $r_m^2 = .040$, provides a reasonable fit to the estimated covariances by size of

---

[205] If the initially selected I_p, J_p and K_p did not seem to perform well here, then we could modify them somewhat so they performed better here. Then we would go back and check the performance in fitting the intercepts of the regressions fit to the correlations. We could iterate in this manner until we arrived at the best set of parameters.

FIGURE 27

Mixed "Intercepts" Estimated From Regressing Experience Rating Data



Model using values for the parameters of Ip = 18,000, Jp = .1, Kp = 80,000, Ix = 20,000, Jx = .15, Kx = 315,000, Im = 20,000, rp^2 =.015, rx^2 = .26, and rm^2 = .040.

risk.[206] Bear in mind that the largest size category has a limited number of risks and so the resulting estimate of the covariance is not very accurate.

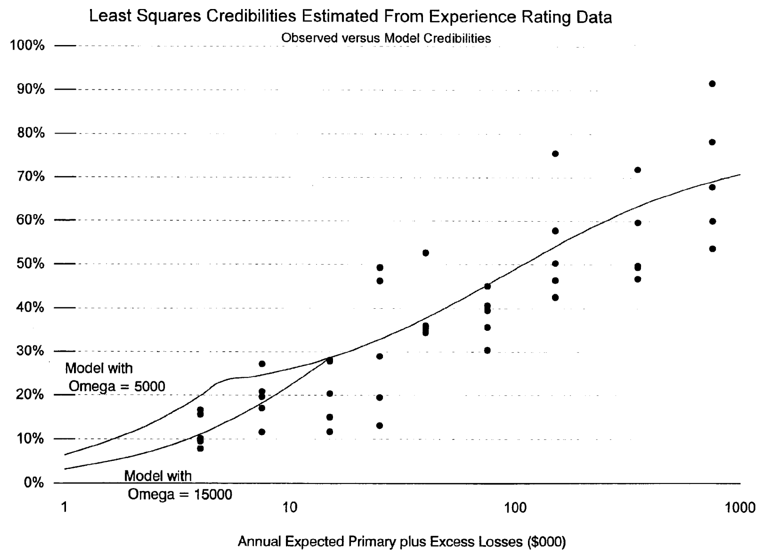
Using the selected set of parameters, we can compare the theoretical values from Equation 10.15 to the observed intercepts from the regressions fit to the mixed correlations. As seen in Figure 27, the fit is not unreasonable. Thus, the selected values of I_p, J_p, K_p, I_x, J_x, K_x, I_m, r_p^2, r_x^2, and r_m^2 seem consistent with the observed mixed intercepts.

10.8. Selecting Ω

The final parameters to be selected are Ω_p, Ω_x and Ω_m. Based on the reasonable fits obtained so far, Ω should be near the

[206]The value of J_m was selected to be between the selected J_p and J_x. More data on extremely large risks would improve the estimates of all the J parameters.

FIGURE 28



2 years of Losses used to predict a 3rd year of Losses. Model adjusted for 2 years of data used to predict the very next year. Model using values for the parameters of Ip = 18,000, Jp = .1, Kp = 80,000, Ix = 20,000, Jx = .15, Kx = 315,000, Im = 20,000, Jm = .13, Km=140,000, rp^2 = .015, rx^2 = .26, and rm^2 = .040.

smaller sizes of risk examined or below the eligibility level for experience rating in Massachusetts.[207] Due to limited information, one value will be selected for all three parameters, $\Omega_p = \Omega_x = \Omega_m$.

Figure 28 displays least squares credibilities estimated from the 3 years of data used to experience rate policies. The credibilities are those that produced the smallest squared error when the first 2 years of data were used to predict the third.[208] These are compared to the model credibilities that result from the estimated parameters and the use of Equations 10.12 and 10.13.[209]

[207] If Ω were in the middle of the range of sizes examined, the observed covariance structure should have been affected.

[208] For each size category there are five estimates, one for each "rating year."

[209] The primary and excess credibilities were averaged: $Z = DZ_p + (1-D)Z_x$, with $D = .22$.

The model credibilities are in the range indicated by the data.[210]

The credibilities for smaller size risks are shown for two values of Ω, $\Omega = \$5{,}000$ and $\Omega = \$15{,}000$. Based on Figure 28, $\Omega = \$15{,}000$ does a better job than $\Omega = \$5{,}000$. However, a better estimate of Ω would result from a more detailed analysis of data from risks barely eligible for or too small to be experience rated in Massachusetts.[211] While it is beyond the scope of this paper, a preliminary review of merit rating data for Massachusetts workers compensation indicates that $\Omega \approx \$5{,}000$ or perhaps even a little less. In any case, for illustrative purposes the selected values will be $\Omega_p = \Omega_x = \Omega_m = \$5{,}000$.[212]

10.9. Estimated Credibilities

The selected parameter values are:

$$
\begin{array}{llll}
I_p = \$18{,}000 & J_p = .10 & K_p = \$80{,}000 & r_p^2 = .015 \\
I_x = \$20{,}000 & J_x = .15 & K_x = \$315{,}000 & r_x^2 = .26 \\
I_m = \$20{,}000 & J_m = .13 & K_m = \$140{,}000 & r_m^2 = .040
\end{array}
$$

$$
\begin{array}{lll}
\gamma_p = .80 & \rho_p = .85 & \Omega_p = \$5{,}000 \\
\gamma_x = .80 & \rho_x = .80 & \Omega_x = \$5{,}000 \\
\gamma_m = .80 & \rho_m = .83 & \Omega_m = \$5{,}000.
\end{array}
$$

Using the above parameter values and Equations 10.12 and 10.13, credibilities were calculated for 3 years of data being used

---

[210]Since the parameters were estimated by a different analysis of this exact same data, this serves as a consistency check rather than an independent test of the results.

[211]For example, for a risk with \$1,000 in expected annual losses, with $\Omega = \$5{,}000$ $Z = 6.7\%$, while with $\Omega = \$15{,}000$ $Z = 3.3\%$. Thus an examination of the credibilities indicated by the data for smaller risks should help to determine the appropriate Ω.

[212]This would correspond to a minimum ballast value of $K_p\Omega_p/(I_p + \Omega_p) = (\$80{,}000)(5/(18+5)) \approx \$17.3$ thousand. Interestingly, for $g = 7$ as in Massachusetts, the NCCI minimum ballast value would be (7) (2,500) = \$17,500. The corresponding minimum weighting value would be $(K_p/K_x)(\Omega_p/\Omega_x)(I_x + \Omega_x)/(I_p + \Omega_p) = (80/300)(1)(25/23) = .29$. The NCCI minimum W is .07.

TABLE 10

EXPERIENCE RATING CREDIBILITIES[1]
Using Parameters Listed in Section 10.9

Expected Annual Losses ($000)	Primary[2]				Excess[3]				Combined[4]
	Z_1	Z_2	Z_3	Z_p	Z_1	Z_2	Z_3	Z_x	Z
.1	.9%	1.1%	1.3%	3.3%	.02%	.02%	.03%	.1%	.8%
.5	4.0	5.0	6.2	15.2	.09	.11	.14	.3	3.6
1	7.2	9.1	11.7	28.1	.2	.2	.3	.7	6.8
2	12.1	15.6	20.8	48.4	.5	.6	.7	1.7	12.0
3	15.3	20.3	28.1	63.7	.7	.9	1.1	2.7	16.2
4	17.5	23.8	34.1	75.4	1.0	1.3	1.5	3.8	19.6
5	19.0	26.4	39.2	84.5	1.3	1.7	2.0	5.0	22.5
7.5	19.9	27.8	41.6	89.3	1.5	1.8	2.2	5.4	23.9
10	20.6	29.0	43.9	93.5	1.6	2.0	2.4	5.9	25.2
25	22.7	33.5	54.5	110.6	2.3	2.9	3.6	8.8	31.2
50	22.0	35.8	65.4	123.3	3.4	4.4	5.5	13.3	37.5
100	17.3	34.7	77.3	129.3	5.0	6.6	8.7	20.3	44.3
250	6.9	27.0	90.6	124.4	6.8	10.0	15.3	32.1	52.4
500	.8	19.6	95.5	115.9	7.0	11.8	21.5	40.3	56.9
1,000	− 1.3	14.5	94.8	108.1	6.1	12.4	27.9	46.4	59.9
2,500	− .2	12.6	89.1	101.5	4.3	11.6	35.1	51.0	62.1
5,000	1.0	13.1	84.6	98.8	3.3	10.8	38.8	52.9	63.0
10,000	1.9	14.0	81.4	97.3	2.7	10.1	41.1	53.8	63.4
∞	2.9	15.5	77.4	95.7	2.0	9.1	43.7	54.8	63.8

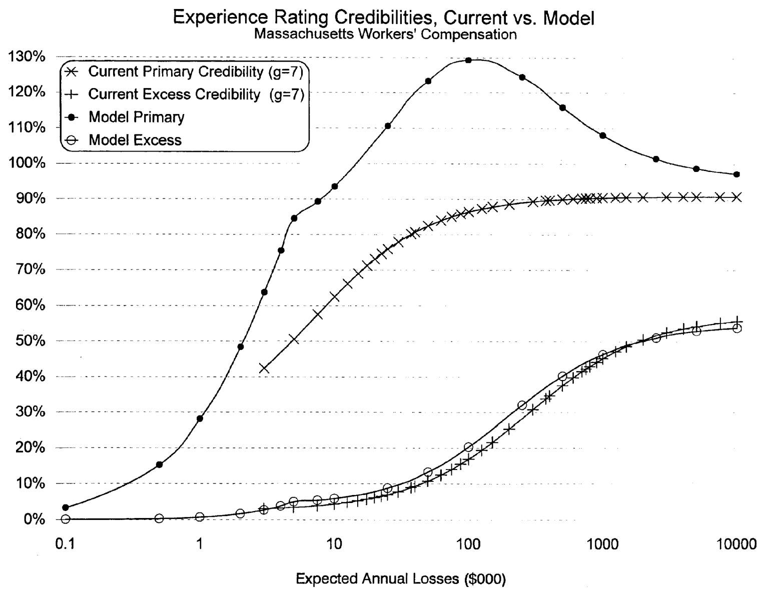
[1]Using data from years 1, 2 and 3 to predict year 5.
[2]Z_p is the sum of the primary credibilities for the three years.
[3]Z_x is the sum of the excess credibilities for the three years.
[4]$Z = DZ_p + (1 - D)Z_x$, for $D = .22$.

to predict the fifth year.[213] Table 10 displays the primary and excess credibilities assigned to each of the three years of data as well as the sum. Note that the primary credibilities can sum to greater than 100%. As pointed out in Mahler [11] and Mahler [12], this is not unusual when we take into account the covariance of the primary and excess losses.[214] In such circumstances

[213]For example, 1994, 1995 and 1996 data is being used to experience rate a policy written during 1998.

[214]Note that the numerator of π_i involves the primary losses while the denominator is the sum of the primary and excess expected losses. If instead the denominator had been just primary expected losses, then the ratio would be larger and the weight assigned to it would be smaller by a factor of the D-ratio. Then the primary weights would sum to less than 100%.

FIGURE 29

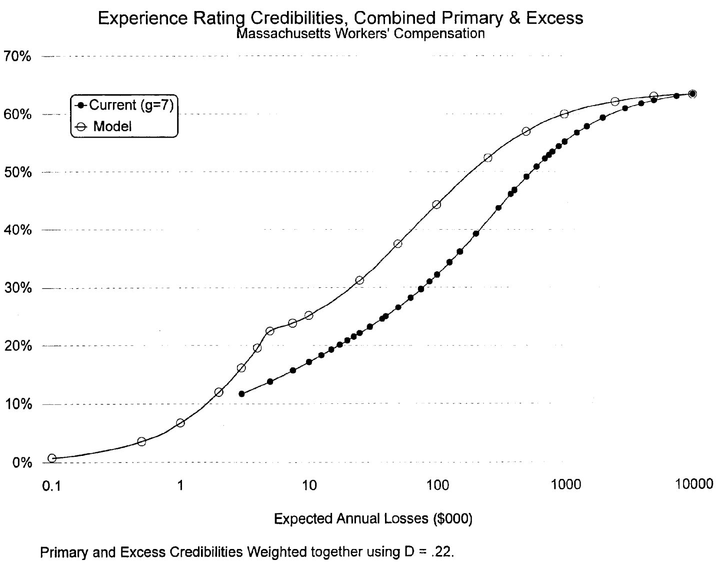


we could constrain the primary credibilities to be equal to unity, as shown in Mahler [11] and Mahler [12]. In any case, the combined credibility is between 0 and 1. It should also be noted that the uncertainty in the estimated J and Ω parameters produces uncertainty in the credibilities for large and small risks respectively.

As the size of risk increases, the weight assigned to the most recent year increases relative to that for the most distant year. For very large risks, we can rely almost solely on the latest year of data. For very small risks, it would be reasonable to rely on more than three years of data, since the older years would have credibilities close to that for the more recent years of data.

Figure 29 compares the primary and excess credibilities from Table 10 to those currently used in Massachusetts workers com-

FIGURE 30



pensation experience rating.[215] Figure 30 does the same comparison for the weighted average of the primary and excess credibilities. As in Mahler [12], the indicated primary and combined credibilities are generally higher than those from the NCCI plan.

At least part of this difference is due to the fact that Massachusetts average claim costs are higher than the national average. Using the same \$5,000 split point between primary and excess in every state results in lower than average D-ratios in Massachusetts. Thus, the primary losses in Massachusetts are "very primary," while the excess losses are only "mildly excess." Thus, both the Massachusetts primary losses and excess losses contain more useful information and less random noise than in

[215]The NCCI Revised Experience Rating program, with $g = 7$. Here we have ignored the All Risk Rating Program (ARAP) which is currently applied on top of experience rating in Massachusetts and in combination produces more responsiveness to the insured's losses.

the average state. This would not be the case if the split point depended on the state specific parameter g.

On the other hand, due to the consideration here of the covariance between the primary and excess losses, the primary credibilities are higher and the excess credibilities are lower than they would otherwise be. The primary losses contain valuable information for predicting both the future primary and excess losses.

On balance, the excess credibilities for the current model are fairly close to those from the NCCI plan, while the primary credibilities are much greater. As stated before, the results would be expected to differ somewhat in low severity states.

In any case, the combined model credibilities are more similar to what would be obtained in other states.[216] The combined credibilities are between 0 and 1. In this case, they increase smoothly from zero to a maximum of about 63% for the largest risks.[217] Due to shifting risk parameters and parameter uncertainty, the maximum credibility is less than 100%.

The model combined credibilities are generally larger than those from the current NCCI plan. For example, for $100,000 in expected annual losses, the model has a combined credibility of 44.3%, while the current plan has 32.2%. While there are significant differences,[218] the overall magnitude and pattern of credibilities is very similar.

Note that model credibilities are also shown for risk sizes below the current eligibility level for experience rating.[219] Recall

---

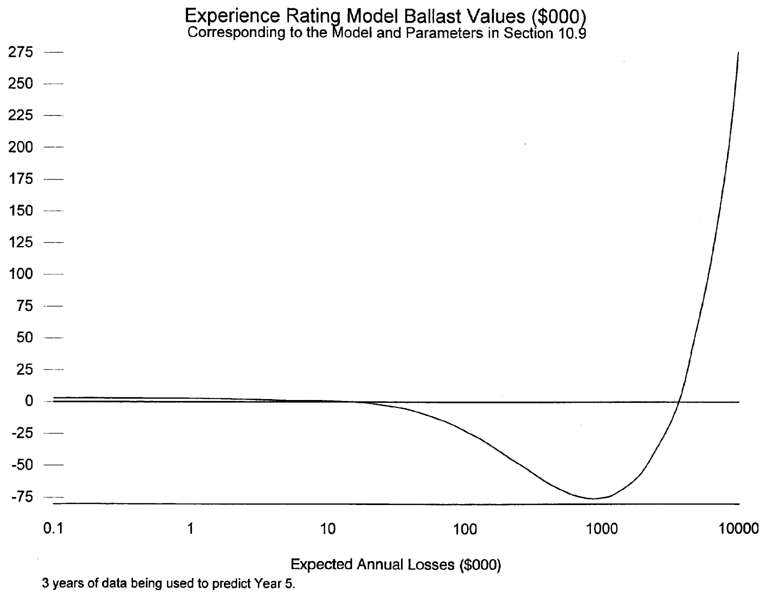
[216]The D-ratio is lower in Massachusetts than in the average state, so the primary credibilities receive less weight. This would result in lower combined credibilities, except that the primary credibilities are larger than average in Massachusetts.

[217]Due to the limited data for very large risks, the model parameters were chosen to some extent so that the maximum credibility would be close to that from the NCCI plan. In an average state the NCCI plan has a maximum credibility of about 67%, as shown in Mahler [12].

[218]For most insureds a 13% difference in credibilities would have a less than 3% effect on their experience modifications.

[219]The NCCI formulas for credibility are not intended to be applied to very small risks. As discussed in Mahler [12], minimum B values, etc., are used to deal with this problem. The NCCI credibilities graphed here are prior to any such refinements.

FIGURE 31

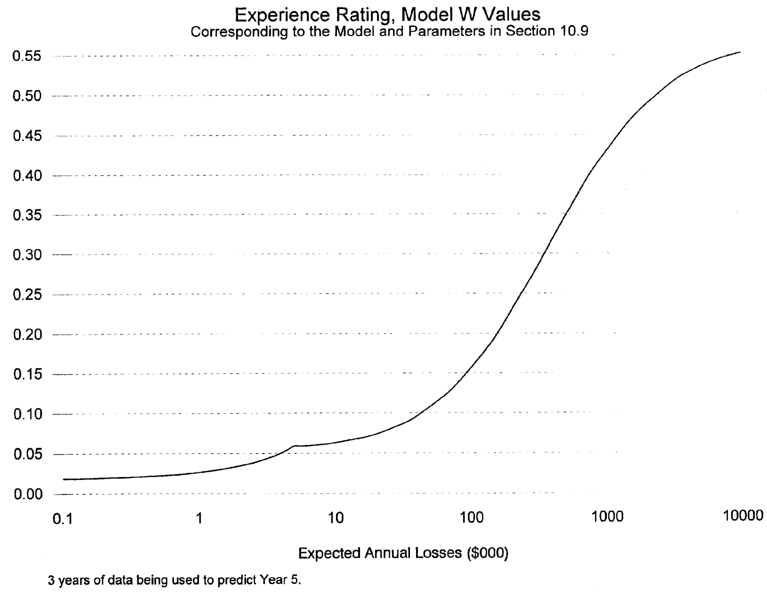


that these credibilities for very small risks depend very significantly on the estimated Ω parameter. An analysis of data from these very small risks would refine this estimate.

Figure 31 shows the ballast values corresponding to the model primary credibilities shown in Table 10. Since $B = E((1/Z_p) - 1)$, when $Z_p >> 100\%$, it follows that $B < 0$. While it is currently the case that B is positive, there is no mathematical reason why B cannot be negative.[220] Small risks have $B \approx 3{,}000$. B declines and becomes negative before increasing to very large positive values. Figure 32 shows W (weighting) values corresponding to the model credibilities shown in Table 10. Other than a discontinuity in the derivative of W that occurs at $\Omega = 5{,}000$, the W values increase smoothly with size of risk.

[220] $B = 0$ would correspond to $Z_p = 100\%$.

FIGURE 32

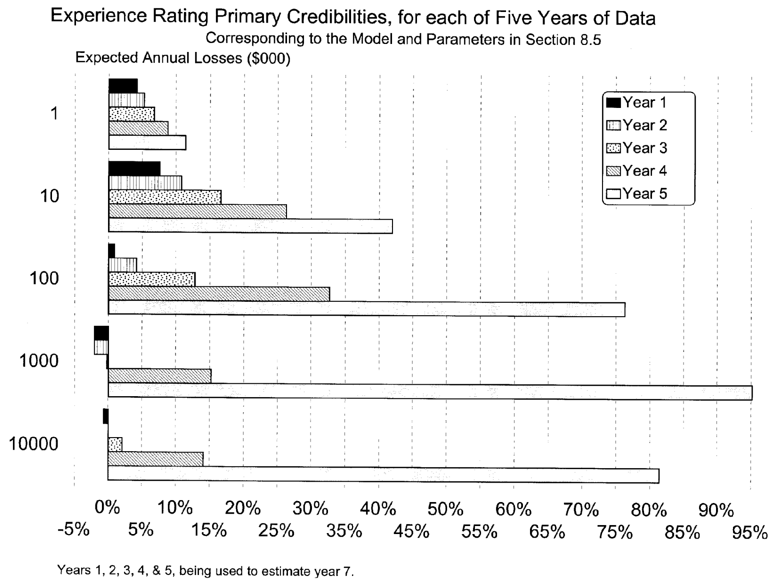


Currently each of three years used for experience rating is treated similarly. Instead each year could receive different credibilities. Figure 33 displays the model primary credibilities assigned to each of five years of data for various sizes of risks. Note that the weights assigned to an individual year of primary losses can be negative.[221] Figure 34 similarly displays the excess credibilities. The same pattern is observed in each figure, although for a given size of risk the weights given to different years are more similar to each other in the excess case than in the primary case.

It should be noted that for simplicity, equal expected losses have been assumed for each year. Equations 10.6 to 10.11 and Equations 10.12 and 10.13 apply equally well when the expected

[221] Also, the weights assigned to individual years of primary losses can in theory be greater than 100%.

FIGURE 33

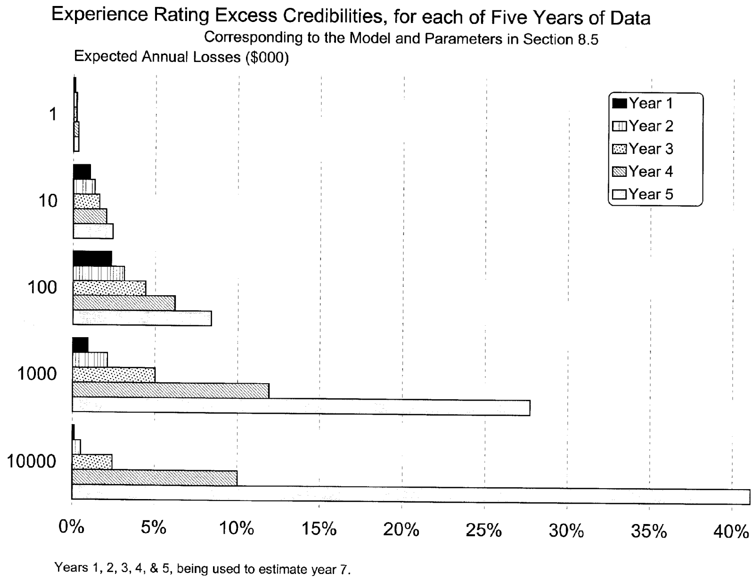


losses differ by year. In that case, years with more expected losses get more credibility than they would otherwise receive. The pattern due to varying volumes of data by year would be superimposed on that shown in Figures 33 and 34.

10.10. Taking Into Account Differences in the Maturity of the Experience Rating Data

Generally the data used for experience rating is at different reports. For workers compensation, generally three years of Unit Statistical Plan data is used for experience rating. For example, 1995 at first report, 1994 at second report, and 1993 at third report, might be used to experience rate a 1997 policy. The fact that the data are not at ultimate can affect the credibilities in two ways. First, as in Section 7.10, since the 1995 data is at an earlier report than the 1993 data, the 1995 data is a poorer estimator of 1997 ultimate losses compared to the 1993 data,

FIGURE 34



than if both 1993 and 1995 were at the same maturity.[222] Thus, the lack of maturity of the 1995 data reduces its value relative to the 1993 data and thus the credibility assigned to the 1995 data. In addition, all of the years of data are not at ultimate. Thus, they are all somewhat worse estimators than if they were available at ultimate. Thus, they all receive somewhat less credibility.[223]

As in Section 7.10, it will be assumed that the effect of loss development is to reduce the covariances between data at different reports. This refinement to the covariance structure will have the expected impact on the credibilities.

[222]Since 1995 is more recent than 1993, it is a better estimator of 1997, all other things being equal.
[223]Recall that unlike in Section 7.10, here the complement of credibility is given to the grand mean.

The first step is to estimate the correlations between the same experience rating data but at different reports. As before, various size categories will be used. Also, we can look at correlations between Primary Deviation Ratios, between Excess Deviation Ratios and between Primary and Excess Deviation Ratios.

For example, for risks with expected annual losses between \$10,000 and \$20,000, for the data at first report during Rating Year 91 and second report during Rating Year 92, the correlation of Primary Deviation Ratios is .937. Similar correlations can be calculated for Rating Year 92 vs. Rating Year 93, Rating Year 93 vs. Rating Year 94, and Rating Year 94 vs. Rating Year 95.[224] These four correlations have been averaged and are displayed in Table 11 as .942.

Table 11 displays similar correlations for other size categories as well as correlations for 2nd report vs. 3rd report and 1st report vs. 3rd report data.

The correlations between Primary Deviation Ratios and the correlations between Excess Deviation Ratios for different reports can be used directly, since for the same reports the correlation is one. However, for the mixed correlation between Primary Deviation Ratios and Excess Deviation Ratios, one would have to compare the correlation for different reports to the correlation for the same report, appropriately adjusted. Unfortunately this is not a practical solution,[225] therefore, the observed mixed correlations will not be used.

The primary and excess correlations in Table 11 do not display an obvious dependence on size of risk over the size categories examined.[226]

---

[224]These correlations are .952, .929, .949, illustrating the random fluctuation in the individual estimates for a given size category based on data from a single state.

[225]The actual correlations for a single report include a term involving the process variance. Unlike what was done in Section 7.11, we should not just totally remove this piece for comparison purposes since the different reports are not independent realizations of the same risk process, nor are the primary and excess losses independent.

[226]It is expected that the correlations will get closer to unity for very large risks, based on the analysis of classification data in Section 7.11.

TABLE 11

EXPERIENCE RATING, CORRELATIONS BETWEEN SAME DATA AT DIFFERENT REPORTS

Expected Annual Losses ($000)	1st to 2nd Corr.	1st to 2nd # Risks	2nd to 3rd Corr.	2nd to 3rd # Risks	1st to 3rd Corr.	1st to 3rd # Risks
			Primary Deviation Ratios			
3 to 5	.929	3404	.956	4052	.923	3447
5 to 10	.948	4212	.966	5247	.924	4229
10 to 20	.942	2742	.962	3458	.914	2764
20 to 30	.960	1078	.975	1359	.942	1073
30 to 50	.964	909	.972	1144	.941	900
50 to 100	.954	700	.975	900	.942	696
100 to 200	.955	330	.974	437	.940	332
200 to 500	.897	172	.958	222	.913	175
500 to 1,000	.890	46	.987	60	.877	52
1,000 to 2,000	.986	16	.987	19	.974	14
			Excess Deviation Ratios			
3 to 5	.833	3404	.893	4052	.758	3447
5 to 10	.836	4212	.910	5247	.785	4229
10 to 20	.861	2742	.897	3458	.769	2764
20 to 30	.852	1078	.930	1359	.804	1073
30 to 50	.877	909	.912	1144	.803	900
50 to 100	.858	700	.925	900	.798	696
100 to 200	.864	330	.919	437	.809	332
200 to 500	.779	172	.932	222	.720	175
500 to 1,000	.831	46	.930	60	.762	52
1,000 to 2,000	.924	16	.918	19	.857	14
			Mixed Deviation Ratios			
3 to 5	.561	3404	.574	4052	.552	3447
5 to 10	.564	4212	.584	5247	.554	4229
10 to 20	.546	2742	.569	3458	.546	2764
20 to 30	.585	1078	.610	1359	.605	1073
30 to 50	.529	909	.538	1144	.515	900
50 to 100	.585	700	.596	900	.576	696
100 to 200	.574	330	.607	437	.580	332
200 to 500	.530	172	.606	222	.540	175
500 to 1,000	.588	46	.648	60	.504	52
1,000 to 2,000	.551	16	.557	19	.504	14

Excluding the smallest and largest size intervals (with the least data), the average correlations between the different reports are:

	1–2	2–3	1–3
Primary	.939	.971	.924
Excess	.845	.919	.781

For illustrative purposes the following adjustments for loss development up to third report will be made to the primary, excess, and mixed covariances:[227]

	1–2	2–3	1–3
Primary Adjustment Factor	.94	.97	.92
Excess Adjustment Factor	.84	.92	.78
Mixed Adjustment Factor	.89	.94	.85

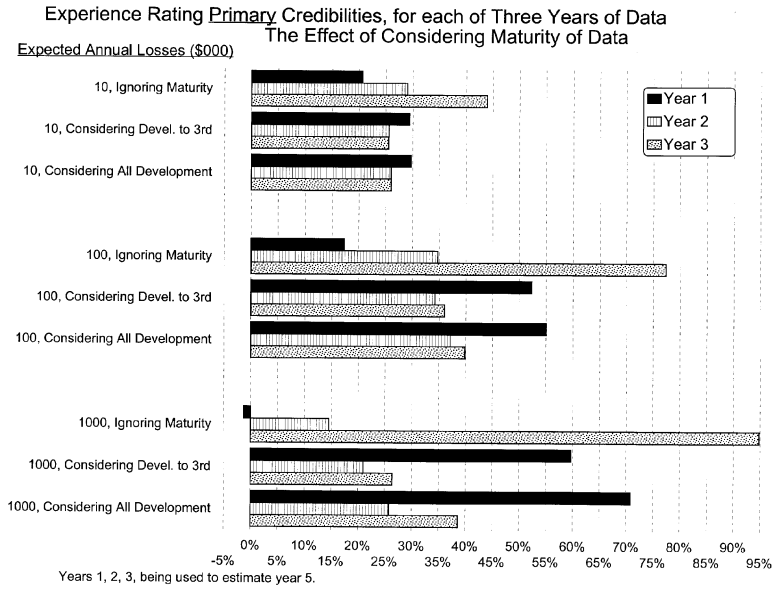
Using the parameters in Section 10.9, prior to any adjustment for differences in maturity, for $100,000 in expected annual losses we obtain the credibilities for Years 1, 2 and 3 predicting Year 5 shown both in Table 10 and in the first row below.

	Primary				Excess				Combined
	Year 1	Year 2	Year 3	Total	Year 1	Year 2	Year 3	Total	
No Adjustment for Maturity[228]	17.3%	34.7%	77.3%	129.3%	5.0%	6.6%	8.7%	20.3%	44.3%
Adjusting for Development to Third Report	25.3%	36.5%	67.9%	129.8%	5.7%	6.3%	5.8%	17.8%	42.4%
Adjusting for Development Both to Third Report and Beyond Third Report	27.4%	39.0%	69.7%	136.1%	4.1%	4.5%	3.7%	12.3%	39.5%

[227]These adjustment factors will only be applied to risks with expected annual losses between about $5,000 and $1 million. Risks outside that range would have adjustment factors that have not been estimated.

[228]See Table 10.

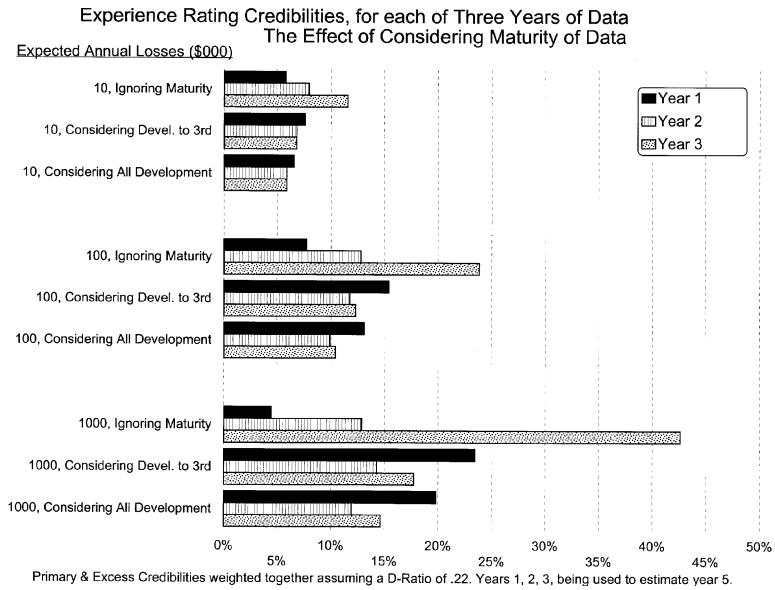
FIGURE 35



If we multiply all of the primary covariances between a year at first report and second report by an adjustment factor of .94, between second and third by .97, and between first and third by .92, with similar adjustments applied to the excess and mixed covariances, then the calculated credibilities are revised as shown in the second row above. The primary losses for Year 3 data at first report received less weight than when the maturity of the data was ignored. The primary losses for Year 1 at third report receive more weight. The overall credibility goes down somewhat.

Figure 35 displays the impact on primary credibilities for various sizes of risk, for each year separately. Figure 36 displays the impact on combined primary and excess credibilities for each year separately. Taking into account development up to third report alters the credibilities assigned to individual years; more

FIGURE 36



mature years get more weight while less mature years get less weight. Figure 37 shows the effect on the combined credibilities summed for the three years. The overall credibility is reduced by about 2% due to the consideration of development to third report.

The covariances are also affected by loss development beyond third report. The vast majority of such development affects excess losses rather than primary losses.[229] For illustrative purposes it will be assumed that development beyond third report reduces all the excess covariances between the data years and the year to be predicted (at ultimate) by .84, the adjustment factor

[229]In Massachusetts workers compensation, most claims of size less than $5,000 are closed by third report. Most claims open at third report have incurred amounts at third report that exceed $5,000 and also settle for more than $5,000.

for development from the first to the third report. The mixed covariances will be adjusted by a factor of .92, while the primary covariances are not adjusted at all.

The resulting credibilities were shown in the third row above for a risk with $100,000 in expected annual losses. Taking into account loss development beyond third report in this manner reduced the relative value of the excess losses as a predictor. Therefore, the credibilities assigned to the excess losses decreased, while those assigned to the primary losses increased.

Figures 35 to 37 compare the credibilities including the impacts of loss development to ultimate to those excluding any consideration of maturity as well as those including the impacts of loss development to third report. As expected, the inclusion of all loss development generally lowers the credibilities.[230]

10.11. Conclusions-Experience Rating

While similar analyses of experience rating have been made in the past,[231] the present analysis incorporates shifting risk parameters, risk heterogeneity and parameter uncertainty in a comprehensive and integrated manner. While the example was for a single split experience rating plan for workers compensation, a similar analysis should be valuable for any experience rating type situation where the volume of data varies significantly between entities. For example, general liability experience rating or the assignment of towns to territories[232] would fall into this category.

On the other hand, situations such as private passenger automobile Safe Driver Insurance Plans or Bonus–Malus plans would allow a somewhat simpler analysis, since the size of the insured

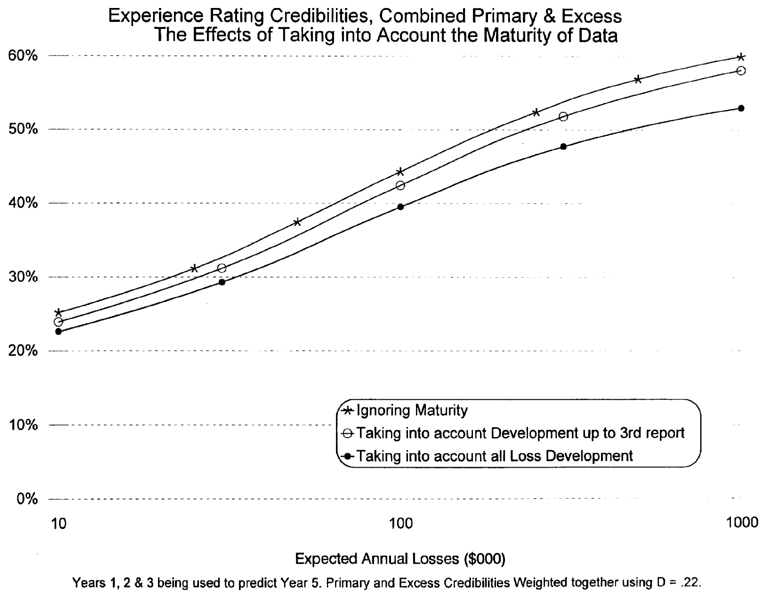
---

[230]Recall that the adjustment factors were illustrative and not based on any specific experience rating data beyond third report.

[231]See, for example, Mahler [11] and Finger [29].

[232]See Conger [30].

FIGURE 37



is not significant.[233] The phenomenon of risk heterogeneity is not important in that case. Thus, the situation is the special case examined in Section 3.6, where parameter uncertainty and shifting risk parameters are important. In that case, we expect a covariance structure of the form:[234]

$$\mathrm{Cov}[X_i, X_j] = r^2 \rho^{|i-j|} + \delta_{ij} d^2. \tag{10.19}$$

The size of risk E has been suppressed as not important in this case, and therefore the variance due to parameter uncertainty and that due to the expected value of the process variance can be combined into one term d^2. Equation 10.19 has the same form

[233] In medical malpractice, as discussed in Finger [29], the simpler situation is that of experience rating individual doctors, while the more general situation would involve experience rating groups of doctors.

[234] Compare to Equation 3.19.

as Equation 7.1 in Mahler [1]. Thus, the form of analysis in Mahler [1] should suffice in the case of frequency based private passenger automobile experience rating and similar situations.

11. MISCELLANEOUS

In this section the methods of Mahler [1] will be applied to the estimation of the market risk premium, the baseball models from Mahler [1] and Mahler [20] will be revisited, and the results in Boor [31] will be related to those herein.

11.1. Market Risk Premium

The market risk premium, an important economic concept used in the Capital Asset Pricing Model, is the excess return on stocks expected beyond the risk-free rate. A common estimate of the market risk premium is the difference between the return on large company stocks and the return on three-month U.S. Treasury Bills.[235] Table 12 shows this series from 1926 through 1995.

This series is very volatile. Ibbotson [32] recommends using a long-term (unweighted) average based on a belief that the expected real returns have been reasonably consistent over time. Using the currently available data from 1926 to 1995, the unweighted average is 8.76%.

While the risk parameters underlying this process are relatively stable, they are unlikely to have absolutely no shifting over time. The methods developed in Mahler [1] can be used to estimate the sensitivity of the estimated market risk premium to assumptions about the stability of the risk process.

Let X_i be the observed difference between the return on large company stocks and U.S. Treasury Bills for year i. Then one estimate of the market risk premium is to take all Y years of data

[235] See Chapter 8 of Ibbotson [32]. The market risk premium is referred to as the equity risk premium.

TABLE 12

PART 1

TOTAL RETURN

Year	Large Company Stocks	U.S. Treasury Bills	Difference
1926	11.62	3.27	8.35
1927	37.49	3.12	34.37
1928	43.61	3.56	40.05
1929	−8.42	4.75	−13.17
1930	−24.90	2.41	−27.31
1931	−43.34	1.07	−44.41
1932	−8.19	0.96	−9.15
1933	53.99	0.30	53.69
1934	−1.44	0.16	−1.60
1935	47.67	0.17	47.50
1936	33.92	0.18	33.74
1937	−35.03	0.31	−35.34
1938	31.12	−0.02	31.14
1939	−0.41	0.02	−0.43
1940	−9.78	0.00	−9.78
1941	−11.59	0.06	−11.65
1942	20.34	0.27	20.07
1943	25.90	0.35	25.55
1944	19.75	0.33	19.42
1945	36.44	0.33	36.11
1946	−8.07	0.35	−8.42
1947	5.71	0.50	5.21
1948	5.50	0.81	4.69
1949	18.79	1.10	17.69
1950	31.71	1.20	30.51
1951	24.02	1.49	22.53
1952	18.37	1.66	16.71
1953	−0.99	1.82	−2.81
1954	52.62	0.86	51.76
1955	31.56	1.57	29.99
1956	6.56	2.46	4.10
1957	−10.78	3.14	−13.92
1958	43.36	1.54	41.82
1959	11.96	2.95	9.01
1960	0.47	2.66	−2.19

TABLE 12

PART 2

TOTAL RETURN

Year	Large Company Stocks	U.S. Treasury Bills	Difference
1961	26.89	2.13	24.76
1962	−8.73	2.73	−11.46
1963	22.80	3.12	19.68
1964	16.48	3.54	12.94
1965	12.45	3.93	8.52
1966	−10.06	4.76	−14.82
1967	23.98	4.21	19.77
1968	11.06	5.21	5.85
1969	−8.50	6.58	−15.08
1970	4.01	6.52	−2.51
1971	14.31	4.39	9.92
1972	18.98	3.84	15.14
1973	−14.66	6.93	−21.59
1974	−26.47	8.00	−34.47
1975	37.20	5.80	31.40
1976	23.84	5.08	18.76
1977	−7.18	5.12	−12.30
1978	6.56	7.18	−0.62
1979	18.44	10.38	8.06
1980	32.42	11.24	21.18
1981	−4.91	14.71	−19.62
1982	21.41	10.54	10.87
1983	22.51	8.80	13.71
1984	6.27	9.85	−3.58
1985	32.16	7.72	24.44
1986	18.47	6.16	12.31
1987	5.23	5.47	−0.24
1988	16.81	6.35	10.46
1989	31.49	8.37	23.12
1990	−3.17	7.81	−10.98
1991	30.55	5.60	24.95
1992	7.67	3.51	4.16
1993	9.99	2.90	7.09
1994	1.31	3.90	−2.59
1995	37.43	5.60	31.83
Average	**12.52%**	**3.77%**	**8.76%**

Source: Ibbotson [23], Table 2–5.

and average:

$$\text{Estimate} = \sum_{i=1}^{Y} \frac{1}{Y} X_i.$$

More generally, we can weight together the X_i using weights Z_i such that $\sum Z_i = 1$:

$$\text{Estimate} = \sum_{i=1}^{Y} Z_i X_i.$$

The unweighted average is just a special case, with $Z_i = 1/Y$ for all years.

When parameters shift over time we would expect to have a covariance structure as per Equation 2.1:

$$\mathrm{Cov}[X_i, X_j] = e^2 \delta_{ij} + r^2 \rho^{|i-j|}, \qquad \text{where} \quad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}.$$

Equations 6.7 for the weights Z_i that minimize the expected squared error of the estimate of year $Y+1$ are:

$$\sum_{i=1}^{Y} Z_i \mathrm{Cov}[X_i, X_k] = \mathrm{Cov}[X_i, X_{Y+1}] + \lambda/2, \qquad k = 1, 2, \ldots Y,$$

where λ is the Lagrange multiplier. We can solve these Y linear equations plus the constraint equation for the desired weights Z_i. Given an assumed covariance structure, we can obtain weights and in turn use them to estimate the market risk premium.

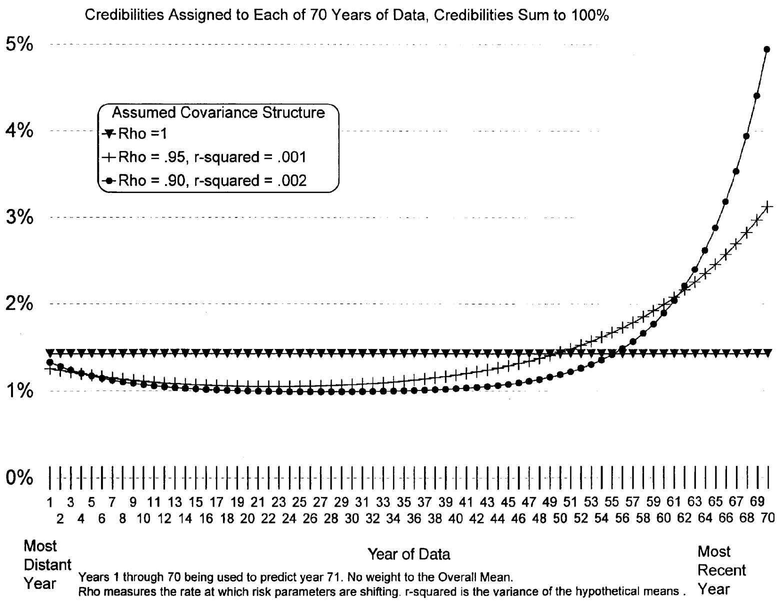
The variance of X is very large, about .0427.[236] The correlations are small.[237] Due to the large random fluctuations there is insufficient data to estimate the parameters of the covariance structure. However, we do have:

$$\mathrm{Cov}[X, X] = \mathrm{Var}[X] = e^2 + r^2 \approx .0427,$$

[236]The standard deviation is .207 compared to the mean of .0876.

[237]They are not statistically different from zero.

FIGURE 38



$$\begin{aligned}
\mathrm{Corr}[X_i, X_{i+1}] &= \mathrm{Cov}[X_i, X_{i+1}]/\sqrt{\mathrm{Var}[X_i]\mathrm{Var}[X_{i+1}]} \\
&= r^2\rho/(e^2 + r^2) \\
&\approx r^2/(e^2 + r^2).
\end{aligned}$$

Since the correlations between successive years are close to zero, r^2 is much smaller than e^2. For example, if $r^2 = .0005$ and $e^2 = .0422$ then $\mathrm{Corr}[X_1, X_2] \approx 1.2\%$. The effect of varying r^2 between .0005 and .0020 has been tested.

Ibbotson [32] believes that the parameters are *not* shifting rapidly. The parameter ρ measures the rate of shifting. For slow shifting, ρ is near 1. The effect of varying ρ between 1 and .90 has been tested. Figure 38 shows examples of the credibilities for various values of ρ and r^2.

TABLE 13

SENSITIVITY OF ESTIMATED MARKET RISK PREMIUM

$r^2(.0001)$	ρ	Estimated Market Risk Premium
5	1	8.76%[1]
5	.975	8.61
5	.95	8.68
5	.90	8.82
10	1	8.76[1]
10	.975	8.52
10	.95	8.67
10	.90	8.91
20	1	8.76[1]
20	.975	8.47
20	.95	8.75
20	.90	9.13

Based on seventy years of data from 1926–1995, (see Table 12).
Assuming total variance of .0427.
[1]Result of straight average.

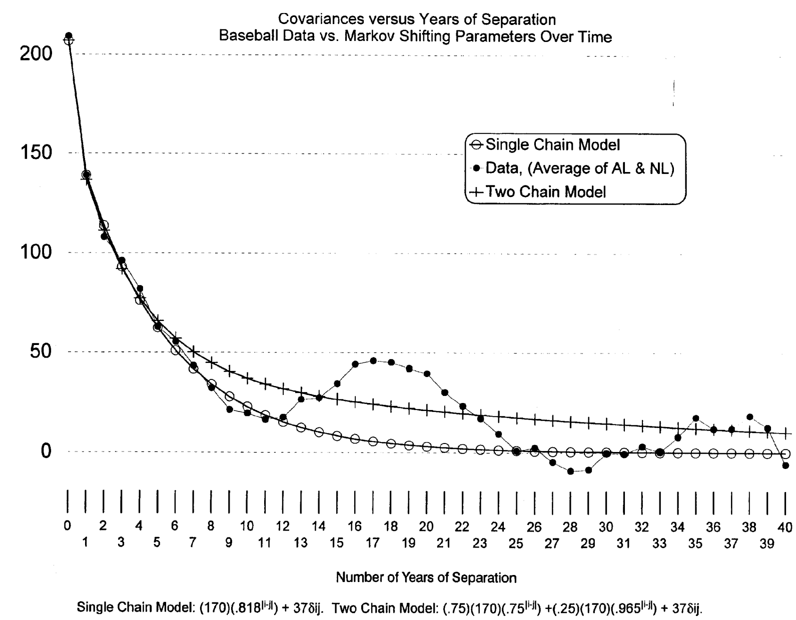
As shown in Table 13, the estimated market risk premium is relatively insensitive to the choice of the parameters of the covariance structure. Any reasonable set of inputs gives an answer in the same range. Bear in mind that just adding the 1995 data point changed the estimate of the market risk premium based on an unweighted average from 8.4% to 8.8%.

In conclusion, taking the straight unweighted average of the available data remains a reasonable method of estimating the market risk premium. While technical refinements could be included to take into account shifting risk parameters, they would not substantially improve or alter the estimate.

11.2. Baseball Example, Revisited

In Mahler [1], the data for the won-loss records of baseball teams was approximated by a model involving a single Markov chain with half-life of about $3\frac{1}{2}$ years. When expressed in terms

FIGURE 39



of games lost, the covariances between years of data X_i and X_j are:

$$\mathrm{Cov}[X_i, X_j] \approx (170)(.818^{|i-j|}) + (37)\delta_{ij}.$$

We could instead use a model involving two Markov chains with different half-lives. This would allow us to approximate the apparent longer term slower shifting as well as the shorter term rapid shifting. While the volume of data is not sufficient to allow us to fit a unique "two-chain" model, as seen in Figure 39 the following does a reasonable job:

$$\mathrm{Cov}[X_i, X_j] = (127.5)(.75^{|i-j|}) + (42.5)(.965^{|i-j|}) + 37\,\delta_{ij}.$$

The more swiftly shifting Markov chain has a dominant eigenvalue of .75 and a half-life[238] of about $2\frac{1}{2}$ years. The more slowly

[238] $(\ln 2) \div (\ln .75) = 2.4$.

shifting Markov chain has a dominant eigenvalue of .965 and a half-life[239] of about $19\frac{1}{2}$ years. The two Markov chains are given 75% weight and 25% weight, respectively.

This is an example of how two or more Markov chains of different half-lives could be used to attempt to model the different sources of shifting parameters over time.[240] Note that this data set does not lend itself to an examination of credibilities versus size of risk, since the seasons do not vary sufficiently in the number of games.

11.3. Boor, "Credibility Based on Accuracy"

As shown in Appendix B, linear Equations 6.7 for the credibilities with no weight to the mean are closely related to those in Boor. One difference is that Boor assumes only two estimators,[241] while Equations 6.7 assume two or more estimators.

A more fundamental difference is that Equations 6.7 assume that each of the estimators is unbiased. In Boor [31] no such assumption is made, so the results in Boor [31] apply in more general situations than Equations 6.7. Since the estimators in Boor [31] are possibly biased, the formulas for credibility involve terms such as $E[X_1X_3]$, rather than $\mathrm{Cov}[X_1,X_3]$ such as in Equations 6.7.

12. SUMMARY AND CONCLUSIONS

In Sections 2 to 5, a general form of the covariances in the presence of shifting risk parameters, parameter uncertainty, and risk heterogeneity was developed. While a simple example using dice[242] was used to develop this covariance structure as shown

---

[239] $(\ln 2) \div (\ln .965) = 19.5$.

[240] Perhaps the chain with the shorter half-life relates to the baseball players while the chain with the longer half-life relates to shifts in management.

[241] The ideas in Boor [31] can be extended to more than two estimators. Boor presents the most common situation where two estimators are being combined.

[242] See Table 1.

in Equations 5.5, the model in Equations 5.10 and 5.11 is in a form appropriate for insurance applications.

Equation 3.10 in Section 3.4 shows that in the presence of parameter uncertainty there is a fundamentally different dependence of the credibility on the number of years of data and the size of risk in a single year. Section 3.7 discusses the fundamentally different dependence of the credibility on the number of years of data in the presence of shifting risk parameters versus parameter uncertainty. We can ameliorate the impact of parameter uncertainty by averaging over many years; in contrast, considering more than one year captures the effects of shifting risk parameters.

Section 5.2 includes a brief discussion of how we might estimate the parameters of the general covariance structure. Sections 7.3 to 7.6 and 7.11 illustrate how this might be done for classification data. Sections 10.4 to 10.8 and 10.10 illustrate how this might be done for experience rating data. While there are difficulties in estimating the required parameters, in general the results of applying credibility are relatively insensitive to the estimated parameters.[243]

Matrix equations are presented for calculating the (least squares) credibilities from the covariance structure. While Equations 2.4, 6.7, 8.1, 10.12 and 10.13 are similar, they each apply in a somewhat different situation. Each such set of linear equations depends on the covariance structure and can be solved for the credibilities using matrix methods.

Section 4.5 presents the credibilities in the important special case of stable (or very slowly shifting) risk parameters. Sections 4.3, 4.6, 4.7 and 5.3 explore the different behavior of credibilities expected for the smallest risks. As discussed in Section 4.8, the general covariance structure predicts the need for a minimum

[243]See Mahler [33].

ballast value in the revised Workers Compensation Experience Rating Plan.[244]

Sections 6.1 to 6.4, 7.7, 7.8, 7.12, 8.5, 8.7, 9.3 to 9.5, 10.9 and 10.10 contain illustrative calculations of credibilities. The general behaviors noted there should carry over to other similar situations.

Section 7 applies the ideas developed in this paper to an illustrative example of classification ratemaking for workers compensation. The parameters of the covariance structure were estimated in Sections 7.3 to 7.6. The behavior of the credibilities[245] when using data from one state was displayed by year[246] and by size of class in Sections 7.7 to 7.9.

Sections 7.10 to 7.12 illustrated the potential impact of the different maturity of the years of data on their credibilities. As expected, the most recent years of data at early reports get somewhat less weight than if we ignored the effects of different maturities.

Section 8 discusses how to incorporate data from outside the state. While the covariance structure has an extra layer of complication, it is still tractable. There are twice as many linear equations in twice as many unknowns, but they can still be easily solved for the credibilities. This general type of treatment should be useful whenever there is supplementary information analogous to the countrywide data.

Section 9 applies the ideas of this paper to an illustrative calculation of an overall rate indication. The effects on the weights

---

[244]The minimum ballast value was used based on practical considerations for almost a decade prior to the developments in this paper. It is pleasant to find an overall theoretical framework into which it fits.

[245]It is assumed in Section 7.9 that the complement of credibility is being given to the prior estimate of the class relativity. Section 7.7 assumed the weights assigned to the data sum to 100%, while Section 7.8 assumed the complement of credibility is given to the grand mean.

[246]The assignment of a separate weight to each year of data is an important refinement compared to the assignment of a single weight to the combined data for all years.

to be assigned to individual years of estimation errors, loss development and trend factors are discussed. Additional work would be required to adopt the general ideas presented to any particular situation. The general conclusions are far from surprising. When we have a smaller volume of data, we choose a more stable method; when we have a larger volume of data, we choose a more responsive method. Data subject to more estimation error is given less weight, all other things being equal.

Section 10 applies the ideas developed in this paper to workers compensation experience rating. This analysis should be useful for any commercial line in which the volume of data varies significantly from insured to insured. Sections 10.4 to 10.8 illustrate how we would estimate the parameters of the covariance structure in the case of a single split experience rating plan. Due to the effects of shifting parameters over time, the complicated behavior by size of risk, and the correlations of the primary and excess losses, the estimation of parameters is difficult and of necessity requires some judgment. Section 10.9 shows a sample calculation of the credibilities. The credibilities are displayed by year and by size of risk. Section 10.10 illustrates how to incorporate the impact of the different maturities of the data.

Section 11 contains miscellaneous results. The methodology is applied to an economic index, generalized to two Markov chains, and related to that in Boor [31].

In each of the various examples presented, there are three steps. First, we must specify the covariance structure.[247] Second, we must estimate the parameters of the covariance structure. Third, we must solve the appropriate set of linear equations for the credibilities.[248]

We live in a dynamic rather than stable environment. Therefore, the ideas presented in this paper on the covariance structure and resulting credibilities should have application in many

[247] See for example Table 2.
[248] See Table 3.

areas of actuarial work where risk parameters shift significantly over time. The methods presented can help answer fundamental questions about how many years of data to use in a particular situation and whether certain years of data should get significantly more weight than others. One needs to estimate how stable is the particular real world situation; how swiftly are risk parameters shifting over time?

REFERENCES

[1] Mahler, Howard C., "A Markov Chain Model for Shifting Risk Parameters," *PCAS* LXXXIV, 1997, pp. 581–659.

[2] Mayerson, Allen L., "A Bayesian View of Credibility," *PCAS* LI, 1964, pp. 85–104.

[3] Hewitt, Charles C., Jr., Discussion of Mayerson: "A Bayesian View of Credibility," *PCAS* LII, 1965, pp. 121–127.

[4] Hewitt, Charles C., Jr., "Credibility for Severity," *PCAS* LVII, 1970, pp. 148–171.

[5] Philbrick, Stephen W., "An Examination of Credibility Concepts," *PCAS* LXVIII, 1981, pp. 195–219.

[6] Herzog, Thomas N., *An Introduction to Credibility*, ACTEX, 1994.

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

[8] Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot, *Loss Models: From Data to Decisions*, New York: John Wiley and Sons, 1998.

[9] Mahler, Howard C., "A Student's Guide to Bühlmann Credibility and Bayesian Analysis," *New England Actuarial Seminars*, 1997.

[10] Meyers, Glenn G., "An Analysis of Experience Rating," *PCAS* LXXII, 1985, pp. 278–317.

[11] Mahler, Howard C., Discussion of Meyers: "An Analysis of Experience Rating," *PCAS* LXXIV, 1987, pp. 119–189.

[12] Mahler, Howard C., Discussion of Gillam: "Parametrizing the Workers Compensation Experience Rating Plan," *PCAS* LXXXX, 1993, pp. 148–183.

[13] Gillam, William R., "Parametrizing the Workers Compensation Experience Rating Plan," *PCAS* LXXIX, 1992, pp. 21–56.

[14] Hewitt, Charles C., Jr., “Loss Ratio Distributions—A Model,” *PCAS* LIV, 1967, pp. 70–88.

[15] Whitney, Albert W., “The Theory of Experience Rating,” *PCAS* IV, 1918, pp. 274–292.

[16] Michelbacher, Gustav F., “The Practice of Experience Rating,” *PCAS* IV, 1918, pp. 293–324.

[17] Mahler, Howard C., “Credibility: Practical Applications,” Casualty Actuarial Society *Forum*, Fall 1989, pp. 187–199.

[18] Mahler, Howard C., “Introduction to Basic Credibility,” Casualty Actuarial Society *Forum*, Winter 1997, pp. 67–103.

[19] Gillam, William R. and Richard H. Snader, “Fundamentals of Individual Risk Rating,” 1992, available from the CAS.

[20] Mahler, Howard C., “An Example of Credibility and Shifting Risk Parameters,” *PCAS* LXXVII, 1990, pp. 225–308.

[21] Finger, Robert J., “Risk Classification,” Chapter 5 of *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, 1990.

[22] Kallop, Roy H., “A Current Look at Workers Compensation Ratemaking,” *PCAS* LXII, 1975, pp. 62–133.

[23] Feldblum, Sholom, “Workers Compensation Ratemaking (selected chapters),” Casualty Actuarial Society *Forum*, February 1993, pp. 241–312.

[24] Harwayne, Frank, “Use of National Experience Indicators in Workers Compensation Insurance Classification Ratemaking,” *PCAS* LXIV, 1977, pp. 74–84.

[25] Meyers, Glenn G., “Empirical Bayesian Credibility for Workers Compensation Classification Ratemaking,” *PCAS* LXXI, 1984, pp. 96–121.

[26] Busche, George R., “A Method to Include Multiple Years of Data in a Company’s Rate Indication,” Casualty Actuarial Society *Forum*, February 1993, pp. 401–420.

[27] Bouska, Amy S., “Exposure Bases Revisited,” *PCAS* LXXVI, 1989, pp. 1–23.

[28] Stern, Phillip K., "Ratemaking Procedures for Automobile Liability Insurance," *PCAS* LII, 1965, pp. 139–202.

[29] Finger, Robert J., "Merit Rating for Doctor Professional Liability Insurance," *PCAS* LXXX, 1993, pp. 291–352.

[30] Conger, Robert F., "The Construction of Automobile Territories in Massachusetts," *PCAS* LXXIV, 1987, pp. 1–74.

[31] Boor, Joseph A., "Credibility Based on Accuracy," *PCAS* LXXIX, 1992, pp. 166–185.

[32] Stocks, Bonds, Bills and Inflation, 1996 Yearbook, Ibbotson Associates, Chicago.

[33] Mahler, Howard C., "An Actuarial Note on Credibility Parameters," *PCAS* LXXIII, 1986, pp. 1–26.

[34] Feller, William, *An Introduction to Probability Theory and its Applications*, New York: John Wiley and Sons, 1968.

[35] Resnick, Sidney I., *Adventures in Stochastic Processes*, Boston: Birkhauser, 1992.

[36] DeVlyder, F. Etienne, *Advanced Risk Theory—A Self-Contained Introduction*, University of Brussels, Belgium, 1996.

APPENDIX A

MARKOV CHAINS[247]

Assume that each year[248] an individual is in a "state." Each state could correspond to a different average claim frequency. Assume that there are a finite number of different states.

Assume with each new year that an individual in State i has a chance P_{ij} of going to State j. This chance is independent of which individual we have picked, what its past history was, or what year it is. The transition probability from State i to State j, P_{ij}, is only dependent on the two States, i and j.

Arrange these transition probabilities, P_{ij}, into a matrix P. This transition matrix P, together with the definition of the states, defines a (finite dimensional) Markov chain.

A vector containing the probability of finding an individual in each of the possible states is called a "distribution." If the distribution in Year 1 is β, then the expected distribution in Year 2 is βP, where βP is the matrix product of the (row vector) distribution β and the transition matrix P. The expected distribution in Year 3 is $(\beta P)P = \beta(PP) = \beta P^2$. The expected distribution in Year $1 + g$ is βP^g.

Let P^T be the matrix transpose of P. Let Λ be the diagonal matrix with entries equal to the eigenvalues of P^T. Let V^T be the matrix each of whose columns are the eigenvectors of P^T. (V has as its rows the eigenvectors of P^T.) Then $(V^T)^{-1}P^T V^T = \Lambda$. Taking the transpose of both sides of this equation and noting that $\Lambda^T = \Lambda$, since Λ is symmetric: $VPV^{-1} = \Lambda$. So the matrix V can be used to diagonalize the transition matrix P:

$$V^{-1}\Lambda^2 V = V^{-1}(VPV^{-1})^2 V = V^{-1}VPV^{-1}VPV^{-1}V = P^2.$$

[247] See Feller [34], Resnick [35], and Appendix A in Mahler [1].

[248] Although in this paper the time interval is a year, in general, it can be anything else.

In general, $P^g = V^{-1}(VPV^{-1})^g V = V^{-1}\Lambda^g V$. So powers of P can be computed by taking powers of the diagonal matrix Λ and using the eigenvector matrix V to transform back. The elements of the diagonal matrix Λ^g are λ_i^g. $\lambda_1 = 1$ (the order of eigenvalues is arbitrary) and $|\lambda_i| < 1$ for $i > 1$ (ignoring the very unusual situation where $\lambda = 1$ is a multiple root of the characteristic equation).[249]

[249]If for any i, $|\lambda_i| > 1$, then there would be no limiting distribution, instead it would blow up. However, a finite dimensional Markov chain such that each state can be reached from every other state and such that no states are periodic has a unique stationary distribution, which is the limit as time goes to infinity. If for all i, $|\lambda_i| < 1$, then again there would be no non-zero limit, instead it would go to zero. Thus, we have all $|\lambda_i| \leq 1$ and at least one $|\lambda_i| = 1$.

APPENDIX B

MATRIX EQUATIONS FOR LEAST SQUARES CREDIBILITY WITH NO WEIGHT TO GRAND MEAN[250]

In this appendix, Equations 6.7 in the main text are derived by minimizing the squared error. The result is one constraint equation plus Y linear equations for the credibilities to be assigned to each of Y years of data. Thus the credibilities can be solved for in terms of the covariance structure. Also, the related result in Boor [31] is derived.

Let

$$\begin{aligned} C_{ij} &= \mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j] \\ &= \text{Covariance of Year } X_i \text{ and Year } X_j, \\ C_{ii} &= \text{Variance of Year } X_i, \qquad \text{and} \\ \mu_i &= \mathrm{E}[X_i] = \text{Expected value for Year } X_i. \end{aligned}$$

Let Z_i be the credibility assigned to Year X_i. We wish to predict Year $X_{Y+\Delta}$ using Y years of data $X_1, X_2, \ldots, X_Y$. Assume $\sum_{i=1}^{Y} Z_i = 1$.

Then the estimate is:

$$F = \sum_{i=1}^{Y} Z_i X_i \qquad \text{and}$$

$$F - X_{Y+\Delta} = \left(\sum_{i=1}^{Y} Z_i X_i\right) - X_{Y+\Delta} = \sum_{i=1}^{Y} Z_i (X_i - X_{Y+\Delta})$$

since $\sum_{i=1}^{Y} Z_i = 1$.

[250] The derivation is along the same lines as those in Mahler [20] and Mahler [1].

Therefore,

$$
\begin{aligned}
(F - X_{Y+\Delta})^2 &= \left(\sum_{i=1}^{Y} Z_i(X_i - X_{Y+\Delta})\right)\left(\sum_{j=1}^{Y} Z_j(X_j - X_{Y+\Delta})\right) \\
&= \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j (X_i - X_{Y+\Delta})(X_j - X_{Y+\Delta}).
\end{aligned}
$$

Then the expected value of the squared difference between the estimate F and $X_{Y+\Delta}$ is, as a function of the credibilities Z,

$$
\begin{aligned}
V(Z) &= \mathrm{E}[(F - X_{Y+\Delta})^2] \\
&= \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j \mathrm{E}[(X_i - X_{Y+\Delta})(X_j - X_{Y+\Delta})].
\end{aligned}
$$

Now

$$
\begin{aligned}
\mathrm{E}[(X_i - X_{Y+\Delta})(X_j - X_{Y+\Delta})] &= \mathrm{E}[X_i X_j] - \mathrm{E}[X_i X_{Y+\Delta}] \\
&\quad - \mathrm{E}[X_j X_{Y+\Delta}] + \mathrm{E}[X_{Y+\Delta}^2] \\
\mathrm{E}[X_i X_j] &= \mathrm{Cov}[X_i, X_j] + \mathrm{E}[X_i]E[X_j] \\
&= C_{ij} + \mu_i \mu_j.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathrm{E}[(X_i - X_{Y+\Delta})(X_j - X_{Y+\Delta})] \\
&\quad = C_{ij} - C_{i,Y+\Delta} - C_{j,Y+\Delta} + C_{Y+\Delta,Y+\Delta} \\
&\qquad + \mu_i\mu_j - \mu_i\mu_{Y+\Delta} - \mu_j\mu_{Y+\Delta} + \mu_{Y+\Delta}^2. \\
&V(Z) = \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j \\
&\qquad \times \{C_{ij} - C_{i,Y+\Delta} - C_{j,Y+\Delta} + C_{Y+\Delta,Y+\Delta} + \mu_i\mu_j \\
&\qquad\quad - \mu_i\mu_{Y+\Delta} - \mu_j\mu_{Y+\Delta} + \mu_{Y+\Delta}^2\}
\end{aligned}
$$

$$
\begin{aligned}
V(Z) = {} & \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j (C_{ij} + \mu_i\mu_j) \\
& - \left(\sum_{i=1}^{Y}(C_{i,Y+\Delta} + \mu_i\mu_{Y+\Delta})Z_i\right)\left(\sum_{j=1}^{Y} Z_j\right) \\
& - \left(\sum_{j=1}^{Y}(C_{j,Y+\Delta} + \mu_j\mu_{Y+\Delta})Z_j\right)\left(\sum_{i=1}^{Y} Z_i\right) \\
& + (C_{Y+\Delta,Y+\Delta} + \mu_{Y+\Delta}^2)\left(\sum_{i=1}^{Y} Z_i\right)\left(\sum_{j=1}^{Y} Z_j\right).
\end{aligned}
$$

The last three terms all simplify since

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

Therefore,

$$
\begin{aligned}
V(Z) = {} & \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j (C_{ij} + \mu_i\mu_j) - 2\sum_{i=1}^{Y}(C_{i,Y+\Delta} + \mu_i\mu_{Y+\Delta})Z_i \\
& + C_{Y+\Delta,Y+\Delta} + \mu_{Y+\Delta}^2.
\end{aligned}
$$

We can minimize $V(Z)$ given the constraint $\sum_{i=1}^{Y} Z_i - 1 = 0$ by using Lagrange multipliers. We set equal to zero the partial derivative with respect to Z_k of $V(Z) - \lambda(\sum_{i=1}^{Y} Z_i - 1)$:

$$2\sum_{i=1}^{Y} Z_i(C_{ik} + \mu_i\mu_k) - 2(C_{k,Y+\Delta} + \mu_k\mu_{Y+\Delta}) - \lambda = 0.$$

Therefore,

$$\sum_{i=1}^{Y} Z_i(C_{ik} + \mu_i\mu_k) = C_{k,Y+\Delta} + \mu_k\mu_{Y+\Delta} + \lambda/2 \qquad k = 1,2,\ldots,Y.$$

Also

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

Thus we obtain $Y+1$ linear equations in $Y+1$ unknowns (the credibility assigned to each of Y years and the Lagrange multiplier λ).

If we assume each of the years X_i is an unbiased estimator of $X_{Y+\Delta}$, then $\mathrm{E}[X_i] = \mathrm{E}[X_{Y+\Delta}]$, or $\mu_i = \mu_{Y+\Delta}$. The above equations reduce to:

$$\sum_{i=1}^{Y} Z_i C_{ik} + \mu_{Y+\Delta}^2 \sum_{i=1}^{Y} Z_i = C_{k,Y+\Delta} + \mu_{Y+\Delta}^2 + \lambda/2.$$

Since $\sum_{i=1}^{Y} Z_i = 1$, this becomes Equations 6.7 in the main text:

$$\sum_{i=1}^{Y} Z_i C_{ik} = C_{k,Y+\Delta} + \lambda/2, \qquad k = 1,2,\ldots,Y$$

Equation 6.7 as well as Equations 2.4, 8.1 and 10.12 to 10.13, as shown in Table 2, are all variations on the so-called "normal equations" for credibilities. See, for example, De Vlyder [36] for an extensive discussion of the relation of the covariance structure to the credibilities.

Boor, "Credibility Based on Accuracy"

The result in Boor [31] can be obtained as a special case of the above development as follows, making no assumption concerning whether $\mathrm{E}[X_i]$ equals $\mathrm{E}[X_{Y+\Delta}]$. Assume we have two estimators X_1 and X_2 that we are using to estimate X_3. Then we get 2 linear equations plus the constraint equation:[251]

$$Z_1(C_{11} + \mu_1\mu_1) + Z_2(C_{12} + \mu_1\mu_2) = C_{13} + \mu_1\mu_3 + \lambda/2,$$

$$Z_1(C_{12} + \mu_1\mu_2) + Z_2(C_{22} + \mu_2\mu_2) = C_{23} + \mu_2\mu_3 + \lambda/2, \qquad \text{and}$$

$$Z_1 + Z_2 = 1.$$

[251] Note that $C_{21} = C_{12}$.

Subtracting the first two equations eliminates the Lagrange multiplier λ:

$$Z_1(C_{11} - C_{12} + \mu_1^2 - \mu_1\mu_2) + Z_2(C_{12} - C_{22} + \mu_1\mu_2 - \mu_2^2)$$
$$= C_{13} - C_{23} + \mu_1\mu_3 - \mu_2\mu_3.$$

Substituting $Z_2 = 1 - Z_1$ and solving for Z_1:

$$Z_1 = \frac{C_{13} - C_{12} - C_{23} + C_{22} + \mu_1\mu_3 - \mu_1\mu_2 - \mu_2\mu_3 + \mu_2^2}{C_{11} - 2C_{12} + C_{22} + \mu_1^2 - 2\mu_1\mu_2 + \mu_2^2}.$$

As in Boor [31], define the following quantities:

$$\begin{aligned}
\tau_1^2 &= \mathrm{E}[(X_1 - X_3)^2] = \mathrm{E}[X_1^2] - 2\mathrm{E}[X_1X_3] + \mathrm{E}[X_3^2] \\
&= C_{11} + \mu_1^2 + C_{33} + \mu_3^2 - 2C_{13} - 2\mu_1\mu_3, \\
\tau_2^2 &= \mathrm{E}[(X_2 - X_3)^2] = \mathrm{E}[X_2^2] - 2\mathrm{E}[X_2X_3] + \mathrm{E}[X_3^2] \\
&= C_{22} + \mu_2^2 + C_{33} + \mu_3^2 - 2C_{23} - 2\mu_2\mu_3, \qquad \text{and} \\
\delta_{12}^2 &= \mathrm{E}[(X_1 - X_2)^2] = \mathrm{E}[X_1^2] - 2\mathrm{E}[X_1X_2] + \mathrm{E}[X_2^2] \\
&= C_{11} + \mu_1^2 + C_{22} + \mu_2^2 - 2C_{12} - 2\mu_1\mu_2.
\end{aligned}$$

Then we can verify that the numerator of Z_1 above is:

$$\tfrac{1}{2}(\tau_2^2 - \tau_1^2 + \delta_{12}^2).$$

The denominator of Z_1 above is δ_{12}^2. Therefore:

$$Z_1 = \frac{\tau_2^2 - \tau_1^2 + \delta_{12}^2}{2\delta_{12}^2}, \qquad \text{and}$$

$$Z_2 = 1 - Z_1 = \frac{\tau_1^2 - \tau_2^2 + \delta_{12}^2}{2\delta_{12}^2},$$

which is the result obtained in Boor [31].[252] We note that the key distinction is that Boor makes no assumption as to whether

[252]See page 169 of *PCAS* 1992.

the estimators are unbiased.[253] Thus his formulas involve terms like $E[X_1 X_3]$ rather than the covariances such as in Equation 6.7 in the main text.

[253]Also, Boor only displays the result for combining two estimators. The development in this appendix works for any number of estimators; we get $Y+1$ linear equations in $Y+1$ unknowns.

APPENDIX C

CLASSIFICATION DATA

Unit Statistical Plan data for Massachusetts workers compensation insurance was examined.[254] A total of 13 composite policy years[255] of data were available at various reports.[256] For each year the latest available report was used: 80/81 to 88/89 @ 5th; 89/90 @ 4th; 90/91 @ 3rd; 91/92 @ 2nd; 92/93 @ 1st report.

For each classification, payrolls and losses were available. The losses were paid losses plus case reserves. Losses were broken down by injury kind and between medical and indemnity, but these splits were not used in the current analysis.

For example, for Class 2003, Bakeries, the experience in composite policy year 92/93 at first report was \$68,928,691 in payroll and \$1,477,837 in losses. This corresponds to a pure premium (per \$100 of payroll) of 2.1440.

Class 2003 is one of 270 classes in the Manufacturing industry group. For composite policy year 92/93 at first report there was \$3,896,021,286 in payroll and \$67,944,193 in losses for the Manufacturing industry group. This corresponds to a pure premium of 1.7439. Thus the relative pure premium for Class 2003 for 92/93 @ 1st is $2.1440/1.7439 = 1.2294$.

Performing similar calculations, we obtain the following relative pure premiums for two example classes:[257]

---

[254]Experience on all insureds in the state was included, except for large deductible policies. (Large deductibles were only available during the most recent three composite policy years.)

[255]A composite policy year runs from July to June. For example, composite policy year 92/93 includes experience from policies with policy effective dates from July 1, 1992 to June 30, 1993.

[256]First report is evaluated 18 months after policy inception. Subsequent reports are made at 12 month intervals, up to and including fifth report.

[257]The similar calculations were done for each class in the Manufacturing industry group. Similar but totally separate calculations were then done for the Goods & Services industry group.

	Relative Pure Premium	
Composite Policy Year	Class 2003	Class 3145[258]
92/93 @ 1st	1.2294	.7931
91/92 @ 2nd	1.2279	.3741
90/91 @ 3rd	1.5828	.5016
89/90 @ 4th	1.3713	.8561
88/89 @ 5th	1.2380	1.4134
87/88 @ 5th	1.7127	.5199
86/87 @ 5th	1.3507	1.0739
85/86 @ 5th	2.0721	1.1651
84/85 @ 5th	1.4784	.7649
83/84 @ 5th	1.6312	.9236
82/83 @ 5th	1.3711	1.6704
81/82 @ 5th	1.0365	1.5151
80/81 @ 5th	1.7196	.9415

The relative pure premiums show considerable fluctuation because these are medium-sized classes and the losses used are unlimited.[259]

In order to divide the classes into size categories, expected losses were calculated. Expected losses for a class for a composite policy year were obtained by multiplying the reported payrolls by three factors. The first factor was the ratio of the State Average Weekly Wage for Composite Policy Year 1992/1993[260] to that for the particular composite policy year. The second factor was the observed pure premium for the industry group for the particular composite policy year and report. The third and final factor was the ratio of the current rate[261] for the class to the average rate for the industry group.

[258]Class 3145 is Screw Manufacturing.

[259]For classification ratemaking individual claims are usually limited. Currently in Massachusetts workers compensation each claim is capped at $200,000 for classification ratemaking. (These excess losses are loaded back via factors which vary by hazard group and injury kind.)

[260]The most recent year of data used.

[261]Rates effective 5/1/96 were used.

For example, for Class 2003 for Composite Policy Year 91/92 @ 2nd report the payroll was $88,136,418. The State Average Weekly Wage during 92/93 was $580.73, while during 91/92 it was $560.28. Thus the first adjustment factor is $\$580.73 \div \$560.28 = 1.036$. The observed pure premium for the Manufacturing industry group for 91/92 @ 2nd report is 2.361. The current manual rate for Class 2003 is $5.77, while the average manual rate for Manufacturing is $4.008. Thus the third adjustment factor is $\$5.77/\$4.008 = 1.43962$.

Thus the expected losses for Class 2003 for 91/92 @ 2nd are

$$(\$88{,}135{,}418 \div 100)(1.036)(2.361)(1.43962) = \$3{,}103{,}552.$$

A similar calculation of expected losses was made for each of the 13 years. Then the average expected annual losses were calculated for each class.[262] It is these average expected annual losses that were used to divide the classes into size categories for purposes of analysis.

[262]The average only included years in which the class had reported payrolls. Some classes were discontinued or newly erected during these 13 years.

APPENDIX D

SPLIT EXPERIENCE RATING PLAN MATRIX EQUATIONS FOR LEAST SQUARES CREDIBILITY

In this appendix, Equations 10.12 and 10.13 in the main text will be derived for the optimal primary and excess credibilities for a split experience rating plan.

Assume we have two well-defined portions of the total losses, which can be thought of as primary and excess.[263] Assume we have Y years of data being used to predict year $Y + \Delta$.[264] We wish to determine credibilities to apply to the primary and excess data for each year.

Define the following quantities:

$$\begin{aligned}
E_{Pi} &= \text{Expected Primary Losses for Year } i, \\
E_{Xi} &= \text{Expected Excess Losses for Year } i, \\
E_i &= E_{Pi} + E_{Xi} = \text{Expected Losses for Year } i, \\
A_{Pi} &= \text{Actual Primary Losses for Year } i, \\
A_{Xi} &= \text{Actual Excess Losses for Year } i, \\
D_i &= E_{Pi}/E_i = \text{D-ratio for Year } i, \\
P_i &= A_{Pi}/E_i, \\
X_i &= A_{Xi}/E_i, \\
\pi_i &= P_i - D_i = (A_{Pi} - E_{Pi})/E_i \\
&= \text{Primary ``Deviation Ratio'' for Year } i, \text{ and} \\
\xi_i &= X_i - (1 - D_i) = (A_{Xi} - E_{Xi})/E_i \\
&= \text{Excess ``Deviation Ratio'' for Year } i.
\end{aligned}$$

---

[263] For workers compensation insurance, currently the first \$5,000 of each claim is primary, while the remainder up to a claim limit is excess. The claim limit for experience rating varies by state.

[264] Typically $Y = 3$ and $\Delta = 2$ currently. Years 1, 2 and 3 are predicting Year 5.

The quantity of interest in experience rating is how the insured's future losses will compare to the expected losses for the average insured in that class or mixture of classes. That estimate, the experience modification, can be written as:[265]

$$F = 1 + \sum_{i=1}^{Y} \pi_i Z_{Pi} + \sum_{i=1}^{Y} \xi_i Z_{Xi}.$$

This differs somewhat from the usual notation in, for example, Gillam and Snader [19] or Mahler [4], since each individual year of data will be assigned a separate credibility of each type, rather than adding the years of data together and having one overall Z_P and Z_X.

If we use the data from years 1 to Y in order to predict $P_{Y+\Delta} + X_{Y+\Delta}$, the ratio of actual to expected losses for year $Y + \Delta$, then the error is:

$$\begin{aligned} F - (P_{Y+\Delta} + X_{Y+\Delta}) &= F - (\pi_{Y+\Delta} + \xi_{Y+\Delta} + 1) \\ &= \sum_{i=1}^{Y} (\pi_i - \pi_{Y+\Delta}) Z_{Pi} + \sum_{i=0}^{Y} (\xi_i - \xi_{Y+\Delta}) Z_{Xi} \\ &\quad - \pi_{Y+\Delta} \left(1 - \sum_{i=1}^{Y} Z_{Pi}\right) - \xi_{Y+\Delta} \left(1 - \sum_{i=1}^{Y} Z_{Xi}\right). \end{aligned}$$

Define $\pi_0 = \xi_0 = 0$ and $Z_{P0} = 1 - \sum_{i=1}^{Y} Z_{Pi}$ and $Z_{X0} = 1 - \sum_{i=1}^{Y} Z_{Xi}$. Then the error is:

$$\sum_{i=0}^{Y} (\pi_i - \pi_{Y+\Delta}) Z_{Pi} + \sum_{i=0}^{Y} (\xi_i - \xi_{Y+\Delta}) Z_{Xi}.$$

[265] Equation 10.1 in the main text.

The squared error is:

$$\sum_{i=0}^{Y}\sum_{j=0}^{Y}(\pi_i - \pi_{Y+\Delta})(\pi_j - \pi_{Y+\Delta})Z_{Pi}Z_{Pj}$$
$$+ 2\sum_{i=0}^{Y}\sum_{j=0}^{Y}(\pi_i - \pi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})Z_{Pi}Z_{Xj}$$
$$+ \sum_{i=0}^{Y}\sum_{j=0}^{Y}(\xi_i - \xi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})Z_{Xi}Z_{Xj}.$$

Thus, the expected value of the squared error is:

$$\sum_{i=0}^{Y}\sum_{j=0}^{Y}Z_{Pi}Z_{Pj}\mathrm{E}[(\pi_i - \pi_{Y+\Delta})(\pi_j - \pi_{Y+\Delta})]$$
$$+ 2\sum_{i=0}^{Y}\sum_{j=0}^{Y}Z_{Pi}Z_{Xj}\mathrm{E}[(\pi_i - \pi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})]$$
$$+ \sum_{i=0}^{Y}\sum_{j=0}^{Y}Z_{Xi}Z_{Xj}\mathrm{E}[(\xi_i - \xi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})].$$

Define the following quantities in terms of covariances:

$$S_{ij} = \mathrm{Cov}[\pi_i, \pi_j] = S_{ji},$$
$$T_{ij} = \mathrm{Cov}[\xi_i, \xi_j] = T_{ji}, \qquad \text{and}$$
$$U_{ij} = \mathrm{Cov}[\pi_i, \xi_j].$$

Note that since $\mathrm{E}[\pi_i] = 0 = \mathrm{E}[\xi_j]$, $S_{ij} = \mathrm{E}[\pi_i\pi_j]$, $T_{ij} = \mathrm{E}[\xi_i\xi_j]$, and $U_{ij} = \mathrm{E}[\pi_i\xi_j]$. We can rewrite the expression for the expected value of the squared error in terms of S_{ij}, T_{ij} and U_{ij}.

For example, the "cross term" can be written as:

$$\mathrm{E}[(\pi_i - \pi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})]$$
$$= \mathrm{E}[\pi_i\xi_j] - \mathrm{E}[\pi_i\xi_{Y+\Delta}] - \mathrm{E}[\pi_{Y+\Delta}\xi_j] + \mathrm{E}[\pi_{Y+\Delta}\xi_{Y+\Delta}]$$
$$= U_{ij} - U_{i,Y+\Delta} - U_{Y+\Delta,j} + U_{Y+\Delta,Y+\Delta}.$$

Similarly,

$$\mathrm{E}[(\pi_i - \pi_{Y+\Delta})(\pi_j - \pi_{Y+\Delta})] = S_{ij} - S_{i,Y+\Delta} - S_{Y+\Delta,j} + S_{Y+\Delta,Y+\Delta}$$

$$\mathrm{E}[(\xi_i - \xi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})] = T_{ij} - T_{i,Y+\Delta} - T_{Y+\Delta,j} + T_{Y+\Delta,Y+\Delta}.$$

Note that:

$$\begin{aligned}\mathrm{E}[(\pi_0 - \pi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})] &= -\mathrm{E}[\pi_{Y+\Delta}(\xi_j - \xi_{Y+\Delta})] \\ &= -U_{Y+\Delta,j} + U_{Y+\Delta,Y+\Delta},\end{aligned}$$

but $U_{0j} = \mathrm{Cov}[\pi_0, \xi_j] = \mathrm{Cov}[0, \xi_j] = 0$. Thus,

$$\begin{aligned}&\mathrm{E}[(\pi_0 - \pi_{Y+\Delta})(\xi_j - \xi_{Y+\Delta})] \\ &\quad = U_{0,j} - U_{0,Y+\Delta} - U_{Y+\Delta,j} + U_{Y+\Delta,Y+\Delta}.\end{aligned}$$

Thus, the same notation works for index values of zero. Therefore, the expected value of the squared error is the following quadratic function of the primary and excess credibilities:

$$\begin{aligned}&\sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Pi}Z_{Pj}(S_{ij} - S_{i,Y+\Delta} - S_{Y+\Delta,j} + S_{Y+\Delta,Y+\Delta}) \\ &\quad + 2\sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Pi}Z_{Xj}(U_{ij} - U_{i,Y+\Delta} - U_{Y+\Delta,j} + U_{Y+\Delta,Y+\Delta}) \\ &\quad + \sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Xi}Z_{Xj}(T_{ij} - T_{i,Y+\Delta} - T_{Y+\Delta,j} + T_{Y+\Delta,Y+\Delta}).\end{aligned}$$

Some simplification is possible using the facts that:

$$\sum_{i=0}^{Y} Z_{Pi} = 1 = \sum_{i=0}^{Y} Z_{Xi}.$$

Thus, the expected value of the squared error is:

$$\sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Pi}Z_{Pj}S_{ij} - 2\sum_{i=0}^{Y} Z_{Pi}S_{i,Y+\Delta} + S_{Y+\Delta,Y+\Delta}$$
$$+ 2\sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Pi}Z_{Xj}U_{ij} - 2\sum_{i=0}^{Y} Z_{Pi}U_{i,Y+\Delta}$$
$$- 2\sum_{i=0}^{Y} Z_{Xi}U_{Y+\Delta,i} + 2U_{Y+\Delta,Y+\Delta}$$
$$+ \sum_{i=0}^{Y}\sum_{j=0}^{Y} Z_{Xi}Z_{Xj}T_{ij} - 2\sum_{i=0}^{Y} Z_{Xi}T_{i,Y+\Delta} + T_{Y+\Delta,Y+\Delta}.$$

In order to minimize the expected value of the squared error we set each of the $2Y$ partial derivatives with respect to one of the credibilities equal to zero. We get $2Y$ linear equations in $2Y$ unknowns.

Taking the partial derivative of the expected squared error with respect to Z_{Pk} and setting it equal to zero yields:[266]

$$\sum_{i=0}^{Y} Z_{Pi}(S_{ik} - S_{k,Y+\Delta}) + \sum_{i=0}^{Y} Z_{Xi}(U_{ki} - U_{k,Y+\Delta}) = 0.$$

Taking the partial derivative of the expected squared error with respect to Z_{Xk} and setting it equal to zero yields:

$$\sum_{i=0}^{Y} Z_{Pi}(U_{ik} - U_{Y+\Delta,k}) + \sum_{i=0}^{Y} Z_{Xi}(T_{ki} - T_{k,Y+\Delta}) = 0.$$

Again some simplification is possible using the facts that:

$$\sum_{i=0}^{Y} Z_{Pi} = 1 = \sum_{i=0}^{Y} Z_{Xi}.$$

[266]Dividing each term by a factor of two.

The linear equations become:

$$\sum_{i=0}^{Y}(Z_{Pi}S_{ik} + Z_{Xi}U_{ki}) = S_{k,Y+\Delta} + U_{k,Y+\Delta}, \qquad \text{and}$$

$$\sum_{i=0}^{Y}(Z_{Pi}U_{ik} + Z_{Xi}T_{ki}) = U_{Y+\Delta,k} + T_{k,Y+\Delta}.$$

Since $S_{0k} = U_{0k} = T_{0k} = 0 = S_{k0} = U_{k0} = T_{k0}$, the summation on the left hand side can start at $i = 1$ rather than 0. The resulting $2Y$ linear equations are Equations 10.12 and 10.13 in the main text.

An important special case occurs in the absence of shifting parameters over time. Further assume we either use one year of data or combine several years of data together.

Then Equations 10.12 and 10.13 become:

$$Z_P S_{11} + Z_X U_{11} = S_{1,1+\Delta} + U_{1,1+\Delta}, \qquad \text{and}$$

$$Z_P U_{11} + Z_X T_{11} = U_{1+\Delta,1} + T_{1,1+\Delta}.$$

The solutions are:

$$Z_P = \frac{(S_{1,1+\Delta} + U_{1,1+\Delta})T_{11} - (U_{1+\Delta,1} + T_{1,1+\Delta})U_{11}}{S_{11}T_{11} - U_{11}^2}, \qquad \text{and}$$

$$Z_X = \frac{(T_{1,1+\Delta} + U_{1+\Delta,1})S_{11} - (U_{1,1+\Delta} + S_{1,1+\Delta})U_{11}}{S_{11}T_{11} - U_{11}^2}.$$

This matches the result in Mahler [11],[267] with:

S_{11} = Total variance of the primary losses,

T_{11} = Total variance of the excess losses,

[267] See Equations 5.3 and 5.4 in Mahler [11].

$S_{1,1+\Delta}$ = Variance of the hypothetical means of the primary losses,

$T_{1,1+\Delta}$ = Variance of the hypothetical means of the excess losses,

U_{11} = Total covariance of the primary and excess losses, and

$U_{1,1+\Delta} = U_{1+\Delta,1}$ = Covariance of the hypothetical means of the primary and excess losses.

In the notation in Mahler [11]:

a = Total variance of the primary losses,

b = Total variance of the excess losses,

c = Variance of the hypothetical means of the primary losses,

d = Variance of the hypothetical means of the excess losses,

r = Total covariance of the primary and excess losses, and

s = Covariance of hypothetical means of the primary and excess losses.

And in the absence of shifting risk parameters over time the optimum Z_P and Z_X are:

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

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

APPENDIX E

USE OF COUNTRYWIDE CLASSIFICATION DATA, MATRIX EQUATIONS FOR LEAST SQUARES CREDIBILITY

This appendix will discuss Equations 8.1 in the main text for the optimal least squares credibility when combining classification data from more than one state.

Assume we have a series of observations of X_i, for example, the class relativities in Massachusetts for each of several years, $i = 1$ to Y. Assume we also have a related series of observations of A_i, for example, the relativities for the same class calculated from data from some other states.[268] Finally, assume we wish to predict $X_{Y+\Delta}$, the class relativity in Massachusetts in year $Y + \Delta$ in the example in the main text, using a weighted average of the X_i and A_i.

More specifically the predictor $F = \sum_{i=1}^{Y} Z_i X_i + \sum_{i=1}^{Y} W_i A_i$ and $\sum_{i=1}^{Y} Z_i + \sum_{i=1}^{Y} W_i = 1$.

Note that here the weights sum to 100%; there is no weight being given to the grand mean. Note that since we are predicting $X_{Y+\Delta}$, X and A will not enter into the matrix equations in a symmetric fashion.[269]

Let the covariances be:

$$S_{ij} = \mathrm{Cov}[X_i, X_j] = S_{ji},$$
$$T_{ij} = \mathrm{Cov}[A_i, A_j] = T_{ji}, \qquad \text{and}$$
$$U_{ij} = \mathrm{Cov}[X_i, A_j].$$

[268] It is not necessary that A_i be available for exactly the same years as X_i, $i = 1$ to Y, but the presentation is easier to follow if we assume that this is the case. (Years with no available data can be treated by giving them a weight of zero.)

[269] In contrast, the primary and excess losses did enter into the equations in Appendix D in a mathematically symmetric manner.

As in Appendices B and D, assume each X_i or A_i is an unbiased estimator of the quantity of interest, $X_{Y+\Delta}$. Then the expected value of the squared error is:

$$\begin{aligned}
V(Z,W) &= \mathrm{E}[(F - X_{Y+\Delta})^2] \\
&= \mathrm{E}\left[\left\{\sum_{i=1}^{Y} Z_i(X_i - X_{Y+\Delta}) + \sum_{i=1}^{Y} W_i(A_i - X_{Y+\Delta})\right\}^2\right] \\
&= \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j (S_{ij} - S_{i,Y+\Delta} - S_{Y+\Delta,j} + S_{Y+\Delta,Y+\Delta}) \\
&\quad + 2\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i W_j (U_{ij} - S_{i,Y+\Delta} - U_{Y+\Delta,j} + S_{Y+\Delta,Y+\Delta}) \\
&\quad + \sum_{i=1}^{Y}\sum_{j=1}^{Y} W_i W_j (T_{ij} - U_{Y+\Delta,i} - U_{Y+\Delta,j} + S_{Y+\Delta,Y+\Delta}).
\end{aligned}$$

Some simplification of the expression for $V(Z,W)$ is possible. Since $S_{i,Y+\Delta} = S_{Y+\Delta,i}$

$$\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{i,Y+\Delta} = \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{Y+\Delta,j}.$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{i,Y+\Delta} + \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{Y+\Delta,j} + 2\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i W_j S_{i,Y+\Delta} \\
&\quad = 2\sum_{i=1}^{Y}\sum_{j=1}^{Y} (Z_i Z_j + Z_i W_j) S_{i,Y+\Delta} \\
&\quad = 2\left(\sum_{i=1}^{Y} Z_i S_{i,Y+\Delta}\right)\left(\sum_{j=1}^{Y}(Z_j + W_j)\right) = 2\sum_{i=1}^{Y} Z_i S_{i,Y+\Delta},
\end{aligned}$$

since $\sum_{j=1}^{Y}(Z_j + W_j) = 1$.

Similarly,

$$2\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i W_j U_{Y+\Delta,j} + \sum_{i=1}^{Y}\sum_{j=1}^{Y} W_i W_j U_{Y+\Delta,i} + \sum_{i=1}^{Y}\sum_{j=1}^{Y} W_i W_j U_{Y+\Delta,j}$$

$$= 2\left(\sum_{i=1}^{Y}(Z_i + W_i)\right)\left(\sum_{j=1}^{Y} W_j U_{Y+\Delta,j}\right) = 2\sum_{i=1}^{Y} W_i U_{Y+\Delta,i}.$$

Also we have:

$$\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{Y+\Delta,Y+\Delta} + 2\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i W_j S_{Y+\Delta,Y+\Delta}$$

$$+ \sum_{i=1}^{Y}\sum_{j=1}^{Y} W_i W_j S_{Y+\Delta,Y+\Delta}$$

$$= S_{Y+\Delta,Y+\Delta}\left(\sum_{i=1}^{Y}(Z_i + W_i)\right)\left(\sum_{j=1}^{Y}(Z_j + W_j)\right) = S_{Y+\Delta,Y+\Delta}.$$

Thus, $V(Z,W)$ simplifies to:

$$V(Z,W) = \sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i Z_j S_{ij} + 2\sum_{i=1}^{Y}\sum_{j=1}^{Y} Z_i W_j U_{ij}$$

$$+ \sum_{i=1}^{Y}\sum_{j=1}^{Y} W_i W_j T_{ij} - 2\sum_{i=1}^{Y} Z_i S_{i,Y+\Delta}$$

$$- 2\sum_{i=1}^{Y} W_i U_{Y+\Delta,i} + S_{Y+\Delta,Y+\Delta}.$$

The constraint equation $\sum_{i=1}^{Y} Z_i + \sum_{i=1}^{Y} W_i - 1 = 0$ is incorporated via the Lagrange multiplier λ. We minimize $V(Z,W)-$

$\lambda(\sum_{i=1}^{Y} Z_i + \sum_{i=1}^{Y} W_i - 1)$, by taking each of $2Y$ partial derivatives with respect to the Zs and Ws and setting them equal to zero. Setting the partial derivative with respect to Z_k equal to zero:

$$2\sum_{j=1}^{Y} Z_j S_{kj} + 2\sum_{j=1}^{Y} W_j U_{kj} - 2S_{k,Y+\Delta} - \lambda = 0.$$

This equation can be rewritten as:

$$\sum_{j=1}^{Y} Z_j S_{kj} + \sum_{j=1}^{Y} W_j U_{kj} = \lambda/2 + S_{k,Y+\Delta}.$$

Similarly, by setting the partial derivative with respect to W_k equal to zero we get:

$$\sum_{j=1}^{Y} Z_j U_{jk} + \sum_{j=1}^{Y} W_j T_{kj} = \lambda/2 + U_{Y+\Delta,k}.$$

The above $2Y$ linear equations (one from the partial derivative of each Z_k and each W_k) plus the constraint equation are Equations 8.1 in the main text. Note the similarities to the Equations 2.4, 6.7, 10.12 and 10.13 in the main text. Each set of equations applies to a somewhat different situation. However, each such set of linear equations depends on the covariance structure and can be solved for the credibilities using matrix methods.

In a situation where there were different years of Massachusetts and countrywide data, Equations 8.1 would be somewhat different in form. For example, assume Massachusetts data for years 1, 2, 3 and 4 plus countrywide data for years 2 and 3 were being used to predict Massachusetts relativities for year 8. The Equations 8.1 would become seven linear equations in seven unknowns:

$$Z_1S_{11} + Z_2S_{12} + Z_3S_{13} + Z_4S_{14} + W_2U_{12} + W_3U_{13} = \lambda/2 + S_{18},$$
$$Z_1S_{21} + Z_2S_{22} + Z_3S_{23} + Z_4S_{24} + W_2U_{22} + W_3U_{23} = \lambda/2 + S_{28},$$
$$Z_1S_{31} + Z_2S_{32} + Z_3S_{33} + Z_4S_{34} + W_2U_{32} + W_3U_{33} = \lambda/2 + S_{38},$$
$$Z_1S_{41} + Z_2S_{42} + Z_3S_{43} + Z_4S_{44} + W_2U_{42} + W_3U_{43} = \lambda/2 + S_{48},$$
$$Z_1U_{12} + Z_2U_{22} + Z_3U_{32} + Z_4U_{42} + W_2T_{22} + W_3T_{32} = \lambda/2 + U_{82},$$
$$Z_1U_{13} + Z_2U_{23} + Z_3U_{33} + Z_4U_{43} + W_2T_{23} + W_3T_{33} = \lambda/2 + U_{83},$$
$$\text{and} \qquad Z_1 + Z_2 + Z_3 + Z_4 + W_2 + W_3 = 1.$$

In this case, the equations each have four terms involving the four weights to each of the years of Massachusetts data, but only two terms involving the two weights to each of the years of countrywide data. There are $4 + 2 + 1 = 7$ unknowns, including the Lagrange multiplier.

APPENDIX F

ESTIMATING PARAMETERS OF BETWEEN STATE COVARIANCES

In order to calculate credibilities when using data from one or more outside states to calculate classification relativities, it is necessary to estimate the variance-covariance structure. This appendix will present an example of how to estimate the parameters of the between state covariances.

Assume as in Section 8 we are estimating Massachusetts class relativities and will use New York data in addition to that from Massachusetts.

Then there are three types of variance-covariance matrices. The first type involves covariances between data from Massachusetts. The second type involves covariances between data from New York. The third type of covariance is that involving data from Massachusetts versus New York. It is expected that for a given volume of data, the correlation of relativities between states is less than the correlation of relativities within states. This is what is observed.

For three years of data combined for each state, adjusted as it would be for classification ratemaking,[270] the relative pure premiums were calculated for the classes in the Manufacturing industry group. Then the correlations between the New York and Massachusetts relative pure premiums were calculated for various sizes of class.[271] The results are in the table below.

How does this compare to the correlation within a single state that we would expect if we could run the risk process twice

[270] Claims sizes are limited. Losses would be adjusted for law changes and loss development. See Kallop [12] and Feldblum [13]. In this case, these adjustments were performed by whichever rating bureau has responsibility for that state.

[271] The size of class was taken as the square root of the product of the payroll (summed over three years) for each state. In other words, the geometric average of the payroll for the two states was used for each class.

Manufacturing Industry Group				
Three Years of Payroll ($ million)			Correlations NY vs. MA	
Minimum	Maximum	Number of Classes	Observed	Capped[272]
1	3	12	.047	.041
3	10	35	.348	.382
10	30	44	.106	.127
30	100	45	.596	.592
100	300	29	.623	.623
300	1,000	5	.818	.823

and create two parallel universes?[273] In that case the portion of the covariance related to the expected value of the process variance, the term involving e^2 or K, would vanish.[274]

Ignoring shifting risk parameters,[275] the covariances between single years of data are given by Equation 4.13:

$$\mathrm{Cov}[X_i, X_j] = r^2\{1 + I/E + \delta_{ij}(K/E + J)\}, \qquad E \geq \Omega.$$

Then for the sum of three years of data, X_1, X_2 and X_3:

$$\mathrm{Cov}[X_1 + X_2 + X_3, X_1 + X_2 + X_3] = r^2\{9 + 9I/E + 3K/E + 3J\}, \qquad E \geq \Omega.$$

If we have data from two parallel versions of Massachusetts, we set $K = 0$, and J' to represent the possibly different J param-

[272]The relativity for each class was capped between 5 and 1/5, in order to limit the impact of any one class on the computed correlation.

[273]In the dice examples in Sections 3 and 4, one just rerolls the dice keeping everything else constant.

[274]For example, if X_1 and X_2 are each independent results of rolling 10 six-sided dice, then their covariance is zero, while the process variance of X_1 or X_2 is positive. The usual Bühlmann covariance structure is $\mathrm{Cov}[X_i, X_j] = \tau^2 + \delta_{ij}\eta^2$. For $i \neq j$, $\mathrm{Cov}[X_i, X_j] = \tau^2$; the term involving the expected value of the process variance, η^2, vanishes.

[275]Setting $\rho = 1$ and $\gamma = 1$.

eter when taking the covariance between two parallel versions of Massachusetts.

Then the correlation for three years of data from each of two parallel versions of Massachusetts is:

$$\frac{9+9I/E+3J'}{9+9I/E+3K/E+3J} = \frac{(3+J')E+3I}{(3+J)E+3I+K}, \qquad E \geq \Omega.$$

If we include shifting risk parameters, we get a slightly different expression for the correlations. The covariances are given by Equation 5.8:

$$\mathrm{Cov}[X_i, X_j] = r^2\{\rho^{|i-j|} + \gamma^{|i-j|} I/E + \delta_{ij}(K/E + J)\}, \qquad E \geq \Omega.$$

For the sum of three years of data, the terms involving ρ sum to $3+4\rho+2\rho^2$. Similarly, the terms involving γ sum to $(3+4\gamma+2\gamma^2)I/E$. Using the values from Section 7.7, for $\rho = .98$, $3+4\rho+2\rho^2 = 8.84$, while for $\gamma = .85$, $3+4\gamma+2\gamma^2 = 7.85$.

Thus, the correlations equal:

$$\frac{8.84+7.85I/E+3J'}{8.84+7.85I/E+3K/E+3J} = \frac{(2.95+J')E+2.62I}{(2.95+J)E+2.62I+K}, \qquad E \geq \Omega.$$

Similarly for $E \leq \Omega$ starting with Equation 5.9, we obtain a correlation of:

$$\frac{8.84+7.85I/\Omega+3J'}{8.84+7.85I/\Omega+3K/E+3J} = \frac{(2.95+J')E+2.62IE/\Omega}{(2.95+J)E+2.62IE/\Omega+K}, \qquad E \leq \Omega.$$

Depending on whether or not the effects that are responsible for parameter uncertainty are reproduced,[276] the term involving u^2 or J may or may not vanish. In the case of MA vs. NY, the two states would be affected by some of the same macroe-

[276] In the dice examples in Section 3, do we maintain the same coin flip or is the coin flipped again?

conomic and other forces that produce parameter uncertainty. Thus, for covariances between MA and NY, we would expect that a portion of the term involving J would remain. Since the two parallel versions of Massachusetts will be used to compare to the interstate situation, we will also assume that in that case a portion of the term involving J will remain. For illustrative purposes take $J' = .05$ for the covariances between two parallel versions of Massachusetts, one-half of the assumed value for the intrastate covariances.[277]

Using the estimated parameters from Section 7.7, with $J' = .05$, we get the correlations shown below as "Expected Intrastate."[278] These have been compared to the observed correlations between New York and Massachusetts.

Payroll 3 Years	Correlations			Ratio to Expected Intrastate	
($ million)	Expected Intrastate	NY vs. MA	NY vs. MA, Capped[279]	NY vs. MA	NY vs. MA, Capped
2	.215	.047	.041	.22	.19
6.5	.458	.348	.382	.76	.83
20	.600	.106	.127	.18	.21
65	.782	.596	.592	.76	.76
200	.900	.623	.623	.69	.69
650	.955	.818	.823	.86	.86

Similar comparisons were done for other large states and for the Goods and Services industry group. Here are the ratios of the observed interstate correlations to the expected intrastate corre-

[277]The credibilities are relatively insensitive to this choice.

[278]In order to translate payrolls into expected losses the payrolls were multiplied by the observed pure premium for the Manufacturing industry group of about $2.50 per $100 of payroll. Thus $1 million of payroll for 3 years corresponds to $8,333 of annual expected losses.

[279]The relativity for each class was capped between 5 and 1/5, in order to limit the impact of any one class on the computed correlation.

lations:

RATIO OF INTERSTATE TO EXPECTED INTRASTATE CORRELATIONS[280] BY THREE YEARS OF PAYROLL ($ MILLION)

	Manufacturing				Goods and Services	
State	3 to 10	10 to 30	30 to 100	100 to 300	10 to 100	100 to 1,000
Connecticut	−.07	.58	.39	.74	.90	.86
Florida	.08	.68	.81	.66	.23	.85
Georgia	.11	.34	.23	.40	.61	.84
Illinois	.50	.10	.60	.81	.58	.92
Michigan	—	.09	.59	.56	1.09	.82
Missouri	.12	.31	.45	.67	.94	.88
New Jersey	.72	.03	.53	.63	—	—
New York	.76	.18	.76	.69	1.09	.96
Oregon	.78	.88	.73	—	.95	.87
Wisconsin	.18	−.04	.69	.66	.86	.89
Average	**.35**	**.32**	**.58**	**.65**	**.81**	**.87**

Generally, the between state correlations are lower than the within state correlations, for a given volume of data. In this case, the between state correlations are perhaps 55% of the within state correlations for Manufacturing[281] and perhaps 85% for Goods and Services. A ratio of 70% would result if the r^2 factor multiplying the interstate covariances were 70% of the r^2 factor multiplying the intrastate covariances.[282]

This ratio of 70% will be used for illustrative purposes in Section 8. As is shown in Section 8.7, the credibilities are relatively insensitive to this choice for values within this general range.

[280]Only results for categories with 15 or more classes are displayed.

[281]The correlations for the smaller size categories for Manufacturing were affected by two classes whose observed Massachusetts relativities were vastly different than those observed in most other states.

[282]Recall that in the intrastate situation, the r^2 factor did not affect the calculated credibilities. In the interstate situation with two (or more) values for r^2, the relative size of the r^2 values will affect the credibilities.

For the interstate covariances the K parameter will be zero.[283] Also, the interstate covariances will use $J = .05$, one-half of the assumed value for the intrastate covariances.

[283]As discussed above, the portion of the covariance related to the expected value of the process variance would vanish when taking covariances between data from different states.