TITLE: PARAMETER UNCERTAINTY IN THE COLLECTIVE RISK MODEL

AUTHOR: Mr. Glenn Meyers

Company Affiliation - Statistical Research Analyst,
CNA Insurance Companies
CAS Status - Fellow, May 1982
AAA Status - Member, April 1981
Education - Ph.D. State University of New York at
Albany, 1972
M.A. Oakland University, 1968
B.S. Alma College, 1967
Major Field, Mathematics
Previous Paper - "An Analysis of Retrospective Rating",
PCAS LXVII
Committees - Credibility Subcommittee ISO
Increased Limits Subcommittee ISO
Table M. Subcommittee NCCI
Individual Risk Rating Plans Subcommittee NCCI
Member - American Statistical Association

Mr. Nathaniel Schenker

Nathaniel Schenker is a graduate student in statistics at the University of Chicago. He received an A.B. degree from Princeton University in 1979. He is a member of the American Statistical Association and the Institute of Mathematical Statistics, and an editorial collaborator for the Journal of the American Statistical Association. He is co-author, with Henry Braun, of "New Models for Reliability Growth", Technical Report No. 174, Princeton University, October, 1980.

REVIEWER: Roy K. Morell

Roy Morell is Associate Actuary with Liberty Mutual Insurance Company. He received his FCAS designation in 1980 and has been a member of the American Academy of Actuaries since 1979. Roy earned a BS degree from the University of New Hampshire.

I. Introduction

This paper discusses the role of Collective Risk Theory in making insurance pricing decisions. Actuaries are making increasing use of Collective Risk Theory to derive aggregate loss distributions which in turn are used to measure the risk of an insurance contract, or to calculate the pure premium of an aggregate excess insurance contract for a large insured.

One of the main pillars of insurance theory has been the Law of Large Numbers. In the context of large account pricing, this law can be stated as follows. "As the risk size increases, we expect, ..., the variance (of the loss ratio) to approach zero ultimately." [1] Large insureds are typically written on a retrospective rating plan or an aggregate excess policy. The effect of the Law of Large Numbers would be that for any entry ratio greater than one, the excess pure premium ratio would approach zero as the size of the insured becomes large.

The practical underwriter would feel very uncomfortable with an agreement to provide coverage for all losses above the expected loss for a zero or nominal premium. His complaint would be that the expected loss cannot be estimated with the necessary precision.

What is of interest is the distribution of actual losses about an unbiased estimate of expected losses. Actual losses depend upon future economic conditions, changes in the insured's operations and changes in loss control procedures. Many of these factors are independent of the size of the insured. Thus one should not expect the Law of Large Numbers to hold when the expected loss cannot be estimated with precision.

For the small insured, the variance due to the random nature of the loss process is large compared to the variance due to misestimation of expected loss. But as the insured increases in size, the variance due to the random nature of the loss process decreases in accordance with the Law of Large Numbers. For the large insured, the variance due to the misestimation of expected losses will dominate.

Below, we will supply a more precise version of this statement. But first, a fundamental question should be addressed. Why use a model?

2. The Actuary's Dilemma

It has been a long standing debate among actuaries as to whether one should use empirical data or theoretical models to derive aggregate loss distributions. After completing the mammoth task of tabulating empirical aggregate loss distributions in 1965, LeRoy Simon wrote: "To avoid the difficulties and pitfalls of empiricism, we should borrow from the mathematical theory of risk, from Monte Carlo techniques and from operations research,---. Let's begin pushing some frontiers today, so that we'll be ready to solve tomorrow's problems."[2]

Officially, it appears that those who favor the empirical approach have prevailed, and Mr. Simon's advice has gone unheeded. Skurnick constructed a table in 1974 based on empirical observations.[3] In 1980, a National Council subcommittee constructed yet another table based on empirical observations. It should also be noted that Mr. Simon's table is still in effect after some sixteen years.

While the use of empirical distributions does not require one to make the assumptions that are necessary in the theoretical approach, there are fundamental problems with the empirical approach. It is conceded that the variance of the loss ratio distribution decreases as the size of the insured increases. It is also conceded that the variance of the distribution of loss ratios should increase as the average claim size increases . But it is necessary to combine the results of insureds of different sizes and average claim severities to get a sufficiently large sample. For example, the tables constructed recently by the National Council combined all insureds on a countrywide basis into expected loss ranges which include \$25000 to \$50000, \$50000 to \$100000 and \$100000 to \$200000.

Thus the actuary is faced with the dilemma of choosing between two undesirable alternatives. If he elects the empirical approach, he must take a sample from a heterogeneous population. If he elects the theoretical approach, he must make a number of simplifying assumptions.

By proposing a mathematical model, we do not advocate abandoning the use of empirical data. Once a model has been constructed, it should be possible to form hypotheses which can be tested by live data. If it can be shown, by using statistical tests, that the model is consistent with the data, the dilemma will be resolved.

It will become apparent that this is easier said than done, but this is the goal toward which we all must strive.

If this goal is reached, there will be many advantages to the theoretical approach. Since the size of insured and the claim severity distribution are input variables, the model should give a better representation of the insured aggregate loss distribution than the empirical approach. It will not be necessary to combine experience from a heterogeneous group of insureds. In addition, it will be possible to adjust the parameters of the model to account for situations when there is little or no empirical data available. For example, it would be a simple matter to find the aggregate loss distribution that results when all claims are subject to an accident limitation.

The Collective Risk Model

Collective Risk Theory started by considering the Generalized Poisson distribution. However it soon became apparent that the assumptions of this distribution are violated for many applications. In this section we will discuss the assumptions underlying the use of the Generalized Poisson distribution and indicate some common violations of these assumptions. We will then state a version of the Collective Risk Model which can deal with certain violations of these assumptions.

We start by considering the Poisson distribution. The assumptions underlying this distribution are as follows.[4]

1. Claims occurring in two disjoint time intervals are independent.

2. The expected number of claims in a time interval (t_1, t_2) is dependent only on the length of the interval and not on the initial value t_1.

3. No more than one claim can occur at a given time.

There are situations when these assumptions are violated. We give three examples.

1. <u>Positive Contagion</u>

   A manufacturer can be held liable for defects in its products which, in many cases, are mass produced. A successful claim against the manufacturer may result in several other claims against the manufacturer. The notion that a higher than expected number of claims in an earlier period can increase the expected number of claims in a future period is called positive contagion.

2. <u>Negative Contagion</u>

   Consider a group life insurance policy. A death in an earlier period will reduce the expected number of deaths in a later period. One does not die twice. The notion that a higher than expected number of claims in an earlier period can decrease the expected number of claims in a future period is called negative contagion.

3. Parameter Uncertainty

There are many cases when one feels that a Poisson distribution is appropriate, but one does not know the expected number of claims. Two options are available under these circumstances. The first option is to estimate the expected number of claims using historical experience. Parameter uncertainty can come from sampling variability. A second option is to use the average number of claims for a group of insureds which are similar to the insured under consideration. Parameter uncertainty arises when some members of the group have different expected numbers of claims.

The effect of parameter uncertainty is similar to that of positive contagion. We give a heuristic argument for this which appeals to modern credibility theory. Suppose one observes N claims during a certain time period. Then one can estimate the number of claims, x, in a future period of equal length using the following forumla:

$$x = Z \cdot N + (1-Z) \cdot E.$$

Where E = Prior estimate

Z = Credibility factor.

Note that if the estimate of the expected number of claims is precise or the group of insureds is homogeneous, the credibility factor will be 0.

If N is greater than expected, the number of claims expected in the future will be greater than the prior estimate.

It should be emphasized that we are not arguing that claims in an earlier period will cause claims in a later period, as in the positive contagion case. We are stating only that the claim count distributions observed under the conditions of parameter uncertainty and positive contagion should be similar.

We now turn to specifying the claim count distributions we shall use for each of the above situations. We shall adopt the following notation.

N - A random variable denoting the number of claims

λ - The expected number of claims ($\lambda = E[N]$)

χ - A random variable with $E[\chi] = 1$ and $Var[\chi] = c$

Parameter uncertainty can be modeled by the following algorithm.

Algorithm 3.1

1. Select χ at random from the assumed distribution.

2. Select the number of claims, N, at random from a Poisson distribution with parameter $\chi\lambda$.

We have the following relationships.

$$E[N] = E_{\chi}[E(N\,|\,\chi)] = E_{\chi}[\chi\lambda] = \lambda \tag{3.1}$$

$$\begin{aligned} \mathrm{Var}\,[N] &= E_{\chi}\ [\mathrm{Var}\ (N\,|\,\chi)] + \mathrm{Var}_{\chi}\,[E(N\,|\,\chi)] \\ &= E_{\chi}\ [\chi\lambda] + \mathrm{Var}_{\chi}\,[\chi\lambda] \\ &= \lambda + c\,\lambda^2 \end{aligned} \tag{3.2}$$

If χ has a Gamma distribution, the claim count distribution described by Algorithm 3.1 is the Negative Binomial distribution.[5] We shall use the Negative Binomial distribution to model both the positive contagion and the parameter uncertainty cases. The Binomial distribution can be used to model the negative contagion case.

We shall call the parameter c the contagion parameter for the claim count distribution. We should note that c has also been called the contamination parameter by some authors.[7] If c = 0, Algorithm 3.1 yields the Poisson distribution.

We now adopt the following notation.

- Z - A random variable denoting the amount of a claim.
- $S(z)$ - The cumulative distribution function of the claim severity, z.
- x - A random variable denoting the aggregate loss for an insured.

Aggregate losses can be generated by the following algorithm.

Algorithm 3.2

1. Select the number of claims, N, at random from the assumed claim count distribution.

2. Do the following N times

   2.1 Select the claim amount, Z, at random from the assumed claim severity distribution.

3. The aggregate loss amount, X, is the sum of all claim amounts, Z, selected in step 2.1.

We now give expressions for the mean and the variance of the aggregate loss distribution generated by Algorithm 3.2.

$$E[X] = E[N] \;.\; E[Z] = \lambda \;.\; E[Z] \tag{3.3}$$

$$\mathrm{Var}\,[X] = E_N \;[\mathrm{Var}\,(X \mid N)] + \mathrm{Var}_N\,[N \;.\; E(Z)]$$

$$= \lambda \;.\; \mathrm{Var}\,[Z] + (\lambda + c\,\lambda^2)\;,\; E^2\,[Z]$$

$$= \lambda \;.\; E[Z^2] + c\,\lambda^2 \;.\; E^2\,[Z] \tag{3.4}$$

Implicit in the use of Algoritm 3.2 is the assumption that the claim severity distribution, $S(z)$, is known. In practice this distribution must be estimated from historical observations, or it must be simply assumed. Parameter uncertainty of the claim severity distribution can significantly affect the predictions of the Collective Risk Model, and it should not be ignored. Our response to this problem is to make the simplifying (and we think reasonable) assumption that the shape of the distribution is known, but there is uncertainty on the scale of the distribution.

More precisely, we specify parameter uncertainty of the claim severity distribution in the following manner.

Let β be a random variable satisfying the conditions $E[1/\beta] = 1$ and $\mathrm{Var}[1/\beta] = b$. We then model aggregate losses by the following algorithm.

Algorithm 3.3

1. Select the number of claims, N, at random from the assumed claim count distribution.

2. Select the scaling parameter, β , at random from the assumed distribution.

3. Do the following N times.

   3.1 Select the claim amount Z at random from the assumed claim severity distribution.

4. The aggregate loss amount, X, is the sum of all claim amounts, Z, divided by the scaling parameter β .

We now give formulas for the mean and the variance for the aggregate loss distribution generated by Algorithm 3.3.

$$
\begin{aligned}
E[X] &= E_{\beta}\ [E(X\,|\beta)] \\
&= E_{\beta}\ [\lambda .\ E(Z)/\beta\,] \\
&= \lambda .\ E[Z]\ .\ E[1/\beta\,] \\
&= \lambda .\ E[Z] \qquad (3.5)
\end{aligned}
$$

$$
\begin{aligned}
Var[X] &= E_{\beta}[Var(X\,|\beta)] + Var_{\beta}\ [E(X\,|\beta)] \\
&= E_{\beta}[(\lambda .\ E(Z^2) + c\lambda^2\ E^2\ (Z))/\beta^2] + Var_{\beta}[\lambda .\ E(Z)/\beta\,] \\
&= (\lambda .\ E[Z^2] + c\ \lambda^2\ .\ E^2\ [Z])\ .\ E[1/\beta 2] + \lambda 2\ .\ E^2[Z] \\
&\qquad .Var[1/\beta] \\
&= (\lambda .\ E[Z^2] + c\ \lambda^2\ .\ E^2\ [Z])\ .\ (1+b) + \lambda^2\ .\ E^2[Z]\ .\ b \\
&= \lambda E[Z^2]\ (1+b) + \lambda^2\ .\ E^2\ [Z]\ .\ (b+c+bc) \qquad (3.6)
\end{aligned}
$$

Let R denote the loss ratio X/E[X]. From Equations 3.5 and 3.6 we have the following results.

$$
\begin{aligned}
Var\ [R] &= Var[X/E[X]] \\
&= (1+b) .\frac{E[Z^2]}{E^2[Z]}\ .\ \frac{1}{\lambda} + b + c + bc \qquad (3.7)
\end{aligned}
$$

Thus this model implies a linear relationship between the Var [R] and $1/\lambda$. If either b or c is nonzero, the Law of Large Numbers does not hold. As the insured gets large, the expected number of claims, λ, gets large, and the variance of the loss ratio distribution approaches b + c + bc. However, for a small insured, this limit is small compared to Var[R].

In this paper we shall assume that β has a Gamma distribution. We shall call b the mixing parameter. The mixing parameter is a measure of parameter uncertainty for the claim severity distribution.

Under the above assumptions on parameter uncertainty, it is possible to calculate the cumulative distribution function and the excess pure premium ratio for aggregate losses in an efficient manner.[8] It should be noted that we have chosen mathematically convenient distributions to model parameter uncertainty. We do not want to imply that these distributions are in any way the "correct" ones. Since parameter uncertainty is not directly observable it is difficult to discover what the correct distribution should be. As we shall show, it is possible to infer the variance of the parameter uncertainty through the use of Equations 3.6 and 3.7. But until statistical methodology has advanced to the point where the proper distributions can be determined, it should be acceptable to use ones which are mathematically convenient.

4. Estimating the Parameters of the Model

In the previous section we proposed a modification of the collective risk model which accounts for parameter uncertainty. This model depends upon the claim severity distribution, $S(z)$, the expected number of claims, λ, the mixing parameter b, and the contagion parameter c.

A complete discussion of estimating claim severity distributions is beyond the scope of this paper.[9] In our work we typically obtain claim severity distributions from bureau circulars, or we estimate claim severity distributions from company data.

Since these claim severity distributions are derived from experience other than that of the insured under consideration, we frequently adjust the scale of the claim severity distribution so that the average claim size matches that which we project for the insured.

The expected number of claims can be obtained by dividing the expected losses by the average claim size.

Two years ago, Gary Patrik and Russell John presented a serious attempt to deal with parameter uncertainty.[10] We summarize their approach. They pick a finite set of parameters (claim severity distributions and claim count distributions) for the collective risk model. They then combine the various outputs of the model by taking a weighted average. The weights are probabilities which they assign subjectively.

The use of subjective probabilities has always been controversial. Many consider the word "guess" to be more appropriate. It is unfortunate that in many situations an answer is demanded, but no data is available. Under these circumstances, the use of subjective probabilities may be acceptable.

Regardless of how one feels toward the use of subjective probabilities, one should always consider the possibility of estimating b and c from observations of aggregate loss data. The remainder of this section will develop ways of doing this.

The estimates will be based on several observations for which a single b and a single c is appropriate. These observations may come from more than one insured. Ideally, one would like to have premium (or exposure), incurred losses, claim count and the mean and variance of the claim severity distribution for each observation. In practice, data is usually not available in this detail. Since one wants to use all available data, we will provide several estimators. The choice of estimators will depend on the available data.

Estimation of c

Let $N_1, \ldots, N_r$ be r independent claim counts. Let $e_1, \ldots, e_r$ be numbers such that $e_i = K \cdot E[N_i]$ for all i. K is a constant of proportionality. While we will refer to e_i as a premium, it could just as easily represent exposure.

Let $E[N_1] = \lambda_1$. Then $E[N_i] = \frac{e_i}{e_1}\lambda_1$. It is a consequence of Equation 3.2 that

$$\mathrm{Var}[N_i] = \frac{e_i}{e_1} \cdot \lambda_1 + c\left(\frac{e_i}{e_1} \cdot \lambda_1\right)^2, i = 1, \ldots r.$$

Let $\hat{\lambda}_1 = \frac{1}{r} \sum_{i=1}^{r} \frac{e_1}{e_i} N_i$ and

$$V = \sum_{i=1}^{r} \left(\frac{e_1}{e_i} N_i - \hat{\lambda}_1 \right)^2 .$$

We have $E(\hat{\lambda}_1) = \lambda_1$.

Now, $V = \sum_{i=1}^{r} \left[\frac{e_1}{e_i} N_i - \lambda_1 - \left(\hat{\lambda}_1 - \lambda_1 \right) \right]^2$

$$= \sum_{i=1}^{r} \left(\frac{e_1}{e_i} N_i - \lambda_1 \right)^2 - 2 \left(\hat{\lambda}_1 - \lambda_1 \right) \sum_{i=1}^{r} \left(\frac{e_1}{e_i} N_i - \lambda_1 \right)$$

$$+ r (\hat{\lambda}_1 - \lambda_1)^2$$

$$= \sum_{i=1}^{r} \left(\frac{e_1}{e_i} N_i - \lambda_1 \right)^2 - r(\hat{\lambda}_1 - \lambda_1)^2 .$$

Thus, $E[V] = \sum_{i=1}^{r} \mathrm{Var} \left[\frac{e_1}{e_i} N_i \right] - r \, \mathrm{Var} \, [\hat{\lambda}_1]$

$$= \sum_{i=1}^{r} \mathrm{Var} \left[\frac{e_1}{e_i} N_i \right] - \frac{1}{r} \sum_{i=1}^{r} \mathrm{Var} \left[\frac{e_1}{e_i} \cdot N_i \right]$$

$$= \frac{r-1}{r} \sum_{i=1}^{r} \left(\frac{e_1}{e_i} \right)^2 \cdot \mathrm{Var} \, [N_i]$$

$$= \frac{r-1}{r} \sum_{i=1}^{r} \left[\frac{e_1}{e_i} \lambda_1 + \lambda_1^2 \, c \right]$$

$$= \frac{r-1}{r} \lambda_1 \sum_{i=1}^{r} \frac{e_1}{e_i} + (r-1) \lambda_1^2 \, c.$$

It follows that

$$c = \frac{E[V] - \frac{r-1}{r} \lambda_1 \sum_{i=1}^{r} \frac{e_1}{e_i}}{(r-1) \, . \, \lambda_1^2} .$$

Substituting V for E[V] and $\hat{\lambda}_1$ for λ_1 yields

$$\hat{c} = \frac{V - \frac{r-1}{r} \cdot \hat{\lambda}_1 \sum_{i=1}^{r} \frac{e_1}{e_i}}{(r-1) \cdot \hat{\lambda}_1^2} \qquad (4.1)$$

as an estimator for c.

Estimation of b

Let $X_1, \ldots, X_r$ be r independent loss amounts associated with $N_1, \ldots, N_r$ respectively. Let β denote the severity scaling coefficient. Let $A_i = X_i/N_i$, $i=1,\ldots,r$ be average claim costs. Let $N = \sum_{i=1}^{r} N_i$

Then $E[A_i \mid N_i, \beta] = 1/\beta \cdot E[Z]$

and $Var[A_i \mid N_i, \beta] = (1/\beta^2) \cdot Var[Z]/N_i$

Thus, $E[A_i \mid N_i] = E_\beta[E[A_i \mid N_i, \beta]] = E[Z]$

and $Var[A_i \mid N_i] = E_\beta[Var[A_i \mid N_i, \beta]] + Var_\beta[E[A_i \mid N_i, \beta]]$

$$= (1+b)\, Var[Z]/N_i + b \cdot E^2[Z] \qquad (4.2)$$

Let $T_1 = \sum_{i=1}^{r} N_i \, (A_i - E[Z])^2$.

Then $E[T_1 | N_1,..,N_r] = \sum_{i=1}^{r} N_i \text{ Var } [A_i | N_i]$

$$= r(1+b) \,.\, \text{Var } [Z] + N \cdot b \cdot E^2[Z].$$

It follows that

$$b = \frac{E[T_1 | N_1,...,N_r] - r \,.\, \text{Var } [Z]}{r. \text{ Var } [Z] + N \,.\, E^2 [Z]} .$$

Substituting T_1 for $E[T_1 | N_1,..,N_r]$ yields

$$\hat{b} = \frac{T_1 - r \,.\, \text{Var } [Z]}{r.\text{Var } [Z] + N.E^2[Z]} \tag{4.3}$$

as an estimator for b.

The above estimator for b assumes that the mean and variance of the claim severity distribution are known. This estimator is appropriate when the claim severity distribution is derived from sources other than our r observations of actual loss data. We now provide an estimator for b which uses the observed claim severity.

Let $\overline{Z} = \sum_{i=1}^{r} X_i/N$, and $T_2 = \sum_{i=1}^{r} N_i\,(A_i - \overline{Z})^2$. We have
$E[\overline{Z} | N_1,\ldots N_r] = E[Z]$.

$$\text{Now } T_2 = \sum_{i=1}^{r} N_i[(A_i - E[Z]) - (\overline{Z} - E[Z])]^2$$

$$= \sum_{i=1}^{r} N_i(A_i - E[Z])^2 - N.\,(\overline{Z} - E[Z])^2.$$

Thus

$$E[T_2 | N_1,\ldots,N_r] = \sum_{i=1}^{r} N_i \text{ Var } [A_i | N_i] - N.\text{Var } [\overline{Z} | N_1,\ldots,N_r]$$

$$= \sum_{i=1}^{r} N_i \text{ Var}[A_i | N_i] - \frac{1}{N} \sum_{i=1}^{r} \text{Var } [N_i A_i | N_i]$$

$$= r(1+b)\ .\ \text{Var } [Z] + N.b.E^2\,[Z]$$

$$- \frac{1}{N}\ .\sum_{i=1}^{r} (N_i\ (1+b)\ \text{Var } [Z] + N_i^2\ [E^2\ [Z])$$

$$= (1+b)(r-1)\ .\ \text{Var } [Z] + b\ E^2[Z]\left(N - \frac{\sum_{i=1}^{r} N_i^2}{N}\right)$$

It follows that

$$b = \frac{E[T_2 | N_1,\ldots,N_r) - (r-1)\ \text{Var } [Z]}{(r-1).\text{Var}[Z] + E^2[Z]\left(N - \frac{\sum_{i=1}^{r} N_i^2}{N}\right)}\ .$$

Let $\widehat{Var}$ [Z] be an estimator for Var [Z]. In practice $\widehat{Var}$ [Z] could be the sample variance of the claim severity distribution for the r observations. It could also be derived from $\bar{Z}$ and an assumed coefficient of variation. Then substituting T_2 for $E[T_2|N_1,\ldots,N_r]$, $\bar{Z}$ for E[Z] and $\widehat{Var}$ [Z] for Var [Z] yields.

$$\hat{b} = \frac{T_2 - (r-1).\widehat{Var}[Z]}{(r-1).\widehat{Var}[Z] + \bar{Z}^2 \left(\dfrac{N - \sum_{i=1}^{r} N_i^2}{N} \right)} \tag{4.4}$$

as an estimator for b.

We now consider the case when we do not have the claim count for each observation.

Let $R_i = X_i/e_i$ for $i = 1,\ldots,r$ and let $\mu = E[R_i]$.

Then $Var[R_i] = Var\,[X_i]/e_i^2$

$$=[(1+b).E[Z^2].E[N_i]+E^2[Z].E^2[N_i](b+c+bc)]/e_i^2 \quad .$$

Now, $E[N_i] = \dfrac{e_i\ .E[N_i].E[Z]}{e_i\ E[Z]} = e_i\ .\ \dfrac{E[R_i]}{E[Z]} = \dfrac{e_i}{E[Z]}\ .\mu\ .$

Thus,

$$\mathrm{Var}\,[R_i] = \frac{(1+b)\,.\,E[Z^2]\,.\,e_i.\mu}{E[Z]} + e_i^2\,.\mu^2(b+c+bc)\,\frac{1}{e_i{}^2}$$

$$= \frac{(1+b)\,.\,E[Z^2]}{E[Z]}.\frac{\mu}{e_i} + \mu^2\,(b+c+bc)$$

Let $\hat{\mu} = \frac{1}{r} \sum_{i=1}^{r} R_i$ and $W = \sum_{i=1}^{r} (R_i - \hat{\mu})^2$. Then $E\,[\hat{\mu}] = \mu$ and

$$W = \sum_{i=1}^{r} [\,(R_i - \mu) - (\hat{\mu} - \mu)\,]^2$$

$$= \sum_{i=1}^{r} (R_i - \mu)^2 - r\,(\hat{\mu} - \mu)^2$$

Thus,

$$E[W] = \sum_{i=1}^{r} \mathrm{Var}\,[R_i] - r\,\mathrm{Var}\,[\hat{\mu}]$$

$$= \frac{r-1}{r} \sum_{i=1}^{r} \mathrm{Var}[R_i]$$

$$= (r-1)\left[(1+b)\,.\,\frac{E[Z^2]}{E[Z]}\cdot\mu\,.\,\frac{1}{r}\sum_{i=1}^{r}\frac{1}{e_i} + \mu^2(b+c+bc)\right]$$

It follows that

$$\frac{E[W]}{(r-1)\mu^2} = (1+b).\,\frac{E[Z^2]}{E[Z]}\;\frac{1}{r}\cdot\sum_{i=1}^{r}\frac{1}{e_i}\;.\,\frac{1}{\mu} + b + c + bc. \qquad (4.5)$$

Given several loss ratios, divide them into n groups of size r_j, $j = 1, \ldots, n$. For the j^{th} group, let the loss ratios and premiums be denoted by R_{ij} and e_{ij}, $i = 1, \ldots, r_j$, respectively. Let

$$\hat{\mu}_j = \frac{1}{r_j} \sum_{i=1}^{r_j} R_{ij}, \quad E_j = \frac{1}{r_j} \sum_{i=1}^{r_j} \frac{1}{e_{ij}} \text{ and } W_j = \sum_{i=1}^{r_j} (R_{ij} - \hat{\mu}_j)^2$$

It is assumed that $E[R_{ij}] = \mu_j$ for each j.

Let $A = (1+b) \cdot \frac{E[Z^2]}{E[Z]}$ and $B = b + c + bc$.

From Equation 4.5 we have

$$\frac{E[W_j]}{(r_j-1)^2 \cdot \mu_j^2} = A \cdot \frac{E_j}{\mu_j} + B.$$

Estimates of A and B can be obtained by performing a linear regression of $\frac{W_j}{(r_j-1)^2 \hat{\mu}_j^2}$ on $\frac{E_j}{\hat{\mu}_j}$. It should be noted that $\frac{W_j}{(r_j-1)^2 \hat{\mu}_j^2}$ is the sample squared coefficient of variation for the j^{th} group.

we then have $\hat{A} = (1 + \hat{b}) \cdot \frac{E[Z^2]}{E[Z]}$

and $\hat{B} = \hat{b} + \hat{c} + \widehat{bc}$.

This yields

$$\hat{b} = \frac{\hat{A} \cdot E[Z]}{E[Z^2]} - 1 \tag{4.6}$$

and

$$\hat{c} = \frac{\hat{B}-\hat{b}}{1+\hat{b}} \tag{4.7}$$

as estimators for b and c.

E[Z] and $E[Z^2]$ can either be assumed, or estimated from the data.

We have given several estimators of b and c. The following tables summarize the data requirements for the various estimates.

Table 1 - Estimators of b

Estimator	Data Requirements for Each Observation
Equation 4.3	Premium Losses Claim count Assumed severity mean and variance
Equation 4.4	Premium Losses Claim count Severity mean and variance estimated from combined experience of all observations. (One may use an assumed coefficient of variation to estimate the variance)
Equation 4.6	Premium Losses Assumed severity mean and variance

Table 2 - Estimators of c

Estimator	Data Requirements for Each Observation
Equation 4.1	Premium Claim count
Equation 4.7	Premium Losses Assumed severity mean and variance Estimate of b (Equation 4.6)

5. Testing the Collective Risk Model on Live Data

In 1980 a National Council committee constructed a new Table M based on empirical data. At this time we became aware of a very large data base of individual insured experience. We requested and received from the National Council a tape containing the experience of well over 1,000,000 insureds. This data forms the basis of our analysis.

The data used in this study was NCCI policy year 1973-74 interstate data developed to the third report. This data consisted of premium and losses for each insured. At the time this study was done the National Council did not have a claim severity distribution available. The closest thing we had to a comparable claim severity distribution was estimated from our own company data for the National Council states for accident year 1975 developed to 42 months. We then changed the scale of the distribution to match the average claim size which was reported by the National Council for policy year 1973-74. The resulting claim severity distribution is given in Exhibit I.

It should be stressed that the National Council data base is made up of insureds from many different states and many different Hazard Groups. Also, the claim severity distribution is a composite distribution of the many different insureds from a single company. These are serious defects. But testing the Collective Risk Model on live data is extremely important, and so we proceed.

Estimating b and c

As noted above, we have premium and loss information for each insured and an assumed claim severity distribution. This requires us to use a linear regression and Equations 4.6 and 4.7 to estimate b and c. A problem with the regression is how to choose the groups in which the squared coefficients of variation of the loss ratios are to be computed. We decided to choose the loss ratios corresponding to the r_1 lowest premiums the first group, the loss ratios corresponding to the next r_2 lowest premiums in the second group and so on. The problem remained of choosing the r_j, $j = 1, \ldots n$ for the n groups. If the r_j's are equal for all j, we observed that the variance of the residuals of the regression decreased as the premium became larger. In statistical terminology this is known as heteroscedasticity. We dealt with this problem in two ways. One way was to have r_j decrease as the premium increases. The other was to use a weighted regression.

The weighted regression can be described as follows. If the model $Y = A X + B + \varepsilon$ is to be fitted, but it appears that the standard deviation of ε is proportional to X, then let $Y' = Y/X$ and let $X' = 1/X$. In the new model $Y' = A' + B' X' + \varepsilon'$, ε', will have approximately constant variance. A' will be an estimate of A and B' will be an estimate of B.

Exhibit II gives the various sets of r_j's that we considered. Exhibit III gives the resulting estimates of b and c. The following comments should be made about these estimates.

1. It is possible for estimates of b and c to be negative. $\hat{b}$ will be negative whenever

$$\frac{\hat{A} \cdot E[Z]}{E[Z^2]} < 1$$

and $\hat{c}$ will be negative whenever

$$\frac{(\hat{B} + 1) \cdot E[Z^2]}{E[Z]} < \hat{A} .$$

This can happen if the assumed mean and variance of the claim severity distribution are not appropriate for the given observations. Negative estimates of b and c can also occur because of random variation of the regression coefficients. Examination of the standard errors of estimates for A suggests that random variation could explain the two negative estimates of c.

If a negative estimate of b or c is obtained, we suggest using 0 in place of the negative estimate.

2. If one uses Equations 4.6 and 4.7 to estimate b and c, the variance of the aggregate loss distribution is not sensitive to the assumed mean and variance of the claim severity distribution. If a small error is made in selecting the mean and the variance of the claim severity distribution, the estimates of b and c will compensate to give the variance of the aggregate loss distribution that was determined by the regression.

3. The estimate $\hat{B}$ of b + c + bc appears to be decreasing as the size of the insured increases. This can be seen by comparing the pairs of estimates #2 (j = 1-11) with #2 (j = 12-22), #3 (j = 1-8) with #3 (j = 9-15) and #4 with #5. In all three comparisons the estimate of B corresponding to higher premium observations was lower than the estimates corresponding to lower premium observations.

   This seems to be a reasonable conclusion. Because of experience rating, one would expect to be able to estimate the losses of a large insured with greater precision than for a small insured. If losses are estimated using manual rates, we would expect B to be the same for all premium sizes.

4. The estimates of A and B vary by the set of r_j's chosen. Examination of the standard errors of the coefficient for $\hat{A}$ indicates that this variation is random. However, the variation in the estimates for B cannot be explained by random variation. It should be noted that the regressions which have a relatively high number of points corresponding to large premium sizes have lower $\hat{B}$'s. Compare regression #1 with regression #2. The variation in $\hat{B}$ can thus be explained by a decreasing B as premium increases.

<u>Comparison of Expected with Actual Results</u>

Using the estimates of b and c we obtained in the preceding section, it is possible to calculate the cumulative probabilities and the excess pure premium implied by the model. In this section we compare the results predicted by the model with the actual results in the National Council data base. This comparison will take two forms. We will first perform chi-square tests on the data. We will then compare excess pure premium ratios predicted from the model with excess pure premium ratios calculated from the data base.

We chose three sets of parameter values for our testing. In our first test we set b=0 and c=0 as a control hypothesis. We chose the estimates $\hat{b}$ = .258 and $\hat{c}$ = .037 from regression #3 since it produced the best fit over all the points. As our third set of parameters we chose $\hat{b}$ = .184 and $\hat{c}$ = .220 for the model when the premium was less than \$125000 (regression #4) and $\hat{b}$ = .263 and $\hat{c}$ = .068 for the model when the premium was greater than \$125000. This enabled us to test the hypothesis that B (and therefore b and c) decrease as the premium increase.

Since the model predicts that the variance of the loss ratio distribution changes with the size of the insured, we decided to perform the chi-square tests on several groups of insureds. Each group was to have a fairly narrow range of premium sizes. The results are given in Exhibit IV.

No set of parameter values performed well when the premium was less than \$15000. While the second and third sets did a much better job than the first, all sets severely underestimated the number of small loss ratios. The following table shows a typical result.

Table 3

Number of Observations: 364
Range of Premiums: 5001 to 5025
Average Loss: 4209.34

Upper Boundary	Observed	Expected	Contribution	
42.09	72	21.79	115.71	(+)
84.19	15	21.53	1.98	(-)
147.33	19	29.82	3.93	(-)
252.56	42	38.32	0.35	(+)
420.93	30	39.24	2.18	(-)
841.87	34	49.97	5.10	(-)
1262.80	20	27.03	1.83	(-)
2104.67	32	31.75	0.00	(+)
3788.41	26	31.63	1.00	(-)
9260.55	30	36.64	1.20	(-)
18521.10	19	18.46	0.02	(+)
Infinity	25	17.82	2.90	(+)

Chi-Square = 136.19

It would appear that higher values of b and c are required for small premium sizes.

If one looks at the results on individual groups, it is difficult to note a pattern in the results. The chi-square test is simply not powerful enough to distinguish between the various sets of parameter values on the individual groups. However, the chi-square test permits the combining of the results of independent tests.[11] When this is done a clear pattern emerges. The results predicted by the second and third sets of parameter values are significantly better than the results predicted by the parameters $b = 0$ and $c = 0$. Allowing for parameter uncertaintly significantly improves the performance of the Collective Risk Model. Comparing results for insureds in the \$15000-50000 range with those in the \$50000 - 200000 range make it equally clear that b and c must vary by the size of the insured.

It should come as no surprise that the chi-square test indicates that we must reject the hypothesis that aggregate losses have the distribution predicted by the model. We have made a number of simplifying assumptions about the nature of the data.

In 1980 a committee of actuaries at the National Council produced tables of excess pure premium ratios based on the same data we have used. Exhibit V provides a comparison of excess pure premiums produced by the model with those produced by the committee.

We believe the excess pure premium ratios are reasonably close. The differences between the model and the 1980 Table M are much less than the differences that can be attributed to known differences among individual insureds, such as claim severity distributions and loss limits [12]

6. The Calculation of b and c for an Individual Insured

The preceding section demonstrated that we should expect b and c to decrease as the size of the insured increases. The estimates of b and c were tested for insureds whose standard premium was below $200000. No measurement of b and c was given for the very large insured.

Large accounts receive extensive analysis by highly paid professionals representing both the insurer and the insured. A great deal of quantitative and qualitative information goes into the analysis. This results in an insurance contract that is tailored to the needs of the individual insured.

It seems highly unlikely that the values of b and c should be simply a function of the size of the insured. Many insureds operating in a radically different environment from year to year, while many others operate in a comparatively stable environment. Large account underwriters recognize this and they design their contracts accordingly.

Because of the intense scrutiny afforded large insureds, large insureds have traditionally been priced on the basis of their own experience.

In this section we test the feasibility of estimating b and c from the historical observations of a single insured. Our test proceeds in the following manner.

1. Select a claim severity distribution.
2. Select the expected losses.
3. Select b and c.
4. Select the number of observations
5. Generate the given number of observations using Algorithm 3.3
6. Calculate $\hat{b}$ and $\hat{c}$.
7. Compare $\hat{b}$ with b and $\hat{c}$ with c.

This test was done repeatedly for a variety of conditions. The results are summarized in Exhibits VI and VII.

Exhibits VI and VII show that while the accuracy of the estimate improves with an increasing number of observations and with increasing size of insured, errors of over 100% can be quite common. The number of observations required for an accurate estimate is unrealistically high. Estimating b and c from historical individual insured data can hardly be called reliable.

Exhibit VIII shows that there is a substantial difference in excess pure premium ratios for large insureds that can be attributed to differences in b and c. We conclude that the primary risk in pricing aggregate excess for a large insured is not in the random nature of the loss process. Instead the risk lies in the selection of the expected losses, or more specifically, in the selection of the parameters for the Collective Risk Model.

We would come to the same conclusion if the excess pure premium ratios were derived by empirical methods rather than by Collective Risk Theory. The large volume of data required would have to come from a group of heterogeneous insureds. It is highly unlikely that the group experience would be a good indicator of what one should expect for an individual large account.

It should be pointed out that underwriters have a considerable amount of flexibility in designing insurance contracts. Frequently, the maximum premium in a retrospective rating plan will be set high enough to absorb errors in estimating the expected losses. Another option is to design a contract which will spread the losses over a period of years.

We would not rule out the possibility of estimating b and c by some other means. It may be possible to estimate b from the standard error of a regression for average claim costs. Using credibility theory to estimate b and c may be possible. Both of these areas require further research.

We believe the proper role for Collective Risk Theory in large account pricing lies in sensitivity testing. The Collective Risk Model can be a very useful tool for determining how much is at risk for various values of b and c. Estimates of b and c based on several insureds may provide a benchmark to work from, but they should be used carefully. This information will be very useful to both the insurer and the insured in designing the proper insurance contract.

Conclusion

The Collective Risk Model can be a very useful tool for pricing insurance contracts. Allowing for parameter uncertainty significantly improves the accuracy of the model. In deciding how to use it, one must also consider the available alternatives.

It is not cost effective to provide an extensive analysis of the individual small insured. Thus class rating and statistical methods are the primary tools for rating the small insured.

For small insureds the Collective Risk Model can reproduce the results of empirical data with reasonable accuracy. Since the model can adjust for known differences between the underlying data in current tables and the proposed insurance contract, such as changes in claim severity and loss limitations, it should be used in place of current tables.

For large insureds the inability to form a precise estimate of expected losses is the major element of risk. The statistical methods of quantifying parameter uncertainty that are presented in this paper are extremely volatile when applied to observations of an individual insured. It seems likely that a sound (and perhaps expensive) qualitative analysis can do a better job. While the Collective Risk Model can provide useful information, the primary burden of responsibility rests with sound underwriting judgment and proper design of the insurance contract. It is questionable that published tables of excess pure premium ratios have any real meaning for the large insured.

8. Acknowledgements

We would like to thank the National Council of Compensation Insurance for providing us the empirical data that was used in this paper. We would also like to thank Bradley Alpert, Philip Heckman and Edward Seligman for their helpful comments.

Exhibit I The Claim Severity Distribution

Loss Amount	Cumulative Probability
0.0	0.0
19.79	0.21384
39.57	0.51025
79.15	0.74056
118.72	0.79959
158.29	0.82665
197.86	0.84450
277.01	0.86657
395.73	0.88626
593.59	0.90606
791.45	0.91797
1187.18	0.93388
1582.91	0.94464
1978.63	0.95223
2770.09	0.96242
3957.27	0.97156
5935.90	0.97998
7914.54	0.98476
9893.17	0.98785
11871.80	0.99001
15829.07	0.99281
19786.34	0.99452
27700.87	0.99649
39572.68	0.99790
59359.02	0.99890
79145.31	0.99934
98931.69	0.99956
118718.00	0.99970
158290.69	0.99983
197863.37	0.99990
277008.75	0.99996
395726.56	0.99998
593590.00	0.99999
791453.44	1.00000

Summary Statistics:

Severity Mean = 632.56
Severity Std Dev = 5407.69

Exhibit IIa Groupings used for the Regressions

	r_j*				
j	#1	#2	#3	#4	#5
1	46800	45000	87922	13632	350
2	34221	39000	60000	11500	325
3	26730	34000	50000	9500	300
4	21008	29000	40000	8000	275
5	31690	25000	30000	7000	250
6	22575	21000	20000	6000	225
7	29264	17500	12800	5000	200
8	18702	15000	6400	4500	175
9	12955	13000	3200	4000	150
10	9478	11000	1600	3500	
11	7304	10000	800	3000	
12	5477	9000	400	2500	
13	4592	8000	200	2000	
14	8573	7000	100		
15	5684	6000	50		
16	7318	5000			
17	4461	4500			
18	2933	4000			
19	3638	3500			
20	2265	3000			
21	1516	2500			
22	1057	1472			
23	724				
24	610				
25	453				
26	827				
27	541				
28	351				
29	231				
30	351				
31	188				
32	222				
33	129				
34	65				
35	41				
36	29				
37	33				
38	9				
39	23				
40	12				
41	8				
42	7				
43	10				

* Range of premium for each set is given in Exhibit IIb

Exhibit IIb Groupings used for the Regressions

Grouping	Range of Premium Sizes
#1	Premium $\geq$ 1000
#2	Premium $\geq$ 1000
#3	Premium $\geq$ 1000
#4	5001 $\leq$ Premium $\leq$ 125000
#5	Premium $\geq$ 125001

The following table should provide one with an indication of how the premium sizes were spread among the various rj's.

Premium Upper Boundary	Insured Count	Average Loss	CV^2
1000	99714	907	65.860
1100	15818	1250	62.606
1250	19515	1367	47.696
1500	25548	1632	31.696
1750	19311	1960	23.433
2000	14742	2345	24.807
2700	27986	2885	17.424
3000	8470	3320	13.802
4000	19568	4696	15.356
5000	12232	5579	10.878
7500	16834	7611	6.330
10000	8603	11176	4.679
15000	8858	15294	4.278
25000	7000	23616	2.710
35000	3129	35732	2.093
50000	2238	51152	1.369
60000	829	61635	1.518
75000	827	74536	1.063
100000	788	93348	0.926
150000	686	136603	0.936
200000	296	172329	0.527
300000	243	262231	0.609
400000	91	312604	0.428
500000	57	385255	0.551
750000	52	480894	0.373
1000000	13	586174	0.292
∞	24	798100	0.303
Total	313472		

Exhibit III Estimates of b and c

Set r_j's	A	SE(A)	B	SE(B)	R^2	$\hat{b}$	$\hat{c}$
1	62831	5746	.275	.020	.820	.340	-.048
2	85383	1609	.443	.038	.870	.181	.222
2 (j =1-11)	55378	3786	1.054	1.269	.071	.181	.740
2 (j =12-22)	52831	2728	.469	.046	.921	.127	.304
3	58997	2967	.305	.013	.976	.258	.037
3 (j =1-8)	55418	2226	.495	.176	.570	.182	.265
3 (j =9-15)	64572	8328	.291	.025	.964	.377	-.062
4	55539	4430	.445	.164	.401	.184	.220
5*	59311	13939	.350	.103	.721	.265	.068

*Used unweighted regression.

Exhibit IV Chi-Square Tests

Premium Range	Sample Size	b = 0 c = 0			b =.258 c =.037			b=.184 c=.220 pr<125000 / b=.263 c=.058 pr>125000		
		X^2	DF	P*	X^2	DF	P*	X^2	DF	P*
200001-205000	35	1.54	4	.82	5.20	3	.33	5.44	4	.16
175001-178000	41	4.02	5	.55	9.57	4	.05	7.18	4	.13
150001-153000	50	2.21	5	.82	5.68	5	.34	8.16	6	.23
125001-127500	58	17.61	7	.01	7.57	6	.27	2.82	7	.90
100001-102500	129	34.76	10	.00	15.76	10	.11	16.57	10	.08
90001-91500	64	14.27	7	.05	4.22	7	.75	2.38	8	.97
80001-81000	70	13.69	8	.09	11.54	10	.17	17.05	8	.03
70001-70700	48	4.18	5	.52	2.82	5	.73	5.44	5	.36
65001-65550	75	12.61	7	.08	13.34	8	.10	18.17	8	.02
60001-60600	73	8.76	8	.36	10.94	8	.20	14.17	8	.08
55001-55550	94	36.77	8	.00	15.22	8	.05	9.03	8	.34
50001-50500	103	25.82	8	.00	22.45	8	.00	19.63	8	.01
Subtotal		176.24	80	.00	124.32	82	.00	126.04	84	.00
45001-45450	99	34.28	8	.00	13.12	8	.11	7.82	8	.45
40001-40400	99	6.81	8	.44	6.68	8	.57	7.24	8	.51
35001-35350	119	76.36	10	.00	38.25	10	.00	15.01	10	.11
30001-30300	146	55.40	10	.00	31.45	10	.00	7.22	10	.71
25001-25250	195	118.57	10	.00	56.35	10	.00	12.03	10	.28
20001-20200	222	60.48	10	.00	31.94	10	.00	18.70	10	.04
15001-15075	148	46.82	10	.00	24.36	9	.01	14.96	10	.13
Subtotal		391.91	66	.00	202.15	65	.00	82.98	66	.07
Total		568.15	146	.00	326.47	147	.00	209.02	150	.00

*P = Probability that X^2 is greater than the observed X^2 if the hypothesis is true.

Exhibit V Comparison of Model and Empirical Excess Pure Premium Ratios.

Empirical Excess Pure Premium Ratios (NCCI 1980 Table M)

	Expected Losses					
Entry Ratio	25000	50000	75000	100000	150000	200000
0.25	.789	.771	.764	.760	.756	.753
0.50	.631	.592	.574	.562	.545	.537
0.75	.516	.458	.431	.414	.388	.373
1.00	.430	.360	.330	.310	.280	.260
1.25	.365	.290	.258	.238	.208	.184
1.50	.316	.239	.208	.187	.159	.133
1.75	.277	.200	.171	.154	.128	.098
2.00	.246	.171	.144	.128	.104	.074
2.25	.220	.148	.123	.110	.087	.057
2.50	.198	.129	.106	.094	.073	.044
2.75	.180	.114	.091	.080	.061	.033
3.00	.163	.101	.080	.069	.051	.026

Model Excess Pure Premium Ratios (the parameters are from regressions #4 and #5)

	Expected Losses					
Entry Ratio	25000	50000	75000	100000	150000	200000
0.25	.7849	.7708	.7653	.7623	.7533	.7523
0.50	.6326	.5972	.5812	.5717	.5422	.5361
0.75	.5216	.4699	.4453	.4304	.2879	.3766
1.00	.4381	.3758	.3458	.3275	.2810	.2670
1.25	.3734	.3051	.2724	.2525	.2074	.1926
1.50	.3222	.2511	.2175	.1973	.1561	.1417
1.75	.2809	.2092	.1759	.1562	.1197	.1062
2.00	.2471	.1762	.1440	.1252	.0934	.0810
2.25	.2190	.1499	.1192	.1014	.0739	.0628
2.50	.1954	.1287	.0996	.0831	.0593	.0494
2.75	.1754	.1113	.0840	.0687	.0482	.0394
3.00	.1583	.0970	.0714	.0573	.0396	.0317

Model Excess Pure Premium Ratios ($b = c = 0$)

	Expected Losses					
Entry Ratio	25000	50000	75000	100000	150000	200000
0.25	.7637	.7531	.7509	.7503	.7500	.7500
0.50	.5883	.5455	.5275	.5179	.5086	.5045
0.75	.4654	.3975	.3636	.3423	.3165	.3011
1.00	.3774	.2961	.2539	.2267	.1921	.1702
1.25	.3122	.2258	.1816	.1535	.1185	.0969
1.50	.2626	.1760	.1333	.1071	.0757	.0573
1.75	.2239	.1399	.1005	.0771	.0503	.0355
2.00	.1931	.1132	.0775	.0570	.0347	.0230
2.25	.1681	.0931	.0610	.0432	.0247	.0154
2.50	.1475	.0776	.0489	.0335	.0180	.0106
2.75	.1304	.0655	.0398	.0265	.0134	.0075
3.00	.1161	.0559	.0328	.0212	.0102	.0054

Exhibit VI Tests of Estimators for c (Equation 4.1)

(c, b)	Expected Losses	Obs	Average $\hat{c}$	S.D. of $\hat{c}$
(.1,.1)	250,000	5	.1004	.0720
"	"	25	.1002	.0326
"	"	100	.1005	.0153
"	"	400	.1013	.0086
"	1,000,000	5	.0957	.0672
"	"	25	.0973	.0305
"		100	.1014	.0155
"	"	400	.0986	.0077
"	5,000,000	5	.0973	.0670
"	"	25	.1001	.0308
"	"	100	.0992	.0218
"	"	400	.0988	.0084
	"			
(.03,03)	250,000	5	.0298	.0237
"	"	25	.0300	.0095
"	"	100	.0306	.0043
"	"	400	.0299	.0022
"	1,000,000	5	.0290	.0211
"	"	25	.0294	.0092
"	"	100	.0300	.0042
"	"	400	.0303	.0023
"	5,000,000	5	.0307	.0220
"	"	25	.0297	.0090
"	"	100	.0303	.0042
"	"	400	.0300	.0019

Exhibit VII Tests of Estimators for b

(c, b)	Expected Losses	Obs	Equation 4.3 Mean = 633 Std Dev= 5408: Average $\hat{b}$	S.D. of $\hat{b}$	Equation 4.4 Mean estimated from simulated observation Dev/Mean = 8.55: Average $\hat{b}$	S.D. of $\hat{b}$	Equation 4.4 Mean and Std Dev estimated from simulated observation: Average $\hat{b}$	S.D. of $\hat{b}$
(.1,.1)	250,000	5	.0957	.3689	.0364	.1749	.0612	.1340
"	"	25	.1035	.1878	.0759	.1203	.0714	.0791
"	"	100	.1207	.0981	.1017	.0798	.0768	.0477
"	"	400	.1191	.0554	.0999	.0441	.0710	.0281
"	1,000,000	5	.1058	.1945	.0787	.1080	.0783	.0979
"	"	25	.1069	.0931	.0956	.0624	.0897	.0528
"		100	.1048	.0384	.0961	.0273	.0920	.0231
"	"	400	.1081	.0197	.1006	.0145	.0932	.0121
"	5,000,000	5	.1074	.1504	.0873	.0847	.0868	.0835
"	"	25	.1031	.0503	.0956	.0390	.0944	.0387
"	"	100	.1078	.0358	.0995	.0252	.0982	.0249
"	"	400	.1033	.0111	.0993	.0098	.0979	.0098
(.03, .03)	" 250,000	5	.0246	.3518	.0115	.1420	.0152	.0958
"	"	25	.0263	.1197	.0185	.0901	.0236	.0550
"	"	100	.0309	.0687	.0287	.0582	.0217	.0311
"	"	400	.0365	.0438	.0334	.0388	.0263	.0169
"	1,000,000	5	.0290	.0718	.0228	.0575	.0252	.0484
"	"	25	.0299	.0333	.0284	.0287	.0284	.0241
"	"	100	.0316	.0160	.0304	.0143	.0298	.0120
"	"	400	.0308	.0083	.0304	.0076	.0284	.0061
"	5,000,000	5	.0302	.0329	.0283	.0288	.0282	.0285
"	"	25	.0314	.0134	.0302	.0116	.0300	.0114
"	"	100	.0315	.0078	.0306	.0072	.0299	.0071
"	"	400	.0303	.0047	.0296	.0044	.0292	.0043

Exhibit VIII The Effect of Parameter Uncertainty on Large Accounts

Excess Pure Premium Ratios

Expected Loss = 1,000,000

Entry Ratio	b=c=0.0	b=c=.01	b=c=.03	b=c=.05	b=c=.10
0.25	0.7500	0.7500	0.7500	0.7500	0.7503
0.50	0.5000	0.5001	0.5014	0.5040	0.5131
0.75	0.2552	0.2631	0.2784	0.2923	0.3219
1.00	0.0829	0.1004	0.1277	0.1494	0.1914
1.25	0.0204	0.0306	0.0514	0.0706	0.1110
1.50	0.0048	0.0085	0.0193	0.0320	0.0639
1.75	0.0010	0.0023	0.0070	0.0143	0.0370
2.00	0.0002	0.0006	0.0025	0.0064	0.0216
2.25	0.0000	0.0001	0.0009	0.0029	0.0128
2.50	0.0000	0.0000	0.0003	0.0013	0.0077
2.75	0.0000	0.0000	0.0001	0.0006	0.0047
3.00	0.0000	0.0000	0.0000	0.0003	0.0030

Expected Loss = 5,000,000

Entry Ratio	b=c=0.0	b=c=.01	b=c=.03	b=c=.05	b=c=.10
0.25	0.7500	0.7500	0.7500	0.7500	0.7501
0.50	0.5000	0.5000	0.5003	0.5016	0.5089
0.75	0.2500	0.2529	0.2659	0.2797	0.3105
1.00	0.0383	0.0680	0.1040	0.1297	0.1764
1.25	0.0006	0.0084	0.0314	0.0528	0.0968
1.50	0.0000	0.0006	0.0080	0.0200	0.0526
1.75	0.0000	0.0000	0.0019	0.0074	0.0287
2.00	0.0000	0.0000	0.0004	0.0027	0.0158
2.25	0.0000	0.0000	0.0001	0.0010	0.0089
2.50	0.0000	0.0000	0.0000	0.0004	0.0051
2.75	0.0000	0.0000	0.0000	0.0001	0.0030
3.00	0.0000	0.0000	0.0000	0.0001	0.0018

Notes

1. Simon, L. J. "The 1965 Table M." PCAS LII, 1965. p. 44.

2. ibid p. 14.

3. Skurnick, D., "The California Table L" PCAS LXI, 1974, P. 117.

4. Beard, R. E., Pentikainen, T. and Pesonen, E. Risk Theory, 2nd Edition, Chapman and Hall, 1977. p.18.

5. ibid p. 110.

6. Heckman, P. and Meyers, G. "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions." (submitted for publication) p.10.

7. Beard, R. E. et al op. cit. p 126.

8. Heckman, P. and Meyers, G. op. cit.

9. Patrik, G. "Estimating Casualty Loss Amount Distributions." PCAS LXVII 1980. p. 57.

10. Patrik, G. and John, R. "Pricing Excess-of-Loss Casualty Working Cover Reinsurance Treaties." Pricing Property and Casualty Insurance Products. 1980, Casualty Actuarial Society Call Paper Program. p. 399.

11. The tests are not independent since the estimates of b and c were based on the entire data base. But it is unlikely that excluding the data used in the tests would have produced significantly different estimates of b and c.

12. Meyers, G. "An Analysis of Retrospective Rating" PCAS LXVII 1980. p.110.