A MARKOV CHAIN MODEL OF SHIFTING RISK PARAMETERS

HOWARD C. MAHLER

Abstract

In this paper, a practical and flexible model involving simple Markov chains is developed that incorporates the phenomenon of shifting risk parameters. One can view this model as a generalization of the gamma-Poisson, beta-binomial, and similar models.

The model is applied to a variety of examples in order to illustrate its possible uses:

- *dice,*
- *a mixture of four Poissons,*
- *California driving data (modeled by a gamma-Poisson), and*
- *baseball data (modeled by a mixture of binomials).*

The model is sufficiently flexible to be applied to other situations.

In each case, the Markov chain model is used to explore the effects of shifting risk parameters over time. A formula is developed and used to calculate covariances. Based on the Markov chain model, when shifting risk parameters over time are significant, the logs of the covariances between years of data are expected to decline linearly as the separation between years increases.

A formula is developed and used to calculate credibilities from the variances and covariances. When shifting risk parameters are significant, older years receive less credibility and as more and more years of data are added, the sum of the credibilities goes to a limit less than one.

1. INTRODUCTION

The phenomenon of shifting risk parameters over time has been explored in past *Proceedings* papers by Venezian [14, 15] and Mahler [7, 9, 10]. It has been shown that this phenomenon can significantly impact the relative value of data for use in predicting the future. Specifically, it can significantly affect the credibility assigned to data to be used for experience rating.

In this paper, a practical model involving simple Markov chains is developed that incorporates the phenomenon of shifting risk parameters. One can view this model as a generalization of the gamma-Poisson, beta-binomial, and similar models. The model is applied to a variety of examples in order to illustrate its possible uses.

Bühlmann credibility[1] is discussed, for example, in Mayerson [11], Hewitt [4, 5], Philbrick [12], and Herzog [3]. Bühlmann derived, under certain assumptions, the linear least squares estimator; a similar derivation is performed for the more general situation in this paper in Appendix C. In order to apply Bühlmann credibility, the Bühlmann credibility parameter is calculated as

$$K = \frac{\text{expected value of process variance}}{\text{variance of hypothetical means}},$$

where the expected value of process variance and the variance of hypothetical means are each calculated for a single observation of the risk process. Then for N observations, the Bühlmann credibility is $Z = N/(N + K)$.

[1] Bühlmann credibility is also referred to as Bayesian credibility or least squares credibility.

2. SIMPLE EXAMPLE INVOLVING DICE

2.1. Bühlmann Credibility

Assume Joe selects N dice of the same type and rolls them. Assume Joe selected either four-sided, six-sided, or eight-sided dice, 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 the type of dice Joe selected. Joe will roll the same dice again.

You can use Bühlmann credibility to predict the sum of that next roll. The expected value of the process variance[2] (for one die) is 3.08. The variance of the hypothetical means[3] is .500. Therefore K = expected value of process variance/variance of hypothetical means = 6.16. The credibility assigned to the observation is $Z = N/(N + K) = N/(N + 6.16)$. Thus for example, if Joe rolls 3 dice which sum to 14, then $Z = 33\%$ and the credibility estimate of the sum of the next roll of three dice[4] is $(14)(33\%) + (10.5)(67\%) = 11.7$.

The credibility can also be written as:

$$Z = \frac{.5N^2}{.5N^2 + 3.08N} \tag{2.1}$$

$$Z = \frac{\text{variance of hypothetical means for the sum of } N \text{ dice}}{\begin{array}{c}(\text{variance of hypothetical means for the sum of } N \text{ dice} \\ + \text{ expected value of the process variance for} \\ \text{the sum of } N \text{ dice})\end{array}}$$

$$= \frac{\text{variance of hypothetical means for the sum of } N \text{ dice}}{\text{total variance for the sum of } N \text{ dice}},$$

[2]The process variances for 4, 6, and 8-sided dice are, respectively, 1.25, 2.92, and 5.25.

[3]The means for 4, 6, and 8-sided dice are, respectively, 2.5, 3.5, and 4.5.

[4]The complement of credibility of $1 - .33 = .67$ is assigned to the overall a priori mean of 3.5 per die times 3 dice.

where we have used the fact that the total variance is the sum of the expected value of process variance and variance of hypothetical means. Also note that the variance of the hypothetical means for the sum of N identical dice is simply N^2 times the variance of hypothetical means for one die since each of the means is multiplied by N. This is a special case of the general result for any random variable Y, $\mathrm{Var}[NY] = N^2\mathrm{Var}[Y]$. In this case, Y is the hypothetical mean for a single roll of each type of die. In contrast, the expected value of the process variance for the sum of N identical dice is just N times the expected value of process variance for a single die. This is a special case of the general result, $\mathrm{Var}[X_1 + X_2 + \cdots + X_N] = N\,\mathrm{Var}[X]$ for X_i independent and identically distributed.

This simple example has so far been a review of basic[5] Bühlmann credibility. Next we will complicate the risk process by adding shifting risk parameters over time.

2.2. *Dice Example, Shifting Parameters Over Time*

Let's introduce a somewhat different risk process. 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.[6]
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 actions will eventually scramble the

[5]This material is currently included on the Part 4B Exam syllabus.

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

information one could obtain in her absence by summing the results of many trials. However, if one uses the most recent trial's result, it is unlikely that Beth will have affected the situation. Thus more recent trials provide more valuable information for predicting the future. Therefore, more recent trials of data should be given more credibility than less recent trials of data.

This is generally the case when one has shifting risk parameters over time. We will determine how to calculate the credibilities for this example as well as in more general situations applicable to insurance.

2.3. Markov Chains

Beth's risk process is a simple example of a Markov chain.[7] 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. Beth's Markov chain was completely described by the "transition probabilities" between the states.

Label the states 1, 2, and 3 corresponding to 4-sided, 6-sided, and 8-sided die. Then let P_{21} = the probability of being in state 1 given that the previous trial was in state 2. This is the probability of Beth replacing a 6-sided die with a 4-sided die, or 10%. Thus $P_{21} = 10\%$. Similarly, P_{23} = the chance of Beth replacing a 6-sided die with an 8-sided die = 15%. P_{22} = the chance of Beth leaving a 6-sided die alone = 75%. Note that $P_{21} + P_{22} + P_{23} = 10\% + 75\% + 15\% = 100\%$. The probabilities of all the things Beth can do to a 6-sided die add up to 100%.

Generally, the transition probabilities for a Markov chain are arranged in a matrix $\boldsymbol{P}$. For Beth's "risk process," the matrix of

[7]Feller [2], Resnick [13].

transition probabilities is:

$$\begin{pmatrix} .80 & .20 & 0 \\ .10 & .75 & .15 \\ 0 & .30 & .70 \end{pmatrix}$$

where we have previously discussed the second row. Note we have assumed no chance of Beth's "risk process" replacing an eight-sided die with a four-sided die so that $P_{31} = 0$.

We note that each of the rows of the matrix sums to unity. As discussed previously, this is a general property of transition matrices. In addition, this matrix was chosen to have a special property.

We have assumed Joe's probability of initially picking each of three types of dice is 25%, 50%, and 25%. Thus the initial probability vector is $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. The Markov chains we will be dealing with will, in the limit, go to a so-called stationary distribution. For the chosen transition probabilities, $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ is that stationary distribution.[8] We expect this initial distribution to, on average, continue over time.

We can see this by thinking of the expected number of each type of die Beth adds or subtracts.

1. There is a $\frac{1}{4}$ chance that Joe picks a 4-sided die. There is a $\frac{1}{4} \times 20\% = 5\%$ chance that Beth adds a 6-sided die and subtracts a 4-sided die.

2. There is a $\frac{1}{2}$ chance that Joe picks a 6-sided die. There is a $10\%/2 = 5\%$ chance that Beth adds a 4-sided die and subtracts a 6-sided die and a $15\%/2 = 7\frac{1}{2}\%$ chance Beth adds an 8-sided die and subtracts a 6-sided die.

[8]The transition probabilities were chosen so that the initial state would be a stationary distribution. See Appendix D for a discussion of how such a transition matrix can be constructed from a given stationary distribution.

3. There is a $\frac{1}{4}$ chance that Joe picks an 8-sided die. There is a $30\%/4 = 7\frac{1}{2}\%$ chance that Beth adds a 6-sided die and subtracts an 8-sided die.

In summary, the change in the probability of a 4-sided die is expected to be $5\% - 5\% = 0$. The change in the probability of a 6-sided die is expected to be $5\% - 5\% + 7\frac{1}{2}\% - 7\frac{1}{2}\% = 0$. The change in the probability of an 8-sided die is expected to be $7\frac{1}{2}\% - 7\frac{1}{2}\% = 0$. Thus we indeed have a stable situation on an expected basis.

Let $\boldsymbol{\alpha}$ be the vector of a priori probabilities. All we have done is verify the matrix equation that $\boldsymbol{\alpha P} = \boldsymbol{\alpha}$. This is the definition of a stationary distribution.

In general, if $\boldsymbol{\beta}$ is a vector of the initial probabilities of being in each state, then the matrix product of $\boldsymbol{\beta}$ and the transition matrix $\boldsymbol{P}$, $\boldsymbol{\beta P}$, is the vector of probabilities after one trial.

One last important point is how we would calculate, for example, the probability, if Joe initially picked a 4-sided die, of Beth after 2 trials replacing this 4-sided die with an 8-sided die. This would be the product of the probabilities of replacing a 4-sided die with a 6-sided die after the first trial and then replacing the 6-sided die with an 8-sided die after the second trial. In this case, that probability is $P_{12}P_{23} = (.20)(.15) = 3\%$.

If Joe initially picked a 4-sided die, what is the probability of having a 4-sided die after two trials? Either Beth did not replace the die at both trials or she replaced the 4-sided die at trial one with a 6-sided die and then at trial two replaced the 6-sided die with a 4-sided die. These probabilities are $P_{11}P_{11} + P_{12}P_{21} = (.80)(.80) + (.20)(.10) = .66$.

Similarly, if Joe initially picked a 4-sided die, the chance of having a 6-sided die after two trials is $P_{11}P_{12} + P_{12}P_{22} = (.80)(.20) + (.20)(.75) = .31$. Note that given that Joe picked a 4-sided die, the probabilities of the three possible situations after two trials add up to unity: $.66 + .31 + .03 = 1.00$. One can also verify that these are the entries of the first row of $\boldsymbol{P}^2$.

In general, one could easily compute such probabilities by taking matrix products of $\boldsymbol{P}$. $\boldsymbol{P}^2 = \boldsymbol{P} \times \boldsymbol{P}$ contains the transition probabilities for two trials. $\boldsymbol{P}^3 = \boldsymbol{P} \times \boldsymbol{P} \times \boldsymbol{P}$ contains the transition probabilities for three trials, etc. Thus if β is a vector of the initial probabilities of being in each state, then $\beta \boldsymbol{P}^N$ is the vector of probabilities after N trials.

2.4. *Eigenvectors and Eigenvalues*

In order to more easily compute credibilities as well as gain a better understanding of the behavior in specific examples, eigenvectors and eigenvalues are useful. An eigenvector $\boldsymbol{v}_i$ and related eigenvalue λ_i of a matrix $\boldsymbol{M}$ are such that

$$\boldsymbol{M}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i.$$

Appendix B contains a brief discussion of eigenvectors and eigenvalues. If the transpose of $\boldsymbol{P}$ has an eigenvector $\boldsymbol{v}_i$, then

$$\boldsymbol{P}^T \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad \text{or} \quad \boldsymbol{v}_i \boldsymbol{P} = \lambda_i \boldsymbol{v}_i.$$

Recall Beth's transition matrix:

$$\begin{pmatrix} .80 & .20 & 0 \\ .10 & .75 & .15 \\ 0 & .30 & .70 \end{pmatrix}.$$

Its transpose has eigenvalues[9] of 1, .769, and .481. It has corresponding eigenvectors[10] of (1,2,1), (1,−.314,−.686), and (1,−3.186,2.186). The eigenvalue 1 corresponds to the stationary distribution; its corresponding eigenvector (1,2,1) is proportional to the stationary distribution.[11] By the definition of an eigenvector with an eigenvalue of 1: $\boldsymbol{v}_1 \boldsymbol{P} = 1\boldsymbol{v}_1 = \boldsymbol{v}_1$.

[9] We have chosen to list the eigenvalues starting with unity for the sake of convenience.
[10] Note, the eigenvectors may each be multiplied by any constant and remain eigenvectors.
[11] Recall that in this example, the stationary distribution was $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

Let λ be the vector of eigenvalues of P^T. Let V be the matrix of corresponding eigenvectors, with each row being an eigenvector. In this case $\lambda = (1, .769, .481)$, while

$$V = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -.314 & -.686 \\ 1 & -3.186 & 2.186 \end{pmatrix} \quad \text{and}$$

$$V^{-1} = \begin{pmatrix} .250 & .658 & .092 \\ .250 & -.103 & -.147 \\ .250 & -.451 & .201 \end{pmatrix}.$$

The elements of the first column of V^{-1} are all equal, and are the proportionality constant to convert the first eigenvector (the elements of the first row of V) into the stationary distribution α. In our example, the first column of V^{-1} is $(.25, .25, .25)$ where each element is the inverse of the sum of the first eigenvector[12] $(1,2,1)$. The sum of any eigenvector but the first is zero.

We have the following result of multiplying matrices[13]:

$$VPV^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & .769 & 0 \\ 0 & 0 & .481 \end{pmatrix}.$$

So the matrix of eigenvectors of P^T can be used to convert P to a diagonal matrix whose elements are the eigenvalues of P^T. Let this diagonal matrix be Λ.

2.5. Limits

$$VPV^{-1} = \Lambda. \tag{2.2}$$

---

[12] We could have just as easily chosen the first eigenvector as $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, in which case since it sums to unity, it is the stationary distribution.

[13] This follows from the matrix equation $VP = \Lambda V$, which taking each row in turn says $v_i P = \lambda_i v_i$.

In general, for any matrix $\boldsymbol{P}$ and any invertible matrix $\boldsymbol{V}$,

$$\boldsymbol{P}^2 = \boldsymbol{V}^{-1}(\boldsymbol{VPV}^{-1})(\boldsymbol{VPV}^{-1})\boldsymbol{V} = \boldsymbol{V}^{-1}(\boldsymbol{VPV}^{-1})^2\boldsymbol{V}.$$

This result extends similarly to higher powers:

$$\boldsymbol{P}^g = \boldsymbol{V}^{-1}(\boldsymbol{VPV}^{-1})^g\boldsymbol{V}.$$

Substituting the particular expression for Λ from Equation 2.2, one obtains

$$\boldsymbol{P}^g = \boldsymbol{V}^{-1}\Lambda^g\boldsymbol{V}. \tag{2.3}$$

So taking powers of the transition matrix corresponds to taking powers of the diagonal matrix Λ. We use the matrix of eigenvectors $\boldsymbol{V}$ to translate back and forth. Λ^g is diagonal with elements λ_i^g. As $g \to \infty$, $\lambda_i^g \to 0$ for $|\lambda_i| < 1$. Since $|\lambda_i| < 1$ for $i > 1$, Λ^g approaches a matrix, all but one of whose elements is zero, and element $(\Lambda^g)_{11} = 1^g = 1$.

As discussed in Appendix A, $\boldsymbol{P}^g \to \boldsymbol{A}$, as $g \to \infty$, where $\boldsymbol{A}$ is a matrix each of whose rows is proportional to the first eigenvector; each row of $\boldsymbol{A}$ is the stationary distribution.

For any initial distribution $\boldsymbol{\beta}$, $\lim_{g\to\infty}\boldsymbol{\beta P}^g = \boldsymbol{\beta A} = \boldsymbol{\alpha}$ since the sum of the elements of $\boldsymbol{\beta}$ is unity and since the rows of $\boldsymbol{A}$ are each the stationary distribution $\boldsymbol{\alpha}$. Thus for any initial distribution, after enough time passes, we approach the stationary distribution:

$$\boldsymbol{\beta P}^g \to \boldsymbol{\alpha}, \qquad \boldsymbol{P}^g \to \boldsymbol{A}, \qquad \text{and} \qquad \Lambda^g \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.4}$$

The speed with which this convergence takes place is dependent on $|\lambda_i|$ for $i \neq 1$. The smaller $|\lambda_i|$ for $i \neq 1$, the larger the effect of shifting parameters over time. In the current example, $\lambda_2 = .769$ and $\lambda_3 = .481$, so convergence takes a while.

For $N = 5$,

$$\Lambda^5 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & .269 & 0 \\ 0 & 0 & .026 \end{pmatrix}, \qquad \text{and}$$

$$P^5 = V^{-1}\Lambda^5 V = \begin{pmatrix} .429 & .437 & .134 \\ .219 & .521 & .261 \\ .134 & .521 & .344 \end{pmatrix}.$$

For $N = 20$,

$$\Lambda^{20} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & .005 & 0 \\ 0 & 0 & 4 \times 10^{-7} \end{pmatrix}, \qquad \text{and}$$

$$P^{20} = V^{-1}\Lambda^{20} V = \begin{pmatrix} .253 & .499 & .248 \\ .249 & .500 & .250 \\ .248 & .501 & .252 \end{pmatrix}.$$

Thus after five trials we expect to have retained a small amount of information about Joe's initial pick. For example, if Joe initially picked a 4-sided die, after five trials there is .429 chance of a 4-sided die, a .437 chance of a 6-sided die, and a .134 chance of an 8-sided die. After 20 trials, for all practical purposes the probabilities are independent of Joe's initial pick. Beth's process has scrambled things sufficiently in order to remove any trace of the initial pick.

We conclude that the outcome of the first trial would provide no useful information for the prediction of the 21st trial. On the other hand, the outcome of the 16th trial would provide some small amount of useful information for the prediction of the 21st trial, being only five trials apart. Thus, we would expect to give the 16th trial some small credibility and the first trial virtually zero credibility when predicting the outcome of the 21st trial.

2.6. *Covariances*

In insurance applications, a year of data takes the place of a trial in the example involving dice. In order to calculate credibilities, we need to calculate the variances as well as the covariances between different years of data. As developed in Appendix E,

let ζ be the vector such that

$$\zeta_i = ((\boldsymbol{\mu} \times \boldsymbol{\alpha})^T \boldsymbol{V}^{-1})_i (\boldsymbol{V}\boldsymbol{\mu})_i. \tag{2.5}$$

Then, for $g > 0$, the covariance of two years of data separated by g years is given by

$$\mathrm{Cov}[X,U] = \sum_{i>1} \zeta_i \lambda_i^g. \tag{2.6}$$

Note that λ_i and ζ_i which determine the behavior of the covariances are each directly and easily calculable[14] from the assumed transition matrix and the means of the states. The steps developed in Appendix E are:

1. Assume[15] a transition matrix $\boldsymbol{P}$ corresponding to the assumed states with means given by the vector $\boldsymbol{\mu}$.
2. Calculate the eigenvalues and eigenvectors of the transpose of the transition matrix $\boldsymbol{P}^T$.
3. Arrange the eigenvalues in descending order with the first one unity; this is the vector $\boldsymbol{\lambda}$.
4. $\boldsymbol{V}$ is the matrix whose rows are the eigenvectors corresponding (in the same order) to the eigenvalues λ_i.
5. The stationary distribution $\boldsymbol{\alpha}$ is proportional to the eigenvector corresponding to the eigenvalue of unity; the elements of $\boldsymbol{\alpha}$ should sum to unity since it is a probability distribution.

[14] Assuming the calculations will be performed on a computer.

[15] In many of the examples, we will assume a stationary distribution $\boldsymbol{\alpha}$ and then construct a transition matrix $\boldsymbol{P}$ such that $\boldsymbol{\alpha}\boldsymbol{P} = \boldsymbol{\alpha}$, using the method in Appendix D.

6. $(\mu \times \alpha)$ is the vector whose ith element is $\mu_i \alpha_i$.

7. V^{-1} is the matrix inverse of V.

8. ζ is the vector whose ith element is the product of the ith element of the vector $(\mu \times \alpha)^T V^{-1}$ and the ith element of the vector $V\mu$.

9. For X and U separated by g years, $g > 0$:

$$\mathrm{Cov}[X,U] = \sum_{i>1} \zeta_i \lambda_i^g .$$

The vector ζ is defined in Equation 2.5 in terms of μ, α, and V. μ, the vector of means for each state, and α, the distribution of probabilities for the states, are not dependent on the rate of shifting parameters.

V is a matrix whose rows are the eigenvectors of P^T. The eigenvectors of $(P^g)^T = (P^T)^g$ are the same as those of P^T. By raising P to a power, one can alter the rate at which parameters shift over time without changing the eigenvectors. Therefore, since V does not depend on the power to which P is raised, it does not reflect the speed of shifting risk parameters.

Therefore, ζ, which is calculated from μ, α, and V, reflects the "structure" of the Markov chain rather than the rate of shifting risk parameters. In contrast, the eigenvalues λ_i *do* reflect the rate at which risk parameters shift. If P is raised to the power g, so are the eigenvalues.

Thus writing the covariance between two years of data in terms of ζ and λ as in Equation 2.6 isolates the effect of the rate of shifting parameters into λ_i^g.

2.7. *The Variance-Covariance Matrix*

For a single year, the variance of X (the covariance of X with itself) is not affected directly by shifting risk parameters over time.[16] With stationary probabilities of being in the different states, the variance of X can be calculated ignoring shifting risk parameters.

The variance of X is computed in the usual way as the sum of the variance of hypothetical means and the expected value of process variance. As discussed previously in Section 2.1, for Joe rolling a single die, the variance of hypothetical means is .50 and the expected value of process variance is 3.08. Therefore, the total variance of X is 3.58.

For this example, as calculated in Appendix E, the covariances for trials separated by given amounts are:

Separation	Covariance[17]	Covariance ÷ Variance of Hypothetical Means
0	3.5833[18]	
1	.3750	.750
2	.2837	.568
3	.2159	.432
4	.1649	.330
5	.1263	.253
6	.0968	.194
7	.0743	.149
8	.0570	.114
9	.0438	.088
10	.0337	.067
20	.0024	.0048
30	.0002	.0004

In this case for $g > 0$, the covariances are $(.468)(.769^g) + (.032)(.418^g)$. Therefore, the variance of hypothetical means is .5, we expect the covariance ÷ variance of hypothetical means to be approximately $.77^g$.

[16]To the extent the probabilities of being in different states is not stationary, the expected value of the process variance may move over time. That is not the situation here.

[17]Based on exact calculation with no intermediate rounding.

[18]Total variance = sum of the variance of hypothetical means of .50 and the process variance of 3.08.

Note that setting $g = 0$ in the formula for $\text{Cov}[X,U]$ gives $.468 + .032 = .500$, the variance of hypothetical means. In general, one can calculate the variance of X by getting the variance of hypothetical means in this manner and adding the expected value of process variance. The process variance will depend, among other things, on the particular type of process, e.g., binomial, Poisson, negative binomial, rolling a die, spinning a spinner, etc.

In summary, as shown in Appendix E, the variance-covariance matrix between years of data is given by

$$\text{Cov}[X,U] = (.468)(.769^g) + (.032)(.481^g) + 3.08 \quad (\text{if } g = 0).$$

In general, for years of data X_i and X_j:

$$\text{Cov}[X_i,X_j] = \sum_{k>1} \zeta_k \lambda_k^{|i-j|} + \delta_{ij} \quad (\text{EPV}) \tag{2.7}$$

where

$$\text{EPV} = \text{expected value of the process variance}$$

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

2.8. *Credibilities*

Assume we have data from years $1,2,\ldots,Y$ and we wish to predict the outcome in year $Y + \Delta$. Then, as shown in Appendix C, the least squares credibilities are given by solving the Y linear equations in Y unknowns:[19]

$$\sum_{j=1}^{Y} \text{Cov}[X_i,X_j]Z_j = \text{Cov}[X_i,X_{Y+\Delta}] \qquad i = 1,2,\ldots,Y. \tag{2.8}$$

---

[19]The equations are those in Mahler [10].

Given values for the variances and covariances, one can solve for the credibilities to assign to each year by simple matrix techniques.

In our example, assume we use the outcome of one trial to predict the outcome of the next trial. Then we get one equation:

$$\mathrm{Cov}[X_1,X_1]Z_1 = \mathrm{Cov}[X_1,X_2], \qquad \text{and}$$

$$Z_1 = \frac{\mathrm{Cov}[X_1,X_2]}{\mathrm{Var}[X]} = .3750/3.5833 = 10.5\%.$$

Note that this is lower than the credibility for a single trial in the absence of shifting parameters, which is $.50/3.58 = 14.0\%$.[20]

If we use two years of data to predict the subsequent year, then we get two equations in two unknowns:

$$Z_1\,\mathrm{Cov}[X_1,X_1] + Z_2\,\mathrm{Cov}[X_1,X_2] = \mathrm{Cov}[X_1,X_3], \qquad \text{and}$$

$$Z_1\,\mathrm{Cov}[X_1,X_2] + Z_2\,\mathrm{Cov}[X_2,X_2] = \mathrm{Cov}[X_2,X_3].$$

For this example,

$$3.5833Z_1 + .3750Z_2 = .2837, \qquad \text{and}$$

$$.3750Z_1 + 3.5833Z_2 = .3750.$$

The solution is

$$\begin{pmatrix} 3.5833 & .3750 \\ .3750 & 3.5833 \end{pmatrix}^{-1} \begin{pmatrix} .2837 \\ .3750 \end{pmatrix} = \begin{pmatrix} .069 \\ .097 \end{pmatrix}.$$

We would give 9.7% credibility to the most recent year of data, 6.9% to the second most recent year, and the complement of credibility, 83.4%, to the overall a priori mean of 3.5.

[20]The ratio of credibilities is .75, the ratio of the covariance (with shifting parameters) to the variance of hypothetical means.

Similarly for three years of data, the equations are:

$$3.5833Z_1 + .3750Z_2 + .2837Z_3 = .2159,$$

$$.3750Z_1 + 3.5833Z_2 + .3750Z_3 = .2837, \qquad \text{and}$$

$$.2837Z_1 + .3750Z_2 + 3.5833Z_3 = .3750,$$

which has the solution $Z_1 = 4.6\%$, $Z_2 = 6.4\%$, $Z_3 = 9.4\%$.

If, instead of using years 1 through 3 to predict year 4, we were using them to predict year 5, the right hand sides of the equations would be instead .1649, .2159, and .2837. This would instead result in a solution $Z_1 = 3.5\%$, $Z_2 = 4.9\%$, $Z_3 = 7.1\%$. The additional year of delay in the availability of data has resulted in lower credibilities.[21]

2.9. *Varying the Rate at which Parameters Shift*

One can easily modify this example to either slow down or speed up the rate at which parameters shift over time. For example, the transition probabilities could be revised so it is one-fifth as likely for Beth to switch the type of die after each trial. Such a revised transition matrix

$$\begin{pmatrix} .96 & .04 & 0 \\ .02 & .95 & .03 \\ 0 & .06 & .94 \end{pmatrix}$$

has the same stationary distribution .25, .5, .25, but the parameters shift about one-fifth as fast.

One can speed up the rate at which parameters shift by raising the transition matrix to a power. For example, squaring the given transition matrix yields a new transition matrix in which the parameters shift exactly twice as fast.[22]

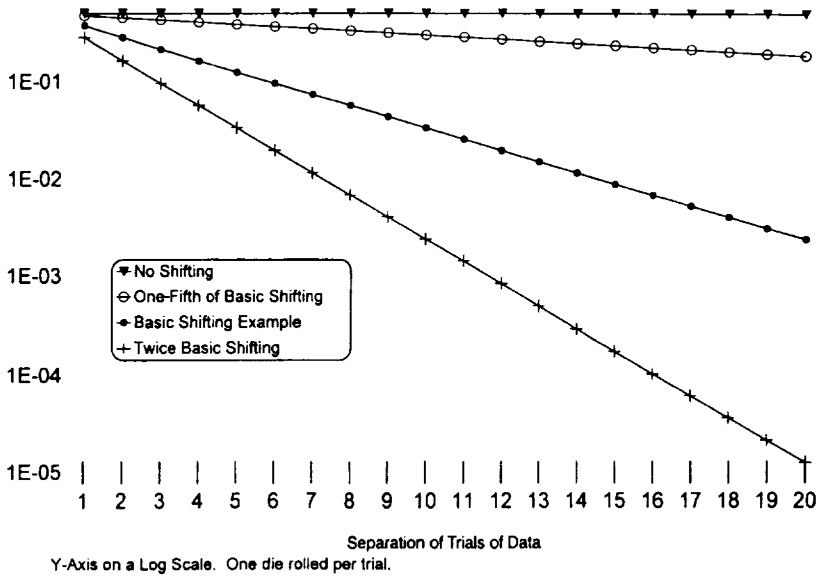
---

[21] Note the Equations 2.8 are sufficiently general to accommodate gaps between the years of data as well as a gap between the last year of data and the year being predicted.

[22] If $\alpha P = \alpha$, we have $\alpha P^2 = (\alpha P)P = \alpha P = \alpha$. Therefore, if α is a stationary distribution for P, it is a stationary distribution for P^2 (or P^3, or P^4, etc.).

FIGURE 1

COVARIANCES BETWEEN DATA SEPARATED BY GIVEN NUMBER OF TRIALS, EXAMPLES WITH DICE, SHIFTING PARAMETERS OVER TIME



The Markov chain that corresponds to the transition matrix $\boldsymbol{P}^2$ results in covariances of data X_i' that follow from those for data X_i from the Markov chain corresponding to $\boldsymbol{P}$:

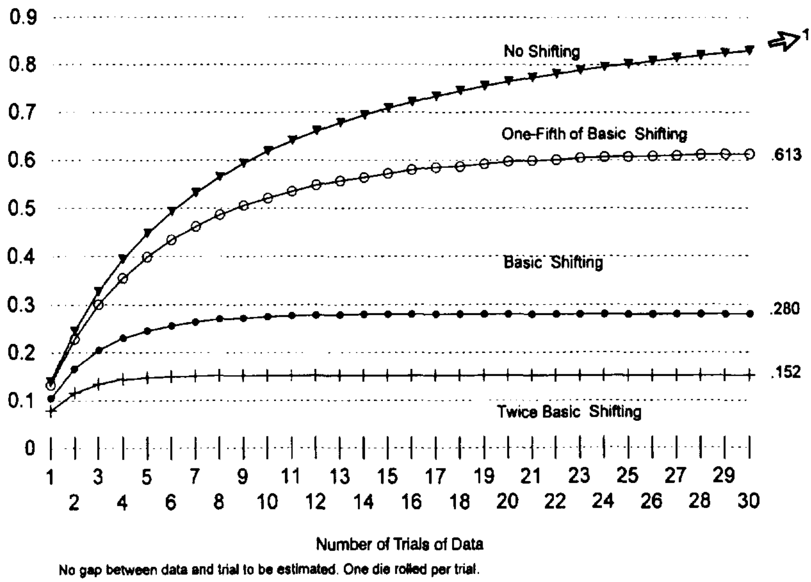
$$\mathrm{Cov}[X_1', X_{1+g}'] = \mathrm{Cov}[X_1, X_{1+2g}].$$

The covariance for a separation of ten years with transition matrix $\boldsymbol{P}^2$ is the same as that for a separation of 20 years with transition matrix $\boldsymbol{P}$.

Figure 1 compares the covariance structure for the basic example to one with no shifting, one-fifth the amount of shifting, and twice the shifting. The vertical axis is on a logarithmic scale. As expected on this logarithmic scale, the covariances decline ap-

FIGURE 2

SUM OF CREDIBILITIES, EXAMPLES WITH DICE, SHIFTING PARAMETERS OVER TIME, VARYING RATES



proximately linearly, with the slope of the decline approximately proportional to the amount of shifting. In the absence of shifting risk parameters over time, the covariances do not decline, rather they are the same regardless of the number of years of separation.[23]

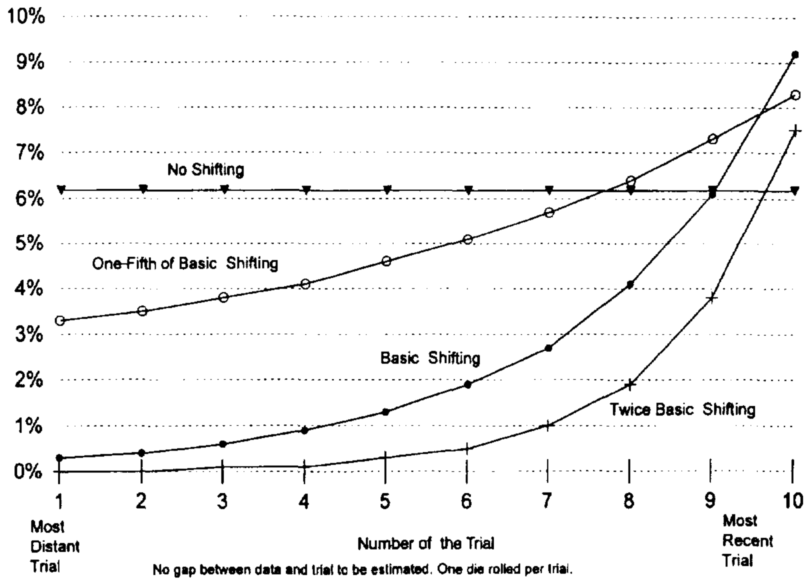
Figure 2 compares the sum of the credibilities one would assign to individual years of data for different amounts of shifting. In the case of no shifting, the sum of the credibilities approaches unity as the number of years approaches infinity.[24] The greater

[23]In many practical applications, the decline will be so small over the time periods of interest that it makes sense to ignore the decline. Thus while the case of no decline *forever* is not realistic, it is a very good approximation for many practical applications.

[24]For this example with no shifting and Y years of data, each year is assigned $1/(Y + 6.1666)$ credibility. The sum is $Y/(Y + 6.1666)$.

FIGURE 3

CREDIBILITIES ASSIGNED TO EACH OF TEN TRIALS OF DATA EXAMPLES WITH DICE, VARIOUS RATES OF SHIFTING PARAMETERS OVER TIME



the shifting, the smaller the sums of the credibilities. The sum of credibilities approaches a value less than unity as the number of years of data increases. The greater the shifting, the lower the limit and the faster it is reached.

Figure 3 compares the credibilities that would be assigned to individual years of data when using ten years of data. In the absence of shifting, each year is assigned equal credibility.[25] The greater the shifting, the greater the difference in credibilities as-

[25]For $Y = 10$, each year is assigned credibility of $1/(10 + 6.1666) = 6.2\%$.

signed to the different years of data. When risk parameters shift rapidly over time, the value of recent information is greater relative to older information.

We note that the most recent year of data is assigned more credibility for the basic example than it is in the absence of shifting. This reflects the fact that in the former situation the *relative* value of the most recent year's data is large compared to the data available from other years. When using ten years of data in the absence of shifting, the total value of the available information is higher, as is the value of the most recent year. However, the value of the most recent year's data relative to all the information available is lower without shifting than with shifting. In contrast, as was shown previously, when using only one year of data, the credibility is lower in the presence of shifting.

Finally, Figure 4 compares the effects of delays in gathering the data. For the basic example, we see how the credibilities decrease as the delay increases. When risk parameters are changing quickly over time, the effect of any delay in collecting data can be very substantial.

2.10. *Size of Risk and Shifting Risk Parameters*

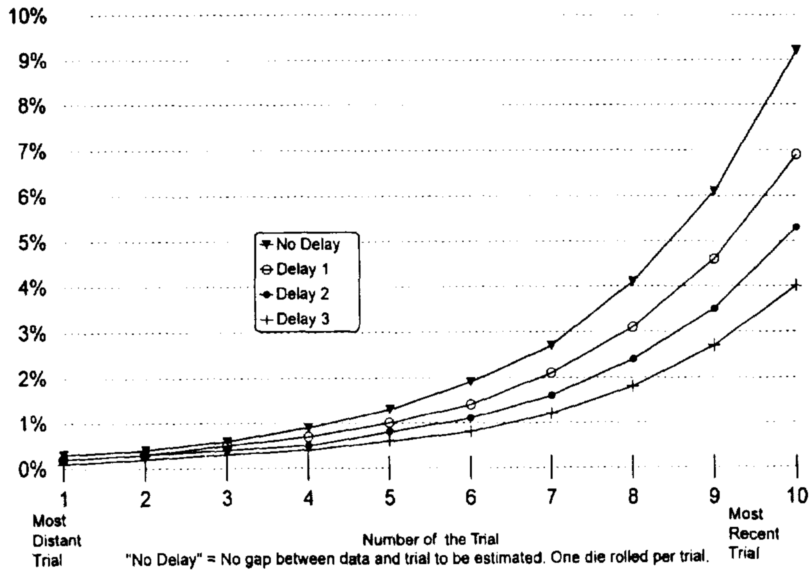
Assume that Joe selects N dice (of the same kind) and rolls them. The resulting sum is the result of one trial or year. After each trial, Beth (possibly) changes the type of dice with transition matrix $\boldsymbol{P}$. (We assume Beth either changes the type of all N dice or leaves them all alone.)

Since we are just adding the results of rolling N identical dice in each year, the covariance between two separate years is given by N^2 times what it was for the case with a single die:

$$\mathrm{Cov}[X_1, X_{1+g}] = N^2 \sum_{i>1} \zeta_i \lambda_i^g.$$

FIGURE 4

CREDIBILITIES ASSIGNED TO EACH OF TEN TRIALS OF DATA, EXAMPLES WITH DICE, EFFECTS OF VARIOUS DELAYS, SHIFTING PARAMETERS OVER TIME



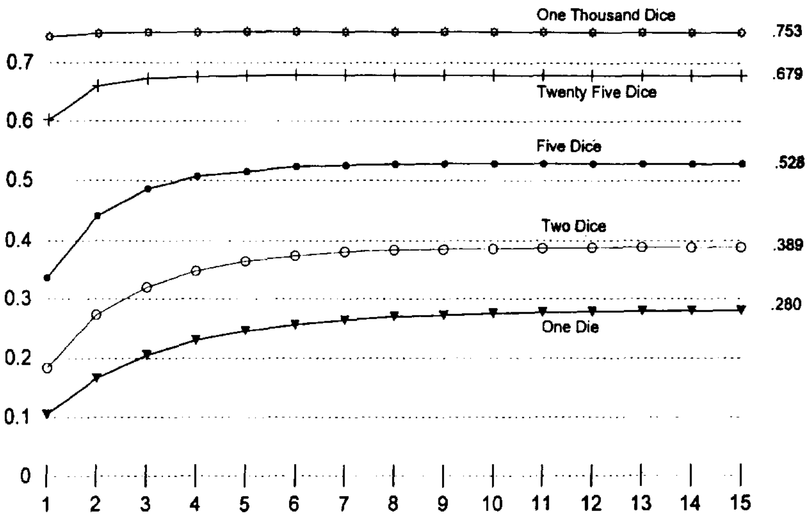
The variance for a year is given by N (expected value of process variance for a single die) + N^2 (variance of hypothetical means for a single die).

As before, given Y years of data, we can solve Y equations in Y unknowns[26] for the credibilities assigned to each year. As in the case of standard Bühlmann credibility, since the expected value of process variance increases with N rather than N^2, as N increases, so does the credibility.

[26]The same Equations 2.8 apply, but the actual values of the variance and covariances depend on N, the number of dice.

FIGURE 5

SUM OF CREDIBILITIES, SHIFTING PARAMETERS OVER TIME, VARIOUS NUMBERS OF DICE PER TRIAL



No gap between data and trial to be estimated. Various numbers of identical dice rolled per trial.

Figure 5 compares the sum of credibilities assigned to Y years for different numbers of dice, N, for the transition matrix discussed previously:

$$\begin{pmatrix} .80 & .20 & 0 \\ .10 & .75 & .15 \\ 0 & .30 & .70 \end{pmatrix}.$$

As expected, the more dice used per roll, the higher the credibility. Also, the more dice, the quicker the limit is approached as the number of years of data increases. For a fixed amount of shifting, for larger risks, the more recent years are relatively more valuable compared to older years than is the case for smaller risks. For larger risks, the random noise in the observation of a single year is less, so one can rely on fewer years of information.

As the number of dice approaches infinity, one relies almost solely on the most recent year of data. In this case, the sum of credibilities approaches 75.6%, with about 74% credibility being assigned to the most recent year.

In general, as the number of dice approaches infinity, the sum of the credibility approaches a number less than unity in the presence of shifting risk parameters. Having Joe roll more dice per trial does not get rid of the effect of Beth (possibly) shifting all the dice between trials. Increasing the size of the risk will not eliminate the uncertainty caused by shifting risk parameters over time.[27]

3. SIMPLE POISSON EXAMPLE

3.1. *Bühlmann Credibility*

To take a simplified insurance example, assume that for individual insureds the claim frequency in each year is given by a Poisson distribution.[28] Assume that there are four types of insureds with different frequencies:

Type of Insured	A Priori Probability	Mean Frequency
Excellent	40%	.25
Good	30%	.50
Bad	20%	.75
Ugly	10%	1.00

Then the overall mean is .50. The variance of hypothetical means is 1/16. The expected value of process variance is the

[27]Note in the model used here, the rate of shifting was assumed to be independent of the size of risk. This was a simplifying assumption which may or may not be a reasonable approximation to a particular real world application.

[28]For parameter θ, $f(n) = e^{-\theta}\theta^n/n!$. The mean and variance of the Poisson are each equal to θ.

expected value of hypothetical means[29] which is the overall mean of .50.

The Bühlmann credibility parameter is $K = .50/(\frac{1}{16}) = 8$. Therefore, the credibility of Y years of data from an individual insured is:[30]

$$Z = \frac{Y}{Y+8}. \tag{3.1}$$

For example, one year of data would be assigned a credibility of about 11%. Note that the total variance is $.5 + \frac{1}{16} = .5625$. The credibility of a single year is the variance of hypothetical means divided by the total variance $= .0625/.5625 = \frac{1}{9}$.

3.2. *Shifting Risk Parameters, Simple Poisson Example*

Assume that in the previous example, the individual insured has a chance of shifting states each year. For example, an excellent insured might have an 18% chance of switching to a good insured the following year,[31] and an 82% chance of remaining an excellent insured. Assume the following transition matrix for illustrative purposes:

$$\begin{pmatrix} .820 & .180 & 0 & 0 \\ .240 & .592 & .168 & 0 \\ 0 & .252 & .608 & .140 \\ 0 & 0 & .280 & .720 \end{pmatrix}.$$

This transition matrix has the selected initial distribution (.4,.3,.2,.1) as a stationary distribution.[32]

---

[29] Since for the Poisson, the mean is equal to the variance.

[30] We assume that we do not know what type of risk the individual is and that the complement of credibility is to be assigned to the overall mean.

[31] Note that we are referring to presumed changes in the unobservable expected claim frequency rather than observed changes in the actual number of claims from year to year.

[32] It can be easily verified that $(.4,.3,.2,.1)\boldsymbol{P} = (.4,.3,.2,.1)$. See Appendix D for how this transition matrix was constructed.

As with the dice example, one can compute the variance-covariance matrix and thus the credibilities.

The expected value of process variance for a single year in this example is .50. Note that this depends on the fact that for each insured for each year we have assumed a Poisson process.

The transpose of the transition matrix has eigenvalues of

$$\lambda = (1, .855, .580, .305).$$

The eigenvectors are the rows of:

$$V = \begin{pmatrix} 1 & .75 & .5 & .25 \\ 1 & .1456 & -.5623 & -.5833 \\ 1 & -1 & -.6667 & .6667 \\ 1 & -2.1456 & 1.729 & -.5833 \end{pmatrix} \quad \text{and}$$

$$V^{-1} = \begin{pmatrix} .4 & .3309 & .2 & .0691 \\ .4 & .0643 & -.2667 & -.1976 \\ .4 & -.3722 & -.2666 & .2388 \\ .4 & -.7722 & .5333 & -.1612 \end{pmatrix}$$

μ = (.25, .50, .75, 1.00) = the assumed means
α = (.40, .30, .20, .10) = the stationary distribution
$(\mu \times \alpha)$ = (.10, .15, .15, .10)
$(\mu \times \alpha)V^{-1}$ = (.2, −.0903, −.0067, −.0030)
$V\mu$ = (1.25, −.6822, −.0833, −.1094)

ζ has as its ith element the product of the ith element of the above two vectors, as shown in Equation 2.5; therefore,

$$\zeta = (.25, .0616, .0006, .0003).$$

Therefore, for $g > 0$, the covariance of two different years is given by Equation 2.6:

$$\mathrm{Cov}[X_1, X_{1+g}] = \sum_{i>1} \zeta_i \lambda_i^g$$

$$\mathrm{Cov}[X_1, X_{1+g}] = (.0616)(.855^g) + (.0006)(.580^g)$$

$$+ (.0003)(.305^g) \qquad \text{for} \quad g > 0. \qquad (3.2)$$

For a single year, we set $g = 0$ and add the expected value of process variance:[33]

$$\mathrm{Var}(X) = .0625 + .5 = .5625. \qquad (3.3)$$

One can use this variance-covariance structure in the Equations 2.8 for the credibilities. For example, if using three years of data X_1, X_2, X_3 to estimate the next year, X_4, then the three equations in three unknowns are:[34]

$$.5625Z_1 + .0531Z_2 + .0453Z_3 = .0386,$$

$$.0531Z_1 + .5625Z_2 + .0531Z_3 = .0453, \qquad \text{and}$$

$$.0453Z_1 + .0531Z_2 + .5625Z_3 = .0531.$$

The solution is $Z_1 = 5.6\%$, $Z_2 = 6.7\%$, and $Z_3 = 8.4\%$. Table 1 displays the solutions for various numbers of years of data.

Figure 6 shows the sum of the credibilities both in the presence of shifting risk parameters and in the absence of shifting risk parameters.[35] Also shown are credibilities corresponding to twice the original rate of shifting[36] and to five times the original rate of shifting.[37] As was seen before, the presence of shifting risk parameters lowers the credibilities. The more rapid the shifting, the greater the effect on the credibilities.

---

[33]This matches the result prior to considering shifting parameters over time, as it should.

[34]$\mathrm{Cov}[X_1, X_2] = .0531$. $\mathrm{Cov}[X_1, X_3] = .0453$. $\mathrm{Cov}[X_1, X_4] = .0386$. $\mathrm{Var}[X] = .5625$.

[35]As was seen in the previous section, in the absence of shifting risk parameters, $1/(Y + 8)$ credibility is assigned to each of Y years for a total of $Y/(Y + 8)$.

[36]Based on using the square of the original transition matrix.

[37]Based on using the fifth power of the original transition matrix.

TABLE 1

CREDIBILITY
SIMPLE POISSON EXAMPLE WITH SHIFTING RISK PARAMETERS
(No Delay in Receiving Data)

Years Between Data and Estimate	Number of Years of Data Used					
	1	2	3	4	5	10
1 (Most Recent)	9.4%	8.8%	8.4%	8.1%	8.0%	7.8%
2		7.2%	6.7%	6.4%	6.3%	6.0%
3			5.6%	5.2%	5.0%	4.7%
4				4.3%	4.0%	3.7%
5					3.3%	2.9%
6						2.2%
7						1.8%
8						1.4%
8						1.1%
10						0.9%
Total Credibility	9.4%	16.0%	20.7%	24.0%	26.6%	32.5%

With shifting risk parameters, as the number of years of data approaches infinity, the sum of the credibilities approaches a limit less than unity. For faster shifting, this limit is lower and it is approached more rapidly.

Since the first term in the covariance in Equation 3.2 dominates, the variance-covariance structure in Equations 3.2 and 3.3 can be approximated by:

$$\mathrm{Cov}[X_i,X_j] = (.0625)(.85^{|i-j|}) + .5\delta_{ij} \tag{3.4}$$

where $\delta_{ij} = 0$ for $i \neq j$ and 1 for $i = j$.

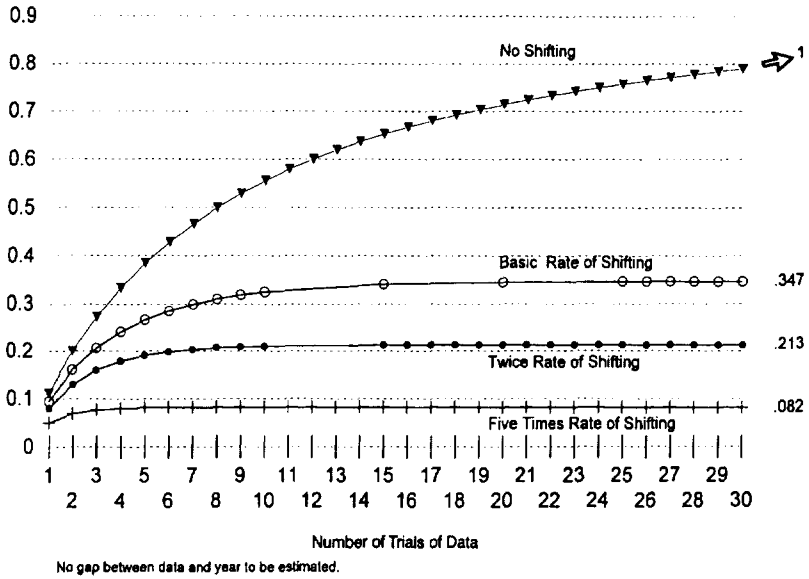
In general,

$$\mathrm{Cov}[X_i,X_j] = \tau^2\lambda^{|i-j|} + \delta_{ij}\eta^2 \tag{3.5}$$

where η^2 is the expected value of process variance, τ^2 is the variance of hypothetical means and λ is the dominant eigen-

FIGURE 6

SUM OF CREDIBILITIES, SIMPLE POISSON EXAMPLE, VARIOUS RATES OF SHIFTING PARAMETERS OVER TIME



value (other than unity) of the transpose of the transition matrix of the Markov chain.

For one year of data, predicting year $1 + \Delta$, the credibility is obtained by solving Equation 2.8:

$$Z(\tau^2 + \eta^2) = \tau^2\lambda^\Delta \qquad Z = \lambda^\Delta/(1 + K) \tag{3.6}$$

where $K = \eta^2/\tau^2$ = Bühlmann credibility parameter.

As shown in Mahler [9], when one has years 1 to Y predicting year $Y + \Delta$, the sum of the credibilities is approximately:[38]

[38] As discussed on pages 162–164 of Mahler [9], this approximation underestimates the credibilities. However, here we have also approximated the covariances, therefore the approximation can go in either direction.

$$\sum_{i=1}^{Y} Z_i \approx \frac{\lambda^{\Delta}\left(\sum_{i=1}^{Y} \lambda^{i-1}\right)}{\sum_{i=1}^{Y} \lambda^{i-1} + K}. \tag{3.7}$$

In the absence of shifting risk parameters, $\lambda = 1$ and the sum of the credibilities given by Equation 3.7 becomes the familiar $Y/(Y+K)$.

In the current example, $\lambda = .855$, $\eta^2 = .5$, $\tau^2 = .0625$, and $K = \eta^2/\tau^2 = 8$. Thus Equation 3.7 becomes for $\Delta = 1$

$$\sum_{i=1}^{Y} Z_i \approx \frac{(.855)\left(\sum_{i=1}^{Y} (.855)^{i-1}\right)}{\left(\sum_{i=1}^{Y} (.855)^{i-1}\right) + 8}. \tag{3.8}$$

For $Y = 3$, Equation 3.8 gives $\sum Z_i \approx 20.9\%$. As seen above, the exact solution gives $Z_1 + Z_2 + Z_3 = 5.6\% + 6.7\% + 8.4\% = 20.7\%$, which happens to be somewhat lower in this case.

As the number of years of data increases in Equation 3.7, the approximate sum of the credibilities approaches:

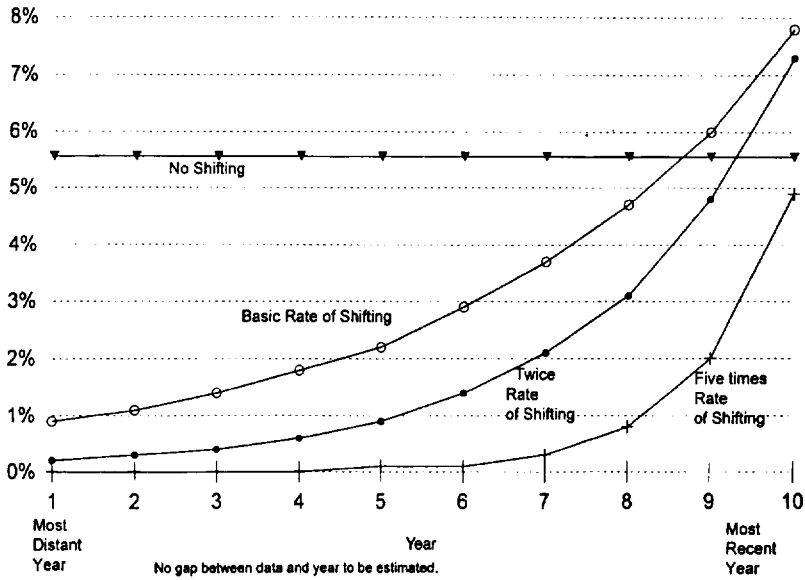
$$\frac{\lambda^{\Delta}\left(\frac{1}{1-\lambda}\right)}{\left(\frac{1}{1-\lambda}\right) + K} = \frac{\lambda^{\Delta}}{1 + K(1-\lambda)}.$$

In this example, for $\Delta = 1$, $\lambda = .855$ and $K = 8$, the sum of the credibilities approaches approximately 39.6%. As seen in Figure 6, the sum of the credibilities actually approaches 34.7%. Thus while this approximation is conceptually useful, one should be cautious in using it for precise numerical results.

Figure 7 displays the credibilities that would be assigned to each of ten years of data. In the absence of shifting risk parameters, each year of data is assigned equal credibility. With shifting

FIGURE 7

CREDIBILITIES ASSIGNED TO EACH OF TEN YEARS OF DATA, SIMPLE POISSON EXAMPLE, VARIOUS RATES OF SHIFTING PARAMETERS OVER TIME



risk parameters, more recent years of data are given more weight than older years of data. The faster the shifting, the less weight is given to the older years of data.

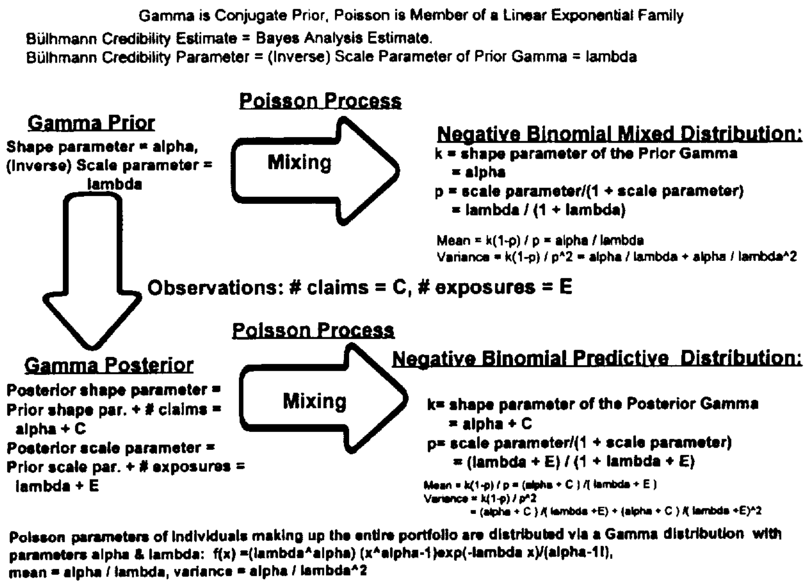
4. CALIFORNIA DRIVING DATA

Mahler [7] examined California driving data. Two sets of data were examined: male and female drivers. The latter set showed more significant evidence of shifting parameters over time. The Markov chain model will be used to model the data for female drivers.[39]

[39]The techniques could be applied in a similar manner to the male drivers.

FIGURE 8

GAMMA-POISSON FREQUENCY PROCESS



For 23,872 female drivers over a period of nine years, there were 7,988 accidents, for an annual accident frequency of .0372. The average variance of a year of data was .0386.

4.1. *Gamma-Poisson*

Such data can be commonly fit with a "gamma-Poisson" in which each insured's frequency is a Poisson process and the Poisson parameters vary over the portfolio via a gamma distribution.[40] Key features of the gamma-Poisson are displayed in Figure 8. The frequency distribution for the portfolio is negative

---

[40]See for example, Mayerson [11], Dropkin [1], Herzog [3], and Hossack, Pollard, and Zehnwirth [6].

TABLE 2

NUMBER OF DRIVERS WITH VARYING NUMBERS OF CLAIMS OVER NINE YEARS

Number of Claims	California Female Drivers	Maximum Likelihood Negative Binomial*	Markov Chain Simulation
0	17,649	17,654	17,695
1	4,829	4,822	4,852
2	1,106	1,101	1,029
3	229	235	239
4	44	48	49
5	9	10	6
6	4	2	2
7	1	0	0
8	1	0	0
9+	0	0	0
	23,872	23,872	23,872

*Negative binomial distribution with parameters $p = .8164$, $k = 1.4876$. The California data has a total of 7,988 accidents while the simulated data has a total of 7,865.

binomial. As shown in Table 2, a negative binomial is a reasonable fit to this data.[41] The overall mean is the mean of the gamma distribution. The total variance minus the mean is the variance of the gamma distribution. Thus one can use the method of moments to determine the parameters of the gamma distribution; for this data, the mean of the gamma would be .0372 and the variance of the gamma would be $.0386 - .0372 = .0014$.

For a gamma distribution with shape parameter α and (inverse) scale parameter λ, this would lead to two equations:

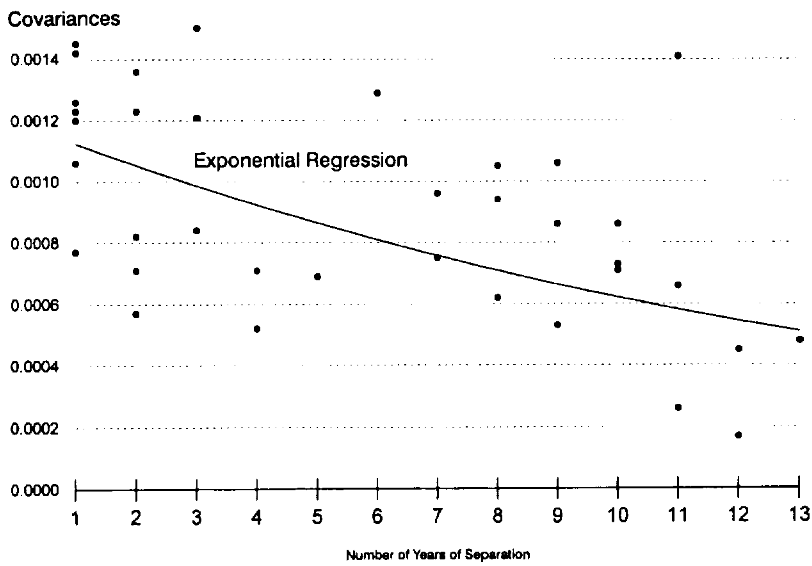
$$\alpha/\lambda = 0.372, \qquad \text{and}$$

$$\alpha/\lambda^2 = .0014.$$

[41] As discussed in reviews of Dropkin [1], this does not imply that the gamma-Poisson model is appropriate.

FIGURE 9

COVARIANCES VERSUS YEARS OF SEPARATION
CALIFORNIA FEMALE DRIVER DATA



This would give values of $\lambda = 26.6$ and $\alpha = .988$. If the shape parameter $\alpha = 1$, one would get an exponential distribution. As a first approximation, assume the frequencies are given by an exponential density function with parameter 26.9: $f(\theta) = 26.9e^{-26.9\theta}$ with mean $1/26.9 = .03717$.

4.2. *Shifting Parameters*

As stated above, the data for California female drivers shows evidence of shifting risk parameters over time. The covariances between years of data with given separations is shown in Figure 9. The covariances appear to decrease for larger separations. The observed covariances were fit to an exponential regression: $\mathrm{Cov}[X_i, X_{i+g}] = .00120e^{-.066g}$, where g is the years of separation.

This is the same general type of behavior one would expect from a Markov chain model of shifting parameters over time. In order for a Markov chain model to fit the observed covariances, the variance of hypothetical means should be about .0012, since setting $g = 0$ in the Markov model gives the variance of hypothetical means.[42] The factor in the exponent, $-.066$, should approximate the log of the dominant eigenvalue (other than unity) of the transition matrix, since this is the approximate rate of decline of the log of the covariances in the model. Thus, in order to match the observed decline, the dominant eigenvalue(s) (other than unity) must be about $e^{-.066} \approx .94$.

4.3. Markov Chain Model

In order to apply the Markov chain model, one has to convert the assumed continuous distribution of frequency parameters into a discrete approximation. For example, take mean frequencies of:

$$\theta_i = .0025, .0075, .0125, \ldots, .3975 \qquad i = 1, 2, \ldots, 80.$$

Take the (initial) probabilities of being in each of these 80 states as α_i proportional to $e^{-26.9\theta_i}$, such that the sum of the α_i is unity.[43] Then, as shown in Appendix D, one can construct an (80×80) transition matrix that has these α as a stationary distribution. For illustrative purposes, assume about $\frac{2}{3}$ chance of shifting up or down a state per year.[44] For this transition matrix,

---

[42]Given the random fluctuation in the data, this estimate of .0012 is not inconsistent with the previous estimate of .0014.

[43]This is a discrete approximation to the selected exponential distribution. The technique will work exactly the same for a gamma distribution with a shape parameter other than unity (which is an exponential).

[44]In order to match the observed covariance structure, this transition matrix will be taken to an appropriate power.

the first ten elements of ζ and λ are:[45]

i	ζ_i	λ_i
1	.00139	1
2	.00059	.9980
3	.00042	.9964
4	.00019	.9939
5	.00008	.9903
6	.00004	.9857
7	.00002	.9801
8	.00001	.9735
9	.00001	.9659
10	.00000	.9574

with all the remaining elements of $\zeta < .0001$. Since only the first few terms contribute significantly to the sum that calculates the model covariances,

$$\mathrm{Cov}[X_1, X_{1+g}] = \sum_{i>1} \zeta_i \lambda_i^g \approx .0013(.997^g) \qquad g > 0.$$

As discussed previously, in order to approximate the observed covariance structure for California female drivers, one would want a decline in the log covariances of about $-.066g$. The above transition matrix has a decline of the log covariances of about $-.003g$. Raising the above transition matrix to the 20th power[46] will multiply the decline in log covariances by about a factor of 20, producing a decline of about $-.060g$, and so should roughly approximate the observed decline.

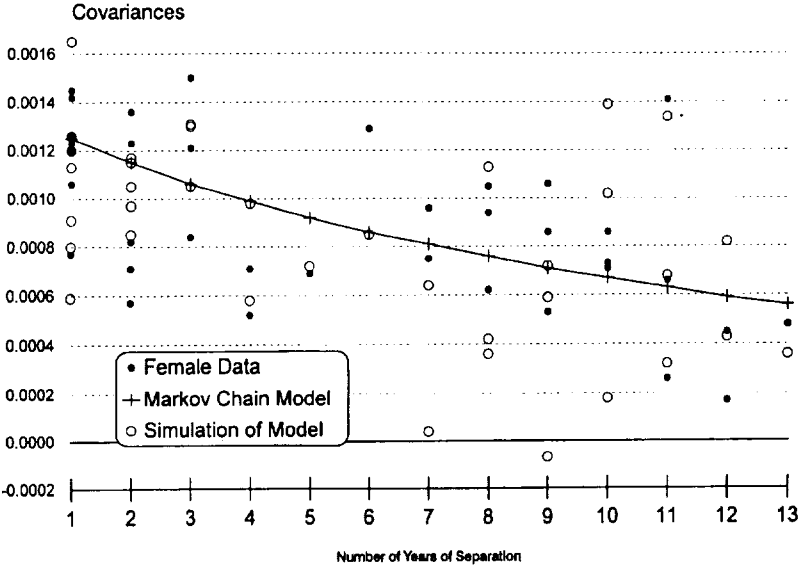
The model covariances for such a transition matrix are shown in Figure 10. The model covariances are a reasonable fit to the

[45] See the previous discussion and Equation 2.5 for the definition of ζ. λ is the set of eigenvalues of the transpose of the transition matrix.

[46] Taking the transition matrix to the 20th power gives a matrix whose eigenvalues are all taken to the 20th power. Thus the logs of all the eigenvalues are multiplied by 20. Since the log covariances decline approximately proportionally to the log of the dominant eigenvalue (other than one), they will decline about 20 times as fast for the new transition matrix as for the original.

FIGURE 10

COVARIANCES VERSUS YEARS OF SEPARATION, FEMALE DRIVER DATA VS. SIMULATED DATA, SHIFTING PARAMETERS OVER TIME



observed data.[47] While the observations extend out to a 13 year separation, one can calculate the model covariances for any number of years of separation.[48]

4.4. *Simulation*

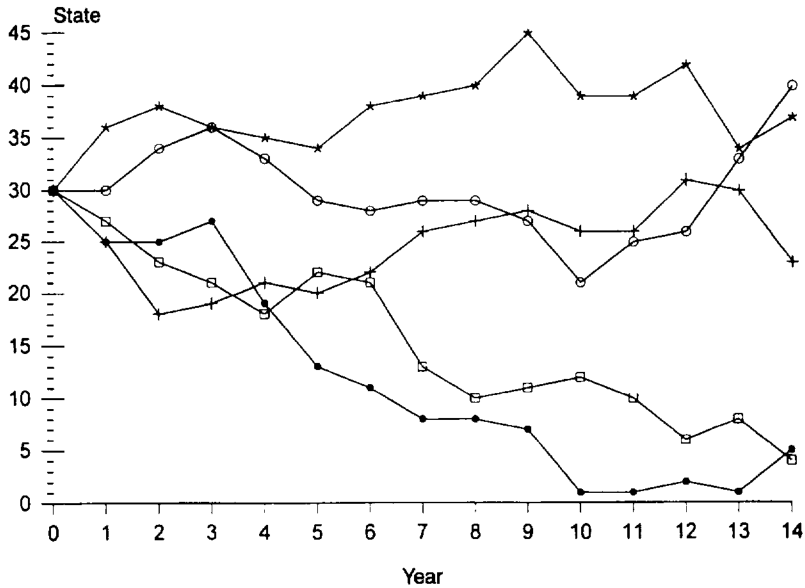
A simulation of this Markov chain model was performed. The first step is to simulate the movement of the Poisson parameters

[47]The limited amount of data would allow other models to fit reasonably well. The observed fit indicates that the form of the proposed model might be useful. It falls well short of demonstrating that it is superior to some other form of model. However, it is clearly superior to a static model without shifting risk parameters.

[48]Beyond 13 years, the model covariances follow a power curve, declining slowly towards zero.

FIGURE 11

MARKOV CHAIN
ILLUSTRATIVE EXAMPLES OF MOVEMENT FOR FIVE RISKS



from year to year, using the probabilities in the transition matrix.[49] Figure 11 shows the result for five risks, each of which started out in State 30, in Year 0. Over the course of 14 years, these risks randomly moved up and down from state to state, with corresponding changes in their assumed expected claim frequency.[50]

The initial configuration of 23,872 drivers by state in Year 0 was chosen to match the selected probability distribution. Then

[49]The selected 80 × 80 transition matrix was the constructed transition matrix to the 20th power. The constructed transition matrix had an average chance of shifting of about $\frac{2}{3}$ per year and had the selected discrete exponential distribution as a stationary distribution.
[50]State 30 corresponds to a Poisson frequency of .1475.

each of the risks moved randomly each year from state to state via the Markov chain. The five risks shown in Figure 11 for illustrative purposes ended up in vastly different states at the end of the 14 year period. They each started with the same assumed annual Poisson frequency of 14.75% in Year 0. Over the course of the 14 year period, they had Poisson parameters ranging from .25% to 22.25%.

There were 61 risks initially in State 30. Over the course of time, their expected claim frequency declined towards the average of 3.7% for the portfolio:

Risks in State 30 in Year 0

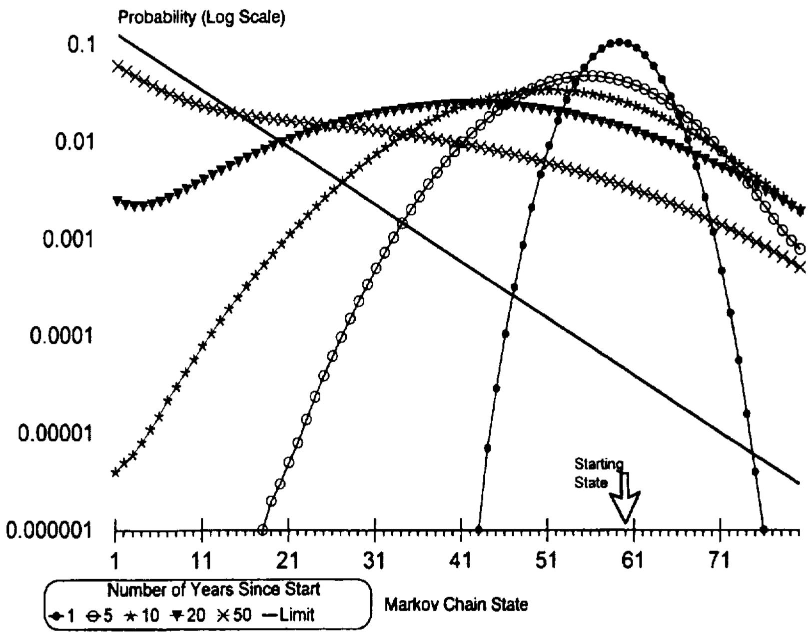
Year	Average Poisson Parameter
0	14.75
1	13.87
2	13.29
3	13.27
4	12.89
5	12.84
6	12.24
7	11.98
8	11.66
9	11.33
10	11.76
11	10.39
12	9.84
13	9.25
14	9.15

After 14 years, the average frequency for these risks moved reasonably towards the overall average.[51] Given enough time, the

[51] The speed at which this occurred was dependent on the particular speed of the shifting parameters over time selected for this example.

FIGURE 12

HAVING STARTED IN STATE 60, CHANCE OF BEING IN A CERTAIN STATE MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVERS

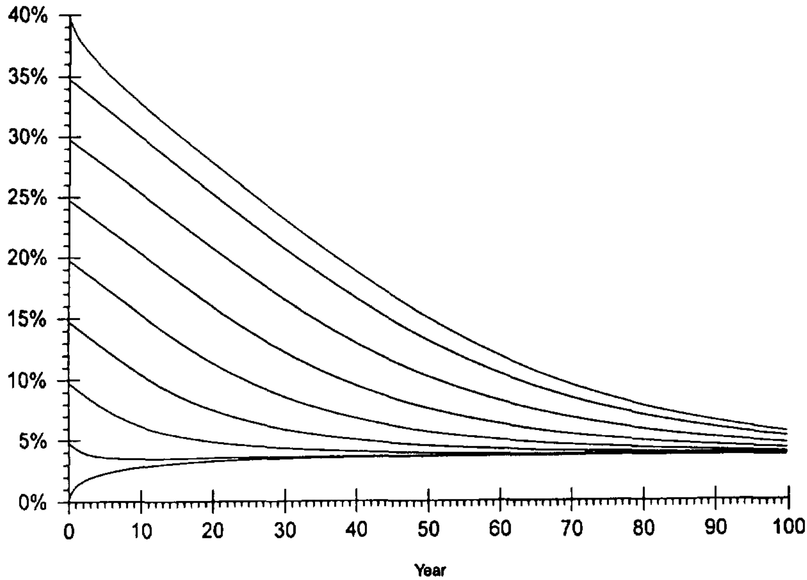


average frequency would have become virtually indistinguishable from the overall average. Figure 12 shows how the distribution evolves over time for risks that start in State 60, (with an initial expected claim frequency of 29.75%). Over time, the distribution approaches the assumed stationary exponential distribution.

This illustrates a general feature of Markov chains: initial information fades over time. For each risk, we have *not* modeled a very long term expected risk propensity that is different than

FIGURE 13

AVERAGE CLAIM FREQUENCY OVER TIME OF DRIVERS STARTING IN VARIOUS STATES MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVERS



average. Rather, the very long term expected risk propensity is the same for each risk. This is again illustrated in Figure 13, which shows how the expected claim frequency approaches the overall average of 3.7% regardless of which Markov chain state the insured started in. In some applications, this may prevent the model from being useful.

Looking at all 23,782 risks, the simulation resulted in a generally similar mix of Poisson parameters each year. So while individual risks' Poisson parameters changed, the portfolio as a whole was approximately "stationary" over time. For example, the mean frequencies and variances of the portfolio for this simulation were:

Year	Mean Poisson Parameter	Variance of Poisson Parameters	Number of Risks in State 30
1	.03724	.001384	72
2	.03715	.001390	72
3	.03697	.001393	63
4	.03701	.001388	69
5	.03707	.001386	62
6	.03706	.001386	64
7	.03691	.001376	58
8	.03712	.001391	60
9	.03703	.001371	65
10	.03707	.001366	51
11	.03684	.001361	47
12	.03687	.001356	51
13	.03684	.001360	53
14	.03689	.001364	47

Also shown for illustrative purposes is the number of risks in State 30 in each (simulated) year. It fluctuates considerably around its expected value of 61. When looked at in this level of detail, the simulation of Poisson parameters results in some differences in the portfolio composition from year to year. In this case, the states are only .5% apart in annual claim frequency, so exactly how many risks are in any single state is of no practical importance, as well as being unobservable in the real world.

The covariances between the Poisson parameters for the simulation decrease approximately in the manner expected by the model (see Table 3). Thus the simulation of the Poisson parameters in this example does not introduce much random fluctuation into the covariances between years.[52]

[52]With a different number of drivers or different transition matrix, the result could differ.

TABLE 3

COVARIANCES (.00001)

Years of Separation	Model	Simulated Poisson Parameters
1	125	126,126,127,126,126,125,126,125,124,123,123,123,123,124
2	115	116,116,116,116,115,116,115,115,114,113,113,113
3	106	107,107,107,107,106,106,106,105,105,105
4	99	100,100,99,100,99,99,98,98,98,98
5	92	93,92,93,92,92,92,92,92,91
6	86	86,87,86,87,86,86,86,86
7	81	81,81,81,81,81,81,81
8	76	75,76,76,76,76
9	71	71,71,71,71,72
10	67	67,67,67,67
11	63	63,63,63
12	59	60,60
13	56	56

Unfortunately, the second step of the simulation does introduce considerable fluctuation into this example. Once one has a set of Poisson parameters (one for each driver), one can simulate the number of accidents that each driver had in a year. In the particular example, since the annual accident frequencies are so low, there is a lot of noise relative to the information. Any one simulation of a year of accident data does not provide much information. In particular, the covariances between simulated years of accident data are subject to considerable random fluctuation.

For example, for two years of Poisson parameters[53] with a covariance of .00086 between the two years, the covariances between years for seven simulated sets of accident data were:

.00115, .00091, .00113, .00040, .00121, .00118, and .00050.

[53]Of 23,872 drivers distributed as per the model of female drivers in California.

This large amount of random fluctuation implies that one should not draw very precise conclusions from the limited available data.

Figure 10 compares the observed covariances for the California female driving data and those for a set of data simulated using the Markov chain model. Within the context of the large amount of random fluctuation, the actual and simulated data sets look generally similar.

Table 2 compares the numbers of insureds with various numbers of accidents over nine years. The simulated data seems to have a somewhat lighter tail than the observed data, although the overall fit is not unreasonable.[54] One could revise the particular inputs used here to attempt to get a somewhat heavier tail. One could increase the variance of hypothetical means[55] and/or have relatively less shifting over time for high frequency drivers.[56] However, these details are beyond the scope of this paper.

One should note that adding shifting risk parameters in the manner done here reduces the probability of an extremely large number of accidents for an insured over an extended period, since the Poisson parameter for an insured tends towards the overall average over time. The most likely insureds to have extremely large numbers of accidents are those whose Poisson parameters are high for all the observed years.

Overall, the Markov chain model presented here does a reasonable job of fitting the female driver data from California. On the other hand, due to the limited amount of data, one should be cautious in drawing any definitive conclusions. There are un-

[54]The negative binomial fit to the data also seems to have a slightly light tail compared to the data, indicating that perhaps a gamma-Poisson model might be improved upon.

[55]For the gamma distribution, the variance of hypothetical means is the overall mean divided by the shape parameter of the gamma. Thus for a fixed overall mean, the smaller the shape parameter, the larger the variance of hypothetical means.

[56]The particular transition matrix (which was raised to the 20th power) assumed approximately $\frac{2}{3}$ chance of shifting per year regardless of the state. One could have had the amount of shifting depend on the accident frequency.

TABLE 4

CREDIBILITY
FEMALE CALIFORNIA ACCIDENT DATA
MARKOV CHAIN MODEL
(No Delay in Receiving Data)

Years Between Data and Estimate	Number of Years of Data Used					
	1	2	3	4	5	10
1 (Most Recent)	3.2%	3.1%	3.1%	3.0%	3.0%	2.8%
2		2.9%	2.8%	2.7%	2.7%	2.5%
3			2.6%	2.5%	2.4%	2.3%
4				2.3%	2.2%	2.1%
5					2.1%	1.9%
6						1.7%
7						1.6%
8						1.5%
8						1.4%
10						1.3%
Total Credibility	3.2%	6.0%	8.5%	10.5%	12.4%	19.1%

doubtedly refinements that would allow a somewhat better fit to the observed data.

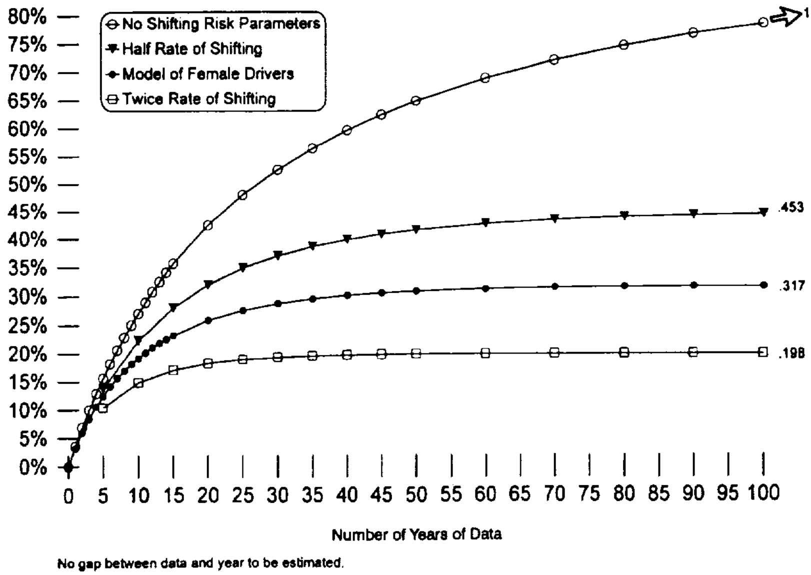
4.5. *Credibilities*

The covariance-variance structure for the Markov chain model fit to the data for the female drivers from California can be used together with Equations 2.8 to solve for the credibilities of different numbers of years of data. These credibilities have the same pattern as in Mahler [7] although the magnitudes are different. The latter appears to be due to a mistake in Mahler [7] in computing the credibilities.[57] In any case, note that the current method has the advantage that it does not require the dividing of the

[57] Unfortunately, it appears that a mistake was made in Mahler [7] in adopting the work in Mahler [10]. The step in Mahler [10] of dividing the variance into three pieces: within

FIGURE 14

SUM OF CREDIBILITIES
MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVER DATA



covariances (or variances) into separate pieces, some of which must be inferred rather than observed. The current method relies on the observable total variances and covariances.[58]

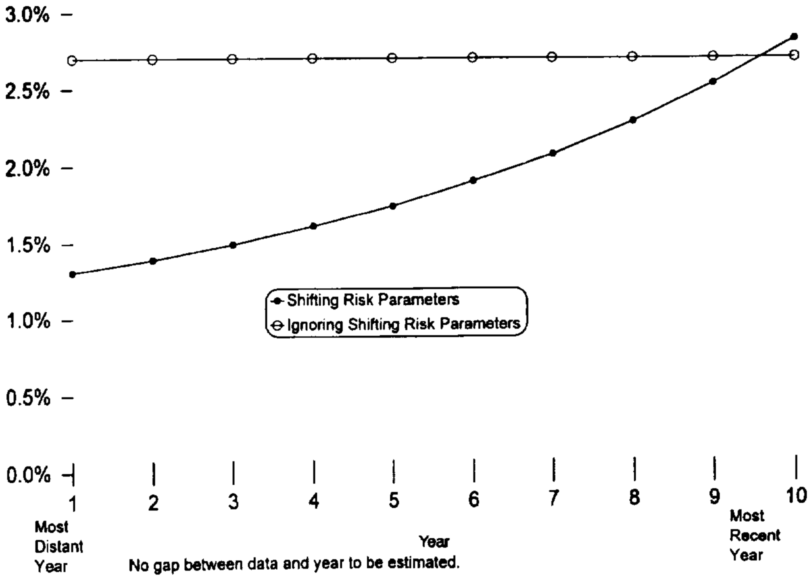
Table 4 displays the credibilities assigned to individual years as well as the sum of the credibilities. Figure 14 compares the sum of the credibilities for the Markov chain model to those

variance, between variance, and the variance due to shifting parameters over time, was not performed in Mahler [7]. This led to an inappropriate total covariance between years being used in the equations for credibility; these covariances were too big by an amount equal to the between variance.

[58]This difference from Mahler [10] is, to a large extent, a matter of presentation and emphasis. (See for example, *PCAS* LXXVII 1990, p. 297.)

FIGURE 15

CREDIBILITIES ASSIGNED TO EACH OF TEN YEARS OF DATA
MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVER DATA



that would result from ignoring shifting risk parameters.[59] With shifting risk parameters, the credibilities are lower.[60] As the number of years approaches infinity, the sum of the credibilities approaches 31.7% rather than 100%.[61] Also shown are credibil-

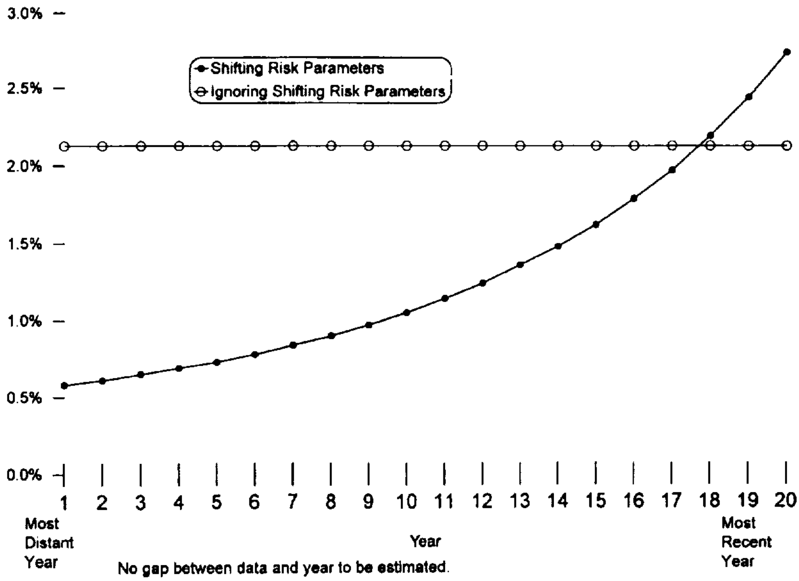
[59]In the absence of shifting risk parameters over time, one has the gamma-Poisson situation summarized in Exhibit 9. The credibility assigned to each of Y years is $1/(Y + \lambda)$ where λ is the scale parameter of the gamma distribution. In the present example λ was taken equal to 26.9. However, the discrete approximation in Section 4.1 produces an expected value of process variance of .037222, and variance of hypothetical mean frequencies of .0013765. Their ratio is a credibility parameter of 27.04. Therefore, in the absence of shifting risk parameters, each of Y years of data would be given a credibility of $1/(Y + 27.04)$ for a sum of credibilities of $Y/(Y + 27.04)$.

[60]The effect of shifting risk parameters in this case starts to have a significant impact after 10 or 15 years.

[61]While the model can be run for more than 50 years of data, it is unclear what the connection to reality is in this case.

FIGURE 16

CREDIBILITIES ASSIGNED TO EACH OF TWENTY YEARS OF DATA
MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVER DATA



ities for half this rate of shifting as well as twice this rate of shifting.

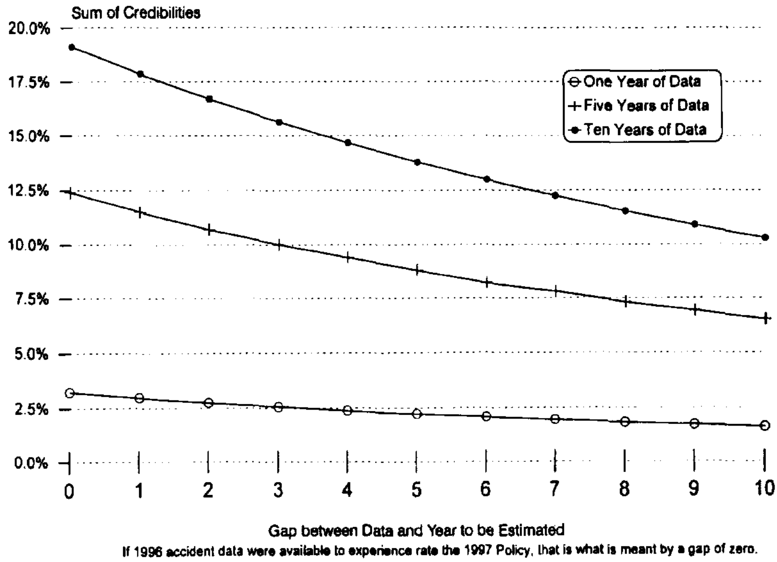
Figure 15 displays the individual credibilities for ten years of data. Figure 16 is similar, but for 20 years of data. In each case, the credibilities assigned to older years of data are significantly lower than those for more recent years of data. While the total credibility is less than in the absence of shifting risk parameters, the most recent year actually receives more credibility.[62]

Figure 17 displays the effect of delays in receiving data. Even in this situation with a relatively slow shifting of risk parameters, the effects of delays are noticeable.

[62]This is the same pattern as was displayed in the simple Poisson example.

FIGURE 17

EFFECTS OF DELAYS IN COLLECTING DATA ON SUM OF CREDIBILITIES, MARKOV CHAIN MODEL OF CALIFORNIA FEMALE DRIVER DATA



5. BASEBALL DATA

Mahler [10] examines the won-lost records of baseball teams. The Markov chain model developed here can be fit to this data.

There are two data sets, American League (AL) and National League (NL), each covering a 60 year period. As in Mahler [10], we will assume for simplicity 150 games per team per year, and convert the losing percentages to numbers of games lost. Table 5 displays the covariances between years of data separated by different amounts.[63] It is evident that the covariances decline as

[63]The separate observations of covariances were averaged. For example, there are 59 pairs of years separated by one year. There is considerable random fluctuation. For example, the covariances for the 59 pairs of years separated by one year for the AL data average to 138.7 with a standard deviation of 78.7.

TABLE 5

COVARIANCES VERSUS YEARS OF SEPARATION, BASEBALL DATA*

Number of Years Separation	American League	National League	Number of Years Separation	American League	National League
0	213.6	205.2	20	45.8	33.2
1	138.7	139.3	21	33.4	26.9
2	109.8	106.2	22	27.4	19.1
3	92.8	99.4	23	14.1	19.9
4	77.7	86.2	24	3.2	15.7
5	55.0	70.5	25	−2.7	4.7
6	45.3	65.2	26	4.0	1.2
7	33.5	53.1	27	3.6	−12.9
8	23.5	40.7	28	0.4	−18.4
9	12.1	30.7	29	−5.4	−11.4
10	15.4	23.8	30	3.4	−3.7
11	12.1	20.9	31	5.5	−6.8
12	9.9	25.2	32	9.4	−3.1
13	18.4	34.7	33	9.7	−8.9
14	17.6	37.1	34	28.3	−12.4
15	26.0	42.9	35	37.7	−2.3
16	36.1	52.2	36	32.6	−8.6
17	34.5	57.4	37	40.8	−16.6
18	42.9	47.2	38	53.4	−16.4
19	43.5	40.6	39	33.2	−7.9
			40	21.4	−33.2

*Covariances between number of games lost per team, based on observed losing percentage and assuming 150 games per team per year.

TABLE 6

CORRELATIONS VERSUS YEARS OF SEPARATION, BASEBALL DATA

Number of Years Separation	American League	National League	Number of Years Separation	American League	National League
0	1.000	1.000	20	0.225	0.136
1	0.633	0.651	21	0.159	0.090
2	0.513	0.498	22	0.125	0.065
3	0.438	0.448	23	0.093	0.055
4	0.360	0.386	24	0.048	0.004
5	0.265	0.312	25	0.006	−0.024
6	0.228	0.269	26	0.010	−0.028
7	0.157	0.221	27	−0.002	−0.095
8	0.124	0.190	28	−0.013	−0.128
9	0.078	0.135	29	−0.032	−0.107
10	0.090	0.100	30	0.006	−0.062
11	0.058	0.083	31	−0.019	−0.061
12	0.063	0.103	32	0.027	−0.028
13	0.101	0.154	33	0.002	−0.015
14	0.104	0.176	34	0.088	0.017
15	0.141	0.180	35	0.143	0.038
16	0.178	0.246	36	0.156	−0.014
17	0.166	0.278	37	0.214	−0.024
18	0.198	0.219	38	0.238	−0.012
19	0.219	0.176	39	0.138	−0.017
			40	0.093	−0.095

the years get further apart. As discussed in Mahler [10], these data display a relatively large impact of shifting risk parameters over time. Table 6 shows the similar pattern for the correlations.

Fitting an exponential regression to the covariances for separations of one to ten years, one obtains:

$$\text{NL}: \quad \text{Cov}[X_1, X_{1+g}] = \exp(5.156 - .185g),$$

$$\text{AL}: \quad \text{Cov}[X_1, X_{1+g}] = \exp(5.317 - .272g).$$

5.1. Markov Chain Model

To fit a Markov chain model to this data, one would want the log of the covariances to decline at a slope of about .23.

The first step in modeling the baseball data is to assume for simplicity that each team's number of games lost in a year is approximately binomial with parameters p and 150. The mean number of games lost, $p150$, will be assumed to have the following discrete distribution:[64]

Expected Number of Games Lost (μ)										
50	55	60	65	70	75	80	85	90	95	100
Probability (α)										
4%	6%	10%	11%	12%	14%	12%	11%	10%	6%	4%

Then using the technique of Appendix D, one can construct an 11×11 transition matrix that has the above α as a stationary distribution:[65]

[64]This simple distribution was chosen for illustrative purposes and is intended to approximate the observed spread of results. The chosen distribution has the desired mean of 75 and together with a binomial risk process would produce a total variance of about 207 compared to the observed total variance of about 209.

[65]The particular matrix was constructed in order to have about a $\frac{1}{2}$ chance of shifting either up or down one state per year.

$$\begin{pmatrix}
.7000 & .3000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2000 & .4875 & .3125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & .1875 & .5506 & .2619 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .2381 & .5010 & .2609 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & .2391 & .4916 & .2692 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .2308 & .5385 & .2308 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & .2692 & .4916 & .2391 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .2609 & .5010 & .2381 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .2619 & .5506 & .1875 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .3125 & .4875 & .2000 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .3000 & .7000
\end{pmatrix}$$

The eigenvalues and ζ vector are:

i	ζ_i	λ_i	i	ζ_i	λ_i
1	5625	1	7	0	.4292
2	169.6	.9670	8	.004	.2846
3	0	.9034	9	0	.1881
4	1.36	.8119	10	.007	.0966
5	0	.7154	11	0	.0330
6	.069	.5708			

$$\mathrm{Cov}[X_1, X_{1+g}] = \sum_{i>1} \zeta_i \lambda_i^g \approx 170(.967^g).$$

Therefore, the log of the covariances would decline at a slope of about .033. To match the baseball data, we desire a decline at a slope of about .23 or about 6 or 7 times as much. Therefore, this transition matrix raised to the 6th power should roughly match the baseball data.[66]

[66]This will correspond to about a 1 in 2,000 chance of a team moving up or down 6 states ($\pm .2$ in expected losing percentage) in a single year.

FIGURE 18

COVARIANCES VERSUS YEARS OF SEPARATION, BASEBALL DATA VS. MARKOV SHIFTING PARAMETERS OVER TIME

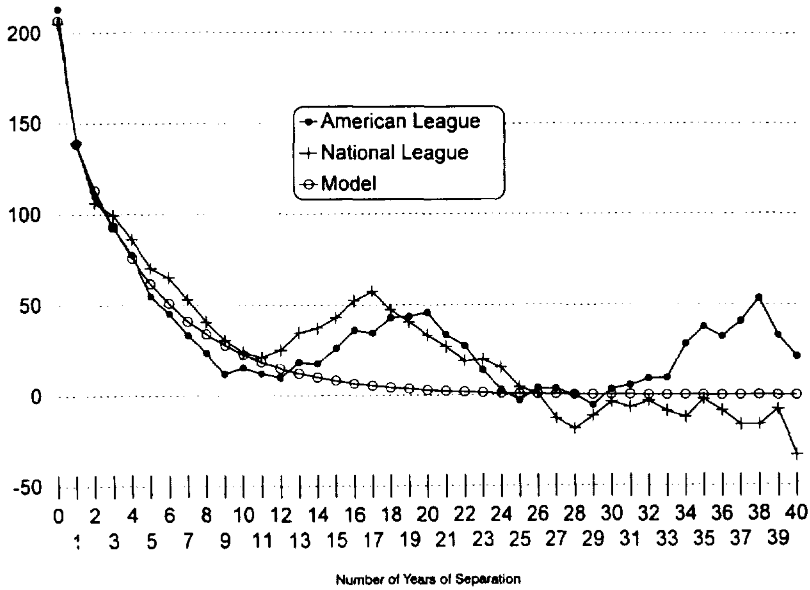


Figure 18 compares the covariances observed for the baseball data and those for the Markov chain model. There is an overall reasonable fit. There are higher covariances than would be predicted by the model for separations of about 15 to 23 years. This may be due to some long term cycle in the data, but in any case is beyond the scope of this paper.[67]

[67] Some factors which remained relatively stable over this 60 year period of time might lead to a tendency for an individual team's expected losing percentage to revert to a long term average different than the overall average of .5. The Markov chain model does not capture any such behavior. Rather, it assumes that given sufficiently long time periods, the average for each risk will be the same. Yet in Section 4.1 of Mahler [10], it is demonstrated that over the 60 year data period, the teams are significantly different. Thus while, as will be shown below, the estimated credibilities are reasonable, the Markov chain model is far from a complete description of the risk process that produced this baseball data.

TABLE 7

CREDIBILITY
BASEBALL DATA MARKOV CHAIN MODEL
(No Delay in Receiving Data)

Years Between Data and Estimate	Number of Years of Data Used					
	1	2	3	4	5	10
1 (Most Recent)	67.0%	55.1%	54.3%	54.2%	54.2%	54.2%
2		17.7%	15.0%	14.8%	14.8%	14.8%
3			4.9%	4.2%	4.1%	4.1%
4				1.4%	1.2%	1.2%
5					.4%	.3%
6						.1%
7						0
8						0
8						0
10						0
Total Credibility	67.0%	72.8%	74.2%	74.6%	74.7%	74.7%

5.2. *Credibilities*

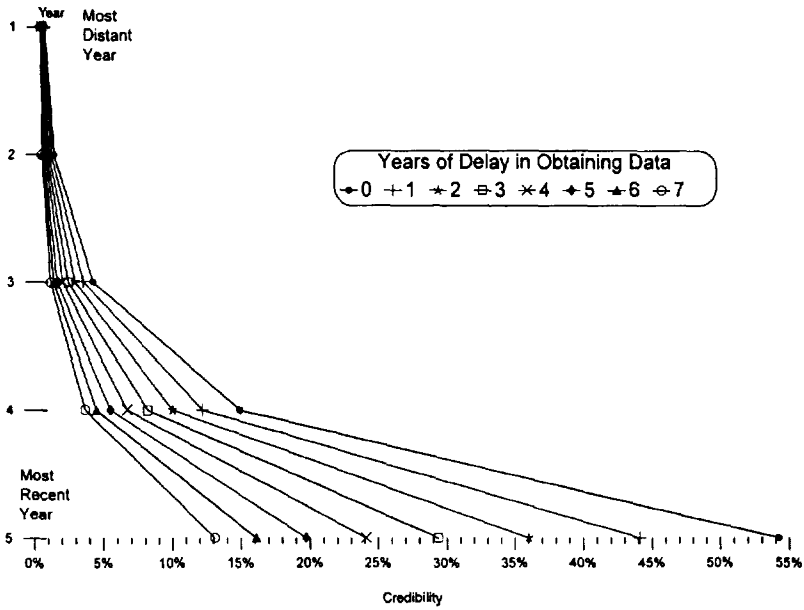
The covariances calculated in the Markov chain model can be used to calculate the credibilities to be assigned to individual years of data. Table 7 displays these credibilities, assuming no delay in receiving the information.

The sum of the credibilities[68] quickly reaches a limit of 74.7% as the number of years of data is increased. Due to the quickly shifting risk parameters over time, the amount of credibility assigned to more distant years of data is small. The credibilities in Table 6 are generally similar to those in Table 16 of Mahler [10]. However, due to the structure imposed by the Markov chain model, the credibilities in Table 6 have a more reasonable pattern when looked at in detail. The credibilities are all between 0 and

[68]The complement of credibility is given to the grand mean.

FIGURE 19

CREDIBILITIES ASSIGNED TO EACH OF FIVE YEARS OF DATA
MARKOV CHAIN MODEL OF BASEBALL DATA



1. They decrease for years more distant in time. The credibility assigned to any individual year of data declines as more years are added. The sum of the credibilities increases smoothly as years of data are added.

Figure 19 displays the credibilities assigned to five separate years of data for various delays in obtaining information. Due to the quickly changing risk parameters, the effect of any delay in obtaining data is significant. As the delay increases, the credibility assigned to any individual year decreases. The smooth pattern shown in Figure 19 demonstrates the effect of the structure imposed by the Markov chain model.

FIGURE 20

CREDIBILITIES ASSIGNED TO EACH OF TEN YEARS OF DATA VARYING THE RATE OF SHIFTING PARAMETERS IN THE MARKOV CHAIN MODEL OF BASEBALL DATA

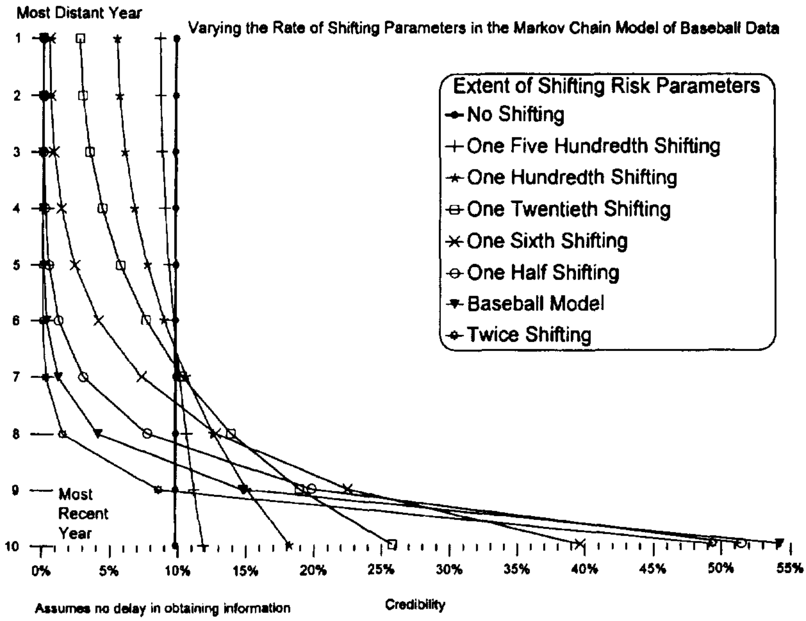


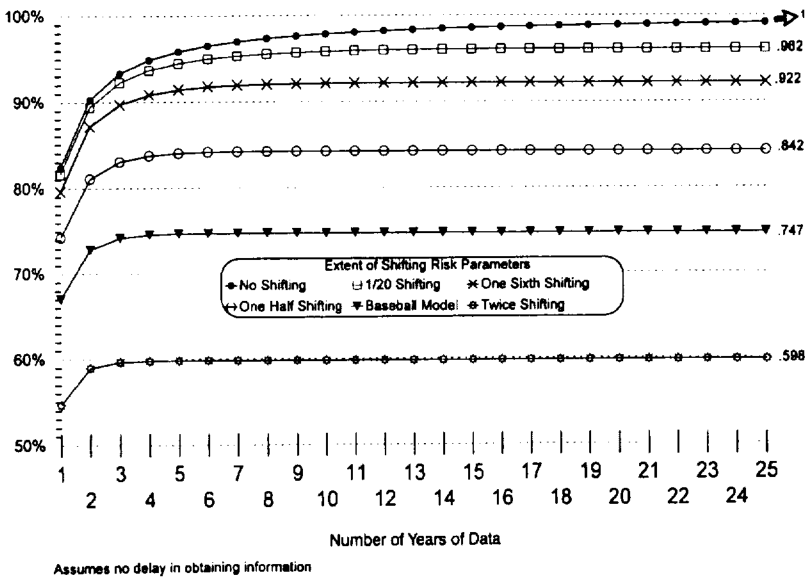
Figure 20 compares the credibilities one would assign to ten individual years of data with either more or less quickly shifting risk parameters than in the baseball data. If a major change in circumstances leads one to believe there has been a significant change in the rate at which parameters shift,[69] then the Markov chain model can be easily adjusted to incorporate one's estimate of the rate at which parameters will shift in the future.

---

[69]In the baseball example, many changes have occurred since 1960, the last year used to calibrate the model. For example, free agency might allow a more frequent movement of players between teams leading to a somewhat quicker rate of shifting of risk parameters.

FIGURE 21

SUM OF CREDIBILITIES, VARYING THE RATE OF SHIFTING PARAMETERS IN THE MARKOV CHAIN MODEL OF BASEBALL DATA



In the absence of shifting risk parameters, the same credibility is assigned to each of the ten years of data. As more and more shifting is introduced, the credibilities for the older years decline. (The curve on Figure 20 gets further and further from a vertical line.) This illustrates the effect of fine tuning the rate of shifting in the Markov chain model.

Figure 21 compares the sums of the credibilities for various numbers of years of data for various amounts of shifting. In the absence of shifting risk parameters, the sum of the credibilities approaches unity as the number of years increases. As the amount of shifting increases, the limit of the sum of the credibilities decreases.

If parameters are shifted at one-hundredth of the rate at which parameters shifted in the baseball example, the maximum sum of the credibilities is 98.4%. For the baseball example, it is 74.7%. For twice the shifting, it is 59.8%. The greater the rate of shifting of risk parameters, the lower the limit and the faster the convergence.

6. AREAS FOR POSSIBLE FUTURE REFINEMENTS

The model presented here was applied to claim frequency situations. It would probably be valuable to extend this to situations involving claim severity or pure premiums.

The model presented here did not fully explore the impact of size of risk. In order to properly explore the impact of size of risk on insurance situations, one would probably have to incorporate the effects of parameter uncertainty and risk heterogeneity as well as shifting risk parameters over time.[70]

The model presented here does not allow for an expected long term difference between risks. Averaged over a sufficiently long period of time, every risk's average frequency is the same. This is undoubtedly a poor model of certain situations.

There is no specific treatment of the entry of new insureds or the exit of current insureds from the database. Venezian [15] specifically models the change in accident propensity of new drivers entering the system as they gain experience and get older. The model as presented here would not accommodate this phenomenon.

Thus while the model presented here is practical and flexible, it would require further work to adapt it to many situations of potential interest.

---

[70] See for example, Mahler [8].

7. SUMMARY

The effects of shifts over time in the risk process of an insured can be quantified in the covariances between years of data. For this Markov chain model, in most cases the covariances can be approximated by[71]:

$$\mathrm{Cov}[X_i,X_j] = \tau^2\lambda^{|i-j|} + \delta_{ij}\eta^2$$
$$\text{where} \quad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i \neq j \end{cases} \tag{7.1}$$

η^2 is the expected value of the process variance,

τ^2 is the variance of the hypothetical means,

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

One has $\mathrm{Var}[X] = \mathrm{Cov}[X,X] = \tau^2 + \eta^2$. This is the usual relationship that the total variance can be split into the expected value of the process variance and the variance of the hypothetical means.

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

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.

Four examples have been considered, involving dice, a mixture of four Poisson distributions, California driving data (modeled by a gamma-Poisson), and baseball data (modeled by a mixture of binomials). The Markov chain model was applied to each of these situations. The resulting values of λ

[71] This is Equation 3.5.

were:

Example	λ	"Half-Life"
Dice[72]	.769	2.6 trials
Mixture of 4 Poissons[73]	.855	4.4 years
Female California Drivers[74]	.961	17.3 years
Baseball Team Results[75]	.818	3.4 years

where the "half-life" is the length of time for the correlations between years to decline by a factor of one-half:

$$\lambda^{\text{half-life}} = .5$$
$$\text{half-life} = \frac{\ln .5}{\ln \lambda} = \frac{-.693}{\ln \lambda}. \qquad (7.2)$$

The longer the half-life, the slower the rate of shifting parameters over time. Thus, the impact of shifting parameters was most significant in the dice example, followed by the baseball data and the mixture of four Poissons example.[76] The female California drivers data with a half-life of about 17 years has much less impact from shifting risk parameters.[77]

If the Markov chain model holds, the correlations between different years of data should decline approximately exponentially. For $i \neq j$, Equation 7.1 gives $\text{Cov}[X_i, X_j] = \tau^2 \lambda^{|i-j|}$.

[72] See Section 2.7.

[73] See Section 3.2.

[74] The dominant eigenvalue shown in Section 4.3 is .998. This transition matrix is then taken to the 20th power, therefore so are the eigenvalues. $(.998)^{20} = .961$.

[75] The dominant eigenvalue shown in Section 5.1 is .967. However, this transition matrix is taken to the 6th power, therefore so are all the eigenvalues. $(.967)^6 = .818$.

[76] The dice example and mixture of four Poissons example were specifically designed to have a significant effect of shifting risk parameters for illustrative purposes. One of the reasons the baseball data was selected for presentation was because it showed a significant impact.

[77] The male driving data displayed even less impact from shifting risk parameters than female driving data. See Mahler [7].

Also, $\mathrm{Var}[X_i] = \mathrm{Var}[X_j] = \eta^2 + \tau^2$. Therefore,

$$\begin{aligned} \mathrm{Corr}[X_i, X_j] &= \left(\frac{\tau^2}{\tau^2 + \eta^2}\right)\lambda^{|i-j|} \\ \ln \mathrm{Corr}[X_i, X_j] &= \ln\left(\frac{\tau^2}{\tau^2 + \eta^2}\right) + |i-j| \ln \lambda \qquad i \neq j. \end{aligned} \tag{7.3}$$

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

Thus given a data set, one can determine whether this (simple) Markov chain model might be appropriate. One determines whether the log correlations as a function of the separation between years (not including zero separation) can be approximated by a straight line.[78] Then one can estimate the parameter λ and the ratio $\tau^2/(\tau^2 + \eta^2)$ from the slope and intercept of the fitted straight line.

These estimates can be used in turn to estimate credibilities. If one has data from years $1, 2, \ldots, Y$ and is estimating year $Y + \Delta$, then the least squares credibilities are given by solving the Y linear equations in Y unknowns:[79]

$$\sum_{j=1}^{Y} \mathrm{Cov}[X_i, X_j] Z_i = \mathrm{Cov}[X_i, X_{Y+\Delta}], \qquad i = 1, 2, \ldots, Y. \tag{7.4}$$

[78]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.

[79]See Equations 2.8.

8. CONCLUSIONS

A Markov chain model has been developed and applied to a number of different examples in which risk parameters shift over time. The model is sufficiently flexible to be applied to other situations.

In each case, the Markov chain model was used to explore the effects of shifting risk parameters over time. Covariances are calculated. Based on the Markov chain model, when shifting risk parameters over time are significant, the logs of the covariances between years of data are expected to decline linearly as the separation between years increases.

Credibilities are calculated from the variances and covariances. When shifting risk parameters are significant, older years receive less credibility and as more years of data are added, the sum of the credibilities goes to a limit less than one. The longer the delay in collecting data, the lower the credibilities.

The Markov chain model can be used to simulate the claims process when there are shifting risk parameters over time in the same manner as the gamma-Poisson and similar models can be used in the absence of shifting parameters. The Markov chain model should aid the actuary's understanding of situations in which shifting risk parameters are significant. It is both practical and sufficiently flexible to be applied in a wide variety of circumstances.

REFERENCES

[1] Dropkin, Lester B., "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* LXXIV, 1987, p. 391. (Reprinted from *PCAS* XLVI, 1959.) Note discussions by Robert A. Bailey, Sholom Feldblum, Stephen W. Philbrick, and Chris Svendsgaard.

[2] Feller, William, *An Introduction to Probability Theory and Its Applications*, New York, Wiley, 1968.

[3] Herzog, Thomas N., *An Introduction to Credibility*, ACTEX, 1994.

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

[5] Hewitt, Charles C., Jr., Discussion of Mayerson: "A Bayesian View of Credibility," *PCAS* LII, 1965, p. 121.

[6] Hossack, I. B., J. H. Pollard, and B. Zehnwirth, *Introductory Statistics with Applications to General Insurance*, New York, Cambridge University Press, 1983.

[7] Mahler, Howard C., "The Credibility of a Single Private Passenger Driver," *PCAS* LXXVIII, 1991, p. 146.

[8] Mahler, Howard C., Discussion of Gillam: "Parametrizing the Workers' Compensation Experience Rating Plan," *PCAS* LXXX, 1993, p. 148.

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

[10] Mahler, Howard C., "An Example of Credibility and Shifting Risk Parameters," *PCAS* LXXVII, 1990, p. 225.

[11] Mayerson, Allen L., "A Bayesian View of Credibility," *PCAS* LI, 1964, p. 85.

[12] Philbrick, Stephen W., "An Examination of Credibility Concepts," *PCAS* LXVIII, 1981, p. 195.

[13] Resnick, Sidney I., *Adventures in Stochastic Processes*, Boston, Birkhauser, 1992.

[14] Venezian, Emilio C., "The Distribution of Automobile Accidents—Are Relativities Stable Over Time?" *PCAS* LXXVII, 1990, p. 309.

[15] Venezian, Emilio C., "Good and Bad Drivers—A Markov Model of Accident Proneness," *PCAS* LXVIII, 1981, p. 65.

APPENDIX A

MARKOV CHAINS[80]

Assume each year[81] an individual is in a "state." In this paper, each state corresponds to a different average claim frequency. In this paper, 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 his past history was, or what year it is. The transition probability from state i to state j, P_{ij}, is dependent on only the two states, i and j.

Arrange these transition probabilities P_{ij} into a matrix $\boldsymbol{P}$. This transition matrix $\boldsymbol{P}$, together with the definition of the states, defines a (finite dimensional) Markov chain.

If an individual is in state i, P_{ii} is the probability that he remains in state i. $1 - P_{ii}$ is the probability that he changes his state.

$\sum_{j=1}^{n} P_{ij}$ is the sum of the probability of this individual changing to each of the possible states (including remaining in state i). Since all the possibilities are exhausted, $\sum_{j=1}^{n} P_{ij} = 1$. Each of the rows of the transition matrix $\boldsymbol{P}$ for a Markov chain must sum to unity.

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 $\beta\boldsymbol{P}$, where $\beta\boldsymbol{P}$ is the matrix product of the (row vector) distribution β and the transition matrix $\boldsymbol{P}$. The expected distribution in year 3 is $(\beta\boldsymbol{P})\boldsymbol{P} = \beta(\boldsymbol{P}\boldsymbol{P}) = \beta\boldsymbol{P}^2$. The expected distribution in year $1 + g$ is $\beta\boldsymbol{P}^g$.

A stationary distribution is a vector α such that $\alpha\boldsymbol{P} = \alpha$. On an expected basis, the portfolio of risks stays in the stationary

[80]See Feller [2] and Resnick [13].

[81]Although in this paper the time interval is a year, in general, it can be anything.

distribution over time. Note that, from its definition, a stationary distribution (if it exists) is an eigenvector corresponding to an eigenvalue of unity. When it exists, the quickest way to compute α, the stationary distribution of $\boldsymbol{P}$, is via:

$$\alpha = (1,1,\ldots,1)(\boldsymbol{I} - \boldsymbol{P} + \mathbf{ONE})^{-1}$$

where $\boldsymbol{I}$ is the identity matrix and **ONE** is the n by n matrix, all of whose entries are one.[82]

All of the Markov chains in this paper have been specifically constructed to have a stationary distribution using the techniques in Appendix D.

For a finite dimensional Markov chain such that each state can be reached from every other state and such that no states are periodic,[83] a unique stationary distribution α exists and for any initial distribution β, $\beta \boldsymbol{P}^g \to \alpha$ as $g \to \infty$. Thus eventually the distribution of risks in the portfolio is α (for all practical purposes) regardless of the initial distribution β.

Taking $\beta = (1,0,0\cdots)$, $\beta = (0,1,0\cdots)$, etc., implies that the rows of $\boldsymbol{P}^g \to \alpha$ as $g \to \infty$. If A is a matrix all of whose rows are the stationary distribution α, $\boldsymbol{P}^g \to A$ as $g \to \infty$.

Let $\boldsymbol{P}^T$ be the matrix transpose of $\boldsymbol{P}$. Let Λ be the diagonal matrix with entries equal to the eigenvalues of $\boldsymbol{P}^T$. Let V^T be the matrix, each of whose columns are the eigenvectors of $\boldsymbol{P}^T$. (V has as its rows the eigenvectors of $\boldsymbol{P}^T$.) Then, as stated in Appendix B, $(V^T)^{-1}\boldsymbol{P}^T V^T = \Lambda$. Taking the transpose of both sides of this equation and noting that $\Lambda^T = \Lambda$, since Λ is symmetric: $V\boldsymbol{P}V^{-1} = \Lambda$. So the matrix V can be used to diagonalize the transition matrix $\boldsymbol{P}$:

$$V^{-1}\Lambda^2 V = V^{-1}(V\boldsymbol{P}V^{-1})^2 V = V^{-1}V\boldsymbol{P}V^{-1}V\boldsymbol{P}V^{-1}V = \boldsymbol{P}^2.$$

---

[82] See Section 2.14 of Resnick [13].

[83] See Section 2.13 of Resnick [13].

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.

If A is matrix whose rows are the stationary distribution α, then:

$$P^g \to A.$$

$$\therefore \quad \Lambda^g = VP^g V^{-1} \to VAV^{-1}.$$

But Λ^g has diagonal element λ_i^g. These only converge to a limit as $g \to \infty$ if $\lambda_i = 1$ or $|\lambda_1| < 1$. Let $\lambda_1 = 1$, since the order of eigenvalues is arbitrary. Then $|\lambda_1| < 1$ for $i > 1$ (ignoring the very unusual situation where $\lambda = 1$ is a multiple root of the characteristic equation).

Let the limit of Λ^g as $g \to \infty$ be denoted by Λ^∞. Then $(\Lambda^\infty)_{ij} = 0$ for $i \neq 1$ or $j \neq 1$, and $(\Lambda^\infty)_{1,1} = 1$.

Therefore $V^{-1}\Lambda^\infty V = \lim_{g\to\infty} P^g = A$.

Thus $(V^{-1})_{i1} V_{1j} = A_{ij} = \alpha(j)$, since the rows of A are the stationary distribution, and $(\Lambda^\infty)_{ij} = 0$ for $i \neq 1$ or $j \neq 1$.

Thus $(V^{-1})_{i1} = \alpha(j)/V_{1j}$.

Note that the left hand side is independent of j, while the right hand is independent of i. Since the equation holds for all i and j, both sides must be independent of i and j. Therefore, the elements of the first column of V^{-1} are all equal. The elements of the first row of V are proportional to the stationary distribution α.

Since $VV^{-1} = I$, where I is the identity matrix, the product of the first column of V^{-1} with any row of V other than the first is zero. But the product of the first column of V^{-1} with any row is proportional to the sum of that row since all the elements of the first column of V^{-1} are equal. Therefore, the sum of any row of V other than the first is zero. Therefore, the sum of any eigenvector other than the first is zero, since the rows of V are the eigenvectors of P^T.

APPENDIX B

EIGENVALUES AND EIGENVECTORS

Given a square matrix $\boldsymbol{P}$, if for a vector, $\boldsymbol{v}$, $\boldsymbol{Pv} = \lambda \boldsymbol{v}$, then $\boldsymbol{v}$ is called an eigenvector of $\boldsymbol{P}$ with eigenvalue λ. Note that if $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{P}$, so is $\boldsymbol{v}$ times any non-zero constant. So eigenvectors can be determined only up to a proportionality constant.

One can find the eigenvalues and thus the corresponding eigenvectors by solving the characteristic equation:[84]

Determinant $(\boldsymbol{P} - \lambda \boldsymbol{I}) = 0$, where $\boldsymbol{I}$ is the identity matrix.

If V is a matrix whose columns are the eigenvectors of $\boldsymbol{P}$, then $V^{-1}\boldsymbol{P}V$ is a diagonal matrix Λ whose elements are the eigenvalues of $\boldsymbol{P}$. ($V^{-1}\boldsymbol{P}V = \Lambda$ follows from the matrix equation $\boldsymbol{P}V = V\Lambda$, which when we take each column reduces to the eigenvalue equation: $\boldsymbol{P}\boldsymbol{v}_i = \boldsymbol{v}_i\lambda_i$.)

If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue λ, then $\boldsymbol{Pv} = \lambda \boldsymbol{v}$. Therefore,

$$\boldsymbol{P}^2\boldsymbol{v} = \boldsymbol{P}(\boldsymbol{Pv}) = \boldsymbol{P}(\lambda \mathrm{v}) = \lambda \boldsymbol{Pv} = \lambda^2 \boldsymbol{v}.$$

Thus $\boldsymbol{v}$ is also an eigenvector of $\boldsymbol{P}^2$ with eigenvalue λ^2. In general, $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{P}^g$ with eigenvalue λ^g. Raising a matrix to a power does not alter the eigenvectors and raises the eigenvalues to the same power.

[84]Eigenvalues and eigenvectors are calculated by many computer software packages. The author used the APL program EIG provided by Manugistics (formerly STSC).

APPENDIX C

MATRIX EQUATIONS FOR LEAST SQUARES CREDIBILITY[85]

In this Appendix, Equations 2.8 in the main text are derived by minimizing the squared error. The result is N linear equations for the credibilities to be assigned to each of N years of data. Thus, the credibilities can be solved for in terms of the covariance structure.

Let

$$C_{ij} = \mathrm{Cov}[X_i, X_j]$$
$$= \text{Covariance of year } X_i \text{ and year } X_j, \qquad \text{and}$$
$$C_{ii} = \text{Variance of year } X_i.$$

Let Z_i be the credibility assigned to year X_i. We wish to predict year $X_{N+\Delta}$ using N years of data $X_1, X_2, \ldots, X_N$ and the grand mean M. Let $Z_0 = 1 - \sum_{i=1}^{N} Z_i$ = complement of credibility.

Then the estimate is:

$$F = \sum_{i=1}^{N} Z_i X_i + Z_0 M.$$

Let

$$X_0 = M, \qquad \text{then} \quad F = \sum_{i=0}^{N} Z_i X_i$$

$$F - X_{N+\Delta} = \left(\sum_{i=0}^{N} Z_i X_i\right) - X_{N+\Delta} = \sum_{i=0}^{N} Z_i (X_i - X_{N+\Delta}),$$

since $\sum_{i=0}^{N} Z_i = 1$.

---

[85] The derivation is adopted from that in Mahler [10].

Therefore,

$$(F - X_{N+\Delta})^2 = \left(\sum_{i=0}^{N} Z_i(X_i - X_{N+\Delta})\right)\left(\sum_{j=0}^{N} Z_j(X_j - X_{N+\Delta})\right)$$

$$= \sum_{i=0}^{N}\sum_{j=0}^{N} Z_i Z_j (X_i - X_{N+\Delta})(X_j - X_{N+\Delta}).$$

Then the expected value of the squared difference between the estimate F and $X_{N+\Delta}$ is, as a function of the credibilities Z:

$$V(Z) = \mathrm{E}[(F - X_{N+\Delta})^2]$$

$$= \sum_{i=0}^{N}\sum_{j=0}^{N} Z_i Z_j \mathrm{E}[(X_i - X_{N+\Delta})(X_j - X_{N+\Delta})].$$

Now

$$\mathrm{E}[(X_i - X_{N+\Delta})(X_j - X_{N+\Delta})] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i X_{N+\Delta}]$$

$$- \mathrm{E}[X_j X_{N+\Delta}] - \mathrm{E}[X^2_{N+\Delta}]$$

$$\mathrm{E}[X_i X_j] = \mathrm{Cov}[X_i, X_j] + \mathrm{E}[X_i]\mathrm{E}[X_j]$$

$$= C_{ij} + M^2,$$

where

$$C_{0j} = \mathrm{Cov}[M, X_j] = 0.$$

Thus

$$\mathrm{E}[(X_i - X_{N+\Delta})(X_j - X_{N+\Delta})]$$

$$= C_{ij} - C_{i,N+\Delta} - C_{j,N+\Delta} + C_{N+\Delta,N+\Delta}, \qquad \text{and}$$

$$V(Z) = \sum_{i=0}^{N}\sum_{j=0}^{N} Z_i Z_j \{C_{ij} - C_{i,N+\Delta} - C_{j,N+\Delta} + C_{N+\Delta,N+\Delta}\}.$$

$$V(Z) = \sum_{i=0}^{N}\sum_{j=0}^{N} Z_i Z_j C_{ij} - \left(\sum_{i=0}^{N} C_{i,N+\Delta} Z_i\right)\left(\sum_{j=0}^{N} Z_j\right)$$

$$- \left(\sum_{j=0}^{N} C_{j,N+\Delta} Z_j\right)\left(\sum_{i=0}^{N} Z_i\right)$$

$$+ C_{N+\Delta,N+\Delta}\left(\sum_{i=0}^{N} Z_i\right)\left(\sum_{j=0}^{N} Z_j\right).$$

The last three terms all simplify, since

$$\sum_{i=0}^{N} Z_i = Z_0 + \sum_{i=1}^{N} Z_i$$

$$= 1 - \sum_{i=1}^{N} Z_i + \sum_{i=1}^{N} Z_i = 1.$$

Therefore,

$$V(Z) = \sum_{i=0}^{N}\sum_{j=0}^{N} Z_i Z_j C_{ij} - \sum_{i=0}^{N} C_{i,N+\Delta} Z_i$$

$$- \sum_{j=0}^{N} C_{j,N+\Delta} Z_j + C_{N+\Delta,N+\Delta}.$$

Also, since $C_{0j} = 0 = C_{i0}$, the elements involving $i = 0$ or $j = 0$ drop out, leaving

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

Taking the partial derivative of $V(Z)$ with respect to Z_k and setting it equal to zero:

$$2\sum_{i=1}^{N} Z_i C_{ik} - 2C_{k,N+\Delta} = 0$$

$$\sum_{i=1}^{N} Z_i C_{ik} = C_{k,N+\Delta}.$$

This results in N equations in N unknowns for $k = 1,2,\ldots,N$. These are Equations 2.8 in the main text.

APPENDIX D

CONSTRUCTING A TRANSITION MATRIX

Assume we are given a set of probabilities corresponding to a set of n states:

$$\alpha_i, \qquad i = 1,2,\ldots,n \qquad \sum \alpha_i = 1.$$

There are many transition matrixes, $\boldsymbol{P}$, such that α is a stationary state, $\alpha\boldsymbol{P} = \alpha$. A method of constructing one such matrix from α will be shown.

The constructed transition matrix will be such that most of its elements are zero. The only non-zero elements will be on the main diagonal, just above the main diagonal or just below the main diagonal. Such a matrix is sometimes referred to as "tri-diagonal."

Such a transition matrix corresponds to each year, an insured either staying in the same state or possibly moving up or down by a single state in a single year.[86]

As a concrete example, take the simple Poisson example in the main text, with four states and $\alpha = (.4,.3,.2,.1)$.

The equation $\alpha\boldsymbol{P} = \alpha$ becomes

$$(.4 \quad .3 \quad .2 \quad .1)\begin{pmatrix} P_{11} & P_{12} & 0 & 0 \\ P_{21} & P_{22} & P_{23} & 0 \\ 0 & P_{32} & P_{33} & P_{34} \\ 0 & 0 & P_{43} & P_{44} \end{pmatrix} = (.4 \quad .3 \quad .2 \quad .1).$$

[86]If this transition matrix were raised to a power, then one could move more than a single state per year. Also, the overall speed of parameter shifting would be increased.

So,

$$
\begin{aligned}
.4P_{11} + .3P_{21} &= .4 \\
.4P_{12} + .3P_{22} + .2P_{32} &= .3 \\
.3P_{23} + .2P_{33} + .1P_{43} &= .2 \\
.2P_{34} + .1P_{44} &= .1.
\end{aligned}
$$

In addition, each row of any transition matrix sums to unity. (Every risk ends up in some state.)

$$
\begin{aligned}
P_{11} + P_{12} &= 1 \\
P_{21} + P_{22} + P_{23} &= 1 \\
P_{32} + P_{33} + P_{34} &= 1 \\
P_{43} + P_{44} &= 1.
\end{aligned}
$$

Thus

$$P_{21} = \frac{.4(1 - P_{11})}{.3} = \frac{4}{3}P_{12}.$$

Similarly

$$
\begin{aligned}
P_{32} &= \frac{.3(1 - P_{22}) - .4P_{12}}{.2} \\
&= \tfrac{3}{2}(P_{21} - P_{23}) - 2P_{12} \\
&= \tfrac{3}{2}P_{23} + 2P_{12} - 2P_{12} = \tfrac{3}{2}P_{23}.
\end{aligned}
$$

Similarly, one gets

$$P_{43} = (\tfrac{2}{1})P_{34}.$$

In general, we need:

$$\alpha_i P_{i,i+1} = \alpha_{i+1} P_{i+1,i}.$$

The left hand side of this equation is the probability of being in state i times the probability of going from state i to state $i + 1$. Thus, this is the expected number of transitions from state i to state $i + 1$. Similarly, the right hand side is the expected number of transitions from state $i + 1$ to state i. In this case, these expected numbers of transitions will cancel and on average will result in no net change.

There are still arbitrary scale factors. (Within large bounds one can pick P_{12} and then P_{21} follows.) For purposes of illustration,[87] let $P_{i,i+1} + P_{i+1,i} = v < 1$ for all i, where v is a parameter that controls the amount of shifting. It represents the approximate probability of shifting either up or down one state; $1 - v$ is the approximate probability of remaining in the same state.

Then once $0 < v < 1$ is chosen, one constructs the transition matrix:

$$P_{i,i+1} = \frac{\alpha_{i+1}}{\alpha_i + \alpha_{i+1}} v,$$

$$P_{i+1,i} = \frac{\alpha_i}{\alpha_i + \alpha_{i+1}} v,$$

$$P_{ii} = 1 - P_{i,i-1} - P_{i,i+1}, \qquad \text{and}$$

$$P_{ij} = 0 \qquad \text{for} \quad |i - j| > 1.$$

This results in a transition matrix with the given α as a stationary distribution and with about a $(1 - v)$ chance of remaining in the same state per year.

This construction algorithm is relatively simple and easily programmable.

In the particular example with $\alpha = (.4,.3,.2,.1)$, taking $v = .42$, the algorithm produces a transition matrix of:

$$\begin{pmatrix} .820 & .180 & 0 & 0 \\ .240 & .592 & .168 & 0 \\ 0 & .252 & .608 & .140 \\ 0 & 0 & .280 & .720 \end{pmatrix}.$$

This is the transition matrix shown in the main text, which has a stationary distribution of $(.4,.3,.2,.1)$.

[87] For certain applications, one may choose to vary the probability of remaining in a state among the different states.

APPENDIX E

COVARIANCES

Let μ_i be the mean for state i. Let X and U be two different years of data separated by g years, $g > 0$. Let X have probability vector $\boldsymbol{\beta}$ for the probability of being in a given state. Then $\boldsymbol{\beta P^g}$ is the probability vector for year U. Then

$$\begin{aligned}\mathrm{E}[XU] = \sum_{i,j} & \Pr(X \text{ in state } i \text{ and } U \text{ in state } j)\\ & \times \mathrm{E}[XU \mid X \text{ in state } i \text{ and } U \text{ in state } j].\end{aligned}$$

If X is in state i and U is in state j, then

$$\mathrm{E}[XU] = \mathrm{E}[X \mid X \text{ in state } i]\mathrm{E}[U \mid U \text{ in state } j] = \mu_i\mu_j,$$

since the die rolls in year X and U are independent.

$$\begin{aligned}&\Pr(X \text{ in state } i \text{ and } U \text{ in state } j)\\ &\quad = \Pr(X \text{ in state } i)\Pr(U \text{ in state } j \mid X \text{ in state } i)\\ &\quad = \beta_i(P^g)_{ij},\end{aligned}$$

since the transition matrix from X to U which are g years apart is $\boldsymbol{P}^g$. Thus

$$\begin{aligned}\mathrm{E}[XU] &= \sum_{ij} \mu_i\mu_j\beta_i(P^g)_{ij}\\ &= (\mu \times \boldsymbol{\beta})^T \boldsymbol{P}^g \mu,\end{aligned}$$

where $\mu \times \boldsymbol{\beta}$ is the vector whose ith element is $\mu_i\beta_i$ and we have taken the matrix product of the transpose of this vector with the matrix $\boldsymbol{P}^g$ and then with the (column) vector μ, so,

$$\boldsymbol{P}^g = \boldsymbol{V}^{-1}\Lambda^g\boldsymbol{V}, \qquad \text{and}$$

$$\therefore\ \mathrm{E}[XU] = (\mu \times \boldsymbol{\beta})^T\boldsymbol{V}^{-1}\Lambda^g\boldsymbol{V}\mu.$$

Now assume that we are in the stationary distribution α; i.e., either the process has been going on long enough that the initial state no longer has any practical importance or the initial distribution was chosen to be equal to α. If $\beta = \alpha$, then

$$\mathrm{E}[XU] = (\mu \times \alpha)^T V^{-1} \Lambda^g V\mu.$$

Let C be the vector given by $(\mu \times \alpha)^T V^{-1}$ and let D be the vector given by $V\mu$, then since g is diagonal with $\Lambda^g_{ii} = \lambda^g_i$; $\mathrm{E}[XU] = \sum_k C_k D_k \lambda^g_{kk}$.

Thus we have written $\mathrm{E}[XU]$ as a sum of coefficients (independent of g) times the eigenvalues raised to the power g.

In the dice example in Section 2:

$\mu = (2.5, 3.5, 4.5) =$ means

$\alpha = (.25, .50, .25) =$ stationary distribution

$$V = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -.314 & -.686 \\ 1 & -3.186 & 2.186 \end{pmatrix} = \text{matrix whose rows are eigenvectors of the transpose of the transition matrix}$$

$$V^{-1} = \begin{pmatrix} .250 & .658 & .092 \\ .250 & -.103 & -.147 \\ .250 & -.451 & .201 \end{pmatrix}$$

$$C = (2.5 \times .25, 3.5 \times .5, 4.5 \times .25) \begin{pmatrix} .250 & .658 & .092 \\ .250 & -.103 & -.147 \\ .250 & -.451 & .201 \end{pmatrix}$$

$$= (.875, -.277, .027)$$

$$D = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -.314 & -.686 \\ 1 & -3.186 & 2.186 \end{pmatrix} \begin{pmatrix} 2.5 \\ 3.5 \\ 4.5 \end{pmatrix} = \begin{pmatrix} 14.000 \\ -1.686 \\ 1.186 \end{pmatrix}$$

$C \times D = (12.25, .468, .032).$

Thus

$$\begin{aligned}
\mathrm{E}[XU] &= 12.25(1^g) + .468(.769^g) + (.032)(.481^g) \\
&= 12.25 + (.468)(.769^g) + (.032)(.481^g) \\
\mathrm{E}[X] &= \mathrm{E}[U] = \mathrm{Sum}(\alpha \times \mu) = 3.5 \\
\therefore\ \mathrm{E}[X]\mathrm{E}[U] &= 3.5^2 = 12.25 \\
\mathrm{Cov}[X,U] &= \mathrm{E}[XU] - \mathrm{E}[X]\mathrm{E}[U] \\
&= (.468)(.769^g) + (.032)(.481^g).
\end{aligned}$$

Note how the first term of $\mathrm{E}[XU]$ cancels with $\mathrm{E}[X]\mathrm{E}[U]$; this will happen if the eigenvalue of unity is placed first. In general, the covariance of X and U is a sum of coefficients times eigenvalues (other than unity) raised to the power g.

Since $|\lambda_i| < 1$ for $i > 1$, the covariance will converge to zero as $g \to \infty$, because it is limited by a constant times the largest λ_i in magnitude (other than unity) raised to the power g.

Let ζ be the vector such that:

$$\zeta_i = C_i D_i = ((\mu \times \alpha)^T V^{-1})_i (V\mu)_i.$$

This is Equation 2.5 in the main text.

Then we have, for $g > 0$, Equation 2.6 in the main text:

$$\mathrm{Cov}[X,U] = \sum_{i>1} \zeta_i \lambda_i^g.$$

Note that λ_i and ζ_i which determine the behavior of the covariance are each directly and easily calculable[88] from the assumed transition matrix and the means of the states.

[88] Assuming the calculations will be performed on a computer.