PARAMETER UNCERTAINTY IN THE
COLLECTIVE RISK MODEL

GLENN MEYERS
NATHANIEL SCHENKER

Abstract

This paper proposes a new version of the collective risk model that allows for uncertainty in selecting the expected number of claims and the claim severity distribution. We provide two different methods of estimating the parameters of this model. It is demonstrated by computer simulation that one must combine the experience of several insureds in order to accurately quantify parameter uncertainty. Tests on a very large sample of individual insured data show a significant improvement in the accuracy of the collective risk model when parameter uncertainty is taken into account. The tests do not show perfect agreement between the model and the empirical data, but the agreement is close enough to be useful in many applications.

1. INTRODUCTION

This paper discusses the role of collective risk theory in making insurance pricing decisions. Collective risk theory provides a way of calculating the probability that a loss arising out of an insurance contract will exceed a given amount. The calculation is done in terms of the underlying claim severity and claim count distributions. Related to this is the excess pure premium, which is the cost of insuring all losses above a given amount. Also of interest is the excess pure premium ratio, which is the excess pure premium divided by the expected loss.

Of principal concern is the relationship between the variance of the loss ratio and the size of the insured. A common assumption was stated by Simon [10, p. 44] as follows: "As the risk size increases, we expect the variance (of the loss ratio) to approach zero." Large insureds are typically written on a retrospective rating plan or an aggregate excess contract. The effect of this assumption would be that for all loss amounts greater than the expected loss, 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 for even a very large insured. His complaint would be that the expected loss cannot be estimated with the necessary precision.

Of interest is the distribution of actual losses about an unbiased estimate of the expected losses. Estimates of the expected losses vary because of many things such as future economic conditions, changes in loss development patterns, changes in the insured's operations, and changes in loss control procedures. Many of these changes are independent of the size of the insured. Thus, one should not expect the variance of the loss ratio to approach zero as the size of the insured becomes large.

The traditional models used in collective risk theory, such as the generalized Poisson distribution, do not allow for uncertainty in estimating the expected loss. This may be acceptable for the small insured, since the variance of the losses due to the random nature of the loss process is large compared to the variance due to the misestimation of the expected loss. As the insured increases in size, however, the variance due to the misestimation of the expected loss will dominate.

Below we will propose a version of the collective risk model that allows for uncertainty in estimating the expected loss. Most excess pure premium ratios

used in the United States are calculated from the well known "Table M" [7]. This table is based on empirically derived excess pure premium ratios. Because of the predominant use of this table we must first address the following question: 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 aggregate loss distributions, Simon [10, p. 14] wrote: "To avoid the difficulties and pitfalls of empiricism we should borrow from the 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."

Officially, it appears that those who favor the empirical approach have prevailed and Mr. Simon's advice has gone unheeded. Skurnick [11] constructed a table for the state of California based on empirical observations. In 1980, a National Council subcommittee constructed another table based on empirical observations. Mr. Simon's table was in effect for seventeen years.

While the use of empirical distributions does not require one to make the assumptions that are necessary with the theoretical approach, there are some fundamental problems with the empirical approach. It is generally agreed that the variance of the loss ratio distribution decreases as the size of the insured increases. It is also agreed that the variance of the loss ratio distribution increases as the average claim severity increases. But it is necessary to combine the experience of insureds of different sizes and average claim severities in order to get a sufficiently large sample. For example, the tables constructed by the National Council on Compensation Insurance in 1980 combined all insureds on a countrywide basis into expected loss ranges that include $25,000 to $50,000, $50,000 to $100,000 and $100,000 to $200,000.

Thus, the actuary is faced with the dilemma of choosing between two undesirable alternatives. If the empirical approach is chosen, a sample from a heterogeneous population is required. If the theoretical model is used, a number of simplifying assumptions must be made.

By proposing a mathematical model, we do not advocate abandoning the use of empirical data. Once a model has been constructed, one should form hypotheses which can be tested on live data. If statistical tests demonstrate 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 are many advantages to the theoretical approach. Since the size of the insured and the claim severity distribution are input variables, it is possible to adjust the parameters of the model to account for situations when there is little or no 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.

3. THE COLLECTIVE RISK MODEL

In this section, we propose a version of the collective risk model that allows for uncertainty in estimating the expected loss. Heckman and Meyers [2] discuss this model in great detail, and so a full mathematical description will not be given here. Much of what follows is taken from their paper and is included here for the sake of completeness.

We start by considering the Poisson distribution. In their classic book on risk theory, Beard, Pentikainen, and Pesonen [1, p. 18] give the assumptions underlying this distribution as follows:

1. Claims occurring in two disjoint time intervals are independent.
2. The expected number of claims in a time interval (t_1, t_2) depends only on the length of the time interval and not the initial value of t_1.
3. No more than one claim can occur at a given time.

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 from historical experience. Parameter uncertainty can arise from sampling variability and changes in claim frequency over time. A second option is to use the average number of claims for a group of insureds that are similar to the insured under consideration. Parameter uncertainty arises when some members of the group have different expected numbers of claims.

We now turn to specifying the claim count distributions that we shall use when parameter uncertainty is present. We shall adopt the following notation.

Let N be a random variable denoting the claim count,
λ be the expected number of claims, and
χ be a random variable with $E[\chi] = 1$ and $\text{Var}[\chi] = c$.

The claim count distribution 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 $\chi \cdot \lambda$.

We have the following relationships.

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

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

If χ has a Gamma distribution, the claim count distribution described by Algorithm 3.1 is the negative binomial distribution (Beard et al., [1, p. 110]). We shall use the negative binomial distribution to model the claim count distribution when parameter uncertainty is present.

We shall call the parameter c the contagion parameter for the claim count distribution. If $c = 0$, Algorithm 3.1 yields the Poisson distribution.

We now adopt the following notation.

Let Z be a random variable denoting claim severity,
   $S(z)$ be the cumulative distribution function for the claim severity, z, and
   X be a random variable denoting the aggregate loss for an insured.

Aggregate losses can then 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 variance of the aggregate loss distribution generated by Algorithm 3.2.

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

$$
\begin{aligned}
\text{Var}[X] &= E_N[\text{Var}(X|N)] + \text{Var}_N[E(X|N)] \\
&= \lambda \cdot \text{Var}[Z] + (\lambda + c \cdot \lambda^2) \cdot E^2[Z] \\
&= \lambda \cdot E[Z^2] + c \cdot \lambda^2 \cdot E^2[Z]
\end{aligned}
\tag{3.4}
$$

Implicit in the use of Algorithm 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. Under such conditions, errors in selecting the parameters of the claim severity are inevitable. To model parameter uncertainty in the claim severity distribution, we make the simplifying assumption that the shape of the distribution is known, but there is uncertainty in the scale of the distribution. Venter [12] makes the same assumption in his treatment of parameter uncertainty.

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 $\text{Var}[1/\beta] = b$. 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 variance for the aggregate loss distribution generated by Algorithm 3.3.

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

$$
\begin{aligned}
\text{Var}[X] &= E_\beta[\text{Var}(X|\beta)] + \text{Var}_\beta[E(X|\beta)] \\
&= E_\beta[(\lambda \cdot E(Z^2) + c \cdot \lambda^2 \cdot E^2(Z))/\beta^2] + \text{Var}_\beta[\lambda \cdot E(Z)/\beta] \\
&= (\lambda \cdot E[Z^2] + c \cdot \lambda^2 \cdot E^2[Z]) \cdot E[1/\beta^2] \\
&\quad + \lambda^2 \cdot E^2[Z] \cdot \text{Var}[1/\beta]
\end{aligned}
$$

$$= (\lambda \cdot E[Z^2] + c \cdot \lambda^2 \cdot E^2[Z]) \cdot (1 + b) + \lambda^2 \cdot E^2[Z] \cdot b$$
$$= \lambda \cdot E[Z^2] \cdot (1 + b) + \lambda^2 \cdot E^2[Z] \cdot (b + c + b \cdot c) \qquad (3.6)$$

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.

Let R denote the ratio $X/E[X]$. From Equations 3.5 and 3.6, we get the following result:

$$\text{Var}[R] = (1 + b) \cdot E[Z^2]/(\lambda \cdot E^2[Z]) + b + c + b \cdot c \qquad (3.7)$$

Under the above assumptions on parameter uncertainty, it is possible to calculate the cumulative probabilities and excess pure premium ratios in an efficient manner (Heckman and Meyers [2]). 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 values of b and c 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.

Interpreting the Model

As mentioned in the introduction, we are concerned with the relationship between the variance of the loss ratio and the size of the insured. Parameter uncertainty can perhaps best be understood in terms of how it affects this relationship.

In what follows, it will be helpful to recognize that $\text{Var}[R]$ is equal to the squared coefficient of variation of the loss ratio.

It can be seen from Equation 3.7 that this model implies a linear relationship between $\text{Var}[R]$ and $1/\lambda$.

Figure 1 illustrates the effect of parameter uncertainty on $\text{Var}[R]$. If $b = c = 0$, the aggregate loss distribution is the generalized Poisson distribution. In this case $\text{Var}[R]$ approaches zero as $1/\lambda$ approaches zero (or as the insured becomes larger). If we introduce parameter uncertainty in the claim count distribution, $\text{Var}[R]$ lies on a line parallel to that implied by the generalized Poisson distribution, with the variance approaching c as $1/\lambda$ approaches zero. If we add parameter uncertainty in the claim severity distribution, the slope of

the line increases by a factor of $1 + b$. The Var[R] approaches $b + c + b \cdot c$
as $1/\lambda$ approaches zero.

FIGURE 1.
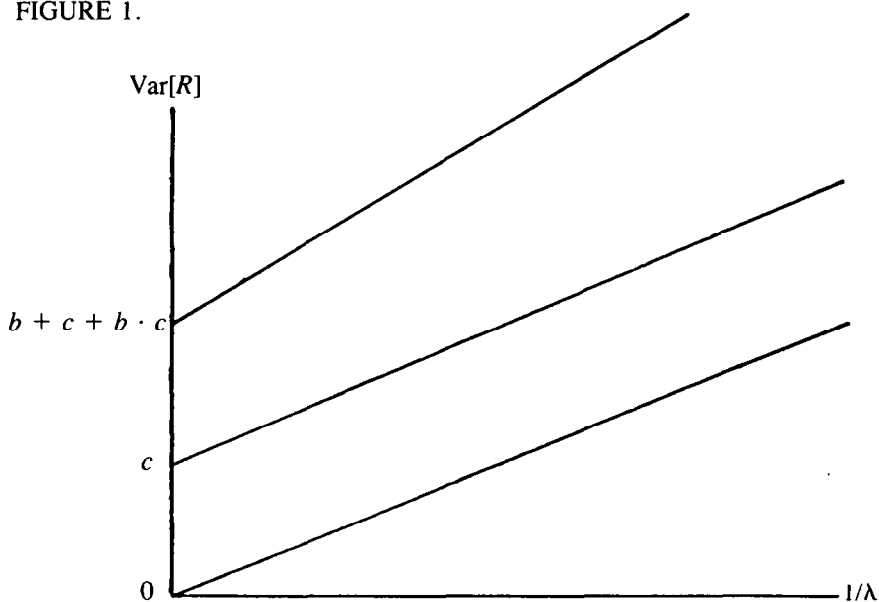
4. ESTIMATING THE PARAMETERS OF THE MODEL

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

The expected number of claims can be estimated from historical claim
frequencies and estimates of current exposure.

A complete discussion of estimating claim severity distributions is beyond the scope of this paper. In our work, we typically obtain claim severity distributions from bureau circulars, or we estimate them from company data using methods similar to those described by Patrik [8]. These claim severity distributions are often derived from experience other than that of the insured under consideration. For this reason, we adjust the scale of the distribution to match the average claim size that we project for the insured.

Before discussing our methods for estimating b and c, we should mention the work of Patrik and John [9]. They deal with parameter uncertainty by picking a finite set of 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.

We will describe two approaches for estimating b and c. The choice of estimators will depend on the kind of data available. The general idea underlying both of these approaches will be to evaluate an expression which "resembles" Var[N], Var[X] or Var[R]. Using Equations 3.2, 3.6 or 3.7, we can then set the expression equal to its expected value, which depends upon b and c. The estimates are then obtained by solving for b and c. The details are given in the appendices.

We first consider the case of a single insured for which we have r years of experience. We assume that all systematic adjustments of the data, such as trend and loss development, have been made.

The following estimators for b and c are derived in Appendix A.

For $j = 1,\dots,r$, let N_j be the claim count for year j. Let e_j be a number such that $e_j = K \cdot E[N_j]$ for each year j. K is a constant of proportionality. The number e_j represents either exposure or premium.

Let $\hat{\lambda}_1 = (1/r) \sum\limits_{j=1}^{r} N_j \cdot (e_1/e_j)$ and

$$V = \sum_{j=1}^{r} ((e_1/e_j) \cdot N_j - \hat{\lambda}_1)^2.$$

Then an estimator for c is given by

$$\hat{c} = \frac{V - (r-1)/r \sum\limits_{j=1}^{r} (e_1/e_j) \cdot \hat{\lambda}_1}{(r-1) \cdot \hat{\lambda}_1^2}. \qquad (4.1)$$

Let $X_1,\ldots,X_r$ be r independent aggregate loss amounts associated with $N_1,\ldots,N_r$, respectively, and let $A_j = X_j/N_j$ be average claim costs. Furthermore let

$$N = \sum_{j=1}^{r} N_j,$$

$\hat{\mu} = \sum\limits_{j=1}^{r} X_j/N$ be an estimate of $E[Z]$,

$\hat{\sigma}^2$ be an estimate of $Var[Z]$, and

$$W = \sum_{j=1}^{r} N_j \cdot (A_j - \hat{\mu})^2.$$

Then an estimator for b is given by

$$\hat{b} = \frac{W - (r-1) \cdot \hat{\sigma}^2}{(r-1) \cdot \hat{\sigma}^2 + \hat{\mu}^2 \cdot (N - (1/N) \sum\limits_{j=1}^{r} N_j^2)}. \qquad (4.2)$$

We used Monte Carlo simulation to test the accuracy of these estimators. Specifically, we selected a claim severity distribution (given in Exhibit I) along with $c = 0.1$ and $b = 0.1$ We then simulated aggregate losses and claim counts using Algorithm 3.3, and tested how well the estimates $\hat{b}$ and $\hat{c}$ compared with the selected b and c. We obtained $\hat{\sigma}^2$ by multiplying the coefficient of variation of our claim severity distribution by the estimate of $E[Z]$ obtained from the simulation. The results of these tests are given in Table 1.

TABLE 1

SIMULATED ESTIMATES OF b AND c

Expected Losses	r	Average $\hat{c}$	Std Dev $\hat{c}$	Average $\hat{b}$	Std Dev $\hat{b}$
250,000	5	.1004	.0720	.0364	.1746
"	25	.1002	.0326	.0759	.1203
"	100	.1005	.0153	.1017	.0798
"	400	.1013	.0086	.0999	.0441
1,000,000	5	.0957	.0672	.0787	.1080
"	25	.0973	.0305	.0956	.0624
"	100	.1014	.0155	.0961	.0273
"	400	.0986	.0077	.1006	.0145
5,000,000	5	.0973	.0670	.0873	.0847
"	25	.1001	.0308	.0956	.0390
"	100	.0992	.0218	.0995	.0252
"	400	.0988	.0084	.0993	.0098

The averages and standard deviations of the estimates for $r = 5, 25, 100,$ and 400 were based on 2000, 400, 100, and 25 trials, respectively. While the estimates of the errors are subject to simulation error, the above table suggests that one may need several hundred observations to accurately estimate b and c. Clearly, this is impossible for a single insured. We have repeated this experiment for different values of b and c and have gotten similar results.

Table 1 does not tell how much accuracy is necessary. To answer this, one must first ask how the collective risk model will be used. An almost certain use is the calculation of excess pure premium ratios. How much error one may tolerate for b and c will then depend on how much error one may tolerate for excess pure premium ratios.

Table 2 gives excess pure premium ratios for various sizes of insureds and various b's and c's. The method of calculating the excess pure premium ratios is that of Heckman and Meyers [2].

TABLE 2

EXCESS PURE PREMIUM RATIOS
Expected Loss = 1,000,000

Entry Ratio	$b = c = 0.0$	$b = c = .01$	$b = c = .05$	$b = c = .10$
0.50	0.500	0.500	0.504	0.513
1.00	0.083	0.100	0.149	0.191
1.50	0.005	0.009	0.032	0.064
2.00	0.000	0.001	0.006	0.022
2.50	0.000	0.000	0.001	0.007

Expected Loss = 5,000,000

Entry Ratio	$b = c = 0.0$	$b = c = .01$	$b = c = .05$	$b = c = .10$
0.50	0.500	0.500	0.502	0.509
1.00	0.038	0.068	0.130	0.176
1.50	0.000	0.001	0.020	0.053
2.00	0.000	0.000	0.003	0.016
2.50	0.000	0.000	0.000	0.005

Table 2 shows that significant differences in excess pure premium ratios result from different values of b and c. Taking this result along with the results indicated by Table 1, we are forced to the rather unpleasant conclusion that parameter uncertainty cannot be adequately quantified on the basis of the experience of a single insured.

If we are to estimate b and c from empirical data, it would appear that our only alternative is to combine the experience of several insureds, and assume that the same b and c are appropriate for all of them. It is this question that we now address.

5. ESTIMATING THE PARAMETERS OF THE MODEL—CONTINUED

If we were to combine the experience of several insureds to estimate b and c, we might consider using the above estimators and treating the combined observations as annual observations of a single insured. We feel, however, that this would be inappropriate for the following reasons.

First, a key assumption in the estimation procedure for c is that the expected number of claims is directly proportional to the measurement of exposure. While this assumption may be appropriate for a single insured, different insureds may have different exposure bases or different inherent claim frequencies.

Second, a key assumption in the estimation procedure for b is that the same claim severity distribution is appropriate for all observations of a single insured. Different insureds can expect to have different claim severity distributions.

Below we will give estimators for b and c. The data requirements for these estimators are as follows.

For each insured and each year we require three items:

1. exposure,
2. incurred losses, and
3. incurred claim count.

Some remarks concerning the data requirements are in order.

First, it will be assumed that the exposure is directly proportional to the expected claim count for each insured. The constant of proportionality may vary with the insured. Many exposure bases, such as payroll, are inflation sensitive. Thus, trends in the exposure base that do not reflect expected claim count should be removed.

Second, it will be assumed that the expected claim severity is the same for all observations of a single insured. The expected claim size need not be the same for all insureds. Incurred losses must be adjusted for trends in claim severity. The trend factors must be derived from external sources so as not to introduce bias in the estimates.

Third, every effort should be made to get the maximum number of observations. A minimum of two observations per insured will be required. We should strive for the maximum number of observations per insured and the maximum number of insureds.

The following estimators for b and c are derived in Appendix A:

T = number of insureds

r_i = number of observations for insured i

N_{ij} = number of claims for observation j of insured i

e_{ij} = exposure for observation j of insured i

X_{ij} = incurred loss for observation j of insured i

$A_{ij} = X_{ij}/N_{ij}$

Z_i = a random variable denoting claim severity for insured i.

For each i, let

$$\hat{\lambda}_{i1} = (1/r_i) \sum_{j=1}^{r_i} N_{ij} \cdot (e_{i1}/e_{ij}) \quad \text{and}$$

$$V = \sum_{i=1}^{T} \sum_{j=1}^{r_i} (N_{ij} \cdot e_{i1}/e_{ij} - \hat{\lambda}_{i1})^2.$$

Then, an estimator for c is given by

$$\hat{c} = \frac{V - \sum_{i=1}^{T} (r_i - 1)/r_i \sum_{j=1}^{r_i} (e_{i1}/e_{ij}) \cdot \hat{\lambda}_{i1}}{\sum_{i=1}^{T} (r_i - 1) \cdot \hat{\lambda}_{i1}^2}. \tag{5.1}$$

For each i, let

$$N_{i.} = \sum_{j=1}^{r_i} N_{ij},$$

$$\hat{\mu}_i = \sum_{j=1}^{r_i} X_{ij}/N_{i.} \text{ be an estimate for } E[Z_i],$$

$\hat{\sigma}_i^2$ be an estimate for $\text{Var}[Z_i]$, and

$$W = \sum_{i=1}^{T} \sum_{j=1}^{r_i} N_{ij} \cdot (A_{ij} - \hat{\mu}_i)^2.$$

Then, an estimator for b is given by

$$\hat{b} = \frac{W - \sum_{i=1}^{T} (r_i - 1) \cdot \hat{\sigma}_i^2}{\sum_{i=1}^{T} ((r_i - 1) \cdot \hat{\sigma}_i^2 + \hat{\mu}_i^2 \cdot (N_{i.} - (1/N_{i.}) \sum_{j=1}^{r_i} N_{ij}^2))}. \tag{5.2}$$

Note that Equations 5.1 and 5.2 reduce to Equations 4.1 and 4.2, respectively, when $T = 1$.

Many companies and rating bureaus do not have the data required for the

above estimators. However, all is not lost. We shall now show that it is possible to get rough estimates of b and c with quite a bit less data.

One of the predictions of this version of the collective risk model is the linear relationship of the squared coefficient of variation of the loss ratio with $1/\lambda$ (or equivalently, 1/expected losses). See Equation 3.7 and Figure 1. The following estimators of b and c will exploit this relationship.

The data requirements for these estimators will be loss ratios and premiums for each insured, and a single claim severity distribution that represents all the insureds. Divide the insureds into T groups of size r_i. For reasons stated above, we would prefer that the groups consist of multiple observations of the same insured. We shall say something about this in the next section. For observation j of group i, let

$e_{ij} =$ exposure (premium),
$X_{ij} =$ incurred loss, and
$R_{ij} = X_{ij}/e_{ij}$.

For each group i, let

$$\hat{\mu}_i = (1/r_i) \sum_{j=1}^{r_i} R_{ij},$$

$$E_i = (1/r_i) \sum_{j=1}^{r_i} 1/e_{ij}, \text{ and}$$

$$W_i = \sum_{j=1}^{r_i} (R_{ij} - \hat{\mu}_i)^2.$$

An estimate of the squared coefficient of variation for the ith group is given by the expression

$$\frac{W_i}{(r_i - 1) \cdot \hat{\mu}_i^2} .$$

An estimate of 1/expected losses for the ith group is given by the expression

$$\frac{E_i}{\hat{\mu}_i} .$$

Using linear regression, we can find an approximate relationship of the following form:

$$\frac{W_i}{(r_i - 1) \cdot \hat{\mu}_i^2} = \hat{A} \cdot \frac{E_i}{\hat{\mu}_i} + \hat{B}. \tag{5.3}$$

Estimators for b and c are then given by

$$\hat{b} = \hat{A} \cdot E[Z]/E[Z^2] - 1 \text{ and} \qquad\qquad (5.4)$$
$$\hat{c} = (\hat{B} - \hat{b})/(1 + \hat{b}). \qquad\qquad (5.5)$$

These estimators for b and c are derived in Appendix B. As mentioned above, we must select a single claim severity distribution that represents all insureds. In practice, it is questionable that this can be done. If estimates are obtained in this manner, then the variance of the aggregate loss distribution will be overstated for the low severity insured and understated for the high severity insured. Meyers [4] discusses several problems associated with this. It should be noted that the linear relationship between the squared coefficient of variation and 1/expected losses derived from Equation 5.3 will be preserved in the model if any reasonable claim severity distribution is selected. However, one should not put undue faith in the particular estimates of b and c.

6. TESTING THE MODEL

Thus far we have proposed a version of the collective risk model which allows for parameter uncertainty, and we have given ways to estimate the parameters for this model. We now turn to the crucial question, how well does it fit empirical data?

In 1980, a National Council committee assembled a large sample of individual insured data for the purpose of constructing a new table of excess pure premium ratios, otherwise known as Table M. This sample contained the standard premium and the incurred losses for all insureds during the policy year beginning July 1, 1973 for all states in which the National Council had jurisdiction.

The data was grouped by premium size, and the empirical loss ratio distributions were used to calculate the excess pure premium ratios for the smaller premium sizes. For those insureds with premium of \$200,000 or more, it was felt that the empirical excess pure premiums were not credible, and so a combination of modeling and empirical data was used.

After the table was completed, we requested and received from the National Council a tape containing this experience. This data forms the basis of our analysis.

Estimating the Parameters of the Model

Since we had only premium and loss data, we used Equations 5.4 and 5.5 with the e_{ij}'s representing standard premiums. Since we had only one observation for each insured, we chose our groups on the basis of premium size. Thus there are two main sources of parameter uncertainty. The first source is heterogeneity of insureds, and the second is differences in premium adequacy. This is consistent with the construction of the new Table M.

It should be noted that if we choose our groups on the basis of multiple observations on a single insured with the e_{ij}'s representing exposure, the sources of parameter uncertainty are quite different. Heterogeneity of insureds is not involved. Changes in the insured's operations, changes in economic conditions, changes in loss development patterns and other changes over time are the main sources of parameter uncertainty. Thus, estimates of b and c under these conditions could be quite different.

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 accident year 1975 developed to 42 months. We chose 42 months because it matched the average maturity of the NCCI data. We then changed the scale of the distribution to match the average claim size which was reported by the National Council for the policy year 1973–74. The resulting claim severity distribution is given in Exhibit I.

In choosing the groups, we put those observations with the lowest r_1 premiums in the first group, those observations with the next lowest r_2 premiums in the second group, and so on. The problem remained of choosing the r_i's, $i = 1, \ldots, n$ for the n groups. We observed that when the r_i's were equal for all i the variance of the residuals of the regression decreased as the premium increased. In statistical terminology, this is known as heteroscedasticity. We dealt with this problem in two ways. One way was to have r_i decrease as the premium increases. The other way was to use a weighted regression.

The weighted regression can be described as follows. If the model $Y = AX + B + \epsilon$ is to be fitted, but if 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 $Y' = A' + B'X' + \epsilon'$, 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_i's that we considered. Table 3 gives the resulting estimates of b and c.

TABLE 3

Estimates of b and c

Set of r_i's	$\hat{A}$	Std Err $\hat{A}$	$\hat{B}$	Std Err $\hat{B}$	R^2	$\hat{b}$	$\hat{c}$
1	62831	5746	.275	.020	.820	.340	$-.048$
2	85383	1609	.443	.038	.870	.181	.222
2(1–11)	55378	3786	1.054	1.269	.071	.181	.740
2(12–22)	52831	2728	.469	.046	.921	.127	.304
3	58997	2967	.305	.013	.976	.258	.037
3(1–8)	55418	2226	.495	.176	.570	.182	.265
3(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

There are several points that should be made about these estimates. First, negative estimates of b and c are possible; $\hat{b}$ will be negative whenever

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

$\hat{c}$ will be negative whenever

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

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 $\hat{A}$ suggests that random variation could explain the two negative estimates of c.

If a negative estimate of b is obtained, we suggest setting $\hat{b}$ equal to zero and setting $\hat{c} = \hat{B}$. If a negative estimate of c is obtained, we suggest setting $\hat{c} = 0$ and $\hat{b} = \hat{B}$. If $\hat{B}$ is negative, we suggest setting $\hat{b}$ and $\hat{c}$ equal to zero.

Second, the estimates of A and B vary by the set of r_i's chosen. Examination of the standard errors of the coefficient $\hat{A}$ indicates that this variation could be random. However, the variation in the estimates of B cannot be explained by random variation. It would appear that the estimate of B is decreasing as the size of the insured increases. This can be seen by comparing the pairs of estimates #2(1–11) with #2(12–22), #3(1–8) with #3(9–15), and #4 with #5. In all three comparisons, the estimate of B corresponding to higher premium

observations is lower than the estimate of B corresponding to lower premium observations.

This means that the sum, $b + c + b \cdot c$, is decreasing as the premium increases. As mentioned above, the division of parameter uncertainty between b and c is suspect for the estimators used. It seems that parameter uncertainty decreases as the size of the insured increases.

This seems to be a reasonable conclusion. Because of experience rating, one would expect the standard premium to be more accurate for large insureds than for small insureds.

Comparison of Expected with Actual Results

Using the estimates of b and c obtained above, it is possible to calculate the cumulative probabilities and the excess pure premium ratios implied by the model. We now 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 goodness of fit tests. A description of the chi-square goodness of fit test can be found in Hoel [3, p. 226]. We will then compare excess pure premium ratios predicted by the model with those of the new Table M.

We chose three sets of parameters for our testing. In the first test, we set $b = 0$ and $c = 0$ because it represents the case with no parameter uncertainty. For the second test, we chose the estimates $\hat{b} = .258$ and $\hat{c} = .037$ from regression #3 since it produced the highest R^2 over all the points. For the third test, we chose the estimates $\hat{b} = .184$ and $\hat{c} = .220$ when the premium was less than \$125,000 (regression #4), and $\hat{b} = .263$ and $\hat{c} = .068$ when the premium was greater than \$125,000 (regression #5). This enabled us to test if B decreases as the premium increases.

Since the variance of the loss ratio distribution changes with the size of the insured, we decided to estimate the distribution implied by the model and perform the chi-square test on each of several groups of insureds. Each group was to have a fairly narrow range of premium sizes. The results are given in Exhibit III.

No set of parameter values performed well when the premium was less than \$15,000. While the second and third sets of parameters performed better than the first, all sets severely underestimated the number of zero loss ratios. It appears that higher values of c are needed for small premium sizes.

It is difficult to note a pattern in the results of the chi-square tests on

individual groups. The chi-square test is simply not powerful enough to distinugish between the various sets of parameter values on individual groups. However, the chi-square test permits combining the results of independent tests. (Actually, the tests are not independent since the parameters b and c were estimated from all the observations. Since the number of observations used in estimating the parameters was far greater than the number of observations in each chi-square test, however, the tests are very nearly independent.) When the results are combined, a clear pattern emerges.

The results predicted by the second and third sets of parameters are better than the results predicted by the parameters $b = 0$ and $c = 0$. Allowing for parameter uncertainty significantly improves the performance of the collective risk model. The results predicted by the third set of parameters are better than the results predicted by the second set for the smaller premium sizes. This is consistent with our hypothesis that B decreases as the size of the insured increases.

Comparisons with the new Table M are given in Exhibit IV. Again we see that allowing for parameter uncertainty significantly improves the performance of the collective risk model. While the model does not fit the new Table M perfectly, it does come reasonably close.

Interpretation of the Results

The combined chi-square statistic calculated in Exhibit III indicates that we should reject the hypothesis that aggregate losses have the distribution predicted by the model. This shows that we have indeed made a number of simplifying assumptions.

This brings us back to the "Actuary's Dilemma." As noted above, the construction of an empirical Table M is suspect because of the necessity of using heterogeneous groups of insureds. It is extremely difficult to tell which is the more accurate. One must look to the applications in order to determine which to use.

Through the end of 1982, Table M was used to determine insurance charges in retrospective rating plans. The same insurance charges were used regardless of what claim severity distributions were appropriate for the insured and what accident limit was selected. Meyers [4] demonstrated that the claim severity distribution and the accident limit have a significant effect on the insurance charge. By examining Meyers' tables, one can see that these differences are much larger than the differences between the collective risk model with parameter uncertainty and the new Table M.

A revision of the retrospective rating plan is currently being considered by the National Council. This revision contains an adjustment for the "overlap" between the insurance charge and the excess loss premium factor. This adjustment was derived using the collective risk model with parameter uncertainty.

While the "Actuary's Dilemma" is not resolved, we see that the collective risk model with parameter uncertainty can make a significant contribution to the solution of today's problems.

7. LARGE INSUREDS

As demonstrated above, it is necessary to combine the experience of several insureds to get stable estimates of b and c. The methods given for estimating b and c assume that these parameters are the same for all insureds. It seems unlikely that b and c are the same for all insureds. For example, a stable company that has been working in the same line of business for many years should have a lower b and c than a company that has recently made material changes to its operations. A detailed examination of a company's operations may reveal additional sources of parameter uncertainty.

For small insureds, it may not be cost effective to conduct such an examination. Thus, it should be acceptable to assume that b and c are the same for all small insureds.

For large insureds, close examinations are routine. It seems quite possible that an underwriter could more accurately quantify parameter uncertainty on the basis of judgmental factors. However, skeptical actuaries respond that while underwriters are very sensitive to both the natural desire to sell insurance and aversion to risk, their quantitative estimates depend very much on what the competition offers. We regard it as an open question as to which method performs the best.

What is an actuary to do under these circumstances? First, we should provide estimates of b and c based on the combined experience of several large insureds. As Morell [6] remarked in his review of the first version of this paper, "We owe them at least that much." Furthermore, the data should contain several years of experience for each insured and the appropriate estimators for b and c should be used. Parameter uncertainty arising from heterogeneity between members of a group of insureds is not applicable for large account pricing.

If a close examination reveals additional sources of parameter uncertainty, sensitivity testing should be done to determine the effect of this uncertainty.

Quite often, the results of such testing can aid in designing a contract that is agreeable to all parties.

The remarks in this section are quite speculative. But they point out the need for extreme caution in using the collective risk model with large accounts.

8. CONCLUSION

This paper proposes a new version of the collective risk model that allows for uncertainty in selecting the expected number of claims and the claim severity distribution. We provide two different methods of estimating the parameters of this model. It is demonstrated by computer simulation that one must combine the experience of several insureds in order to accurately quantify parameter uncertainty. Tests on a very large sample of individual insured data show a significant improvement in the accuracy of the collective risk model when parameter uncertainty is taken into account. The tests do not show perfect agreement between the model and the empirical data, but the agreement is close enough to be useful in many applications.

9. ACKNOWLEDGEMENTS

APPENDIX A — DERIVATION OF EQUATIONS 4.1, 4.2, 5.1 AND 5.2

The equations derived in this appendix require the following data.

T = number of insureds
r_i = number of observations for insured i, $(r_i > 1)$
N_{ij} = number of claims for observation j of insured i
e_{ij} = exposure for observation j of insured i
X_{ij} = incurred loss for observation j of insured i
$\hat{\sigma}_i^2$ = an estimate of Var[Z_i], where Z_i is a random variable denoting claim severity for insured i.

Estimating c

Let λ_{ij} be the expected number of claims for insured i and observation j. Assume $\lambda_{ij} = K_i \cdot e_{ij}$. Then

$$\lambda_{ij} = \lambda_{i1} \cdot e_{ij}/e_{i1}. \tag{A.1}$$

It follows from Equations 3.2 and A.1 that

$$\text{Var}[N_{ij}] = \lambda_{i1} \cdot e_{ij}/e_{i1} + c \cdot (\lambda_{i1} \cdot e_{ij}/e_{i1})^2. \tag{A.2}$$

Let

$$\hat{\lambda}_{i1} = (1/r_i) \sum_{j=1}^{r_i} N_{ij} \cdot e_{i1}/e_{ij} \text{ and}$$

$$V = \sum_{i=1}^{T} \sum_{j=1}^{r_i} (N_{ij} \cdot e_{i1}/e_{ij} - \hat{\lambda}_{i1})^2.$$

Adding and subtracting λ_{i1} inside the parentheses gives us

$$V = \sum_{i=1}^{T} \left(\sum_{j=1}^{r_i} (N_{ij} \cdot e_{i1}/e_{ij} - \lambda_{i1})^2 - r_i \cdot (\hat{\lambda}_{i1} - \lambda_{i1})^2 \right).$$

Thus,

$$E[V] = \sum_{i=1}^{T} (r_i - 1)/r_i \sum_{j=1}^{r_i} (e_{i1}/e_{ij})^2 \cdot \text{Var}[N_{ij}].$$

Using Equation A.2 we get

$$E[V] = \sum_{i=1}^{T} (r_i - 1)/r_i \sum_{j=1}^{r_i} (e_{i1}/e_{ij}) \cdot \lambda_{i1} + c \cdot \sum_{i=1}^{T} (r_i - 1) \cdot \lambda_{i1}^2.$$

Solving for c we get

$$c = \frac{E[V] - \sum_{i=1}^{T} (r_i - 1)/r_i \sum_{j=1}^{r_i} (e_{i1}/e_{ij}) \cdot \lambda_{i1}}{\sum_{i=1}^{T} (r_i - 1) \cdot \lambda_{i1}^2}.$$

Equation 5.1 is obtained by substituting V for $E[V]$ and $\hat{\lambda}_{i1}$ for λ_{i1} ($i = 1, \ldots, T$). Equation 4.1 is simply Equation 5.1 with $T = 1$.

Estimating b

Let $\mu_i = E[Z_i]$,

$\sigma_i^2 = \text{Var}[Z_i]$,

$N_{i.} = \sum_{j=1}^{r_i} N_{ij}$, and

$A_{ij} = X_{ij}/N_{ij}$.

Let β denote the severity scaling factor.

Then $E[A_{ij}|N_{ij}, \beta] = (1/\beta) \cdot \mu_i$ and

$\text{Var}[A_{ij}|N_{ij}, \beta] = (1/\beta^2) \cdot \sigma_i^2/N_{ij}$.

Thus $E[A_{ij}|N_{ij}] = \mu_i$ and

$$\text{Var}[A_{ij}|N_{ij}] = (1 + b) \cdot \sigma_i^2/N_{ij} + b \cdot \mu_i^2. \tag{A.3}$$

Let

$$\hat{\mu}_i = \sum_{j=1}^{r_i} X_{ij}/N_{i.}, \text{ and}$$

$$W = \sum_{i=1}^{T} \sum_{j=1}^{r_i} N_{ij} \cdot (A_{ij} - \hat{\mu}_i)^2.$$

Adding and subtracting μ_i inside the parentheses gives us

$$W = \sum_{i=1}^{T} \left(\sum_{j=1}^{r_i} N_{ij} \cdot (A_{ij} - \mu_i)^2 - N_{i.} \cdot (\hat{\mu}_i - \mu_i)^2 \right).$$

Thus,

$$E[W|N_{ij}\text{'s}] = \sum_{i=1}^{T} \sum_{j=1}^{r_i} (N_{ij} \cdot \text{Var}[A_{ij}|N_{ij}] - (1/N_{i.}) \cdot N_{ij}^2 \cdot \text{Var}[A_{ij}|N_{ij}]).$$

Using Equation A.3 we get

$$E[W|N_{ij}\text{'s}] = (1 + b) \sum_{i=1}^{T} (r_i - 1) \cdot \sigma_i^2 + b \sum_{i=1}^{T} (N_{i.} - (1/N_{i.}) \sum_{j=1}^{r_i} N_{ij}^2) \cdot \mu_i^2.$$

Solving for b we get

$$b = \frac{E[W|N_{ij}\text{'s}] - \sum_{i=1}^{T} (r_i - 1) \cdot \sigma_i^2}{\sum_{i=1}^{T} ((r_i - 1) \cdot \sigma_i^2 + \mu_i^2 \cdot (N_{i.} - (1/N_{i.}) \sum_{j=1}^{r_i} N_{ij}^2))}.$$

Equation 5.2 is obtained by substituting W for $E[W|N_{ij}\text{'s}]$, $\hat{\mu}_i$ for μ_i and $\hat{\sigma}_i^2$ for σ_i^2. Equation 4.2 is simply Equation 5.2 with $T = 1$.

APPENDIX B—DERIVATION OF EQUATIONS 5.4 AND 5.5

The equations derived in this appendix require that individual observations be divided into T groups of size r_i. They also require the first moment, $E[Z]$, and the second moment, $E[Z^2]$, of an assumed claim severity distribution.

For observation j of group i, let

e_{ij} = exposure (premium),
X_{ij} = incurred loss, and
$R_{ij} = X_{ij}/e_{ij}$.

For each group i, let

$\mu_i = E[R_{ij}]$, and

$\hat{\mu}_i = (1/r_i) \sum_{j=1}^{r_i} R_{ij}.$

It follows from Equation 3.6 that

$$\text{Var}[R_{ij}] = \frac{\mu_i \cdot (1 + b) \cdot E[Z^2]}{e_{ij} \cdot E[Z]} + \mu_i^2 \cdot (b + c + b \cdot c). \tag{B.1}$$

For each group i, let

$$E_i = (1/r_i) \sum_{j=1}^{r_i} (1/e_{ij}), \text{ and}$$

$$W_i = \sum_{j=1}^{r_i} (R_{ij} - \hat{\mu}_i)^2.$$

Adding and subtracting μ_i inside the parentheses gives us

$$W_i = \sum_{j=1}^{r_i} (R_{ij} - \mu_i)^2 - r_i \cdot (\hat{\mu}_i - \mu_i)^2.$$

It then follows that

$$E[W_i] = (r_i - 1)/r_i \sum_{j=1}^{r_i} \text{Var}[R_{ij}]. \tag{B.2}$$

Combining Equations B.1 and B.2 we get

$$\frac{E[W_i]}{(r_i - 1) \cdot \mu_i^2} = \frac{(1 + b) \cdot E[Z^2]}{E[Z]} \cdot \frac{E_i}{\mu_i} + b + c + b \cdot c. \tag{B.3}$$

If one finds an approximate relationship of the form

$$\frac{W_i}{(r_i - 1) \cdot \hat{\mu}_i^2} = \hat{A} \cdot \frac{E_i}{\hat{\mu}_i} + \hat{B},$$

Equation 5.4 follows by equating $\hat{A}$ with the coefficient of E_i/μ_i in Equation B.3. Equation 5.5 follows by equating $\hat{B}$ with the constant term in Equation B.3 and using Equation 5.4.

REFERENCES

[1] R. E. Beard, T. Pentikainen, E. Pesonen (1977), *Risk Theory,* 2nd edition, Chapman and Hall.

[2] P. Heckman and G. Meyers (1983), "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX.

[3] P. G. Hoel (1971), *Introduction to Mathematical Statistics,* 4th edition, John Wiley and Sons.

[4] G. Meyers (1980), "An Analysis of Retrospective Rating," *PCAS* LXVII, p. 110.

[5] G. Meyers, and N. Schenker (1982), "Parameter Uncertainty in the Collective Risk Model," *Pricing, Underwriting and Managing the Large Risk,* Casualty Actuarial Society Discussion Paper Program, p. 253.

[6] R. Morell (1982), "Discussion of Parameter Uncertainty in the Collective Risk Model," *Pricing, Underwriting and Managing the Large Risk,* Casualty Actuarial Society Discussion Paper Program, p. 300.

[7] National Council on Compensation Insurance.

[8] G. Patrik (1980), "Estimating Casualty Loss Amount Distributions," *PCAS* LXVII, p. 57.

[9] G. Patrik, and R. John (1980), "Pricing Excess-of-Loss Casualty Working Cover Reinsurance Treaties," *Pricing Property and Casualty Insurance Products,* Casualty Actuarial Society Discussion Paper Program, p. 399.

[10] L. J. Simon (1965), "The 1965 Table M," *PCAS* LII, p. 1.

[11] D. Skurnick (1974), "The California Table L," *PCAS* LXI, p. 117.

[12] G. Venter (1982) "Transformed Beta and Gamma Distributions and Aggregate Loss," *Pricing, Underwriting and Managing the Large Risk,* Casualty Actuarial Society Discussion Paper Program, p. 395.

EXHIBIT I

THE CLAIM SEVERITY DISTRIBUTION

Loss Amount	Cumulative Probability
0.00	0.00000
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
277708.75	0.99996
395726.56	0.99998
593590.00	0.99999
791453.44	1.00000

Summary Statistics:

Severity Mean   =   632.56

Severity Standard Deviation   = 5704.69

EXHIBIT IIa

GROUPINGS USED FOR THE REGRESSIONS

Set of r_i's

i	#1	#2	#3	#4	#5
1	48600	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	29624	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				

Grouping	Range of Premium Sizes
#1	Premium $\geq$ 1000
#2	Premium $\geq$ 1000
#3	Premium $\geq$ 1000
#4	5000 $\leq$ Premium $\leq$ 125000
#5	Premium $>$ 125000

EXHIBIT IIb

GROUPINGS USED FOR THE REGRESSIONS

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

Premium Lower Boundary	Insured Count	Average Loss Ratio	Squared C. V. of Loss Ratio
1000	20444	.706	105.78
1100	26275	.689	45.16
1250	34277	.698	73.15
1500	26756	.724	45.29
1750	21024	.710	45.67
2000	31724	.711	37.34
2500	22588	.730	23.94
3000	29273	.745	20.28
4000	18709	.728	15.64
5000	26329	.800	13.99
7500	13759	.743	9.03
10000	14261	.764	5.38
15000	11779	.757	4.57
25000	4994	.733	2.69
35000	3843	.760	2.53
50000	2946	.722	1.49
75000	1414	.696	1.48
100000	1368	.655	1.02
150000	582	.684	.96
200000	539	.645	.76
300000	221	.614	.68
400000	130	.654	.49
500000	122	.538	.47
750000	55	.503	.43
1000000	60	.432	.39
Total	313472		

EXHIBIT III

CHI-SQUARE TESTS

Premium Range	Sample Size	$b = 0$ and $c = 0$			$b = .258$ and $c = .037$			$b = .184$ and $c = .220$ for Prem $\leq$ 125000; $b = .263$ and $c = .058$ for Prem $>$ 125000		
		χ^2	DF	P*	χ^2	DF	P*	χ^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 χ^2 is greater than the observed χ^2 if the hypothesis is true.

PARAMETER UNCERTAINTY

EXHIBIT IV

EMPIRICAL AND MODEL EXCESS PURE PREMIUM RATIOS

Empirical—NCCI 1980 Table M

Entry Ratio	Expected Loss					
	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—$b = .184$ and $c = .220$ for premium ≤ 125000
$b = .263$ and $c = .058$ for premium > 125000

Entry Ratio	Expected Loss					
	25000	50000	75000	100000	150000	200000
0.25	.785	.771	.765	.762	.753	.752
0.50	.633	.597	.581	.572	.542	.536
0.75	.522	.470	.445	.430	.288	.377
1.00	.438	.376	.346	.328	.281	.267
1.25	.373	.305	.272	.252	.207	.193
1.50	.322	.251	.218	.197	.156	.142
1.75	.281	.209	.176	.156	.120	.106
2.00	.247	.176	.144	.125	.093	.081
2.25	.219	.150	.119	.101	.074	.063
2.50	.195	.129	.100	.083	.059	.049
2.75	.175	.111	.084	.069	.048	.039
3.00	.158	.097	.071	.057	.040	.032

EXHIBIT IV (CONT.)

EMPIRICAL AND MODEL EXCESS PURE PREMIUM RATIOS

Empirical—NCCI 1980 Table M

Model—$b = 0$ and $c = 0$

Entry Ratio	Expected Loss					
	25000	50000	75000	100000	150000	200000
0.25	.764	.753	.751	.750	.750	.750
0.50	.588	.546	.528	.518	.509	.505
0.75	.465	.398	.364	.342	.317	.301
1.00	.377	.296	.254	.227	.192	.170
1.25	.313	.226	.182	.154	.119	.097
1.50	.263	.176	.133	.107	.076	.057
1.75	.224	.140	.101	.077	.050	.036
2.00	.193	.113	.078	.057	.035	.023
2.25	.168	.093	.061	.043	.025	.015
2.50	.148	.078	.049	.034	.018	.011
2.75	.130	.066	.040	.027	.013	.008
3.00	.116	.056	.033	.021	.010	.005