VOLUME LXXIV NUMBERS 141 AND 142

PROCEEDINGS

OF THE

Casualty Actuarial Society

ORGANIZED 1914

1987
VOLUME LXXIV
Number 141 — May 1987
Number 142 — November 1987

Library of Congress Catalog No. HG9956.C3
ISSN 0893-2980

Printed for the Society by
Recording and Statistical Corporation
Boston, Massachusetts

FOREWORD

The Casualty Actuarial Society was organized in 1914 as the Casualty Actuarial and Statistical Society of America, with 97 charter members of the grade of Fellow; the Society adopted its present name on May 14, 1921.

Actuarial science originated in England in 1792, in the early days of life insurance. Due to the technical nature of the business, the first actuaries were mathematicians; eventually their numerical growth resulted in the formation of the Institute of Actuaries in England in 1848. The Faculty of Actuaries was founded in Scotland in 1856, followed in the United States by the Actuarial Society of America in 1889 and the American Institute of Actuaries in 1909. In 1949 the two American organizations were merged into the Society of Actuaries.

In the beginning of the twentieth century in the United States, problems requiring actuarial treatment were emerging in sickness, disability, and casualty insurance—particularly in workers' compensation—which was introduced in 1911. The differences between the new problems and those of traditional life insurance led to the organization of the Society. Dr. I. M. Rubinow, who was responsible for the Society's formation, became its first president. The object of the Society was, and is, the promotion of actuarial and statistical science as applied to insurance other than life insurance. Such promotion is accomplished by communication with those affected by insurance, presentation and discussion of papers, attendance at seminars and workshops, collection of a library, research, and other means.

Since the problems of workers' compensation were the most urgent, many of the Society's original members played a leading part in developing the scientific basis for that line of insurance. From the beginning, however, the Society has grown constantly, not only in membership, but also in range of interest and in scientific and related contributions to all lines of insurance other than life, including automobile, liability other than automobile, fire, homeowners and commercial multiple peril, and others. These contributions are found principally in original papers prepared by members of the Society and published in the annual *Proceedings*. The presidential addresses, also published in the *Proceedings*, have called attention to the most pressing actuarial problems, some of them still unsolved, that have faced the insurance industry over the years.

The membership of the Society includes actuaries employed by insurance companies, ratemaking organizations, national brokers, accounting firms, educational institutions, state insurance departments, and the federal government; it also includes independent consultants. The Society has two classes of members, Fellows and Associates. Both classes are achieved by successful completion of examinations, which are held in May and November in various cities of the United States and Canada.

The publications of the Society and their respective prices are listed in the *Yearbook* which is published annually. The *Syllabus of Examinations* outlines the course of study recommended for the examinations. Both the *Yearbook*, at a $10 charge, and the *Syllabus of Examinations*, without charge, may be obtained upon request to the Casualty Actuarial Society, One Penn Plaza, 250 West 34th Street, New York, New York 10119.

JANUARY 1, 1988

*EXECUTIVE COUNCIL

DAVID G. HARTMAN	*President*
KEVIN M. RYAN	*President-Elect*
RICHARD H. SNADER	*Vice President-Administration*
CHARLES A. BRYAN	*Vice President-Development*
MICHAEL L. TOOTHMAN	*Vice President-Membership*
MICHAEL FUSCO	*Vice President-Programs*

THE BOARD OF DIRECTORS

**Officers*

DAVID G. HARTMAN	*President*
KEVIN M. RYAN	*President-Elect*

†Immediate Past President:

MICHAEL A. WALTERS	1988

†Elected Directors:

JAMES R. BERQUIST	1988
CHARLES F. COOK	1988
DAVID P. FLYNN	1988
MAVIS A. WALTERS	1988
IRENE K. BASS	1989
ALLAN M. KAUFMAN	1989
LEROY J. SIMON	1989
DAVID SKURNICK	1989
ALBERT J. BEER	1990
ALAN C. CURRY	1990
CHARLES L. MCCLENAHAN	1990
JEROME A. SCHEIBL	1990

* Term expires at 1988 Annual Meeting. All members of the Executive Council are officers. The Vice President-Administration also serves as Secretary and Treasurer.

† Term expires at Annual Meeting of year given.

1987 PROCEEDINGS
CONTENTS OF VOLUME LXXIV

1987 PROCEEDINGS
CONTENTS OF VOLUME LXXIV—CONTINUED

Page

1987 PROCEEDINGS
CONTENTS OF VOLUME LXXIV—CONTINUED

Page

NOTICE

Papers submitted to the *Proceedings of the Casualty Actuarial Society* are subject to review by the members of the Committee on Review of Papers and, where appropriate, additional individuals with expertise in the relevant topic. In order to qualify for publication, a paper must be relevant to casualty actuarial science, include original research ideas and/or techniques or have special educational value, and must not have been previously published or be concurrently considered for publication elsewhere. Specific instructions for preparation and submission of papers are included in the *Yearbook* of the Casualty Actuarial Society.

The Society is not responsible for statements of opinions expressed in the articles, criticisms, and discussions published in these *Proceedings*.

Volume LXXIV, Part 1 No. 141

PROCEEDINGS

May 5, 6, 7, 8, 1987

THE CONSTRUCTION OF AUTOMOBILE RATING TERRITORIES IN MASSACHUSETTS

ROBERT F. CONGER

Abstract

In Massachusetts, the past ten years have witnessed the evolution of an increasingly sophisticated system of methodologies for determining the definitions of rating territories for private passenger automobile insurance. In contrast to territory schemes in other states, which tend to group geographically contiguous towns, these Massachusetts methodologies have had as their goal the grouping of towns with similar expected losses per exposure, regardless of the geographic contiguity or non-contiguity of the grouped towns. This paper describes the evolving Massachusetts methodologies during that ten year period.

The paper includes the latest methodology, which was employed to establish rating territories for use in Massachusetts in 1986. That methodology evaluates by-town claim frequency and by-town claim severity separately and then combines the results. The claim frequency approach is to compile detailed insurance data by town, and to compare those actual observations to an a priori *model of the expected insurance losses in each town. The model and the actual observations are blended using empirical Bayesian credibility procedures. The claim severity analysis* *uses a two layer hierarchical empirical Bayesian method in which countywide and statewide severity data supplement less than fully credible town severity data. The combined results of the frequency and severity analyses serve as the basis for ranking the towns according to expected losses per exposure and for placing the towns into rating groups.*

The evolution of the current Massachusetts methodology chronicled in this paper has involved the development and exchange of ideas by a number of individuals in addition to the author. The other principal players have been Richard Derrig, William DuMouchel, Howard Mahler, Stefan Peters, Peter Siczewicz, and Richard Woll. Credit is due also to Ronald Dennis and Lesley Phipps, who played invaluable supporting roles.

1. INTRODUCTION

Classification of risks, including classification of risks by territory, plays an important part in the determination of private passenger automobile insurance premiums in the United States (Stern [29]). In Massachusetts, for example, an experienced driver in Boston may pay more than $400 for a package of compulsory liability coverages costing less than $200 in the territory with the lowest rates. In addition to the magnitudes of the premium differences that depend on risk classification, there are significant public policy issues related to classification (or categorization) of the driving public. As a consequence, private passenger automobile insurance risk classifications have long been a focus of debate in Massachusetts (Massachusetts Division of Insurance [16]) and elsewhere (SRI International [28]). As long ago as 1950, the electorate of Massachusetts specifically voted on a classification issue; in that year, a proposal to eliminate automobile insurance territorial rate variations was placed on the ballot as a referendum question (but was defeated). While the maintenance and pricing of automobile rating territories is just one of many classification issues, it is a very important one.

In recent years the debate about Massachusetts automobile insurance territories has shifted to the technical arena. Mathematicians, statisticians, and actuaries have labored to develop procedures that are practical and workable, and that produce territories best satisfying the criteria suggested nine years ago

(State Rating Bureau (SRB) [23]): equity, homogeneity, discrimination, reliability, stability and compatibility. Briefly, in the context of territory definitions, these criteria were defined as follows:

Equity	The costs of insurance should be distributed fairly among different classes of insureds. "Statistical" equity refers to pricing in accordance with expected losses, while social equity refers to public policy concepts of "fairness." The latter concept is viewed as a series of constraints that perhaps would require recombining statistically justifiable classification separations.
Homogeneity	All the towns in a territory should have approximately the same expected insured losses per car.
Discrimination	The probability of a town being placed in the wrong territory should be minimized.
Reliability	The index used to categorize a town should be a good estimator of the expected insured losses per car in the town.
Stability	The assignments of towns to territories should not change dramatically over time.
Compatibility	A single set of territory definitions should be established so as to be reasonably appropriate for each of the insurance coverages.

The satisfaction of these criteria generally has been sought through efforts to develop an effective way to estimate the expected insured losses per car in each town. These estimates provide a basis for identifying towns in which expected losses are similar, and for grouping towns which are as homogeneous as practicable. The evolution has yielded a territory review methodology that has several interesting features.

- A regular review, typically biennial, of all territory definitions.
- The use of detailed insurance data, by town, as the basic information underlying the determination of the town groupings.
- The development of a model that predicts variations in claim frequency among towns, and the use of empirical Bayesian credibility procedures to combine the predicted claim frequency patterns with the actual by-town claim frequencies.

- The implementation of an empirical Bayesian credibility procedure that estimates the average claim severity in a town by credibility weighting (a) the observed claim severity in the town with (b) the claim severity in the town's county, and (c) the statewide claim severity.
- The development of several measures of the homogeneity of various groupings of towns into territories.

This paper describes the latest territory review methodology and describes the development of that methodology. All of the evolutionary steps described in this paper reflect methods evaluated for and/or included in actual filed recommendations to the Massachusetts Commissioner of Insurance. Thus, while various methodological advances have been accomplished, the parties necessarily have observed a constraint that any methodology be sufficiently practical to include in a Massachusetts rate filing.

The details of the latest methodology, which is described in this paper, are set forth in a rate filing of the Massachusetts Automobile Rating and Accident Prevention Bureau[1] (MARB [14]) and in the resulting decision of the Commissioner (Massachusetts Division of Insurance [20]). This paper relies in part on that Bureau filing, which is quoted or paraphrased without specific attribution at several points in this paper (see Appendices A and B).

2. HISTORY

Automobile insurance rates have varied according to location of garaging for many years. Shortly after the turn of the century, automobile insurers recognized variations in accident frequency from one area to another and divided the United States into two rating territories (All-Industry Research Advisory

[1] The reader may be assisted by a brief description of the regulatory process that governs Massachusetts private passenger automobile insurance and a brief description of the parties involved. The Commissioner of Insurance, who is the state regulator, affirmatively establishes rates, territories, rating procedures, and so forth, effective January 1 each year. The Commissioner has statutory authorization to allow insurance companies to set rates competitively, but has chosen to retain the rate setting authority himself in each of the recent years, following a brief experiment with competitive rating in 1977. In establishing the various rating components, the Commissioner must rely on recommendations from participants in the annual rate hearing process. With regard to the establishment of territory definitions, three participants have offered the principal recommendations. First, the Massachusetts Automobile Rating and Accident Prevention Bureau (MARB), also known as the Massachusetts Rating Bureau, represents the insurance industry. Second, the State Rating Bureau (SRB), which is an arm of the Division of Insurance, the state regulatory body on insurance matters, participates routinely. Third, the Attorney General (AG) intervenes in the hearing process, ostensibly on behalf of the motoring public.

Council (AIRAC) [1]): Greater New York, Boston and Chicago; and Remainder of the United States.

By 1917, the country was divided into eleven rating territories (DuMouchel [3]): Greater New York; Chicago and St. Louis; Boston; Philadelphia; Providence; Baltimore, Washington, D.C., and Pittsburgh; Detroit, Indianapolis, and Milwaukee; Minneapolis and St. Paul; Alabama, Kentucky, Tennessee; Arkansas and portions of other states; Arizona and other states.

Over the years, the system of territories proliferated, and as the patterns of state definition of automobile insurance laws and state regulation of automobile insurance rates solidified by 1950, it became clearly appropriate for each state to have unique rates. In addition, most states were subdivided into a number of territories, as is the case today; the average number of territories per state is fourteen (AIRAC [1]).

The early territory definitions apparently were established largely by judgment, but typically many rating territories were subdivided into two or more statistical territories, so that possible alterations to the existing scheme of rating territories could be studied in a systematic fashion.

In recent years, various methods have been used in different states to review and revise territory definitions. Those methods are beyond the scope of this paper but are described in other sources (California Department of Insurance [2]; McDonald and Thornton (Texas) [24]; New Jersey Department of Insurance [25]; Rhode Island Ad Hoc Committee on Territorial Rating [26]; AIRAC [1]).

3. THE EVOLUTION OF MASSACHUSETTS METHODOLOGIES

Perhaps nowhere has the problem of establishing territory definitions been subjected to the frequent review and pace of methodological development that have occurred in Massachusetts over the past ten years. Several factors have contributed to this history, including:

- The availability of a long and continuing history of detailed insurance data by town, for each of the 351 cities and towns that comprise Massachusetts.[2] This data base provides ready building blocks for alternative territory schemes, and the continuity of reporting of town data facilitates regular reviews and revisions of such schemes.

[2] In Massachusetts, unlike some states, all land falls inside the boundaries of cities and towns. Note that references below to 360 "towns" include a subdivision of Boston into ten "towns" for automobile rating purposes.

- Regulatory and statutory pressures to flatten rate differentials between territories, which have led to an increased interest by insurers in at least knowing the indicated rates for each geographic cell of the state.
- Regulatory demands for "scientific" approaches to all aspects of ratemaking.

Although the start of the evolution of the current territorial review process was stimulated by the revision of territories that took effect in 1977, some mechanisms for the regular review of territories were in place prior to 1977.

As recently as 1976, two sets of Massachusetts automobile rating territories existed, one set for liability coverages and one set for physical damage coverages.[3] As if the existence of two sets of territories were not sufficiently confusing, liability territory 1 was charged the *highest* liability rates, while physical damage territory 1 was charged the *lowest* physical damage rates. Since 1977, the various parties have unanimously agreed that a single set of territories should apply to all coverages (the SRB's "compatibility" criterion), and that the potential marginal actuarial precision to be gained by maintaining separate territories did not merit the accompanying additional administrative costs and confusion. This position is supported by the fact that most drivers purchase physical damage coverages and increased limits liability coverage in addition to compulsory liability coverages.

Prior to the 1977 rate revision, the *methodologies* used for devising liability and physical damage rating territories also were independent (SRB [23]). For physical damage coverages, twenty-four territories had been established on a geographic basis similar to that used in other states currently. For liability coverages, towns were grouped together into six territories based on the similarity of historical loss pure premiums[4] for the two principal compulsory injury

---

[3] The existence of separate territory definitions for different coverages was due, at least in part, to the fact that the two different sublines were under the jurisdiction of two different insurance industry rating bureaus in that era.

[4] Loss pure premium is defined as: (a) the claims dollars associated with claims against policies insuring cars in the town, divided by (b) the number of exposures, or insured cars, in the town. All data—exposures, claim counts, and losses—are coded to the town in which the car is garaged (not, for example, to the town in which the accident occurs). As is fairly common in the actuarial techniques applied to classification issues, loss development and trend are ignored on the assumption that they will not have measurably different effects in the different towns.

coverages, no-fault and liability; the two coverages were combined into a single pure premium in a somewhat complicated fashion that is beyond the scope of this paper.[5] A classical credibility factor was assigned to each town's data, based on a full credibility standard of 1000 claims. For any town with less than full credibility, the historical town pure premium was credibility weighted against the underlying pure premium for the territory to which the town had been assigned previously. The resulting "formula pure premiums" were used to rank the towns and to group each town with other towns having similar formula pure premiums, so as to produce six territories. Finally, various constraints were imposed to prevent a town from moving too many territories in any one revision or reversing direction from its movement in the previous revision.

The term "territory" in the 1976 liability methodology, and in all methodologies adopted since then, was purely an historical convention; no geographical constraints were imposed on the selection of towns to be included in a territory. Thus, each of the six territories could contain a variety of non-contiguous towns from all parts of the state. This approach is potentially somewhat confusing to the motoring public, who might hold a more geographically-based concept of territory; in recent years, the Commissioner has been offered proposals for partial imposition of geographic constraints (Massachusetts Division of Insurance [17], AG [10]). Each of the reviews since 1977, however, has indicated substantial variations in pure premiums among neighboring towns. Thus, imposition of geographical constraints would carry a cost: a reduction in the claims-experience homogeneity of the resulting territories. The Commissioner, since 1977, has maintained the freedom from geographic constraint in grouping towns into territories, and many of the territories include towns from all corners of the state.

The 1977 Revision of Territories

The review of territories for 1977 (SRB [23]) indicated that the historical methodologies were failing to produce homogeneous territories comprised of towns having similar pure premiums; rather, the town pure premiums within a territory varied widely. Several methodological sources of the inadequacy of traditional review techniques were identified:

- Excessive reliance on geographical factors in the establishment of physical damage territories;

[5] No use was made of data for the property damage liability coverage. The bodily injury territory definitions applied to this coverage as well. The exclusion of PDL data apparently was attributable in large part to the frequent enactment of statutes changing the nature of this coverage.

- Reliance on a subset of liability coverages to formulate liability territories, particularly since the subset used (bodily injury coverages) was perceived in 1976 as being subject to relatively great volatility in claim severity;
- Inadequate credibility treatment; and
- Excessive application of constraints on town movements. The constraints applied included both direct constraints—actual restrictions on town movements—and indirect constraints, such as assigning the complement of credibility to the town's former territory.

For 1977, an entirely new algorithm was introduced by the Massachusetts State Rating Bureau (SRB [23]). The new approach diverged from past methods in several respects.

First, claim frequency[6] rather than pure premium data were used. The exclusion of claim severity data was justified on the basis of the relatively great variability that the SRB perceived in such data, and the difficulty of studying a phenomenon whose distributions are "poorly known, badly skewed, and difficult to estimate from samples of actual experience" (SRB [23]). Although the possibility of systematic variations in claim severity from town to town was not denied, apparently the value of any information in the historical severity data was believed to be overwhelmed by the instability introduced by the use of such data. Preliminary tests underlying the 1977 review indicated to the actuaries at the State Rating Bureau that the use of claim frequency alone produced satisfactorily discriminatory territories. Subsequent reviews (SRB [22]); MARB [13]; MARB [14]; see below) have developed methodologies for extracting claim severity information from the historical data without also capturing undesirable chance variations in severity. These reviews have indicated that, with the benefit of the new methodologies, claim severity patterns are quite significant and should be reflected in the analysis of town data; but these new methodologies had not been developed by 1976.

Second, the review for 1977 relied on claim frequencies for the physical damage coverages (comprehensive and collision) only; no liability data were used, even though the resulting territories applied to all coverages.[7] A combined

[6] Claim frequency is defined as: (a) the number of claims against policies insuring cars in the town, divided by (b) the number of exposures, or insured cars, in the town. See also the definition of loss pure premium, above.

[7] By 1977, a single insurance industry rating bureau (MARB) had jurisdiction over all coverages; as a result, a unified approach to territories could be implemented more readily than in prior years.

"frequency" was constructed as the sum of comprehensive and collision claim counts, divided by comprehensive exposures. Concerns with the stability of bodily injury data, particularly for small towns, apparently contributed to the decision to exclude these data; the impact of this concern was amplified by the difficulty at that time of identifying an appropriate data element to which the complement of credibility could be applied systematically. The property damage liability (PDL) data, as in earlier years, was tainted by the effects of numerous statutory coverage changes, and thus was excluded from the methodology. The SRB analysts tested the performance of the constructed frequency, however, and concluded that this constructed physical damage frequency could be used to establish a single set of territories that would be acceptably homogeneous for every coverage. Later analyses reached a different conclusion and developed approaches that could successfully employ data from all coverages.

Third, the graduated credibility approach used prior to 1977 was replaced by a decision to assign zero credibility to the 72 smallest towns (based on their exposure volume) and *full* credibility to all larger towns. The small towns were assigned judgmentally to the same territory as a nearby larger "mother" town having similar demographic, economic, and industrial characteristics. This approach represented a rejection of the former complement rather than a rejection of the former credibility formula itself. In prior liability reviews, the complement of credibility was assigned to the data for a town's existing territory. For 1977, the existing territories were seen as being too out-dated to perform this function. Further, the existing territories for physical damage had been based on geographical contiguity, and thus did not necessarily provide an appropriate point of departure for the development of territories based on expected losses. Finally, the prior approach was seen as being structurally too restrictive on town movements. The 1977 resolution of the credibility issue was not entirely satisfactory, however, in that it provided no partial credibility and provided no systematic basis for the treatment of the "noncredible" towns. These issues were the focus of considerable analysis in subsequent reviews.

Fourth, as in the review for 1976 liability territories, the review for 1977 ranked towns according to the selected data element (in this case, the constructed physical damage claim frequency) and then towns having similar values were combined into territories. The review for 1977, however, introduced a more systematic method (which is beyond the scope of this paper) for deciding where to make the cutoff between one group of towns and the next. The result was one set of twenty-four territories used for all coverages[8] in 1977.

[8] Except for a few coverages that have rates not varying by territory.

Finally, the numerous constraints on town movements were removed, and as a result many towns were affected sharply by the territory reassignments. In later reviews, constraints were reimposed. These constraints were intended primarily to avoid sudden rate changes.

The method used for 1977, while lacking many of the important features of the later methodologies, can be credited with four significant achievements. First, it produced territories that were more homogeneous than the predecessor territories. Second, it highlighted the potential perils of including claim severity results in an assessment of the claims experience of smaller towns. Third, it pointed to the need for a credibility procedure that could deal with the small towns. Finally, and more generally, by dislodging the embedded process, the review of 1977 served to stimulate the ongoing research that followed.

The MARB Review for 1981 Territories

The 1977 territories remained intact through 1980. During 1980, the staff of the Massachusetts Automobile Rating and Accident Prevention Bureau (MARB), working with the Class-Territory Subcommittee of the Bureau's Private Passenger Actuarial Committee, conducted an extensive review of the data that had emerged since the 1977 revision, a review of the methods used in the 1977 revision, and research into possible methodological improvements. That research and review culminated in a filing (MARB [11]) that recommended a revision to the territory definitions based on a method that addressed some of the perceived shortcomings of the techniques used to construct the 1977 territories and that utilized the latest data. The key aspects of that proposed method are discussed below.

The MARB proposal for 1981 continued to rely on town-to-town differences in claim frequency rather than on town-to-town pure premium patterns. For each coverage, each town was assigned a severity equal to the statewide severity. A synthetic pure premium for the town was calculated as the product of (a) the town claim frequency, and (b) the statewide claim severity:

$$PP_{t,c} = Y_{t,c} \times X_c,$$

where $Y_{t,c}$ is the claim frequency for town t, coverage c;

X_c is the statewide average claim severity for coverage c; and,

$PP_{t,c}$ is the synthetic pure premium for town t, coverage c.

The inclusion of the statewide average claim severity served only to introduce a measure of the relative importance to overall premium of the various coverages. This approach, then, continued to ignore any town-to-town differences in claim severity. As in the 1977 review, the practitioners at this time believed that claim frequency effects explained most of the significant variation in pure premiums. The exclusion of the severity information was also based on concerns about the instability of the severity data, and on the absence of a credibility or modeling approach capable of separating information from noise in the severity data. While later reviews filled this void and indicated the significance of severity differences between towns, the later reviews also confirmed that the claim frequency effects were the dominant elements in defining town-to-town differences in pure premiums.

A major difference between the MARB proposal for 1981 and its predecessor methodologies was the inclusion of data for all the coverages for which rates varied by territory.[9] The use of data for all coverages has been retained in subsequent territory reviews. The MARB cited several reasons for this change in approach. First, public policy considerations seem to indicate, *a priori*, that the motorists in a town ought to bear *more* responsibility, not less, for the at-fault (liability) claims than for the physical damage claims; thus, the liability coverages ought to be returned to the territory review process. Second, the review for 1981 indicated that liability claim frequency patterns among towns did *not* parallel physical damage claim frequency patterns (contrary to the conclusions implicit in the preceding methodology), and thus that physical damage data could not be used as a proxy for all coverages. Third, the review indicated that, contrary to prior expectations, instability in liability claim frequencies was not a serious problem, so that there was no need to exclude them. Fourth, the statutory definition of PDL had finally stabilized (in 1977), so a usable data series for that coverage could, at last, be compiled. Fifth, the liability coverages are too large a component of overall rates to be ignored. Finally, the MARB review for 1981 introduced an empirical Bayesian credibility procedure that seemed to be capable of accommodating any inherent variations in claim frequencies. The several coverages were incorporated in the territory review process for 1981 by creating an overall average synthetic pure premium for each town that is simply the exposure-weighted average of the synthetic pure premium for each coverage:

$$I_t = \frac{\sum_c E_{t,c} \times PP_{t,c}}{\sum_c E_{t,c}},$$

[9] Compulsory Bodily Injury Liability (known as coverage A-1), compulsory No-Fault BI (A-2), compulsory Property Damage Liability (PDL), Collision, and Comprehensive.

where $E_{t,c}$ is insured exposures for town t, coverage c;

$PP_{t,c}$ is the synthetic pure premium for town t, coverage c (see above); and,

I_t is the all-coverages synthetic pure premium for town t.

This formula not only returns liability data to the analysis; it actually accords them dominant weight (since the insured exposures are greater for the compulsory liability coverages than for the optional physical damage coverages). This weighting scheme simply reflects the contribution of each coverage to overall premium rather than any conclusion that liability data are inherently more suitable for territory analyses.

A major area reviewed for 1981 was the treatment of credibility and the element to which the complement of credibility is assigned. Concerns with these aspects of the 1977 review included the absence of any systematic basis for assigning a complement to non-credible towns; the determination of a point of full credibility; and, the absence of any partial credibility treatment.

The MARB review for 1981 introduced a significant new element to supplement actual town data, to the extent town data were judged to be less than fully credible. As described above, the 1976 procedure assigned the complement of the town credibility to data from the town's previous territory, and the 1977 procedure judgmentally assigned the indications from a nearby "mother" town to a town whose data was judged not credible. In its proposal for 1981, the MARB introduced a claim frequency *model* that was assigned the complement of the town's credibility. This model estimated the claim frequency (or, more properly, the all-coverages synthetic pure premiums, I_t) in each town as a linear function of traffic density[10] in each town. For each town, the model M_t was calculated as:

$$M_t = aD_t + b,$$

where D_t is the town "traffic density"

$= E_t \div$ Road-miles in town;

E_t is the PDL exposure in the town;

a and b are regression coefficients; and,

M_t is the model synthetic pure premium for the town.

[10] The relationship of traffic density to geographical variations in insurance experience has been observed in the literature (e.g., HLDI [9]; AIRAC [1]), as well as in some of the methodologies used in other states.

The regression parameters a and b were calibrated by a weighted linear least squares regression of I_t against D_t (weighted by compulsory coverage exposures in each town). A similar (but not identical) model has been used in all subsequent reviews.

The "traffic density" variable does not measure all components of traffic density. The numerator includes only a count of vehicles insured in a town. Unfortunately, reliable vehicle count data are not available from the Registry of Motor Vehicles, so the insured exposures were utilized, and any town-to-town variations in compliance with the state's compulsory insurance laws are assumed away. This is not perceived as a major modeling problem in Massachusetts. The vast majority of motorists (on the order of 95%) do purchase compulsory insurance. Furthermore, the insurance statistical plan does properly match insured exposures and insured losses, so that any systematic patterns of coverage should be captured by other elements of the analysis. The traffic density variable also omits the effect of one town's residents driving in another town. This omission was purposeful, as there was no intent to directly attribute to residents of one town the effects of congestion caused by non-resident drivers. Thus, while D_t is not traffic density in a wholly traditional sense, the MARB concluded that it was adequate and appropriate for the task at hand.

The calculated traffic densities vary significantly—by a factor of 50—from town to town, as illustrated in Exhibit 2 for a sample of towns, and the regression relationship explained a significant portion of the town-to-town variations[11] in I_t.

Other explanatory variables were explored. For the most part, these variables related to the size or socio-economic characteristics of a town. It did not prove possible at that time to identify a variable for which data were available and that contributed meaningfully to the explanatory power of the regression.

M_t, then, is an estimate of the town's claim frequency based on a modeling process; I_t reflects the actual claim frequency. The analysis utilized I_t, to the extent credible, and assigned the complement of the credibility to M_t.

---

[11] Boston data did not fit the regression relationship and thus were omitted from the calibration of regression parameters. The assignment of the Boston subdivisions to territories was judgmental, placing each section of Boston in an independent territory, as had been done for 1977.

EXHIBIT 1

1985 MASSACHUSETTS PRIVATE PASSENGER BASE RATES FOR EXPERIENCED OPERATORS

	Coverage				
Territory*	BI Liability (A-1)	No-Fault BI (A-2)	PDL	Collision	Compre-hensive
1	$ 43	$10	$113	$147	$ 67
2	46	11	122	154	68
3	48	12	123	153	69
4	49	12	127	158	71
5	55	14	129	159	74
6	54	13	134	162	73
7	55	14	142	167	76
8	59	15	146	172	85
9	62	16	151	175	84
10	60	15	155	187	94
11	64	16	157	179	87
12	73	19	164	187	91
13	75	19	172	195	110
14	74	19	175	200	125
15	75	19	177	217	129
16	83	22	189	234	154
17	63	16	155	190	97
18	77	20	179	226	123
19	84	22	182	238	129
20	78	20	175	229	125
21	107	27	204	310	158
22	117	31	229	329	158
23	89	23	195	283	152
24	72	19	174	221	115
25	82	21	189	238	137
26	91	24	212	259	159

* Territories 17–26 are subdivisions of Boston.

EXHIBIT 2

EXAMPLE OF TRAFFIC DENSITY CALCULATIONS

Town Name	(1) 1983 PDL Exposures	(2) Road Miles	(3) Traffic Density (1) ÷ (2)
Hampden	2,968.9	57	52.1
Holland	914.0	37	24.7
Montgomery	374.4	31	12.1
Tolland	165.8	38	4.4
Wales	641.7	28	22.9
Amherst	10,022.2	121	82.8
Easthampton	8,556.3	83	103.1
Northampton	13,633.4	178	76.6
South Hadley	8,218.0	99	83.0
Ware	4,745.5	121	39.2
Belchertown	4,679.7	147	31.8
Hadley	3,034.7	80	37.9
Hatfield	2,023.0	59	34.3
Huntington	1,098.6	54	20.3
Williamsburg	1,492.7	48	31.1
Chesterfield	555.1	56	9.9
Cummington	439.5	62	7.1
Goshen	403.7	43	9.4
Granby	3,104.1	70	44.3
Middlefield	212.1	37	5.7
Pelham	658.7	43	15.3
Plainfield	294.4	49	6.0
Southampton	2.527.3	71	35.6
Westhampton	678.1	49	13.8
Worthington	622.1	64	9.7
Cambridge	30,201.7	142	212.7
Lowell	37,722.5	239	157.8
Everett	15,385.0	63	244.2
Malden	23,400.1	107	218.7
Medford	26,637.5	151	176.4
Newton	44,546.0	311	143.2
Somerville	26,169.1	105	249.2
Waltham	27,703.2	152	182.3
Watertown	17,042.1	77	221.3
Arlington	25,049.7	121	207.0
Belmont	14,010.8	82	170.9
Chelmsford	19,038.3	186	102.4
Concord	9,960.8	109	91.4
Dracut	12,546.6	112	112.0
Framingham	35,207.4	233	151.1
Hudson	9,345.9	79	118.3
Lexington	18,274.3	153	119.4
Marlborough	17,001.5	130	130.8
Melrose	15,376.9	82	187.5
Maynard	5,420.1	41	132.2

The credibility-weighted formula pure premium, F_t, for each town was calculated as

$$F_t = Z_t I_t + (1 - Z_t) M_t,$$

where Z_t is the credibility assigned to the data for town t.

Finally, the MARB review for 1981 territories introduced empirical Bayesian credibility procedures to assess the credibility to be assigned to the actual town data. Conceptually, the procedure treats the model pure premiums, M_t, as a "prior" estimate of the town experience, and the calculated synthetic pure premiums, I_t, as a subsequent observation. The credibility assigned to town data, I_t, was

$$Z_t = \frac{P_t}{P_t + K},$$

where P_t is an estimate of the town premium, and

K is the empirically determined credibility constant.

The credibility constant, K, is the ratio:

$$\frac{\text{an overall measure of year-to-year variations in town experience}}{\text{a measure of the extent to which actual town data, } I_t \text{, deviate consistently from the model, } M_t}.$$

This same *conceptual* formulation of K has been used in the subsequent territory reviews, although the actual procedures for estimating K have changed.[12] In each of these reviews, the derivation of K (or rather, the numerator and denominator of K) has relied on empirical methods that utilize the actual numerical values of the prior estimates and the observations.

The derivation of the credibility constant is beyond the scope of this paper (but see MARB [11]). The following interpretations, however, may be placed on the credibility formula and formula for K (see, for example, Hewitt [7]; and Hickman [8]):

(1) The magnitude of K is affected directly by the extent to which the density model, M_t, fits the actual data, I_t. If the model fits well, then the credibility algorithm concludes that little additional information is available from I_t. The denominator of K is small, K is large, and the credibilities assigned to I_t are relatively small.

[12] Only the current procedures for estimating K are detailed in this paper.

(2) Conversely, if the model, M_t, fits the data poorly, then the denominator of K is large, K is small, the credibilities assigned to I_t are relatively large, and the weights assigned to M_t are relatively small.

(3) If the town experience, I_t, varies significantly from year to year, the formulation concludes that I_t should not receive much weight. The numerator of K is large, K is large, and the credibilities assigned to I_t are small.

(4) The credibility formula structurally resembles the familiar $Z = P/(P+K)$ formula, which assigns more credibility to larger towns.

The factors described in (1), (2), and (3) are relative, not absolute. This highlights a major difference between the Bayesian credibility procedures used here and classical credibility: in the approaches used here, the credibility assigned to a set of data depends not only on the characteristics of that data, but also on the characteristics of the information that will be accorded the complement of the credibility.

The MARB proposal for 1981 continued the procedure of grouping together towns with similar values of the one-dimensional index (in this case, F_t) chosen to reflect town claims experience, although the details of the grouping procedure were somewhat different than in prior years.[13] Like the procedure used for 1977, the result was twenty-four rating territories: Territory 1 was the lowest rated territory, Territory 14 was the highest rated non-Boston territory, and Territories 15–24 were the ten subdivisions of Boston (not ranked in any particular order). Constraints on the movements of towns from their old territory assignments were reintroduced; however, restrictions applied only if an otherwise-indicated reassignment of a town would produce an unacceptably large rate change.[14]

In addition to identifying aspects of the territory analysis procedure in which methodological changes were needed, and proposing such changes, the data analyses undertaken in connection with the MARB proposal for 1981 territories indicated that the claims experience for towns shifted with sufficient rapidity that territory realignments should be evaluated regularly—preferably every other year.

[13] The details of the grouping procedure were virtually identical to those used by the Commissioner in subsequent years and in the MARB proposal for 1986, described below.

[14] Exhibit 1 displays the 1985 base rates by territory for experienced drivers, and provides a perspective on the rate implications of changing territories.

The State Rating Bureau recommendations for 1981 (as described in Massachusetts Division of Insurance [17]) concurred in the need for an updating of the 1977 territories, but did not embrace the methodological changes proposed by MARB. Rather, the SRB proposed either: (a) a simple updating of the 1977 territories based on later data, or (b) an updating of the town rankings based on later data and the introduction of a "territory within region" concept. Under this concept, each territory would be comprised of all towns having similar claims experience *and* located within a common geographic region of the state.

The Commissioner of Insurance, faced with this methodological dispute, chose the simple updating for 1981 and directed the parties to undertake a cooperative review and development of methodological changes (Massachusetts Division of Insurance [17]).

Review for 1982 Territories

For the development of 1982 territories, the parties did join in a cooperative effort, as well as continuing independent research efforts. Not the least of these research efforts was a master's thesis by one of the State Rating Bureau staff members (Siczewicz [27]). In this joint study for 1982, the work of Siczewicz provided most of the technical refinements to the treatment of credibility that had been developed in the MARB proposal for 1981. In general, the joint MARB-SRB-AG components of the proposal for the modification of rating territories for 1982 bore a strong resemblance to the MARB proposal for 1981. The major differences are summarized below.

In the review for 1982, the tabulation of the actual town data claims experience, the calibration of the density model, and the empirical determination of credibility parameters were conducted for each coverage separately, rather than for all coverages combined. This separate approach was intended to allow the credibility procedure to deal more fully with any differences between coverages in the stability of town claims experience. The town claim frequency (by coverage) was not converted into a pure premium at this stage, but rather was expressed as a claim frequency index.[15]

With the benefit of further study, the density model of claim frequency patterns was expanded to include two additional explanatory variables besides traffic density: a measure of the mix of driver classes in a town, derived from

[15] Claim frequency index = Town claim frequency/Statewide claim frequency.

the average classification relativity (*ACRF*) in the town; and a dummy variable that allowed the aberrant data in Boston to be included in the parameterization calculations without distorting the density regression coefficient.[16]

The *ACRF* variable is intended to reflect the fact that the claim frequency of the insureds in a town is affected by the mix of driver classifications in the town. For example, a town populated solely by senior citizens would be expected to have a lower claim frequency than an otherwise similar town comprised solely of operators with less than three years of experience. Actual towns fall somewhere between these extremes.

The ten subdivisions of Boston were observed to have claim frequencies significantly different from the claim frequencies of the 350 remaining towns in Massachusetts. These differences were not explained by the density and the class mix variables. In fact, differences *between* the ten subdivisions of Boston depart from the patterns which would be predicted by the traffic density model.

The form of the model proposed for 1982, and still utilized today, is:

$$\begin{aligned}\text{Model Frequency Index}_{c,t} = {} & A_{0,c} \\ & + A_{1,c} \times \text{Density}_t \\ & + A_{2,c} \times ACRF_{c,t} \\ & + A_{3,c} \times \text{Boston Dummy}_t;\end{aligned}$$

where A_0, A_1, A_2, A_3 are the regression parameters;
Boston Dummy = 1 in Boston,
0 elsewhere; and,
c refers to coverage, t refers to town.

The credibility procedure was refined so that the credibility parameters and the model regression parameters were determined simultaneously. As noted above in the discussion of the MARB proposal for 1981, the value of the credibility parameter depends on the characteristics of the claim frequency model, since the credibility parameter K depends on differences between the model and actual claim frequencies. In turn, in the review for 1982, the model regression parameters were determined by a weighted least squares regression, where the weights depended on the credibility assigned to the towns' data.

[16] The abnormalities of the Boston data were attributed to the high density of commercial vehicles in Boston (commercial vehicles are not captured in the traffic density variable or in any of the insurance data used in the territory analyses) and the small geographic size of the Boston subdivisions, which suggests that most driving is between subdivisions, not within a single subdivision.

In a broad sense, the use of regression weights dependent on the credibility assigned to a town is similar to the use of exposures, since exposures are a key factor in calculating credibility for a town.

The credibility for a particular town utilized a formula similar to 1981:

$$Z_{c,t} = H_{c,t}/[H_{c,t} + (\tau_c^2/\sigma_c^2)]$$

where H_t = exposures divided by claim frequency;

τ^2 is a measure of year-to-year variation in claim frequency; and,

σ^2 is a measure of the extent to which actual claim frequencies differ from model claim frequencies.

The use (for the 1981 review) of premiums to calculate the town credibility, Z_t, is replaced in this formula by H_t. Like premium, H_t produces larger credibility for towns with more exposures. With H_t, however, the higher the claim frequency, the less credibility is attributed to the actual data. This formulation of H_t assumes that the variability of claim frequency is proportional to claim frequency itself, and that the actual frequency in a town should be given less weight (credibility) as the variability of that claim frequency increases. This approach parallels the overall interpretation of the credibility constant, which is that less weight should be given to a body of data that exhibits instability. The specific methodology used to estimate σ^2 and τ^2 also was changed from the MARB review for 1981, based on Siczewicz [27]. That new methodology has been retained in subsequent reviews and is described below in connection with the MARB proposal for 1986 territories (see Appendix A).[17]

In the review for 1982, the formula frequency index in each town (for each coverage) was, as in the 1981 proposal, calculated as the credibility-weighted average of the actual claim frequency index and the model claim frequency index. In the next step, after this calculation, the effects of the class mix in each town were removed from the town's formula frequency index, since class effects are captured by classifications and classification relativities. The procedure for removing classification effects has been retained in subsequent reviews and is detailed in the Appendix A description of the latest methodology. A final formula claim frequency for each town and coverage is estimated by applying the town claim frequency index to the statewide claim frequency for the coverage.

[17] This credibility methodology also has been adapted for use in calculating Massachusetts private passenger class-territory rate relativities (MARB [15]).

The treatment of claim severity in the review for 1982 paralleled the implicit treatment in the previous year's review: for each coverage the statewide average claim severity, X_c, was assigned to each town, recognizing no variations in claim severity. This statewide average severity, applied to the town formula claim frequency, produced, for each coverage, a formula "pure premium" by town.

Finally, a one-dimensional index that combined all coverages was calculated for each town as

$$\frac{\sum_c E_{c,t} \times \text{Formula Pure Premium}_{c,t}}{\sum_c E_{c,t} \times \text{Statewide Pure Premium}},$$

where $E_{c,t}$ is the insured exposures for coverage c, town t.

This index calculates an all-coverages formula pure premium for the town and compares it to the statewide pure premium that would be observed if the town's coverage purchase patterns were observed statewide. The intent is to ascertain the extent to which a town is above or below average for the coverages purchased in that town.

An alternative formulation using actual statewide exposures in the denominator was rejected, since this alternative formulation would improperly differentiate between two towns identical in all respects except the extent to which physical damage coverages are purchased. Viewed another way, the residents of the town in which physical damage coverages are purchased heavily pay for those coverages directly and should not also pay indirectly by being placed in a higher rating territory.

The MARB, AG, and SRB joined in recommending this final index as the basis for establishing 1982 automobile rating territories (MARB [12]; AG [10]; SRB [21]), and the Commissioner adopted that recommendation (Massachusetts Division of Insurance [18]). With the exception of the treatment of claim severity, which has been refined in the subsequent two reviews, this methodology developed for 1982 has been retained in subsequent reviews and thus is set forth in greater detail in the Appendix A description of the most recent methodology.

The AG differed from the other parties in the method of using the final index to group towns. The AG proposed a clustering algorithm that would have placed *two* constraints on the towns in a territory: (1) the towns should have similar final index values, and (2) all the towns in a territory must be contiguous (AG [10]).

The addition of the contiguity constraint reflected, and imposed, the expectation that two adjacent towns would tend to have similar expected losses. This constraint was also intended to address concerns expressed by members of the driving public that sharp rate differentials between neighboring areas were unfair.

The resulting territories, while comprised of chains of contiguous towns, did not resemble tight clusters, as might have been hoped. In addition, the addition of the contiguity constraint cost a significant loss of homogeneity in the expected losses of towns in each territory. Various technical problems, beyond the scope of this paper, were also identified with the cluster algorithm.

Thus, the SRB and MARB recommended the continued use of town groupings based solely on similarity of town index values, and the Commissioner followed this recommendation. As in the prior revision, the reassignments of individual towns were constrained to avoid any unacceptably large rate changes from 1981 to 1982. The combination of the later data and the methodological changes resulted in territory reassignments for more than 250 towns.

Review for 1984 Territories

During the discussions that led to the joint recommendations for 1982, the parties agreed that a biennial review of territories would be appropriate. The agreement to follow a biennial schedule was based on several considerations:

(1) The claims experience of towns, relative to the statewide average, changes significantly over time. For example, one analysis performed by the MARB indicated that two years of later data (with no methodological changes) would produce indications that over 160 towns should be assigned to new territories, including 35 towns whose territory assignments should change by more than one territory. Thus, delaying a review beyond two years would allow miscategorization of many towns, and might necessitate unacceptably large rate effects when a territory revision did occur.

(2) A two-year interval provides adequate time for the parties to consider methodological improvements.

(3) Because of statistical coding procedures used in Massachusetts, insurance companies can accommodate territory realignments fairly easily, so that biennial revisions are not burdensome.

(4) Annual repetition of the entire territory review and decision process was viewed as impractical.

In accordance with the agreed biennial review schedule, representatives of the MARB, SRB, and AG met during 1983 to consider a possible revision of the territory definitions for 1984. Again the goals of the group were to review the methodologies previously used; to consider alterations and refinements to those methodologies; to review the data that had emerged since the prior review; and, to present to the Commissioner recommendations that had some common bases, even though it was not expected that complete unanimity would be achieved.

As in the previous review, the territory realignment process divided naturally into two major components: the determination of an index for ranking the towns, and the grouping of the ranked towns into territories. The work of the group led to a refinement in the index calculation and to complete unanimity as to the best index and rankings that could be devised for the 1984 review. The process of using the resulting index to group the towns into territories remained an area of some disagreement among the parties.

The index procedure agreed to by the parties, which was documented in the MARB filing [13], recommended only one methodological change to the approach used for the 1982 revision. Specifically, the treatment of severity was modified by assigning to each town the average claim severity for the town's *county*[18], rather than the *statewide* average claim severity. This refinement reflected the clear regional differences in average claim severity, but did so without introducing the instability observed in town claim severities. At that time, the parties had not been able to develop a credibility or modeling procedure that was satisfactory for incorporating by-town claim severities.

This methodological change had a significant impact on the final town index values, because the county average claim severities differ significantly from the statewide average claim costs, as shown in Exhibit 3. This exhibit, which displays the ratios of county average claim costs to statewide average claim costs, reveals for each coverage differences of at least 20% between the county with the lowest average claim severity and the county with the highest average claim severity.

[18] Or county group: in some cases small counties were combined.

EXHIBIT 3

ILLUSTRATIVE CLAIM COST VARIATIONS AMONG COUNTIES

	County Claim Cost Indices (1979–81)				
County Group	BI	PIP	PDL	Collision	Comprehensive
Barnstable, Dukes, Nantucket	.9983	1.0332	.9917	1.0761	.8250
Berkshire	1.0389	.9611	.9257	1.0608	.5895
Bristol	.8713	.9221	.9147	.9305	.9764
Essex	.9693	.9776	.9945	.9903	1.0315
Franklin, Hampshire	.9963	.9599	.9084	1.0672	.6030
Hampden	.9453	.9330	.9270	.9148	.8109
Middlesex	1.0070	1.0069	1.0261	.9902	1.0304
Norfolk	1.0628	1.0043	1.0468	1.0403	1.0309
Plymouth	1.0293	1.0329	1.0437	1.0978	1.0288
Suffolk	1.0718	1.1734	1.0983	.9480	1.3419
Worcester	1.0271	.9840	.9745	1.0355	.7701

Note: Indices calculated as follows:

A. For each year and each coverage, divide each county group average claim cost by statewide average claim cost.

B. For each county group and coverage, calculate an exposure weighted average of the resulting 1979, 1980 and 1981 indices.

The Commissioner (Division of Insurance [19]) adopted the parties' joint recommendations as to the calculation of the final town index values, and, as in the prior revision, selected town groupings based solely on similarity of town index values. This approach created sixteen non-Boston territories; the ten Boston territories were retained, as recommended by the SRB and MARB.[19]

Review for 1986 Territories

In preparing its recommendations for 1986 territories, the MARB retained the 1984 treatment of claim frequency, but again reviewed the handling of claim severity, in addition to incorporating updated data in the analysis.[20] This analysis introduced a newly-developed credibility procedure for claim severity which allowed, for the first time, the utilization of claim severity information by town. Of course, these data still were viewed as being less-than-fully credible, so that complementary data sources were also employed. The selected sources were the countywide[21] average claim severity and the statewide average claim severity.[22] Within each coverage, the claim severity relativity for a town was determined as a credibility-weighted average of the indications from the three sources. The credibility parameters were determined by a two layer hierarchical empirical Bayesian method, described more fully in Appendix A. The empirical Bayesian method compares the variation in relative severity within a town across years; the variation in relative severity across towns within a county; and, the variation in relative severity across counties within the state.

In this approach, the estimated severity for a town is the combination of the severity for the town, the severity for the county that contains the town, and the overall statewide severity. The town's own severity is used to the extent it is credible, with the complement of credibility being given to the estimated severity for the county. In turn, the estimated severity for the county is the credibility-weighted mean of the county to the extent it is credible, with the complement of credibility being given to the credibility-weighted severity overall.

---

[19] The AG recommended combining the ten subdivisions of Boston into three territories.

[20] These recommendations were developed by the MARB [14]. Recommendations of the AG and SRB were prepared and submitted separately.

[21] Actually, county groups in some cases. This component, taken alone, is equivalent to the standalone severity treatment used in the revision for 1984.

[22] This component, taken alone, is equivalent to the claim severity treatment used in the revision for 1982.

The introduction of this new procedure makes very little difference, of course, for small towns whose data is given little credibility and which therefore are assigned approximately the county average claim severity, as they were in the review for 1984. Similarly, the new procedure makes very little difference for a town with claim severities close to the county average. For larger towns that have average claim severities differing significantly from their county taken as a whole, the partial recognition of the town data can make a significant difference. In a few cases the credibility-weighted town severity is as much as 7% different from the county severity. Exhibit 4 illustrates the change in final town index values (for a selection of towns) due to this methodological change.

The other details of the MARB's methodology for 1986 are substantially the same as in the methodology used in the review for 1984. The entire procedure proposed by the MARB for calculating the town index values for use in establishing 1986 automobile rating territories is detailed in Appendix A.

The final town index values produced by the methodology are displayed in Exhibit 5 for a sample of towns. In this exhibit the towns are displayed in rank order, according to the final town index values, ranging from Buckland with a final index of .5034 (expected losses per car are about half the statewide average), to Chelsea with a final index of 1.9318 (expected losses per car are nearly twice the statewide average). The ten subdivisions of Boston are shown at the end of the exhibit and have final index values ranging from 1.2311 to 2.7791. These index values were used by the MARB in proposing 1986 territory definitions. As in prior years, the MARB recommended grouping towns having similar index values.

The AG's recommendations concurred with the MARB's new index calculations. The SRB did not offer a single specific index methodology, but rather expressed a concern that the revision of territories was occurring too frequently and that too many towns were being reassigned in each revision. The Commissioner's Decision (Massachusetts Division of Insurance [20]) employed the MARB index but, mirroring the SRB's concerns, imposed tight constraints on allowed reassignments of towns.

EXHIBIT 4

SAMPLE COMPARISON OF TOWN INDEX VALUES PRODUCED BY TWO METHODS (1980–1983 DATA)

Town Name	1984 Method	1986 Method	Difference
Hampden	.7659	.7708	.0049
Holland	.6475	.6567	.0092
Montgomery	.6424	.6503	.0079
Tolland	.6488	.6509	.0021
Wales	.8009	.8072	.0063
Amherst	.7211	.7298	.0087
Easthampton	.7546	.7325	−.0222
Northampton	.7512	.7299	−.0213
South Hadley	.7535	.7366	−.0169
Ware	.7604	.7694	.0090
Belchertown	.7820	.7879	.0060
Hadley	.6149	.6024	−.0125
Hatfield	.6583	.6548	−.0035
Huntington	.7498	.7923	.0425
Williamsburg	.7065	.7091	.0026
Chesterfield	.7251	.7339	.0087
Cummington	.6561	.6618	.0057
Goshen	.6323	.6565	.0243
Granby	.7600	.7763	.0163
Middlefield	.5793	.5786	−.0006
Pelham	.6884	.6680	−.0204
Plainfield	.6693	.6646	−.0047
Southampton	.6912	.7071	.0158
Westhampton	.7513	.7548	.0035
Worthington	.6231	.6292	.0061
Cambridge	1.3202	1.3130	−.0073
Lowell	1.1582	1.1488	−.0094
Everett	1.3963	1.4757	.0793
Malden	1.2804	1.3440	.0636
Medford	1.2249	1.2576	.0327
Newton	.9684	.9076	−.0609
Somerville	1.5165	1.5588	.0423
Waltham	1.0395	1.0412	.0017
Watertown	1.0894	1.0819	−.0075
Arlington	.9562	.9330	−.0233
Belmont	.9184	.8772	−.0412
Chelmsford	.7610	.7438	−.0171
Concord	.7150	.6968	−.0183
Dracut	.9456	1.0045	.0588
Framingham	.9564	.9359	−.0204
Hudson	.8916	.8935	.0019
Lexington	.7993	.7612	−.0381
Marlborough	.9501	.9526	.0024
Melrose	1.0178	1.0369	.0191
Maynard	.7719	.7571	−.0147

EXHIBIT 5
SHEET 1

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(For Selected Towns)

Town Name	Index						PDL	Observed Pure Premium (latest three years)	
	BI	PIP	PDL	Coll.	Comp.	Combined	Exposures	Liability	Package
TERRITORY 1									
Buckland	0.5218	0.7366	0.4851	0.5470	0.3501	0.5034	952	$40.66	$109.89
Middlefield	0.6031	0.8149	0.5193	0.6633	0.4216	0.5786	212	32.76	100.02
Hadley	0.6028	0.7590	0.6461	0.6664	0.3850	0.6024	3,035	68.80	156.29
Worthington	0.4848	0.7371	0.5603	0.8248	0.5624	0.6292	622	45.35	164.78
Montgomery	0.6284	0.8956	0.5669	0.8384	0.4171	0.6503	374	61.44	213.45
Tolland	0.6041	0.8572	0.6676	0.7491	0.4391	0.6509	166	65.13	251.95
Hatfield	0.7726	0.7326	0.6637	0.7104	0.4177	0.6548	2,023	77.94	165.05
Goshen	0.5767	0.7435	0.6027	0.8800	0.4543	0.6565	404	29.52	199.34
Holland	0.5397	0.8712	0.6477	0.8328	0.4253	0.6567	914	69.30	190.59
Cummington	0.5592	0.7008	0.6663	0.8621	0.4390	0.6618	440	74.37	174.67
Plainfield	0.6870	0.7631	0.5609	0.8870	0.4990	0.6646	294	87.60	196.53
TOTAL (48 towns)						0.6235	81,045	$63.90	$161.92
TERRITORY 2									
Colrain	0.6147	0.7210	0.6560	0.8020	0.5095	0.6651	924	$81.64	$192.84
Pelham	0.7262	0.6530	0.6822	0.7369	0.4839	0.6680	659	75.40	163.79
Concord	0.6206	0.6773	0.7900	0.8217	0.4302	0.6968	9,961	73.91	198.98
Westminster	0.8263	0.7813	0.6824	0.8120	0.4102	0.7041	3,378	79.75	186.29
TOTAL (33 towns)						0.6900	94,683	$77.00	$188.88
TERRITORY 3									
Rockport	0.5922	0.7192	0.8276	0.7717	0.5292	0.7068	3,920	$69.03	$187.59
Southampton	0.8541	0.9494	0.6603	0.7553	0.4671	0.7071	2,527	69.55	179.61
Williamsburg	0.7943	0.7382	0.7546	0.7765	0.4272	0.7091	1,493	83.44	187.00
Amherst	0.6887	0.6539	0.7270	0.8517	0.6172	0.7298	10,022	78.67	204.44
Northampton	0.7930	0.9033	0.8373	0.7261	0.4467	0.7299	13,633	85.53	180.99
Easthampton	0.8141	0.8874	0.8341	0.7618	0.3950	0.7325	8,556	87.51	184.02
Chesterfield	0.7669	0.9490	0.5845	0.8806	0.6588	0.7339	555	75.75	217.71
South Hadley	0.8665	0.9271	0.8201	0.7660	0.4004	0.7366	8,218	85.80	183.24
Chelmsford	0.7609	0.6756	0.8297	0.7881	0.5589	0.7438	19,038	86.94	218.69
Brimfield	0.9462	0.9789	0.7243	0.7987	0.3833	0.7467	1,336	96.42	195.70
TOTAL (43 towns)						0.7313	192,994	$81.48	$200.11
TERRITORY 4									
Dalton	0.8673	0.7354	0.8972	0.7410	0.4462	0.7518	3,427	$86.45	$185.69
Westhampton	0.7902	0.9104	0.7261	0.9326	0.4372	0.7548	678	113.21	249.89
Maynard	0.7847	0.7828	0.8802	0.7766	0.5040	0.7571	5,420	89.74	203.76
Lexington	0.7151	0.6436	0.8919	0.8103	0.5826	0.7612	18,274	88.87	222.39
Ware	0.9327	0.9597	0.7733	0.8745	0.3557	0.7694	4,746	85.02	194.39
Hampden	0.8769	1.0627	0.7932	0.8385	0.4141	0.7708	2,969	100.69	219.45
Granby	0.8929	1.0815	0.7670	0.8488	0.4533	0.7763	3,104	93.16	205.79
Belchertown	0.9374	1.0049	0.7505	0.8813	0.4590	0.7879	4,680	95.83	212.63
TOTAL (52 towns)						0.7688	244,421	$88.71	$219.64

EXHIBIT 5
SHEET 2

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(FOR SELECTED TOWNS)

	Index							Observed Pure Premium (Latest Three Years)	
Town Name	BI	PIP	PDL	Coll.	Comp.	Combined	PDL Exposures	Liability	Package
TERRITORY 5									
Freetown	0.7255	1.0134	0.7048	0.8121	0.8952	0.7923	4,429	$77.97	$214.40
Huntington	0.8181	1.0210	0.7228	0.9715	0.5001	0.7923	1,099	97.82	259.53
Wales	0.9791	1.1784	0.7330	0.8517	0.4766	0.8072	642	103.84	229.27
Cheshire	1.1968	0.8297	0.8353	0.7898	0.5492	0.8393	1,858	114.34	230.16
TOTAL (58 towns)						0.8133	377,184	$93.36	$225.84
TERRITORY 6									
Pittsfield	0.9237	1.0037	1.0512	0.7716	0.4955	0.8405	23,215	$103.34	$213.53
Belmont	0.8108	0.7705	0.9853	0.9192	0.7551	0.8772	14,011	94.29	242.68
Groveland	1.0068	0.9376	0.9187	0.9202	0.6729	0.8897	2,997	121.45	275.98
TOTAL (29 towns)						0.8623	246,318	$99.29	$237.42
TERRITORY 7									
Lynnfield	0.7559	0.7815	0.8793	0.9499	0.9536	0.8903	7,022	$89.51	$278.18
Hudson	1.0119	0.9559	0.9603	0.9667	0.5376	0.8935	9,347	113.63	262.89
Newton	0.8075	0.7563	1.0137	0.9686	0.7953	0.9076	44,546	97.82	265.62
Arlington	0.8520	0.8296	1.0094	0.9641	0.8836	0.9330	25,050	97.88	257.57
Framingham	0.8947	0.9556	1.0863	0.9804	0.6821	0.9359	35,207	110.40	270.40
Taunton	0.9870	1.1305	0.8896	0.8941	0.9970	0.9433	21,884	108.03	254.07
TOTAL (29 towns)						0.9155	316,422	$103.49	$259.77
TERRITORY 8									
Norwood	1.0023	0.8238	0.9998	0.9462	0.8767	0.9473	16,092	$107.88	$267.20
Marlborough	1.0850	1.0861	1.0098	0.9880	0.6390	0.9526	17,002	117.22	271.45
Wilmington	1.0063	0.9694	1.0451	0.9645	0.9674	0.9938	10,232	124.82	299.34
Tewksbury	1.0542	1.0173	1.0043	1.0518	0.8375	0.9973	13,205	118.43	301.59
TOTAL (11 towns)						0.9701	133,231	$110.64	$280.22
TERRITORY 9									
Marshfield	0.8233	1.0737	0.9463	1.0817	1.0827	1.0006	11,537	$104.63	$298.99
Dracut	1.0308	1.2132	1.0090	0.9948	0.9184	1.0045	12,547	128.22	295.68
Melrose	1.0408	0.9126	1.0519	0.9753	1.1626	1.0369	15,377	110.18	290.56
Waltham	1.0843	1.0354	1.1057	1.0299	0.9376	1.0412	27,703	118.94	288.15
Holyoke	1.3525	1.1436	1.1678	0.9313	0.8089	1.0597	17,069	132.77	261.19
TOTAL (21 towns)						1.0345	351,465	$116.81	$287.84
TERRITORY 10									
Haverhill	1.2649	1.1987	1.0834	0.9720	0.9866	1.0679	21,905	$132.68	$294.11
Watertown	1.0245	0.8988	1.1275	1.1080	1.0918	1.0819	17,042	111.42	293.96
Worcester	1.2873	1.0872	1.2669	1.0435	0.8536	1.1113	63,452	134.67	292.29
TOTAL (6 towns)						1.0914	149,630	$126.23	$297.59

EXHIBIT 5
SHEET 3

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(For Selected Towns)

	Index						PDL	Observed Pure Premium (Latest Three Years)	
Town Name	BI	PIP	PDL	Coll.	Comp.	Combined	Exposures	Liability	Package
TERRITORY 11									
Holbrook	1.1936	1.0925	1.1007	1.1166	1.1444	1.1276	6,136	$130.64	$325.24
Lowell	1.2058	1.2939	1.1749	1.1143	1.0629	1.1488	37,723	138.40	324.54
Quincy	1.1494	1.0244	1.1279	1.1782	1.4034	1.1907	39,832	121.99	330.82
TOTAL (7 towns)						1.1604	138,765	$124.57	$330.05
TERRITORY 12									
Springfield	1.6355	1.6528	1.2426	1.0985	0.9862	1.2463	62,300	$153.62	$309.41
Medford	1.0835	1.0678	1.1427	1.1778	1.7861	1.2576	26,638	125.71	365.73
Brockton	1.3352	1.2859	1.2262	1.2281	1.4207	1.2832	41,920	143.62	360.58
TOTAL (3 towns)						1.2604	130,857	$144.74	$337.27
TERRITORY 13									
Cambridge	1.1122	1.1053	1.1675	1.3506	1.7395	1.3130	30,202	$124.92	$362.73
Malden	1.2725	1.2228	1.2689	1.2197	1.7768	1.3440	23,400	140.38	378.93
Lynn	1.2561	1.1420	1.3693	1.1942	1.8444	1.3690	33,217	144.02	369.84
TOTAL (4 towns)						1.3415	91,504	$136.84	$370.15
TERRITORY 14									
Lawrence	1.4980	1.6148	1.3072	1.1900	1.9389	1.4421	23,935	$162.76	$394.42
Everett	1.3023	1.2718	1.3755	1.2769	2.2025	1.4757	15,385	147.74	424.12
TOTAL (3 towns)						1.4549	48,036	$154.29	$407.97
TERRITORY 15									
Somerville	1.3551	1.2398	1.3910	1.4373	2.3460	1.5588	26,169	$151.50	$436.19
TOTAL (1 town)						1.5588	26,169	$151.50	$436.19
TERRITORY 16									
Revere	1.5148	1.4559	1.4578	1.5443	3.1677	1.7956	17,396	$162.89	$547.65
Chelsea	1.6758	1.6544	1.5957	1.6970	3.3012	1.9318	7,307	183.39	573.08
TOTAL (2 towns)						1.8359	24,704	$168.95	$555.17
TERRITORY 17—West Roxbury (Boston)									
	1.2349	1.0351	1.1328	1.1949	1.4951	1.2311	12,867	$118.80	$335.09
TERRITORY 18—Roslindale (Boston)									
	1.4514	1.2973	1.4038	1.5675	2.4886	1.6536	10,769	$150.58	$457.66
TERRITORY 19—Jamaica Plain (Boston)									
	1.5896	1.6845	1.4404	1.7914	3.1083	1.8918	10,400	$155.81	$512.46
TERRITORY 20—Hyde Park (Boston)									
	1.3877	1.6163	1.3734	1.5910	2.4247	1.6534	11,481	$151.63	$463.63
TERRITORY 21—Dorchester (Boston)									
	2.1882	2.3175	1.6118	2.3901	4.2874	2.4664	33,479	$213.34	$727.76
TERRITORY 22—Roxbury (Boston)									
	2.4487	2.8154	1.8756	2.6857	4.8291	2.7791	6,538	$245.62	$818.45
TERRITORY 23—Boston Central (Boston)									
	1.6152	1.6337	1.4941	1.9746	3.6202	2.0513	16,904	$162.83	$613.69

EXHIBIT 5
SHEET 4

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(FOR SELECTED TOWNS)

Town Name	Index						PDL	Observed Pure Premium (Latest Three Years)	
	BI	PIP	PDL	Coll.	Comp.	Combined	Exposures	Liability	Package
TERRITORY 24—Brighton (Boston)	1.2748	1.2898	1.3588	1.5606	2.0844	1.5361	15,873	$137.77	$422.43
TERRITORY 25—South Boston (Boston)	1.6416	1.6759	1.4753	1.9179	3.2148	1.9269	6,491	$168.47	$572.51
TERRITORY 26—East Boston (Boston)	1.7229	1.6995	1.8241	1.9951	3.9162	2.2038	10,780	$197.54	$669.83
TOTALS ALL TERRITORIES						1.0000	2,783,010	$109.48	$279.68

Perspective

The continuing evolution since 1976 of the improved methods described above, which are used for calculating a one-dimensional town index that reflects for each town the relative expected insured losses per car, has contributed to a trend towards satisfying the criteria set forth by the SRB in 1976 (and described in the introduction to this paper). Specifically, the criteria of (statistical) equity and reliability depend directly on the quality of the estimation of expected losses; the compatibility criterion has been satisfied by the decision to maintain a single set of territories; and the stability criterion has been addressed by scheduling regular territory reviews to minimize the number of dramatic territory changes, and by imposing constraints on any large changes that are indicated by the data. The criteria of homogeneity and discrimination depend on the accuracy of the estimation of expected losses, which serves as the basis for making territory assignments, but also depends on the selection of a grouping process, given the final town index. The next section discusses the grouping process.

4. GROUPING TOWNS INTO TERRITORIES

The presentation of territory recommendations in Massachusetts in the last nine years generally has involved two principal steps: first, developing a one-dimensional index that quantifies the relative claims experience in each town; and second, using the one-dimensional index to group towns into territories. The preceding section focused on the first step, from the use of composite physical damage claim frequencies in 1977 to the use for 1986 of a synthetic pure premium index computed from Bayesian estimates of town claim frequencies and claim severities by coverage.

This section discusses the methodology used to group towns into territories, given the one-dimensional final town index. Although various techniques have been discussed and proposed, the Commissioner has used basically the same approach in each of the territory revisions since 1977.

Principal Considerations

The principal considerations that have governed the proposals for groupings of towns into territories are:

(1) The homogeneity of competing territory configurations.

(2) The possible reintroduction of proximity constraints.

(3) The handling of the ten subdivisions of Boston.

(4) The magnitude of rate differentials between territories.

(5) The magnitude of individual town rate changes that would result from proposed realignments of territories.

(6) The number of territories.

(7) The size (number of exposures) of each territory.

The first of these considerations, homogeneity, has been defined in practice to refer to the extent to which individual town claims experience differs from the average claims experience for all towns in a territory. Several quantitative measures have been developed, as discussed below, to compare the overall homogeneity of competing territory configurations.

The second consideration, reintroduction of geographical constraints, has been suggested for several reasons, including improved public understanding of territories and social equity advantages of increasing the probability that apparently similar towns in the same area of the state would be placed in the same territory.

Suggested alternatives to the current ten independent territories for the ten subdivisions of Boston have involved combining some of the sections of Boston with one another and/or with non-Boston territories. Doing so would increase the exposure base used for pricing and resulting territories, would provide a degree of cross-subsidization between the combined Boston subsections, and would degrade the homogeneity of the territory configuration.

The fourth and fifth considerations, the magnitude of rate differentials between territories and the magnitude of individual town rate changes from year to year, have principally acted as constraints on otherwise-indicated territory changes.[23] In the grouping procedures actually adopted, these considerations generally have been incorporated by partially tempering the reassignments of a few towns for which the analysis indicated substantial changes, although a more restrictive constraint was employed by the Commissioner for 1986. These considerations have also contributed to the rejection of some proposals to reintroduce geographical constraints, since: (a) some of the geographic proposals could not be introduced without causing unacceptably large rate changes for certain towns (Massachusetts Division of Insurance [19]), and (b) with the large rate differentials between territories that would be implied by restrictions on the number of territories available to towns in a particular geographical area, small changes in data or methodology could cause a large rate change for a town. These implications of the geographic proposals follow from the fact that each geographic region of the state contains towns from a wide range of current territories.

The sixth consideration, the number of territories, is largely a practical one. Approximately two dozen territories have been viewed as enough to maintain a reasonable degree of homogeneity without the system becoming administratively cumbersome.

The final consideration, the number of exposures in each territory, has two aspects: each territory should provide a sufficient data base for the ratemaking process, and no territory should be dramatically larger than the remaining territories (since homogeneity might suffer). These factors have guided the development of proposals for grouping towns into territories.

[23] The rate changes discussed here affect individual towns only. All territory proposals are implemented so as to have no overall rate level effects.

Selected Grouping Methodology

The town grouping methodologies used in the 1977, 1982, 1984, and 1986 revisions all are generally similar (but with details differing). In essence, the towns are ranked in accordance with their final town index values, and index value breakpoints are selected. A territory is then defined as including all towns having index values between two consecutive breakpoints.

For the most part, the town index values form a continuum, with few obvious breakpoints, so that the breakpoints generally have been selected by a numerical algorithm. In the MARB proposal for 1986, for example, breakpoints initially were selected at an index value of unity (which is the statewide average index value) and at each integer power of 1.06; all towns with an index value below .665 (1.06^{-7}) are placed in Territory 1, and all towns with an index value above 1.504 (1.06^{7}) are combined in a single territory.[24]

The selection of the 1.06 factor was based on: (a) the number of territories it produced, (b) the sizes of the resulting territories, and (c) the homogeneity of the resulting territories (see below). Judgment, however, is superimposed on the territories at the high end of the index value range, where natural breakpoints are evident. Further, a judgment was made to continue the ten independent Boston territories.

Finally, a capping algorithm is applied to determine the rate impact on each town of the territory realignment. In the 1982 and 1984 revisions, any town seen as being subjected to an unacceptably large rate increase due to the realignment is reassigned to a territory closer (in territory number) to its current territory placement.

In the 1986 revision, the Commissioner imposed additional constraints: any town proposed by the MARB to move one territory up or down was not moved at all, while any town proposed by the MARB to move more than one territory up or down was constrained to move one territory (in the direction indicated). With these additional constraints, only 22 towns changed territories, and thus the 1986 territories are nearly identical to the 1984–85 territories.

Homogeneity Measures

Appendix B details several quantitative measures that have been designed to compare the relative homogeneity of alternative Massachusetts automobile

[24] The algorithm used for 1982 and 1984 was similar. The revision for 1977 used a more complex algorithm to select breakpoints.

territory configurations. Each of the measures captures a slightly different dimension of homogeneity or heterogeneity, and no attempt has been made to calibrate the measures so that one measure can be compared to another; nor is there an absolute scale against which a territory configuration can be judged "homogeneous" or "not homogeneous." Rather, the appropriate comparison is among the results of a single homogeneity measure applied to various territory configurations. The territory configuration with a homogeneity value closer to zero is considered relatively more homogeneous by the standards of a particular measure.

These homogeneity measures have been used in three aspects of the territory review process in Massachusetts. First, they have been used to determine whether existing territory configurations are showing a significant[25] deterioration in homogeneity as they become outdated. Second, the measures have been used to compare different methods of constructing the final town index, to see which produces more homogeneous territories. Finally, and most obviously, the homogeneity measures have been used to evaluate alternative proposals for selecting territory groupings, given a set of final town index values. Exhibit 6 illustrates these uses of the homogeneity measures.

Outdated Territories

Exhibit 6 displays homogeneity measures for the 1982–83 territories, the 1984–85 territories, and the territories proposed by MARB for 1986.[26] The results indicate clearly that the 1982–83 and 1984–85 territories are significantly less homogeneous than are the territories proposed for 1986. It is not immediately evident from Exhibit 6, whether this difference is due to shifting claims experience or due to improving methodologies, but the inclusion on Exhibit 6, of the updated calculations based on the 1984 methodology makes it apparent that much of the difference is due to shifting claims experience.

Index Methodology

Exhibit 6 compares the homogeneity of territories produced by the 1984 town index methodology and by the 1986 town index methodology. Each is

[25] "Significant" in this context is a qualitative term, as statistical significance levels have not been determined for the homogeneity measures.

[26] Prior to the additional constraints that the Commissioner imposed on town movements.

EXHIBIT 6
SHEET 1

HOMOGENEITY MEASURES FOR TERRITORY GROUPINGS INDEX BASED ON 1984 INDEX METHOD UPDATED FOR NEW DATA VS. 1986 INDEX METHOD

	Territory Division Index Intervals; no capping								
	5%		5.5%		6%		6.5%		1984–85
Homogeneity Measure*	1984 Method	1986 Method	1984 Method	1986 Method	1984 Method	1986 Method	1984 Method	1986 Method	Territory Grouping
1. P.P. Squared Diff. (Absolute)									
a) Liability	78.18	61.71	75.98	61.06	74.57	60.54	68.22	59.56	88.03
b) Package	370.20	345.74	403.21	301.84	373.66	277.07	320.06	288.38	427.04
2. P.P. Squared Diff. Cred. Weighted (%)									
a) Liability	.008722	.007414	.008370	.006662	.007982	.006506	.007391	.006314	.009365
b) Package	.005353	.004372	.005492	.004013	.005302	.003877	.004954	.003904	.007294
2A. P.P. Squared Diff. (%)									
a) Liability	.009681	.008415	.009276	.007451	.008769	.007234	.008047	.006994	.01035
b) Package	.005652	.004638	.005780	.004228	.005568	.004084	.005193	.004080	.007665
3. Index Squared Diff. (Absolute)	.000768	.001161	.000719	.000852	.000611	.000645	.000328	.000650	.001271
4. P.P. Maximum Diff. (Absolute)									
a) Liability	25.22	26.35	23.30	23.68	23.18	23.35	20.67	22.14	25.59
b) Package	52.41	46.59	52.72	45.48	51.99	45.91	48.79	43.07	51.43
5. P.P. Maximum Diff. Cred. Weighted (%)									
a) Liability	.2283	.2378	.2261	.2088	.2201	.2143	.2167	.2052	.3068
b) Package	.1613	.1493	.1567	.1444	.1533	.1337	.1450	.1417	.2198
5A. P.P. Maximum Diff. (%)									
a) Liability	.3170	.3700	.3028	.2831	.2691	.2761	.2489	.2522	.3859
b) Package	.1938	.1857	.1890	.1735	.1819	.1634	.1700	.1683	.2565
6. Index Maximum Diff. (Absolute)	.03572	.04294	.03416	.03335	.03104	.03114	.02983	.03096	.08221
7. Error Entropy	1.9366	1.9019	1.8101	1.9290	1.7993	1.9718	1.9038	1.9709	2.6141

* Refer to Appendix B for formulas.

EXHIBIT 6
SHEET 2

HOMOGENEITY MEASURES FOR TERRITORY GROUPINGS

Homogeneity Measure***	Using 1986 Index Methodology Territory Division Index Intervals; No Capping				MARB Recommendation		1984–85	1982–83
	5%	5.5%	6%	6.5%	Uncapped Terr.*	Capped Terr.**	Territory Grouping	Territory Grouping
1. P.P. Squared Diff. (Absolute)								
a) Liability	61.71	61.06	60.54	59.56	59.15	60.05	88.03	127.07
b) Package	345.74	301.84	277.07	288.38	224.99	234.38	427.04	718.08
2. P.P. Squared Diff. Cred. Weighted (%)								
a) Liability	.007414	.006662	.006506	.006314	.006456	.006537	.009365	.01276
b) Package	.004372	.004013	.003877	.003904	.003738	.003885	.007294	.01176
2A. P.P. Squared Diff. (%)								
a) Liability	.008415	.007451	.007234	.006994	.007183	.007273	.01035	.01401
b) Package	.004638	.004228	.004084	.004080	.003941	.004101	.007665	.01226
3. Index Squared Diff. (Absolute)	.001161	.000852	.000645	.000650	.000315	.000346	.001271	.005063
4. P.P. Maximum Diff. (Absolute)								
a) Liability	26.35	23.68	23.35	22.14	23.20	24.02	25.59	27.27
b) Package	46.59	45.48	45.91	43.07	44.49	48.29	51.43	62.16
5. P.P. Maximum Diff. Cred. Weighted (%)								
a) Liability	.2378	.2088	.2143	.2052	.2151	.2183	.3068	.3635
b) Package	.1493	.1444	.1337	.1417	.1318	.1343	.2198	.2849
5A. P.P. Maximum Diff. (%)								
a) Liability	.3700	.2831	.2761	.2522	.2769	.2804	.3859	.6260
b) Package	.1857	.1735	.1634	.1683	.1614	.1700	.2565	.3349
6. Index Maximum Diff. (Absolute)	.04294	.03335	.03114	.03096	.02854	.04292	.08221	.1581
7. Error Entropy	1.9019	1.9290	1.9718	1.9709	1.9522	1.9844	2.6141	3.1465

* Reflecting judgmental adjustments to territories at high end of scale.
** Prior to imposition of Commissioner's additional constraints (see text).
*** Refer to Appendix B for formulas.

displayed with various alternatives to the 1.06 index value boundaries actually used for 1986. The results indicate that:

(a) For most of the homogeneity measures based on actual loss pure premiums, the 1986 method of treating claim severity substantially improves the homogeneity of the territories.

(b) For the homogeneity measures based on the constructed index values and for the error entropy measure, the 1986 and 1984 methodologies produce similar homogeneity values. However, since the dispersion of the index values has been increased by the recognition of claim cost variations by town, the index-based homogeneity measures and the error entropy measures probably have little useful value in comparing the homogeneity of the final territories produced by the two methods.

Territory Groupings

Exhibit 6, displays the homogeneity measures produced by territory groupings based on the selected 1.06 breakpoint factor as well as those based on alternative breakpoint factors of 1.05, 1.055, and 1.065. Generally, the homogeneity measures indicate that the breakpoint factors of 1.06 and 1.065 are to be preferred, with the 1.06 factor performing best on the measures that reflect a package of all major insurance coverages. The 1.06 factor actually was selected for the several reasons indicated above. Exhibit 6 also indicates that the MARB's judgmental adjustments to the territory breakpoints and the MARB's application of the "traditional" capping process produce only minor changes in the homogeneity measures.

By all measures, the proposed territories are far more homogeneous than the 1984–85 territories. However, the additional constraints imposed by the Commissioner nearly recreate, in 1986, the 1984–85 territory definitions and thus bear a nontrivial cost in terms of homogeneity.

Perspective

This loss of homogeneity usefully may be viewed as the cost of shifting the regulatory emphasis from the homogeneity criterion towards the stability criterion. This trade-off illustrates two general principles often encountered in classification issues (and other issues): that not all constraints can be satisfied simultaneously; and that the relative emphasis placed on the different constraints ultimately must be resolved by the application of judgment, even if complex methodologies are available to clarify the nature and implications of the necessary choices.

5. SUMMARY

The methodologies described in this paper may be useful specifically to practitioners in the automobile insurance field. In addition, particularly with regard to the empirical Bayesian credibility techniques, the formulas—or the concepts they implement—may be useful in other fields as well.

Two conclusions of the Massachusetts territory analysis are of particular interest in that they suggest a change to the conventional structure of automobile rating territories and a change to the frequency with which territories are reviewed. These two conclusions are:

(1) That claims experience varies significantly from town to town, even among neighboring towns with generally similar characteristics; and

(2) That claims experience of towns shifts materially over time and, therefore, that territory definitions should be reviewed regularly.

While the author expects that Massachusetts methodologies will continue to evolve in the future, the procedures and results of the current Massachusetts state of the art may prove useful elsewhere in the meantime.

REFERENCES

[1] All-Industry Research Advisory Council, *Geographical Differences in Automobile Insurance Costs*, Research Report A82-1, October, 1982.

[2] California Department of Insurance, Rate Regulation Division, *Study of California Driving Performance by Zip Code*, 1978.

[3] William H. DuMouchel, "The 1982 Massachusetts Automobile Insurance Classification Scheme," *The Statistician* 32, 1983.

[4] William H. DuMouchel and Jeffrey E. Harris, "Bayes and Empirical Bayes Methods for Combining Cancer Experiments in Man and Other Species," Massachusetts Institute of Technology Technical Report No. 24, February, 1981.

[5] C. B. Garrison and A. S. Paulson, "An Entropy Measure of the Geographic Concentration of Economic Activity," *Economic Geography*, Vol. 49, 1973, p. 319.

[6] James F. Gill, "An Approximation for the Testing of Private Passenger Liability Territorial Rate Levels Using Statewide Distribution of Classification Data," *PCAS* LI, 1964.

[7] C. C. Hewitt, "Credibility for the Layman," in *Credibility Theory and Applications*, ed. by P. M. Kahn, Academic Press, New York, 1975.

[8] J. C. Hickman, "Introduction and Historical Overview of Credibility," in *Credibility Theory and Applications*, ed. by P. M. Kahn, Academic Press, New York, 1975.

[9] Highway Loss Data Institute, *Relationships Between Losses and Vehicle Density, 1982–84 Models*, Insurance Special Report A-25, October, 1985.

[10] Massachusetts Attorney General, *Preliminary Filing for the 1981 Hearing on Automobile Insurance Territories* (proposed territories to be used in 1982), August, 1981.

[11] Massachusetts Automobile Rating and Accident Prevention Bureau, *Filing for 1981 Massachusetts Automobile Rating Territories*, July, 1980.

[12] Massachusetts Automobile Rating and Accident Prevention Bureau, *Filing for 1982 Massachusetts Automobile Rating Territories*, August, 1981.

[13] Massachusetts Automobile Rating and Accident Prevention Bureau, *Filing for 1984 Private Passenger Territory and Classification Definitions*, July, 1983.

[14] Massachusetts Automobile Rating and Accident Prevention Bureau, *Filing for 1986 Private Passenger Territory and Classification Definitions*, July, 1985.

[15] Massachusetts Automobile Rating and Accident Prevention Bureau, *Recommendations with Respect to Rate Relativities for Private Passenger Vehicles* (for 1985), November, 1984.

[16] Massachusetts Division of Insurance, *Automobile Insurance Risk Classification: Equity and Accuracy*, 1978.

[17] Massachusetts Division of Insurance, *Opinion, Finding and Decision Regarding the Establishing of Territorial Boundaries to be Used in Rate Making for Automobile Insurance* (territories to be used in 1981), August, 1980.

[18] Massachusetts Division of Insurance, *Opinion, Finding and Order Relative to Rating Territories for Automobile Insurance for 1982*, December, 1981.

[19] Massachusetts Division of Insurance, *Opinion, Finding and Order Relative to Rating Territories and Classifications for Automobile Insurance for 1984*, January, 1984.

[20] Massachusetts Division of Insurance, *Opinion, Finding and Order Relative to Rating Territories and Classifications for Automobile Insurance for 1986, Part I*, November, 1985.

[21] Massachusetts State Rating Bureau, *1982 Massachusetts Automobile Rating Territories—Recommendation* (proposed territories to be used in 1982), August, 1981.

[22] Massachusetts State Rating Bureau, *Recommendations With Respect to Rating Territories and Classification for 1984 Automobile Insurance*, July, 1983.

[23] Massachusetts State Rating Bureau, "Territories for Automobile Insurance in Massachusetts" (procedures for determining territories for use in 1977), 1976.

[24] James L. McDonald and John H. Thornton, "A Possible Technique for Automobile Insurance Territorial Classification," American Risk and Insurance Association Annual Meeting, August, 1983.

[25] New Jersey Department of Insurance, *In re: Hearing on Automobile Insurance Classifications and Related Methodologies*, April, 1981.

[26] Rhode Island Ad Hoc Committee on Territorial Rating: *Report of the Actuarial Subcommittee*, December, 1980.

[27] Peter Siczewicz, "A Procedure to Determine Massachusetts Automobile Insurance Territories," Master's Thesis, Massachusetts Institute of Technology, Statistics Center Technical Report No. NSF-30, June, 1981.

[28] SRI International, *The Role of Risk Classifications in Property and Casualty Insurance: A Study of the Risk Assessment Process*, May, 1976.

[29] Philipp K. Stern, "Ratemaking Procedures for Automobile Liability Insurance," *PCAS* LII, 1965.

APPENDIX A
CALCULATION OF THE TOWN INDEX AND TOWN RANKING[27]

A1. SUMMARY

This appendix describes the calculation of the index that is intended to reflect a town's overall loss potential relative to the statewide average loss potential. The calculation methodology described here is that underlying the 1986 Massachusetts automobile rating territories, as described in the body of the paper. Exhibit 7 schematically displays the process of deriving the final town index used to rank the towns.

The starting point for the calculation of the town index is the actual experience (exposures, number of claims, and loss payments) of the vehicles insured in each town. This actual experience may be expressed in terms of claim frequency (average number of claims per insured exposure) and average claim cost (average cost per claim).

The analysis uses the actual claim frequency by coverage of each town, credibility weighted with model claim frequencies by town and coverage; the parameters of the model and the calibration of the credibility functions are based on an analysis of patterns and variations in claim frequency across towns and years. The claim frequency method of analysis is detailed in Section A2.

The analysis also utilizes average claim cost data by town, credibility weighted with average claim cost data by county and statewide. The procedure used to estimate the relative average claim cost by town is detailed in Section A3.

The resulting claim frequency and claim cost indications by town are combined to produce a pure premium index by town and coverage. These pure

[27] This Appendix was excerpted, with editing, from sections of the Massachusetts Automobile Rating and Accident Prevention Bureau's (MARB) *Filing for 1986 Private Passenger Territory and Classification Definitions*, July 1985, which was written by Dr. Richard Derrig, Howard Mahler, and the author of this paper. The Bayesian credibility procedures used in the claim frequency analysis were developed by the MARB and by Peter Siczewicz. The Two Layer Hierarchical Empirical Bayesian Method of analyzing claim severities (see below) was developed and prepared by Howard Mahler for the MARB *Filing for 1986 Private Passenger Territory and Classification Definitions*. Howard Mahler's work on the Two Layer Hierarchical Empirical Bayesian Method was based on P. Heckman, "Credibility and Solvency," *Pricing Property and Casualty Insurance Products*, CAS Discussion Paper Program, May 1980; and G. Venter, "Structured Credibility in Applications—Hierarchical, Multidimensional, and Multivariate Models," 1984 (unpublished).

EXHIBIT 7
SHEET 1

OUTLINE OF TOWN INDEX CALCULATIONS

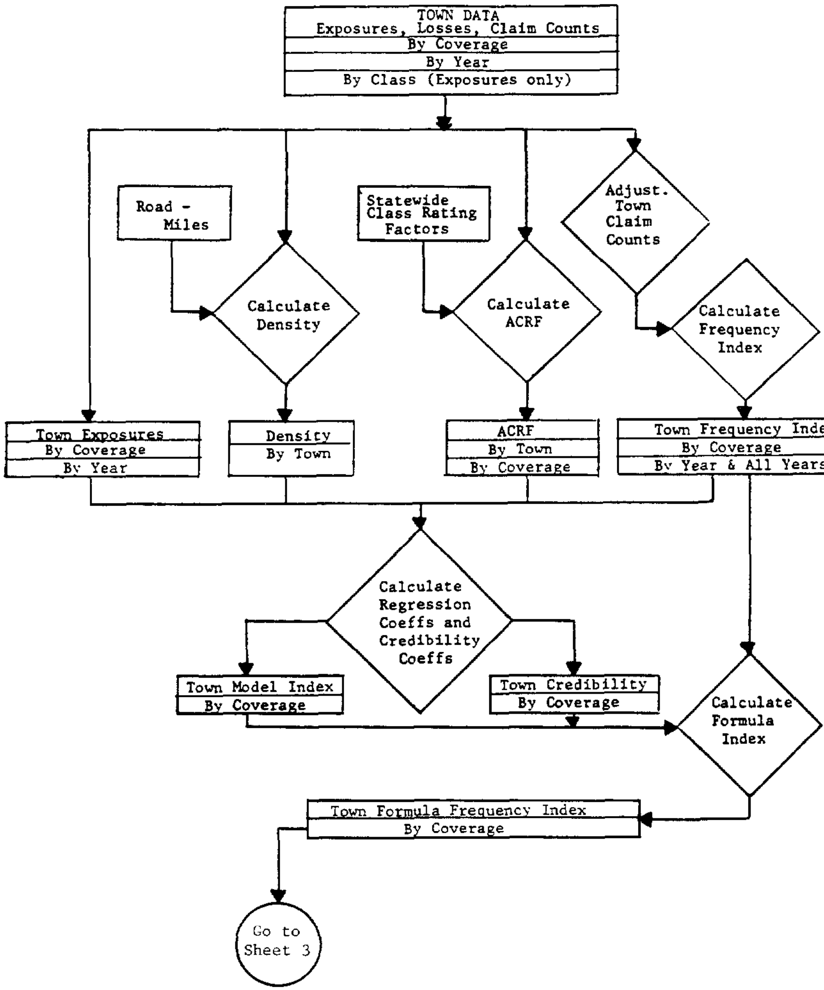


EXHIBIT 7
SHEET 2

OUTLINE OF TOWN INDEX CALCULATIONS

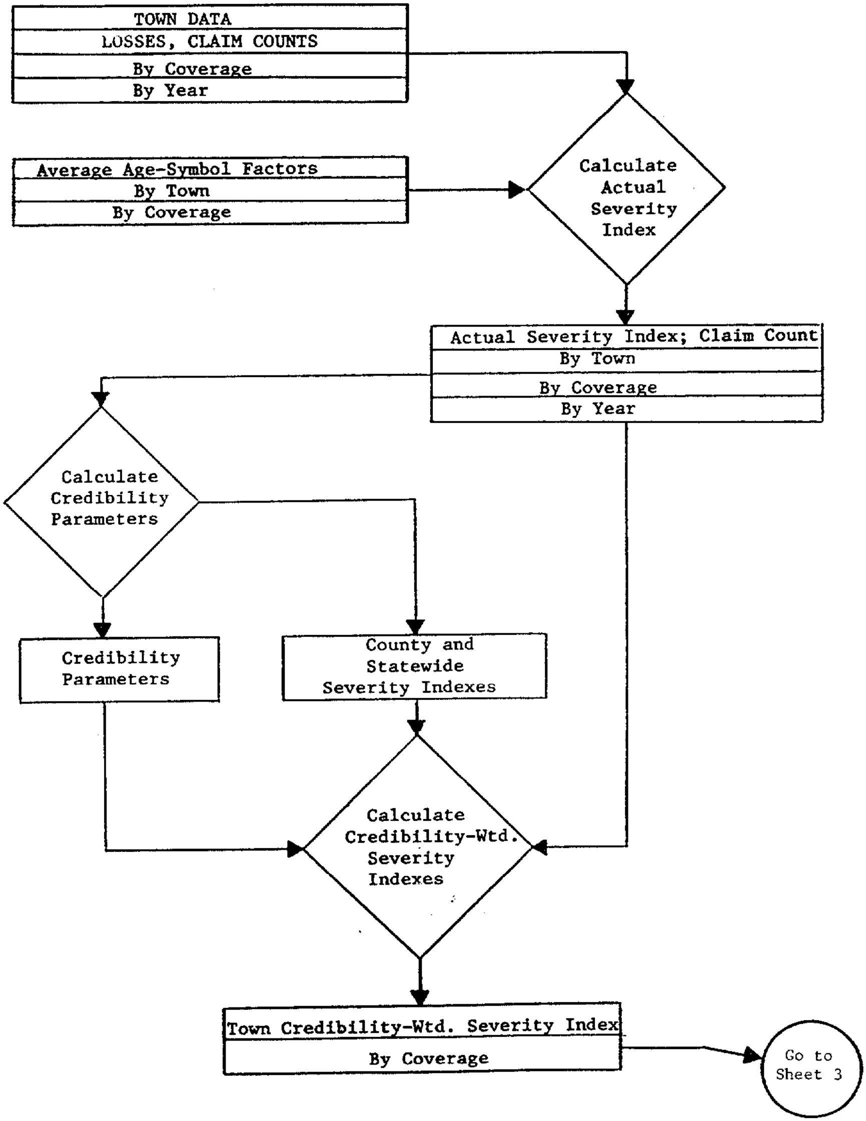
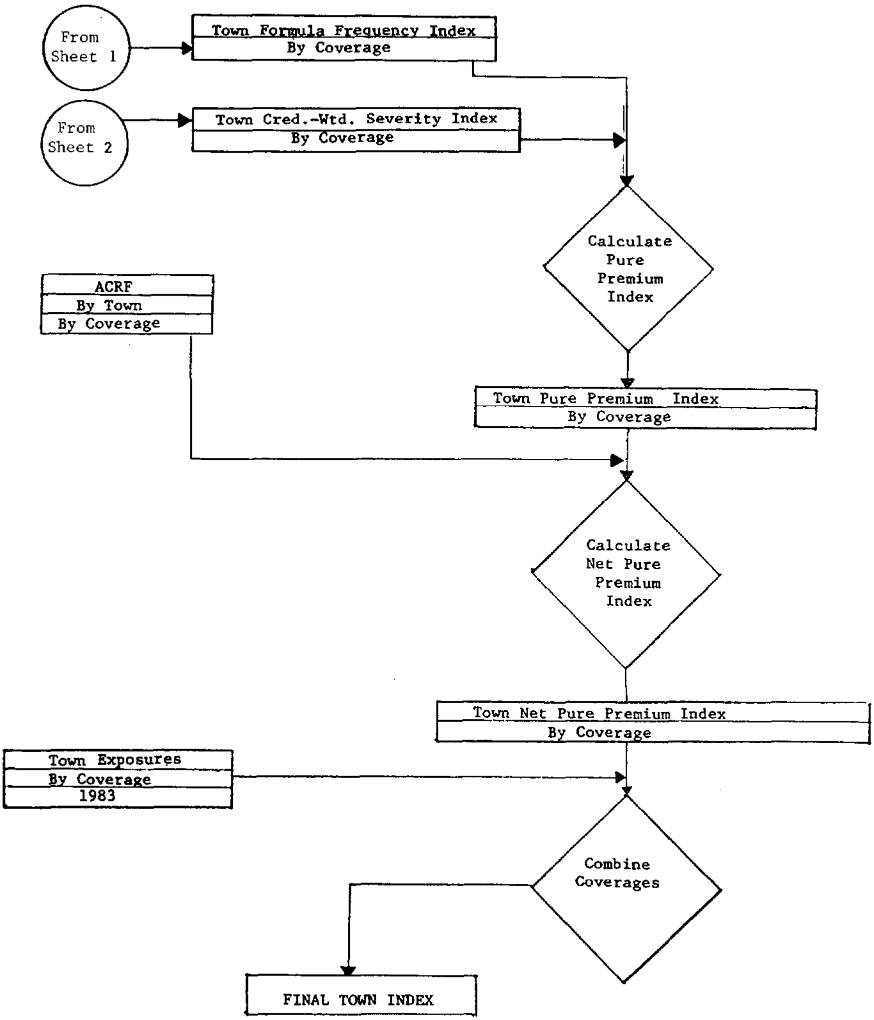


EXHIBIT 7
SHEET 3

OUTLINE OF TOWN INDEX CALCULATIONS



premium indices are then modified to the extent they reflect components of the town's driver classification mix already captured by other elements of the rating system.

As described in Section A4, the final town index is a weighted average of the pure premium indices for the five major coverages for which rates vary by territory.

A2. BUILDING THE CLAIM FREQUENCY INDEX

The details of the methodology used to determine the claim frequency index are described and illustrated in this section. Exhibit 8 details the formulas used.

a. Data

Exposures and claim counts by town and year (latest four years) for each of the coverages A-1, A-2, PDL, Collision, and Comprehensive are used. In order to ensure that the ultimate ranking of an individual town is not adversely affected by a single natural catastrophe, a listing of physical damage experience for each town by month is reviewed and compared with a list obtained from the Insurance Services Office of catastrophes assigned serial numbers during the experience period. The current review indicated that none of the serialized catastrophes produced unusual claim counts that might require adjustment or special treatment.

b. Actual Claim Frequency

The claim frequency in a town for a particular coverage and year is calculated as claims divided by exposures. The claim frequency *index* in a town for a particular coverage and year is the ratio of the town's claim frequency to the statewide claim frequency for the same coverage and year.

A claim frequency index for a town and coverage for all years combined is calculated as the average of the claim frequency index for each year, weighted by the town's exposure by year for the specified coverage. The resulting indices are re-balanced to produce an average index of unity across all towns.

c. Claim Frequency Model

Three explanatory variables affecting the claim frequency in a town are used in the claim frequency model: the traffic density in the town, the class mix in the town, and whether the town is part of Boston. The effect of each of these variables differs from coverage to coverage.

EXHIBIT 8
SHEET 1

CALCULATION OF CLAIM FREQUENCY INDEX MODEL PARAMETERS AND CREDIBILITY PARAMETERS FORMULAS[28]

The basic structure of the claim frequency model is:

$$\begin{aligned}\text{Frequency Index}_{c,t} = {} & A_{0,c} \\ & + A_{1,c} \times \text{Density}_t \\ & + A_{2,c} \times ACRF_{c,t} \\ & + A_{3,c} \times \text{Boston Dummy}_t\end{aligned}$$

where the subscripts c and t refer to coverage and territory;
Density$_t$ is the town density for non-Boston towns;
$ACRF_{c,t}$ is the average class rating factor for the coverage and town;
Boston Dummy$_t$ = 1 in towns which are part of Boston, 0 elsewhere; and
$A_{0,c}$; $A_{1,c}$; $A_{2,c}$; $A_{3,c}$ are regression coefficients.

The regression coefficients are determined separately for each coverage, so the c subscripts will be dropped in the remaining formulas.

With 360 towns in Massachusetts, it is convenient to perform the algebra in matrix notation, which parallels the structure of the APL program used in the analysis:

y is a 360 × 1 vector of actual town claim frequency indices.
$\hat{y}$ is a 360 × 1 vector of model claim frequency indices.
x is a 360 × 4 matrix of the independent variables in the claim frequency model, where
Column 1 is unity;
Column 2 is Density$_t$;
Column 3 is $ACRF_t$;
Column 4 is Boston Dummy$_t$.

[28] For a more detailed exposition of these formulas refer to Siczewicz [27]. Also see DuMouchel and Harris [4].

EXHIBIT 8
SHEET 2

x^T is the 4×360 transpose of x;
A is a 1×4 matrix of the regression coefficients;
W is a 360×360 diagonal matrix of the weights to be applied to each town in the weighted least squares regression; and,
W^{-1} is the inverse of W.

In practice, W is determined from W^{-1}; the entry for each town in W^{-1} is an estimate of the variance of the claim frequency in the town.

The first estimate of this variance in town t is

τ^2/H_t,

where τ^2 is a statewide measure of the year to year variations in claim frequency (see below), and H_t is

$$H_t = \frac{\text{Exposures}_t}{\text{Actual Claim Frequency Index}_t} \quad \text{(all years combined).}$$

That is, given the statewide claim frequency variation, a town with more exposures is estimated to have a lower variance, while a town with a high claim frequency is estimated to have a high variance in claim frequency.

The statewide value of τ^2 is calculated as:

$$\tau^2 = \frac{\sum_t \sum_y H_{t,y} (Y_{t,y} - Y_t)^2}{360 \times (\text{number of years of data} - 1)}$$

where t = town;
y = year;
$H_{t,y}$ = exposures$_{t,y}/Y_t$;
Y_t = claim frequency index for town t, all years combined; and,
$Y_{t,y}$ = claim frequency index for town t, year y.

The first estimate of the regression coefficients A is calculated using a weighted least squares regression

$$A = (X^T WX)^{-1} \quad X^T WY.$$

EXHIBIT 8
SHEET 3

The second estimate of the variance in town t is:

$$(\tau^2/H_t) + \sigma^2,$$

where σ^2 is a measure of the variation of the model from the data. σ^2 is calculated as:

$$\sigma^2 = (\text{RSS} - (n-m))/\left(\left(\sum_t H_t/\tau^2\right) - \text{trace}\,(X^T W^2 X)\,((X^T W X)^{-1})\right),$$

where RSS = Residual sum of squares

$$= \sum_t ((Y_t - \hat{Y}_t)^2/\,(\tau^2/H_t));$$

$\hat{Y}_t$ = Model claim frequency index for town t

$= (XA)_t;$

$n = 360 =$ number of towns; and,

$m = 4 =$ number of years of data.

With the revised values of W^{-1}, W is recalculated, and the final estimate of the regression parameters is

$$A^1 = (X^T WX)^{-1} \quad X^T WY.$$

The credibility assigned to the actual town frequency index is:

$$Z_t = \frac{\sigma^2}{\sigma^2 + \tau^2 \times (\hat{Y}_t \div \text{Exposures}_t)}$$

$$= \frac{(\text{Exposures}_t \div \hat{Y}_t)}{(\text{Exposures}_t \div \hat{Y}_t) + (\tau^2/\sigma^2)}\,.$$

The traffic density in a town is calculated as the ratio of insured exposures[29] in the town to road miles in the town.

The class mix in a town is quantified as the average of the rating class relativities underlying the current rates, weighted by the exposure distribution by class within the town; class mix factors (*ACRF*'s) are calculated by town separately for each coverage.

In order to reduce the possibility of Boston claim frequency patterns distorting the model for the remaining 350 towns, a "dummy" variable is introduced into the model; this variable has a value of unity in Boston, zero elsewhere. In addition, the traffic density variable is set equal to zero in Boston.

The structure of the claim frequency model is

$$\begin{aligned}\text{Model Frequency Index}_{c,t} = {} & A_{0,c} \\ & + A_{1,c} \times \text{Density}_t \\ & + A_{2,c} \times ACRF_{c,t} \\ & + A_{3,c} \times \text{Boston Dummy}_t,\end{aligned}$$

where the subscripts c and t refer to coverage and town, respectively.

d. Model and Credibility Parameters

The values of the model coefficients (the A values in the above equations) are determined empirically for each coverage using the latest four years of data. In addition, the credibility attributable to the actual claim frequency is determined by an analysis of the extent to which the actual claim frequency index contains meaningful information about town frequencies not captured by the model.

The values of the model coefficients are determined for each coverage separately by a weighted least squares regression of actual claim frequencies on density, *ACRF*, and the Boston Dummy variable. The weight applied in the regression analysis to the data for each town is essentially proportional to the credibility assigned to that data. The specific formulas used in this analysis, which determines both the regression parameters and the credibility parameters, are outlined in Exhibit 8.

---

[29] The latest year's PDL exposures are used.

The regression model parameters estimated in the latest review are:

	A-1	A-2	PDL	COMPRE-HENSIVE	COLLISION
Intercept ($A_{0,c}$)	−1.1233	−0.5227	−0.07902	−0.5680	−0.2486
Density coefficient ($A_{1,c}$)	.002142	.0007907	.002672	.002625	.002647
ACRF coefficient ($A_{2,c}$)	1.8124	1.3714	0.7270	1.1949	0.8816
Boston Dummy ($A_{3,c}$)	0.8052	0.6200	0.7320	1.7224	1.3393

For illustrative purposes, collision model claim frequency indices for Holland (rural), Wilmington (suburban), and Brighton (part of Boston) are calculated below:

	HOLLAND	WILMINGTON	BRIGHTON
(1) Town Density (×.002647)	24.7	97.5	0.0
(2) *ACRF* (×.8816)	.9682	1.0528	1.0014
(3) Boston Dummy (×1.3393)	0	0	1
(4) Intercept (−.2486)	−0.2486	−0.2486	−0.2486
(5) Model Claim Freq. Index	.6703	.9376	1.9735
(6) Balancing Factor to Produce Average Index of 1.000 (averaged over all towns, 4 years)	.98704	.98704	.98704
(7) Model Claim Frequency Index, Balanced	.6791	.9499	1.9994

The credibility to be assigned to the actual claim frequency index for a particular town and coverage is calculated as:

$$Z_{c,t} = \frac{H_{c,t}}{H_{c,t} + (\tau^2{}_c/\sigma^2{}_c)}$$

where $Z_{c,t}$ = credibility assigned to actual frequency index for coverage c, town t;

$H_{c,t} = E_{c,t} \div MFI_{c,t}$;

$E_{c,t}$ = exposures for coverage c, town t, all years combined;

$MFI_{c,t}$ = model claim frequency index for coverage c, town t;

τ_c^2 = a measure of the year to year variation in claim frequencies (see Exhibit 8); and,

σ_c^2 = a measure of the extent to which actual claim frequencies differ from model claim frequencies (see Exhibit 8).

The credibility parameters (τ_c^2, σ_c^2) determined in the latest review in accordance with the formulas outlined in Exhibit 8, are

	σ^2	τ^2
A-1	.03898	194.66
A-2	.03574	112.12
PDL	.01327	22.94
Comprehensive	.04078	25.01
Collision	.01816	13.59

Continuing the three town example, credibilities for collision are calculated as follows:

	HOLLAND	WILMINGTON	BRIGHTON
(1) Model Claim Frequency ($MFI_{c,t}$)	.6791	.9499	1.9994
(2) Exposures ($E_{c,t}$)	1,629.5	21,480.6	35,513.6
(3) Collision Credibility	.7623	.9680	.9596

$$\frac{(2) \div (1)}{((2) \div (1)) + (\tau^2/\sigma^2)}$$

e. Formula Frequency Index by Coverage and Town

The formula frequency index for each coverage is the weighted average of the actual frequency index and the model frequency index. The weight accorded the actual frequency index is the credibility, Z, determined in accordance with the above procedure; the model frequency index is calculated using the model parameters determined above.

Algebraically, the formula frequency is calculated as:

$$FF_{c,t} = (Z_{c,t} \times AFI_{c,t}) + ((1 - Z_{c,t}) \times MFI_{c,t})$$

where

$FF_{c,t}$ = Formula frequency index for coverage c, town t; and

$AFI_{c,t}$ = Actual claim frequency index for coverage c, town t.

Continuing the three town example, the formula frequency index values are:

	HOLLAND	WILMINGTON	BRIGHTON
(1) Actual Claim Frequency Index ($AFI_{c,t}$)	.7636	.9671	1.7234
(2) Model Frequency Index ($MFI_{c,t}$)	.6791	.9499	1.9994
(3) Credibility ($Z_{c,t}$)	.7623	.9680	.9596
(4) Formula Frequency Index ($FFI_{c,t}$) = (3)×(1) + (1.0−(3))×(2)	.7435	.9665	1.7346

A3. CALCULATING THE CLAIM COST INDEX

Separately for each coverage, claim severity relativities for each town are estimated. These relativities compare the estimated average claim severity for the town to the statewide average claim severity.

These claim severity relativities for each town are determined as a credibility-weighted average of the town, the county[30], and the statewide claim severity relativities indicated by historical data. The credibility parameters are determined by a Two Layer Hierarchical Empirical Bayes Method.

[30] Barnstable, Dukes, and Nantucket Counties are grouped together as Dukes and Nantucket are too small to remain ungrouped.

The estimated severity for a town is the combination of the severity for the town, the severity for the county that contains the town, and the overall statewide severity. The town's own severity is used to the extent it is credible, with the complement of credibility being given to the estimated severity for the county. In turn, the estimated severity for the county is the credibility-weighted mean of the county severity to the extent it is credible, with the complement of credibility being given to the credibility-weighted statewide severity.

The mechanics of the process are described in Exhibits 9 and 10. The calculated parameters are shown in Exhibit 11. Illustrative examples of the credibility-weighting process are included in Exhibit 12.

The input variables needed for this method of evaluating claim severities by town are:

1. Claims by coverage by year by town.
2. Relative average claim cost by coverage by year by town, modified by average age/symbol relativity by coverage by town[31]

$$= \frac{\text{average claim cost by coverage by year by town}}{\text{average claim cost by coverage by year, statewide}}$$

$$\div \frac{\text{average age/symbol relativity by coverage by town}}{\text{average age/symbol relativity by coverage statewide}}$$

3. County or county group assignments of the towns.

As shown in Exhibit 12, for the three town collision example the methodology yields:

	Holland	Wilmington	Brighton
Estimated Relative Severity	1.0846	1.0508	.9011

[31] The modification for average age/symbol is only needed for comprehensive and collision. This modification is intended to remove from the territory analysis variations between towns that are captured by another rating variable, age/symbol factors.

EXHIBIT 9
SHEET 1

CALCULATION OF THE CLAIM COST INDEX USING THE TWO LAYER HIERARCHICAL EMPIRICAL BAYESIAN CREDIBILITY MODEL SUMMARY OF FORMULAS[32]

Assume a *nested* series of groupings. In this specific implementation, the nested series of groupings is: towns, groups of these towns into counties (actually county groups), and the statewide group of all counties.

Assume an observed variable, X, for each town, for several time periods. (In the territory analysis, X is the relative claim severity.)[33] The intent is to estimate X in the future.

Let $X_{Cg}(t)$ represent X for time t, town g, in county C.

Similarly, let $P_{Cg}(t)$ represent the corresponding measure of exposure (in our case, number of claims).

The use of a dot, instead of a variable, denotes summation over that variable. For example:

$$P_{C.}(.) = \sum_{g,t} P_{Cg}(t), \text{ and}$$

$$W_{C.} = \sum_{g} W_{Cg}.$$

The mean of X, weighted by P, is denoted by $\overline{X}$. For example:

$$\overline{X}_{Cg} = \frac{\sum_{t} X_{Cg}(t)\, P_{Cg}(t)}{\sum_{t} P_{Cg}(t)}, \text{ and}$$

[32] These formulas and their derivatives and implementation were developed and prepared by Howard Mahler and included in Massachusetts Automobile Rating and Accident Prevention Bureau, *Filing for 1986 Private Passenger Territory and Classification Definitions*, July, 1985.

[33] For the physical damage coverages, X is the relative claim severity divided by the relative average age/symbol relativity.

EXHIBIT 9
SHEET 2

$$\bar{X}_C = \frac{\sum_{g,t} X_{Cg}(t)\, P_{Cg}(t)}{\sum_{g,t} P_{Cg}(t)} .$$

Then, given certain assumptions, the least squares estimate for the variable X in the future, denoted by $t = 0$, is given by:

$$X_{Cg}(0) = W_{Cg}\bar{X}_{Cg} + (1 - W_{Cg})m_C$$

where $m_C = V_C M_C + (1 - V_C)\hat{m}$ = estimated relative severity for the county;

$$\hat{m} = \frac{\sum_C V_C M_C}{V_.} = \text{credibility weighted mean overall;}$$

$$W_{Cg} = \frac{P_{Cg}}{P_{Cg} + k_C} = \text{credibility for the town;}$$

$$V_C = \frac{W_{C.}}{W_{C.} + k/k_C} = \text{credibility for the county; and,}$$

$$M_C = \sum_g \frac{W_{Cg}\bar{X}_{Cg}}{W_C} = \text{credibility weighted mean for the county.}$$

The parameters k, k_C are to be estimated from the observed data.

It should be noted that the estimated severity for a town is the combination of the severity for the town, the severity for the county that contains the town, and the overall severity. The town's own severity is used to the extent it is credible, with the complement of credibility being given to the estimated severity for the county. In turn, the estimated severity for the county is the credibility weighted mean of the county to the extent it is credible, with the complement of credibility being given to the credibility weighted severity overall.

EXHIBIT 9
SHEET 3

Let $I(P) = 0$ if $P = 0$
$\quad\quad\quad\; 1$ if $P \neq 0$

and $D_{1C} = \sum_{g,t} P_{Cg}(t)\,(X_{Cg}(t) - \overline{X}_{Cg})^2$

$$D_{2C} = \sum_{g,t} P_{Cg}(t)\,(X_{Cg}(t) - \overline{X}_{C})^2$$

$$D_3 = \sum_{C,g,t} P_{Cg}(t)\,(X_{Cg}(t) - \overline{X})^2$$

Let $E(Y)$ represent the expected value of Y; $E(Y)$ will be *estimated* by the observed value of Y. For example, $E(D_{1C})$ will be estimated by the observed value for D_{1C}.

The estimates of the parameters are as follows:

$$s_C^2 = \frac{E(D_{1C})}{\left[\sum_{g,t} I(P_{Cg}(t))\right] - \left[\sum_{g} I(P_{Cg}(.))\right]}$$

$$k_C = \frac{P_{C}.(.) - \sum_{g} \dfrac{P_{Cg}^2(.)}{P_{C}.(.)}}{\dfrac{E\,(D_{2C})}{s_C^2} - \left[\sum_{g,t} I(P_{Cg}(t)) - I(P_{C}.(.))\right]}$$

$$k = \frac{\sum_{C} s_C^2 \left[P_{C}.(.) - \dfrac{P_{C}^2.(.)}{P_{..}(.)}\right]}{E\,(D_3) - \sum_{Cgt} s_C^2 \left[I\,(P_{Cg}(t)) - \dfrac{P_{Cg}(t)}{P_{..}(.)}\right] - \sum_{C} \left[s_C^2 \dfrac{P_{C}.(.) - \sum_{g} \dfrac{P_{Cg}^2(.)}{P_{..}(.)}}{k_C}\right]}$$

EXHIBIT 10
SHEET 1

CALCULATION OF THE CLAIM COST INDEX USING THE TWO LAYER HIERARCHICAL EMPIRICAL BAYESIAN CREDIBILITY MODEL IMPLEMENTATION

In the use of the Empirical Bayesian Credibility Model described in Exhibit 9 to calculate average claim costs by town, some fluctuations in the calculated values would be expected, since the parameters of the model are being calculated from only a limited quantity of data. For the practical implementation of the model, it is desirable to eliminate undue fluctuations.

Limitations on s_g^2

The parameters s_C^2 are estimated separately for each county (and each coverage). Since certain counties are relatively small, the computed value of s_C^2 can be subject to undue fluctuations.

$$s_C^2 = \frac{\sum_{g,t} P_{Cg}(t)\ (X_{Cg}(t) - \overline{X}_{Cg})^2}{\left[\sum_{g,t} I(P_{Cg}(t))\right] - \left[\sum_{g} I(P_{Cg}(.))\right]}$$

s_C^2 can be viewed as a weighted average of s_g^2 for each town g in the county C, where

$$s_g^2 = \frac{\sum_{t} P_{Cg}(t)\ (X_{Cg}(t) - \overline{X}_{Cg})^2}{\left[\sum_{t} I(P_{Cg}(t))\right] - I(P_{Cg}(.))};$$

the weights[34], w_g, are

$$w_g = \frac{\left[\sum_{t} I(P_{Cg}(t))\right] - I(P_{Cg}(.))}{\sum_{g,t} I(P_{Cg}(t)) - \sum I(P_{Cg}(.))};\ \text{and,}$$

[34] The weights for all towns are equal if every town has at least one claim in each year. Those towns in which no claims occurred in some years would receive less weight.

EXHIBIT 10
SHEET 2

s_C^2 is defined as

$$s_C^2 = \sum_g w_g s_g^2 .$$

Since s_C^2 is a weighted average of s_g^2 for individual towns, a reasonable way to limit variations in s_C^2 is to limit the contribution made by any individual town. This can be accomplished by restricting the value of s_g^2 that enters the computation of s_C^2 to lie between chosen minimum and maximum values. The minimum and maximum values can be chosen as a factor times the overall s^2 (which is a weighted average of s_g^2 over all towns in the state). Factors of 1/5 and 5 were chosen judgmentally.

Thus, in computing s_C^2 for each county, s_g^2 for each town was restricted to be within a range of 1/5 or 5 times the overall s^2 for all counties.

$$s^2 = \frac{\sum_{Cgt} P_{Cg}(t)\,(X_{Cg}(t) - \overline{X}_{Cg})^2}{\sum_{Cgt} I(P_{Cg}(t)) - \sum I(P_{Cg}(.))}$$

$$\tilde{s}_g^2 = \begin{cases} 1/5\ s^2 & \text{if } s_g^2 \leqq 1/5\ s^2 \\ s_g^2 & \text{if } 1/5 s^2 \leqq s_g^2 \leqq 5s^2 \\ 5s^2 & \text{if } s_g^2 \geqq 5s^2 \end{cases}$$

$$\tilde{s}_C^2 = \sum w_g \tilde{s}_g^2$$

The resulting values of $\tilde{s}_C^2$ which were used in the review for 1986 are displayed in Exhibit 11.

Limitations on k values

Even with the application of these limitations, calculated k values may exhibit some fluctuations. Therefore, for each coverage, the credibility parameters k and k_C (k applies to the state, while there is a k_C for each county) are limited by the imposition of a maximum value and a minimum value. When

EXHIBIT 10
SHEET 3

the calculated value was less than the minimum, the value of the parameter was set equal to that minimum.[35] When the calculated value was more than the maximum value, the value of the parameter was set equal to the maximum value.

The choice of maximum and minimum values for k and k_C involves the use of some actuarial judgment, although tests indicated that the resulting combined indices for towns are relatively insensitive to these choices. A maximum value of 2500 claims and a minimum value of 100 claims were used for all coverages. The resulting values of k and k_C which were used in the latest territory review are displayed in Exhibit 11.

[35] In certain cases, the calculated value of the parameter k_c was a large negative number. This occurred when the calculated denominator was negative because the observed variations of the average claim costs between the towns within a county were small relative to the observed variation of the average claim costs within the individual towns from year to year. (For the overall k this would have occurred if the observed variations of the average claim costs between the different counties were small relative to the observed variation within a county from year to year.) This case was treated as an extension of the case where the calculated denominator was a very small positive quantity, and the calculated parameter was a very large positive quantity. Thus in those cases where the calculated parameter was negative, its value was set equal to the ,chosen maximum value. This choice has the appropriate effect on credibilities: it will assign less credibility to the towns within a county and more to the county.

EXHIBIT 11
SHEET 1

CALCULATION OF THE CLAIM COST INDEX USING THE TWO LAYER HIERARCHICAL EMPIRICAL BAYESIAN CREDIBILITY MODEL

	Credibility Parameter K, Severity				
County Group	BI	PIP	PDL	Comp.	Coll.
Overall	548	740	515	143	1026
Barnst., Dukes, Nant.	132	571	501	766	2500
Berkshire	100	268	631	1276	100
Bristol	309	339	165	929	153
Essex	1503	703	1319	100	387
Franklin	2500	2500	2500	317	211
Hampden	1812	356	2500	185	342
Hampshire	2500	2500	1731	898	202
Middlesex	2500	544	2500	125	199
Norfolk	2500	667	1272	245	269
Plymouth	421	409	978	380	646
Suffolk	1005	332	1157	696	2500
Worcester	2500	227	631	719	177

EXHIBIT 11
SHEET 2

CALCULATION OF THE CLAIM COST INDEX USING THE TWO LAYER HIERARCHICAL EMPIRICAL BAYESIAN CREDIBILITY MODEL

	Credibility Parameter S^2, Severity				
County Group	BI	PIP	PDL	Comp.	Coll.
Overall	1.7	2.0	1.2	4.1	1.6
Barnst., Dukes, Nant.	1.4	2.3	1.3	2.7	1.6
Berkshire	1.4	2.2	1.3	2.5	1.4
Bristol	1.8	2.6	1.1	7.2	1.7
Essex	1.5	1.7	1.1	3.4	1.1
Franklin	2.1	1.5	1.3	1.8	1.9
Hampden	1.7	1.7	1.0	4.2	2.1
Hampshire	1.6	2.2	1.3	2.0	1.8
Middlesex	1.8	2.0	1.3	4.4	1.3
Norfolk	1.7	1.7	1.4	3.9	1.6
Plymouth	1.6	1.7	1.2	5.4	2.0
Suffolk	1.4	1.8	1.7	13.6	2.0
Worcester	1.9	1.6	1.0	3.0	1.7

EXHIBIT 12
SHEET 1

PRICING EXAMPLES—SEVERITY METHODOLOGY

BRIGHTON
SUFFOLK — TOWN'S PDL EXPOSURES 15,872.8

	BI	PIP	PDL	COMP.	COLL.
(1) Claims for Town	657.0000	1,197.0000	5,569.0000	6,770.0000	6,388.0000
(2) Cred. Weighted Mean for County Group	1.0944	1.1421	1.0718	1.3749	.9381
(3) Overall K (Claims)	548.0728	740.3648	515.3687	142.6349	1,026.1405
(4) Credibility for County Group	.9034	.8148	.9580	.9825	.9526
(5) Cred. Weighted Mean Overall	.9968	.9891	.9900	.8527	1.0684
(6) Est. Rel. Sev., Cnty. = (2) × (4) + (5) × 1 − (4)	1.0849	1.1138	1.0684	1.3657	.9443
(7) Actual Relative Sev. for Town	1.1094	1.0858	1.0178	1.1788	.8842
(8) K for County Group (Claims)	1,004.7278	331.8594	1,156.6019	696.4471	2,500.0000
(9) Cred. for Town = (1)/((1) + (8))	.3954	.7829	.8280	.9067	.7187
(10) Est. Rel. Sev. for County Group = (6)	1.0849	1.1138	1.0684	1.3657	.9443
(11) Est. Rel. Sev., Town = (7) × (9) + (10) × 1 − (9)	1.0946	1.0919	1.0265	1.1962	.9011

EXHIBIT 12
SHEET 2

PRICING EXAMPLES—SEVERITY METHODOLOGY

HOLLAND
HAMPDEN Town's PDL Exposures 914.0

	BI	PIP	PDL	Comp.	Coll.
(1) Claims for Town	14.0000	56.0000	164.0000	105.0000	128.0000
(2) Cred. Weighted Mean for County Group	.9334	.9513	.9325	.7344	1.0438
(3) Overall K (Claims)	548.0728	740.3648	515.3687	142.6349	1,026.1405
(4) Credibility for County Group	.9167	.8349	.9743	.9514	.8162
(5) Cred. Weighted Mean Overall	.9968	.9891	.9900	.8527	1.0684
(6) Est. Rel. Sev., Cnty. = (2) × (4) + (5) × 1 − (4)	.9387	.9575	.9340	.7402	1.0483
(7) Actual Relative Sev. for Town	.9478	.9118	1.0217	.6832	1.1810
(8) K for County Group (Claims)	1,812.2624	355.5324	2,500.0000	184.7182	342.3312
(9) Cred. for Town = (1)/((1) + (8))	.0077	.1361	.0616	.3624	.2737
(10) Est. Rel. Sev. for County Group = (6)	.9387	.9575	.9340	.7402	1.0483
(11) Est. Rel. Sev., Town = (7) × (9) + (10) × 1 − (9)	.9388	.9513	.9394	.7195	1.0846

EXHIBIT 12
SHEET 3

PRICING EXAMPLES—SEVERITY METHODOLOGY

WILMINGTON
MIDDLESEX   TOWN'S PDL EXPOSURES 10,232.4

	BI	PIP	PDL	COMP.	COLL.
(1) Claims for Town	419.0000	734.0000	3,010.0000	2,386.0000	2,170.0000
(2) Cred. Weighted Mean for County Group	1.0063	.9865	1.0308	.8848	1.0383
(3) Overall K (Claims)	548.0728	740.3648	515.3687	142.6349	1,026.1405
(4) Credibility for County Group	.9726	.9524	.9923	.9772	.9014
(5) Cred. Weighted Mean Overall	.9968	.9891	.9900	.8527	1.0684
(6) Est. Rel. Sev., Cnty. = (2) × (4) + (5) × 1 − (4)	1.0060	.9866	1.0305	.8841	1.0413
(7) Actual Relative Sev. for Town	1.0411	.9717	1.0661	.9980	1.0517
(8) K for County Group (Claims)	2,500.0000	543.7111	2,500.0000	124.9303	198.6265
(9) Cred. for Town = (1)/((1) + (8))	.1435	.5745	.5463	.9502	.9161
(10) Est. Rel. Sev. for County Group = (6)	1.0060	.9866	1.0305	.8841	1.0413
(11) Est. Rel. Sev., Town = (7) × (9) + (10) × 1 − (9)	1.0111	.9781	1.0499	.9923	1.0508

A4. FINAL TOWN INDEX

This section describes the combining of the claim frequency indices (from Section A2) and claim cost indices (from Section A3) and the determination of an overall index that incorporates all coverages.

The first step is to calculate a pure premium index by town and coverage. This index is simply the product of the claim frequency index and the claim cost index, and is interpreted as being a measure of the town's pure premium (average insurance loss dollars per vehicle) relative to the statewide average pure premium.

Any town-to-town variation in pure premiums that is captured by other rating variables, however, should not also influence a town's territory assignment. Therefore, each town's pure premium index is adjusted to remove the effects of the mix of insured drivers by driver classification[36] as measured by the *ACRF* described above. The resulting town net pure premium indices are re-balanced to unity within each coverage. In the three town example for collison:

	HOLLAND	WILMINGTON	BRIGHTON
(1) Claim frequency index	.7435	.9665	1.7346
(2) Claim cost index	1.0846	1.0508	.9011
(3) Pure premium index $= (1) \times (2)$	.8064	1.0156	1.5630
(4) Average class rating factor	.9682	1.0528	1.0014
(5) Net pure premium index, re-balanced to unity $= ((3) \div (4)) \div 1.00015$	.8328	.9645	1.5606

Finally, an average index across all coverages (c) is calculated for each town (t) by weighting the coverage net pure premium indices. The weight assigned to each coverage depends on the number of exposures purchasing the coverage and on the statewide pure premium for the coverage:

$$\frac{\sum_c \text{Exposures}_{c,t} \times \text{Statewide Pure Prem}_c \times \text{Net Pure Prem Index}_{c,t}}{\sum_c \text{Exposures}_{c,t} \times \text{Statewide Pure Premium}_c}.$$

[36] As noted in Section A3, a corresponding adjustment to remove the effects of varying distributions by age and symbol is incorporated in the claim severity index calculation.

The resulting index is balanced to unity (on the latest year's PDL exposures) across all towns.

Applying the above formula to the three towns:

	HOLLAND	WILMINGTON	BRIGHTON
(1) Exposure (latest year)			
A-1, A-2, PDL	914.0	10,232.4	15,872.8
Comprehensive	533.7	7,176.6	11,806.0
Collision	422.2	5,794.7	9,679.2
(2) Net Pure Premium Index			
A-1	.5397	1.0063	1.2748
A-2	.8712	.9694	1.2898
PDL	.6477	1.0451	1.3588
Comprehensive	.4253	.9674	2.0844
Collision	.8328	.9645	1.5606
(3) Statewide Average Pure Premium			
A-1	38.61	38.61	38.61
A-2	14.92	14.92	14.92
PDL	62.01	62.01	62.01
Comprehensive	57.28	57.28	57.28
Collision	120.00	120.00	120.00
(4) Balancing Factor	1.0011	1.0011	1.0011
(5) Weighted average net pure premium index	.6567	.9938	1.5361

The resulting index is used to rank the 360 towns according to their loss potential. For the three town example the ranks are:

	RANK
Holland	30
Brighton	349
Wilmington	302

APPENDIX B
HOMOGENEITY AND HOMOGENEITY MEASURES[37]

B1. INTRODUCTION

As discussed in the body of this paper, one of the criteria by which alternative territory schemes are assessed is homogeneity; i.e., towns within the same territory grouping should possess similar inherent loss potential. If the territories are to be homogeneous then no town's loss potential measure should differ substantially from the average loss potential measure of all towns in that territory. This notion can be used formally to construct several quantitative indices which then can be used to guide the ratemaker in some of the grouping judgments which need to be made.

This appendix defines the indices that have been constructed for use in Massachusetts; all of them are referred to as *homogeneity measures* and are displayed in Exhibit 6.

B2. LOSS POTENTIAL

There are two readily available data sources which can be used to indicate a town's loss potential. One is the value of the combined index produced by the procedure described in Appendix A and displayed in Exhibit 5 for a sample of towns. Another is the actual latest three year experience pure premiums for the liability coverages and for the typical package of coverages.[38] Exhibit 5 also displays these pure premiums for a sample of towns. Each measure has relevance. The combined index is a true credibility weighted estimate of a synthetic pure premium relationship between towns, while the actual three year pure premiums are the data used to set territory relativities in the ratemaking process. Rather than choose between these two measures, both are used as homogeneity indicators.

---

[37] This appendix was taken, with minor editing, from sections of the Massachusetts Automobile Rating and Accident Prevention Bureau's *Filing for 1986 Private Passenger Territory and Classification Definitions*, July, 1985. These sections of the MARB's filed analysis, including the specific homogeneity measures, were developed and prepared by Dr. Richard Derrig.

[38] The liability coverages consist of basic limits (10/20) A-1, PDL (5,000) and A-2. The package coverages consist of A-1 (10/20), PDL (5,000), A-2, Collision, and Comprehensive.

Homogeneity Measures

This section defines several measures of the homogeneity of a territory grouping procedure. In general, the measures test the difference between the town's loss potential and the average of the entire territory's loss potential. The measures utilize both the actual pure premium and the combined index values of loss potential. The first tests calculate both the average absolute squared difference (measure 1) and the percentage squared difference for the pure premium values. Since the latter will measure the percentage difference from the town's actual pure premium, which might be unstable for small towns, this measure is calculated with (measure 2) and without (measure 2A) a credibility weight for the reliability of the actual data. In order to test the average spread of the territory grouping, the next measures rely on the average maximum deviations of the town value from the territory average both using the absolute difference (measure 4), percentage difference with (measure 5) and without (measure 5A) a credibility weight, and the model combined index (measure 6). The precise definitions are listed in Exhibit 13. For all these measures, a homogeneity value closer to 0 indicates a more homogeneous set of territories.

B3. ERROR ENTROPY

One further measure of homogeneity can be defined based upon the information-theoretic concept of entropy. In general, entropy quantifies the degree of disorder or uncertainty in a system. An entropy-like measure is applied to determine the disorder or uncertainty in the difference between a town's combined index and the territory average index. In a sense, that difference is the "error" which results when the territory average index is assigned to the town. This is the assumption of perfect homogeneity. The entropy measure will then quantify the relative information about the concentration of these errors among territory grouping procedures. The notion of entropy has been used in a somewhat similar way by Garrison and Paulson [5] to compare concentrations in economic activity over time.

Consider a set of k categories $C_1, \ldots, C_k$ and a random sample of size n. Each observation of the sample falls into one of the categories C_i with some fixed probability $p_i > 0$; $i = 1, 2, \ldots, k$ with $\Sigma p_i = 1$, and in the sample a total of n_i observations fall into category C_i. Then the *entropy* or expected information of the system is defined by

$$H = \sum_{i=1}^{k} p_i \log p_i \, .$$

EXHIBIT 13
SHEET 1

HOMOGENEITY MEASURE DEFINITIONS

MEASURE	DEFINITION
1. Pure Premium Squared Diff.	$\sum_{\text{Town}_i} 83\ \text{EXP}_i\ (\text{Town PP}_i - \text{Terr PP}_i)^2 \div \sum_{\text{Town}_i} 83\ \text{EXP}_i$
2. Pure Premium Cred. Wgtd. Percentage Squared Diff.	$\sum_{\text{Town}_i} 83\ \text{EXP}_i\ \text{Max Cred}_i \left(\frac{\text{Town PP}_i - \text{Terr PP}_i}{\text{Town PP}_i}\right)^2 \div \sum_{\text{Town}_i} 83\ \text{EXP}_i$
2a. Pure Premium Percentage Squared Diff.	$\sum_{\text{Town}_i} 83\ \text{EXP}_i \left(\frac{\text{Town PP}_i - \text{Terr PP}_i}{\text{Town PP}_i}\right)^2 \div \sum_{\text{Town}_i} \text{EXP}\ 83_i$
3. Index Squared Diff.	$\sum_{\text{Town}_i} 83\ \text{EXP}_i\ (\text{Town Ind}_i - \text{Terr Ind}_i)^2 \div \sum_{\text{Town}_i} 83\ \text{EXP}_i$
4. Pure Premium Maximum Diff.	$\sum_{\text{Terr}_i} 83\ \text{EXP}_i\ \max_i \lvert \text{Town PP}_i - \text{Terr PP}_i \rvert \div \sum_{\text{Town}_i} 83\ \text{EXP}_i$

EXHIBIT 13
SHEET 2

HOMOGENEITY MEASURE DEFINITIONS

MEASURE	DEFINITION
5. Pure Premium Cred. Wgtd. Percentage Max. Diff.	$\sum_{\text{Terr}_i} 83 \text{ EXP}_i \operatorname*{Max}_i \text{ Max Cred}_i \left\lvert \frac{\text{Town PP}_i - \text{Terr PP}_i}{\text{Town PP}_i} \right\rvert \div \sum_{\text{Town}_i} 83 \text{ EXP}_i$
5a. Pure Premium Percentage Max. Diff.	$\sum_{\text{Terr}_i} 83 \text{ EXP}_i \operatorname*{Max}_i \left\lvert \frac{\text{Town PP}_i - \text{Terr PP}_i}{\text{Town PP}_i} \right\rvert \div \sum_{\text{Town}_i} 83 \text{ EXP}_i$
6. Index Max. Diff.	$\sum_{\text{Terr}_i} 83 \text{ EXP}_i \operatorname*{Max}_i \lvert \text{Town Ind}_i - \text{Terr Ind}_i \rvert \div \sum_{\text{Town}_i} 83 \text{ EXP}_i$
7. Error Entropy	$-\sum_{e_i} (\text{EXP}_{(e_i)}/\text{EXP}) \log (\text{EXP}_{(e_i)}/\text{EXP})$

EXHIBIT 13
SHEET 3

HOMOGENEITY MEASURE DEFINITIONS

Notational Conventions

1. 83 EXP_i	means the 1983 PDL exposure in earned car years for Town i.
2. Town PP_i	means the pure premium of 1981–1983 losses divided by 1981–1983 earned car years for Town i.
3. Terr PP_i	means the pure premium of 1981–1983 losses divided by 1981–1983 earned car years for all towns in the territory containing Town i.
4. Max $Cred_i$	means the maximum of the Empirical Bayes produced credibility values for all coverages (5 or 6) for Town i.
5. Town Ind_i	means the model combined index for Town i.
6. $EXP_{(e_i)}$	means the total earned car years of exposure for all towns whose "error," Town Ind_i − Terr $Ind_i = e_i$, lies in the interval (e_i).
7. EXP	means total exposure in earned car years.

The underlying probabilities p_i indicate the strength or concentration of the category C_i. On a sampling basis, for purposes of the current analysis, entropy is defined by the approximation[39]

$$h = - \sum_{i=1}^{k} (n_i/n) \log (n_i/n).$$

The greatest uncertainty occurs when H (or h) is the maximum value of log k, while the least uncertainty (most categorial information) occurs when H (or h) equals zero.

The construction of territories seeks the information content for the per exposure error in territory index assignment to towns. Assuming homogeneous towns, the sample size is the total exposure n. The categories are intervals of errors. (For this application, intervals of .01 were chosen to define categories.)

C_{-2}		C_{-1}		C_0		C_1		C_2
	$-.01$		0		.01		.02	

Thus, define:

$$n_i = \sum_t \text{Town Exposure}_t$$

when e_t = Town Index$_t$ − Territory Index$_t$ falls into C_i. Then, the entropy measure h will define the "concentration" of the errors e_t. The smaller the value of h, the more homogeneous the territory grouping will be. This is designated as homogeneity measure 7 and labelled the "Error Entropy" measure.

39 As usual, if $n_i = 0$ then $(n_i/n) \log (n_i/n) = 0$.

REVISIONS IN LOSS RESERVING TECHNIQUES NECESSARY TO DISCOUNT PROPERTY-LIABILITY LOSS RESERVES

STEPHEN P. D'ARCY

Abstract

Statutory accounting principles for property-liability insurers in the United States, in all but very special circumstances, do not recognize the time value of money in the establishment of loss reserves. The Tax Reform Act of 1986 stipulates an interest rate and a methodology for discounting loss reserves for tax purposes. The National Association of Insurance Commissioners (NAIC) is studying the discounting issue. Insurers need to consider the appropriate procedures and interest rates to be used in discounting loss reserves. This paper proposes a method of calculating loss payout patterns based on paid loss development data combined with other reserving techniques that would minimize the additional effort involved in adopting discounting. It also analyzes the repercussions of adopting discounting for statutory accounting purposes.

Discounting loss reserves would have both positive and negative effects on the property-liability insurance industry. Discounting at an appropriate interest rate would increase the usefulness of the combined ratio as a profitability measure, with values less than 100 indicating profits and in excess of 100 indicating losses, subject to the accuracy of loss reserves. Statutory surplus would increase as a result of discounting, which, although having no real economic effect, might provide more capacity for the insurance industry due to regulatory reliance on statutory values. Conversely, discounting would increase the complexity of loss reserving, create a dependence of reserve adequacy on future interest rate levels, and increase the expenses of insurers by raising tax levels. Discounting would have its greatest impact on commercial and professional liability insurers.

1. INTRODUCTION

The Revenue Act of 1921 established the statutory accounting principles of the property-liability insurance industry as the basis for determining federal income taxes. These accounting principles include the provision for an unearned premium reserve that ignores prepaid expenses, thus leading to an equity in the unearned premium reserve. These principles also establish that the loss reserves represent the best estimate of total future payments on losses that have already occurred regardless of when the payment is to be made. Discounting, although allowed in specific instances of periodic payments, is generally not used. Statutory accounting principles are based on the need to assure company solvency and, in most instances, are recognized as being conservative.

Several recent developments led the federal government to reconsider the provisions of the Revenue Act of 1921. The property-liability insurance industry has been extremely unprofitable from 1982 through 1986, based on statutory accounting principles, reducing federal income tax receipts. The industry received tax refunds of approximately $1.7 billion in 1984 and $2.0 billion in 1985 for taxes paid in prior years [16, 21]. New forms of insurance transactions also demonstrate that, in times of high interest rates, the opportunity to use undiscounted loss reserves can lead to tax driven financial transactions. A group of insurers provided retroactive liability insurance at a price below expected losses to MGM Grand Hotels after a major fire had occurred. Leading to this below full cost pricing was the knowledge that the underwriting loss created by this transaction would shelter other income from taxes and the premium income would be invested for a number of years before the loss would be paid [28]. In another case, a large insurer with a surfeit of tax losses sold loss reserves to an insurer in a tax paying situation by transferring responsibility for paying losses to the other insurer and paying that insurer a sum less than the value of the loss reserves. The first insurer immediately booked an underwriting profit and the second an underwriting loss on the transaction [15]. Finally, an important motive behind the development of captive insurers is for noninsurance corporations to obtain the right to use insurance accounting techniques for their self insurance programs by meeting whatever legal constraints apply [27].

The combined ratio is the total of the loss ratio and the expense ratio. Traditionally, an insurer is considered profitable as long as the combined ratio is below 100 percent. The use of an undiscounted loss ratio generates problems with this benchmark because insurers can operate profitably with combined ratios well in excess of 100 percent. An alternative profitability measure is the operating ratio, which subtracts the ratio of investment income to earned premium from the combined ratio. Often an operating ratio less than 100 percent is considered profitable for the insurer in total by combining underwriting and investment results. Two problems arise from this measure. First, the investment income value includes interest and dividend income and realized capital gains and losses, but does not include unrealized gains or losses. The realized gains may have been generated in the current period or in prior years. Thus the investment income does not really reflect the achieved rate of return in the current period. Second, the investment income is based to a large extent on prior periods' premiums collected, loss reserves established, and investments made. It does not reflect the future investment experience on the current book of business as it develops. Therefore, the operating ratio is an inexact profitability measure.

Although the emphasis of the discounting issue has involved loss reserves, premiums may also need discounting. If the premium is paid after the coverage period, as is the case for paid loss retrospective contracts, premiums must be discounted if losses are discounted.

The General Accounting Office (GAO) proposed requiring property-liability insurers to discount loss reserves for determining federal income taxes [10, 14]. This provision would immediately boost insurer taxable income which would increase the amount of Federal taxes payable by the property-liability insurance industry. Use of tax loss carry-forwards could delay the impact of the increased tax level. Under the GAO proposal, loss reserves would be discounted based on the average pre-tax investment income rate achieved by each insurer over the preceding five years. The Treasury Department recommended requiring property-liability insurers to establish qualified reserve accounts (QRA) as a method of discounting loss reserves for all policies issued on or after January 1, 1986 [13, 23]. This proposal allows insurers to establish their own procedures and interest rates for the QRA, subject to approval of the Internal Revenue Service. Under certain circumstances, the QRA method is equivalent to applying a cash accounting system to losses.

The Tax Reform Act (TRA) of 1986 includes five changes in property-liability insurance taxation in addition to the general corporate tax changes. Starting in 1987, loss reserves are to be discounted using the applicable federal rate on midmaturity (three to nine year) securities based on the five year period prior to the calendar year for which discounting is applied. Months prior to August, 1986, however, are not included in determining the discount rate. A "fresh-start" approach applies under which beginning reserves are treated as having been discounted, but the change in accounting profits generated by applying discounting to previously undiscounted loss reserves is not taxed. Insurers can use either loss payout patterns calculated by the Treasury Department or company payout patterns. In addition to discounting loss reserves, 20 percent of the change in unearned premium reserve is included in taxable income, the loss reserve deduction is reduced by 15 percent of tax-exempt interest and dividends received on investments made after August 7, 1986, the protection against loss account (PAL) for mutuals is eliminated, and special deductions for small mutual insurers are rescinded. Of the general corporate tax provisions included in TRA, applying the alternative minimum tax to book earnings, which include tax-exempt income, will also significantly affect property-liability insurance operations.

All federal discounting provisions apply only to loss reserve deductions used in determining taxable income. They do not address the issue of discounting statutory loss reserves, which have always been subject to state regulation. The current situation requires maintaining statutory loss reserves as stipulated by state insurance law and separately calculating the discounted loss reserves for income tax purposes. The National Association of Insurance Commissioners is also considering loss reserve discounting, although no model regulations have been adopted. A number of industry trade associations have raised issues related to discounting [1, 9].

By not discounting loss reserves, insurers are maintaining a safety margin, which varies by reserve accuracy, interest rates, and loss payout patterns. There is no formal recognition of this safety margin and it is not generally quantified. If loss reserves were discounted, this safety margin would be eliminated. In its place, some actuaries propose the establishment of a formal risk loading. This risk loading would vary with the size and degree of accuracy of the loss reserve. It could vary by line and by insurer. If such a risk loading were adopted as an allowable deduction, it would serve to reduce the tax impact of discounting and improve the theoretical support for conservatism in statutory accounting.

The purposes of this paper are to determine what steps property-liability insurers would have to take in order to comply with loss reserve discounting and to analyze the repercussions of these changes. This research demonstrates the effect of discounting on the industry and proposes a methodology for insurers to calculate loss payout patterns based on company data.

2. LOSS RESERVING TECHNIQUES

Currently a number of loss reserving techniques are used to determine the value for the loss reserve. For statutory accounting purposes, actuaries need only project the total amount to be paid in the future for losses that have already occurred (or have been reported for claims-made coverage), without any concern about when the loss will be paid. The one exception is for periodic payments under workers' compensation. The difficulty of achieving this goal is apparent by observing the accuracy of past loss reserve figures. Numerous studies have indicated that large errors in loss reserves, either under or overreserving, have occurred from the 1960s through the most recent reserves tested. Forbes [12], Anderson [2], and Balcarek [3] demonstrate that loss reserves for the industry were progressively less adequate through the 1960s. Smith [26] determines a pattern of overreserving during the period 1955–1961, underreserving for 1962–1970, overreserving for 1971–1972, and underreserving for 1973–1974, for a sample of insurers' automobile liability loss reserves. Weiss [30] shows that reserving errors tend to stabilize insurer profitability.

A number of specific loss reserving techniques are described and critiqued in the actuarial literature [24, 25]. Among the more commonly used reserving procedures are individual case estimates, the average value method, the loss ratio method, incurred loss development, and paid loss development. Also, for each basic technique a number of enhancements have been proposed to deal with special circumstances. Each technique has its advantages and disadvantages. Generally actuaries recommend using more than one technique and establishing the loss reserve at the level about which several methods cluster.

The paid loss development reserving technique, described in detail later, is readily adaptable to discounting. However, insurers should not emphasize this reserving technique and dismiss the other reserving methods simply due to this feature. Actuaries should continue to determine loss reserves based on a variety of reserving techniques and then apply the paid loss development data, as demonstrated in this paper, to establish the loss payout pattern. The primary loss reserving techniques will be presented and critiqued to demonstrate the need for reliance on a manner of calculations in establishing the loss reserve.

Individual Case Estimates

Under the individual case estimates method of loss reserving, claims department personnel assign an individual value to each known claim. The total loss reserve is the sum of all the individual claim estimates, with an adjustment to reflect historical differences between the total case reserve and ultimate loss development. This adjustment covers the incurred but not reported loss reserve plus or minus any systematic underreserving or overreserving on the case estimates. The individual case estimates method is accurate only if any bias in individual case reserving estimates is consistent and if claim reporting patterns do not change. The case reserve value is based on the presumed final settlement value of the claim and does not consider the length of time until settlement. This method does not provide any information concerning when the loss is likely to be paid.

One problem with this reserving methodology is the learning process of claim personnel. As these individuals develop more expertise in settling claims, any consistent bias they may have reflected in prior years could be corrected. For example, a claims person who consistently underreserved losses is likely to increase reserve values. If this change occurred throughout the claims department, the adjustment made to total case reserves based on historical factors would prove to be inaccurate.

Another problem is the effect of shifts in reporting patterns. If new claim procedures increase the speed of entering claims into the system, or if a weekend or other work interruption delays recording claims at the end of a reporting period, this method could be incorrect. Consistency in both claim estimation and reporting is necessary for the individual case estimate method to be accurate.

Average Value Method

The average value method of loss reserving uses claim counts and average claim values to determine the loss reserve. If this method is used to value reported claims only, the number of reported but unsettled claims is multiplied by an estimate for the average cost of settling the claims. Individual loss estimates are not material. If this method is used to value the total reserve, the total number of claims is projected from reported claims based on historical claim reporting patterns. Average claim values are projected from prior claim payments, with the recognition that larger claims tend to be settled more slowly than smaller claims.

The average value loss reserve method provides no information on when a claim is to be paid. Although this procedure does not depend on consistency in claims department reserving estimates, it does depend on consistency in reporting and settlement patterns. Also, the projection of average values, based on historical averages and trends, must be accurate. Changes in the rate of inflation or other factors that affect claim severity, such as deductibles or policy limits, must be considered.

A commonly used combination of reserving techniques is for insurers to use the average value reserving method for quickly settled claims. After a claim has been open for a period of time, a case estimate method is used. In this situation, the strengths and weaknesses of each method apply depending on the length of time the claim is open. For claims that have not been open long, on which information is likely to be incomplete, average values are used to establish the reserve. The simple cases that are settled quickly never change value using this reserving method. As a case remains open and the opportunity exists for more information to be collected, individual case reserve estimates are used. During the average reserve period, reporting patterns must be consistent for this method to produce accurate reserves. Also, the method used to determine average claim values must be accurate. For the time that the case estimate method is used, reserving bias and reporting patterns have to be consistent for the method to generate accurate reserves. The major advantage of this combination of reserving methods is that claims personnel need not maintain reserving consistency prior to the investigation of the claim.

Loss Ratio Method

The loss ratio method of loss reserving determines the reserve by subtracting the losses paid to date from the total expected losses. Total expected losses are calculated by multiplying the expected loss ratio by the earned premium. Changes in claim reporting patterns, bias in establishing case reserves, and shifts in average claim values do not affect the accuracy of this reserving procedure. As long as the ultimate loss ratio estimate is accurate, this procedure will be correct. Any inaccuracy in the loss ratio estimate, however, generates inaccurate loss reserves.

This method of loss reserving does not provide any information on when the loss is to be paid. It is a useful method when the expected loss ratio can be projected accurately, and claim reporting and reserving patterns have not been consistent. For lines of business with long loss payment tails, this method can be risky for an insurer, since rates are established from past loss experience and any inaccuracy in this loss reserving procedure would not be apparent for a long time.

Incurred Loss Development

The incurred loss development method of loss reserving calculates the loss reserve by projecting current incurred losses, which are paid losses plus outstanding case reserves, to ultimate incurred loss levels based on historical development patterns. The loss reserve is the total projected incurred losses minus losses paid to date. Outstanding reserves may be established on an average value basis, by individual case estimates, or by a combination of these methods. Unlike the case estimate reserving method, losses paid to date are also used in projecting ultimate losses.

Partial and ultimate incurred loss development factors are calculated from historical information. Partial loss development factors are generally determined by examining the change in incurred losses for a specific accident year (or other exposure period) from one report period to the next. The ultimate incurred losses are not known until all losses are settled which, for liability lines, can take decades. Reliance on loss development factors based on an era when conditions may have been considerably different from the current time introduces substantial risk into the reserving process. A commonly used technique in this reserving method is to combine partial incurred loss development factors with ultimate development factors. This technique combines the currency of recent development experience for the most volatile segment of the reserve period with the stability of older values for the remaining period.

This method of loss reserving does not provide information on when losses are to be paid. The accuracy of this method depends on consistency in loss reporting, settlement, and reserving. It is less sensitive to changes in loss reserving than the case estimate methodology since paid losses are also included. This reserving procedure is widely used by insurers and is useful for long tailed lines.

Paid Loss Development

The paid loss development method of loss reserving calculates the reserve by projecting ultimate losses from losses paid to date based on historical development patterns. The loss reserve is the total projected losses less the losses paid to date. This method of loss reserving can easily be used to indicate when losses will be paid in the future. A number of variations of paid loss development are described in Berquist and Sherman [4], all of which could be used to calculate when losses will be paid.

The accuracy of this reserving technique depends on consistency in loss settlement patterns. It is not dependent on consistent reporting patterns or case reserve estimates. Changes in the rate of inflation, which can affect loss payments, shifts in company procedures that infuence settlement patterns, or societal shifts such as changes in court backlog can all cause inaccuracies in this reserving method. This procedure is widely used by insurers. The major drawback for this technique is the length of time necessary to determine ultimate loss payments for long tailed lines and the likelihood of changes in factors that influence payment patterns occurring during this time. A possible combination of reserve procedures is to use payment development for a number of years and then incurred development to ultimate subsequent to that period. When losses will be paid cannot be determined directly from the loss development data for the time incurred loss development is applied.

An example of the method used to calculate paid loss development values is illustrated on Exhibit I.

Ultimate paid losses for accident year i, C_{iu}, are projected from losses paid through development year j, C_{ij}, by

$$C_{iu} = C_{ij}\left(\frac{C_{\cdot u}}{C_{\cdot j}}\right),$$

where $C_{\cdot u}/C_{\cdot j}$ is the standard paid loss development factor from development year j to ultimate. The standard paid loss development factor is calculated from historical experience. The most recent ultimate experience, average values for a number of years, or trended values could be used to determine the standard factors. Once the ultimate paid losses are projected, the outstanding reserves are determined by subtracting paid losses to date, C_{ij}, from the estimate of ultimate paid losses, C_{iu}. Partial paid loss development factors are often used to modify indications produced by the use of ultimate paid loss development factors. This technique, similar to the use of partial incurred loss development factors, is useful when changes in the loss payment pattern have occurred.

In order to determine when losses will be paid in the future, loss payout patterns can be calculated from paid loss development factors. Let P_{ij} equal the percent of ultimate paid losses for accident year i paid in development year j. P_{ij} is calculated by

$$P_{ij} = (C_{ij} - C_{i,j-1})/C_{iu}.$$

EXHIBIT I

Accident Year	Paid Losses — Development Year 1	2	3	4	5	6	7	8	Incurred Losses
1976	$C_{76,1}$	$C_{76,2}$	$C_{76,3}$	$C_{76,4}$	$C_{76,5}$	$C_{76,6}$	$C_{76,7}$	$C_{76,8}$	$C_{76,8} + R_{76,8}$
1977	$C_{77,1}$	$C_{77,2}$	$C_{77,3}$	$C_{77,4}$	$C_{77,5}$	$C_{77,6}$	$C_{77,7}$		$C_{77,7} + R_{77,7}$
1978	$C_{78,1}$	$C_{78,2}$	$C_{78,3}$	$C_{78,4}$	$C_{78,5}$	$C_{78,6}$			$C_{78,6} + R_{78,6}$
1979	$C_{79,1}$	$C_{79,2}$	$C_{79,3}$	$C_{79,4}$	$C_{79,5}$				$C_{79,5} + R_{79,5}$
1980	$C_{80,1}$	$C_{80,2}$	$C_{80,3}$	$C_{80,4}$					$C_{80,4} + R_{80,4}$
1981	$C_{81,1}$	$C_{81,2}$	$C_{81,3}$						$C_{81,3} + R_{81,3}$
1982	$C_{82,1}$	$C_{82,2}$							$C_{82,2} + R_{82,2}$
1983	$C_{83,1}$								$C_{83,1} + R_{83,1}$

where C_{ij} = cumulative paid losses for accident year i through the end of development year j, and

R_{ij} = reserves for accident year i as of the end of development year j.

The more mature an accident year, the more accurate the estimate of ultimate losses is likely to be. The paid loss development factors can be used to project when the outstanding reserves will be paid. The outstanding reserve for accident year i at the end of development year j is represented by

$$R_{ij} = \left(\sum_{k=j+1}^{u} P_{ik} \right) C_{iu} \, .$$

This equation states that the outstanding reserve is the sum of the percentage of losses to be paid in each subsequent development year times ultimate losses. The amount to be paid in the next development year, $j+1$, can be determined by

$$C_{i,j+1} - C_{ij} = R_{ij} \left(\frac{P_{i,j+1}}{\sum_{k=j+1}^{u} P_{ik}} \right) .$$

Similarly, subsequent years of loss payments can be determined. Thus, this method of loss reserving can be used to project when losses will be paid for use in discounting loss reserves.

3. PROPOSED REVISION IN RESERVING TECHNIQUES

In order to discount loss reserves, it is necessary to estimate both the total future payments on losses that have already occurred and when the loss payments will be made. Since most insurance accounting occurs on an annual basis, projecting the year of loss payment will usually be sufficient. This paper assumes annual periods for loss payment patterns. More accurate determination of the proper discounting reserve level could be made if a shorter unit of time were used. McClenahan has proposed a reserving methodology based on monthly periods that would allow discounting [18].

If insurers relied solely on paid loss development to establish reserves, shifts in loss settlement patterns could lead to inaccurate reserves. Although this loss reserving technique directly projects when losses will be paid, a combination of paid loss development and other reserve procedures can be used to estimate loss reserves and to project when losses will be paid.

In order to discount loss reserves without reducing the accuracy of loss reserving methods, the loss reserve should be established based on the best reserving methods available without regard to discounting. This approach will generally involve selecting a value from a number of reserve indications determined by applying several methods of loss reserving. The payment pattern for the outstanding reserves can then be determined as follows:

Let R_{ij} = the outstanding reserve for accident year i as of the end of development year j, and

$P_{\cdot j}$ = the standard percentage of losses paid during development year j.

The standard percentage of losses paid, $P_{\cdot j}$, can be determined by a number of methods, subject to the constraint that $\Sigma_{j=1}^{n} P_{\cdot j} = 1$. Averages, least squares regression, trending, or use of the most recent values are all potential methods to determine $P_{\cdot j}$.

The losses for accident year i to be paid within one year of the evaluation date j can be calculated by

$$E_{i,j+1} = R_{ij} \left(P_{\cdot j+1} / \sum_{k=j+1}^{u} P_{\cdot k} \right) ,$$

where $E_{i,j+1}$ are the losses for accident year i projected to be paid in development year $j+1$.

The best estimate of the loss reserve as of evaluation date j for accident year i is multiplied by the proportion of outstanding losses based on the paid loss development method that will be paid during the next, $j+1$, development year. The paid loss development method is used to project the payout pattern, but not necessarily the loss reserve. Similarly, the losses for accident year i to be paid in the second year after the evaluation date j are determined by

$$E_{i,j+2} = R_{ij} \left(P_{\cdot j+2} / \sum_{k=j+1}^{u} P_{\cdot k} \right) .$$

To determine the total losses from all accident years to be paid in the year following evaluation date j, the following calculation should be performed

$$T_1 = \sum_{i=f}^{l} R_{i,l-i+1} \left(P_{\cdot l-i+2} / \sum_{k=l-i+2}^{u} P_{\cdot k} \right) ,$$

where f is the first accident year with losses still outstanding;

l is the latest accident year; and

T_1 is the total losses from prior accident years to be paid in the following development year.

4. INDUSTRY IMPACT

Assuming that property-liability insurers do not implicitly discount loss reserves now, the adoption of discounting would result in a number of changes. Loss reserves would be lower, surplus would increase, and loss reserves would decline [17]. To examine the effect of discounting on the industry, the 1983 Industry Total Annual Statement, provided by A. M. Best Company, was analyzed. The loss development data included on Schedules O and P were used to project industry loss payment patterns for the Schedule O lines, automobile liability, other liability, medical malpractice, workers' compensation, and the multiple peril lines. These payment patterns were then applied to the outstanding reserves to project when the outstanding losses would be paid. The future payments were then discounted.

Determination of the appropriate discount rate is a crucial problem in implementing loss reserve discounting. No consensus yet exists on the correct methodology. The GAO proposal relies on an individual insurer's past investment income rate. The TRA dictates use of the historical interest rate on midmaturity U.S. securities. Cummins and Chang propose use of the current risk-free interest rate, which is generally considered the rate on short term U.S. government issues [5]. Myers and Cohn propose use of the risk adjusted rate of return based on the capital asset pricing model [19]. The risk adjustment factors, however, are not constant over time or consistent across insurers, which leads to severe implementation problems [6].

The discount rates as of 1987 determined by the various approaches described above range from approximately 5 percent for the risk free rate to 10 percent for some insurers' historical values. A rate of approximately 7 percent will be required by the TRA method for 1987 and prior accident years. The two endpoints are used to illustrate the ramifications of loss reserve discounting. The results are extremely sensitive to the selected discount rate, indicating that much additional research should focus on the proper methodology for determining the discount rate. The rate mandated under the Tax Reform Act of 1986 does not have any theoretical support and was chosen primarily for revenue producing considerations [20].

As discussed earlier, a number of methods exist for determining loss payment patterns based on historical data. The 1983 Annual Statement blank provides for information on cumulative paid losses and loss adjustment expense for the most recent eight years as shown on Table I. Losses paid in a particular

TABLE I

ANNUAL STATEMENT INFORMATION
CUMULATIVE PAID LOSSES AND LOSS ADJUSTMENT EXPENSE

Accident Year	Development Year							
	1	2	3	4	5	6	7	8
1976								*Y*
1977	*X*	*X*	*X*	*X*	*X*		*Y*	
1978	*X*	*X*	*X*	*X*	*X*	*Y*		
1979	*X*	*X*	*X*	*X*	*X+Y*			
1980	*X*	*X*	*X*	*X+Y*				
1981	*X*	*X*	*X+Y*					
1982	*X*	*X+Y*						
1983	*X+Y*							

Source:
X Schedule P, Part 3
Y Schedule P, Part 1; Schedule O, Part 3

development year can be determined by subtracting adjacent cumulative values, if both are available. The percent of ultimate losses can be determined by dividing the losses paid in a development year by the total accident year losses, which can be estimated by adding the outstanding reserve for a given accident year to the cumulative paid losses through the latest available development year.

For this project, the loss payment pattern was determined by using the cumulative paid loss value for each accident year as of the latest development period. This method assumes that all years develop similarly and all future paid loss development will be consistent with the latest year's experience. Use of averages or trended values can produce more stable results, but the Annual Statement does not provide enough information to use a better method for all development years and for all lines. For the five years that multiple development is available, paid loss development factors have been fairly consistent for automobile liability, workers' compensation, and multiple peril lines. Other liability and medical malpractice both indicate a shift to greater loss payments in the early development years starting in 1982. Introduction of claims-made policies may have caused this shift in payment pattern or underreserving for these years may be indicated.

Paid loss development must be projected for each development year until all losses are paid. The Annual Statement shows only eight years of development. Based on the outstanding reserves after eight years, Schedule O lines have 2.85 percent of losses unpaid, automobile liability 1.74 percent, other liability 16.19 percent, medical malpractice 32.16 percent, workers' compensation 13.69 percent, and multiple peril lines 1.63 percent. For all except the Schedule O lines, the same percent of losses paid in development year eight are assumed to be paid in subsequent years until all losses are settled. This assumption is conservative since losses are likely to be paid at a decreasing rate. This method results in all losses being settled by development year 18. Unpaid losses after eight years of development on Schedule O lines generally represent reinsurance involving lines that would normally appear on Schedule P. The same 18 year maximum settlement time is applied to Schedule O development. The calculated percent of losses and loss adjustment expenses paid in each development year by line is shown on Table II.

Assuming that the payment patterns by line projected from the 1983 Industry Total Annual Statement apply to accident year 1983, a discounted accident year loss and loss adjustment expense ratio by line can be calculated. Losses paid in the first development year, 1983, are undiscounted. Losses to be paid in the second development year, 1984, are discounted by $(1+d)^{1/2}$, where d is the

TABLE II

PERCENT OF ULTIMATE LOSS AND LOSS ADJUSTMENT EXPENSE PAID IN EACH DEVELOPMENT YEAR BY LINE PROPERTY-LIABILITY INDUSTRY TOTALS

Development Year	Schedule O Lines	Automobile Liability	Other Liability	Medical Malpractice	Workers' Compensation	Multiple Peril
1	58.90%	35.95%	12.10%	5.80%	27.42%	56.18%
2	29.37	29.75	15.56	8.59	24.80	26.87
3	4.53	14.38	11.38	9.00	12.71	5.12
4	2.00	9.00	13.09	12.17	8.75	4.46
5	1.44	4.49	9.91	10.34	4.84	2.26
6	0.59	2.58	8.25	10.58	3.51	1.44
7	0.18	1.19	6.98	8.07	2.88	1.31
8	0.14	0.92	6.54	3.29	1.40	0.73
9	0.29	0.92	6.54	3.29	1.40	0.73
10	0.29	0.82	6.54	3.29	1.40	0.73
11	0.29		3.11	3.29	1.40	0.17
12	0.29			3.29	1.40	
13	0.29			3.29	1.40	
14	0.29			3.29	1.40	
15	0.29			3.29	1.40	
16	0.29			3.29	1.40	
17	0.29			3.29	1.40	
18	0.24			2.55	1.09	
	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

interest rate at which losses are discounted. The use of this factor assumes that losses to be paid in the second development year will be paid halfway through the year or equally throughout the year. Losses to be paid in the third development year, 1985, are discounted by $(1+d)^{3/2}$, and so forth, with losses to be paid in the 18th development year, 2000, discounted by $(1+d)^{33/2}$. The undiscounted loss and loss adjustment expense ratios by line for 1983 and the corresponding discounted loss and loss adjustment expense ratios based on 5 percent and 10 percent interest rates are shown in Table III.

Discounting reduces the total loss and loss adjustment expense ratio from 82.43 percent to 77.67 at a 5 percent discount rate and to 74.18 percent at a 10 percent discount rate. The combined ratio, based on the 28.44 percent industry expense ratio, is 110.87 percent undiscounted, but only 102.62 if loss and loss adjustment expense reserves are discounted at 10 percent. Even with discounting at a rather high rate, the industry did not earn an underwriting profit based on discounted loss reserves for 1983.

Several caveats should be emphasized at this point. Calculation of these discounted loss and loss adjustment expense ratios assumes that the outstanding reserves for accident year 1983 are correct. Many observers feel these reserves are inadequate [22]. Second, it is assumed that current reserves are not discounted. If they are already discounted, this calculation indicates the effect of additional discounting. At the end of 1983, most insurers were not explicitly discounting any reserves except some periodic payments under workers' compensation. Some medical malpractice writers now do discount loss reserves, but the insurer used as an illustration was not explicitly discounting at the end of 1983.

The procedure used to discount all years' loss reserves is similar to the method used to discount accident year 1983 loss and loss adjustment expense reserves. For accident year 1982 outstanding reserves, two years of payments have already occurred by the end of 1983. Thus, the outstanding losses are projected to be settled based on payment development from year three to ultimate. Similarly, outstanding reserves for accident years 1976 through 1981 are projected to be paid based on the remaining payment tail values. The Annual Statement blank combines all accident years prior to 1976; for this project these reserves are treated as accident year 1975 losses.

The effect on the industry of discounting all years' loss and loss adjustment expense reserves but not including any increase in income taxation (based on the "fresh-start" provision) is shown in Table IV. The loss and loss adjustment

TABLE III

ACCIDENT YEAR 1983 LOSS AND LOSS ADJUSTMENT EXPENSE RATIOS PROPERTY-LIABILITY INDUSTRY TOTALS

	Undiscounted	Discounted at 5%	Discounted at 10%
Schedule O	78.03%	75.75%	74.10%
Automobile Liability	88.78	84.29	80.59
Other Liability	93.40	79.71	69.68
Medical Malpractice	117.41	90.70	73.92
Workers' Compensation	84.35	75.10	68.97
Multiple Peril	75.13	72.73	70.79
Total	82.43%	77.67%	74.18%
Expense Ratio	28.44%	28.44%	28.44%
Combined Ratio	110.87%	106.11%	102.62%

TABLE IV

NET WRITTEN PREMIUM TO SURPLUS RATIOS PROPERTY-LIABILITY INDUSTRY TOTALS (000 OMITTED)

	Undiscounted	Discounted at 5%	Discounted at 10%
Loss and Loss Adjustment Expense Reserve	$121,205,523	$105,534,079	$ 94,449,381
Policyholders' Surplus	65,835,979	81,507,423	92,592,121
Net Written Premium	109,263,815	109,263,815	109,263,815
Premium/Surplus	1.66	1.34	1.18

expense reserve declines from $121 billion undiscounted to $106 billion if discounted at 5 percent and $94 billion if discounted at 10 percent. Discounting reserves would increase policyholders' surplus which would affect premium to surplus ratios. The 1983 industry premium to surplus ratio is 1.66 without discounting, 1.34 discounting reserves at 5 percent, and 1.18 discounting reserves at 10 percent. The industry's *reported* financial position would be dramatically different if loss reserves were discounted. In economic terms, no real change would occur. Statutory values would be different, but no change in the economic value of the industry would take place.

5. INDIVIDUAL COMPANY IMPACT

The impact of discounting loss reserves varies markedly by company based on line of business mix, claim settlement patterns, and individual financial position. Three companies were selected to illustrate the differing impact. Company A is a multiline insurer, Company B specializes in personal lines, and Company C writes only medical malpractice insurance. The effect of discounting loss reserves on the loss and loss adjustment expense ratio, the combined ratio, and the net written premium to surplus ratio for each company is shown on Table V.

In calculating the effect of discounting for individual insurers, two differences from the industry method were used. First, cumulative paid loss development for each of the first eight development years is the average of values shown in the 1982 and 1983 Annual Statements. Prior years are not available for the industry aggregate experience. Second, Schedule P experience for that insurer in total, rather than by line, is used to avoid distortions of a single line's payout pattern of an insurer.

For the multiline insurer, Company A, discounting at a 10 percent rate reduces the accident year loss and loss adjustment expense ratio from 95.7 percent to 79.1 percent. The combined ratio is still unprofitable at 111.3 percent, reduced from 127.9 percent. The personal lines carrier, Company B, shows a much smaller reduction in loss and loss adjustment expense ratio, from 85.8 percent to 82.0 percent. The smaller reduction results from faster loss payments in these lines. Even this minor reduction is enough to reduce the combined ratio below 100 from 103.0 percent to 99.2 when loss reserves are discounted at a 10 percent rate. For Company C, the medical malpractice insurer, discounting reduces the loss and loss adjustment expense ratio significantly, from 156.8 percent to 96.1 percent when discounted at a 10 percent rate. The combined ratio decreases from 161.5 percent to an almost profitable 100.8 percent.

TABLE V

IMPACT OF DISCOUNTING ON INDIVIDUAL INSURERS
ACCIDENT YEAR 1983

	Company A			Company B			Company C		
Discount Rate	0%	5%	10%	0%	5%	10%	0%	5%	10%
Loss and Loss Adjustment Expense Ratio	95.7%	86.3%	79.1%	85.8%	83.7%	82.0%	156.8%	121.0%	96.1%
Expense Ratio	32.2	32.2	32.2	17.2	17.2	17.2	4.7	4.7	4.7
Combined Ratio	127.9%	118.5%	111.3%	103.0%	100.9%	99.2%	161.5%	125.7%	100.8%
Net Written Premium to Surplus Ratio	1.60	1.24	1.06	0.96	0.93	0.90	3.71	0.68	0.43

Similar differences in the impact on the premium to surplus ratio occur. On the extremes, Company B shows only a modest shift in this ratio, whereas for Company C the premium to surplus ratio plummets from 3.71 to 0.43 when reserves are discounted at the 10 percent rate. It should be remembered that these values are correct only if current reserves are accurate and undiscounted, and loss payment patterns are consistent.

6. REPERCUSSIONS FROM ADOPTING DISCOUNTING

Discounting property-liability insurance loss reserves would have a number of effects on the industry, some favorable and some unfavorable. Among the favorable results would be:

1) Reestablish the value of the combined ratio as a profitability indicator. Investment earnings would be directly included in this ratio. Hence, levels under 100 would be profitable and levels over 100 would be producing losses, assuming the proper discount rate is used and reserve accuracy is consistent at the beginning and end of the year.
2) Increase the statutory capacity of the industry. Statutory surplus would increase as loss reserve liabilities were reduced. To the extent that statutory surplus values serve as a constraint on an insurer's ability to write more business, this accounting change would indicate that there is more surplus available to write additional business or to shift to other uses. Current concerns over capacity shortages may be alleviated by this accounting change [29]. Many insurance conventions, including allowable premium to surplus ratios, have evolved from historical periods when economic conditions were significantly different from today. Compared with any time prior to the 1970s, interest rates are now higher and loss payout patterns longer. Both of these changes serve to reduce the value of discounted loss reserves compared to undiscounted values. Thus statutory surplus, which is calculated based on undiscounted loss reserves, is reduced well below the level that would have been determined based on a market value accounting for loss reserves. When interest rates were low and loss payments relatively short, discounted loss reserves did not differ much from the undiscounted values. Thus, statutory surplus was a reasonable estimate of the insurer's economic worth. The higher interest rates and slower loss payment patterns have, in effect, made statutory surplus a far more conservative estimate, but allowable premium

to surplus ratios have not been adjusted to offset this development. Adopting loss reserve discounting for statutory accounting would correct this distortion that has gradually crept into insurance accounting.

Among the unfavorable effects of discounting would be the following:

1) Complicate the reserving process by requiring estimates of the total value of losses to be paid in the future, the timing of those payments, and the discount rate. The process, which is currently a time consuming calculation, will become even more involved, delaying the production of operating results.
2) Create a dependence on future interest rates. Discounting loss reserves is reasonable only if the insurer can earn interest on invested assets supporting the reserves in line with projected values. Volatile interest rates create the risk that the insurer may earn a rate less than that projected. To the extent that actual earnings fall below the interest rate used to discount loss reserves, loss reserves would be inadequate. Currently, for almost all cases, changes in interest rates do not affect the accuracy of statutory loss reserve levels. It is conceivable that future insurance insolvencies could result from falling interest rates if discounting is adopted for statutory accounting, as this would cause the loss reserves to be inadequate. Several authors have suggested that property-liability insurers could match assets and liabilities, as is common for life insurers and banks, to eliminate interest rate risk [8, 11]. Liabilities of property-liability insurers vary stochastically, in some cases in line with changes in inflation. Therefore, it is impossible to match those liabilities with bond investments [7].
3) Increase taxation. The purpose of discounting proposals for the federal government is to raise additional tax revenue from the property-liability insurance industry. Additional taxes would simply be an expense passed on to the policyholders. Raising expenses would make the insurance product less attractive to consumers with a viable alternative to insuring.

7. SUMMARY AND CONCLUSIONS

Federal government pressure to raise revenues collected from the insurance industry has led to discounting loss reserves for income tax purposes. Arguments for a uniform accounting system and the desire to constrain rate levels may in turn lead regulators to impose discounting requirements for statutory accounting. This paper indicates some of the complications raised by discounting loss reserves. The effect of discounting loss reserves is significant. Current combined ratios decrease toward 100 percent when discounting at market rates is applied. Premium to surplus ratios also decline drastically, potentially indicating the presence of additional insurance capacity that was not evident under statutory accounting conventions. The reported financial position of the property-liability insurance industry would look very different if discounting for statutory accounting were adopted.

The property-liability insurance industry officially ignores the concept of the time value of money and publicly declares that undiscounted values are the best indicators of industry results. Although many insurers do reflect the time value of money for internal reporting purposes, little uniformity in techniques exists. Lengthening loss payouts and high interest rates, in addition to the TRA provisions, are bound to increase pressure on regulators to extend this concept. Including investment income in rate calculations is one method of recognizing the time value of money. Discounting loss reserves is another. Insurers should initiate a more open discussion of the various techniques for dealing with discounting. This paper presents a method for calculating discounted loss reserves that can be implemented without disrupting the current loss reserving calculations. Hopefully, this research will encourage greater discussion and debate about incorporating the time value of money into insurance calculations.

REFERENCES

[1] Alliance of American Insurers, National Association of Independent Insurers, National Association of Mutual Insurance Companies and Reinsurance Association of America, "Comments on Proposed Taxation Report by GAO," February 24, 1984.

[2] D. R. Anderson, "Effects of Under and Overevaluations in Loss Reserves," *Journal of Risk and Insurance,* Vol. XXXVIII, No. 4, December, 1971, p. 585.

[3] R. J. Balcarek, "Loss Reserving in the Sixties," *PCAS* LIX, 1972, p. 18.

[4] J. R. Berquist and R. E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," *PCAS* LXIV, 1977, p. 123.

[5] J. David Cummins and Lena Change, "An Analysis of the New Jersey Formula for Including Investment Income in Property-Liability Insurance Ratemaking," *Journal of Insurance Regulation,* Vol. I, No. 4, June, 1983, p. 555.

[6] J. David Cummins and Scott Harrington, "Property-Liability Insurance Rate Regulation: Estimation of Underwriting Betas Using Quarterly Profit Data," *Journal of Risk and Insurance,* Vol. LII, No. 1, March, 1985, p. 16.

[7] Stephen P. D'Arcy, Discussion of "Duration" by Ronald E. Ferguson, *PCAS* LXXI, 1984, p. 8.

[8] Ronald E. Ferguson, "Duration," *PCAS* LXX, 1983, p. 265.

[9] M. J. Fisher, "Industry Trade Groups Hit GAO Report on Taxation," *National Underwriter Property & Casualty Insurance Edition,* Vol. 88, No. 9, March 2, 1984, p. 1.

[10] M. J. Fisher, "P&C Tax Study Questions Loss Reserves, Expenses," *National Underwriter Property & Casualty Insurance Edition,* Vol. 88, No. 5, February 3, 1984, p. 1.

[11] H. Russell Fogler, "Bond Portfolio Immunization, Inflation, and the Fisher Equation," *Journal of Risk and Insurance,* Vol. LI, No. 2, June, 1984, p. 244.

[12] S. W. Forbes, "Loss Reserving Performance Within the Regulatory Framework," *Journal of Risk and Insurance,* Vol. XXXVII, No. 4, December, 1970, p. 527.

[13] "From Treasury's Proposal on Discounting of Loss Reserves," *National Underwriter Property & Casualty Insurance Edition,* Vol. 88, No. 50, December 14, 1984, p. 3.

[14] General Accounting Office, *Congress Should Consider Changing Federal Income Taxation of the Property/Casualty Insurance Industry,* U.S. Government Printing Office, January 24, 1984.

[15] J. Greenwald, "Insurers' Sad Song May Play One More Time," *Business Insurance,* Vol. 18, No. 13, March 26, 1984, p. 1.

[16] Insurance Information Institute, *Insurance Facts,* 1985–86, *Property Casualty Fact Book,* p. 6.

[17] Yehuda Kahane and M. Moshe Porat, "Financial Analysis of Underwriting Results—A Corrected Approach to Loss Ratio Analysis, with a Special Reference to Inflation," *Journal of Risk and Insurance,* Vol. LI, No. 4, December, 1984, p. 710.

[18] Charles L. McClenahan, "A Mathematical Model for Loss Reserve Analysis," *PCAS* LXII, 1975, p. 134.

[19] Stewart Myers and Richard Cohn, "Insurance Rate of Return Regulation and the Capital Asset Pricing Model," Massachusetts Automobile Rate Hearings, Boston, August, 1981.

[20] Thomas S. Neubig, "The Effect of Tax Reform on Insurance Companies," Presented at the National Tax Association's Tax Institute of America Conference, November, 1986.

[21] S. Pointek, "Despite Record Underwriting Loss, Industry Netted $2 Billion in 1985," *National Underwriter Property & Casualty Insurance Edition,* Vol. 90, No. 13, March 28, 1986, p. 1.

[22] S. Pointek, "Industry Loss Reserves Are 10 Percent Short: ISO," *National Underwriter Property & Casualty Insurance Edition,* Vol. 88, No. 3, January 20, 1984, p. 1.

[23] President of the United States, *The President's Tax Proposals to the Congress for Fairness, Growth and Simplicity,* U.S. Government Printing Office, 1985.

[24] R. E. Salzmann, *Estimated Liabilities for Losses & Loss Adjustment Expenses,* Englewood Cliffs, NJ, Prentice-Hall, Inc., 1984.

[25] David Skurnick, "A Survey of Loss Reserving Methods," *PCAS* LX, 1973, p. 16.

[26] B. D. Smith, "An Analysis of Auto Liability Loss Reserves and Underwriting Results," *Journal of Risk and Insurance,* Vol. XLVII, No. 2, June, 1980, p. 305.

[27] Barry D. Smith, "Analyzing the Tax Deductibility of Premiums Paid to Captive Insurers," *Journal of Risk and Insurance,* Vol. LIII, No. 1, March, 1986, p. 85.

[28] Michael L. Smith and Robert C. Witt, "An Economic Analysis of Retroactive Liability Insurance," *Journal of Risk and Insurance,* Vol. LII, No. 3, September, 1985, p. 379.

[29] *Wall Street Journal,* January 24, 1985, p. 1, column 5.

[30] Mary Weiss, "A Multivariate Analysis of Loss Reserving Estimates in Property-Liability Insurers," *Journal of Risk and Insurance,* Vol. LII, No. 2, June, 1985, p. 199.

ADJUSTING LOSS DEVELOPMENT PATTERNS FOR GROWTH

CHARLES L. McCLENAHAN

Abstract

This paper examines the impact of changes in exposure growth on loss development patterns. An adjustment methodology for use in cases where growth patterns have changed materially during the observation period is proposed and an example is presented.

1. INTRODUCTION

The vast majority of pricing and reserving analysis performed by casualty actuaries is based, at least in part, upon the construction of loss development triangles and the projection of "loss development factors" (or "link ratios"). Where these factors are based upon historical development patterns there is an underlying, and generally unstated, assumption that each historical exposure period at a given point of development represents a body of claim experience at a consistent average age. In practice, the average age of the exposure period may change over time as a result of variations in inflation, settlement practices, reporting patterns, and exposure growth. The purpose of this short paper is to examine the impact of exposure growth changes upon the development patterns and to propose a method for the adjustment of historical patterns where such impact is material.

While this paper deals with the impact of exposure growth upon the loss development patterns, an earlier paper by LeRoy J. Simon [1] deals with the specific impact of such growth patterns upon exposure-based IBNR factors.

2. GROWTH AND DEVELOPMENT PATTERNS

In order to understand the relationship between exposure growth and loss development, let us look at a highly simplified development pattern. We will assume that losses only occur on the first day of a month and are always reported on the first day of the month immediately following occurrence. Each claim has an associated indemnity benefit of $300 with $100 being paid on the first day of each of the three months immediately following reporting. Case reserves are assumed to be exactly adequate on an undiscounted basis. The following example will summarize the assumed pattern for a single claim occurring on 7/1/86:

Date	Cumulative Reported	Cumulative Paid	Case Reserve
7/1/86	$ 0	$ 0	$ 0
8/1/86	300	0	300
9/1/86	300	100	200
10/1/86	300	200	100
11/1/86	300	300	0

Let us now look at three companies, each having 156 claims occurring during accident year 1986. Company A has increasing exposure, and therefore increasing monthly claims. Company B has stable exposure and Company C has declining exposure. The assumed claim counts are as follows:

Accident Date	Company A	Company B	Company C
1/1/86	2	13	24
2/1/86	4	13	22
3/1/86	6	13	20
4/1/86	8	13	18
5/1/86	10	13	16
6/1/86	12	13	14
7/1/86	14	13	12
8/1/86	16	13	10
9/1/86	18	13	8
10/1/86	20	13	6
11/1/86	22	13	4
12/1/86	24	13	2
Total	156	156	156

For accident year 1986, the three companies have the following situations as of 12/31/86:

	Company A	Company B	Company C
Paid Loss	$27,200	$35,100	$43,000
Case Reserve	12,400	7,800	3,200
Case Incurred	39,600	42,900	46,200
IBNR	7,200	3,900	600
Ultimate Loss	46,800	46,800	46,800
Ultimate/Paid	1.721	1.333	1.088
Ultimate/Case Inc.	1.182	1.091	1.013

In practice, of course, the ultimate values will not be known with certainty at 12/31/86. For the sake of illustration we are assuming perfect knowledge.

Here we have three hypothetical companies writing the same line of business with identical accident year claim counts and very different accident year development patterns. The differences, of course, arise from the varying distributions of the claims in time over the accident year. The average age of claim at 12/31/86 is 4.67 months for Company A, 6.50 months for Company B, and 8.33 months for Company C. Inasmuch as claims growth can be generally expected to reflect exposure growth, the exposure growth pattern can be seen to have a potentially significant impact upon the loss development pattern.

This relationship between exposure growth and development pattern is not, in and of itself, a problem. Should either Company A or Company B continue to experience consistent exposure patterns, the indicated loss development patterns would produce reliable estimates for unpaid and for unreported losses. When exposure growth is inconsistent, however, an adjustment to historical indications may be warranted.

3. HYPOTHETICAL CASE STUDY

Appendix A contains the assumptions and data underlying a somewhat more complex example for a hypothetical company. A totally fictitious reporting pattern has been assumed along with uniform exponential pure premium trend. The exposure growth assumption is a period of uniform positive growth followed by a period of declining growth with the final exposure growth rate being negative. The observed loss development factors are as follows:

Accident Year	Age-to-Age Factors (Age in Years) 1-2	2-3	3-4
1983	1.8699	1.1144	1.0009
1984	1.8697	1.1143	
1985	1.8537		
Weighted Average	1.8635	1.1144	1.0009
To Ultimate:	2.0785	1.1154	1.0009

Using ultimate factors based upon observed weighted averages:

Accident Year	Reported 12/31/86	Ultimate Factor	Projected Ultimate	"Actual" Ultimate
1984	$1,469,650	1.0009	$1,470,973	$1,470,979
1985	1,542,366	1.1154	1,720,355	1,718,089
1986	875,722	2.0785	1,820,188	1,755,193

While it may be argued that the use of the weighted average factors is inappropriate in light of the observed "trend" in the 1-2 factors, it is unlikely that the selected factor for 1-2 would have been as low as the 1.7971 required to generate the "actual" ultimate value had the "trend" been projected to continue. Comparing the projected and "actual" IBNR needs:

Accident Year	Projected IBNR	"Actual" IBNR	Percent Error
1984	$ 1,323	$ 1,329	−0.5%
1985	177,989	175,723	1.3
1986	944,466	879,471	7.4
Total	$1,123,778	$1,056,523	6.4%

Since we have used a consistent monthly reporting pattern along with constant pure premium change, the error in projection, other than rounding error, is due entirely to our inability to accurately reflect the impact of the varying rate of exposure growth on the development pattern.

4. PROPOSED ADJUSTMENT TO DEVELOPMENT FACTORS

Assume that in a growth-free environment, observed losses at accident year age x are $1 - a^x$ of ultimate. (Note that if a is replaced with $e^{-\alpha}$ this becomes $1 - e^{-\alpha x}$, the standard single parameter exponential decay function. While the author does not contend that any single parameter function can be expected to provide a good fit to an entire development pattern, the assumption is sufficiently reasonable for use in calculating adjustment factors within the context of this paper. Appendix B contains information relating to the indicated values of a for various industry data.)

Further assume that exposure growth is at a rate of $100g\%$ per annum. Let us now define L_i^g to be the observed proportion of ultimate losses at accident year age i:

$$L_i^g = \int_{i-1}^{i} (1+g)^{i-x}(1-a^x)dx \qquad i \geq 1$$

$$= \frac{g}{\ln(1+g)} + \frac{a^{i-1}(1+g-a)}{\ln(a) - \ln(1+g)} \qquad i \geq 1;\ g \neq 0 \tag{4.1}$$

If we now define the age-to-age development factor from age $i - 1$ to i as ${}_{i-1}F_i^g$:

$${}_{i-1}F_i^g = \frac{L_i^g}{L_{i-1}^g} \qquad i \geq 2;\ g \neq 0$$

$$= \frac{g\{\ln[(1+g)/a]\} + \ln(1+g)\{1 - [(1+g)/a]\}a^i}{g\{\ln[(1+g)/a]\} + \ln(1+g)\{1 - [(1+g)/a]\}a^{i-1}} \qquad i \geq 2;\ g \neq 0$$

Or, letting $c = g\{\ln[(1+g)/a]\}$ and $b = -\ln(1+g)\{1 - [(1+g)/a]\}$,

$${}_{i-1}F_i^g = \frac{c - ba^i}{c - ba^{i-1}} \tag{4.2}$$

In the special case where $g = 0$:

$$L_i^0 = 1 + \frac{a^{i-1}(1-a)}{\ln(a)} \qquad i \geq 1$$

$${}_{i-1}F_i^0 = \frac{\ln(a) + a^{i-1}(1-a)}{\ln(a) + a^{i-2}(1-a)} \qquad i \geq 2 \tag{4.3}$$

It is proposed that, where growth has been erratic, an attempt be made to estimate the value of a and that historical development patterns be adjusted to a growth-free basis. After selection of factors, growth would be re-introduced into the projected ultimates.

5. EXAMPLE OF PROCESS

Going back to the hypothetical case outlined in Appendix A, the first requirement is an estimate of the parameter a. Looking at the 1983 accident year, we note that at accident year age 1, .479 (589,380/1,229,203) of "ultimate" losses were observed. Using 1/83 to 1/84 earned exposure growth, the observed growth rate was .127 [(1,062/942) − 1]. Setting (4.1) equal to .479, and substituting .127 for g yields an estimate for a of .251. (Of course, we don't know the true ultimate losses in actual practice. The goal here is to attempt, by the best means available, to estimate the parameter a. By using a reasonably well-developed year, or group of years if available, where exposure growth is known or can be reasonably estimated, an approximate value for a can be derived.) Using (4.2) we can now generate the following:

Accident Year	a	g	b	c
1983	.251	.127	.417	.191
1984	.251	.126	.414	.189
1985	.251	.060	.188	.086
1986	.251	−.138	−.361	−.170

Accident Year	Theoretical Development Factors		
	1-2	2-3	3-4
1983	1.908	1.119	1.027
1984	1.915	1.120	1.027
1985	1.911	1.120	1.027
1986	1.855	1.116	1.026

Note that the growth factors (g) for 1984 through 1986 are based upon the December-to-December growth from Appendix A.

Application of (4.3) provides the following "growth-free" factors:

1-2	2-3	3-4
1.886	1.118	1.026

The following factors adjust to a "growth-free" basis:

Accident Year	1-2	2-3	3-4
1983	.988	.998	1.000
1984	.985	.998	
1985	.987		

The following factors adjust back to a "growth-inclusive" basis:

Accident Year	1-2	2-3	3-4
1984			1.000
1985		1.002	1.000
1986	.984	.998	1.000

Next we adjust the observed development factors to a "growth-free" basis and project the remainder of the development to ultimate (brackets indicate projected factors). In this example the projection is assumed to be the beginning-incurred-weighted "growth-free" factor:

Accident	Growth-Free Development Factors		
Year	1-2	2-3	3-4
1983	1.8475	1.1133	1.0009
1984	1.8417	1.1121	[1.0009]
1985	1.8296	[1.1126]	[1.0009]
1986	[1.8385]	[1.1126]	[1.0009]
Weighted Average	1.8385	1.1126	1.0009

Now we readjust the projected "growth-free" factors back to a "growth-inclusive" basis:

Accident Year	1-2	2-3	3-4	To Ultimate
1984			[1.0009]	[1.0009]
1985		[1.1148]	[1.0009]	[1.1158]
1986	[1.8072]	[1.1104]	[1.0009]	[2.0085]

Finally, we calculate the adjusted projected ultimate losses:

Accident Year	Reported 12/31/86	Ultimate Factor	Projected Ultimate
1984	$1,469,650	1.0009	$1,470,973
1985	1,542,366	1.1158	1,720,972
1986	875,722	2.0085	1,758,888
Total	$3,887,738		$4,950,833

Looking at the efficacy of the projections:

Accident Year	Adjusted IBNR	Actual IBNR	Percent Error
1984	$ 1,323	$ 1,329	−0.5%
1985	178,606	175,723	1.6
1986	883,166	879,471	0.4
Total	$1,063,095	$1,056,523	0.6%

Obviously this represents an improvement over the unadjusted error of 6.4%.

6. WHEN TO USE ADJUSTMENT PROCESS

The reader will have noted that where changes in growth are small or where development factors are close to unity there is little impact of the adjustment process. In order to help the user decide when it may be appropriate to utilize the proposed adjustment process, Appendix C contains "growth-free" adjustment factors for various values of a and g. Note how insensitive the factors are to the underlying value of a. In order to use this table, the appropriate factor for the "old" growth rate should be divided by the factor for the "new" growth rate. The resultant factor represents the approximate impact on the unadjusted age-to-age factor. For example:

Auto Liability-Paid Loss Development (a = .600)
Observed 1-2 Factor = 2.100
Growth Underlying Observation = +15% Per Year
Current Exposure Growth Rate = −5% Per Year
Approximate 1-2 Factor = 2.100 (.984/1.006) = 2.054

7. CONCLUSION

This method is intended to produce appropriate adjustments to indicated loss development factors in situations where there have been material changes in exposure growth patterns. While frequency and severity changes can produce variations in development patterns as well, this method does not address those situations. Where frequency and/or severity changes are observed concurrently with exposure growth changes, this method can be used to eliminate the impact of the exposure growth changes in order to facilitate the analysis of frequency and severity.

In most cases, exposure growth will have been sufficiently consistent to obviate the need for the approach outlined in this paper. For new lines of business or where rapid growth or withdrawal occur, however, this approach provides a relatively simple and efficacious basis for improving estimates of ultimate losses.

REFERENCE

[1] LeRoy J. Simon, "Distortion in IBNR Factors," *PCAS* LVII, 1970, p. 64.

APPENDIX A
HYPOTHETICAL REPORTED LOSS DEVELOPMENT

Assume the following loss reporting pattern (ages in months):

Age	Incremental Reports	Cumulative Reports
1	5.0%	5.0%
2	5.0	10.0
3	15.0	25.0
4	10.0	35.0
5	10.0	45.0
6	7.5	52.5
7	7.5	60.0
8	5.0	65.0
9	4.0	69.0
10	3.0	72.0
11	2.5	74.5
12	2.5	77.0
13	2.5	79.5
14	2.5	82.0
15	2.0	84.0
16	2.0	86.0
17	2.0	88.0
18	2.0	90.0
19	1.5	91.5
20	1.5	93.0
21	1.5	94.5
22	1.5	96.0
23	1.0	97.0
24	1.0	98.0
25	1.0	99.0
26	1.0	100.0

Assume further that exposure in force during January, 1983 was 942 units and that exposure grew between January, 1983 and December, 1984 at a monthly rate of 1.0% (12.7% per annum), and then grew at a declining rate such that growth was zero at December, 1985 and −25.0% per annum by December,

1986. Finally, assume that the January, 1983 pure premium per exposure unit was $100.00 and that pure premium grew between January, 1983 and December, 1986 at a monthly rate of 0.5% (6.2% per annum).

As detailed below, the observed reported loss development pattern would be as follows:

Accident Year	Age 12	Age 24	Age 36	Age 48
1983	$589,380	$1,102,063	$1,228,092	$1,229,203
1984	705,364	1,318,846	1,469,650	
1985	832,041	1,542,366		
1986	875,722			

HYPOTHETICAL REPORTED LOSS DEVELOPMENT

Month	Earned Exposure	Pure Premium	Ultimate Incurred	Reported Losses as of Date: 12/83	12/84	12/85	12/86
1/83	942	$100.00	$ 94,200	$ 72,534	$ 92,316	$ 94,200	$ 94,200
2/83	952	100.50	95,676	71,279	92,806	95,676	95,676
3/83	961	101.00	97,061	69,884	93,179	97,061	97,061
4/83	971	101.51	98,566	68,011	93,145	98,566	98,566
5/83	980	102.02	99,980	64,987	92,981	99,980	99,980
6/83	990	102.53	101,505	60,903	92,877	101,505	101,505
7/83	1,000	103.04	103,040	54,096	92,736	103,040	103,040
8/83	1,010	103.56	104,596	47,068	92,044	104,596	104,596
9/83	1,020	104.08	106,162	37,157	91,299	106,162	106,162
10/83	1,031	104.60	107,843	26,961	90,588	107,843	107,843
11/83	1,041	105.12	109,430	10,943	89,733	109,430	109,430
12/83	1,052	105.65	111,144	5,557	88,359	110,033	111,144
1/84	1,062	106.18	112,763		86,828	110,508	112,763
2/84	1,073	106.71	114,500		85,303	111,065	114,500
3/84	1,083	107.24	116,141		83,622	111,495	116,141
4/84	1,094	107.78	117,911		81,359	111,426	117,911
5/84	1,105	108.32	119,694		77,801	111,315	119,694
6/84	1,116	108.86	121,488		72,893	111,162	121,488
7/84	1,127	109.40	123,294		64,729	110,965	123,294
8/84	1,139	109.95	125,233		56,355	110,205	125,233
9/84	1,150	110.50	127,075		44,476	109,285	127,075
10/84	1,162	111.05	129,040		32,260	108,394	129,040
11/84	1,173	111.61	130,919		13,092	107,354	130,919
12/84	1,185	112.17	132,921		6,646	105,672	131,592
1/85	1,196	112.73	134,825			103,815	132,129
2/85	1,206	113.29	136,628			101,788	132,529
3/85	1,216	113.86	138,454			99,687	132,916
4/85	1,224	114.43	140,062			96,643	132,359
5/85	1,232	115.00	141,680			92,092	131,762
6/85	1,238	115.58	143,088			85,853	130,926
7/85	1,244	116.16	144,503			75,864	130,053
8/85	1,248	116.74	145,692			65,561	128,209
9/85	1,252	117.32	146,885			51,410	126,321
10/85	1,254	117.91	147,859			36,965	124,202
11/85	1,256	118.50	148,836			14,884	122,046
12/85	1,256	119.09	149,577			7,479	118,914
1/86	1,255	119.69	150,211				115,662
2/86	1,251	120.29	150,483				112,110
3/86	1,244	120.89	150,387				108,279
4/86	1,236	121.49	150,162				103,612
5/86	1,224	122.10	149,450				97,143
6/86	1,211	122.71	148,602				89,161
7/86	1,195	123.32	147,367				77,368
8/86	1,177	123.94	145,877				65,645
9/86	1,157	124.56	144,116				50,441
10/86	1,134	125.18	141,954				35,489
11/86	1,110	125.81	139,649				13,965
12/86	1,083	126.44	136,935				6,847
AY 83	11,950	$102.86	$1,229,203	$589,380	$1,102,063	$1,228,092	$1,229,203
AY 84	13,469	109.21	1,470,979		705,364	1,318,846	1,469,650
AY 85	14,822	115.91	1,718,089			832,041	1,542,366
AY 86	14,277	122.94	1,755,193				875,722

APPENDIX B

a VALUES IMPLIED BY INDUSTRY PAID LOSS AND LOSS EXPENSE DATA A.M. BEST 200 COMPANY SCHEDULE P DATA AS OF 12/31/85

Accident Year	Auto Liability	Workers' Compensation	General Liability	Multi-Peril
		Paid-to-Incurred Percentage		
1976	99.12%	89.59%	87.96%	99.12%
1977	98.83	88.95	87.15	98.78
1978	98.55	87.47	85.05	98.08
1979	97.88	85.77	80.59	97.72
1980	96.65	83.86	75.40	96.65
1981	93.94	80.31	66.40	94.19
1982	89.18	75.81	55.11	91.14
1983	80.38	68.04	39.68	86.48
1984	65.28	54.66	24.94	79.15
1985	34.27	26.04	8.81	55.80
		Implied a to Generate Observed Cumulative Percentage		
1976	.6226	.7975	.8092	.6233
1977	.6097	.7829	.7961	.6131
1978	.5893	.7713	.7886	.6103
1979	.5768	.7569	.7912	.5826
1980	.5678	.7379	.7916	.5679
1981	.5709	.7225	.8040	.5660
1982	.5735	.7013	.8185	.5455
1983	.5811	.6837	.8449	.5133
1984	.5892	.6734	.8664	.4566
1985	.6573	.7396	.9119	.4420

Method: 1980 Workers' Compensation
1980 is age 6 years at 12/31/85
Set $1 - a^6 = .8386$ thus, $a = .7379$

APPENDIX C

FACTORS TO ADJUST TO "GROWTH-FREE" BASIS

	a = .250			a = .600			a = .800		
g	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4
−.250	1.033	1.004	1.001	1.033	1.006	1.002	1.032	1.006	1.003
−.200	1.025	1.003	1.001	1.025	1.005	1.002	1.025	1.005	1.002
−.150	1.018	1.002	1.000	1.019	1.003	1.001	1.018	1.004	1.001
−.100	1.012	1.001	1.000	1.012	1.002	1.001	1.012	1.002	1.001
−.050	1.006	1.001	1.000	1.006	1.001	1.000	1.006	1.001	1.000
.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.050	.994	.999	1.000	.994	.999	1.000	.994	.999	1.000
.100	.989	.999	1.000	.989	.998	.999	.989	.998	.999
.150	.984	.998	1.000	.984	.997	.999	.984	.997	.999
.200	.979	.998	.999	.979	.996	.999	.979	.996	.998
.250	.974	.997	.999	.974	.995	.998	.974	.995	.998
.300	.970	.996	.999	.969	.994	.998	.970	.994	.998
.350	.965	.996	.999	.965	.994	.998	.965	.993	.997
.400	.961	.995	.999	.961	.993	.997	.961	.993	.997
.450	.957	.995	.999	.956	.992	.997	.957	.992	.997
.500	.953	.994	.999	.952	.991	.997	.953	.991	.996

DISCUSSION BY DANIEL F. GOGOL

Mr. McClenahan presents a method that can be very useful when there has been a recent, significant change in exposure. The average accident date in the most recent accident year may be considerably different from what it was for previous accident years at the same stage of development. Both loss reserving and pricing decisions can benefit greatly from an accurate estimate of the effect of changing exposure on loss development patterns. For a method that is so simple, the one presented by Mr. McClenahan seems to apply very well to a fairly large portion of the exposure and development patterns encountered in practice.

The mathematical derivation of the method applies to a development pattern that has the property that for some $a < 1$, the observed losses at accident age x are $1 - a^x$ of ultimate. Actual development patterns sometimes poorly fit curves of this form. Exposure changes during an accident year are represented in the paper by the function $(1 + g)^x$, and it also may poorly fit actual patterns.

Another problem is the following. Mr. McClenahan defines the observed proportion of ultimate losses at accident year age i, if exposure growth is at a rate of $100g\%$ per annum, by:

$$L_i^g = \int_{i-1}^{i} (1 + g)^{i-x}(1 - a^x)dx \qquad i \geq 1$$

If $1 - a^x$ is the proportion of ultimate losses at accident age x (not accident year age x), then $L_i^g \div \int_{i-1}^{i} (1 + g)^{i-x}dx$ would be the proportion of ultimate losses at accident year age i. The divisor was omitted from Mr. McClenahan's expression. This does not affect the development factors since they are of the form $L_i^g \div L_{i-1}^g$ and the factor $\int_{i-1}^{i} (1 + g)^{i-x}dx$ cancels out. But the curve $1 - a^x$ should represent the proportion of ultimate losses at accident age x, and the curve that is calculated in Appendix B represents the proportion at accident year age x instead. The proportion at accident age x is closer to the proportion at accident year age $x + .5$, since the average accident is approximately one-half year old at the end of accident year age 1. In order to produce a curve, $1 - a^x$, that would be a good fit for the recent accident years, which are generally the most important in loss reserving, it would probably be better to use a much smaller value for x than 6, which is the value Mr. McClenahan uses in Appendix B.

DISCUSSION BY DANIEL F. GOGOL

Mr. McClenahan presents a method that can be very useful when there has been a recent, significant change in exposure. The average accident date in the most recent accident year may be considerably different from what it was for previous accident years at the same stage of development. Both loss reserving and pricing decisions can benefit greatly from an accurate estimate of the effect of changing exposure on loss development patterns. For a method that is so simple, the one presented by Mr. McClenahan seems to apply very well to a fairly large portion of the exposure and development patterns encountered in practice.

The mathematical derivation of the method applies to a development pattern that has the property that for some $a < 1$, the observed losses at accident age x are $1 - a^x$ of ultimate. Actual development patterns sometimes poorly fit curves of this form. Exposure changes during an accident year are represented in the paper by the function $(1 + g)^x$, and it also may poorly fit actual patterns.

Another problem is the following. Mr. McClenahan defines the observed proportion of ultimate losses at accident year age i, if exposure growth is at a rate of $100g\%$ per annum, by:

$$L_i^g = \int_{i-1}^{i} (1 + g)^{i-x}(1 - a^x)dx \qquad i \geq 1$$

If $1 - a^x$ is the proportion of ultimate losses at accident age x (not accident year age x), then $L_i^g \div \int_{i-1}^{i} (1 + g)^{i-x}dx$ would be the proportion of ultimate losses at accident year age i. The divisor was omitted from Mr. McClenahan's expression. This does not affect the development factors since they are of the form $L_i^g \div L_{i-1}^g$ and the factor $\int_{i-1}^{i} (1 + g)^{i-x}dx$ cancels out. But the curve $1 - a^x$ should represent the proportion of ultimate losses at accident age x, and the curve that is calculated in Appendix B represents the proportion at accident year age x instead. The proportion at accident age x is closer to the proportion at accident year age $x + .5$, since the average accident is approximately one-half year old at the end of accident year age 1. In order to produce a curve, $1 - a^x$, that would be a good fit for the recent accident years, which are generally the most important in loss reserving, it would probably be better to use a much smaller value for x than 6, which is the value Mr. McClenahan uses in Appendix B.

The numbers e_i can be chosen to reflect changes in exposure, frequency (e.g., seasonal changes), and severity (e.g., due to claim cost inflation) that are estimated to be representative of the loss development data. The numbers A_k can be based on the loss development data and whatever curve fitting seems appropriate. The numbers x_j that are derived from the e_i's and A_k's can then be used to produce yearly development patterns resulting from a different pattern of change in exposure, frequency, and severity (i.e., a different e_1, e_2, e_3, e_4). By subdividing the year into quarters or months, the problem of variable expected losses between quarters or months is dealt with, but not the variability during quarters or months. However, the overall variability is decreased by subdividing. Two different patterns of exposure during a year can theoretically cause a difference of almost twelve months between the expected average accident dates. But, if the two patterns have the same total amount of exposure during each quarter, or during each month, then the difference between the expected average accident dates must be less than three months, or less than one month, respectively.

In order to use the method presented, it is necessary to choose some n such that $A_k = 1$ for $k > n$. This can be done for some n that is not so large as to be impractical. Some adjustment to the actual estimates of some of the later A_k may be necessary, but it does not have to significantly affect the early development factors derived from the method. These early factors are the ones that are most significantly affected by changes in exposure during an accident year.

Example

Suppose that an insurance company has started writing a new line of business and that the line's estimated ultimate losses for the year's accident quarters are .05, .12, .27, and .56, respectively, of the estimated ultimate losses for the accident year. Suppose reported loss development factors at the end of the year are based on industry-wide data for the line, and that the estimated average industry losses for the four quarters of the accident years on which the data is based are .238, .246, .254, and .262, respectively, of the estimated average accident year losses. Also, assume that the following smoothed progression is selected as a good fit to the industry data: $A_4 = .662$, $A_5 = .832$, $A_6 = .935$, $A_7 = .987$, $A_8 = 1.000$.

Since $A_8 = 1.000$, it is assumed that $x_j = 1.000$ for $j \geq 5$. Therefore, the equations

$$\begin{aligned}
.238\,x_4 + .246\,x_3 + .254\,x_2 + .262\,x_1 &= .662\\
.238\,x_5 + .246\,x_4 + .254\,x_3 + .262\,x_2 &= .832\\
.238\,x_6 + .246\,x_5 + .254\,x_4 + .262\,x_3 &= .935\\
.238\,x_7 + .246\,x_6 + .254\,x_5 + .262\,x_4 &= .987
\end{aligned}$$

can be solved, giving $x_1 = .330$, $x_2 = .600$, $x_3 = .800$, $x_4 = .950$. So the portion of ultimate accident year losses for the company's new line of business that is reported as of the end of the year is estimate by

$$.05(.95) + .12(.80) + .27(.60) + .56(.33) = .490.$$

So the development factor to ultimate for the company's new line is estimated as 2.041 (i.e., 1/.490) as compared to 1.511 from the industry data.

DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXII

AN ANALYSIS OF EXPERIENCE RATING

GLENN G. MEYERS
VOLUME LXXII

DISCUSSION BY HOWARD C. MAHLER

The author wishes to thank Robert Conger, Glenn Meyers, Stephen Philbrick, and Gary Venter for reading an earlier version of this discussion and providing helpful comments. I also wish to thank Lesley Phipps for typing this discussion.

1. INTRODUCTION

This paper is another valuable contribution by Glenn Meyers to the actuarial literature [1]. In it, the author analyzes many aspects of experience rating formulas. Mr. Meyers's paper contains a remarkable amount of material.

It can be divided into four parts, each of which would have been a useful paper on its own. His first two sections give an introduction to experience rating. His third section examines private passenger automobile merit rating data, illustrating a general result in credibility theory with important practical implications. Meyers's fourth, fifth, and sixth sections examine commercial lines experience rating in terms of a useful general concept which Meyers has called efficiency. His seventh section gives a generally applicable method of applying statistical tests to choose the most appropriate form of an experience rating plan.

Although I will concentrate my discussion on certain portions of Mr. Meyers's paper, this in no way reflects upon the importance of the other portions of this paper. Rather, it reflects the large amount of significant material Mr. Meyers has presented, and the inability of this author to analyze it all thoroughly in a single discussion of tractable length.

Section 2 of this discussion concerns Meyers's discussion of the Bailey-Simon results [2]. Meyers proposes an explanation for the observed credibilities based on parameter uncertainty. I also discuss two other similar phenomena—risk heterogeneity and shifting parameters over time. Section 3 of this discussion presents simple examples of the phenomena discussed in Section 2.

Section 4 of this discussion summarizes Meyers's fourth section on the efficiency criterion.

Section 5 of this discussion presents the Bühlmann credibility result for a split experience rating plan.[1] It gives the general formulas to use to assign credibility to the primary and excess losses so as to maximize efficiency.

In Section 6 of this discussion, the formulas derived in Section 5 are used to analyze Meyers's General Liability example. Among the important points discussed is the use of credibilities other than the optimal credibilities from Section 5.

Section 7 of this discussion continues that analysis in more detail. The loss in efficiency due to the use of other than the optimal credibilities is shown to be small for this example. Also, the effects of the choices of different loss limits is explored.

Section 8 of this discussion points out that under certain circumstances it is theoretically valid to have a self-rating point.

Section 9 of this discussion contains the conclusions I draw from my analysis of Meyers's General Liability example.

Section 10 of this discussion analyzes Meyers's Workers' Compensation example. The analysis of the multi-split plan parallels that of the General Liability single split plan. In addition, the multi-split plan is compared to a single split plan, and is found not to perform significantly better for this example.

Section 11 of this discussion summarizes Meyers's seventh section, which gives a generally applicable method of testing experience rating plans.

Section 12 of this discussion gives my conclusions. I believe that some of Mr. Meyers's conclusions do not follow from the work he presents even though they may well turn out to apply in many specific cases encountered in real world applications.

Following the main body of this discussion, there are ten appendices which provide the mathematical support. Appendix A contains two results for covariances which should be more widely known among actuaries. Appendix J contains an interesting example with continuous mixing functions. The other appendices should be of interest to those with a serious interest in credibility theory.

[1] The mathematical derivation is in Appendix F.

2. FORMULAS FOR THE CREDIBILITY

In the third section of his paper, Mr. Meyers discusses two formulas for the credibility. The first, Meyers's formula 3.2, is the usual Bayesian credibility formula

$$Z = \frac{N}{N + K}, \qquad K \geqq 0. \tag{2.1}$$

The second is Meyers's formula 3.3

$$Z = \frac{N}{JN + K}, \qquad J \geqq 1, K \geqq 0. \tag{2.2}$$

which the author derives assuming parameter uncertainty. (See Appendix B for a further discussion of this formula.)

2.1 Parameter Uncertainty and the Bailey-Simon Data

He goes on to see how well the two formulas fit data from the classic paper by Bailey and Simon on the credibility of a single private passenger car [2].[2] He estimates values of J and K from the credibility for one and two years of data. He finds that formula 2.2 does a better job of fitting the credibility observed for three years. In itself, this should not be surprising since formula 2.1 is a special case of formula 2.2, and the extra choice of parameter available should allow a better fit for formula 2.2. Nevertheless, the resulting fit for Classes 1 and 2 is quite impressive.[3] Even for the other classes the fit is a substantial improvement over that for formula 2.1. It should be noted that Class 1, with 3 million car years, has over ten times the data in any of the other classes.

There is an explanation for the poor fit of formula 2.2 to the Bailey-Simon data for Class 4; this same explanation applies, to a lesser extent, to Class 5. The key point is that one cannot have three clean years of experience unless one has been licensed for at least three years. Class 4 includes many drivers who have less than three years of driving experience. Those risks with one year of experience go into Merit Rating Class Y (clean for one year) if they are clean, and Merit Rating Class B (clean for less than one year) if they are not,

[2] This section of Mr. Meyers's paper constitutes a discussion of this remarkable paper by Bailey and Simon written over a quarter of a century ago.

[3] The definitions of the classes are given in the Bailey-Simon paper. Class 1 is Pleasure—No Male Operator under 25. Class 2 is Pleasure—Non-principal Male Operator under 25. Class 3 is Business Use. Class 4 is Unmarried Owner or Principal Operator under 25. Class 5 is Married Owner or Principal Operator under 25.

as explained in Wittick [3]. Both Merit Rating Class A (clean for three years) and Merit Rating Class X (clean for two years) contain no risks with only one year of experience. We expect drivers with only one year of experience to be worse than the average for Class 4. Thus Merit Rating Class A (clean for three years) for driving Class 4, will have a lower frequency than the average for driving Class 4, merely because all of its drivers have at least three years of experience. Thus when we compare it to the remainder of driving Class 4, the resulting Bailey-Simon credibility for three years of data is overstated. The same is true to a lesser extent for the Bailey-Simon credibility for two years of data.[4]

2.2 Practical Implications of Parameter Uncertainty

As noted by the author, formula 2.2 has a maximum credibility of $1/J$. Based on the fit to the Bailey-Simon data, this implies maximum credibilities between 7% and 13%.[5] This implies that no private passenger automobile merit rating scheme can ever attain extremely large credits regardless of how many years of data are used. More generally, when parameter uncertainty is present ($J>1$), then the maximum credibility is less than 100%.

If formula 2.2 holds, the law of diminishing returns sets in very quickly. Using Mr. Meyers's parameters, roughly two-thirds[6] of the theoretical maximum credibility has been achieved using three years of data.

2.3 Shifting Parameters Over Time

An important conceptual distinction should be made between adding up separate units during the same time period (e.g., a large commercial risk) and adding up different years of experience (e.g., a private passenger automobile merit rating plan). While similar formulas might fit the observations in both cases, they do not have exactly the same meaning.

---

[4] This problem, which applies to an analysis of many merit rating plans, could have been avoided if it were possible to remove from the data all risks for which the insured and/or principal operator has been licensed for less than three years.

[5] If a more refined class plan were used, the credibilities would be lower. If the number of accidents rather than just the number of years since the last accident were taken into account, the credibilities would differ. If severity were taken into account, the credibilities would differ. The credibilities will differ depending on whether just accidents or accidents and convictions are taken into account.

[6] The value differs by class. It is 62% for Class 1, and 75% or greater for the other classes.

There are other similar phenomenon which, when important, cause formula 2.1 to no longer apply. One phenomenon is the shifting of parameters over time, which is discussed briefly by both Bailey-Simon and Meyers. Bailey and Simon put this forward as one possible explanation for the observation that extra years of data add relatively little credibility. "It can be fully accounted for only if an individual insured's chance for an accident changes from time to time within a year and from one year to the next, or if the risk distribution of individual insureds has a marked skewness reflecting varying degrees of accident proneness."[7,8]

In Appendix C, a formula is derived for the credibility when the parameters shift over time.[9] The exact solution is complicated for $N \geqq 3$. However, the following formula is approximate for $N = 3$, and exact for $N = 1$ or $N = 2$. (For $N \geqq 3$ this formula produces credibilities slightly too high.)

$$Z = \frac{\rho^{\Delta} \sum_{i=1}^{N} \rho^{i-1}}{\left(\sum_{i=1}^{N} \rho^{i-1} \right) + K} \tag{2.3}$$

where $\rho \leqq 1$ is the covariance between the risk processes one year apart[10] and Δ is the time between the mid-point of the last year of experience used in the rating and the mid-point of the policy year to which the rating will be applied.[11]

---

[7] Bailey and Simon explain in their subsequent paper [4] that what they meant by "marked skewness" leads to formula 2.1.

[8] Bailey and Simon also put forward as a partial explanation the fact that risks enter and leave the various classes. In addition, their use of a premium basis for frequency does not completely eliminate the maldistribution that would result from the use of an imperfect exposure base, as pointed out in the discussion by Hazam [5]. Finally, the Bailey-Simon credibilities are estimated by only looking at the indicated claims-free discounts. In contrast, the optimal credibility is a least squares fit to the Bayesian result for all the observed levels of claims.

[9] A simple assumption is made to quantify the impact of the shift. Other assumptions could be made which lead to other formulas. However, the basic idea remains, if the parameters shift over time, then data from far in the past can be of minimal value in predicting the future.

[10] ρ would capture some aspects that might be considered to be due to parameter uncertainty.

[11] Typically, $\Delta = 2$ for workers' compensation, and $\Delta = 1$ for private passenger automobile merit rating.

We see that the $\sum_{i=1}^{N} \rho^{i-1} \leqq N$ has replaced N in formula 2.1. Also, there is a maximum credibility of

$$\frac{\rho^{\Delta}}{1 + K(1 - \rho)}. \tag{2.4}$$

For ρ considerably less than one, adding more years of data quickly reaches the point of having no practical advantage.

Using the Bailey-Simon credibilities for one and two years of data, we can solve for the parameters in formula 2.3 ($\Delta = 1$). The results are

TABLE 2.1

Class	K	ρ	Three Year Credibility Predicted	Three Year Credibility Observed
1	10.9	.55	8.0%	8.0%
2	7.7	.39	6.5%	6.8%
3	6.9	.40	7.4%	8.0%
4	3.0	.28	8.7%	9.9%
5	8.6	.37	5.5%	5.9%

Formula 2.3 produces a good fit for Class 1, a fair fit to Classes 2, 3, and 5, but it is unacceptable for Class 4. Formula 2.2 does considerably better for Classes 2 and 3. As has already been explained, we do not expect a good fit for Classes 4 and 5. The maximum credibilities indicated range from 8% to 13%, roughly the same range as indicated by formula 2.2.

One can use all three years of data in an attempt to estimate the parameters in either formula 2.2 or formula 2.3. Using a least squares fit, the results for formula 2.2 are given in Table 2.2 and for formula 2.3 in Table 2.3. We note that overall the fits of the formulas to the Bailey-Simon data, which is reproduced for convenience in Table 2.4, are as good as can be expected given the nature of the data. While the assumptions behind formula 2.3 seem more applicable to the situation here, formula 2.2 does at least as good a job of fitting the observed data. We note that the indicated maximum credibilities ($N = \infty$) are consistently lower for formula 2.3 than for formula 2.2.

TABLE 2.2

FORMULA 2.2 FIT TO THE DATA IN TABLE 2.4

Class	J	K	$N = 1$	$N = 2$	$N = 3$	$N = \infty$
1	7.88	13.78	4.62%	6.77%	8.02%	12.7%
2	10.96	11.30	4.49	6.02	6.79	9.1
3	9.00	10.85	5.04	6.93	7.93	11.1
4	8.30	6.06	6.96	8.83	9.69	12.0
5	12.37	14.33	3.75	5.12	5.83	8.1

TABLE 2.3

FORMULA 2.3 ($\Delta = 1$) FIT TO THE DATA IN TABLE 2.4

Class	J	ρ	$N = 1$	$N = 2$	$N = 3$	$N = \infty$
1	11.14	.557	4.59%	6.83%	8.00%	9.4%
2	8.61	.428	4.45	6.09	6.75	7.2
3	8.47	.473	4.99	7.01	7.89	8.7
4	4.51	.381	6.91	8.93	9.63	10.0
5	11.09	.448	3.71	5.17	5.80	6.3

TABLE 2.4

EMPIRICAL CREDIBILITIES FROM BAILEY-SIMON PAPER

Class	$N = 1$	$N = 2$	$N = 3$
1	4.6%	6.8%	8.0%
2	4.5	6.0	6.8
3	5.1	6.8	8.0
4	7.1	8.5	9.9
5	3.8	5.0	5.9

The extent to which parameters actually shift over time for any given line of insurance is an important empirical question worthy of further investigation. One would examine the correlations between years of data separated from each other by different time spans. One would also examine the empirical credibility of one year of data being used to predict some later year of data, for different separations between the two years.

The results of such an investigation should be quite useful in the design of experience rating plans. It would help to decide how many years of data should go into the plan. Also, it would help decide whether it is worthwhile, i.e. produces a significant increase in efficiency (as defined by Meyers), to give more weight to the more recent years of data.[12] It would also help in deciding what those relative weights should be.

2.4 Risk Heterogeneity

Another phenomenon is risk heterogeneity. In other words, a large risk may be made up of smaller risks. If we treat the smaller risks within a single large risk as independent observations from the same distribution we get the usual Bayesian formula 2.1. However, if smaller risks were grouped together in a totally random fashion to give larger risks, then there would be no increase in credibility between a small risk and a large risk. The actual situation is generally somewhere between those two extremes.

As shown in Appendix D, this would lead to a formula for credibility of the form

$$Z = \frac{N + I}{N + K}, \qquad 0 \leqq I \leqq K. \tag{2.5}$$

It should be noted that the value of K in formula 2.5 differs from that in formula 2.1. Formula 2.5 does not fit the Bailey-Simon data.

Formula 2.5 was derived for large risks. It would not apply for small risks, i.e., those too small to have separate and distinct subunits.[13] Specifically, no conclusion should be drawn from the fact that formula 2.5 has a minimum credibility of I/K.

[12] For example, the Massachusetts Private Passenger Automobile Safe Driver Insurance Plan currently gives less weight to older incidents via a so-called aging process.

[13] It should be noted that generally experience rating plans have an eligibility requirement which excludes very small risks.

If both parameter uncertainty and risk heterogeneity are important, as shown in Appendix D, the formula for credibility takes the form

$$Z = \frac{N + I}{JN + K}, \qquad 0 \leqq I \leqq K, J \geqq 1. \tag{2.6}$$

Formula 2.6 can be fit to the Bailey-Simon data. However, using three years of data to fit three parameters I, J, and K leaves no way to test the predictions.

Let M be a measure of the (average) size of the risk in each year. Let N be the number of years of data used for experience rating. Then if all three phenomena are taking place, we get the following formula in Appendix E:[14]

$$Z = \frac{\rho^{\Delta}\left(\sum_{i=1}^{N} \rho^{i-1}\right)(M + I)}{\left(\sum_{i=2}^{N} \rho^{i-1}\right)(M + I) + JM + K}. \tag{2.7}$$

For $\rho = 1$, formula 2.7 reduces to

$$Z = \frac{N(M + I)}{(N-1)(M + I) + JM + K}. \tag{2.8}$$

For $N = 1$, formula 2.8 reduces to formula 2.6, as it should.

2.5 Conclusions

I suspect that each of the three phenomena discussed is taking place to some extent. It would be worthwhile to obtain a more current set of private passenger automobile data that followed a risk for more than three years. Then one could determine the relative importance of the three phenomena. It would also be worthwhile to investigate the effects of these phenomena on other lines of insurance. For example, parameter uncertainty and risk heterogeneity would be expected to be particularly important for large commercial risks.

More generally, it would be worthwhile to determine empirically the credibility associated with each size of risk.[15] For such an investigation, identifying

[14] As stated previously, ρ captures certain aspects that might be labeled parameter uncertainty. Here J captures only those aspects of parameter uncertainty that relate to adding up subunits at the same point in time.

[15] The National Council on Compensation Insurance is currently doing so for workers' compensation.

the underlying causes would be helpful but not necessary. However, the above reasoning leads to useful candidates to check against the observed behavior with size of risk.[16]

3. AN EXAMPLE ILLUSTRATING PARAMETER UNCERTAINTY, SHIFTING PARAMETERS OVER TIME, AND RISK HETEROGENEITY

This example will try to illustrate what is meant by the three related but somewhat different concepts of parameter uncertainty, shifting parameters over time, and risk heterogeneity. The mathematics are developed and discussed in Appendices B, C, D, and E.

Assume we have the legendary little old lady from Pasadena who only uses her car to drive back and forth to church on Sundays. Let us ignore any seasonal variations in driving conditions. Further assume she always travels at the same time of day and always uses the same route.

One year actually consists of 52 observations of her claims process. If we treat these as 52 independent observations from the same distribution then we would get formula 2.1 for the credibility.

3.1 Parameter Uncertainty

However, we note that there are factors outside of her risk process that will vary her loss potential randomly, i.e., change the parameters of her risk process on any given day. To take one example, whether or not it is raining would affect her risk process.

Assume that there is a higher chance of an accident when it rains. Further, assume whether or not it rains during her trip is a random variable. Then we have a risk process such as described by Mr. Meyers's algorithm A.1 (ignoring, for simplicity, claim severity). This is an example of parameter uncertainty.

There is a kernel of uncertainty in the number of accidents she has in a year due to the variability caused by the different possible states of the universe. Extra observations will not reduce the effect of this kernel of uncertainty. This is why we get the lower credibilities indicated by formula 2.2. This is also why the maximum credibility is less than one.

3.2 Shifting Parameters Over Time

Let us change the example to illustrate risk parameters shifting over time.

[16] The table at the end of Appendix G gives a good list of such candidates.

(They shift in a definite direction, but we cannot predict beforehand in which direction.) Assume that the little old lady changes churches, and that her trip is significantly different (e.g., longer or shorter). Then her accident potential is different. Experience based on old data when she was driving to her old church is less useful for predicting her expected future claims experience driving to her new church. Thus the credibility assigned to it would be lower. The credibility would be given by formula 2.3. Once again, the maximum credibility is less than one. Many observations in a single year will not take this shift into account, and observations over a longer period of time include, of necessity, "stale" data.

3.3 Risk Heterogeneity

Finally, let us change the example to illustrate risk heterogeneity. Let us assume we have a classification of risks which is made up solely of cars which are only driven by little old ladies, who only use them to drive to church on Sundays.[17] Further, let us assume that in each case the car is jointly owned by two little old ladies. Finally, assume that they take turns driving, each one driving every other Sunday. In this case, the average process variance per car for the class remains the same as the case where each car was driven by one little old lady.[18]

However, since the distribution of the loss potential of cars has become more concentrated toward the mean, the variance between the cars making up the class would be less than in the case where each car was driven by one little old lady.[19] Thus, the claims data for a car is less credible than the similar situation where we had only one driver of each car. It would be even less credible if they alternated churches each Sunday, as well as drivers.

The formula for credibility when there were heterogeneous risks was given by formula 2.5. Another simple but illustrative example is given in Appendix D of this discussion.

---

[17] While this is clearly unrealistic, many class plans for private passenger automobiles do have a senior citizens class.

[18] This would not be the case if they flipped a coin in order to decide who did the driving each Sunday. In that case, the process variance would be greater. This can be usefully thought of as a case of parameter uncertainty.

[19] This would not be the case if for each pair of little old ladies who jointly own a car, the two drivers in each pair have the exact same loss potential as each other.

3.4 Conclusions

It should be noted that part of the difficulty in assigning credibilities to one car year of exposure is that a car year can mean considerably different things depending on how the car is used, how far it is driven, and how many drivers it has. Part of the purpose of experience rating is to make up for any such inadequacies in the exposure base or risk classification system.

The interested reader would probably find it useful to construct a similar example of his own for a large commercial risk. One could take a workers' compensation insured consisting of ten separate locations of equal size, and give examples of each of the three phenomena.

4. THE EFFICIENCY CRITERION

In Section 4 of his paper, Mr. Meyers defines the efficiency of an experience rating plan as the reduction in the expected squared error. (See formula 5.2 below.) The higher the efficiency, the more accurate the experience rating plan.

The author defines the efficiency so that it is never more than 100%.[20] The efficiency can only reach 100% if all the risks in the class have the same mean. Since classes are usually not perfectly homogeneous, the efficiency obtainable by any estimator is usually less than 100%. The author shows that the maximum efficiency using credibility is achieved when the credibility is equal to the Bühlmann (i.e., Bayesian credibility) result. For this case, the efficiency equals the credibility.

The author also shows that the efficiency as a function of the credibility is a parabola. Thus, even if the credibility used is not quite the Bühlmann result, there is still a substantial improvement in accuracy due to the use of credibility.[21] In the next section of his paper, the author shows how this general principle applies to the use of a self-rating point.

5. MAXIMIZING EFFICIENCY, PRIMARY AND EXCESS LOSSES

It is possible to generalize the Bühlmann result to the cases Meyers examines. Assume we have an experience rating plan, and our estimate of the mean

[20] The efficiency can be negative for a particularly poor choice of estimator.

[21] This pleasant and useful property of credibility estimates is explored in more detail in Mahler [6].

F is given by

$$F = (1 - Z_p)\,E_p + Z_pA_p + (1 - Z_e)\,E_e + Z_eA_e\,. \tag{5.1}$$

The subscripts p and e will stand for primary and excess; however, for now they can be treated as any two well defined portions of the total losses. E_p and E_e are the expected losses of each type. A_p and A_e are the actual losses of each type. Z_p and Z_e will be thought of as the credibilities assigned to each portion of the losses; however, for now they can be treated as just numbers to be determined.

In accordance with Meyers, define the efficiency of F by the expression

$$1 - \frac{E[(F - \mu)^2]}{E[(M - \mu)^2]}\,, \tag{5.2}$$

Where M is the grand mean, and μ is the mean for individual risks. In this case, $M = E_p + E_e$.

In order to maximize the efficiency, one must minimize $E[(F - \mu)^2]$. In Appendix F, Z_p and Z_e are determined so as to maximize the efficiency. Let:

a = total variance of the primary losses
b = total variance of the excess losses
c = variance of the hypothetical means of the primary losses
d = variance of the hypothetical means of the excess losses
r = total covariance of the primary and excess losses
s = covariance of hypothetical means of the primary and excess losses.

Then the optimum Z_p and Z_e are

$$Z_p = \frac{(c + s)b - (d + s)r}{ab - r^2}\,,\ \text{and} \tag{5.3}$$

$$Z_e = \frac{(d + s)a - (c + s)r}{ab - r^2}\,. \tag{5.4}$$

It is interesting to note that if we set the primary losses equal to the total losses and thus the excess losses equal to zero, then the solution to the equations becomes

$$Z_p = \frac{c}{a}\,,$$

which is the usual expression for credibility, as in Meyers's equation 3.1.

However, unlike the usual case for credibilities, formulas 5.3 and 5.4 do not have the property of restricting Z_p or Z_e to the closed interval between zero and one. Thus, although it may merely be a matter of semantics, some caution is required before labelling Z_p and Z_e as credibilities. For simplicity of exposition, I will refer to them as credibilities, but perhaps a more precise term to apply would be weights.

The maximum efficiency that results from the optimal values of Z_p and Z_e given by formulas 5.3 and 5.4 is

$$\text{Maximum Efficiency} = \frac{Z_p(c + s) + Z_e(d + s)}{c + d + 2s} . \tag{5.5}$$

Thus, the maximum efficiency is a weighted average of the two credibilities that produce this maximum.[22]

In Appendix G, the dependence of the credibilities and efficiency on the size of risk is explored. One does not get the familiar formula 2.1 that we had for the no-split situation.[23]

6. SINGLE SPLIT PLANS

In Section 5 of his paper, Mr. Meyers illustrates the advantage of having a loss limit as per the General Liability single split plan. He does this by means of an example in which he assumes four types of risks.[24] It is useful to think of these risks as excellent, good, bad, and terrible. While this choice simplifies the computations, it still captures the essence of experience rating, which is to distinguish between risks to the extent that they are otherwise not distinguished by the class plan.

The claim count distribution is chosen as a binomial with N trials. N is used as a measure of the size of the insured. Once again this simplifies the computations, but captures the essential features. There are high and low frequency risks and the process variance increases linearly with N.

[22] The denominator is the variance of the hypothetical means of the total losses (primary plus excess).

[23] Of course as discussed above, one can make alternative assumptions, and get alternative formulas for the no-split situation, as for example formulas 2.2, 2.3, and 2.5.

[24] In Appendix J, an example is given of a continuous distribution of types of risks.

The severity distribution is chosen as a discrete version of the Shifted Pareto.[25] The use of the discrete version again simplifies the computations while maintaining the essential features. There are high and low severity risks. Most of the claims are small; however, the large claims contribute a large part of the mean and most of the variance.[26]

Mr. Meyers employs the Panjer algorithm to derive the aggregate loss distribution from the assumed frequency and severity distributions.[27] The reader should note that, for these simple examples, it is relatively simple to calculate the aggregate distributions directly via convolutions. Also, all of the calculations necessary to explore the behavior of the credibility results can be done from the separate frequency and severity distribution without first obtaining the aggregate loss distribution.[28]

6.1 Frequency versus Severity

Mr. Meyers looks at three different examples. In his first example, which is displayed in his Table 5.2, only the frequency distributions vary between the risks. In his second example, which is displayed in his Table 5.3, only the severity distributions vary between risks. In his third example, which is displayed in his Table 5.4, both the frequency and severity distributions vary between risks. In actual applications, which example is a better approximation to reality will depend on the relative importance of the variance between risks of the frequency and the variance between risks of the severity. For each example, Mr. Meyers displays the results of using Bayes Theorem as well as credibility. I will only discuss the results of using credibility.

Mr. Meyers points out the conflicting roles of the frequency and severity variances between risks in the choice of a loss limit. If only the frequency distributions vary, then the loss limit should be low. The assumption in Meyers's

[25] While the Pareto has been extensively used as a size of loss distribution for a group of risks, it is unclear whether or not the Pareto is an appropriate size of loss distribution for individual risks.

[26] In fact, for Mr. Meyers's choice of parameters, $q = 1.25$, the unlimited Pareto has an infinite variance.

[27] The Panjer algorithm is explained in Venter [7]. It is simpler than the Heckman-Meyers algorithm and designed to handle the case where one has a discrete severity distribution. The Heckman-Meyers algorithm is explained in Heckman and Meyers [8].

[28] In fact, they need only be done completely for $N = 1$, with the results for other values of N following from the results in Appendix G. Mr. Meyers's results for the use of Bayes Theorem do require that the aggregate loss distributions be calculated.

first example is that the size of a claim is completely random and does nothing to distinguish good and bad risks. On the other hand, if only the severity distributions vary, we want a higher loss limit. In Meyers's second example, we want to capture as much of the valuable information contained in the size of claim as is useful. When both the severity and frequency distributions vary, the optimal loss limit is somewhere between the results for the first two cases.

In his third example, Mr. Meyers takes the frequency and severity as highly correlated.[29] Those risks with a high mean frequency also have a high mean severity. Thus, although frequency and severity are assumed to be independent for a given risk, they cannot be treated as independent when looking at all risks combined.

This high correlation chosen by Mr. Meyers, as well as the particular choice of parameters, affects the particular results obtained. Thus, when attempting to apply Mr. Meyers's method of analysis to a particular real world situation, it is important to carefully choose those assumptions which most closely match that situation. With this caveat, the method of analysis should be widely applicable. I will analyze Meyers's third example extensively below in Section 7.

6.2 Basic Limits versus Total Limits

Sometimes the question that is asked is as important as the answer. Mr. Meyers poses the question in Section 5 of his paper as trying to maximize the efficiency, where only the error in predicting basic limits losses is considered in the efficiency. A useful extension might be to consider the error in predicting total limits losses,[30] which would produce different results. If you assume that each of these risks would receive the same increased limits factor, then one could explore the behavior of the efficiency for various total limits using the methods discussed below.

6.3 Primary and Excess Credibilities

Formulas 5.3 and 5.4 give the primary and excess credibilities which will maximize the efficiency. (However, these "credibilities" do not necessarily lie between zero and one.) It is possible to use other values for the credibility, but, of course, the efficiencies will be lower.

[29] In Meyers's Section 6, when examining the Workers' Compensation Experience Rating Plan, frequency and severity are taken as independent of each other.

[30] In the next section, Mr. Meyers deals with unlimited losses while exploring the features of the Workers' Compensation Experience Rating Plan.

As shown in Appendix F, if the excess variance is relatively large, as it is for this example, then it is a good approximation to the optimal credibilities to take

$$Z_p = \frac{c + s}{a}\text{, and} \tag{6.1}$$

$$Z_e = 0\,.$$

In fact, the General Liability Plan sets $Z_e = 0$. Subject to that constraint, formula 6.1 gives the maximum efficiency. As shown in Appendix F, this is equivalent in the General Liability Plan to taking

$$Z = \left(\frac{c + s}{a}\right)\left(\frac{E_p}{E_p + E_e}\right). \tag{6.2}$$

In actual application, formula 6.2 can lead to values greater than one. Thus, it might be more practical to employ

$$Z = \text{MIN}\left[1, \left(\frac{c + s}{a}\right)\left(\frac{E_p}{E_p + E_e}\right)\right]. \tag{6.3}$$

In his tests, Mr. Meyers uses

$$Z = \frac{c}{a}\,. \tag{6.4}$$

The ratio of the credibility given by formula 6.2 to that given by formula 6.4 is

$$\left(1 + \frac{s}{c}\right) \div \left(1 + \frac{E_p}{E_e}\right).$$

For the assumptions used here, this is independent of N. For the Meyers example, this expression is greater than one, so that the credibilities given by formula 6.2 are larger than those given by formula 6.4.

In Appendix F, a formula is given for the loss in efficiency due to using formula 6.4 rather than formula 6.2.[31] As will be seen below for the cases explored by Mr. Meyers, the loss in efficiency is relatively small. Nevertheless, for certain applications, it may be significant.

[31] The relative loss in efficiency turns out to be independent of N.

7. MEYERS'S GENERAL LIABILITY EXAMPLE IN MORE DETAIL

Meyers displayed in more detail in Exhibit 5.1 the case of varying frequencies and severities when $N = 4$ with a loss limit of 4.[32] For this case we have $a = 4.744$, $b = 29.740$, $c = 1.026$, $d = .753$, $r = 4.752$, and $s = .874$.

Using formulas 5.3 and 5.4 gives $Z_p = 41.2\%$, and $Z_e = -1.1\%$, with a resulting efficiency of 21.7%.

Using formula 6.2 results in $Z = 26.3\%$ and an efficiency of 21.6%.[33] In this case $Z \leqq 1$, so formula 6.3 gives the same result as formula 6.2.

Formula 6.4 (used by Meyers) gives $Z = 21.6\%$ and an efficiency of 20.9%, which matches Meyers's result.[34]

The behavior observed by Mr. Meyers for credibilities in Table 5.4 can be explained in terms of formula 6.4 being an approximation to formula 6.3, which is in turn an approximation to formula 6.2, which in turn is an approximation to formulas 5.3 and 5.4, the true optimal credibility result.[35]

7.1 Results of the Various Formulas for Credibility

In Tables 7.1 to 7.4, I have calculated the equivalent of Meyers's Table 5.4 (frequency and severity both vary) for these various different credibility formulas. I have extended the tables to cover more values of N and loss limits.

---

[32] There are four prior distributions given equal weight. They have binomial parameters p of .2, .3, .4, and .5 respectively. They have first parameters of the Pareto, which Meyers calls b, of .25, .50, .75, and 1.00 respectively. They all have the second parameter of the Pareto, which Meyers calls q, equal to 1.25.

[33] Note that this is approximately equal to the Bühlmann credibility from formula 6.4 given below. Why this is the case is explained in Appendix H.

[34] The efficiency is not equal to the credibility as one might expect from Meyers's Appendix B, since we are measuring the error in predicting the basic limits losses rather than just the primary portion of the basic limits losses.

[35] Which is in turn a linear approximation to the optimal Bayesian result.

TABLE 7.1

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
PRIMARY AND EXCESS CREDIBILITIES AS PER FORMULAS 5.3 AND 5.4

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	19.9%	33.1%	49.6%	66.2%	79.4%	88.2%	93.4%	100.0%
2	22.7	37.0	53.9	69.9	82.1	90.0	94.5	100.0
2.5	22.8	37.1	54.1	70.1	82.2	90.1	94.6	100.0
3	22.5	36.7	53.6	69.7	82.0	90.0	94.6	100.0
4	21.7	35.6	52.5	68.8	81.4	89.6	94.4	100.0
6	20.0	33.3	49.9	66.5	79.9	88.7	93.9	100.0
8	18.5	31.3	47.6	64.5	78.4	87.8	93.5	100.0
12	16.4	28.2	44.0	61.1	75.9	86.2	92.6	100.0
16	15.0	26.1	41.4	58.5	73.8	84.9	91.8	100.0

TABLE 7.2

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.2

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	18.6%	31.4%	47.6%	64.3%	77.9%	87.1%	92.6%	98.8%
2	22.6	36.9	53.8	69.8	82.0	89.9	94.4	99.4
2.5	22.8	37.1	54.0	70.0	82.2	90.0	94.5	99.5
3	22.5	36.7	53.6	69.7	82.0	90.0	94.5	99.6
4	21.6	35.5	52.3	68.7	81.3	89.6	94.4	99.7
6	19.6	32.8	49.4	66.1	79.5	88.5	93.9	99.9
8	18.0	30.5	46.7	63.7	77.8	87.5	93.3	99.9
12	15.5	26.9	42.4	59.5	74.6	85.4	92.1	100.0
16	13.8	24.3	39.1	56.2	71.9	83.7	91.1	100.0

TABLE 7.3

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.3

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	18.6%	31.4%	47.6%	63.8%	73.2%	77.9%	80.3%	82.7%
2	22.6	36.9	53.8	69.8	80.8	86.3	89.1	91.9
2.5	22.8	37.1	54.0	70.0	81.6	87.7	90.7	93.7
3	22.5	36.7	53.6	69.7	81.8	88.9	91.7	95.0
4	21.6	35.5	52.3	68.7	81.3	89.0	92.8	96.6
6	19.6	32.8	49.4	66.1	79.5	88.5	93.3	98.1
8	18.0	30.5	46.7	63.7	77.8	87.5	93.1	98.8
12	15.5	26.9	42.4	59.5	74.6	85.4	92.1	99.4
16	13.8	24.3	39.1	56.2	71.9	83.7	91.1	99.7

TABLE 7.4

EFFICIENCIES
(COUNT AND SEVERITY DISTRIBUTIONS VARY AS PER MEYERS'S TABLE 5.4)
CREDIBILITY AS PER FORMULA 6.4, I.E. SHOULD MATCH MEYERS'S TABLE 5.4

Loss Limit ($1000)	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = \infty$
1	15.6%	26.2%	39.8%	53.8%	65.2%	72.9%	77.5%	82.7%
2	20.9	34.1	49.7	64.5	75.8	83.1	87.2	91.9
2.5	21.5	34.9	50.9	65.9	77.4	84.7	89.0	93.7
3	21.4	35.0	51.1	66.5	78.2	85.8	90.2	95.0
4	20.9	34.4	50.7	66.5	78.8	86.8	91.4	96.6
6	19.3	32.2	48.5	64.9	78.1	87.0	92.2	98.1
8	17.8	30.1	46.2	62.9	76.9	86.5	92.2	98.8
12	15.4	26.7	42.1	59.2	74.2	85.0	91.6	99.4
16	13.8	24.2	38.9	56.0	71.7	83.4	90.8	99.7

7.2 Efficiency as a Function of Loss Limit

Table 7.4 here should match Meyers's Table 5.4.[36] We see the same behavior noted by Mr. Meyers. Taking larger loss limits gives higher efficiency only up to a point; then it decreases. The optimal loss limit increases with size of risk, as was noted by Mr. Meyers. Table 7.1, using primary and excess credibilities, shares the former feature, but not the latter feature.

The fact that, for fixed N, the efficiencies in Table 7.1 have a maximum somewhere in between $L = 0$ and $L = 50$ (the limit for basic limit losses in Meyers's example) is not surprising. If $L = 50$, then all of the losses are primary. If $L = 0$, then all of the losses are excess. In each case, the solution reduces to that of the no-split plan. Thus, the two endpoints have the same efficiency. For $0 < L < 50$, the special case where we restrain $Z_p = Z_e$ reduces to that of the no-split plan. Thus we know that the split plan allowing Z_p and Z_e to vary independently so as to get the maximum efficiency does at least as well as the no-split plan, which is just a special case.

Thus $0 < L < 50$ does at least as well as $L = 0$ or $L = 50$. In fact, it does better. The efficiency peaks somewhere between the endpoints and decreases as L approaches the two extremes. This behavior carries over to Tables 7.2, 7.3, and 7.4 which can be thought of as successive approximations getting further and further from the optimal results in Table 7.1

In Table 7.1, the optimal loss limit is about 2.5, independent of N.[37] This behavior carries over to Table 7.2. In Table 7.3, the optimal loss limit increases slowly with N, after $N = 32$. The reason is that we have restricted $Z \leqq 1$;[38] if formula 6.2 would indicate $Z > 1$, we set $Z = 1$ instead. Thus we are giving

[36] In fact, it does not for a loss limit of \$1000. Apparently, Mr. Meyers mistakenly set these efficiencies equal to those from his Table 5.2 where only the claim count distributions vary. The efficiencies here should be lower, even though the two cases are very similar. It is true that since the severity distributions are discrete in units of \$1000, choosing a loss limit of \$1000 means that all limited claims are of size \$1000. In other words, the experience rating plan ignores the size of claim in both cases for a loss limit of \$1000. In each case, the experience rating plan is explaining the same amount of variation, based solely on the observed difference in claim counts. However, in the case here, the total observed variation in losses is greater than when only the claim count distributions vary, since here the severity distributions also vary. Thus, a smaller proportion of the total observed variation is explained here. Therefore, the efficiency is lower here.

[37] The reason this is so for Table 7.2 is explained in Appendix H.

[38] This becomes applicable in the upper righthand corner of the table. For example, for $N = 64$ and a loss limit of 2.5, the credibility indicated by formula 6.2 is 109.1%.

less than the optimum weight to the primary losses. We can afford to raise the loss limit somewhat (staying within the range where formula 6.2 would indicate $Z \geqq 1$, thus formula 6.3 indicates $Z = 1$) so as to make the mean primary loss larger, in order to make up for the lessened weight that is being applied to the primary losses.

Finally, Table 7.4 has the optimum loss limit increase faster with N than does Table 7.3. This is so since formula 6.4 is close to formula 6.3, but in Meyers's example yields lower credibilities. Thus, once again, we raise the loss limit to make up for a too low weight applied to the primary losses.

In Table 7.3, we notice that, for a given loss limitation, after a certain point there is little increase in efficiency with increasing size of risk. This is explained in Appendix F.

8. SELF-RATING POINT

Formulas 6.2 and 6.3 for credibility illustrate the theoretical validity of a self-rating point. One can have $Z \geqq 1$. In fact, this is the case for Meyers's third example.

It makes sense to define the self-rating point as the smallest size risk such that $Z = 1$ at the optimum loss limit. Using formula 6.2 this means that

$$\frac{N^2\,\hat{c} + N^2\,\hat{s}}{N^2\,\hat{c} + N\,\hat{t}}\left(\frac{E_p}{E_p + E_e}\right) = 1, \tag{8.1}$$

whereas in Appendix G we let:

$$\hat{c} = \frac{c}{N^2},$$

$$\hat{s} = \frac{s}{N^2}, \text{ and}$$

$$\hat{t} = t/N = (a - c)/N.$$

Formula 8.1 can be solved for

$$N = \frac{\hat{t}\,(E_p + E_e)}{\hat{s}\,E_p - \hat{c}\,E_e} \tag{8.2}$$

In the Meyers example, we saw in Table 7.2 that the optimum loss limit is about 2.5. For this loss limit we have $\hat{c} = .041$, $\hat{s} = .054$, $\hat{t} = .548$, $E_p = .495$, and $E_e = .375$. Thus, formula 8.2 gives $N \cong 42$.

Thus, in this example, the self-rating point would be 42.[39] When using formula 6.3, we expect the optimum loss limit to increase above this self-rating point. This is confirmed by Table 7.3 where the optimum loss limit began to increase for $N \geqq 64$ after remaining constant for $N \leqq 32$.

9. CONCLUSIONS FROM MEYERS'S GENERAL LIABILITY EXAMPLE

Those features and assumptions of Meyers's third example that are probably true for most applications are:

(1) Both the frequency and severity distributions vary between risks (although perhaps not in the same relative importance as in this example).
(2) Frequency and severity are somewhat correlated (although probably not to the extent they are in the example).
(3) The excess losses have a much higher coefficient of variation than do the primary losses.
(4) The primary severity and excess severity are highly correlated.

Based on the above analysis of Meyers's example, when the general assumptions of his example hold, the behavior we expect to see for a single split experience rating plan is as follows:

(1) The optimum loss limit increases slowly as the size of risk gets larger up to the self-rating point.[40] (It shall be shown when examining Meyers's next example that the optimum loss limit remains virtually constant here because of the particular choice of parameters for this example.)
(2) The optimum loss limit increases more rapidly as the size of risk increases beyond the self-rating point.
(3) The more important the differences in severity, the higher the optimum loss limit. The more important the differences in frequency, the lower the optimum loss limit.
(4) The efficiency is very close to optimal for loss limits close to optimal.
(5) The optimal credibilities will not be of the form $N/(N + K)$, although such a formula will give efficiencies close to optimal. (This will not be true if any of the phenomena discussed in Sections 2 and 3 of this discussion are significant.)

---

[39] This corresponds in this example to expected basic limits losses of about $37,000, or about 26 claims on average.

[40] In certain cases it may not be appropriate to have a self-rating point. For example, this will be the case if the phenomena discussed in Sections 2 and 3 of this discussion significantly reduce the credibilities.

10. MULTI-SPLIT PLAN

In Section 6 of his paper, Mr. Meyers constructs an example to illustrate the behavior of a multi-split plan such as that currently used for workers' compensation. As in his Section 5, the risks are divided into a small number of possible types.[41]

He uses a Poisson distribution to model the claim counts. He uses a (continuous) Weibull distribution to model claim severity. The frequency and severity are treated as independent.[42]

No overall limit is applied by the author to the losses. In actual application of the Experience Rating Plan, a per claim accident limit is applied.[43] The value differs considerably by state, but I will use \$100,000 here for illustrative purposes. If the severity distribution has a large tail, this accident limitation can add to the efficiency of the plan.

10.1 The Current Workers' Compensation Experience Rating Plan

Mr. Meyers examines whether the current Workers' Compensation Experience Rating Plan or his formula 6.1 works better, i.e., which produces higher efficiency. Mr. Meyers concludes that his formula 6.1, which gives no credibility to the excess losses, outperforms the current Workers' Compensation plan.

The first thing to note is that the current workers' compensation formula can be written as the new estimate of expected losses equals the old estimate of expected losses times the experience modification,

$$F = \frac{A_p + W A_e + (1 - W)\, E_e + (1 - W)\, K}{E + (1 - W)\, K} E\,. \tag{10.1}$$

Then, following Snader [9], let

$$Z_p = \frac{E}{E + K_p}\,, \tag{10.2}$$

$$Z_e = \frac{E}{E + K_e} = W\, Z_p\,, \tag{10.3}$$

$$K_p = (1 - W)K = B, \text{ and} \tag{10.4}$$

[41] In Appendix J, an example is given of a continuous distribution of risks.

[42] In Section 5 of his paper, Mr. Meyers's example had the frequency and severity highly correlated.

[43] For accidents involving multiple claims, the limitation is twice that for single claim accidents.

$$K_e = \frac{(1 - W)(K + E)}{W} . \tag{10.5}$$

Then formula 10.1 can be written as

$$F = E_p (1 - Z_p) + A_p Z_p + E_e (1 - Z_e) + A_e Z_e . \tag{10.6}$$

Formula 10.6 is the form of F that has been previously discussed. As was shown previously, the solution for the optimal Z_p and Z_e given in formulas 5.3 and 5.4 do not have the form of formulas 10.2 and 10.3, even if K_p and K_e were constants with size of risk. In the current workers' compensation plan, $1 - W$ and thus K_p decreases with increasing size of risk until it is zero for self-rated risks. Below a certain value, $W = 0$ and thus $Z_e = 0$. Above that value, W increases to 1 with increasing size of risk, and K_e decreases with increasing size of risk until it reaches zero for self-rated risks. Values of K_p, K_e, Z_p, and Z_e are displayed in Table 10.1 for a typical choice of parameters.[44]

Mr. Meyers makes the excellent point that there is no theoretical framework in which the standard workers' compensation formula (with this particular variation of K_p and K_e with size of risk) is optimal.[45]

10.2 Meyers's Alternative, Zero Excess Credibility

Meyers's formula 6.1 can be written as

$$F = E \left(\frac{A_p + K}{E_p + K}\right) , \text{ or}$$

$$F = E_p \left\{1 - \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right)\right\} + A_p \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right) + E_e . \tag{10.7}$$

which is a special case of formula 5.1 with

$$Z_p = \left(1 + \frac{E_e}{E_p}\right) \left(\frac{E_p}{E_p + K}\right) , \text{ and} \tag{10.8}$$

$$Z_e = 0$$

[44] These are the values currently in use in Massachusetts.

[45] It is also of interest to note that Z_e increases very quickly with the size of risk.

TABLE 10.1

WORKERS' COMPENSATION EXPERIENCE RATING
CURRENT PLAN
EXAMPLE FOR TYPICAL VALUES*

Expected Losses	Credibility Parameters K (00)		Credibilities (Percent)	
(000)	Primary	Excess	Primary	Excess
5	200	Infinite	20	0
10	200	Infinite	33	0
15	200	Infinite	43	0
20	200	Infinite	50	0
25	200	Infinite	56	0
30	198	49500	60	1
35	196	26950	64	1
40	194	19400	67	2
45	194	21017	70	2
50	192	16800	72	3
60	188	12533	76	5
70	184	10350	79	6
80	182	10111	81	7
90	178	8900	83	9
100	174	8031	85	11
120	168	7350	88	14
140	162	6821	90	17
160	154	6026	91	21
180	148	5692	92	24
200	140	5133	93	28
225	132	4756	94	32
250	124	4405	95	36
275	116	4074	96	40
300	106	3609	97	45
325	98	3315	97	50
350	90	3027	97	54
375	82	2745	98	58
400	72	2363	98	63
425	64	2094	99	67
450	56	1828	99	71
475	48	1563	99	75
500	38	1220	99	80
525	30	962	99	85
550	22	704	100	89
575	14	448	100	93
600	6	192	100	97
625	0	0	100	100

* Self-Rating Point varies by state. Self-Rating Point taken as S = \$615,000. Q = \$25,000. K = \$20,000. See Snader [9].

As we saw in the discussion of Meyers's Section 5, the use of the credibilities in formula 10.8 results in less efficiency than the use of the theoretically optimal credibilities, but for many applications the loss in efficiency may be acceptable.

10.3 A Modification of Meyers's Example

The loss in efficiency will be examined for the example in Meyers's section 6. In order to simplify the calculations, a discrete version of the Weibull will be used.[46] In order to better match the current plan, an accident limitation will be used. The accident limitation will be chosen at \$100,000. The probability, $F(x)$, that a claim will be less than or equal to x is given by:[47]

$$F(x) = 1 - e^{-(x/b)^c} \qquad x = \$100, \$200, \ldots \$100{,}000 \tag{10.9}$$

The remaining probability will be at the accident limitation \$100,000.

The primary portion of each loss will be determined by the multi-split formula

$$x_p = \begin{cases} x, & x \leqq B \\ \dfrac{(B + C)\,x}{x + C}, & x \geqq B\,. \end{cases} \tag{10.10}$$

The current plan has $C = 4B$ and $B = \$2{,}000$.

The frequency will be Poisson as per Mr. Meyers; however, we will take the parameter λ equal to N times 4, 7, 10, 13, and 16. Thus N represents the size of risk, with $N = 10$ in Meyers's example.[48]

Appendix I shows how to calculate the quantities that enter formulas 5.3 and 5.4, when the frequency and severity are independent.

The results for this example with $N = 10$ are $a = 86961$, $b = 349814$, $c = 65952$, $d = 101857$, $r = 123788$, and $s = 77843$, $Z_p = 185.8\%$, $Z_e = -14.4\%$, and Efficiency $= 74.6\%$.

---

[46] The use of the discrete version changes some of the actual values, but the essence of the example is preserved. It should be noted that the continuous Weibull usually does not fit the observed size of loss distribution for small claims. For example, with $b = 50$ and $c = .25$, 31% of the losses will be one dollar!

[47] Meyers takes $c = .25$. He lets $b = 30, 40, 50, 60$, and 70 with equal probability.

[48] It should be noted that really small risks are currently not eligible for experience rating. The eligibility level for a risk with three years of experience eligible for experience rating is generally set so that for each state it approximates the average premium of a risk with 10 full time employees. (For example, in one state, this is currently \$3500 in premium per year.)

Using formula 6.2 would result in $Z = 81.4\%$ and an efficiency of 73.5%. This is quite close to optimal.

Using formula 6.4 would result in $Z = c/a = 75.8\%$ and an efficiency of 73.2%. This is still quite close to optimal. Thus, as we saw in Section 7 of this discussion, the use of formula 6.4, as suggested by Mr. Meyers, results in a relatively small loss in efficiency compared to optimal.

10.4 Results of Using the Various Formulas for Credibilities

Table 10.2 gives, for various sizes of risks and various choices of B (as per formula 10.10), the efficiencies for this example using formulas 5.3 and 5.4 for credibility. We notice that for smaller risks, the optimal choice of B is much lower than the current value of \$2000.[49] The optimal B rises with size of risk. For the largest risks it reaches \$2000.

Table 10.3 is similar to Table 10.2, except that the credibilities are calculated using formula 6.2. There is a similar pattern to Table 10.2. The loss in efficiency is relatively small compared to Table 10.2. Table 10.4 uses formula 6.3 in order to calculate the credibilities. It is virtually identical to Table 10.3.

Table 10.5 is similar to the preceding tables, except that the credibilities are calculated using formula 6.4, as recommended by Meyers. The pattern is again very similar, and the losses in efficiency are probably sufficiently small to be acceptable for most practical applications. Thus, at least for this example, there is little disadvantage to setting the excess credibility equal to zero.

Table 10.6 is similar to Table 10.2, except that a single split plan has been used instead of a multi-split plan. The loss in efficiency is relatively small. Thus, at least for this example, there is little disadvantage to the use of the simpler single split plan.

Table 10.7 is similar to Table 10.2, except that an accident limitation of \$200,000 rather than \$100,000 has been used. The pattern is similar to that in Table 10.2. As expected, the optimal value of B is (slightly) higher, since the tail of the severity distribution is relatively more important. The efficiencies in Table 10.2 are lower than those in Table 10.7, but it is inappropriate to compare

[49] This result depends on Meyers's choice in this example of the relative importance of variation of frequency and variation of severity as well as the use of the Weibull distribution even for small sizes of claims.

them directly. Table 10.7 is the result of attempting to estimate losses capped at $200,000. This is a more difficult task than trying to estimate losses capped at $100,000 as in Table 10.2.[50]

TABLE 10.2

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULAS 5.3 AND 5.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	78.4%	85.9%	92.7%
$200	41.4	63.6	79.1	86.9	93.1
$500	34.6	58.9	78.5	87.8	93.7
$1,000	29.6	54.3	77.0	88.0	94.0
$2,000	25.2	49.4	74.6	87.7	94.3
$5,000	20.3	42.9	70.5	86.5	94.3
$10,000	17.2	38.3	66.8	85.0	94.1

TABLE 10.3

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.2

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	77.1%	81.2%	82.7%
$200	41.0	63.5	78.7	84.4	86.6
$500	33.6	58.3	78.5	87.1	90.6
$1,000	28.2	53.0	76.7	87.9	92.7
$2,000	23.2	47.0	73.5	87.6	93.9
$5,000	17.4	38.7	67.3	85.4	94.3
$10,000	13.9	32.7	61.7	82.6	93.7

[50] One could put the two tables on a comparable basis by adjusting the efficiencies in Table 10.2 to what they would have been if one measures efficiency in terms of the variation of the losses capped at $200,000. However, this is beyond the scope of this discussion.

TABLE 10.4

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.3

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.8%	65.5%	77.1%	81.1%	82.6%
$200	41.0	63.5	78.7	84.3	86.3
$500	33.6	58.3	78.5	87.1	90.3
$1,000	28.2	53.0	76.7	87.9	92.4
$2,000	23.2	47.0	73.5	87.6	93.8
$5,000	17.4	38.7	67.3	85.4	94.3
$10,000	13.9	32.7	61.7	82.6	93.7

TABLE 10.5

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULA 6.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
$100	45.6%	65.3%	76.8%	80.9%	82.5%
$200	40.8	63.2	78.3	84.0	86.2
$500	33.4	58.0	78.1	86.6	90.1
$1,000	28.0	52.7	76.3	87.4	92.2
$2,000	23.0	46.8	73.2	87.2	93.5
$5,000	17.4	38.6	67.1	85.2	94.1
$10,000	13.9	32.6	61.6	82.5	93.6

TABLE 10.6

EFFICIENCIES
SINGLE SPLIT PLAN, FORMULAS 5.3 AND 5.4

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
\$100	47.8%	63.8%	74.8%	83.4%	91.9%
\$200	48.2	65.4	76.8	84.6	92.3
\$500	44.1	64.7	78.6	86.2	92.8
\$1,000	38.8	61.9	78.9	87.2	93.3
\$2,000	33.0	57.5	78.1	87.9	93.8
\$5,000	25.6	49.9	74.9	87.7	94.3
\$10,000	20.6	43.5	70.9	86.6	94.4
\$20,000	16.4	36.9	65.6	84.4	94.0

TABLE 10.7

EFFICIENCIES
MULTI-SPLIT PLAN, FORMULAS 5.3 AND 5.4
\$200,000 ACCIDENT LIMITATION

B	$N = 1$	$N = 3$	$N = 10$	$N = 30$	$N = 100$
\$100	44.7%	64.0%	76.6%	83.9%	91.1%
\$200	40.4	62.2	77.5	85.3	91.7
\$500	33.7	57.7	77.3	86.6	92.5
\$1,000	28.7	53.1	75.8	87.0	93.1
\$2,000	24.2	48.0	73.4	86.7	93.5
\$5,000	19.1	41.1	68.8	85.4	93.7
\$10,000	16.0	36.1	64.7	83.7	93.5

11. TESTING AN EXPERIENCE RATING PLAN ON ACTUAL DATA

In Section 7 of his paper, Mr. Meyers gives a generally applicable method of testing experience rating plans. It is a more modern and statistically sophisticated version of the method presented by Dorweiler. The author uses this method on actual data to test which formula for credibilities performed best, as well as to test which values of parameters worked best given a particular formula.

Meyers looked at formulas 2.1 and 2.2, and found that formula 2.2, which assumes parameter uncertainty, performed better. It would be interesting to perform the same test on other candidates, such as formula 2.5. In fact, given the other features of the plan, one can find by trial and error a relation of credibility with size of risk that works well.

Given the appropriate data, the general method presented by the author should be able to answer the following questions which were not tested in the paper.

1. What is the best loss limit to use?
2. Does a loss limit which increases with the size of risk significantly improve the performance of the plan?
3. Does a multi-split plan perform significantly better than a single split plan?
4. Does assigning non-zero credibility to the excess losses perform significantly better than assigning zero credibility to the excess losses?

Mr. Meyers is able to use the method in this section not only to get a point estimate of the credibility parameter K, but also to get a confidence interval for K. It is interesting to note that the best estimate of K is not at the center of the confidence interval. Rather, the best estimate is nearer the low end of the confidence interval. This is at least partially explained by the fact that it is the ratio of different estimates of K rather than their difference which is important.[51]

[51] This feature of estimates of K is discussed in Mahler [6]. For example, if one either doubles or halves K, the resulting maximum changes in credibility are the same. The connection to the result here was pointed out to this author by Mr. Meyers.

12. CONCLUSIONS AND SUMMARY

While Mr. Meyers's paper is an excellent contribution to the actuarial literature which opens up many areas for further investigation, I think the author goes a little too far in drawing conclusions from his work. I will arrange my conclusions in a manner parallel to the author's, in order to allow ready comparison.

MEYERS	MAHLER
1. A loss limit can be an effective tool for increasing the accuracy of an experience rating formula. Loss limits are particularly helpful when there are differences in claim frequency. Even if the only differences among the insureds are in claim severity, little accuracy will be lost with a loss limit.	1a. A loss limit can be an effective tool for increasing the accuracy of an experience rating formula.[52] 1b. If the differences between risks within rating classes are mainly due to differences in frequency, then a lower loss limit is optimal. If they are mainly due to severity, then a higher loss limit is optimal. 1c. The efficiency is relatively insensitive to the choice of the loss limit. If your chosen loss limit is close to optimal, and assuming you choose credibilities close to optimal, then your efficiency will be very close to optimal. 1d. The larger the maximum loss that can occur,[53] and the thicker the tail of the size of loss distribution,[54] the more important it is to have a loss limit.

[52] Appendix J gives an example where there is no practical advantage to the use of a loss limit.

[53] Here we mean the largest loss considered by the plan, that is $50,000 in the Meyers's general liability example and $100,000 in my version of Meyers's workers' compensation example.

[54] The Pareto has a very thick tail. The Weibull has a thick tail for Meyers's $c < 1$, but not quite as thick as the Pareto. See Hogg and Klugman [10].

MEYERS	MAHLER
	1e. The optimal loss limit increases slowly or remains virtually constant up to a large size of risk. For very large risks, the optimal loss limit increases more quickly.
	1f. The optimal credibilities depend on the loss limit chosen.
	1g. Under certain conditions, there is a theoretical as well as a practical justification for having a self-rating point.
2. The current formula in the Workers' Compensation Experience Rating Plan, which has a separate treatment of primary and excess losses, is less accurate than a formula which uses only primary losses.	2a. The current formula used in the Workers' Compensation Experience Rating Plan can be improved.
	2b. The current manner in which the credibilities in the Workers' Compensation Plan vary with size of risk has no theoretical justification. Empirical studies should be done to come up with more appropriate relationships.
	2c. The gain in efficiency from the use of a multi-split rather than a single split plan may not be large enough to justify the use of the more complicated multi-split plan.
	2d. The excess credibilities are expected to be relatively small. The gain in efficiency may not be large enough to justify the use of the excess losses.
	2e. The per claim limitation in the Workers' Compensation Plan serves a useful purpose.

MEYERS

3. There are some very plausible situations when the standard credibility formula $Z = E/(E + K)$ is not appropriate. These include parameter uncertainty over time and loss limit which increases with the size of the insured. Failure to recognize this will result in overstating credibilities for larger insureds.

MAHLER

3a. The traditional formula for credibility, formula 2.1, applies in only limited special situations.

3b. Three specific phenomena are examined: parameter uncertainty, shifting parameters over time, and risk heterogeneity. Each of these will tend to lower credibilities for large risks compared to those from the traditional formula. (Formulas for each are presented.) One or more of them are expected to be important in many situations.

3c. Under certain conditions, the optimal credibility will remain substantially less than one, regardless of how large the risk gets or how many years of data are used.

3d. Under certain circumstances, older years of data should be given substantially less credibility than more recent years of data. There may be only a minimal gain in efficiency from using additional years of data.

3e. If the loss limit changes, the optimal credibilities also change. This is another reason why formula 2.1 may not apply.

3f. The efficiencies are relatively insensitive to the choice of the credibilities. The credibilities, in turn, are relatively insensitive to the choice of parameters entering the formula.

MEYERS	MAHLER
4. The author would recommend an experience rating formula based on the credibility formula 2.2 $$Z = E/(JE + K).$$ A loss limit that does not vary by size of insured should be a part of the plan. Excess losses should not be a part of the plan. This formula is less complicated than current formulas and should be easier to administer.	4. Any reasonable experience rating plan is expected to achieve a substantial increase in efficiency. However, theoretical and empirical studies should allow a significant improvement in the efficiency of most plans.

Mr. Meyers's paper has already stimulated work on experience rating plans which should lead to substantial improvements in the design of these plans in the near future. The paper also examines some interesting features of credibility which should have implications outside the area of experience rating plans.

REFERENCES

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

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

[3] Herbert E. Wittick, "The Canadian Merit Rating Plan for Individual Automobile Risks," *PCAS* XLV 1958, p.214.

[4] Robert A. Bailey and LeRoy J. Simon, "Two Studies in Automobile Insurance Ratemaking," *PCAS* XLVII 1960, p.1.

[5] William J. Hazam, Discussion of [2], *PCAS* XLVII 1960, p.150.

[6] Howard C. Mahler, "An Actuarial Note on Credibility Parameters," *PCAS* LXXIII 1986, p. 1.

[7] Gary G. Venter, Discussion of [8], *PCAS* LXX 1983, p.62.

[8] Philip E. Heckman and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX 1983, p.22.

[9] Richard H. Snader, "Fundamentals of Individual Risk Rating and Related Topics," Part I of Study Note, Casualty Actuarial Society, 1980.

[10] Robert V. Hogg and Stuart A. Klugman, *Loss Distributions,* John Wiley & Sons, 1984.

[11] George W. Snedecor and William G. Cochran, *Statistical Methods,* The Iowa State University Press, 1980.

[12] Gary G. Venter, "Classical Partial Credibility with Application to Trend," *PCAS* LXXIII 1986, p. 27.

[13] Richard G. Woll, "A Study of Risk Assessment," *PCAS* LXVI 1979, p.84.

[14] Michael A. Walters, "Homeowners Insurance Ratemaking," *PCAS* LXI 1974, p.15.

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

APPENDIX A

TWO RESULTS FOR COVARIANCES

In this appendix will be established two useful results involving covariances. The first result is:

Total covariance = expected value of the process covariance
+ parameter covariance

where the parameter covariance is another term for the covariance of the hypothetical means. It should be noted that the similar result involving variances is just a special case of this result.[55]

If θ represents the set of parameters, this result can be written as follows.

Theorem: $\mathrm{COV}[X,Y] = E_\theta\,[\mathrm{COV}[X,Y|\theta]] + \mathrm{COV}_\theta\,[E[X|\theta], E[Y|\theta]]$

Proof:

$$\begin{aligned}
\mathrm{COV}[X,Y] &= E[XY] - E[X]\,E[Y] \\
&= E_\theta\,[E[XY|\theta]] - E_\theta[X|\theta]\,E_\theta[Y|\theta] \\
&= E_\theta\,[E[XY|\theta]] - E_\theta\,[E[X|\theta]\,E[Y|\theta]] \\
&\quad + E_\theta\,[E[X|\theta]\,E[Y|\theta]] - E_\theta\,[X|\theta]\,E_\theta\,[Y|\theta] \\
&= E_\theta\,[E[XY|\theta] - E[X|\theta]\,E[Y|\theta]] \\
&\quad + E_\theta\,[E[X|\theta]\,E[Y|\theta]] - E_\theta[X|\theta]\,E_\theta[Y|\theta] \\
&= E_\theta\,[\mathrm{COV}[X,Y|\theta]] + \mathrm{COV}_\theta\,[E[X|\theta], E[Y|\theta]]
\end{aligned}$$

The second result puts the covariance of the primary losses and excess losses in terms of the means, variances, and covariance of the primary and excess severity and the frequency. The familiar result for the variance of the losses in terms of frequency and severity is a special case of this result.[56]

Theorem: Assume a claims process in which frequency and severity are independent of each other, and the claim sizes are mutually independent random variables with a common distribution. Then let each claim be divided into two pieces in a well-defined manner not dependent on the number of claims. For convenience, we refer to these two pieces as primary and excess.

---

[55] Both results are familiar to statisticians. See, for example, Snedecor and Cochran [11]. However, the result for variances seems more familiar to actuaries.

[56] The proof given here parallels that for the more familiar result given in Appendix 2 of Venter [12].

Let:

$$\begin{aligned} T_p &= \text{Primary Losses} \\ T_e &= \text{Excess Losses} \\ X_p &= \text{Primary Severity} \\ X_e &= \text{Excess Severity} \\ N &= \text{Frequency} \end{aligned}$$

Then:

$$\text{COV}\ [T_p,T_e] = E[N]\ \text{COV}\ [X_p,X_e] + \text{VAR}[N]\ E[X_p]\ E[X_e]$$

Proof: T_p is the sum of the individual primary portions of claims $X_p(i)$, where i runs from 1 to N, the number of claims. Similarly, T_e is a sum of $X_e(i)$. Since N is a random variable, both frequency and severity contribute to the covariance of T_p and T_e.

To compute the covariance of T_p and T_e, begin by calculating

$$E\ [T_p\ T_e \mid N = n],$$

fix the number of claims n and find

$$E\left[\left(\sum_{i=1}^{n} X_p(i)\right)\left(\sum_{i=1}^{n} X_e(i)\right)\right].$$

Expanding the product yields n^2 terms of the form $X_p(i)\ X_e(j)$. When $i = j$ the expected value of the term is

$$E[X_p(i)\ X_e(i)] = \text{COV}\ [X_p,X_e] + E[X_p]\ E[X_e]$$

from the definition of covariance. Otherwise it is $E[X_p]\ E[X_e]$, since then $X_p(i)$ and $X_e(j)$ are independent. Thus

$$E\left[\left(\sum_{i=1}^{n} X_p(i)\right)\left(\sum_{i=1}^{n} X_e(i)\right)\right] = n\ \text{COV}\ [X_p,X_e] + n^2\ E[X_p]\ E[X_e].$$

Now, by general considerations of conditional expectations,

$$E\ [T_pT_e] = E_n\ [E[T_pT_e \mid N = n]].$$

Thus, taking the expected value of the above equation with respect to N gives

$$\begin{aligned} E\ [T_pT_e] &= E[N]\ \text{COV}\ [X_p,X_e] + E[N^2]\ E[X_p]\ E[X_e], \text{ and} \\ \text{COV}\ [T_p,T_e] &= E[T_pT_e] - E[T_p]\ E[T_e] \\ &= E[N]\ \text{COV}[X_p,X_e] + (\text{VAR}[N] + E^2[N])\ E[X_p]\ E[X_e] \\ &\quad - E[N]\ E[X_p]\ E[N]\ E[X_e] \\ &= E[N]\ \text{COV}[X_p,X_e] + \text{VAR}[N]\ E[X_p]\ E[X_e]. \end{aligned}$$

APPENDIX B

PARAMETER UNCERTAINTY

In this appendix, the effect of parameter uncertainty on the formula for credibility is discussed. This discussion is intended to aid in the understanding of both Meyers's result for this phenomenon and the results obtained here in the later appendices for the two other similar phenomena, risk heterogeneity and shifting parameters over time.

As explained in Mr. Meyers's Appendix A, when there is parameter uncertainty, the credibility as a size of risk no longer follows formula 2.1, but rather follows formula 2.2.

The important point is the behavior of the expected variance within classes, which Meyers labels δ^2.

$$\delta^2 = E\,[(A - \mu)^2]$$

This is normally thought of as the expected value of the process variance. If for each risk the parameters of the risk process themselves vary randomly,[57] then δ^2 really is made up of two pieces.[58] The first piece of δ^2 is due to the variance of the parameters due to different states of the universe.[59] The second piece of δ^2 is due to the process variance, given a specific state of the universe. The first piece is expected to be proportional to N^2, just as was the variance between risks due to different parameters.[60] The second piece is expected to be proportional to N as usual.

In other words, we can write δ^2 as

$$\delta^2 = N^2\,\alpha^2 + N\,\chi^2.$$

[57] An example is given in Section 3 of this discussion.

[58] This is a special case of the first result in Appendix A.

[59] In the example in Section 3, there were for simplicity two states, determined by whether it was raining or not.

[60] Thus somewhat paradoxically, δ^2, which is usually thought of as "process variance" actually includes a piece of "parameter variance," albeit of a very special variety.

The "good" piece of δ^2 goes up only as N, while the "bad" piece goes up as N^2. This bad piece of δ^2 was introduced due to the assumed different possible states of the universe. Unlike the good piece of δ^2, this piece of δ^2 increases as quickly as the variation between risks (Meyers's τ^2), which also increases as N^2.

Thus taking more observations will not get rid of the effect due to the variation inherent in the universe.[61]

If

$$\tau^2 = N^2 \beta^2.$$

then

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N}{N(1 + \alpha^2/\beta^2) + \chi^2/\beta^2},$$

which is of the form

$$Z = \frac{N}{NJ + K}, \qquad J \geqq 1,$$

which is equation 2.2.

[61] The same idea has applications to risk assessment efficiency and class homogeneity. See Woll [13].

APPENDIX C

RISK CHARACTERISTICS CHANGING OVER TIME

In this appendix, the effect of shifting parameters over time on the formula for credibility is discussed. A general formula is derived and the results are applied for a reasonable special case. It is this special case that results in formula 2.3 in the main text.

Assume that the parameters that describe the loss process of a risk are not constant over time.[62] Then the theoretical true mean for each risk is a function of time. For example, the shift might be due to a change in the attitude of management with regard to safety or due to a change in the upkeep of the roads on which the insured usually drives his car. We are not including shifts that are expected to affect all risks in the same manner, for example, claim cost trend.

Assume we have an experience rating plan, and we use N years of data, such that our estimate of the mean F we expect in year $N + \Delta$ is given by:

$$F = \left(1 - \sum_{i=1}^{N} Z_i\right) E_{N+\Delta} + \sum_{i=1}^{N} Z_i A_i$$

where A_i is the actual losses observed for year i, (brought up to the expected level of year $N + \Delta$),[63] Z_i is the credibility assigned to that year losses, and $E_{N+\Delta}$ is the expected losses for year $N + \Delta$.[64]

We assume that the individual years of data are generated by the same size of risk (or same number of risks). Thus except for the assumed shifting parameters over time, we would assign each of the years equal weight.

General Case

We wish to find Z_i for $i = 1$ to N, such that the efficiency is maximized. In order to proceed, we will assume a covariance structure. We will assume that the correlation expected between two years of data separated in time by i

---

[62] We assume that while individual risk parameters shift, the overall distribution of risks remains the same. Also, we assume there is no way to predict the shift of an individual risk, and that the class plan doesn't pick up the shift.

[63] Thus we assume $E_i = E_j = E$.

[64] Typically $N = 3$. For workers' compensation usually $\Delta = 2$. For private passenger auto usually $\Delta = 1$.

years, is a function of i, $l(i)$. It is assumed that the expected correlation decreases (or stays the same) as the years get further apart. Specifically, we assume:

$$l(0) = 1$$

$$l(i) \geqq l(i + 1).$$

The covariance structure assumed is:

$$E[(A_i - E)(A_j - E)] = \beta^2\, l(|i - j|) + \delta_{ij}\chi^2$$

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

where for different years, $i \neq j$, we get the variation of the hypothetical means, β^2, times a factor equal to a correlation $l(|i - j|) \leqq 1$, dependent on the number of years of difference. For $i = j$, we get the variation of the hypothetical means β^2, plus the expected value of the process variance χ^2. The closer $l(|i - j|)$ is to 1, the less shifting of parameters there is over time.

$$\text{Efficiency} = 1 - \frac{E\,[(F - \mu_{N+\Delta})^2]}{E\,[(E - \mu_{N+\Delta})^2]}$$

Substituting our expression for F, letting $K = \chi^2/\beta^2$ and simplifying we get:

$$\text{Efficiency} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} Z_i\, Z_j\,\{l(|i - j|) + K\delta_{ij}\} - 2\sum_{i=1}^{N} Z_i\, l(|N + \Delta - i|)}{E[(E - \mu_{N+\Delta})^2]}$$

We get the maximum efficiency by setting each of the partial derivatives with respect to the different credibilities equal to zero. This gives the following set of N linear equations in N unknowns.

$$\sum_{i=1}^{N} Z_i\,\{l(|i - j|) + \delta_{ij}K\} = l(|N + \Delta - j|) \qquad j = 1, 2, \ldots, N$$

These equations can be solved simply by using the usual matrix methods. However, for $N > 2$ the expressions for the solutions are complicated to actually write out. For $N = 2$ remembering that $l(0) = 1$, we get:

$$Z_1 = \frac{(1 + K)\, l(\Delta + 1) - l(1)\, l(\Delta)}{(1 + K)^2 - l(1)^2}$$

$$Z_2 = \frac{(1 + K)\, l(\Delta) - l(1)\, l(\Delta + 1)}{(1 + K)^2 - l(1)^2}$$

The ratio of Z_1 to Z_2 is given by:

$$\frac{Z_1}{Z_2} = \frac{(1 + K)\, l(\Delta + 1) - l(1)\, l(\Delta)}{(1 + K)\, l(\Delta) - l(1)(\Delta + 1)}$$

For the usual case where $l(\Delta + 1) < l(\Delta)$, this ratio is less than one, and thus $Z_1 < Z_2$. As expected the more recent data (year 2) is given more credibility than the less recent data (year 1).

$$Z_1 + Z_2 = \frac{(1 + K) - l(1))(l(\Delta + 1) + l(\Delta))}{(1 + K)^2 - l(1)^2}$$

$$= \frac{l(\Delta + 1) + l(\Delta)}{1 + l(1) + K}$$

For $l(i) = 1$ for all i, this reduces to the familiar $2/(2 + K)$.

For $N > 2$, one can approximate the exact solution as follows.

Adding up the N linear equations gives after rearranging the order of summation:

$$\sum_{i=1}^{N} Z_i \left\{ \left[\sum_{j=1}^{N} l(|i - j|) \right] + K \right\} = \sum_{j=1}^{N} l(|N + \Delta - j|)$$

The term $\sum_{j=1}^{N} l(|i - j|)$ depends on the value of i. For $i = 1$ and $i = N$ it is equal to $\sum_{j=0}^{N-1} l(j)$. For values of i between 1 and N, smaller values of $|i - j|$ are duplicated in the sum, while larger values of $|i - j|$ no longer enter it.[65] Thus since we have assumed that l is a decreasing function, we have

$$\sum_{j=1}^{N} l(|i - j|) \geqq \sum_{j=0}^{N-1} l(j)$$

Thus if we substitute $\sum_{j=0}^{N-1} l(j)$ into our previous equation wherever $\sum_{j=1}^{N} l(|i - j|)$ appeared we get:

$$\left(\sum_{i=1}^{N} Z_i \right) \left\{ \left(\sum_{j=0}^{N-1} l(j) \right) + K \right\} \leqq \sum_{j=1}^{N} l(|N + \Delta - j|)$$

This suggests the following approximation, which gives lower credibilities than the exact solution.

[65] For example, if $N = 5$ and $i = 2$, then the sum is $l(1) + l(0) + l(1) + l(2) + l(3)$.

$$\sum_{i=1}^{N} Z_i \cong \frac{\sum_{i=0}^{N-1} l(\Delta + i)}{\left(\sum_{i=0}^{N-1} l(i)\right) + K}$$

The sum of the Z_i is the credibility assigned to the data for all the separate years combined. In the main text this is called Z.

For $l(i) = 1$ for all i, this reduces to the familiar $Z = N/(N + K)$.

If the covariance structure has the basic property assumed here, that the correlation between individual years of data is smaller the longer the time span between the years, then the credibilities will have the following properties. The credibility assigned to a more recent year of data should be higher than that assigned to a more distant year of data.[66] If the correlation between distant years of data is significantly lower, i.e., l (i) gets small relatively quickly, then beyond a certain point using more recent years of data will lead to little improvement in the estimate of the mean. If the individual years of data are given the same weight, then using more years of data will eventually lead to a *worse* estimate of the mean, since the very old data provides a poor estimate of the future mean. Finally, the smaller Δ is, i.e., the less the delay is in getting and using the data, the higher the credibilities and the better the resulting estimate of the future mean.[67]

Special Case

Let us take a special case of the general covariance structure that has been assumed. Let $l(i) = \rho^i$, $\rho \leqq 1$. Then:

$$E[(A_i - E)(A_j - \mathrm{E})] = \beta^2\rho^{|i-j|} + \delta_{ij}X^2$$

Thus for two different years of data, their covariance is proportional to a constant ρ taken to the power equal to the number of years separating them. For $\rho < 1$, the covariance decreases the larger the separation, and goes to zero rather quickly.

[66] This concept of giving more recent data more weight than less recent data is a familiar one to actuaries. See for example, "Homeowners Insurance Ratemaking" by Walters [14]. However, when estimating values at ultimate, it might be appropriate in certain circumstances to assign less weight to more recent but immature data.

[67] For experience rating generally $\Delta \geqq 1$.

While for illustrative purposes this is not an unreasonable assumption for the structure of covariances, it is far from the only assumption that could be made. The actual covariance structure for a particular real world application would have to be determined empirically.

For this special case, we have for $N = 2$:

$$Z_1 = \frac{\rho^{\Delta}\rho K}{(1 + K)^2 - \rho^2}$$

$$Z_2 = \frac{\rho^{\Delta}}{(1 + K)^2 - \rho^2}(1 + K - \rho^2)$$

$$\frac{Z_2}{Z_1} = \frac{1 + K - \rho^2}{\rho K} = \frac{1}{\rho} + \frac{1}{K}\left(\frac{1}{\rho} - \rho\right) > 1$$

$$Z_1 + Z_2 = \frac{\rho^{\Delta}(1 + \rho)}{(1 + \rho) + K}$$

For $\rho = 1$, $Z_1 + Z_2$ reduces to the familiar $2/(2 + K)$.

As discussed above, one can approximate the exact solution by[68]:

$$Z = \sum_{i=1}^{N} Z_i \cong \frac{\rho^{\Delta}\left(\sum_{i=1}^{N} \rho^{i-1}\right)}{\left(\sum_{i=1}^{N} \rho^{i-1}\right) + K}$$

This is formula 2.3. For $\rho = 1$, this reduces to the familiar $N/(N + K)$.

[68] The exact solution gives lower credibilities.

APPENDIX D

RISK HETEROGENEITY

In this appendix, the effect of risk heterogeneity on the formula for credibility for large risks is discussed. In addition, the combined effect of risk heterogeneity and parameter uncertainty is discussed.

In general, a large risk is made up of smaller risks. For example, a large commercial risk might consist of a grouping of separate factories. Assume our different risks consist in each case of grouping together N factories of the same size.[69] Then how does the variance between different risks, $E[(E - \mu)^2]$, which Meyers calls τ^2, depend on the size of risk N?[70]

A Simple Example

To illustrate the point, let us examine a very simple example. Assume that half the factories are "good" and half are "bad." The good factories have an expected mean of one, while the bad factories each have an expected mean of two. Depending on how the factories are grouped together to form risks, τ^2 has a different dependence on N.

Case 1

Assume that the risks consist solely of good factories or bad factories, but never a mixture. Then the risks of size N have an expected mean of either N or $2N$, with equal frequency. Thus τ^2 is $N^2/4$.

Case 2

Assume that the risks consist of good and bad factories grouped together totally at random. A risk of size N, is merely a random sample of size N from the set of all factories. Thus in this case, τ^2 is N times the variance between individual factories. $\tau^2 = N/4$.

Case 3

Assume that half the risks are "superior" and half "inferior". Each factory in a superior risk has a 2/3 chance to be good and a 1/3 chance to be bad. The situation is reversed for inferior risks. Then the expected means of the superior

[69] In certain cases, a factory could be usefully broken up into smaller subunits. We are merely presenting a simple example here.

[70] τ^2 measures how homogeneous the classes are. The smaller τ^2, the less the separation between the risks and the more homogeneous the class.

risks of size N extend from N to $2N$, with the probabilities given by the binomial distribution with $p = 1/3$. The inferior risks also have expected means from N to $2N$, but with the probabilities given by the binomial distribution with $p = 2/3$.[71]

One can compute the variance τ^2 for specific values of N in the usual straightforward manner.[72] However, τ^2 can be broken up into two pieces.[73] The first piece is the variation among different superior risks or the variation among different inferior risks. This is just $2N/9$, since the variance of the binomial distribution is just $Np(1 - p)$. The second piece of τ^2 is the variance between the grand mean of superior risks and the grand mean of the inferior risks. Since these grand means are $4N/3$ and $5N/3$ respectively, this piece of τ^2 is just $N^2/36$. Adding the two pieces together, one gets:

$$\tau^2 = 2N/9 + N^2/36.$$

Generalizing the Simple Example

In Case 1, it was assumed that the risks are homogeneous. Each of the factories making up a risk has the same expected mean. In this case, τ^2 is proportional to N^2, since the expected mean for each risk just gets multiplied by N. This special case is the one that is usually dealt with.

On the other hand, in Case 2, we have assumed the other extreme, that the factories are grouped together totally at random.[74] Each risk is merely a sample of size N from the overall set of factories, and thus τ^2 is N times the variance between the individual factories. In this special case τ^2 is proportional to N.

Thus in the two extreme cases, we have either τ^2 proportional to N or τ^2 proportional to N^2. We expect most real world situations to be in the intermediate situation, such as Case 3, where bad factories are more likely to be grouped

[71] The use of the terms superior and inferior could be thought of in terms of some underwriting criterion. While the average superior risk has a lower expected mean than the average inferior risk, there are inferior risks with low means and vice versa.

[72] For example, for $N = 3$, the risks will have expected means of 3, 4, 5, and 6 with probabilities of 1/6, 1/3, 1/3, 1/6. Thus $\tau^2 = 11/12$. The reader can verify that this matches the formula for τ^2 given below.

[73] This is a special case of the first result given in Appendix A.

[74] As demonstrated in Hewitt [15], for loss ratio distribution purposes, the sum of two $50,000 risks doesn't act the same as a single $100,000 risk. Thus, Case 2 is not a good model of the reality; it is an extreme case chosen for illustrative purposes.

together with bad factories, but a single risk can be made up of both good and bad factories. For this intermediate case τ^2 had the form

$$\tau^2 = N\pi^2 + N^2\beta^2$$

This form for τ^2 follows from breaking the variance into two pieces. The first piece is the variation among risks of similar type.[75] This piece of τ^2 is proportional to N. The second piece is the variance between the grand means of different types of risks. This piece of τ^2 is proportional to N^2.

In order to get the credibility we must combine τ^2 with δ^2. As explained in Appendix B, without parameter uncertainty it makes sense to assume

$$\delta^2 = N\chi^2,$$

then

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N + \pi^2/\beta^2}{N + \dfrac{\chi^2 + \pi^2}{\beta^2}}\,.$$

which can be written in the form

$$Z = \frac{N + I}{N + K}\,, \qquad 0 \leqq I \leqq K$$

which is formula 2.5 in the main text.

While it at first appears that $I > 0$ (i.e. risk heterogeneity) leads to higher credibilities than formula 2.1 ($I = 0$), that is not the case. One must remember that the K in formula 2.5 is not equal to the K in formula 2.1. The K here has an additional term of π^2/β^2 compared to the K in formula 2.1. Thus since risk heterogeneity affects both I and K, a more careful analysis is required.

Formula 2.5 can be rewritten in the form

$$Z = 1 - \frac{\chi^2}{\chi^2 + N(\beta^2 + \pi^2) - (N - 1)\pi^2}\,.$$

The variance between single units (for example, between individual factories) is $\beta^2 + \pi^2$. Keeping $\beta^2 + \pi^2$ constant, we can see from the above equation, that as π^2 increases, Z decreases (for $N > 1$). In other words, the

[75] Thus somewhat paradoxically τ^2, which can be thought of as the "between variance," actually includes a piece of "within variance," albeit of a very special variety. In Appendix B, a similar but reversed situation was explained for δ^2, when there is parameter uncertainty.

greater the risk heterogeneity that is present, the lower the credibility, all other things being equal. Looked at another way, the more risk heterogeneity, the smaller τ^2 is, all other things being equal (for $N > 1$). Therefore, the more risk heterogeneity, the smaller the credibility.

Risk Heterogeneity and Parametric Uncertainty

If both risk heterogeneity and parameter uncertainty (see Appendix B) are important then we have:

$$\delta^2 = N^2\alpha^2 + N\chi^2$$
$$\tau^2 = N\pi^2 + N^2\beta^2$$

Thus:

$$Z = \frac{\tau^2}{\tau^2 + \delta^2} = \frac{N + \pi^2/\beta^2}{N\left(1 + \dfrac{\alpha^2}{\beta^2}\right) + \dfrac{\chi^2 + \pi^2}{\beta^2}}$$

This can be written in the form

$$Z = \frac{N + I}{NJ + K}, \qquad 0 \leqq I \leqq K, J \geqq 1$$

which is equation 2.6.

APPENDIX E

PARAMETER UNCERTAINTY, SHIFTING RISK PARAMETERS AND RISK HETEROGENEITY

In this appendix, all three phenomena discussed in Appendices B, C, and D will be assumed to be of importance. The resulting formula for the sum of the credibilities will be a combination of the features of those in the prior appendices.[76] The notation from the previous appendices will be used.

Let M be a measure of the (average) size of the risk in each year. Let N be the number of years of data used for experience rating.

Assume the following covariance structure:

$$E\,[(A_i - E)(A_j - E)] = \rho^{|i-j|}\,(M^2\beta^2 + M\pi^2) + \delta_{ij}\,(M^2\alpha^2 + M\chi^2)$$

$$E\,[(\mu_i - E)(\mu_j - E)] = \rho^{|i-j|}\,(M^2\beta^2 + M\pi^2)$$

Then proceeding as in the previous appendices we get the following set of linear equations for the optimal credibilities Z_m, $m = 1, \ldots, N$.

$$\sum_{i=1}^{N} Z_i\,\{(M^2\beta^2 + M\pi^2)\,\rho^{|i-m|} + \delta_{im}\,(M^2\alpha^2 + M\chi^2)\} = \rho^{\Delta+N-m}$$

This set of equations can be solved by matrix methods. As in Appendix C, we can get an approximate solution that is exact for $N < 3$. (I, J, and K are defined in the previous appendices.)

$$Z = \sum_{i=1}^{N} Z_i \cong \frac{\rho^{\Delta}\left(\sum_{i=1}^{N} \rho^{i-1}\right)(M + I)}{\left(\sum_{i=2}^{N} \rho^{i-1}\right)(M + I) + JM + K}$$

This is formula 2.7.

[76] The formulas from the prior appendices are special cases of the formula presented here.

APPENDIX F

EFFICIENCY AND CREDIBILITIES FOR SPLIT EXPERIENCE RATING PLANS

In this appendix, the optimal primary and excess credibilities for a split experience rating plan are derived. The solution, equations 5.3 and 5.4 in the main text, is a generalization of the familiar Bühlmann result for the no-split situation. The second part of this appendix explores the results of using credibilities other than the optimal ones.

Assume we have an experience rating plan, and our estimate of the mean F is given by

$$F = (1 - Z_p)\, E_p + Z_p A_p + (1 - Z_e)\, E_e + Z_e A_e \,.$$

The subscripts p and e will stand for primary and excess; however, for now they can be treated as any two well-defined portions of the total losses. E_p and E_e are the expected losses of each type. A_p and A_e are the actual losses of each type. Z_p and Z_e will be thought of as the credibilities assigned to each portion of the losses; however, for now they can be treated as just numbers to be determined.

Efficiency

In accordance with Meyers, define the efficiency of F by

$$\text{Efficiency} = 1 - \frac{E[(F - \mu)^2]}{E[(E_p + E_e - \mu)^2]}\,,$$

where μ is the theoretical true mean for each risk. Let μ_p and μ_e be the excess and primary pieces of μ. $\mu = \mu_p + \mu_e$. μ is a function of the parameters that describe each risk.

$$\begin{aligned}(F - \mu)^2 = {} & Z_p^2\,(A_p - E_p)^2 + Z_e^2\,(A_e - E_e)^2 + (E_p + E_e - \mu)^2 \\ & + 2Z_pZ_e\,(A_p - E_p)\,(A_e - E_e) + 2Z_p\,(A_p - E_p)\,(E_p - \mu_p) \\ & + 2Z_p\,(A_p - E_p)\,(E_e - \mu_e) + 2Z_e\,(A_e - E_e)\,(E_e - \mu_e) \\ & + 2Z_e\,(A_e - E_e)\,(E_p - \mu_p)\end{aligned}$$

$$(E_p + E_e - \mu)^2 = (E_p - \mu_p)^2 + (E_e - \mu_e)^2 + 2(E_p - \mu_p)\,(E_e - \mu_e)$$

Let:

a = total variance of the primary losses;
b = total variance of the excess losses;
c = variance of the hypothetical means of the primary losses;
d = variance of the hypothetical means of the excess losses;
r = total covariance of the primary and excess losses; and,
s = covariance of hypothetical means of the primary and excess losses.

Remembering that μ is only a function of the set of parameters ϕ, μ is subject to parameter variance but not process variance. The actual observed losses A_p and A_e are subject to both parameter and process variance. E_p and E_e are the overall grand means and are subject to neither kind of variance.

Then we have

$E[(A_p - E_p)^2] = a;$
$E[(A_e - E_e)^2] = b;$
$E[(A_p - E_p)(\mu_p - E_p)] = E[(\mu_p - E_p)(\mu_p - E_p)] = c;$
$E[(A_e - E_e)(\mu_e - E_e)] = d;$
$E[(A_e - E_p)(A_e - E_e)] = r;$
$E[(A_p - E_p)(\mu_e - E_e)] = E[(\mu_p - E_p)(\mu_e - E_e)] = s;$ and,
$E[(A_e - E_e)(\mu_p - E_p)] = E[(\mu_e - E_e)(\mu_p - E_p)] = s.$

Substituting these values back in the definition of efficiency, we get

$$\text{Efficiency} = \frac{2Z_p(c + s) + 2Z_e(d + s) - Z_p^2 a - Z_e^2 b - 2Z_pZ_e r}{c + d + 2s}$$

Optimal Credibilities

The optimal credibilities are given by the least squares solution, which results in the maximum efficiency. In order to maximize the efficiency, we set the partial derivatives with respect to Z_p and Z_e equal to zero. This gives

$a\,Z_p + r\,Z_e = c + s$, and
$r\,Z_p + b\,Z_e = d + s$.

The solution of this simple set of two equations in two unknowns is

$$Z_p = \frac{(c + s)b - (d + s)r}{ab - r^2}, \text{ and}$$

$$Z_e = \frac{(d + s)a - (c + s)r}{ab - r^2}.$$

This is the desired result, which is equations 5.3 and 5.4 in the main text.

If we substitute these values of $Z_p = Z_{p,m}$ and $Z_e = Z_{e,m}$ into the formula for efficiency and simplify we get

$$\text{Maximum Efficiency} = \frac{Z_{p,m}(c + s) + Z_{e,m}(d + s)}{(c + d + 2s)} .$$

Thus, the maximum efficiency is a weighted average of the two credibilities that produce this maximum. This is similar to Meyers's result in his Appendix B, where the maximum efficiency was equal to the credibility that produces the maximum.

Zero Excess Credibility

Now we will explore the results of using credibilities other than the optimal ones. Let us take a special case when the variance of the excess losses is very large; in other words, $b \rightarrow \infty$. Then

$$Z_p \cong \frac{c + s}{a} \text{, and}$$

$$Z_e \cong 0$$

In fact, if we set $Z_e = 0$, then the formula for efficiency becomes

$$\text{Efficiency} = \frac{2Z_p(c + s) - Z_p^2 a}{c + d + 2s} .$$

For this case, the value of Z_p such that the efficiency is maximized is given by

$$Z_p = \frac{c + s}{a} .$$

This is equation 6.1 in the main text.

The credibility obtained by applying the Bühlmann method to the primary losses alone is c/a. The credibility here is larger (for $s > 0$) since we have taken into account the positive correlation between the primary and excess severities.

When $Z_p = \dfrac{c + s}{a}$ and $Z_e = 0$ we get

$$\text{Efficiency} = \frac{(c + s)^2}{(c + d + 2s)a} .$$

If we let our estimate of the losses be given as in the General Liability Experience Rating Plan, using credibility Z as described by Meyers then

$$F = ZA_p \left(1 + \frac{E_e}{E_p}\right) + (1 - Z)E_p + (1 - Z)E_e.$$

This is a special case of the previous case, with

$$Z_p = Z\left(1 + \frac{E_e}{E_p}\right) .$$

Thus, we know the maximum efficiency occurs when

$$Z \left(1 + \frac{E_e}{E_p}\right) = Z_p = \frac{c + s}{a} \text{, or}$$

$$Z = \left(\frac{c + s}{a}\right) \left(\frac{E_p}{E_p + E_e}\right) .$$

This is equation 6.2 in the main text.

Meyers's Case

Meyers instead uses $Z = c/a$. Putting the corresponding value of Z_p into the formula for efficiency gives, after simplifying terms

$$\text{Meyers's Efficiency} = \frac{(s + c)^2 - \left(s - c\,\dfrac{E_e}{E_p}\right)^2}{(c + d + 2s)\, a} .$$

Comparing this efficiency to that obtained when $Z = \left(\dfrac{c + s}{a}\right) \left(\dfrac{E_p}{E_p + E_e}\right)$,

we get:

$$\frac{\text{Meyers's Efficiency}}{\text{Maximum Efficiency}} = 1 - \frac{\left(s - c\dfrac{E_e}{E_p}\right)^2}{(s + c)^2} .$$

So, using the usual value of Z in accordance with Meyers leads to a loss of efficiency. However, the loss will be small whenever the second term is small. This is the case for all the examples tested in Meyers's Tables 5.2, 5.3, and 5.4.

Credibility Equal to One

Another useful special case is when in formula 6.3 in the main text for large N we take $Z = 1$ and thus

$$Z_p = 1 + \frac{E_e}{E_p}, \text{ and } Z_e = 0.$$

Then the efficiency is given by

$$\text{Efficiency} = \frac{2\left(1 + \frac{E_e}{E_p}\right)(c + s) - \left(1 + \frac{E_e}{E_p}\right)^2 a}{c + d + 2s}.$$

Using the notation of Appendix G, this becomes

$$\text{Efficiency} = 1 - \frac{\hat{c}\,\frac{E_e^2}{E_p^2} + \hat{d} + \frac{\hat{t}}{N}\left(1 + \frac{E_e}{E_p}\right)^2 - 2\,\frac{E_e}{E_p}\,\hat{s}}{\hat{c} + \hat{d} + 2\hat{s}},$$

which has a limit as N gets large of:

$$\text{Efficiency} = 1 - \frac{\hat{c}\,\frac{E_e^2}{E_p^2} + \hat{d} - 2\,\frac{E_e}{E_p}\,\hat{s}}{\hat{c} + \hat{d} + 2\hat{s}}.$$

Thus using formula 6.3 for a fixed loss limit, the maximum efficiency is less than 100%, and we expect to get relatively little improvement in efficiency beyond the point where $Z = 1$. This is the behavior observed in Table 7.3.

APPENDIX G

DEPENDENCE ON SIZE OF RISK

In this appendix, the variation of credibility with size of risk is explored for the cases examined in Meyers's Sections 5 and 6. Also tables of the specific values of the parameters entering the credibility formulas are given for two specific cases from Meyers's Sections 5 and 6. Finally, the general behavior of $N(1 - Z)/Z$ with size of risk is examined. In the examples in Meyers's Sections 5 and 6 there is no parameter uncertainty, the individual risks are homogeneous, and the parameters for an individual do not change over time. Then, as discussed by Meyers in his Section 3, the factors that go into formulas 5.3 and 5.4 for Z_p and Z_e vary as follows with the size of risk N.

Increase as N	Increase as N^2
$a - c = t =$ process variance of primary losses	c
$b - d = u =$ process variance of excess losses	d
$r - s = v =$ process covariance of primary and excess losses	s

$$\text{Let } \hat{t} = \frac{a - c}{N} = \frac{t}{N}$$

$$\hat{c} = \frac{c}{N^2}$$

with similar definitions, for the other quantities, such that we obtain new quantities which are independent of N.

Then substituting into formula 5.3 we get:

$$Z_p = N \frac{N(\hat{c}\hat{d} - \hat{s}^2) + \hat{c}\hat{u} + \hat{s}\hat{u} - \hat{d}\hat{v} - \hat{s}\hat{v}}{N^2(\hat{c}\hat{d} - \hat{s}^2) + N(\hat{c}\hat{u} + \hat{d}\hat{t} - 2\hat{s}\hat{v}) + \hat{u}\hat{t} - \hat{v}^2}$$

Substituting into formula 5.4 would give a similar complicated formula for Z_e. If we set the primary losses equal to the total losses, and set the excess losses equal to zero, then

$$Z_p = \frac{c}{a} = \frac{N}{N + \hat{t}/\hat{c}},$$

which is the familiar expression for credibility, given in Meyers's formula 3.2. However, we notice that in the more general expression, we do not have the familiar simple function of N. Instead we have:

$$Z_p = N \frac{N + K_3}{N^2 + NK_1 + K_2}$$

$$\text{where } K_1 = \frac{\hat{c}\hat{u} + \hat{d}\hat{t} - 2\hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

$$K_2 = \frac{\hat{t}\hat{u} - \hat{v}^2}{\hat{c}\hat{d} - \hat{s}^2}$$

$$K_3 = \frac{\hat{c}\hat{u} + \hat{s}\hat{u} - \hat{d}\hat{v} - \hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

Similarly

$$Z_e = N \frac{N + K_4}{N^2 + NK_1 + K_2}$$

where

$$K_4 = \frac{\hat{d}\hat{t} + \hat{s}\hat{t} - \hat{c}\hat{v} - \hat{s}\hat{v}}{\hat{c}\hat{d} - \hat{s}^2}$$

For large N, if $K_1 > K_3$ we have $Z_p < 1$, but if $K_3 > K_1$, we have $Z_p > 1$.[77] In the latter case, it makes sense to refer to risks such that $Z_p \geqq 1$ as self-rated. Notice, the difference from the usual result, formula 2.1 in the main text, where Z remains strictly less than one, but gets so close to one so as to make no practical difference in the resulting efficiencies. As $N \rightarrow \infty$ we do have $Z_p \rightarrow 1$ and $Z_e \rightarrow 1$.[78]

We can write formula 5.5 for the maximum efficiency as:

$$\text{Maximum Efficiency} = \frac{Z_p(\hat{c} + \hat{s}) + Z_e(\hat{d} + \hat{s})}{\hat{c} + \hat{d} + 2\hat{s}}$$

---

[77] For the example as per Meyers's Table 5.4, the table below uses a loss limit of 2500, K_1 = 8,308, K_2 = 110,847, K_3 = 19,119 and K_4 = 76. For the example as per Meyers's Section 6, the table below uses B = 2000, K_1 = 169, K_2 = 461, K_3 = 407, and K_4 = −42.

[78] Thus, this is a different phenomenon than discussed in Meyers's Section 3, where $N \rightarrow \infty$, $Z \rightarrow 1/J \leqq 1$.

The dependence of the maximum efficiency on N is solely contained in Z_p and Z_e themselves. The weights are independent of N. As $N \to \infty$, we have $Z_p \to 1$, $Z_e \to 1$, Maximum Efficiency $\to 1$.

EXAMPLE AS PER MEYERS'S TABLE 5.4

Loss Limit ($000's)	$\hat{c}$	$\hat{d}$	$\hat{r}$	$\hat{s}$	$\hat{t}$	$\hat{u}$	$\hat{v}$	$N = 1$ E_p	E_e
1	.013	.129	.338	.040	.215	9.303	.299	.350	.520
2	.033	.084	.626	.052	.441	8.525	.574	.463	.407
2.5	.041	.072	.752	.054	.548	8.171	.698	.495	.375
3	.050	.061	.848	.055	.684	7.846	.793	.528	.343
4	.064	.047	1.024	.055	.930	7.247	.970	.572	.299
6	.087	.031	1.287	.051	1.418	6.225	1.236	.631	.240
8	.104	.022	1.470	.047	1.897	5.374	1.422	.670	.200
12	.130	.012	1.678	.039	2.817	4.020	1.639	.722	.149
16	.148	.007	1.749	.033	3.694	2.987	1.717	.756	.115

EXAMPLE AS PER MEYERS'S SECTION 6[79]

B ($000's)	$\hat{c}$	$\hat{d}$	$\hat{r}$	$\hat{s}$	$\hat{t}$	$\hat{u}$	$\hat{v}$	$N = 1$ E_p	E_e
.1	50	2550	784	317	41	35110	467	1.65	9.38
.2	92	2305	1278	419	105	34263	859	2.23	9.26
.5	209	1864	2394	581	365	32094	1813	3.30	8.19
1	380	1454	3698	700	901	29190	2998	4.37	7.12
2	660	1019	5373	778	2101	24796	4595	5.65	5.84
5	1234	499	7580	751	5751	16676	6829	7.53	3.95
10	1803	221	8218	605	11057	9803	7612	8.93	2.56

[79] Meyers's example corresponds to $N = 10$, B = \$2000, except that here a claim limitation of \$100,000 has been used, and the Weibull distribution has been approximated by a discrete analog. See Section 10 of this discussion.

When doing empirical studies it is sometimes useful to focus on the following quantity as a function of size of risk, rather than the credibilities themselves. Define $\kappa = N(1 - Z)/Z$. Then for the usual simple case, formula 2.1 in the main text, $\kappa = K$. With parameter uncertainty, formula 2.2 in the main text, $\kappa = (J - 1)N + K$. For the split plan case handled previously in this appendix, $\kappa_p = [N(K_1 - K_3) + K_2]/(N + K_3)$. One can take other cases and derive the expected behavior of κ as a function of N, the size of risk. For some cases, this requires combining the results of this appendix with those in Appendices B, D, and F. Note that for the case of a split plan, the same general functional form applies to κ_p and κ_e, even though the specific coefficients may make the specific curves look very different. The following table presents the results for the different cases.

$$\kappa = \frac{N(1 - Z)}{Z} \text{ as a function of } N \text{, the size of risk}$$

	No-Split Plan	Split Plan (Primary vs. Excess)
No Parameter Uncertainty No Risk Heterogeneity	Constant	$\frac{\text{Linear}}{\text{Linear}}$
With Parameter Uncertainty No Risk Heterogeneity	Linear	$\frac{\text{Quadratic}}{\text{Linear}}$
No Parameter Uncertainty With Risk Heterogeneity	$\frac{N}{\text{Linear}}$	$N\frac{\text{Linear}}{\text{Quadratic}}$
With Parameter Uncertainty With Risk Heterogeneity	$N\frac{\text{Linear}}{\text{Linear}}$	$N\frac{\text{Quadratic}}{\text{Quadratic}}$

APPENDIX H

MAXIMUM EFFICIENCY AS A FUNCTION OF LOSS LIMIT AND SIZE OF RISK

In this appendix, the behavior of the maximum efficiency with changing loss limit and size of risk is explored. The observed behavior for the examples in Meyers's Sections 5 and 6 is explained in terms of the underlying mathematics and the specific choices of parameters for those examples. Which example is a better approximation to a real world application of experience rating will determine whether the loss limit used should increase significantly for large risks.

Using the credibilities from formula 6.2 in the main text, and the notation of Appendix G, the maximum efficiency is given by

$$\text{Maximum Efficiency} = \frac{(c+s)^2}{(c+d+2s)a} = \frac{(\hat{c}+\hat{s})^2 N^4}{(\hat{c}+\hat{d}+2\hat{s})N^2(\hat{t}N+\hat{c}N^2)}$$

$$= \left[\frac{(\hat{c}+\hat{s})^2}{\hat{c}\,(\hat{c}+\hat{d}+2\hat{s})}\right] \frac{N}{N+\dfrac{\hat{t}}{\hat{c}}}\,.$$

The first term is independent of N.[80] It can be rewritten as

$$\left[\frac{(\hat{c}+\hat{s})^2}{\hat{c}(\hat{c}+\hat{d}+2\hat{s})}\right] = 1 - \frac{\hat{c}\hat{d}-\hat{s}^2}{\hat{c}(\hat{c}+\hat{d}+2\hat{s})}$$

We expect $\hat{s}^2$ to be close to $\hat{c}\hat{d}$, since we expect the hypothetical mean primary losses to be highly correlated with the hypothetical mean excess losses. (If they were perfectly correlated then $\hat{s}^2 = \hat{c}\hat{d}$.) Thus we expect the first term in the expression for the maximum efficiency to be close to one. In fact, for the examples here, that is the case, as shown in the table below.

The second term in the expression for the maximum efficiency is the Bühlmann credibility used by Meyers. It varies with the loss limit only in so far as $K = \hat{t}/\hat{c}$ does. The smaller K, the larger the second term. However, in general, we do not expect the second term to be extremely sensitive to K.

Thus neither term is expected to be extremely sensitive to the loss limit chosen. In fact, the observed optimal efficiencies for fixed N are relatively insensitive to the loss limit.

[80] In fact, $\hat{c} + \hat{d} + 2\hat{s}$, which is part of the denominator, is the variation of the hypothetical means of the total losses, and is thus independent of the loss limit chosen.

For the examples in Meyers's Table 5.4, the variation of the second term with loss limit is more important than that of the first term. Thus, selecting the smallest K will produce the largest efficiency regardless of N. As shown in the table below, a loss limit of 2.5 gives the smallest K. As was seen in Table 7.2 the largest efficiency was indeed obtained by taking a loss limit of 2.5 regardless of N, the size of the risk.

In the example in Meyers's Section 6, the first term increases significantly with the loss limit. Thus in order to get the largest efficiency, we have a conflict between choosing a smaller value of K (low loss limit) and a larger value of the first term (high loss limit). For larger N, the second term depends less on K, thus the first term is relatively more important. Thus we expect the optimal loss limit to increase significantly with size of risk. This was indeed the behavior observed in Table 10.1.

EXAMPLE AS PER MEYERS'S TABLE 5.4

Limit (\$000's)	$\dfrac{(\hat{c} + \hat{s})^2}{\hat{c}\,(\hat{c} + \hat{d} + 2\hat{s})}$	$K = \dfrac{\hat{t}}{\hat{c}}$
1	.988	17.2
2	.994	13.6
2.5	.995	13.5
3	.996	13.7
4	.997	14.5
6	.999	16.3
8	.999	18.2
12	1.000	21.7
16	1.000	24.9

EXAMPLE AS PER MEYERS'S SECTION 6[81]

B ($000's)	$\frac{(\hat{c} + \hat{s})^2}{\hat{c}\,(\hat{c} + \hat{d} + 2\hat{s})}$	$K = \frac{\hat{t}}{\hat{c}}$
.1	.834	.82
.2	.876	1.14
.5	.922	1.74
1	.949	2.37
2	.969	3.19
5	.987	4.66
10	.995	6.13

[81] Meyers's example corresponds to $N = 10$, B = $2000, except that here a claim limitation of $100,000 has been used, and the Weibull distribution has been approximated by a discrete analog. See Section 10 of this discussion.

APPENDIX I

CALCULATION OF THE QUANTITIES ENTERING THE CREDIBILITY FORMULA

In this appendix, expressions for the parameters entering the formulas for credibility will be derived. The results are summarized at the end of the appendix. We will assume that the frequency and severity are independent and that the individual claims are taken independently from a size of loss distribution.[82]

Let ϕ be a set of parameter(s) that describes the claims process.[83]

We assume that ϕ may take on different values, with probability density function $f(\phi)$; f is commonly referred to as the mixing distribution.

Let θ be the parameter(s) which specify the severity distribution. Assume θ takes on different values, with probability density function $g(\theta)$. Similarly, let ψ be the parameter(s) which specify the frequency distribution. Assume ψ takes on different values, with probability density function $h(\psi)$.

Since frequency and severity have been assumed to be independent we have:

$$f(\phi) = g(\theta)\, h(\psi)$$

Let

E_p = expected value of the primary losses taken over all values of ϕ.
$E_p(\phi)$ = expected value of the primary losses given a specific set of parameters ϕ.

Use a similar definition of the corresponding symbols for excess losses.

Definitions

Define the following quantities, which will be useful:

$\bar{n}$ = average frequency
m_p = average primary portion of a claim
$= E_p \div \bar{n}$ = primary severity

---

[82] This assumption is made by Meyers in his Algorithm 6.1. For real risks, this may not be true. In Meyers's Section 5, although for a given risk the frequency and severity are independent, the frequency and severity between risks are not independent, thus the formulas in this appendix do not all apply to that situation.

[83] Following Meyers's Algorithm 3.2, ϕ is assumed constant over time for an individual risk, but may be different for different risks. In fact, for real risks ϕ varies over time, as noted by Meyers in his discussion of the Bailey and Simon results on the credibility of a single private passenger car.

m_e = average excess portion of a claim
= $E_e \div \bar{n}$ = excess severity
α_f = parameter variance of the frequency
α_p = parameter variance of the primary severity
= variance of the hypothetical mean primary severities
α_e = parameter variance of the excess severity
β_f = expected value of the process variance of the frequency
β_p = expected value of the process variance of the primary severity
β_e = expected value of the process variance of the excess severity
γ = covariance of hypothetical mean primary severity and excess severity
ζ = expected value of the process covariance of the primary severity and excess severity

Also let

$m_p(\theta)$ = expected value of primary severity for a specific set of parameters θ

$= E_p(\theta) \div \bar{n}$
$m_e(\theta) = E_e(\theta) \div \bar{n}$

Derivation of Results

Let c = variance of the hypothetical means of the primary losses.

$$\begin{aligned} &= \int E_p^2(\phi)\, f(\phi)\, d\phi - E_p^2 \\ &= \int \bar{n}^2\, (\psi)\, h(\psi)\, d\psi \int m_p^2(\theta)\, g(\theta)\, d\theta - \bar{n}^2 m_p^2 \\ &= (\alpha_f + \bar{n}^2)\, (\alpha_p + m_p^2) - \bar{n}^2 m_p^2 \\ &= \alpha_p\alpha_f + \alpha_p\bar{n}^2 + \alpha_f m_p^2 \end{aligned}$$

Similarly, let d = variance of the hypothetical means of the excess losses.

$$d = \alpha_e\alpha_f + \alpha_e\bar{n}^2 + \alpha_f m_e^2$$

Let a = total variance of the primary losses. By a well-known result used by Meyers,[84] total variance equals parameter variance plus expected value of the process variance.

Thus

$a = c$ + expected value of the process variance of the primary losses.

[84] This result is a special case of a result derived for covariances in Appendix A of this discussion.

The process variance of the primary losses can be put in terms of the frequency and severity in the usual manner.[85]

$$a = c + E_\theta[m_p^2(\theta) \text{ (process variance of frequency) } + \bar{n}(\theta) \text{ (process variance of primary severity)}]$$

$$= c + E_\theta[m_p^2(\theta)\, \overline{\beta}_f + n\beta_p]$$
$$= c + (\alpha_p + m_p^2)\, \beta_f + \bar{n}\beta_p$$
$$= (\alpha_p + m_p^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_p + \bar{n}\beta_p$$

Similarly, let b = total variance of the excess losses.

$$b = (\alpha_e + m_e^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_e + \bar{n}\beta_e$$

Let s = covariance of the hypothetical means of the primary and excess losses

$$s = \int E_p(\phi)\, E_f(\phi)\, f\phi\, d\phi - E_pE_e$$
$$= \int m_p(\theta)\, m_e(\theta)\, g(\phi)\, d\phi \int \bar{n}^2(\psi)\, h(\psi)\, d\psi - \bar{n}^2 m_p \mathrm{m}_e$$
$$= (\gamma + m_pm_e)(\alpha_f + \bar{n}^2) - \bar{n}^2 m_pm_e$$
$$= \gamma\alpha_f + \gamma\bar{n}^2 + \alpha_f m_pm_e$$

Let r = total covariance of the primary and excess losses

The total covariance can be split into two pieces in a manner similar to that for the variance.[86]

Total covariance = parameter covariance + expected value of the process covariance.

Thus, $r = s$ + expected value of the process covariance.

The process covariance can be written in terms of the frequency and severity in a manner similar to the usual formula for the process variance.[86] Given a set of parameters ϕ:

process covariance of the primary and excess losses = (process covariance of the frequency) (mean primary severity) (mean excess severity) + (mean frequency) (process covariance of the primary and excess severity)

$$= \mathrm{VAR}[n|\psi]\; m_p(\theta)\; m_e(\theta) + \bar{n}(\psi)\; \zeta(\theta)$$

---

[85] This result is also a special case of a result derived for covariances in Appendix A.

[86] This result is derived in Appendix A.

Taking the expected value over all values of the parameters gives the expected value of the process covariance of the primary and excess losses equal to

$$\beta_f(\gamma + m_p m_e) + \bar{n}\zeta.$$

Thus

$$r = s + \beta_f(\gamma + m_p m_e) + \bar{n}\zeta$$
$$r = (\alpha_f + \beta_f)(\gamma + m_p m_e) + \bar{n}^2\gamma + \bar{n}\zeta$$

For the special case of a single split plan[87] ζ has a relatively simple form.

Let the probability density function of the severity be $\pi(x;\theta)$. For a fixed value of θ, the process covariance is

$$\begin{aligned}\zeta(\theta) &= \textstyle\int_0^\infty \text{MIN}[x,L]\ \text{MAX}[0,x - L]\ \pi(x;\theta)\ dx - m_p(\theta)m_e(\theta)\\ &= \textstyle\int_0^L 0\ dx + \int_L^\infty L(x - L)\ \pi(x;\theta)\ dx - m_p(\theta)m_e(\theta)\\ &= Lm_e(\theta) - m_p(\theta)m_e(\theta)\end{aligned}$$

Thus, the expected value of the process covariance is:

$$\zeta = Lm_e - (\gamma + m_p m_e)$$
$$= m_e(L - m_p) - \gamma.$$

Thus, for a single split plan we have

$$r = (\alpha_f + \beta_f)(\gamma + m_p m_e) + \bar{n}^2\delta + \bar{n}(m_e(L - m_p) - \gamma).$$

Summary of Results

$$\begin{aligned}a &= (\alpha_p + m_p^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_p + \bar{n}\beta_p\\ b &= (\alpha_e + m_e^2)(\alpha_f + \beta_f) + \bar{n}^2\alpha_e + \bar{n}\beta_e\\ c &= \alpha_p\alpha_f + \alpha_p\bar{n}^2 + \alpha_f m_p^2\\ d &= \alpha_e\alpha_f + \alpha_e\bar{n}^2 + \alpha_f m_e^2\\ r &= (\alpha_f + \beta_f)(\alpha + m_p m_e) + \bar{n}^2\gamma + \bar{n}\zeta\\ s &= \gamma\alpha_f + \gamma\bar{n}^2 + \alpha_f m_p m_e\end{aligned}$$

[87] Those dollars below L are primary; those above are excess. The current Workers' Compensation Experience Rating Plan is a multi-split plan. The difference is discussed in Snader [9].

APPENDIX J

GAMMA-POISSON, GAMMA-EXPONENTIAL PROCESS

In this appendix, the results of Appendix I are carried forward for a specific choice of distributions. In addition, the resulting efficiencies are shown for a specific choice of parameters.

We assume a single split rating plan with a loss limit of L. We assume no overall limitation on claims.[88] The frequency is given by a Gamma-Poisson process.[89] The frequency for an individual risk is Poisson, with parameter ψ:

$$\omega(n;\psi) = e^{-\psi}\frac{\psi^n}{n!}; \qquad \text{mean } \psi, \text{ variance } \psi.$$

In turn, the parameter ψ has a mixing distribution which is Gamma, with parameters of η and ϵ:

$$h(\psi) = \frac{\epsilon^\eta \psi^{\eta-1}\, e^{-\epsilon\psi}}{\Gamma(\eta)}; \qquad \text{mean } \frac{\eta}{\epsilon}, \text{ variance } \frac{\eta}{\epsilon^2}.$$

The severity is given by a Gamma-Exponential process.[90] The severity for an individual risk is exponential, with parameter θ:

$$\pi(x;\theta) = \theta e^{-\theta x}; \qquad \text{mean } 1/\theta, \text{ variance } 1/\theta^2.$$

In turn, the parameter θ has a mixing distribution which is Gamma, with parameters ξ and ν:

$$g(\theta) = \frac{\nu^\xi}{\Gamma(\xi)} e^{-\nu\theta}\theta^{\xi-1}; \qquad \text{mean } \frac{\xi}{\nu}, \text{ variance } \frac{\xi}{\nu^2}.$$

We assume $\xi > 2$ so that the resulting Pareto has finite variance. If $2 > \xi > 1$ the means are finite, but the excess variances are infinite, and the following formulas are not valid.[91]

---

[88] The mathematics are only slightly more complicated for an overall limitation. However, an overall limitation applied in both Sections 7 and 10 of this discussion.

[89] The resulting overall frequency distribution is a negative binomial.

[90] The resulting overall severity distribution is a Pareto. The exponential distribution is a special case of the Gamma. The more general Gamma-Gamma process results in a Generalized Pareto Distribution; see Hogg and Klugman [10]. While the Gamma-Gamma is probably a better model of reality, the mathematics here would be much more complicated.

[91] With an overall limitation on losses, the variances would be finite.

Quantities Entering the Credibility Formulas

$\bar{n} = \eta/\epsilon$

β_f = expected value of the process variance of the frequency
$= \bar{n} = \eta/\epsilon$

α_f = variance of the hypothetical mean frequencies
$= \eta/\epsilon^2$

$$m_p(\theta) = \frac{1}{\theta} - \frac{e^{-\theta L}}{\theta}$$

$$m_e(\theta) = \frac{e^{-\theta L}}{\theta}$$

$$m_p = \frac{\nu}{\xi - 1} - \frac{\nu}{\xi - 1}\left(1 + \frac{L}{\nu}\right)^{1-\xi}$$

$$m_e = \frac{\nu}{(\xi - 1)}\left(1 + \frac{L}{\nu}\right)^{1-\xi}$$

α_p = variance of the hypothetical mean primary severity

$$= \frac{\nu^2}{(\xi - 1)(\xi - 2)} - \frac{2\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} + \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_p^2$$

α_e = variance of the hypothetical mean excess severity

$$= \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_e^2$$

γ = covariance of the hypothetical mean primary and excess severity

$$= \frac{\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - m_p m_e$$

Process Variance of the Primary Severity (θ) =

$$\frac{1}{\theta^2} - \frac{2Le^{-\theta L}}{\theta} - \frac{e^{-\theta L}}{\theta^2}$$

Process Variance of the Excess Severity (θ) =

$$\frac{2e^{-\theta L}}{\theta^2} - \frac{e^{-2\theta L}}{\theta^2}$$

Process Covariance of the Primary and Excess Severity (θ) =

$$\frac{Le^{-\theta L}}{\theta} + \frac{e^{-2\theta L}}{\theta^2} - \frac{e^{-\theta L}}{\theta^2}$$

β_p = Expected Value of the Process Variance of the Primary Severity

$$= \frac{\nu^2}{(\xi - 1)(\xi - 2)} - \frac{2L\nu(1 + L/\nu)^{1-\xi}}{(\xi - 1)} - \frac{\nu^2(1 + 2L)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

β_e = Expected Value of the Process Variance of the Excess Severity

$$= \frac{2\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

ζ = Expected Value of the Process Variance of the Excess Severity

$$= \frac{L\nu(1 + L/\nu)^{1-\xi}}{(\xi - 1)} + \frac{\nu^2(1 + 2L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)} - \frac{\nu^2(1 + L/\nu)^{2-\xi}}{(\xi - 1)(\xi - 2)}$$

A Specific Example

In the Gamma distribution the first parameter controls the shape[92], while the second parameter basically determines the scale once the first parameter is chosen.

The mixing distribution of the frequency is Gamma, with parameters η and ϵ. Let $\eta = 4$ and $\epsilon = 4/N$. Thus since $\bar{n} = \eta/\epsilon$, N is the mean number of claims.

The mixing distribution of the severity is Gamma, with parameters ξ and ν. Let $\xi = 2.5$ and $\nu = 4500$. Thus, since $m = m_p + m_e = \nu/(\xi - 1)$, the average size of claim is $4500/1.5 = 3000$.

Then the resulting efficiencies are as follows:

[92] For parameters η and ϵ, the skewness of the Gamma is two over the square root of η. For η = 1, the Gamma is the exponential distribution. For η large, the Gamma approaches the normal distribution.

EFFICIENCIES

L	$N = 1$	$N = 10$	$N = 100$
0	31.43%	82.09%	97.86%
100	32.11	82.25	97.87
1000	32.01	82.23	97.87
10000	31.67	82.14	97.87
∞	31.43	82.09	97.86

We note that for this example, the efficiencies are almost independent of the loss limit L. In fact, the single split plan has no practical advantage over the no split plan ($L = 0$ or $L = \infty$).

DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXIII

THE COST OF MIXING REINSURANCE

RONALD F. WISER
VOLUME LXXIII

DISCUSSION BY NEAL J. SCHMIDT

Most insurers have complex reinsurance arrangements. Ron Wiser's paper examines the effect of mixing proportional facultative placements with excess of loss treaties. This discussion will review the mixing problem from the treaty reinsurer's perspective. It will be shown that the mixing of facultative and treaty is a specific example of a more general problem underlying all treaty reinsurance. Potential pitfalls in the author's proposals will be discussed and alternative suggestions presented.

We often see excess of loss protection of proportional treaties where the excess treaty is applied before the proportional treaty. Facultative transactions, whether excess or proportional, usually inure to the benefit of treaties and therefore are applied first. The order of application results in the mixing problem that, if unanticipated, can undermine an otherwise well planned reinsurance program.

The main thesis of the paper is well illustrated by specific examples, intuitive argument and mathematical proofs. The underwriter who places proportional facultative reinsurance without regard to the effect of any applicable excess protection will be unpleasantly surprised by the effect the placement will have on the net position. Any placement of proportional reinsurance that inures to the benefit of any existing excess treaty will penalize net results. This follows from the general principle that any action taken by the ceding company that reduces the exposure to the excess treaty by a greater proportion than it reduces premium ceded to the treaty will adversely affect net results. Given the shape of the common size of loss distributions, the effect can be considerable.

The reader must be careful not to conclude that proportional facultative reinsurance should never be considered when an excess of loss treaty is in place. The paper states ". . . the net position after mixed reinsurance will always be worse than under a pure excess reinsurance . . ." An illustrative example and a proof are offered. If we examine the structure of the example, the nature of the results becomes clearer.

The price of an excess of loss treaty is usually determined before the subject business is written and any facultative protection is placed. As the example is constructed, the excess treaty price is fixed and the decision to place proportional facultative reinsurance is variable. The implication is that the excess of loss treaty is priced without any consideration of the benefit of inuring reinsurance. In this case, it follows that the excess treaty is overpriced and net results will suffer. But, as the author points out in his conclusion, excess of loss reinsurance treaties are priced anticipating a certain part of the book will be ceded proportionally before the treaty applies. This is an essential part of treaty negotiations.

If an appropriate credit is applied in the treaty pricing, then the aggregate net position after mixed reinsurance will be no different than under a pure excess situation. It is important to include the phrase "aggregate net position," because the excess treaty is priced to be appropriate for the sum of all subject policies. The excess treaty rate will only coincidentally be appropriate for any individual policy.

The example very effectively demonstrates that treaty pricing should reflect anticipated inuring proportional cessions. This is a point well worth making. It is just as important to emphasize that this is not an appropriate model for individual risk underwriting. There are shortcomings to using this procedure to analyze the marginal effect on results of writing a particular primary policy. Treaty pricing is an aggregate concept and it is inappropriate to include its effects in underwriting individual primary risks.

In opposition to the chronological order of placement, it might be more appropriate to treat proportional facultative placements as fixed and the treaty price as variable. Due to practical considerations, the treaty is placed first. However, the anticipated effect of facultative placements is reflected in the treaty rate and should be considered a prerequisite for treaty placement. Once a treaty is in place it seems inappropriate to change company policy on inuring facultative reinsurance.

The procedure used in the paper allows for the decision on whether to place facultative reinsurance without regard for the long term effects on the treaty. Extensive implementation of this type of analysis could lead to a change in the practice of purchasing facultative protection. Treaty reinsurers would then rightfully charge the cedant with selecting against the treaty. This would surely lead to a breakdown in the relationship required for an effective treaty.

Generalization of the Mixing Problem

The particular mixing situation examined in the paper is one aspect of a more general problem existing in all excess treaty reinsurance. A group of risks are reinsured under one agreement, often with one rate. There are obvious expense benefits to treaty reinsurance over facultative. The cost for these expense savings is a decrease in pricing accuracy.

Determining a single rate for a treaty is difficult because the risks reinsured may vary by line of business, state, classification, policy limit, deductible, inuring facultative reinsurance, etc. It is made even more difficult by the changes in distribution from year to year. The mixing of proportional reinsurance placements is one of many possible distribution changes that must be anticipated to properly price a treaty.

If we look at the treaty rate as a weighted average of correct rates for each anticipated exposure, we can see why it is inappropriate to apply an overall treaty rate to an underwriting analysis of an individual risk. To do so would lead us to write only the most hazardous risks due to the favorable treaty rate for their exposure.

If, as suggested earlier, the example was constructed assuming a fixed distribution of business ceded to the treaty, the results would be consistent with our expectations. It is unanticipated distribution changes in the subject business that affect the adequacy of the treaty rate and, therefore, the net position.

In the example in the paper, no credit was given in the excess treaty rate for proportional facultative placements. Reducing the exposure to the treaty through use of inuring facultative reinsurance benefits the treaty and penalizes net results. This is an unanticipated change in the distribution of the subject business that benefits reinsurers.

It is easy to present an alternative example in which exposure to the treaty is greater than originally anticipated and net results benefit. Assume the cedant's subject book of business is comprised of two lines. The low hazard line has an anticipated subject premium of $30 million. The high hazard line has $45 million in anticipated subject premium. It is determined that the exposure for each line requires a rate of 1% for the low hazard line and 50% for the high hazard line. A weighted average determines an overall rate of 30.4% and a reinsurance premium of $22.8 million. If the actual distribution of business at year end is $25 million in low hazard subject premium and $50 million in high hazard subject premium, the exposure rate is 33.67% producing $25.3 million in required reinsurance premium. The actual premium is $2.5 million less. This may seem an extreme example, but it closely parallels an actual treaty in place during 1985. In practice, changes in distribution can be much greater than shown in this example.

The author is admittedly presenting the ceding company's point of view. His paper focuses on one example that has detrimental effects on the cedant. The reinsurer may be more inclined to believe that, due to the much greater control of the business exercised by the ceding company, distribution changes to the detriment of reinsurers are more prevalent. Whether or not either view reflects reality, the opportunity to abuse a treaty is always present. It is essential to the continuity of a good reinsurance relationship that every effort be made to avoid such a possibility.

In order to minimize the negative effect of distribution changes, both parties must strive to develop an understanding of the book of business and the purpose of treaty reinsurance. Assumptions that underlie treaty pricing must be conveyed from management to individual underwriters and effectively implemented.

In contrast to one of the paper's contentions, it is imperative that the underwriter and actuary involved with pricing individual risks be concerned only with producing rates geared to a profitable direct premium. The effect of ceded reinsurance on net results is the domain of the ceded reinsurance manager, and actuarial involvement should address proper pricing for the aggregate exposure. The direct pricing process should take place without recognition of the treaty but within management guidelines. These guidelines should reflect the understandings that were reached with reinsurers and upon which the rate is based. In this way both parties' concerns are addressed. The cedant is satisfied that the rate reflects any exposure-reducing actions on its part. The reinsurer is satisfied that the danger of selecting against the treaty is reduced.

Conclusion

Proportional facultative reinsurance can have a considerable effect on the exposure ceded to an excess of loss treaty. In order to insure a fair price on an excess treaty, it is necessary to reduce the rate in recognition of anticipated facultative placements. However, incorporating net considerations into the direct pricing system is inappropriate and likely to lead to abuse of the treaty. These issues should be dealt with at the management level to insure compliance with the intent of the treaty.

In practice, we are likely to find that many companies do not effectively communicate the essentials of the treaty agreement to the primary underwriters responsible for individual accounts. Insurers and reinsurers will undoubtedly endorse the author's call for an improvement to this vital link as necessary to maintain stable reinsurance relationships.

I commend Ron Wiser on a very well organized and understandable paper on a difficult subject. It presents many insightful ideas and should stimulate more actuaries to examine the complexities of the many different reinsurance arrangements found in most companies.

ADDRESS TO NEW MEMBERS—MAY 11, 1987

RONALD L. BORNHUETTER

Although my brief remarks this morning are directed to the new Fellows and Associates we honor today, I do hope there will be something for the rest of the audience too.

Each of us here in this room knows how hard he or she has worked to attain Associateship or Fellowship. It is a road we have all followed and there will be many future rewards for each of you in the years ahead. I only cite the recent emergence of casualty actuaries serving as senior management officials in many property and casualty companies as an example.

Your focus for these past several years while you have been taking examinations, has, by necessity, been inward in scope. The CAS has been all consuming for you and, therefore, I would like to briefly turn your attention outward to the actuarial world in which we live and the one which you are entering.

With the addition of this Associate class, the CAS membership now exceeds 1,300, while five years ago the number was just over 900, and ten years ago the CAS was half of today's size—certainly a growing, vibrant organization of which we are all proud to be a part.

I would remind you that there are others here in North America who also call themselves actuaries and some are even beyond these borders. A brief overview may be of enlightenment to you.

First, there is the Society of Actuaries, the other learned actuarial organization and our friendly rival. They number over 10,000 strong, which makes eight-plus life actuaries for every casualty actuary—which is about an equal match. Seriously, though, the Society of Actuaries has many common threads with the CAS and, as you can tell from the President's column in the latest issue of the *Actuarial Review*, we will again be looking to see if we can bring the two organizations closer together. One last point on the Society of Actuaries—only five examinations are needed for membership. I congratulate each of you for taking the longer route to membership.

The two organizations in which you should become involved, or at least follow their activities closely, are the two national organizations—the American Academy of Actuaries, with some 8,500 members, and the Canadian Institute of Actuaries, with its 1,300 members. Each is comprised of life, health, pension, and casualty actuaries. I could spend some time talking about these organizations, but I leave you with just one thought—whether one agrees with the situation or not, these organizations are the spokespersons for our profession in the public arena, and will continue to be. The only way to ensure the quality of these statements is to be an active member and participant. Please put this on your future agenda.

There are more professionals here in the United States who carry the designation of actuary: the Conference of Actuaries in Public Practice with 900 members, who are consulting actuaries; Enrolled Actuaries, some 4,000 strong, who are licensed pension actuaries; and, ASPA—the American Society of Pension Actuaries. Don't panic, the numbers are not all additive. For example, over 1,000 members of the American Academy are CAS members, which is an extremely high percentage of those U.S. casualty actuaries who are eligible.

For a brief moment, let's become global and go beyond North America. The International Actuarial Association, which meets once every four years at its Congress, has 4,600 members. However, only 1,200 come from the United States and Canada. This is a poor showing, considering that over 10,000 actuaries are eligible. ASTIN is the property/casualty section of the IAA and has 1,400 members, with 500 coming from the two North American countries. This organization meets annually, usually in Europe. These organizations are currently dominated by European actuaries; however, this is slowly changing as foreign members are looking to North American actuaries for leadership in the future.

One interesting piece of trivia concerns one of our sister organizations, the Institute of Actuaries in the United Kingdom, which is very similar to the two learned organizations in North America with examinations, etc. Do you realize that over 25% of their membership does not work in the insurance industry? They are employed in the securities industry—an interesting commentary which does make some sense.

That is a very brief thumbnail sketch of the populace of the actuarial world you are entering. Now I would like to turn to another part of your new world which is a little closer to home. I am referring to the concepts of statements of principles and standards of practice. Each of you must take heed of this activity as principles are the "thou shall" and "thou shall not," and standards are the "how to" of our profession. They are the ground rules of the game. The examinations you have successfully completed have, by necessity, concentrated on casualty actuarial principles, but this is only the beginning. In addition, this activity is very much in the forefront right now. Others in this room should also pay attention to this activity as it emerges toward maturity.

There are two elements at work right now—the CAS is again taking the lead in the profession by currently working on the "thou shall and thou shall not"—the statements of principles on ratemaking and reserving. These will serve as the cornerstone for the development of standards of practice. The development of standards is currently being handled by a relatively new organization—the Interim Actuarial Standards Board (IASB)—a group of nine very senior actuaries coming from all disciplines within the profession. All nine members each have at least twenty years' membership in their Society. All but two have served as president of at least one of the actuarial organizations in the United States. There are two other members of the CAS besides myself serving as members of the Board—Tom Murrin and Jim Hickman.

My message to you today is that the IASB takes its charge very, very seriously and intends to actively pursue the development of standards where they are needed. The Board is waiting for the CAS principles to be finalized and will then proceed to promptly work on the needed standards of practice. We are not talking about trivial concepts now—subjects such as the discounting of loss reserves and its ramifications are very high on the agenda. In fact, work on standards by the IASB Casualty Committee has already begun in order that as little time as possible elapses between the "thou shall and thou shall not" and the "how to's."

Most importantly, we are working swiftly towards dropping the "I" from IASB and having the Board become permanent in 1988. To give you some frame of reference, we expect the Board's annual budget to run in the hundreds of thousands of dollars, which includes a paid staff to work full time on Board activities. With 10,000 U.S. actuaries as the support base, we could be talking about \$20.00 to \$30.00 or more per actuary each year to support the principles/standards movement. Please pay attention and participate if you can. It is your lifeblood.

Ladies and gentlemen, I welcome you to the world of actuaries—it is an exciting, ever-changing, challenging, and very rewarding world and you should feel very proud today to be a part of it. The CAS is only an organization of individuals and it is those individuals—its members—that make it great. May your careers blossom and prosper in the years ahead—the opportunity is there.

Thank you for your kind attention.

KEYNOTE ADDRESS—MAY 11, 1987
SHAPING AMERICA'S ECONOMIC FUTURE

PAT CHOATE

What I'd like to do this morning is three things. The first is to describe a little American economic history. The second is to describe the current economic situation in the United States and what it's likely to be. The third is to talk about politics—where we are, where we seem to be going, and what it's going to mean to you and your profession.

In doing this, I think it's useful for us to recognize that you actuaries take a look at uncertainties of the future. I was very impressed with some of the discussions last night when I got to ask some questions. For example, how do you explain to your wife what you do? Or, how do your children explain to their friends what you do? I got some very detailed and complex answers. What I got out of those discussions is that, by and large, this group, whether they will acknowledge it or not, comprises probably the most sophisticated futurists in America, because in effect what you're doing is taking some slice of the future. You're determining what some of the risks are that are associated with that future, and then you're really putting your names and reputations on the line by putting a dollar value on them. You're actually pricing the future. In a very real sense, you are the most sophisticated futurists in America.

One of the things I would suggest for you to consider for your November meeting this year is to acknowledge that this is also the 100th anniversary of the publication of the most profound book on the future ever published in the United States. It was published in 1887 by Edward Bellamy. It is called *Looking Backward: 2000–1887*. It's instructive for people in your profession and mine. Bellamy had an insight. He wanted to talk about what was happening in this country with the industrial revolution. He wanted to talk about how it was changing people's lives. But he wanted to do it in a manner that could reach large numbers of people.

His insight was that he had his principal character in the novel go to sleep in the year 1887 and awaken in the year 2000. When he awoke, he had perfect recall of what the 19th Century was like, and he could then compare what had happened over the 113 year period of time. What he imagined was absolutely phenomenal. He imagined, first of all, the supermarket. He imagined credit cards. He imagined something comparable to the Sony Walkman, where you could walk into a room and dial up music and listen to the music without being disturbed. He imagined a six-day work week—which was revolutionary. He imagined the end of child labor. He imagined retirement at the age of forty-five—early retirement. He imagined child care services. He imagined women being educated and working on the same basis as men. And he imagined that they would be paid on the same basis as men. Maybe in the year 2000 that will be true. His predictions are not over.

What's really striking about Bellamy's book is that in about a year and a half it sold over 3.5 million copies in the United States alone. Let me put that in perspective: Lee Iacocca's book, which I thought would never get off the *New York Times* Top 10 List, sold 2.5 million copies. What Bellamy was able to do was capture the essence of change occurring in our economy. Today his book is largely forgotten. Jules Verne is remembered, H.G. Wells is remembered, but Edward Bellamy is seldom remembered. One of the reasons Bellamy is not remembered is that his ideas were so powerful they got picked up by the political movement. When you take a look at the New Deal, when you take a look at Woodrow Wilson's New Freedoms, you find that the fiction that Edward Bellamy imagined quickly became the reality of contemporary politics. His issue was less that of precedence and more that of defining a set of political philosophies and ideas that could define our life and times. What I'm suggesting is that as discerning Victorians then, they began to determine what was going to be happening in our future, and they began to create ideas, philosophies, and approaches to deal with the future. We're into that same sort of period. We're in the midst, as they were in 1887, of great underlying, far-reaching changes that are affecting our work, our industries, our businesses. What we're really talking about today is identifying those ideas, those concepts, those frameworks, and those decisions that we could take to shape work and life in this country in the balance of this century and well into the 21st Century.

We are at one of those times that come in American history that can truly be called a hinge of history. It's happened before. It happened with Woodrow Wilson when he became President. It happened during the New Deal. It happened in that time period 1943 through 1947. It happened briefly, and to some small degree, in the period 1981–1982 with the Reagan revolution. What I'm suggesting to you is that we're moving in 1987 into another one of those periods of fundamental political and economic change. It's also going to be one of those periods of change where we're going to have a real generational shift in leadership in this country. Think about it. When Gary Hart left the race he was fifty years old; he was one of the oldest Democratic candidates. Only Paul Simon was older at the age of fifty-eight. All of the Democratic candidates for President are under the age of fifty with the exception of Simon. The youngest is thirty-nine—Albert Gore. When you take a look at the Republicans, you're seeing people such as Jack Kemp and others now moving into their ranks of leadership. In a very real sense, this is the same situation that existed in 1959–1960, when Kennedy and Nixon fought it out, the first two presidential candidates born in this century. When they came into power, when they came into their political maturity, they and their colleagues defined economic and political life, the debate in this country for almost three decades. That's the period we're now going into. It becomes very important for us to sort out the choices. What's at stake? What does it mean? Because in large measure, what's done over the next four or five years, in politics and economics, is going to have a major influence on our work and life well into the 21st Century.

Let me put into perspective what's happening. I think the first thing we've got to recognize is that we're moving into this country's third economic era. The first era was from George Washington to Taft. That was the developmental period. It was a period when we closed our borders to everything except people. From 1857 to 1912 we didn't lower the tariff one time in this country.

The second period was Wilson through Ronald Reagan, the free trade era, a period when we opened our borders, accepted imports, and did a great deal of exporting. We moved from a debtor to a creditor nation.

Now we're moving into a third era, what I call an era of interdependence. We have gone from a period of economic isolation to a period of economic interdependence. The point I'm making is that so far we're not making that shift very well. Let me give you four facts that suggest our inability to deal with our realities.

The first fact is in industry such as autos. This country has drifted from half of the global market share twenty years ago to somewhere in the neighborhood of 23% of global market share. When we look at business services, we find that this country has declined from 15% of global market share to 7% of global market share. When we look at high technology trade, we find that we've gone from a $26 billion trade surplus eight years ago to a $2 billion trade deficit last year. And when we look at agricultural trade, in May of last year, the United States actually imported more food than we exported. So when we look at it, what we find (though we've gone into a period of economic interdependence) across the board is that our firms, our industries, our workers, are being pushed out of one market after another. What makes that particularly dangerous for us as a country is the fact that 70% of our industry now faces intense foreign competition in these markets, up sharply from 25% as recently as 1963 when John Kennedy was President.

The difficulty we have is arousing our population to this reality. We have the lowest levels of unemployment that we've had since the 1970s. We have low inflation. Our stock market is proceeding apace very well. What is less obvious to our public is that a major portion of this prosperity is false in the sense that it's being propelled by the largest Keynesian stimulus that this country has ever had: these massive federal budget deficits. This can't continue. The issue is going to be how do we end these deficits and move back to self-sustaining, non-inflationary economic growth, and how, in the process of making this transition, do we do it in a manner so that we don't pitch the world into another global recession, if not a global depression.

In the very real sense, we're in a period of very risky business, as we work our way out of our dilemmas, as we try to move back to a period of self-sustaining growth. What must we do? That is really the central question that all of these presidential candidates and the members of Congress and the business leaders are grappling with.

My advice goes along several lines: the first thing we must do is put out of our minds some of the false choices that are being laid before us today. Let's just speak about the four principal ones. The first of those false choices is manufacturing versus service industries. Yes, it's very true that we're seeing a real decline in employment in manufacturing. And yes, it is true that the overwhelming majority of our jobs are being created in the service industries. But it is not true that the United States can afford to lose its manufacturing base as many now advocate. The fact of the matter is that manufacturing contributes about 30% of all of the value added created in this economy. More importantly, the manufacturing base underpins a major portion of the service industries. When we look at our foreign competitors, what we also find is that their manufacturing and service industries are closely linked. The Japanese, for example, over the next two years will be building over 300 auto parts plants in the United States. Almost without exception, what they're doing is bringing their own engineering services, their own architects and construction companies, to control the design and the construction of those facilities. In other words, you can see the linkages here between the manufacturing and the service industries.

What's happening in our industrial sector is what happened to agriculture seventy-five years ago. We shifted out of high labor intensive agriculture into low labor intensive agriculture. But, because we were able to maintain the productivity and the production, we were able to create the agri-businesses and create millions of jobs in those industries. That's the same route we must take today.

This leads us to the second false choice of industrial production versus industrial employment. The only way that this country is going to be able to sustain its industrial production and be competitive is to reduce the number of manufacturing workers. We must automate, and automate quickly. We must move from 18% of our population in manufacturing to 10–12%, and we must do that very quickly. That's the only way that we can maintain a competitive edge and produce at the high quality levels that are now required in the global marketplace. What that's going to mean is massive disruption in our labor markets. What we're already seeing is two million people a year become unemployed because their jobs have disappeared.

The real question that we face is how we take these workers, the preponderance of whom are middle aged, and get them back in the workforce. How do we do it quickly? And how do we do it with minimal disruption, both economically and politically?

The third false choice that we're presented is big business versus small business. This is a pervasive choice. This is a choice that is almost part of our culture. It is one of our cultural myths. Part of that can be traced back to the robber barons and the reality that this is the only country in the world where big government was created to control big business. In every other nation, big government came first and facilitated the creation of big business. There has been a long, historical, adversarial relationship that is translated into an adversarial relationship between big and small business. But the fact of the matter is that the overwhelming preponderance of our exports are done by big business. They alone have the capital, the necessary resources on a massive scale, to build the plants, and equipment, and to take on foreign competitors from Japan and Korea which are often double and triple the size of our largest corporations.

At the same time, we find a majority of our jobs are being created by small firms—70% of our jobs are being created by our small firms. We also find that big business is the primary customer for these small firms, and we find increasingly that big business is placing a good amount of its business with small firms to get the lower operating cost, to fill out the niches, to get flexibility, much as the Japanese do. The real question is not big versus small business. The question is how do we have a prosperous big and small business section.

The final choice we've been given, again a false choice, is high-tech versus basic industries. The fact is, our basic industries, such as auto and steel, are not going to survive unless they get those advanced technologies that can automate their facilities quickly. The reality is that high-tech is going to be the salvation of big industry in this country. What we need is an economic environment where capital investment can proceed much faster than it is now. Once we put these myths aside, once we put these false choices aside, then the question logically is—how do we proceed? How do we go about creating the type of economy that we need? My argument is that what we really need is an economy that is flexible, an economy that's got vigor, one that's dynamic. In fact, you could almost say that there are two schools of economic thought on how to do it. One was expressed by Damon Runyon in his musical "Guys and Dolls." He had Bat Masterson, the protagonist, say—"The race may not go to the swift or the battle to the strong, but that's how to bet your money." That's opposed to the Mae West school. She said, "If something is worth doing well, it's worth doing very slowly."

I personally am of the Damon Runyon school. If you are going to have a swift and strong economy, how do you go about it? What's keeping us from doing what we should be doing? There's a series of obstacles, or choke points, or bottlenecks, that really keep this economy from moving along. The first and the most important today is the fact that we just haven't recognized the fact that we are in an interdependent global economy. Our policies are still trapped in the 1940s and 1950s when we dominated the global economy. When you go back and look, what you find is that the global trade system we now have was designed by us and the British between 1943 and 1947. We built it on three foundations: the International Monetary Fund, the World Bank, and something called the International Trade Organization. The International Trade Organization was to be a supernational organization whose basic purpose was to control unfair trade practices. It was to knock them down and not permit a re-occurrence of the Great Depression. The United States Senate refused to ratify that treaty. Then what we had to do was go into a very complex, very difficult set of negotiations with other nations and create something called the GATT (General Agreement on Tariffs and Trade). When it was first created it worked fairly well. It was based on the assumption that the rest of the world's economies were, in some degree, like those of the United States and England, that in effect it was an Anglo-American rule-driven economy. The government sets the rules;

then the entrepreneurs, and the market, can work. For twenty years it worked very well. But two or three things have since happened. The first is that the very nature of trade has changed. The GATT only deals with merchandise trade, but, as you know in your business, increasingly trade is now in services, in capital flows and financial flows. In fact, we're at a point today where the GATT only covers 7% of global imports/exports, financial flows, and commerce. 93% of all global exchanges are done without rules. Most of your industry is done without rules in the global marketplace.

The second basic thing that's occurred is that four other types of economic systems have now come into play. What we now see, for example, is that the Marxist economies, Eastern Europe, the People's Republic of China, the Soviet Union, do 20% of all the trade in the world. We also now have Socialist economies in Europe, France, for example, where government owns the businesses, or Sweden where it will be privately owned but heavily government-regulated. They now do about 20% of global trade. We find the developing economies in Brazil, Latin America, Africa, and Southeast Asia. These economies are scrambling to deal with the reality that they're going to have three billion more people over the next forty years. That's more people than there were on the earth when John Kennedy was president.

Then we have the fifth type of economic system, the plan-driven market economies of Asia, Japan, Korea, Taiwan, and Singapore. Here are economies where business and government blend the power of the state with the flexibility of the marketplace. Business and government have a vision. That vision targets certain industries. It may be textiles in the '40s, steel in the '50s, consumer electronics in the '60s, automobiles in the '70s, computers and high-tech in the '80s, or advance material and bio-genetics in the '90s. Credit is given, infant industry protection is given, foreign firms aren't permitted to compete. They take their products to the point where they're competitive on a world class basis, and then they work together, business and government, and surge onto the global marketplace. What is different about that system and our system?

Our system focuses on process. Their system focuses on results. We focus on quarterly earnings. They focus on marketshare—long-term marketshare. They have a different set of relationships between business and government than we do. And equally important, they have a different set of industrial structures than we do. For example, in Japan 40% of all the manufacturing is owned by the banks. Not only will it be owned by banks, but you'll find in those great industrial combines a horizontal and a vertical integration that's not permitted in this country by anti-trust and a variety of other laws. So suddenly what they have is the power of size and the benefits really of a dynamic oligopoly without any of the laziness of monopoly. They have become the most formidable competitors that we face. They're unsentimental, they're predatory, they're fully willing to do whatever it takes to get the business and to take the marketshare. And, as many of you are aware, the financial industry is one that the Japanese have targeted for the late 1980s and into the 1990s. They're tough, aggressive competitors, and for that we must respect them.

At the same time, it is important for us to recognize that there are a number of things that we're going to have to do here to enable us to produce goods and services that are more competitive, that are competitive in terms of price, quality, service, innovation, and marketing. The five foundations: price, service, quality, marketing, and innovation. By and large, I'm absolutely convinced that we can do it.

In 1981 I wrote a book called *America in Ruins*. I'm not pointing that out for any reason other than to say that I'm not inherently an optimist. I point it out in order to say that this is a job that we're fully capable of doing. We've got the capital, even though we've slipped on technology and really can be criticized for not deploying our technology. We really do have first-rate technology, and more importantly, the capacity to create even more technology. We have millions of skilled, dedicated workers. But, when you take a look at our assets, what you find is we're not deploying them, either as well or as rapidly as we should. Equally important, we find that we have some real bottlenecks that keep this great economic engine from performing the way it can and must if we are to realize most of our national aspirations.

Let me talk about just a few of these. These are some of the things that you're now beginning to see the Congress and the presidential candidates focus on. These are some of the issues that you're going to see become very visible in 1988 and certainly visible in 1989. Political policy, economic policy, is a business in which the gestation period is roughly equivalent to that of an elephant. It takes about twenty-four months. What happens is you'll see the ideas thrown up in op-ed pieces; they'll come out in books; they'll be discussed. You'll see a draft piece of legislation; it will be batted around; and then, if it can survive that sort of an intellectual and political salmon run, it stands a chance. Here are some of the ideas that are now in that salmon run, some of which I think have a chance at getting someplace.

For business, I believe that the principal obstacles to meeting the competitive challenge are the enormous pressures that American businesses face for quick results and short-term earnings. I trace that to two different sources. For big business I think the principal cause, or the principal source, of the short-term myopia is the fundamental changes that have occurred in the New York stock market over the past thirty years. Specifically, what we've seen is the shift of ownership of stock from individuals to institutions. In 1952, for example, institutions owned 5% of the stock. Today, they own somewhere in the neighborhood of 35%. What we have seen, and what we did not anticipate as this shift occurred, is the fundamental difference in the way that individuals and institutions treat their stock. Individuals, by and large, will hold their stock for longer periods of time—six or seven years—at least according to the economist on the New York Stock Exchange staff.

What we find is that, increasingly, institutions are turning over their stock—they're going for the earnings. Those institutional fund managers find themselves under intense pressure to get quick results. We can measure this in a number of ways. One is simply the large block trades. In 1965, there were on average nine large block trades; last year there were about 2,400. Another way we can measure it is the total pace in which the New York Stock Exchange turns over. In the mid-1970s, the total value of the New York Stock Exchange was turning over roughly every five years. Now the total value of the New York Stock Exchange is turning over every twenty-two months. You take that rapid turnover; you take this large amount of money in the hands of institutions; you take the fact that of your largest 200 corporations, such as TRW, 50–60% of the stock is held in the hands of institutions, in other words, no long-term loyalty to the company or to its products; and then you take on top of that the raiders that are willing to take a corporation, break it up, and just suck the equity out of it, like a vampire; and suddenly you have a dynamic underway in which American corporations are really holding back: they're moving to the defensive.

What you're seeing across the board is large numbers of American corporations that are simply buying back their own stock. They're pulling back on research. They're holding back much of their capital investment. In other words, they're moving to a defensive posture to prepare themselves. That's not a foolish position, but it imperils our long-term ability to meet the competitive challenge. When you look, for example, at what's happening in research, the National Science Foundation reports that we've seen a fundamental shift in the character of the research that corporations are doing. Specifically, rather than putting their money into those type of activities where you can have a real long-term breakthrough, such as xerography, increasingly what's happened is our American corporations are putting their money on those kinds of activities where they can simply refine existing technologies, in other words, where they can get it to market quickly.

What they're saying is they can't afford to wait. They can't afford to wait five to eight years through the research program. There are very few corporations in this country, such as General Electric or IBM, that have the profitability to permit them to keep their earnings up and, at the same time, to do the research to prepare to compete in the 1990s. For small business, the principal driving force for short-term pressures on them is their inability to get long-term money. What we find is a real boom in venture capital activity in this country over the past ten years. We also find that only about 1% of our startups of small businesses are financed by venture capital. The preponderance, 99% of the small businesses in this country, are financed by people taking their savings, borrowing from their families, taking a second mortgage on their house, scrimping here and there, pulling it together. Even when those firms are able to succeed in the short-term, they quickly move to a point when they've reached some level of success, when they have a plateau when they need longer-term money.

It's at that point that they're really cut off at the knees. While the United States has the largest formation rate of small business, we also have the largest failure rate. One of the problems is that we don't have the ways and means for our small businesses to get patient capital—long-term capital. What we now see moving through the House Small Business Committee is the means to create that. It's called an industrial mortgage corporation. It would have the federal government establish a device much like Fannie Mae or Ginnie Mae in the housing market that in effect would be a secondary mortgage mechanism that would permit banks and other commercial lending organizations to take a portion of a small business industrial loan and sell it off to that instrument where they can put it into larger play. Whether that winds up being the mechanism or not, the fact is it addresses the right issue. That issue is that over 85% of the small business loans of under $1 million are due within five months. The fact is that when you're dealing in this type of economy, you can't make it on five month money. You've got to have some five to ten year long-term money. What this country needs is to find the ways and means to channel lots of money to the small business sector and do it on sound business principles.

Another major problem this country has is within our workforce. By the same token, it is for the foreseeable future one of our prime opportunities. Let me just describe some of this, and some of this is going to directly touch you and your companies in your personal lives. When we take a look at the demography of our workforce, what we find is that because of the post-World War II baby generation, the majority of our workers are now moving into their most productive years. For economists, we say that's generally between the ages of twenty-five and fifty-five. In 1970, we had 61% of our people in that population cohort. Now it's roughly 67%. By 1995, when we take a look with the demographers, we see that roughly 75% of our people will be in that population cohort. That gives us a real demographic advantage if we can deploy and use those people. Flip it around and you say that it's a necessity that we deploy those people and make sure that they're competent. Another simple fact is, roughly nine out of ten of our workers in the year 2001 are already adults and most are at work. We're going to make it or break it with today's workers. That's the message over the next decade: we're going to make it or break it with today's workers. Whatever we do in this K through 12 system, as important as it is, whatever we do on that will not begin to show up until the 1990s, or perhaps the late 1990s. That's for the future. The real issue is how do we retool 110 million American workers, most of whom got their education and thought it would last them a lifetime? How do we deal with a workforce where one out of five workers changes jobs every year?

The first problem we face is that while one out of five of our people have been to college, another one out of five of our adults in this country cannot read, write, or count at a 7th grade level. A fifth of all of our adults are functionally illiterate.

Roughly 10% of our workers are now impaired, to some degree, because of drugs and alcoholism. It's costing our employers roughly $45–60 billion a year in health cost, absenteeism, and accidents. It's tough to tell when somebody is on drugs or alcohol, but it's a major problem that we have.

A third major issue is that we have the least motivated workforce of all industrial nations. Daniel Yankelovich reports out of his surveys that only ten out of every 100 workers believe that if they work harder it will result in any increased pay or benefit to them. They believe that the money will go to the boss, to the top CEO, to the stockholders, to somebody else. I don't know that they're that stupid. They may be actually right, given our structure. For most of our workers we don't have a pay system that's set up to relate reward and effort. Maybe, in effect, these workers are seeing reality. But the fact is when you take a look at the Japanese, what you find when you run surveys on their workforce is that 93% of their workers think if they work harder and smarter that they're going to be beneficiaries. Ten out of 100 for Americans, ninetythree out of 100 for the Japanese. That's a serious indictment of our management in compensation systems.

Take a look at the Toyota-GM plant in Freemont, California. As you know, General Motors was set to build a massive plant in Tennessee, the Saturn Plant, and they were going to automate it and were going to use the technology and do these various things. The Japanese came in and put in their management system in that facility, in that joint venture. What they found is they were getting 20 or 30% greater productivity out of the same workers, with the same equipment, than what they could expect out of the Saturn plant.

Ford Motor Company has made a major change in their management style and their motivation practices since 1980. The chairman of Ford Motor Company just says it flat out—"The only thing at our plants that is not new are the workers and the walls: management, machines, everything." They're getting compound productivity growth rates of over 7% a year. That's one of the reasons, in addition to their quality, why Ford Motor Company was able for the first time this year to surge ahead of General Motors in profits. It's a real challenge that each of us faces in business. How do we take and motivate these workers?

We have two other worker issues that are important, that as a society we have not yet addressed. The first is we've not yet accommodated the reality of large numbers of women in the workforce. Let me put it another way. This country has undergone one of the most formidable workplace revolutions that we've ever had. It's been largely invisible and it's been a very polite revolution. That is the movement of women on a permanent basis into the workforce. In the 1960s and 70s we had a circumstance in which most of the women that went into the workforce went into traditional areas such as teaching. In 1970, for example, we had a circumstance where only eight-tenths of 1% of the engineering graduates were women. It's just now moved up to about 15%. When you move into law and medicine, you're moving into rates that are into the 30s and 40%. And across the board what you're seeing, in most professions, is the same massive movement, on a permanent basis, of women into the workforce. That's important for two reasons. One, it says we have the type of opportunity as an egalitarian society that we should have. But it also says that increasingly a major portion of this country's human capital and knowledge now resides in our women workers. That's important because this quiet revolution has occurred so quickly that nine out of ten women in the workforce today are still in their childbearing years, and seven out of ten of them will have one more child. You can see where that's leading me in the argument—child care services.

Child care services has suddenly become, I believe, one of the major workplace issues in this country. It's increasingly less a social issue and more an economic and workplace issue. Yet what we find is the majority of our employers have not yet recognized this. Given your profession, you'll see it in the benefits, in other plans, but most companies have not yet addressed this reality. If we are to make full use of our limited human capital in this country, in this period of intense competition, we must find the ways and means to assure safe, nurturing, affordable, convenient child care services. Child care services will be one of the major campaign issues in 1988.

The final issue I want to bring to your attention is portable pensions. We're at a point in this country where we increasingly have a middle-aged workforce. When people become middle-aged, they begin to think about retirement and responsibilities. Yet what we also find in this country is a circumstance in which only half of the workers are employed by firms that even offer a private pension plan. Because of the vesting periods, and because of the high turnover rate that we now have, only about a third of our workers are vested in any pension plan, and for many of them, that vesting is of a very short nature and so the value of that plan doesn't mean much. In the past, the issues about pensions have centered around how do you level out benefits for that worker who goes from job to job. How do you manage the plans, etc? For firms it's been an issue of how do you set up a pension plan that will hold the worker. How do you put platinum handcuffs on the worker? How do you keep the worker there? Increasingly, those issues are going to shift; they're going to shift to that of how do you have flexibility within the workforce? How do you create the ways and means whereby the worker can have a pension and move from job to job? Our demand now is not for workers to stay with the same job; our demand now is for people to move. What William White wrote thirty years ago in his book *The Organization Man* is no longer true. Then, as White envisioned it, what we would have in this country is most of our workers, workers of high caliber, going to work for some of the great institutions, business, academia, government, and then they would stay with those institutions for the majority of their lives. They would tend those great organizations, and the compact that was made is the institutions would then provide them services, benefits, and pensions. That compact has broken down. That compact because of external realities no longer exists.

Let me just ask this audience—how many of you in your careers, since past the age of twenty-two, have been with three or more employers? The Harvard Department of Economics estimates workers over thirty who change their pension plans once will lose 27% of their benefits. If they change jobs twice, they'll lose 52% of their benefits. The reality is that most of our workers will change occupations three times and jobs six to ten times during their careers. The time has come in this country for us to devise a private pension system where all workers are covered and where those workers can go from job-to-job, place-to-place, occupation-to-occupation, in and out of the workforce, between the public and the private sector, and all the while build up a safe, sound, well-financed retirement system.

In short, the time has come to tie the pension to the worker rather than to the job. I think we're going to see something on that. One of your speakers in the next panel, Congressman Ed Feighan, has been one of the leaders in Congress and at the national level to make that happen. You might want to talk to him about that.

In summary, what I'm saying is we're coming into a period, one of those hinges of history, one of those great periods of change, where the decisions that we make will be big decisions. They will affect the balance of our lives, and they will have a disproportionate influence on the lives and welfare of our children. The challenge that we face as a nation is how do we deal in an interdependent world. And how do we deal in a world where we no longer have an easy superiority over other nations? How do we deal with other nations, increasingly in Asia and Europe, as equals, rather than supplicants? Equally important, how do we deal domestically, in an environment of great uncertainty, where we have much catch-up business to do from the past, out of the '60s and '70s, things that we didn't do that we must now do? And at the same time, we must deploy our formidable assets of capital, technology, and workers in a manner where we can meet that foreign competition.

If we were to have a motto, I would almost go back and use the one of a friend of mine in college. He was an electrical engineer, very popular, always dating. I was very jealous of him, so I once asked Bartus, I said, "Bartus, you're so popular." He answered, "Well, it's my philosophy. Beauty is only skin deep, but that's thick enough for me." So I propose, as a national motto, that economic prosperity may not be everything, but that's good enough for us.

MINUTES OF THE 1987 SPRING MEETING

May 10–13, 1987

CONTEMPORARY RESORT HOTEL, ORLANDO, FLORIDA

Sunday, May 10, 1987

The Board of Directors held their regular quarterly meeting from 1:00 P.M. to 4:00 P.M.

Registration was held from 4:00 P.M. to 6:30 P.M.

From 5:30 P.M. to 6:30 P.M. there was a special presentation to new Associates and their guests. This session included an introduction to standards of professional conduct and the CAS committee structure.

A general reception for all members and guests was held from 6:30 P.M. to 7:30 P.M.

Monday, May 11, 1987

Registration continued from 7:00 A.M. to 8:00 A.M.

President Michael A. Walters opened the meeting at 8:00 A.M. The first order of business was the admission of members. Mr. Walters recognized the 66 new Associates and presented diplomas to the 22 new Fellows. The names of these individuals follow.

FELLOWS

Neil C. Aldin	Steven A. Gapp	Robert H. Lee
D. Lee Barclay	Charles Gruber	Warren D. Montgomery
Allan Chuck	Denis G. Guenthner	Layne M. Onufer
Frederick F. Cripe	Mark J. Homan	Charles I. Petit
Myron L. Dye	Ruth A. Howald	Rajagopalan J. Raman
Howard M. Eagelfeld	Wayne S. Keller	Timothy L. Schilling
Kenneth Easlon	Paul J. Kneuer	David A. Withers
Grover M. Edie		

ASSOCIATES

Ralph L. Abell
Christiane Allaire
Jean-Luc E. Allard
Paul Boisvert, Jr.
Charles H. Boucek
Joseph J. Boudreau
Theresa A. Bourdon
Malcolm E. Brathwaite
Paul J. Brehm
Richard S. Brutto
John W. Buchanan
Ruy A. Cardoso
Chyen Chen
Walter P. Cieslak
Ann M. Conway
Mark Crawshaw
Susan L. Cross
William Der
Carol Desbiens
Anthony M. DiDonato
Norman E. Donelson
Janet M. Ericson
Pierre Fromentin
Gregory S. Girard
Leonard R. Goldberg
Gregory T. Graves
Alex R. Greene
Ann V. Griffith
Larry A. Haefner
David H. Hays
David R. Heyman
Clive L. Keatinge
Eric R. Keen
Jerome F. Klenow
Kenneth R. Krissinger
Paul E. Lacko
Dean K. Lamb
Pierre G. Laurin
Joseph R. Lebens
Nicholas M. Leccese, Jr.
John J. Lewandowski
Sam F. Licitra
Elise C. Liebers
Brett A. MacKinnon
Sean P. McDermott
Mary F. Miller
Karen J. Pichler
Richard A. Plano
Donald W. Procopio
Mark R. Proska
Sara E. Schlenker
Frederic F. Schnapp
Debbie Schwab
Kim A. Scott
Mark R. Shapland
Peter J. Siczewicz
Craig P. Taylor
R. Glenn Taylor
Andre Veilleux
James C. Votta
Robert A. Weber
Peter G. Wick
Lincoln B. Williams
Robin M. Williams
Ernest I. Wilson
Bill S. Yit

Mr. Walters then introduced Mr. Ronald Bornhuetter, a past President of this society, who addressed the members concerning their professional responsibilities.

Mr. Michael Fusco, Vice President-Programs, gave the highlights of the program.

Mr. Stephen Philbrick, Chairman of the Committee on Review of Papers, summarized the three new *Proceedings* papers and the one review of a previous *Proceedings* paper.

Mr. Gary Venter summarized his review of a *Proceedings* paper.

Ms. Janet Fagan, Chairman of the Committee on Continuing Education, gave a summary of the Discussion Paper program.

Mr. Walters concluded the business session at 9:00 A.M. and introduced Dr. Pat Choate, Director of Policy Analysis, TRW, Inc., who delivered the keynote address. Dr. Choate spoke on a wide range of issues including competitiveness, foreign trade, and the role of government in the regulation of business and industry.

A panel presentation, "The McCarran-Ferguson Act; Have We Seen the Last of It" followed. The panel was moderated by Mr. David G. Hartman, Senior Vice President and Actuary, Chubb Group of Insurance Companies. The panel members were James M. Stone, President, Plymouth Rock Assurance Corp.; Bruce A. Bunner, Principal, Pete Marwick, Mitchell & Company; and, Representative Edward F. Feighan, U.S. House of Representatives.

A luncheon followed from Noon to 1:30 P.M.

The afternoon was devoted to presentations of the twelve discussion papers, four new *Proceedings* papers, and six panel presentations.

The new *Proceedings* papers were:

1. "Adjusting Loss Development Patterns for Growth"
   Author: Charles L. McClenahan
   Coopers & Lybrand

2. "The Construction of Automobile Rating Territories in Massachusetts"
   Author: Robert F. Conger
   Tillinghast/Towers Perrin

3. "Revisions in Loss Reserving Techniques Necessary to Discount Property Liability Loss Reserves"
   Author: Stephen P. D'Arcy
   University of Illinois

4. Discussion of "An Analysis of Experience Rating"
   Author of Discussion: Howard C. Mahler
   Massachusetts Rating Bureau
   Author of Paper: Glenn Meyers
   University of Iowa

The Discussion Papers presented were:

1. "An Analysis of the Impact of the Tax Reform Act on the P/C Industry"
   Authors: Gerald I. Lenrow
   Coopers & Lybrand

   Owen Gleeson
   General Reinsurance Corporation

2. "An Investigation of Methods, Assumptions, and Risk Modeling for the Valuation of P/C Insurance Companies"
   Author: Robert S. Miccolis
   Tillinghast/Towers Perrin

3. "Investor's Valuation of an Insurance Company"
   Author: Joel S. Weiner
   CIGNA Property & Casualty Companies

4. "Underutilization of Capacity"
   Author: Neal J. Schmidt
   St. Paul Reinsurance Management Corporation

5. "Allocation of Surplus for a Multi-Line Insurer"
   Author: Paul J. Kneuer
   Insurance Services Office

6. "A Non-Parametric Approach to Evaluating Reinsurers' Relative Financial Strength"
   Authors: Stephen J. Ludwig
   Hartford Insurance Company

   Robert F. McAuley
   Hartford Insurance Company

7. "Regulatory Standards for Reserves"
   Author: Oakley Van Slyke
   Coopers & Lybrand

8. "Insurance Profits: Keeping Score"
   Author: Richard G. Woll
   Allstate Insurance Company

9. "A Framework for Forecasting P/C Insurers' Financial Results"
   Authors: Paul Braithwaite
   Insurance Services Office

   Isaac Mashitz
   North American Reinsurance Company

10. "Asset/Liability Management: Beyond Interest Rate Risk"
    Author: William H. Panning
    Aetna Life & Casualty Insurance Company

11. "Measuring R.O.E. From a Financial Planning Perspective"
    Authors: Bruce R. Jones
    Travelers Insurance Company

    Neil Aldin
    Travelers Insurance Company

12. "An Analysis of the Capital Structure of an Insurance Company"
    Author: Glenn Meyers
    University of Iowa

The panel presentations covered the following topics:

1. "Actuarial Principles and Standards of Practice"

   The statement of principles and standards of practice affect the professional activities of every practicing actuary. This panel discussed the distinct roles of the CAS and the IASB in developing such guidelines for both the insurance ratemaking and reserving disciplines.

   Moderator: Charles A. Bryan
   Senior Vice President & Actuary, USAA

   Panelists: Chairman of Committee on Ratemaking
   Michael J. Miller
   Consulting Actuary, Tillinghast/Towers Perrin

   Chairman of Committee on Reserves
   James A. Faber
   Principal, Peat, Marwick, Mitchell & Company

2. "Personal Lines Classification Plans: Actuarial and Real World Considerations"

   The panel discussed how companies approach private passenger pricing by classification. The discussion included actuarial techniques and type of data used by the companies, as well as other considerations such as marketing strategies, interaction with underwriting, and regulatory constraints.

   Moderator: Steven F. Goldberg
   Vice President & Actuary, USAA

   Panelists: Irene K. Bass
   Vice President & Senior Actuary, Crum & Forster

   Alan R. Ledbetter
   Vice President, GEICO

   John J. Javaruski
   Secretary, Hartford Insurance Company

3. "Report on Current Activities of the Financial Analysis Committee and its Interaction With the Committee on Valuation Principles and Techniques"

   The CAS Committee on Financial Analysis presented initial results on a study of the problems created by asset/liability mismatch in an environment of unstable interest rates. There was also a discussion of the genesis and goals of the Committee on Valuations and its relationship to other development committees including the Financial Analysis Committee.

   Moderator: Steven Petlick
   Vice President, Continental Reinsurance

   Panelists: Robert P. Eramo
   Vice President & Chief Actuary, Unigard

   Charles H. Berry, III
   Actuary, Aetna Life & Casualty Insurance Company

   Robert A. Miller, III
   Consulting Actuary, Milliman & Robertson, Inc.

4. "Risk Retention Act"

   A well structured Risk Retention entity should be both attractive to its members as a risk bearing entity and profitable to insurers as a risk transferee. This workshop reviewed the key features of the Liability Risk Retention Act of 1986 and, using a case study, presented the critical multi-disciplinary worksteps in evaluating the feasibility of forming a Purchasing Group or Risk Retention Group.

   Moderator: Albert J. Beer
   Principal, Tillinghast/Towers Perrin

   Panelists: Alfred O. Weller
   Vice President & Chief Actuary, Fred S. James & Company

   P. Bruce Wright, Esq.
   Partner, LeBoeuf, Lamb, Leiby & MacRae

5. "How to Start an Insurance Company"

   Two insurance company executives related their motivations and experiences in starting up a new insurance company. In particular, they discussed perceived opportunities in the marketplace, regulatory concerns, financing, marketing strategies, inherent risks, and ultimate success.

   Kenneth R. Rosen
   President & Chief Executive Officer, Victoria Financial Corporation

   E. Dow Walker, Jr.
   Chief Operating Officer, Mutual Assurance

6. "How to End an Insurance Company"

   Two actuaries experienced in rehabilitating and running off troubled insurance companies discussed what must be considered from an economic, legal, and social standpoint in performing their services.

   Dale F. Ogden
   Executive Vice President, Kramer Capital Consultants, Inc.

   David M. Patterson
   Vice President, Philadelphia Insurance Research Group

The officers held a reception for the new Fellows and their spouses from 5:30 P.M. to 6:30 P.M.

The Presidents Reception for all members and guests was held from 6:30 P.M. to 7:30 P.M.

Tuesday, May 12, 1987

Tuesday was devoted to a continuation of the Monday afternoon sessions.

A Reception and Barbecue was held from 6:30 P.M.to 9:00 P.M.

Wednesday, May 13, 1987

The business session resumed at 9:00 A.M. with a presentation of the Harold Schloss Award to Mr. Brett Scrantan.

There was also an award of the Michelbacher prize to Mr. Glenn Meyers.

The business session was convened at 9:30 A.M. and a panel presentation, "Insurance Company Ratings: Do They Mean What We Think They Mean" followed. The panel was moderated by Robert A. Bailey, Senior Vice President, E. W. Blanch & Company. The four panelists included Lawrence A. Hayes, Vice President, Standard & Poors Corporation; Michael Miron, U.S. Editor, Insurance Solvency International; Robert Arvanitis, Senior Analyst, Moodys' Investors Service; and Robert A. Brian, General Partner, Conning & Company. These speakers described how they see their role in the formal financial evaluation process.

The meeting was adjourned at 11:15 A.M.

May, 1987 Attendees

In attendance, as indicated by the registration records, were 335 Fellows; 203 Associates; and 39 guests, subscribers, and students. The list of their names follows.

FELLOWS

Aldin, N. C.	Bailey, R. A.	Bassman, B. C.
Aldorisio, R. P.	Bailey, V. M.	Beer, A. J.
Alfuth, T. J.	Barclay, D. L.	Bell, L. L.
Allaben, M. S.	Barrow, B. H.	Bellinghausen, G. F.
Angell, C. M.	Bashline, D. T.	Belvin, W. H.
Asch, N. E.	Bass, I. K.	Ben-Zvi, P. N.

FELLOWS

Bensimon, A. S.
Berens, R. M.
Berry, C. H., III
Berry, J. L.
Bertrand, F.
Beverage, R. M.
Biegaj, W. P.
Bill, R. A.
Biller, J. E.
Biondi, R. S.
Biscoglia, T. J.
Boccitto, B. L.
Bornhuetter, R. L.
Boulanger, F.
Bouska, A. S.
Bovard, R. W.
Bowen, D. S.
Boyd, W. A.
Braithwaite, P.
Brian, R. A.
Brooks, D. L.
Brubaker, R. E.
Bryan, C. A.
Burger, G.
Cantin, C.
Chansky, J. S.
Chanzit, L. G.
Cheng, J. S.
Chernick, D. R.
Chiang, J. D.
Childs, D. M.
Christiansen, S. L.
Christie, J. K.
Chuck, A.
Coffin, J. D.
Cohen, H. L.
Conger, R. F.
Cook, C. F.
Corr, F. X.
Covney, M. D.
Cripe, F. F.
Crowe, P. J.
Cundy, R. M.
Curran, K. F.
Currie, R. A.
Curry, A. C.
Daino, R. A.
D'Arcy, S. P.
Dawson, J.
Demers, D.
Dempster, H. V.
Deutsch, R. V.
Dodd, G. T.
Doellman, J. L.
Doepke, M. A.
Dolan, M. C.
Donaldson, J. P.
Dorval, B. T.
Drennan, J. P.
Drummond-Hay, E. T.
Duffy, B.
Duffy, T. J.
Dye, M. L.
Eagelfeld, H. M.
Easlon, K.
Edie, G. M.
Ehrlich, W. S.
Eland, D. D.
Engles, D.
Evans, G. A.
Eyers, R. G.
Faber, J. A.
Fagan, J. L.
Fallquist, R. J.
Fein, R. I.
Fisher, R. S.
Fisher, W. H.
Fitzgibbon, W. J., Jr.
Flynn, D. P.
Foote, J. M.
Foster, R. B.
Fresch, G. W.
Friedberg, B. F.
Furst, P. A.
Fusco, M.
Gapp, S. A.
Garand, C. P.
Gersie, M. H.
Gleeson, O. M.
Gluck, S. M.
Goddard, D. C.
Goldberg, S. F.
Goldfarb, I. H.
Gottheim, E. F.
Gottlieb, L. R.
Grace, G. S.
Grady, D. J.
Graves, J. S.
Greco, R. E.
Grippa, A. J.
Gruber, C.
Guenthner, D. G.
Gutterman, S.
Hachemeister, C. A.
Hafling, D. N.
Hall, J. A., III
Hallstrom, R. C.
Haner, W. J.
Hankins, S. E.
Hanson, J. L.
Hapke, A. J.
Hartman, D. G.
Hartman, G. R.
Haseltine, D. S.

FELLOWS

Haskell, G. E.
Hayne, R. M.
Hennessy, M. E.
Henry, D. R.
Henzler, P. J.
Herder, J. M.
Herman, S. C.
Herzfeld, J.
Hewitt, C. C., Jr.
Hibbard, W. J.
Hillhouse, J. A.
Hoffmann, D. E.
Homan, M. J.
Horowitz, B. A.
Hosford, M. T.
Howald, R. A.
Hutter, H. E.
Ingco, A. M.
Irvan, R. P.
Jaeger, R. M.
Jameson, S.
Jean, R. W.
Jerabek, G. J.
John, R. T.
Johnson, A. P.
Johnston, T. S.
Jones, B. R.
Judd, S. W.
Kallop, R. H.
Kaufman, A. M.
Keatts, G. H.
Keller, W. S.
Kelly, A. E.
Khury, C. K.
Kilbourne, F. W.
Kist, F. O.
Kline, D. F.
Klinker, F. L.
Kneuer, P. J.
Koch, L. W.
Kollar, J. J.
Koupf, G. I.,
Kozik, T. J.
Krakowski, I.
Krause, G. A.
Kuehn, R. T.
Lamb, R. M.
Lamonica, M. A.
Lange, D. L.
LaRose, J. G.
Larsen, M. R.
Lattanzio, S. P.
Ledbetter, A. R.
Lee, R. H.
Lehman, M. R.
Levin, J. W.
Lewis, M. A.
Lindquist, P. L.
Lino, R.
Lino, R. A.
Lipton, B. C.
Littmann, M. W.
Livingston, R. P.
Lo, R. W.
Lommele, J. A.
Loper, D. J.
Lotkowski, E. P.
Lowe, R. F.
Ludwig, S. J.
MacGinnitie, W. J.
Mahler, H. C.
Makgill, S. S.
Marks, S. D.
Mashitz, I.
Mathewson, S. B.
Mayer, J. H.
McCarter, M. G.
McClenahan, C. L.
McClenahan, D. L.
McClure, J. W., Jr.
McClure, R. D.
McConnell, C. W.
McDonald, G. P.
McGovern, W. G.
McLean, G. E.
Mealy, D. C.
Mendelssohn, G. A.
Meyer, R. E.
Meyers, G. G.
Miccolis, J. A.
Miccolis, R. S.
Miller, M. J.
Miller, P. D.
Miller, R. A., III
Miller, R. R.
Miner, N. B.
Montgomery, W. D.
Moody, R. A.
Moore, B. C.
Moore, B. D.
Moore, P. S.
Muetterties, J. H.
Mulder, E. T.
Muleski, R. T.
Munt, D. S.
Murad, J. A.
Murrin, T. E.
Myers, N. R.
Neidermyer, J. R.
Neis, A. R.
Nester, K. L.
Newlin, P. R.
Newville, B. S.
Nichols, R. W.

FELLOWS

Nickerson, G. V.
Nikstad, J. R.
Normandin, A.
Noyce, J. W.
Oakden, D. J.
O'Connell, P. G.
Onufer, L. M.
Palczynski, R. W.
Paquette, S. L.
Parker, C. M.
Patrik, G. S.
Pearl, M. B.
Petersen, B. A.
Petit, C. I.
Petlick, S.
Philbrick, S. W.
Phillips, H. J.
Pierce, J.
Pinto, E.
Plunkett, R. C.
Pratt, J. J.
Prevosto, V. R.
Purple, J. M.
Quirin, A. J.
Racine, A. R.
Raman, R. K.
Rapp, J. W.
Reichle, K. A.
Reynolds, J. J., III
Richardson, J. F.
Robertson, J. P.
Rodermund, M.
Rogers, D. J.
Rosenberg, D. M.
Roth, R. J., Jr.
Rowland, W. J.
Scheibl, J. A.
Schilling, T. L.
Schmidt, N. J.
Schneider, H. N.
Schumi, J. R.
Schwartzman, J. A.
Sheppard, A. R.
Sherman, H. A.
Sherman, O. L., Jr.
Shoop, E. C.
Shrum, R. G.
Silver, M. S.
Simon, L. J.
Skurnick, D.
Smith, M. B.
Snader, R. H.
Splitt, D. L.
Stanard, J. N.
Steeneck, L. R.
Stergiou, E. J.
Stewart, C. W.
Strug, E. J.
Suchoff, S. B.
Taht, V.
Tatge, R. L.
Thibault, A.
Thompson, K. B.
Toothman, M. L.
Treitel, N. R.
Tuttle, J. E.
Van Ark, W. R.
Venter, G. G.
Wacek, M. G.
Walters, M. A.
Walters, M. A.
Warthen, T. V., III
Wasserman, D. L.
Weller, A. O.
Westerholm, D. C.
Whatley, P. L.
White, J.
Whitman, M.
Wickwire, J. D., Jr.
Wilson, J. C.
Wilson, R. L.
Wiseman, M. L.
Wiser, R. F.
Withers, D. A.
Woll, R. G.
Yingling, M. E.
Yonkunas, J. P.
Zatorski, R. T.
Zubulake, T. J.

ASSOCIATES

Abell, R. L.
Abramson, G. R.
Allaire, C.
Allard, J. E.
Almagro, M., Jr.
Amoroso, R. C.
Anderson, B. C.
Andler, J. A.
Apfel, K.
Atkinson, R. V.
Austin, J. P.
Bellafiore, L. A.
Blakinger, J. M.
Boisvert, P., Jr.
Boor, J. A.

ASSOCIATES

Boucek, C. H.
Boudreau, J. J.
Bourdon, T. W.
Brahmer, J. O.
Brathwaite, M. E.
Brehm, P. J.
Brutto, R. S.
Buchanan, J. W.
Bujaucius, G. S.
Cadorine, A. R.
Campbell, K. A.
Cardoso, R. A.
Carpenter, W. M.
Cellars, R. M.
Chen, C.
Chorpita, F. M.
Cieslak, W. P.
Connor, V. P.
Conway, A. M.
Costner, J. E.
Crawshaw, M.
Cross, S. L.
Cutler, J. Z.
Davis, R. C.
DeConti, M. A.
Der, W.
Desbiens, C.
DiDonato, A. M.
Donelson, N. E.
Donnelly, V. T.
Doyle, M. J.
Dufresne, J.
Duperreaut, B.
Dupuis, C.
Earwaker, B. G.
Englander, J. A.
Eramo, R. P.
Ericson, J. M.
Esposito, D. L.
Fleming, K. G.
Fletcher, J. E.
Francis, L. A.
Friedman, H. H.
Fromentin, P.
Gaillard, M. B.
Gardner, R. W.
Gebhard, J. J.
Gidos, P. M.
Girard, G. S.
Godbold, M. E.
Godbold, N. T.
Goldberg, L. R.
Goldberg, T. L.
Granoff, G.
Graves, G. T.
Greene, A. R.
Griffith, A. V.
Gunn, C. H.
Gwynn, H. M.
Haefner, L. A.
Halpert, A.
Handte, M. R.
Harrison, D. C.
Harrison, E. E.
Hay, R. S.
Hays, D. H.
Head, T. F.
Henry, T. A.
Heyman, D. R.
Holdredge, W. D.
Jaso, R. J.
Javaruski, J. J.
Johnson, R. W.
Keatinge, C. L.
Keen, E. R.
Kleinberg, J. J.
Klenow, J. F.
Kline, C. D., Jr.
Koegel, D.
Kolk, S. L.
Kolojay, T. M.
Konopa, M. E.
Krissinger, K. R.
Lacek, M.
Lacko, P. E.
Lamb, D. K.
Laurin, P. G.
Lebens, J. R.
Leccese, N. M., Jr.
Leiner, W. W., Jr.
Levine, G. M.
Lewandowski, J. J.
Licitra, S. F.
Liebers, E. C.
MacKinnon, B. A.
Maguire, B. P.
Mair, S. A.
McConnell, D. M.
McCoy, M. E.
McDermott, S. P.
Miller, M. F.
Miller, W. J.
Mittal, M. L.
Miyao, S. K.
Mohrman, D. F.
Moody, A. W.
Morgan, S. T.
Morrow, J. B.
Mucci, R. V.
Mueller, R. A.
Muller, R. G.
Musante, D. R.
Nelson, J. K.
Neuhauser, F., Jr.

ASSOCIATES

Newman, H. E.
Ogden, D. F.
Ostergren, G. V.
Overgaard, W. T.
Peacock, W. W.
Pei, K. J.
Peterson, S. J.
Petrelli, J. L.
Philbrick, P. G.
Placek, A. C.
Post, J. H.
Potok C. M.
Pridgeon, R. D.
Procopio, D. W.
Proska, M. R.
Pulis, R. S.
Quintano, R. A.
Rathjen, R. L.
Rice, D. E.
Rice, J. W.
Rice, W. V.
Robinson, R. D.
Roth, R. J.
Rowland, V. T., Jr.
Sandler, R. M.
Sansevero, M., Jr.
Scheuing, J. R.
Schlenker, S. E.
Schlenz, J. W.
Schnapp, F. F.
Schwab, D.
Schwandt, J. C.
Scott, K. A.
Scully, M. W.
Seiffertt, B. A.
Shapland, M. R.
Siczewicz, P. J.
Silverman, J. K.
Silverman, M. J.
Skolnik, R. S.
Slusarski, J.
Snow, D. C.
Sornberger, G. C.
Stadler, E.
Steingiser, R.
Stroud, R. A.
Sweeney, E. M.
Swisher, J. W., Jr.
Taylor, C. P.
Taylor, R. G.
Theisen, J. P.
Tistan, E. S.
Toczylowski, D. L.
Trudeau, M.
Valenti, K. P.
Varca, J. J.
Veilleux, A.
Vogel, J. F.
Votta, J. C.
Weber, R. A.
Weiner, J. S.
Whatley, M. W.
Wick, P. G.
Williams, L. B.
Williams, R. M.
Willsey, R. L.
Wilson, E. I.
Wilson, O. T.
Wilson, W. F.
Yatskowitz, J. D.
Yau, M. W.
Yit, B. S.
Yow, J. W.
Yunque, M. A.

GUESTS-SUBSCRIBERS-STUDENTS

Brescia, J.
Butsic, R.
Cheng, P. C.
Clause, R.
Cronin, M.
DeMarlie, G.
Derrig, R.
Duitsman, W.
Eversmann, T.
Franz, V.
Gale, E.
Gutman, E.
Jensen, P.
Kaltman, A.
Kartechner, J.
Kaufman, D.
Lacefield, D.
Lenrow, J.
McAuley, R.
Miron, M.
Mitchell, K.
Murphy, K. A.
Nicholson, J. E.
O'Farrell, R.
Radhakrishnan, R.
Santomenno, S.
Smith, D.
Spangler, J.
Stenmark, J.
Stevens, V.

GUESTS-SUBSCRIBERS-STUDENTS

Swenson, R.	Thomas, A.	Whitby, O.
Taylor, R. G.	Van Leer, P.	Wildman, P.
Tetro, J.	Wade, J.	Wootton, G.

Volume LXXIV, Part 2 No. 142

PROCEEDINGS

November 4, 5, 6, 1987

AN ANALYSIS OF EXCESS LOSS DEVELOPMENT

EMANUEL PINTO AND DANIEL F. GOGOL

Abstract

There is very little information available regarding excess loss development, despite its importance in excess of loss pricing and reserving. In this study, paid and reported excess loss development patterns are estimated at various retentions for certain casualty lines of business. The effects of allocated loss adjustment expense and policy limits on excess development are discussed. The pattern of change, as development progresses, of Pareto distributions fitted to casualty loss distributions was considered in developing curve fitting methods. A method is described for determining development factors by layer. Applications to excess loss pricing, loss reserving, and increased limits factors are mentioned.

Special thanks to ISO, which provided us with a great deal of data, and to Susan Greiff, Thomas Highet, Madelyn Esposito and Francine Leong who assisted in the data processing and compilation.

1. INTRODUCTION

Loss development patterns for both reported and paid excess losses are of fundamental importance in excess of loss pricing as well as in estimating loss reserves for excess of loss insurance and reinsurance. Excess of loss reinsurance constitutes a major portion of the business written by reinsurers and is the area involving the greatest degree of independent pricing and reserving activity.

There is a paucity of published information regarding both reported and paid excess loss development. The Reinsurance Association of America (RAA) publishes a study biennially of reported excess casualty loss development patterns for certain lines of business, based on data supplied by member companies. Incurred[1] loss development patterns for automobile liability, general liability, workers' compensation and medical malpractice have been described in these studies. Certain of these lines of business have well over twenty years of significant reported excess loss development, indicating that excess reporting patterns vary significantly from first dollar reporting patterns. In that study, however, excess losses in various layers are all grouped together, so the data does not indicate the development patterns by line for various individual layers. Since the data indicates that excess business generally exhibits much slower reporting than that normally associated with primary business, there appears to be a relationship between the layer for which business is written and the resulting development pattern. It is this relationship that we intend to analyze in this paper for both paid and reported losses. Applications to increased limits and excess of loss pricing are also noted.

The protracted development of excess losses reflected in the RAA study suggests that the development is not only caused by late reported claims and increases in the average reported loss per claim but also by changes at successive maturities in the proportion of claims with losses which are large multiples of the average. Thus, the shape of the size of loss distribution changes at successive valuations. Accordingly, we requested and received from the Insurance Services Office various data comprising size of loss distributions at successive maturities. Specifically, included in the data were size of loss distributions of incurred losses, for policy year evaluations up to 99 months, or the latest evaluation, for policy years 1972 through 1982. This countrywide monoline data was provided separately for OL&T; M&C and Products with each size of loss distribution containing 118 intervals.

These size of loss distributions combine data from business written at different policy limits. Thus, the data includes losses censored at each of the policy limits. While no adjustments were made to this data, the implications of using combined limits data are discussed in Appendix B.

[1] "Incurred" is used in this study to mean the same as reported, i.e., it excludes IBNR.

Finally, the treatment of allocated loss adjustment expense in these distributions should be mentioned. Losses were assigned to a given size of loss interval based on reported loss size (paid plus outstanding) excluding allocated loss adjustment expenses. The total allocated loss adjustment expense associated with the losses in each interval was given separately. As loss adjustment expense is treated in different ways in excess reinsurance, the treatment of these expenses will be discussed further in the context of deriving excess development factors.

Size of loss distributions listing paid losses and outstanding losses separately, as well as paid and outstanding allocated loss adjustment expense separately, were also provided by ISO for OL&T and M&C. The latest valuation available with this policy year data was 63 months. The RAA study provides reported loss development data for over twenty years of development for general liability and other lines on an accident year basis.

2. INCURRED EXCESS LOSS DEVELOPMENT FACTORS

In this section, we will display and discuss the incurred excess loss development factors derived from the size of loss distributions.

In developing these factors, we adjusted the retentions for policy years prior to 1982 to recognize changing levels of average cost per occurrence. For policy year 1982, the retentions used were $10,000, $25,000, $50,000, $100,000, $250,000, $500,000 and $1,000,000. For prior policy years, these retentions were multiplied by relativities reflecting the average cost per occurrence for the given policy year relative to the average cost per occurrence for the 1982 year. (Although ISO has used higher trend for higher layers in determining increased limits factors, we did not find support for this procedure in the data provided. Higher trend for higher layers would produce a trend towards smaller maximum likelihood estimates of the Pareto parameter, but this is not the case, as shown in Reichle and Yonkunas [2].) Thus, the relativity for 1982 was 1.00, while for each prior policy year N it was computed by multiplying the relativity for the policy year $N + 1$ by the ratio of the average cost per occurrence for year N to the average cost per occurrence for year $N + 1$. The ratio was based on the latest available pair of reports at the same stage of development, excluding claims closed without payment. As the resulting deflated retentions did not correspond with endpoints of the 118 size of loss intervals, the closest possible endpoints were selected.

Allocated loss adjustment expense (ALAE) is handled in different ways in excess reinsurance contracts. The three most common treatments are as follows:

1) ALAE is added to the pure loss amount and the total is treated as one in determining coverage.
2) ALAE is assigned to an excess layer on a pro rata basis. That is, the ratio that the excess portion of the pure loss bears to the total loss is applied to the total ALAE to determine the excess ALAE.
3) ALAE is not included in the coverage.

Separate sets of excess loss development factors were calculated to reflect each of the above treatments of ALAE. This was done, respectively, as follows:

1) All ALAE on occurrences with pure loss greater than a given retention was included with the pure losses excess of that retention.
2) The total ALAE on occurrences for which the pure loss exceeded a given retention was multiplied by the ratio of the pure excess losses to the ground up losses on these same occurrences to determine the excess ALAE. This excess ALAE was then included with the pure excess losses.
3) No ALAE was added to the pure excess losses.

A discussion of the degree of accuracy of these methods of assigning ALAE can be found in Appendix A.

The factors shown in Exhibits 1 through 3 are dollar weighted averages of the factors by policy year. The retentions shown are retentions on policy year 1982 level, although they actually correspond to different retentions for different policy years. By estimating the factor for the increase in average cost per occurrence from policy year 1982 to accident year 1987, for example, one could bring the retentions to accident year 1987 level.

A review of the factors will show that the development is not materially affected after 39 months by the treatment of allocated loss adjustment expense. Therefore, future discussion will only deal with the case in which ALAE is included in the limit. This is probably the most common treatment in reinsurance, and it corresponds to the factors for excess losses plus ALAE. It is also clear from these factors that the development increases as the retention increases. Some exceptions to this trend occur at retentions of $500,000 and $1,000,000 for individual stages of development. This may be due to the fact that there is a lesser amount of data at these retentions which increases the variability of the factors. Despite the exceptions, these higher retentions tend to have the largest development factors.

EXHIBIT 1

OL&T BI DEVELOPMENT FACTORS

Excess Losses Plus ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.2113	1.1178	1.0682	1.0437	1.0504	1.0094
10,000	1.3356	1.1799	1.1056	1.0664	1.0710	1.0118
25,000	1.3849	1.2200	1.1402	1.0877	1.0909	1.0146
50,000	1.4055	1.2549	1.1764	1.1128	1.1134	1.0167
100,000	1.4021	1.2942	1.2168	1.1506	1.1424	1.0235
250,000	1.3512	1.3517	1.2963	1.2120	1.2015	1.0383
500,000	1.2742	1.3940	1.4080	1.2787	1.2626	1.0613
$1,000,000	1.0688	1.3061	1.6135	1.3662	1.3534	1.1111

Excess Losses Plus Pro Rata ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.2113	1.1178	1.0682	1.0437	1.0504	1.0094
10,000	1.3437	1.1870	1.1111	1.0695	1.0729	1.0127
25,000	1.3909	1.2291	1.1483	1.0926	1.0938	1.0160
50,000	1.4098	1.2655	1.1860	1.1189	1.1172	1.0191
100,000	1.4023	1.3070	1.2287	1.1573	1.1468	1.0264
250,000	1.3563	1.3611	1.3150	1.2180	1.2077	1.0446
500,000	1.2648	1.3957	1.4292	1.2838	1.2701	1.0684
$1,000,000	1.0503	1.3501	1.6417	1.3731	1.3576	1.1182

Excess Losses Only

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.2064	1.1185	1.0702	1.0458	1.0504	1.0115
10,000	1.3451	1.1940	1.1181	1.0735	1.0737	1.0155
25,000	1.3955	1.2389	1.1578	1.0981	1.0943	1.0193
50,000	1.4148	1.2777	1.1963	1.1249	1.1176	1.0239
100,000	1.4107	1.3191	1.2404	1.1626	1.1474	1.0319
250,000	1.3689	1.3690	1.3277	1.2199	1.2067	1.0517
500,000	1.2753	1.3981	1.4340	1.2832	1.2663	1.0740
$1,000,000	1.0316	1.3888	1.6258	1.3629	1.3504	1.1197

EXHIBIT 2

M&C BI Development Factors

Excess Losses Plus ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.4959	1.2077	1.0865	1.0297	1.0285	1.0210
10,000	1.6246	1.2630	1.1100	1.0401	1.0360	1.0267
25,000	1.6816	1.2974	1.1316	1.0513	1.0449	1.0319
50,000	1.7201	1.3280	1.1509	1.0642	1.0554	1.0382
100,000	1.7528	1.3583	1.1771	1.0788	1.0724	1.0491
250,000	1.7481	1.3775	1.2214	1.1008	1.1194	1.0782
500,000	1.6110	1.3845	1.2520	1.1340	1.1898	1.1192
$1,000,000	1.4056	1.5619	1.2130	1.1942	1.4206	1.2383

Excess Losses Plus Pro Rata ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.4959	1.2077	1.0865	1.0297	1.0285	1.0210
10,000	1.6326	1.2682	1.1128	1.0414	1.0375	1.0274
25,000	1.6909	1.3044	1.1354	1.0531	1.0475	1.0332
50,000	1.7297	1.3353	1.1556	1.0660	1.0594	1.0401
100,000	1.7689	1.3654	1.1828	1.0811	1.0789	1.0525
250,000	1.7652	1.3862	1.2306	1.1049	1.1267	1.0826
500,000	1.6093	1.4190	1.2534	1.1372	1.1993	1.1264
$1,000,000	1.4064	1.5551	1.1934	1.1901	1.4891	1.2350

Excess Losses Only

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.4865	1.2039	1.0838	1.0273	1.0300	1.0216
10,000	1.6294	1.2690	1.1136	1.0410	1.0410	1.0285
25,000	1.6933	1.3090	1.1367	1.0533	1.0519	1.0349
50,000	1.7368	1.3418	1.1587	1.0659	1.0649	1.0423
100,000	1.7835	1.3723	1.1871	1.0814	1.0858	1.0551
250,000	1.7878	1.3927	1.2346	1.1070	1.1300	1.0839
500,000	1.6334	1.4367	1.2555	1.1372	1.2014	1.1250
$1,000,000	1.4010	1.5516	1.1970	1.1846	1.5060	1.2276

EXHIBIT 3

PRODUCTS BI DEVELOPMENT FACTORS

Excess Losses Plus ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.6284	1.1974	1.1032	1.0545	1.0707	1.0332
10,000	1.7891	1.2906	1.1276	1.0632	1.0800	1.0293
25,000	1.9089	1.3561	1.1501	1.0776	1.0932	1.0369
50,000	1.9563	1.3844	1.1736	1.0928	1.1058	1.0405
100,000	2.0207	1.4221	1.1993	1.1165	1.1165	1.0421
250,000	2.1053	1.4790	1.2301	1.1453	1.0944	1.0440
500,000	2.3936	1.5098	1.4073	1.1660	1.1180	0.9605
$1,000,000	1.8026	1.5847	1.9141	1.2074	1.2271	0.7657

Excess Losses Plus Pro Rata ALAE

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.6284	1.1974	1.1032	1.0545	1.0707	1.0332
10,000	1.7995	1.3065	1.1302	1.0653	1.0812	1.0311
25,000	1.8940	1.3571	1.1538	1.0805	1.0939	1.0398
50,000	1.9255	1.3847	1.1777	1.0961	1.1053	1.0443
100,000	1.9550	1.4214	1.2041	1.1203	1.1135	1.0465
250,000	1.9284	1.4790	1.2514	1.1494	1.0924	1.0302
500,000	2.1034	1.5104	1.4556	1.1520	1.1271	0.9303
$1,000,000	1.7797	1.5970	1.9188	1.2199	1.2676	0.7245

Excess Losses Only

Retention	27–39	39–51	51–63	63–75	75–87	87–99
$ -0-	1.5635	1.1844	1.0958	1.0511	1.0636	1.0347
10,000	1.7291	1.2966	1.1266	1.0663	1.0758	1.0403
25,000	1.8118	1.3416	1.1505	1.0810	1.0885	1.0483
50,000	1.8340	1.3699	1.1752	1.0969	1.0993	1.0536
100,000	1.8344	1.4096	1.2034	1.1199	1.1081	1.0546
250,000	1.7100	1.4690	1.2601	1.1528	1.0942	1.0252
500,000	1.5748	1.5052	1.4556	1.1485	1.1267	0.9242
$1,000,000	1.4736	1.5162	1.9311	1.2105	1.2719	0.7226

The excess development factors shown were all derived directly from the underlying size of loss distributions. We now use these factors to estimate curves which, in addition to smoothing the underlying factors, will generate excess development factors beyond 99 months as well as for retentions other than those previously treated. This would be necessary for computing development factors at policy year 1982 retentions that are equivalent to various retentions at accident year 1987 level, for example.

For each development interval, a curve is estimated to fit the excess loss development factors as a function of retention. These curves are then fitted to a smoothly progressing series of curves. The procedure is done separately for each line of business.

The curve selected to fit the excess development factors as a function of retention was $y = ax^b$ where x is the retention divided by \$10,000. Thus, a is the value given by the curve for development excess of \$10,000.

The use of this function was motivated by the qualities of the single parameter Pareto distribution used to model size of loss distributions. This is discussed further in Section 4.

Separate curves of the form $y = a_n x^{b_n}$ were fit to the excess loss development factors by retention for each interval of development of the form 27 to 27 + $12n$ months, for $n = 1, 2, 3, 4, 5$ or 6. These intervals were used rather than individual successive intervals of development in order to stabilize the curve fitting process. Only retentions up to \$250,000 were used, since the data for larger retentions had much less credibility.

The a_n and b_n values were determined from the corresponding data points x,y by fitting the values of $\log y$ and $\log x$ to a least squares line which gives:

$$\log y = \log a_n + b_n \log x.$$

Thus, values for a_n and b_n were determined for each of the development intervals. These values were then separately fit to curves as a function of the stage of development. The method is illustrated in Exhibit 4 for the a_n values for M&C BI.

Thus, it is actually the values of $a_n' - 1$ that are fitted to the curve $y = cx^d$ to obtain the fitted values. Sherman [3] recommends this type of approach for fitting loss development factors. An exactly analogous procedure is used to obtain fitted b_n'' values. The formulas chosen to determine the fitted values a_n'' and b_n'' through 99 months are used to produce the tail beyond 99 months. In

The excess development factors shown were all derived directly from the underlying size of loss distributions. We now use these factors to estimate curves which, in addition to smoothing the underlying factors, will generate excess development factors beyond 99 months as well as for retentions other than those previously treated. This would be necessary for computing development factors at policy year 1982 retentions that are equivalent to various retentions at accident year 1987 level, for example.

For each development interval, a curve is estimated to fit the excess loss development factors as a function of retention. These curves are then fitted to a smoothly progressing series of curves. The procedure is done separately for each line of business.

The curve selected to fit the excess development factors as a function of retention was $y = ax^b$ where x is the retention divided by \$10,000. Thus, a is the value given by the curve for development excess of \$10,000.

The use of this function was motivated by the qualities of the single parameter Pareto distribution used to model size of loss distributions. This is discussed further in Section 4.

Separate curves of the form $y = a_n x^{b_n}$ were fit to the excess loss development factors by retention for each interval of development of the form 27 to 27 + $12n$ months, for $n = 1, 2, 3, 4, 5$ or 6. These intervals were used rather than individual successive intervals of development in order to stabilize the curve fitting process. Only retentions up to \$250,000 were used, since the data for larger retentions had much less credibility.

The a_n and b_n values were determined from the corresponding data points x,y by fitting the values of log y and log x to a least squares line which gives:

$$\log y = \log a_n + b_n \log x.$$

Thus, values for a_n and b_n were determined for each of the development intervals. These values were then separately fit to curves as a function of the stage of development. The method is illustrated in Exhibit 4 for the a_n values for M&C BI.

Thus, it is actually the values of $a_n' - 1$ that are fitted to the curve $y = cx^d$ to obtain the fitted values. Sherman [3] recommends this type of approach for fitting loss development factors. An exactly analogous procedure is used to obtain fitted b_n'' values. The formulas chosen to determine the fitted values a_n'' and b_n'' through 99 months are used to produce the tail beyond 99 months. In

EXHIBIT 5
OL&T BI Excess Loss & ALAE Development Factors
Fitted Factors

Development Interval	Fitted *b* Values	Retention 10,000*	25,000	50,000	100,000	250,000	500,000	1,000,000
27– 39	.01000	1.36556	1.37813	1.38771	1.39736	1.41023	1.42004	1.42991
39– 51	.03986	1.15206	1.19492	1.22839	1.26281	1.30978	1.34647	1.38420
51– 63	.05066	1.08024	1.13157	1.17202	1.21390	1.27158	1.31703	1.36410
63– 75	.03873	1.05099	1.08895	1.11858	1.14901	1.19051	1.22290	1.25617
75– 87	.02528	1.03587	1.06014	1.07889	1.09796	1.12370	1.14356	1.16378
87– 99	.01616	1.02691	1.04222	1.05396	1.06583	1.01873	1.09391	1.10623
99–111	.01055	1.02110	1.03102	1.03859	1.04622	1.05638	1.06414	1.07195
111–123	.00712	1.01710	1.02375	1.02882	1.03391	1.04068	1.04583	1.05100
123–135	.00497	1.01420	1.01882	1.02234	1.02586	1.03054	1.03409	1.03766
135–147	.00357	1.01203	1.01534	1.01785	1.02037	1.02372	1.02625	1.02879
147–159	.00263	1.01035	1.01279	1.01464	1.01649	1.01895	1.02081	1.02267
159–171	.00199	1.00902	1.01086	1.01226	1.01365	1.01550	1.01690	1.01830
171–183	.00153	1.00795	1.00937	1.01044	1.01152	1.01294	1.01401	1.01509
183–195	.00120	1.00708	1.00819	1.00903	1.00987	1.01098	1.01182	1.01267
195–207	.00096	1.00635	1.00723	1.00790	1.00857	1.00946	1.01013	1.01080
207–219	.00077	1.00573	1.00645	1.00699	1.00753	1.00824	1.00878	1.00933
219–231	.00063	1.00521	1.00579	1.00624	1.00668	1.00726	1.00770	1.00815
231–243	.00052	1.00476	1.00524	1.00561	1.00597	1.00646	1.00682	1.00719
243–255	.00044	1.00437	1.00477	1.00508	1.00538	1.00579	1.00609	1.00640
255–267	.00037	1.00403	1.00437	1.00463	1.00489	1.00523	1.00548	1.00574
267–279	.00031	1.00373	1.00402	1.00424	1.00446	1.00475	1.00497	1.00518
279–291	.00027	1.00347	1.00372	1.00390	1.00409	1.00434	1.00452	1.00471
291–303	.00023	1.00323	1.00345	1.00361	1.00377	1.00398	1.00414	1.00430
303–315	.00020	1.00302	1.00321	1.00335	1.00349	1.00367	1.00381	1.00395
315–327	.00018	1.00284	1.00300	1.00312	1.00324	1.00340	1.00352	1.00365
327–339	.00015	1.00267	1.00281	1.00291	1.00302	1.00316	1.00327	1.00338
339–351	.00014	1.00251	1.00264	1.00273	1.00282	1.00295	1.00304	1.00314
351–363	.00012	1.00237	1.00248	1.00257	1.00265	1.00276	1.00284	1.00293
		Actual Factors						
27– 39		1.33560	1.38490	1.40550	1.40210	1.35120	1.27420	1.06880
39– 51		1.17990	1.22000	1.25490	1.29420	1.35170	1.39400	1.30610
51– 63		1.10560	1.14020	1.17640	1.21680	1.29630	1.40800	1.61350
63– 75		1.06640	1.08770	1.11280	1.15060	1.21200	1.27870	1.36620
75– 87		1.07100	1.09090	1.11340	1.14240	1.20150	1.26260	1.35340
87– 99		1.01180	1.01460	1.01670	1.02350	1.03830	1.06130	1.11110
		Cumulative Comparison						
27– 99 Actual		2.01300	2.31900	2.61400	2.97100	3.58000	4.28500	4.62700
27– 99 Fitted		1.90000	2.24200	2.54100	2.88000	3.39900	3.85200	4.36600

* These equal the fitted *a* values.

EXHIBIT 6

M&C BI Excess Loss & ALAE Development Factors

Fitted Factors

Development Interval	Fitted *b* Values	Retention 10,000*	25,000	50,000	100,000	250,000	500,000	1,000,000
27– 39	.02402	1.64008	1.67658	1.70472	1.73334	1.77190	1.80165	1.83189
39– 51	.02784	1.25665	1.28913	1.31425	1.33986	1.37449	1.40127	1.42848
51– 63	.02666	1.09481	1.12188	1.14280	1.16412	1.19290	1.21515	1.23781
63– 75	.02266	1.04677	1.06874	1.08566	1.10285	1.12599	1.14382	1.16193
75– 87	.01867	1.02704	1.04476	1.05836	1.07214	1.09064	1.10484	1.11923
87– 99	.01534	1.01728	1.03168	1.04270	1.05385	1.06876	1.08018	1.09173
99–111	.01270	1.01183	1.02367	1.03272	1.04185	1.05405	1.06337	1.07277
111–123	.01063	1.00852	1.01839	1.02592	1.03351	1.04362	1.05133	1.05911
123–135	.00899	1.00638	1.01471	1.02106	1.02744	1.03594	1.04242	1.04894
135–147	.00769	1.00493	1.01204	1.01745	1.02289	1.03012	1.03563	1.04117
147–159	.00665	1.00390	1.01003	1.01470	1.01938	1.02561	1.03034	1.03510
159–171	.00579	1.00315	1.00849	1.01255	1.01662	1.02203	1.02614	1.03027
171–183	.00509	1.00259	1.00728	1.01084	1.01441	1.01915	1.02276	1.02637
183–195	.00451	1.00216	1.00630	1.00945	1.01261	1.01680	1.01998	1.02317
195–207	.00402	1.00182	1.00551	1.00832	1.01113	1.01486	1.01769	1.02052
207–219	.00360	1.00155	1.00486	1.00737	1.00989	1.01323	1.01576	1.01830
219–231	.00325	1.00134	1.00432	1.00658	1.00885	1.01185	1.01413	1.01642
231–243	.00294	1.00116	1.00386	1.00591	1.00796	1.01068	1.01274	1.01481
243–255	.00267	1.00102	1.00347	1.00534	1.00720	1.00967	1.01155	1.01342
255–267	.00244	1.00090	1.00314	1.00484	1.00655	1.00880	1.01051	1.01223
267–279	.00224	1.00080	1.00285	1.00441	1.00597	1.00804	1.00961	1.01118
279–291	.00206	1.00071	1.00260	1.00404	1.00547	1.00738	1.00882	1.01026
291–303	.00190	1.00064	1.00238	1.00371	1.00503	1.00679	1.00812	1.00945
303–315	.00176	1.00057	1.00219	1.00342	1.00465	1.00627	1.00750	1.00873
315–327	.00164	1.00052	1.00202	1.00316	1.00430	1.00581	1.00695	1.00809
327–339	.00153	1.00047	1.00187	1.00293	1.00399	1.00539	1.00646	1.00752
339–351	.00142	1.00043	1.00174	1.00272	1.00371	1.00502	1.00602	1.00701
351–363	.00133	1.00039	1.00161	1.00254	1.00347	1.00469	1.00562	1.00655
		Actual Factors						
27– 39		1.62460	1.68160	1.72010	1.75280	1.74610	1.61100	1.40560
39– 51		1.26300	1.29740	1.32800	1.35830	1.37750	1.38450	1.56190
51– 63		1.11000	1.13160	1.15090	1.17710	1.22140	1.25200	1.21300
63– 75		1.04010	1.05130	1.06420	1.07880	1.10080	1.13400	1.19420
75– 87		1.03600	1.04490	1.05540	1.07240	1.11940	1.18980	1.42060
87– 99		1.02670	1.03190	1.03820	1.04910	1.07820	1.11920	1.23830
		Cumulative Comparison						
27– 99 Actual		2.52000	2.79900	3.06600	3.40100	3.90800	4.21700	5.59400
27– 99 Fitted		2.46800	2.79300	3.06800	3.36900	3.81300	4.18800	4.59900

* These equal the fitted *a* values.

EXHIBIT 7

Products BI Excess Loss & ALAE Development Factors
Fitted Factors

Development Interval	Fitted *b* Values	Retention 10,000*	25,000	50,000	100,000	250,000	500,000	1,000,000
27– 39	.04877	1.80564	1.88815	1.95307	2.02022	2.11254	2.18517	2.26030
39– 51	.04373	1.27527	1.32740	1.36825	1.41036	1.46802	1.51320	1.55977
51– 63	.02738	1.13277	1.16155	1.18381	1.20649	1.23715	1.26086	1.28502
63– 75	.01617	1.07914	1.09525	1.10759	1.12007	1.36791	1.14960	1.16256
75– 87	.00997	1.05298	1.06265	1.07002	1.07744	1.08733	1.09487	1.10246
87– 99	.00650	1.03817	1.04438	1.04909	1.05383	1.06013	1.06492	1.06973
99–111	.00446	1.02893	1.03314	1.03634	1.03954	1.04380	1.04703	1.05027
111–123	.00318	1.02275	1.02574	1.02801	1.03028	1.03329	1.03557	1.03786
123–135	.00235	1.01841	1.02061	1.02228	1.02395	1.02616	1.02784	1.02951
135–147	.00179	1.01523	1.01690	1.01816	1.01943	1.02110	1.02237	1.02364
147–159	.00140	1.01283	1.01413	1.01511	1.01609	1.01739	1.01838	1.01937
159–171	.00111	1.01097	1.01200	1.01278	1.01356	1.01459	1.01537	1.01616
171–183	.00090	1.00950	1.01033	1.01096	1.01159	1.01242	1.01306	1.01369
183–195	.00074	1.00832	1.00900	1.00951	1.01003	1.01071	1.01123	1.01175
195–207	.00061	1.00735	1.00791	1.00834	1.00877	1.00934	1.00977	1.01020
207–219	.00052	1.00654	1.00702	1.00738	1.00774	1.00821	1.00858	1.00894
219–231	.00044	1.00587	1.00627	1.00658	1.00688	1.00729	1.00759	1.00790
231–243	.00038	1.00529	1.00564	1.00590	1.00616	1.00651	1.00677	1.00704
243–255	.00033	1.00480	1.00510	1.00533	1.00556	1.00585	1.00608	1.00631
255–267	.00028	1.00438	1.00464	1.00484	1.00503	1.00530	1.00549	1.00569
267–279	.00025	1.00401	1.00424	1.00441	1.00459	1.00481	1.00499	1.00516
279–291	.00022	1.00369	1.00389	1.00404	1.00420	1.00440	1.00455	1.00470
291–303	.00019	1.00341	1.00358	1.00372	1.00385	1.00403	1.00417	1.00430
303–315	.00017	1.00316	1.00331	1.00343	1.00355	1.00371	1.00383	1.00395
315–327	.00015	1.00293	1.00307	1.00318	1.00329	1.00343	1.00354	1.00365
327–339	.00014	1.00273	1.00286	1.00296	1.00305	1.00318	1.00328	1.00338
339–351	.00013	1.00255	1.00267	1.00276	1.00284	1.00296	1.00305	1.00313
351–363	.00011	1.00239	1.00250	1.00258	1.00265	1.00276	1.00284	1.00292
		Actual Factors						
27– 39		1.78910	1.90890	1.95630	2.02070	2.10530	2.39360	1.80260
39– 51		1.29060	1.35610	1.38440	1.42210	1.47900	1.50980	1.58470
51– 63		1.12670	1.15010	1.17360	1.19930	1.23010	1.40730	1.91410
63– 75		1.06320	1.07760	1.09280	1.11650	1.14530	1.16600	1.20740
75– 87		1.08000	1.09320	1.10580	1.11650	1.09440	1.11800	1.22710
87– 99		1.02930	1.03690	1.04050	1.04210	1.04400	.96050	.76570
		Cumulative Comparison						
27– 99 Actual		3.07700	3.63700	3.99600	4.47700	5.01200	6.36800	6.20300
27– 99 Fitted		3.07700	3.53900	3.93300	4.37200	5.02800	5.58800	6.21100

* These equal the fitted *a* values.

As has been mentioned, the RAA Loss Development Study combines business written at various retentions. The subline mix underlying the "General Liability Excluding Asbestos" experience is also difficult to estimate. For these reasons, as well as the fact that the RAA experience is accident year, it is difficult to make a precise comparison of our results with those of the RAA. Nevertheless, Exhibit 8 shows a rough comparison based on the following assumptions:

1) A retention of $250,000 is used to reflect the development characteristics of the various retentions and limits underlying the RAA experience.
2) An equal weighting of the excess loss development factors for OL&T, M&C and Products is used to approximate the subline mix of the RAA data.
3) A weighting of 25% of the accident year factor from $12 + 12k$ months to $12 + 12(k + 1)$ months and 75% of the accident year factor from $12 + 12(k + 1)$ months to $12 + 12(k + 2)$ months is used to estimate the policy year factor from $27 + 12k$ months to $27 + 12(k + 1)$ months.
4) Dollar weighted factors are derived using the most recent five years of RAA experience.

EXHIBIT 8

DEVELOPMENT FACTOR COMPARISON

Development Interval	Fitted ISO Data Excess $250,000	RAA (as of 12/84)
27–39	1.765	1.801
39–51	1.384	1.392
51–63	1.234	1.242
63–75	1.151	1.153
75–87	1.101	1.097
87–99	1.070	1.072
99–111	1.051	1.067
111–123	1.039	1.049
123–135	1.031	1.038
135–147	1.025	1.038
147–159	1.021	1.030
159–171	1.017	1.029
171–183	1.015	1.036
183–Ult.	1.105	1.228

The RAA data begins to show higher developments than the curves fitted to ISO data after 99 months. This could be partially due to the effects of reinsurance coverage on an aggregate basis showing up later in the development. Also, the RAA study points out that unidentified longer tailed medical malpractice losses are present in the RAA data, particularly in the older years. This could have a great effect on development at later valuations. It is also possible that the distribution of RAA retentions and limits results in larger development at later stages relative to earlier stages than the development associated with the fixed ISO retention. Higher layer losses have more relative weight at later stages since they develop more slowly. The RAA data, unlike the ISO data, includes excess and surplus lines and umbrella business written with large policy limits. Finally, as mentioned, the curves chosen to fit the ISO data through 99 months are used to produce the tail beyond 99 months. The RAA development, despite its limitations, is based on actual data at all maturities.

It is possible, if so desired, to calculate development factors by retention beyond 99 months that are more consistent with the RAA factors. One simple method is as follows. Suppose the OL&T, M&C, and Products factors for a retention of $250,000 are $1 + a$, $1 + b$ and $1 + c$, respectively, and the RAA factor is $1 + d$. Solve for x such that $(a + b + c)x \div 3 = d$ and let $1 + ax$, $1 + bx$ and $1 + cx$ be the OL&T, M&C, and Products factors for a $250,000 retention. (This is based on the approximation that the 3 sublines comprise equal portions of the RAA data.) Then use the fitted factors by subline for a retention of $10,000 to solve for the b value using $y = ax^b$. Factors at other retentions can then be calculated.

In calculating adjusted development factors at other retentions, this method assumes the fitted factors at the $10,000 retention are accurate. The lower development of the $10,000 retention, as well as the substantial amount of data available for determining factors at the $10,000 retention, support this as a reasonable method. This method operates identically for producing factors to ultimate as for age-to-age factors.

Commercial Auto Liability

The commercial auto liability study was based on a total of almost $4 billion in losses from accident years 1980, 1981 and 1982. These were the only years available to us and our study is of the only available development factors: 21 to 33, 33 to 45, and 45 to 57 months.

The development factors for losses plus ALAE excess of various retentions (on an accident year 1982 level) are:

Retention	21–33	33–45	45–57	21–57	33–57
-0-	1.084	1.031	1.011	1.130	1.042
10,000	1.137	1.044	1.012	1.201	1.057
25,000	1.152	1.050	1.014	1.227	1.065
50,000	1.159	1.053	1.016	1.240	1.070
100,000	1.172	1.058	1.013	1.256	1.072
250,000	1.177	1.030	1.043	1.264	1.074
500,000	1.444	.949	1.168	1.601	1.108

A pattern of increasing development with increasing retentions can be observed, especially in the 21–57 month factors. The factors for the $500,000 retention have limited credibility. Due to the small change in development factors from one retention to another, no curve fitting was performed.

The breakdown of premium by policy limits for accident year 1982 can be approximated as 5% at $100,000, 15% at $300,000, 60% at $500,000, and, 20% at $750,000 or $1,000,000.

Accident year development factors for excess losses based on a weighted average of RAA development data for the last five years as of 12/31/84 for auto liability are:

12–24	24–36	36–48	48–60	60–72	72–84	84–ultimate
1.804	1.204	1.093	1.062	1.052	1.026	1.076

3. EXCESS PAID LOSS & ALAE DEVELOPMENT

In this section, ratios of excess paid losses and ALAE to excess incurred losses and ALAE were determined at policy year valuations from 27 months to ultimate for OL&T BI and M&C BI. (Sufficient data was not available for Products BI.) These ratios of paid to reported, in conjunction with excess incurred loss and ALAE development, will produce excess paid loss and ALAE development factors.

The procedure previously discussed which was used in developing excess incurred losses and ALAE by retention at various valuations was used for both paid and reported losses and ALAE from 27 months to 63 months of development. The resulting ratios of paid to reported are shown in Exhibit 9 for policy year 1982 cost levels.

EXHIBIT 9

RATIO OF PAID TO REPORTED EXCESS LOSS AND ALAE

OL&T BI

Retention	27 mo.	39 mo.	51 mo.	63 mo.
$ 10,000	.1937	.3587	.5041	.6356
25,000	.1616	.3217	.4634	.5964
50,000	.1518	.3080	.4469	.5754
100,000	.1585	.3210	.4519	.5838
250,000	.1852	.3616	.4919	.5640
$500,000	.2269	.3103	.5106	.4205

M&C BI

Retention	27 mo.	39 mo.	51 mo.	63 mo.
$ 10,000	.1417	.2427	.4098	.5350
25,000	.1425	.2358	.4069	.5294
50,000	.1526	.2364	.4054	.5233
100,000	.1751	.2473	.4142	.5279
250,000	.2312	.2924	.4464	.5094
$500,000	.2209	.3586	.4285	.4794

It appears that the paid-to-reported ratios shown for excess loss and ALAE do not vary meaningfully as a function of the retention. Accordingly, we selected the paid-to-reported ratios for loss and ALAE excess of $25,000 as characteristic of the various retentions shown in producing a development pattern of paid-to-reported ratios. It should be noted that small losses exhibit significantly higher paid-to-reported ratios than those shown for the retentions above.

The following ISO excess of $25,000 loss development data was available beyond 63 months for loss and ALAE combined.

OL&T BI Ratios

(1) Months	(2) Paid to Reported	(3) Outstanding to Reported	(4) Ratio of (3) to Prior Value of (3)
63	.5710	.4290	—
75	.6809	.3191	.7438
87	.7768	.2232	.6995
99	.8717	.1283	.5748

M&C BI Ratios

(1) Months	(2) Paid to Reported	(3) Outstanding to Reported	(4) Ratio of (3) to Prior Value of (3)
63	.5660	.4340	—
75	.7091	.2909	.6703
87	.8019	.1981	.6810
99	.8680	.1320	.6663

In light of the column (4) ratios, and the fact that the outstanding to reported ratio will ultimately reach zero, a factor of .67 was selected judgmentally to be repeatedly applied to the outstanding to reported ratios at 63 months. The resulting patterns of paid to reported excess loss and ALAE are shown on Exhibit 10.

EXHIBIT 10

ISO Excess of $25,000 Loss Development Data
Ratios of Paid to Reported Excess Loss and ALAE

OL&T BI		M&C BI	
Valuation	Ratio	Valuation	Ratio
27	.1616	27	.1425
39	.3217	39	.2358
51	.4634	51	.4069
63	.5964	63	.5294
75	.7296	75	.6847
87	.8188	87	.7887
99	.8786	99	.8585
111	.9187	111	.9052
123	.9455	123	.9365
135	.9635	135	.9574
147	.9755	147	.9715
159	.9836	159	.9809
171	.9890	171	.9872
183	.9926	183	.9914
Ult.	1.0000	Ult.	1.0000

Excess paid to reported ratios have been used thus far since they vary less by retention and valuation than paid development factors. Also, they allow for the use of the more extensive reported data in estimating paid development. Excess paid loss and ALAE development factors can be determined simply by multiplying each reported loss development factor linking two valuations by the quotient of the paid to reported ratios for the later and earlier valuations. For example, the estimated paid loss development factors for loss and ALAE excess of $100,000 are as follows (see Exhibits 5 and 6).

OL&T BI		M&C BI	
27– 39	2.7817	27– 39	2.8682
39– 51	1.8190	39– 51	2.3121
51– 63	1.5623	51– 63	1.5146
63– 75	1.4056	63– 75	1.4264
75– 87	1.2322	75– 87	1.2351
87– 99	1.1437	87– 99	1.1470
99–111	1.0940	99–111	1.0985
111–123	1.0641	111–123	1.0692
123–135	1.0454	123–135	1.0504
135–147	1.0331	135–147	1.0379
147–159	1.0249	147–159	1.0293
159–171	1.0192	159–171	1.0232
171–183	1.0152	171–183	1.0188
183–Ult.	1.0872	183–Ult.	1.1152

4. RELATION OF RESULTS TO THE SINGLE PARAMETER PARETO DISTRIBUTION

It has been seen that excess loss development increases as the retention increases. A perspective on this relationship, and excess loss development in general, can be obtained by considering a model that illustrates the two influences underlying loss development:

1) The reporting pattern of claims over time.
2) The changing characteristics of the size of loss distribution at successive reports.

Without the latter influence, the development factors for losses excess of different retentions would be identical.

It has been noted, by both Philbrick [1] and Reichle and Yonkunas [2], that the single parameter Pareto distribution fits the tail of casualty loss distributions fairly well (at least if the interval of loss sizes is not too long), and that the parameter tends to decrease at successive stages of development. This motivated our use of the curve ax^b to fit loss development factors as a function of the retention x, as explained below.

If a series of Pareto distributions with parameters that are decreasing and greater than one were to perfectly represent a series of actual tails of loss distributions at successive development stages, the excess loss development factor from any stage m to stage $m + n$ $(n > 0)$ for retention x (where x is big enough to be included in the tail) would increase as x increased, since it equals ax^b for some fixed $a > 0$ and $b > 0$. The proof follows.

If k is the lower bound of the tail which is represented by a Pareto distribution with parameter q, and x represents the size of loss divided by k, then the density function $qx^{-(q+1)}$, as x ranges from one to infinity, represents the "normalized" (i.e., divided by k) loss distribution. The probability of a loss greater than k being between ak and bk equals $\int_a^b qx^{-(q+1)}dx$, and the losses excess of a retention ck are $nk\int_c^\infty (x - c)qx^{-(q+1)}dx$, where n is the number of losses greater than k. If the distribution of losses greater than k at i^{th} report is represented by a Pareto with parameter q_i, and at j^{th} report $(j > i)$ by a Pareto with parameter q_j, and the numbers of losses greater than k at i^{th} and j^{th} report are n_i and n_j, then the development factor for losses excess of ck from i^{th} to j^{th} report equals

$$\frac{n_j}{n_i}\left(\frac{q_i - 1}{q_j - 1}\right) c^{q_i - q_j}.$$

Therefore, if d is the development factor from i^{th} to j^{th} report for losses excess of k, then $dy^{q_i - q_j}$ is the development factor for losses excess of yk (for $y > 1$).

The development factor for losses excess of x, where $x > k$, is thus

$d\left(\frac{x}{k}\right)^{q_i - q_j}$, which equals $\frac{d}{k^{q_i - q_j}}x^{q_i - q_j}$, and $\frac{d}{k^{q_i - q_j}} > 0$ and $q_i - q_j > 0$.

This completes the proof.

The term $\frac{n_j}{n_i}$ in the expression $\frac{n_j}{n_i}\left(\frac{q_i - 1}{q_j - 1}\right) c^{q_i - q_j}$ represents the development due to additional reportings greater than k. The term $(q_i - 1)/(q_j - 1)$ represents the development arising from the change in the average excess loss above ck for occurrences greater than ck. The term $c^{q_i - q_j}$ reflects the development arising from the increased proportion of occurrences greater than k which are also greater than ck, resulting from the changing shape of the distribution. It can be seen that $c^{q_i - q_j}$ is the only term affected by a change in the retention.

As an example, let:

k = the lower bound of the tail = 25,000;
x = the primary retention = 100,000;
q_1 = the Pareto parameter for 1st report tail losses = 1.75;
q_{10} = the Pareto parameter for 10th report tail losses = 1.25; and,
d = the 1st to 10th development factor for losses excess of \$25,000 = 2.5.

Then the 1st to 10th development factor for losses excess of 100,000 is given by the formula

$$d\left(\frac{x}{k}\right)^{q_i - q_j}, \text{ i.e., } 2.5\ (4)^{.5} = 5.0.$$

Philbrick [1] and Reiche and Yonkunas [2] also noted that when a Pareto is fitted to a distribution of casualty losses greater than some amount k, the tail of the Pareto is thicker than the tail of the empirical loss distribution at very large loss sizes. Nevertheless, the effect of this error is lessened in using a ratio to estimate a development factor if the error is similar in the numerator and denominator.

5. DEVELOPMENT FACTORS BY LAYER, EXCESS LOSS RATIOS, AND INCREASED LIMITS FACTORS

The following method is used to produce development factors by layer, where the layer of losses from a to b is defined as the total of the portions between a and b of every loss. By applying the excess age-to-ultimate loss development factors to the latest available excess losses for each retention for each policy year, we get projected ultimate excess losses for each retention for each policy year. We also have "ground up" development factors, based on the same data, with which we project ultimate ground up losses for each policy year. The ground up age-to-ultimate factors are derived by fitting a curve $1 + ax^b$ to the factors through 99 months.

By taking weighted averages of the ratios of ultimate excess losses to ultimate ground up losses for all policy years for the retentions (in 000's) 10, 25, 50, 100, 250, 500, and 1000, we get ratios that we call $f(10)$, $f(25)$, $f(50)$, $f(100)$, $f(250)$, $f(500)$ and $f(1000)$. An exponential curve could then be fit between any two successive data points to get intermediate values of $f(x)$. This curve gives estimates of the ratios of ultimate excess losses to ultimate ground up losses for each retention. In order to produce the n^{th}-to-ultimate development factor for the layer from c to d, we first divide the curve values $f(c)$ and $f(d)$ by the n^{th} to ultimate development factors for losses excess of c and d, respectively, to get estimates $e_{c,n}$ and $e_{d,n}$ of the ratios of n^{th} report excess losses, for retentions c and d, to ultimate ground up losses.

We then let the development from n^{th}-to-ultimate for the layer from c to d equal $(f(c) - f(d)) \div (e_{c,n} - e_{d,n})$, i.e., the estimated ultimate excess losses in the layer divided by the n^{th} report excess losses in the layer. The n^{th} to

$(n + 1)^{th}$ development factor for a layer is produced by dividing the n^{th} to ultimate factor by the $(n + 1)^{th}$-to-ultimate factor.

The values of $f(x)$ (x is in \$000's) given by the data and derived development factors for losses and ALAE are:

	OL&T BI	M&C BI	Products BI
$f(10)$	.677	.802	.835
$f(25)$	.579	.755	.735
$f(50)$	.484	.674	.617
$f(100)$	.372	.543	.463
$f(250)$	.240	.319	.243
$f(500)$	.144	.148	.125
$f(1,000)$	.076	.041	.032

The OL&T development factors for 27 months to ultimate for retentions of (in 000's) 50, 100, 250, 500 and 1,000 are 3.150, 3.668, 4.485, 5.223, and 6.081, respectively. The factors for the layers 50–100, 50–250, 50–500, and 50–1,000, using the above method, follow:

Layer (in \$000's)	Method and Development Factor
50–1,000	$(.484 - .076) \div ((.484 \div 3.150) - (.076 \div 6.081)) = 2.891$
50– 500	$(.484 - .144) \div ((.484 \div 3.150) - (.144 \div 5.223)) = 2.697$
50– 250	$(.484 - .240) \div ((.484 \div 3.150) - (.240 \div 4.485)) = 2.437$
50– 100	$(.484 - .372) \div ((.484 \div 3.150) - (.372 \div 3.668)) = 2.144$

As with our unlimited development factors by retention, these factors for layers are somewhat lower than the factors would be for losses uncensored by policy limits. (See Appendix B.) Since about 80% of the losses are not censored by policy limits below \$500,000, the factors produced by the above method are more accurate for layers whose upper bound does not exceed \$500,000. The techniques of producing different development factors by retention or layer and projecting development to ultimate could be useful in estimating ultimate uncensored excess loss ratios, which are important in reinsurance pricing. The techniques could also be used in producing increased limits factors, which are an important part of primary insurance pricing. The actual development factors and data from this study concerning excess losses by layer could provide estimates of increased limits factors up to \$100,000 or possibly \$250,000 limits, since the policy limits in effect have little effect on the layer up to $100,000, or even $250,000. We do not present such estimates, however.

6. SUMMARY

The results that have been produced indicate clearly that loss and ALAE development varies significantly by retention. Accordingly, pricing and reserving estimates incorporating development factors may be substantially in error if this is not taken into account. As this applies to paid as well as reported loss development, recognition of retention is also a major factor in estimating discounted losses using paid development factors.

The protracted development of excess losses and the data limitations inherent in this study suggest a need for further study of development factors beyond 99 months. It would also be beneficial to review development by retention for other lines of business such as medical malpractice and workers' compensation.[2]

The results are closely related to the decrease in the Pareto parameter in successive reports and its relationship to loss development by retention. The principles employed would have relevance for other lines for which the Pareto provides a good fit.

With sufficient data, it would be very worthwhile to study excess development for uncensored losses and for higher retentions than those examined here.

[2] Study of New York data appears in Taylor and Lattanzio [4].

REFERENCES

[1] Stephen W. Philbrick, "A Practical Guide to the Single Parameter Pareto Distribution," *PCAS* LXXII, 1985, p. 44.

[2] Kurt A. Reichle and John P. Yonkunas, Discussion of "A Practical Guide to the Single Parameter Pareto Distribution," *PCAS* LXXII, 1985, p. 85.

[3] Richard E. Sherman, "Extrapolating, Smoothing, and Interpolating Development Factors," *PCAS* LXXI, 1984, p. 123.

[4] Frank C. Taylor and Francis Lattanzio, Discussion of "Accident Limitations for Retrospective Rating," *PCAS* LXIV, 1977, p. 96.

APPENDIX A
TREATMENT OF ALAE IN ESTIMATING DEVELOPMENT FACTORS

The type of occurrence excess coverage that is most common in casualty treaty reinsurance covers the amount of the loss and allocated loss adjustment expense combined in excess of the retention for each occurrence. The method of estimating the development factors for this type of reinsurance, however, was based on the development of the amount of the loss and allocated loss adjustment expense combined in excess of the retention for only those occurrences for which the pure loss exceeded the retention.

The error involved in using this approach is relatively small, since the amount in excess of any retention that is produced by the losses plus ALAE for all occurrences for which the losses alone are less than the retention is small compared to the total losses plus ALAE in excess of the retention. In other words, only a small portion of the excess is missing from our development factors.

Suppose, for example, that for every occurrence, the ratio of the loss to the loss plus ALAE is a. If the tail of the "normalized" (see Section 4) loss distribution is represented by the Pareto density function $qx^{-(q+1)}$, with $q > 1$, then the portion of the total losses plus ALAE in excess of the retention x_0 that is produced by occurrences for which the pure loss is greater than the retention equals

$$\int_{x_0}^{\infty} qx^{-(q+1)}\left(\frac{x}{a} - x_0\right) \mathrm{dx} \div \int_{ax_0}^{\infty} qx^{-(q+1)}\left(\frac{x}{a} - x_0\right) \mathrm{dx},$$

which equals $(q + a - qa)/(a^{1-q})$. If $q = 1.5$ and $a = .87$, for example, then the above expression equals .993.

If $q = 1.5$ and $a = .87$ at first report and $q = 1.3$ and $a = .85$ at ultimate report, then the expression changes from .993 to .995. In this case, the estimate of the first to ultimate development factor would be 1.002 times the development that would be computed using a precise treatment of ALAE.

This problem does not apply to the development factors for losses plus prorated ALAE, since occurrences with pure losses below the retention are not covered by reinsurance arrangements with prorated ALAE. Those factors involve a different estimate—use of losses excess of a retention divided by total losses for the occurrences greater than the retention—as a multiplier for the ALAE. To be precise, the ALAE for each occurrence should be multiplied by the loss excess of the retention divided by the total loss for that occurrence.

The distortion in development factors should be small, even in the product of all the development factors. For each loss and corresponding ALAE and each retention, prorated ALAE = (excess loss ÷ loss) ALAE, so prorated ALAE ÷ excess loss = ALAE ÷ loss for each loss. Since the data indicated that ALAE ÷ loss is about .15 on the average, whatever distortion there is in the estimate of the prorated ALAE would cause less than .15 times as much distortion in losses plus prorated ALAE.

APPENDIX B
EFFECT OF POLICY LIMITS ON DEVELOPMENT FACTORS

The general liability sublines studied had policy limits distributions based on policy year 1982 and policy year 1983 data.

Distribution of Premium

Policy Limit (in $000's)	OL&T BI	M&C BI	Products BI
25	.0043	.0034	.0018
50	.0069	.0031	.0042
100	.0366	.0347	.0248
200	.0022	.0010	.0000
250	.0013	.0032	.0025
300	.1351	.1367	.1792
500	.4161	.5334	.6464
1,000	.3609	.2464	.1354
1,500	.0043	.0027	.0005
2,000	.0191	.0136	.0019
3,000	.0132	.0218	.0033
Total	1.0000	1.0000	1.0000

As an illustration of the approximate effect of these policy limits on excess loss development factors, consider the following example of their effect on an unlimited (no policy limits) loss distribution. Let $10,000 be the lower bound of a tail of unlimited losses for which the "normalized" (divided by 10,000) loss distribution is represented by the Pareto density function $qx^{-(q+1)}$.

Let $q = 1.6$ for a policy year as of 27 months and 1.3 for a policy year at ultimate development, and let a represent the development factor from 27 months to ultimate for losses excess of $10,000.

Since $b^{(1-q)} \div (q - 1)$ is the formula for the losses excess of bk, normalized at k and divided by the number of occurrences greater than k, the unlimited losses excess of \$10,000, \$100,000, \$300,000, \$500,000 and \$1,000,000 at 27 months and at ultimate development can be represented as:

Retention	Excess at 27 months	Excess at Ultimate
\$ 10,000	x	ax
100,000	$.251x$	$.501ax$
300,000	$.130x$	$.360ax$
500,000	$.096x$	$.309ax$
\$1,000,000	$.063x$	$.251ax$

From this, the excess losses can be divided into the following layers, by subtracting from each excess amount the amount directly below it:

Layers (in \$000's)	Amount at 27 months	Amount at Ultimate
100–300	$.121x$	$.141ax$
300–500	$.034x$	$.051ax$
500–1000	$.033x$	$.058ax$
over 1000	$.063x$	$.251ax$

Now suppose that the policy limits earned premium distribution corresponding to the time period of the losses is 20% at \$300,000 (per occurrence), 60% at \$500,000, and 20% at \$1,000,000, instead of the losses being unlimited.

The development of the unlimited losses excess of \$100,000 from 27 months to ultimate $= (.501\ ax) \div (.251\ x) = 1.996\ a$, whereas the development of the limited losses $= (.141\ ax + .8(.051\ ax) + .2(.058\ ax)) \div (.121x + .8(.034x) + .2(.033x)) = 1.252a$. This is a big difference, but we should consider that the development factor for the losses limited only by \$500,000 limits $= (.141ax + .051ax) \div (.121x + .034x) = 1.239a$ and that the development factor for the losses limited only by \$1,000,000 limits $= (.141ax + .051ax + .058ax) \div (.121x + .034x + .033x) = 1.330a$. Thus, the limited development is not that different from the development of losses limited only at \$500,000 or only at \$1,000,000. If $a = 3$, which is not unreasonable, then $1.252a = 3.756$, $1.239a = 3.717$, and $1.330a = 3.990$. For retentions less than \$100,000, the difference between these types of development factors is less, since the portion below \$100,000 is not affected by the limits. Similarly, the development factors for losses excess of \$300,000 from 27 months to ultimate for unlimited losses, limited losses, losses limited only at \$500,000 and losses limited only at

\$1,000,000 are 2.769a, 1.559a, 1.500a, and 1.627a, respectively. The development factors for losses excess of \$500,000 are the same for the given policy limit distribution as for losses limited only at \$1,000,000.

For simplicity, we have considered only one policy year rather than a series of policy years with inflation operating on both average cost per occurrence and the average policy limit. But it seems probable that the development factors for retentions up to amounts corresponding to \$500,000 on a 1982 cost level, using actual limited losses for any policy year prior to 1982, are similar to development factors for losses limited only by any single limit which is between amounts corresponding to \$500,000 and \$1,000,000 on a policy year 1982 level. The development factors for limited losses are considerably different from unlimited development factors, but only a small portion of premium is written at policy limits over \$1,000,000, so development factors for limited losses are very useful. Also, the substantial disparity between limited and unlimited losses would be expected given the excessive thickness of the Pareto tail at extremely large loss amounts.

DISCUSSION BY GEORGE M. LEVINE

1. INTRODUCTION

Messrs. Pinto and Gogol have written a paper rich with practical techniques for determining excess loss development by layer of loss for liability lines. I have used their novel approach for analyzing reporting patterns by liability layer, and had success in tailoring their patterns to determine expected development for various reinsurance programs. Before presenting my results, I will summarize their technique and present some goodness of fit tests comparing the actual data to their fitted curves. In addition, some limitations of the use of their method will be offered.

2. SUMMARY OF PINTO/GOGOL TECHNIQUE

The authors begin by describing the lack of available published information by layer for reported and paid excess loss development. Although the Reinsurance Association of America (RAA) publishes accident year reported loss development studies every two years, and the Insurance Services Office (ISO) annually distributes policy year reported loss development patterns, empirical loss detail by layer is generally not available. This lack of published data dictates the use of theoretical loss distributions (like the Pareto distribution). The Pinto/Gogol technique, although theoretically supported by the properties of the Pareto distribution, has the advantage of being applicable to empirical data.

From ISO excess loss development data by subline, Pinto/Gogol smooth the data two ways—by liability limit (retention), and by development interval. From the "Actual Factors" matrices that they present in Exhibits 1 through 3, they initially smooth the data horizontally by retention, for the monthly intervals 27–99, 39–99, 51–99, 63–99, 75–99, and 87–99. The curve selected to fit the data is $y = ax^b$, where x is the retention divided by \$10,000. Next, the authors convert the factors by retention back to the report-to-report intervals (27–39, 39–51, etc.), and smooth the data vertically (by retention) using a normal power approximation developed by Sherman [1].

By fitting the curve ax^b to the excess loss development within each development interval by layer, the authors have provided an easy method to determine development for layers of retention not published (for example, \$75,000 and \$150,000). In Section 4, the authors explain that the motivation of the selection of the curve ax^b to fit loss development factors was the single parameter Pareto distribution's good fit to the tail of casualty loss distributions. Another interesting parallel is the technique that Rosenberg and Halpert applied regarding their analysis of methods to adjust historical loss distributions for trend [2]. In their research, Rosenberg and Halpert chose the model $tr(x_t) = ax_t^b$ over two other models because that model provided the best fit to the trend data via a least squares regression test.

The parallel use of this same curve to fit actual liability data for trend and loss development is noteworthy. After rereading the sections of Rosenberg and Halpert regarding their fitting techniques of their model to trend, one can understand the relevancy of their comments regarding the function ax_t^b for trend to Pinto and Gogol's comments regarding ax^b for loss development. By setting $b > 0$, Rosenberg and Halpert have allowed trend to increase by claim size, and Pinto/Gogol have allowed excess loss development to increase by size of retention. Although much discussion has centered around the alleged "overlap" of trend and loss development, the use (and we will see later, the good fit) of the same theoretical function in both instances illustrates the similarity of the forces impacting trend and loss development.

Through the application of the Sherman normal power curve approximation, tail factors for development beyond 99 months have been determined. The authors offer several reasons why the fitted ISO tails are faster than the development based on RAA data. In addition, Pinto/Gogol offer a method to use the RAA development with the ISO fitted patterns, for development beyond 99 months, if the actuary believes that the RAA development is more appropriate.

3. GOODNESS OF FIT TESTS

The authors' intent is to find a loss distribution which will fit the three actual ISO loss development data matrices reasonably well. In their smoothing technique, cumulative intervals (27–99, 39–99, etc.) are used by the authors for smoothing development by retention. Additionally, Pinto/Gogol present the cumulative comparison of actual and fitted factors, for the 27–99 interval, showing the apparent similarity of those cumulative factors. For that reason, the goodness of fit tests are performed for the development of all six cumulative intervals.

The goodness of fit tests are applied to the actual and fitted cumulative data on Exhibit 1, Sheets 2–4, for OL&T BI, M&C BI, and Products BI, respectively. The formula for percentage error is as follows:

$$\left\{\left(\frac{\text{Fitted Cumulative Factor}}{\text{Actual Cumulative Factor}}\right) - 1\right\} \times 100 \tag{3.1}$$

For each subline, three groups of mean percentage errors are calculated and shown on Exhibit 1, Sheet 1:

- by development interval, retention layers excess of \$10,000 to excess of \$250,000;
- by development interval, retention layers excess of \$10,000 to excess of \$1,000,000; and,
- by retention, all development intervals (27–99 to 87–99).

Restated, the goodness of fit tests are performed by row (development interval) twice, for the \$10,000 to \$250,000 retention columns and \$10,000 to \$1,000,000 retention columns; and by column (retention) once, for all development intervals. Also, mean percentage errors are calculated for the entire data matrices (all development intervals and retentions), to provide an indicator of the goodness of fit for the overall technique.

The conclusions from the goodness of fit tests are as follows:

- Excluding the excess development for retentions greater than \$250,000, the mean percentage errors for OL&T BI, M&C BI, and Products BI are −1.7%, −0.4%, and −0.6%, respectively. Therefore, the fitted cumulative development errors are at most 2% below the actual cumulative development errors, averaged over all retentions and development intervals.
- For every subline, the mean absolute percentage errors (MAPE) by retention (columns) exceed 1.5% for the following areas:
  OL&T BI: \$10,000 and \$250,000 retentions
  M&C BI: \$10,000 and \$250,000 retentions
  Products BI: \$100,000 retention
- For OL&T and M&C, the MAPEs for \$500,000 and \$1,000,000 retentions are between 8% and 27%. The Products MAPE for the \$1,000,000 retention is 8% in contrast to the other retentions' MAPEs of less than 2%. These observations show that the Sherman normal power approximation works reasonably well for the development through \$250,000 retentions, but is inconsistent for retentions in excess of \$250,000.
- The fit of the actual data by development interval (row) for retention intervals \$10,000 through \$250,000 is fine for M&C and Products, with all MAPEs below 1.5%.

EXHIBIT 1
SHEET 1

GOODNESS OF FIT TESTS
MEAN PERCENTAGE ERRORS

OL&T BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

By Development Interval			By Retention	
	Retention Layers			
Development Interval	10,000 to 250,000	10,000 to 1,000,000	Retention	27–99 to 87–99
27–99	−4.0%	−5.1%	10,000	−3.7%
39–99	−4.8%	−10.4%	25,000	−0.8%
51–99	−2.4%	−8.9%	50,000	0.2%
63–99	−1.3%	−5.5%	100,000	−0.3%
75–99	−0.8%	−3.6%	250,000	−3.8%
87–99	3.2%	2.7%	500,000	−10.0%
All	−1.7%	−5.1%	1,000,000	−17.4%

M&C BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

By Development Interval			By Retention	
	Retention Layers			
Development Interval	10,000 to 250,000	10,000 to 1,000,000	Retention	27–99 to 87–99
27-99	−1.1%	−3.4%	10,000	−1.9%
39–99	−1.1%	−7.7%	25,000	0.4%
51–99	−0.3%	−6.4%	50,000	1.2%
63–99	0.9%	−5.3%	100,000	0.7%
75–99	−0.8%	−6.4%	250,000	−2.5%
87–99	−0.2%	−2.3%	500,000	−7.9%
All	−0.4%	−4.5%	1,000,000	−26.8%

PRODUCTS BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

By Development Interval			By Retention	
	Retention Layers			
Development Interval	10,000 to 250,000	10,000 to 1,000,000	Retention	27–99 to 87–99
27–99	−1.3%	−2.6%	10,000	−0.3%
39–99	−1.3%	−4.3%	25,000	−0.9%
51–99	−0.1%	−3.3%	50,000	−1.0%
63–99	−0.8%	3.5%	100,000	−1.6%
75–99	−1.5%	3.8%	250,000	0.6%
87–99	1.0%	8.0%	500,000	1.0%
All	−0.6%	0.8%	1,000,000	7.9%

EXHIBIT 1
SHEET 2

OL&T BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

Fitted Cumulative Factors
Retention

Development Interval	10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	1.89995	2.24203	2.54120	2.88023	3.39887	3.85234	4.36635
39–99	1.39133	1.62686	1.83122	2.06120	2.41015	2.71284	3.05358
51–99	1.20769	1.36148	1.49075	1.63223	1.84012	2.01478	2.20603
63–99	1.11799	1.20318	1.27195	1.34462	1.44711	1.52979	1.61720
75–99	1.06375	1.10490	1.13711	1.17024	1.21554	1.25095	1.28741
87–99	1.02691	1.04222	1.05396	1.06583	1.08173	1.09391	1.10623

Actual Cumulative Factors
Retention

Development Interval	10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	2.01337	2.31925	2.61370	2.97050	3.57977	4.28524	4.62738
39–99	1.50747	1.67467	1.85962	2.11861	2.64932	3.36309	4.32951
51–99	1.27762	1.37268	1.48189	1.63700	1.95999	2.41254	3.31484
63–99	1.15559	1.20390	1.25968	1.34533	1.51199	1.71345	2.05444
75–99	1.08364	1.10683	1.13199	1.16925	1.24752	1.34000	1.50376
87–99	1.01180	1.01460	1.01670	1.02350	1.03830	1.06130	1.11110

	Percentage Errors Retention							Mean Percentage Errors For Retentions:	
Development Interval	10,000	25,000	50,000	100,000	250,000	500,000	1,000,000	10,000 to 250,000	10,000 to 1,000,000
27–99	−5.6%	−3.3%	−2.8%	−3.0%	−5.1%	−10.1%	−5.6%	−4.0%	−5.1%
39–99	−7.7%	−2.9%	−1.5%	−2.7%	−9.0%	−19.3%	−29.5%	−4.8%	−10.4%
51–99	−5.5%	−0.8%	0.6%	−0.3%	−6.1%	−16.5%	−33.4%	−2.4%	−8.9%
63–99	−3.3%	−0.1%	1.0%	−0.1%	−4.3%	−10.7%	−21.3%	−1.3%	−5.5%
75–99	−1.8%	−0.2%	0.5%	0.1%	−2.6%	−6.6%	−14.4%	−0.8%	−3.6%
87–99	1.5%	2.7%	3.7%	4.1%	4.2%	3.1%	−0.4%	3.2%	2.7%
All Intervals Mean Percentage Errors:	−3.7%	−0.8%	0.2%	−0.3%	−3.8%	−10.0%	−17.4%	−1.7%	−5.1%

EXHIBIT 1
SHEET 3

M&C BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

Fitted Cumulative Factors

Development Interval	Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	2.46773	2.79319	3.06752	3.36890	3.81312	4.18770	4.59910
39–99	1.50464	1.66601	1.79943	1.94359	2.15200	2.32437	2.51057
51–99	1.19734	1.29235	1.36917	1.45059	1.56567	1.65876	1.75739
63–99	1.09365	1.15195	1.19808	1.24608	1.31249	1.36506	1.41976
75–99	1.04479	1.07786	1.10355	1.12987	1.16563	1.19343	1.22190
87–99	1.01728	1.03168	1.04270	1.05385	1.06876	1.08018	1.09173

Actual Cumulative Factors

Development Interval	Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	2.51971	2.79852	3.06557	3.40139	3.90759	4.21684	5.59436
39–99	1.55097	1.66420	1.78220	1.94054	2.23533	2.61753	3.98005
51–99	1.22801	1.28272	1.34202	1.42866	1.62275	1.89060	2.54821
63–99	1.10631	1.13355	1.16606	1.21371	1.32860	1.51006	2.10075
75–99	1.06366	1.07823	1.09572	1.12505	1.20694	1.33162	1.75913
87–99	1.02670	1.03190	1.03820	1.04910	1.07820	1.11920	1.23830

Development Interval	Percentage Errors Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000	Mean Percentage Errors For Retentions: 10,000 to 250,000	10,000 to 1,000,000
27–99	−2.1%	−0.2%	0.1%	−1.0%	−2.4%	−0.7%	−17.8%	−1.1%	−3.4%
39–99	−3.0%	0.1%	1.0%	0.2%	−3.7%	−11.2%	−36.9%	−1.1%	−7.7%
51–99	−2.5%	0.8%	2.0%	1.5%	−3.5%	−12.3%	−31.0%	−0.3%	−6.4%
63–99	−1.1%	1.6%	2.7%	2.7%	−1.2%	−9.6%	−32.4%	0.9%	−5.3%
75–99	−1.8%	−0.0%	0.7%	0.4%	−3.4%	−10.4%	−30.5%	−0.8%	−6.4%
87–99	−0.9%	−0.0%	0.4%	0.5%	−0.9%	−3.5%	−11.8%	−0.2%	−2.3%
All Intervals Mean Percentage Errors:	−1.9%	0.4%	1.2%	0.7%	−2.5%	−7.9%	−26.8%	−0.4%	−4.5%

EXHIBIT 1
SHEET 4

PRODUCTS-BI EXCESS LOSS & ALAE DEVELOPMENT FACTORS

Fitted Cumulative Factors

Development Interval	Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	3.07710	3.53866	3.93323	4.37181	5.02759	5.58824	6.21139
39–99	1.70416	1.87414	2.01387	2.16403	2.37988	2.55735	2.74804
51–99	1.33631	1.41189	1.47186	1.53438	1.62115	1.69003	1.76182
63–99	1.17969	1.21552	1.24332	1.27177	1.31039	1.34037	1.37105
75–99	1.09317	1.10981	1.12255	1.13544	1.15271	1.16595	1.17933
87–99	1.03817	1.04438	1.04909	1.05383	1.06013	1.06492	1.06973

Actual Cumulative Factors

Development Interval	Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000
27–99	3.07724	3.63668	3.99647	4.47700	5.01208	6.36789	6.20299
39–99	1.72000	1.90512	2.04287	2.21557	2.38070	2.66038	3.44113
51–99	1.33271	1.40485	1.47564	1.55795	1.60967	1.76208	2.17147
63–99	1.18190	1.22150	1.25736	1.29905	1.30857	1.25210	1.13446
75–99	1.11164	1.13354	1.15058	1.16350	1.14255	1.07384	0.93959
87–99	1.02930	1.03690	1.04050	1.04210	1.04400	0.96050	0.76570

Development Interval	Percentage Errors Retention 10,000	25,000	50,000	100,000	250,000	500,000	1,000,000	Mean Percentage Errors For Retentions: 10,000 to 250,000	10,000 to 1,000,000
27–99	−0.0%	−2.7%	−1.6%	−2.3%	0.3%	−12.2%	0.1%	−1.3%	−2.6%
39–99	−0.9%	−1.6%	−1.4%	−2.3%	−0.0%	−3.9%	−20.1%	−1.3%	−4.3%
51–99	0.3%	0.5%	−0.3%	−1.5%	0.7%	−4.1%	−18.9%	−0.1%	−3.3%
63–99	−0.2%	−0.5%	−1.1%	−2.1%	0.1%	7.1%	20.9%	−0.8%	3.5%
75–99	−1.7%	−2.1%	−2.4%	−2.4%	0.9%	8.6%	25.5%	−1.5%	3.8%
87–99	0.9%	0.7%	0.8%	1.1%	1.5%	10.9%	39.7%	1.0%	8.0%
All Intervals Mean Percentage Errors:	−0.3%	−0.9%	−1.0%	−1.6%	0.6%	1.0%	7.9%	−0.6%	0.8%

Based upon the above observations, several areas for understatement of actual development exist for these three sublines' data. For OL&T, fitted development at \$10,000 and \$250,000 retentions is at least 7% below actual development for the development interval 39–99 months. The M&C actual development is understated at least 3% for these same cells. Products fitted development data is about 2% understated for the \$100,000 retention; this difference is not substantial. These differences might be adjusted for on an adhoc basis after application of the technique.

I initially performed these same tests on the data as of successive (e.g., noncumulative) intervals, and discovered that the goodness of fit can reverse with the accumulation of development. For example, for OL&T BI, the percentage error for development interval 27–39, retention \$250,000, is +4.4%; the corresponding factor for the interval 27–99 is −5.1%. This shows that random variations in reporting do not always get smoothed out when the development for successive intervals is accumulated.

In conversations with the authors, they indicated their goal was to produce an intuitively reasonable, natural, and smooth sequence of curves for development to provide knowledge where published information is not available. Based upon these goodness of fit tests, and ignoring pockets of discrepancies, the authors have met their goal.

4. RETENTIONS IN EXCESS OF \$250,000

Based on the goodness of fit tests, it is obvious that the fit to the actual data for retentions in excess of \$250,000 is poor. The authors mention that the tendency for development to increase as retention increases is *reversed* at \$500,000 and \$1,000,000 for the 27-39 month intervals.

The authors suggest this may be due to a credibility problem of the data for these large claims. However, there may be another reason. Some claims people feel that, for very large claims, an estimate of the loss put up in the first year often is not revised until several years later, closer to a jury trial. The following comparison for Products, based on goodness of fit tests for excess development factors at the $1,000,000 retention, is interesting:

DEVELOPMENT INTERVAL	PERCENTAGE ERROR
27–39 (SUCCESSIVE)	+25.4%
27–99 (CUMULATIVE)	+ 0.1%
63–75 (SUCCESSIVE)	− 3.7%
63–99 (CUMULATIVE)	+20.9%

The inclusion of the later 39–99 month development provides a better fit for the data than the 27–39 month development alone. At 63 months, however, the inclusion of the 75–99 month development provides a much poorer fit than the 63–75 month development alone.

In summary, this "catch-up" theory is supported by the goodness of fit tests. For less mature data, the inclusion of later development tends to smooth out the random variations; for more mature data, including the tail provides a poorer fit. For either reason, a lack of credibility or differing reserving practices, it seems wise to exclude the very high retentions when applying this technique.

5. THE METHOD APPLIED

In Section 5, the authors introduce the formula for the excess development factor as follows:

$$(f(c) - f(d)) \div (e_{c,n} - e_{d,n}) \tag{5.1}$$

with $f(c)$ being the ratio of the excess losses to the "ground-up" projected ultimate losses, and $e_{c,n}$ representing the excess loss ratio divided by the loss development factor to ultimate, for retention c and month n. The function $f(x)$ is a very familiar one to actuaries—it is merely an excess loss function. In standard actuarial terminology, these $f(x)$'s are also noted as $X3(x)$ [3]. For workers' compensation, the excess loss premium factors could be a readily available published source for these excess ratios.

At first, I found this formula to be somewhat problematic, but found the proof to be somewhat straightforward. (The proof is presented in Appendix A.) From that formula, other powerful formulas can be derived to estimate other development patterns.

For example, loss development data sometimes is available only for basic limits and total limits, but is not available for excess limits. Setting the basic limit equal to B (in 000's), the formula for development from 12 months to ultimate in the layer \$0 to \$$B$, the basic limits "layer," is the following:

$$LDF_{BASIC,\,12} = \frac{1 - f(B)}{\dfrac{1}{LDF_{0,12}} - \dfrac{f(B)}{LDF_{B,12}}} \tag{5.2}$$

with $LDF_{BASIC,12}$ representing the basic limits loss development factor, $LDF_{0,12}$ the total limits loss development factor, and $LDF_{B,12}$ being the excess loss development factor, all from 12 months to ultimate. Also, $f(0) = 1$ and $f(B) =$ the excess loss ratio for losses in excess of \$$B$, the basic limits. Here, basic limits development is treated as development for the layer of losses in excess of \$0 retention minus the losses in excess of the basic limit of \$$B$. This also leads to the formula:

$$LDF_{EXCESS} = LDF_{B,12} = \frac{f(B)}{\dfrac{1}{LDF_{0,12}} - \dfrac{(1 - f(B))}{LDF_{BASIC,12}}} \tag{5.3}$$

From this discussion, it can be inferred that primary loss development, from ground up, can be considered a special case of excess development. Therefore, primary development factors by layer can be produced using the Pinto/Gogol formula, as long as total limits "ground-up" loss development is available (that is, retention = \$0). Since the raw ISO development was presented on the exhibits, I extrapolated to 363 months the M&C BI "ground-up" development data using the Sherman method that Pinto and Gogol used for the excess development for other retentions. The application of formula (5.2) for primary development (\$15,000, \$25,000, up to \$250,000) is presented on Exhibit 2.

The results from this technique for M&C primary development are disappointing. The reason for these intuitively disturbing factors may be given by Rosenberg and Halpert. In their study for trend, they found that "a concern with the trend function ax_t^b is that it tends to underestimate the trend for small x_t, that is, small sizes of losses." This conclusion may extend to conclusions for excess loss development as well. Although the ground-up development is not part of the smoothing technique, its use with the excess development (that had been smoothed and tested for goodness of fit) to produce primary development factors does not provide sensible results.

EXHIBIT 2

DETERMINATION OF PRIMARY LOSS LAYER LOSS DEVELOPMENT FACTORS M&C BI BASIC LIMITS LOSSES & ALAE

Loss Layer	Development Interval 27:ult	39:ult	51:ult	63:ult
$0–$15,000	1.201	0.962	0.955	0.936
$0–$25,000 (from ISO)	2.007	1.379	1.164	1.075
$0–$25,000 (computed)	1.123	0.862	0.830	0.818
$0–$35,000	1.137	0.853	0.806	0.792
$0–$50,000	1.189	0.874	0.811	0.793
$0–$75,000	1.288	0.927	0.844	0.819
$0–$100,000	1.385	0.984	0.884	0.888
$0–$250,000	1.727	1.187	1.032	0.975

Formula: $(f(low) - f(high))/((f(low)/low{:}ult) - (f(high)/high{:}ult))$

Development Interval	Cumulative Fitted Factors 0	15000	25000	35000	50000	75000	100000	250000
27:ult	2.2093	2.8639	3.2210	3.4802	3.7778	4.1472	4.4309	5.4702
39:ult	1.4769	1.7293	1.9212	2.0591	2.2161	2.4092	2.5563	3.0872
51:ult	1.2475	1.3607	1.4903	1.5824	1.6862	1.8126	1.9079	2.2461
63:ult	1.1521	1.2295	1.3284	1.3979	1.4755	1.5690	1.6389	1.8829
$f(x)$ ratios:	1.000	0.786	0.755	0.721	0.674	0.605	0.543	0.319

Development Interval	Report-to-Report Fitted Factors 0	15000	25000	35000	50000	75000	100000	250000
27– 39	1.49590	1.65613	1.67658	1.69018	1.70472	1.72141	1.73335	1.77192
39– 51	1.18395	1.27092	1.28912	1.30125	1.31424	1.32916	1.33984	1.37446
51– 63	1.08278	1.10671	1.12188	1.13199	1.14281	1.15523	1.16412	1.19291
63– 75	1.04456	1.05643	1.06873	1.07691	1.08565	1.09567	1.10284	1.12597
75– 87	1.02687	1.03484	1.04476	1.05134	1.05837	1.06641	1.07215	1.09065
87– 99	1.01720	1.02363	1.03168	1.03702	1.04271	1.04921	1.05385	1.06877
99–111	1.01180	1.01705	1.02367	1.02806	1.03272	1.03806	1.04186	1.05405
111–123	1.00850	1.01288	1.01839	1.02204	1.02592	1.03035	1.03351	1.04363
123–135	1.00630	1.01006	1.01470	1.01778	1.02105	1.02478	1.02743	1.03593
135–147	1.00490	1.00807	1.01204	1.01466	1.01744	1.02062	1.02288	1.03012
147–159	1.00380	1.00661	1.01004	1.01230	1.01470	1.01744	1.01939	1.02562
159–171	1.00314	1.00551	1.00849	1.01045	1.01254	1.01492	1.01661	1.02202
171–183	1.00256	1.00466	1.00728	1.00900	1.01084	1.01293	1.01441	1.01908
183–195	1.00211	1.00399	1.00631	1.00784	1.00946	1.01131	1.01262	1.01681
195–207	1.00177	1.00345	1.00552	1.00688	1.00832	1.00997	1.01114	1.01487
207–219	1.00149	1.00301	1.00486	1.00608	1.00737	1.00884	1.00989	1.01322
219–231	1.00127	1.00266	1.00433	1.00543	1.00659	1.00792	1.00886	1.01187
231–243	1.00110	1.00235	1.00386	1.00485	1.00591	1.00711	1.00796	1.01068
243–255	1.00095	1.00210	1.00347	1.00437	1.00533	1.00642	1.00719	1.00966
255–267	1.00083	1.00189	1.00314	1.00396	1.00484	1.00583	1.00654	1.00879
267–279	1.00073	1.00171	1.00286	1.00361	1.00441	1.00533	1.00598	1.00804
279–291	1.00064	1.00155	1.00260	1.00330	1.00403	1.00487	1.00547	1.00737
291–303	1.00057	1.00141	1.00238	1.00302	1.00370	1.00448	1.00503	1.00678
303–315	1.00051	1.00128	1.00218	1.00278	1.00341	1.00412	1.00463	1.00625
315–327	1.00046	1.00119	1.00202	1.00258	1.00316	1.00383	1.00431	1.00582
327–339	1.00041	1.00109	1.00187	1.00239	1.00294	1.00356	1.00400	1.00541
339–351	1.00037	1.00101	1.00173	1.00221	1.00272	1.00330	1.00371	1.00501
351–363	1.00034	1.00093	1.00161	1.00206	1.00253	1.00307	1.00346	1.00468

Rosenberg and Halpert may provide the solution, however, by suggesting that the understatement could "be corrected by changing the model to . . . $a(x_t + c)^b$, or by using the function ax_t^b only for claim sizes greater than a selected value and using empirical data to trend small losses." In this case, the fitted development for lower retentions (say below $35,000) may need to be adjusted to produce reasonable primary development. That study is beyond the scope of this discussion.

However, I also applied their data for some standard excess development layers, and obtained satisfyingly reasonable results for the excess layers. These results are presented in Exhibit 3.

6. SUMMARY

Messrs. Pinto and Gogol have written a fine paper with practical and useful applications. Although excess development at very low retentions (for use in primary development) or large retentions (in excess of $250,000) may be dubious, application for development at retentions between those extremes are easy to apply and tailor for today's "mix and match" reinsurance program environment.

EXHIBIT 3
DETERMINATION OF PRIMARY LOSS LAYER LOSS DEVELOPMENT FACTORS
M&C BI EXCESS LIMITS LOSSES & ALAE

Loss Layer	Development Interval 27:ult	39:ult	51:ult	63:ult
$15,000–$515,000	2.551	1.552	1.234	1.126
$35,000–$535,000	3.143	1.874	1.454	1.297
$50,000–$550,000	3.430	2.027	1.557	1.375
$75,000–$575,000	3.789	2.217	1.683	1.470
$100,000–$600,000	4.066	2.363	1.779	1.542
$250,000–$750,000	5.102	2.897	2.124	1.794
$35,000–$500,000	3.112	1.857	1.442	1.288
$50,000–$500,000	3.386	2.003	1.540	1.363
$75,000–$500,000	3.721	2.181	1.659	1.452
$100,000–$500,000	3.971	2.312	1.745	1.516
$250,000–$500,000	4.851	2.768	2.041	1.733

Formula: $(f(low) - f(high))/((f(low)/low{:}ult) - (f(high)/high{:}ult))$

Development Interval	Cumulative Fitted Factors 500000	515000	535000	550000	575000	600000	750000
27:ult	6.4157	6.4595	6.5164	6.5579	6.6253	6.6905	7.0429
39:ult	3.5610	3.5828	3.6110	3.6316	3.6650	3.6973	3.8712
51:ult	2.5413	2.5547	2.5721	2.5848	2.6054	2.6252	2.7317
63:ult	2.0913	2.1007	2.1129	2.1218	2.1361	2.1499	2.2238
$f(x)$ ratios:	0.148	0.142	0.135	0.130	0.122	0.114	0.078

Development Interval	Report-to-Report Fitted Factors 500000	515000	535000	550000	575000	600000	750000
27– 39	1.80167	1.80295	1.80460	1.85080	1.80772	1.80957	1.81930
39– 51	1.40124	1.40240	1.40389	1.40497	1.40671	1.40837	1.41715
51– 63	1.21516	1.21612	1.21735	1.21825	1.21970	1.22108	1.22837
63– 75	1.14380	1.14457	1.14555	1.14627	1.14743	1.14853	1.15436
75– 87	1.10486	1.10547	1.10626	1.10683	1.10775	1.10863	1.11326
87– 99	1.08020	1.08069	1.08132	1.08178	1.08251	1.08322	1.08694
99–111	1.06337	1.06377	1.06428	1.06466	1.06526	1.06584	1.06886
111–123	1.05134	1.05167	1.05210	1.05241	1.05291	1.05338	1.05588
123–135	1.04240	1.04268	1.04304	1.04330	1.04371	1.04411	1.04621
135–147	1.03562	1.03586	1.03616	1.03638	1.03673	1.03707	1.03886
147–159	1.03036	1.03056	1.03082	1.03101	1.03132	1.03161	1.03314
159–171	1.02613	1.02631	1.02653	1.02670	1.02696	1.02722	1.02854
171–183	1.02267	1.02282	1.02302	1.02316	1.02339	1.02362	1.02477
183–195	1.02000	1.02013	1.02031	1.02044	1.02064	1.02084	1.02187
195–207	1.01770	1.01782	1.01798	1.01809	1.01827	1.01845	1.01936
207–219	1.01575	1.01586	1.01600	1.01610	1.01627	1.01642	1.01724
219–231	1.01415	1.01425	1.01438	1.01447	1.01461	1.01475	1.01549
231–243	1.01274	1.01283	1.01294	1.01302	1.01316	1.01328	1.01395
243–255	1.01153	1.01161	1.01171	1.01179	1.01191	1.01202	1.01263
255–267	1.01050	1.01057	1.01067	1.01073	1.01084	1.01095	1.01150
267–279	1.00961	1.00968	1.00976	1.00982	1.00992	1.01002	1.01053
279–291	1.00881	1.00887	1.00895	1.00901	1.00910	1.00919	1.00965
291–303	1.00811	1.00816	1.00823	1.00829	1.00837	1.00845	1.00888
303–315	1.00748	1.00754	1.00760	1.00765	1.00773	1.00781	1.00820
315–327	1.00696	1.00701	1.00707	1.00712	1.00719	1.00726	1.00763
327–339	1.00648	1.00652	1.00658	1.00662	1.00669	1.00676	1.00710
339–351	1.00600	1.00605	1.00610	1.00614	1.00620	1.00626	1.00658
351–363	1.00561	1.00565	1.00570	1.00574	1.00580	1.00585	1.00615

REFERENCES

[1] Richard E. Sherman, "Extrapolating, Smoothing, and Interpolating Development Factors," *PCAS* LXXI, 1984, p. 122.

[2] Sheldon Rosenberg and Aaron Halpert, "Adjusting Size of Loss Distributions for Trend," *Inflation Implications for Property-Casualty Insurance,* Casualty Actuarial Society 1981 Discussion Paper Program, p. 458.

[3] J. Gary LaRose, "A Note on Loss Distributions," *PCAS* LXIX, 1982, p. 15.

APPENDIX A

PROOF

$f(c)$ = Excess loss ratio at retention c, the lower retention.

$f(d)$ = Excess loss ratio at retention d, the upper retention.

$$e_{c,n} = \frac{f(c)}{LDF_{c,n}}$$

with $LDF_{c,n}$ the excess loss development factor at retention c from n months to ultimate.

$$e_{d,n} = \frac{f(d)}{LDF_{d,n}}$$

with $LDF_{d,n}$ the excess loss development factor at retention d from n months to ultimate.

We know, however, that:

$$f(c) = \frac{\text{Ultimate Losses Excess of Retention } c}{\text{Ultimate Ground-Up Losses}} = \frac{ULT_c}{ULT_g}$$

$$f(d) = \frac{\text{Ultimate Losses Excess of Retention } d}{\text{Ultimate Ground-Up Losses}} = \frac{ULT_d}{ULT_g}$$

$$LDF_{c,n} = \frac{\text{Ultimate Losses Excess of Retention } c}{\text{Reported Losses Excess of Retention } c} = \frac{ULT_c}{REP_c}$$

$$LDF_{d,n} = \frac{\text{Ultimate Losses Excess of Retention } d}{\text{Reported Losses Excess of Retention } d} = \frac{ULT_d}{REP_d}$$

(I have dropped the subscript n for months of development.)

So:

$$\frac{f(c) - f(d)}{e_{c,n} - e_{d,n}}$$

$$= \frac{f(c) - f(d)}{\dfrac{f(c)}{LDF_{c,n}} - \dfrac{f(d)}{LDF_{d,n}}}$$

$$= \frac{\dfrac{ULT_c}{ULT_g} - \dfrac{ULT_d}{ULT_g}}{\left[\dfrac{ULT_c}{ULT_g} \times \dfrac{REP_c}{ULT_c}\right] - \left[\dfrac{ULT_d}{ULT_g} \times \dfrac{REP_d}{ULT_d}\right]}$$

$$= \frac{\dfrac{ULT_c - ULT_d}{ULT_g}}{\dfrac{REP_c}{ULT_g} - \dfrac{REP_d}{ULT_g}} = \frac{\dfrac{ULT_c - ULT_d}{ULT_g}}{\dfrac{REP_c - REP_d}{ULT_g}}$$

$$= \frac{ULT_c - ULT_d}{REP_c - REP_d}$$

It is obvious that the excess loss development factor for the layer c to d is the ultimate losses greater than c minus ultimate losses in excess of d, divided by the reported losses in excess of c minus the reported losses in excess of retention d.

CREDIBILITY FOR CLASSIFICATION RATEMAKING VIA THE HIERARCHICAL NORMAL LINEAR MODEL

STUART KLUGMAN

Abstract

In the past twenty years there has been ever increasing improvement in the techniques of classification ratemaking. Most of this has centered around improvements in credibility procedures and most of the improvements have been due to incorporating aspects of Bayesian analysis. In this paper, I attempt to take this trend to its (perhaps) final stage by developing a true Bayesian approach to the classification ratemaking credibility problem.

The opening section will provide the rationale for the Bayesian approach. I will argue that a hierarchical model with a noninformative prior is the most appropriate general framework. I will argue further that a normal model is a reasonable choice, and this model will provide results at least as good as those currently available. An indication of how the normality condition can be relaxed will also be presented.

The second section contains a general description and analysis of the hierarchical normal linear model (HNLM). Included are point estimation, estimation of the error in the estimator, and prediction intervals for future losses. The last two items are of special interest since current credibility procedures provide little insight with respect to variation.

The next two sections discuss the special case of the one-way model. This is the most common ratemaking model and is the simplest case of the HNLM. In Section 3, the formulas from Section 2 are evaluated for this model. In Section 4, two data sets are analyzed. The first set provides an indication of the computational work required to use the HNLM. The second set provides a comparison of this method with two other ratemaking approaches.

The final section contains a discussion of the more complex models that can be handled with the HNLM.

The majority of this work was supported by a grant from the Society of Actuaries Research Development Fund through the Actuarial Education and Research Fund grants competition. An earlier version of the one-way model was presented at a NATO sponsored conference (Klugman [22]). Many conversations with Glenn Meyers and Gary Venter helped refine the arguments presented here. Two anonymous referees from the CAS provided a number of suggestions that improved the exposition. I thank the National Council on Compensation Insurance and Glenn Meyers for providing the two data sets and especially thank Kathy Hockenberry for writing the programs to do the analyses.

1. JUSTIFICATION FOR BAYESIAN CREDIBILITY WITH A NORMAL MODEL AND A NONINFORMATIVE PRIOR

The historical basis for credibility procedures is long, varied, and generally considered to be one of the major actuarial contributions to statistical data analysis. Virtually from the beginning (Whitney [35] and Bailey [1]), the Bayesian and shrinkage nature of the problem was recognized. In a breakthrough paper, Bühlmann [6] placed the credibility problem in the framework of Bayesian decision analysis. I will begin by reviewing the Bayesian view and then discuss the four schools of Bayesian methodology that are prominent today. As part of this paper, I will argue that one of these methods is superior to the others. Next, I will argue that the normal model is appropriate even though we know that it does not accurately model insurance losses. This part closes with a suggestion for allowing for non-normal losses while retaining the advantages of normal theory. The final element of this section is a discussion of the noninformative prior.

1.1 Credibility as a Bayesian Problem

The basic credibility problem for classification ratemaking can be posed as follows: The population can be separated into k groups, the various rating classes. Our objective is to estimate the mean loss per year generated by a randomly selected member of a particular group. Data is collected from a sample of members from each group. It is usually assumed that the observed losses are independent and that the variances of the observations are proportional to some measure of exposure. If this were all that were known, the most reasonable answer would be to use the sample mean from each group as an estimate of the population mean. Usually, however, we know more. In particular, when individual classes have abnormally good or bad experience, we tend to discount the experience when setting rates. This clearly makes good business sense and with the correct model makes good statistical sense.

The usual way to model this phenomenon is to treat the group means as a random sample from some probability distribution. This implies that experience from the other groups tells us something about the overall level of claims (the mean of this second level distribution), and therefore tells us something about the mean for the group in question. It also sets bounds on how much one can legitimately expect one class to differ from another. If more is known about the relationship among the groups, that knowledge can be incorporated into the second level distibution. Examples of this are presented in Section 5.

The model described in the previous paragraph is a standard Bayesian problem. We have a model given by the p.d.f. $p(x|\theta,g)$ where x represents the data and (θ,g) represents all unknown parameters. The parameters in θ are the ones we want to estimate. The parameters in g are nuisance parameters, usually variances. In the above setting, θ would be the group means. The prior (second level) p.d.f. $p(\theta,g)$ represents our knowledge of (θ,g) before the data are collected. Since the Bayesian approach has now been widely accepted among actuaries (at least for this estimation problem), I will provide no further arguments to support that view. Interested readers who desire a wide ranging discussion of the merits of the Bayesian view are referred to Berger [2].

Given this setup, there are two ways to proceed. If the forms of the two distributions are known, the Bayes estimator is the posterior mean of θ given the data x. Bühlmann [6] took a different approach. To avoid thinking about the distributions, he first restricted himself to estimators that are linear functions of the data. He then searched for the estimator that minimized the mean squared error. This mean would be taken over all possible values of x and θ. For his result it was essential that g be empty. That is, the model variance had to be known. Under this framework, it turned out that the estimator depended only upon the first two moments of the model and prior distributions. To many people, the word credibility is now reserved only for procedures that find linear estimators. In fact, Hewitt [14] compares a credibility estimator to a Bayes estimator (as I have defined it above). In this paper, the objective is to find the best estimator, and I see no reason to restrict attention to those that are linear functions of the data. I use the word credibility to describe any procedure that uses information ("borrows strength") from samples from different, but related, populations.

A larger problem is the fact that the moments of the model and prior are rarely known, and therefore must be estimated. This has led to a number of schools of Bayesian thought. Having agreed to use a Bayesian procedure, the remaining task is to identify the best one.

1.2 Four Schools of Bayesian Analysis

There are at least four different approaches that are currently being used to solve the estimation problem. In this section I briefly outline them and then offer some opinions as to their respective merits.

1.2.1 Pure Bayes with Two Levels

This is the view that has already been mentioned. Here, the prior distribution must be elicited. This is very difficult to do in the insurance setting as one would have to be able to set out a distribution that describes the class-to-class variation in losses. Since we do not even know the means (determining them is the point of the exercise), it is unlikely that we know much about how the means vary.

This problem can be resolved by removing the prior to a higher level of abstraction. This is done in the fourth school discussed in this subsection. To my knowledge, no one today is using the two level approach. At best, it is a starting point for the second method to be discussed here.

1.2.2 Empirical Bayes

This method evolved as an attempt to resolve the problems created by the first method. Although they did not use the phrase "empirical Bayes," Bühlmann and Straub [7] were the first to employ this method in the credibility setting. It remains popular, being advocated in more recent articles by the Insurance Services Office [16] and Meyers [25]. There is considerable evidence that it provides excellent solutions to the estimation problem.

Many people do not consider empirical Bayes methods to be at all Bayesian. Also, there is considerable disagreement as to what the phrase "empirical Bayes" means. To avoid controversy, I will describe an estimation method that corresponds to the approach used in the papers cited above. It will be referred to as the EB approach and the reader can decide what that means. Begin with the density $p(x|\theta,g)$, the first level density (or distribution, when talking about the random variable). The density $p(\theta,g|h)$ will be referred to as the second level density. Note the introduction of h. To the pure Bayesian, the parameters of the second level density must be known, and therefore do not need to be displayed. In reality, that is not true, so we add them to the formulation.

In brief, the EB idea is to first act as if g and h were known and find the Bayes estimate of θ. Next, use the data in some manner to estimate the nuisance parameters g and h and insert these estimates into the Bayes solution. The first thing to note is that upon doing so, we no longer have a Bayesian analysis. The second level distribution was supposed to represent prior opinion, yet here we are unable to establish this distribution until after we have seen the data. The usual justification is to show that as the sample size goes to infinity, the estimates of the second level distribution converge to what they ought to be if we had complete knowledge (which is what one has with an infinite sample size). A thorough discussion of these principles can be found in Norberg [30].

An alternative approach can yield the same solution as the EB approach. Assume again that the nuisance parameters are known and then search for the estimator that is linear in the data and minimizes the expected squared error. Once again, substitute ad hoc estimates of the nuisance parameters into the solution. This has been called least-squares credibility.

There are three major objections to the EB approach. The first is that some external theory must be used to find estimators of the nuisance parameters. Since these parameters are usually variances, it is common to begin with sums of squares that look "right" and then to adjust them to create unbiased estimators of the various parameters. One drawback is that the resulting estimators (even in the simplest cases) can take on negative values. This does not make sense when one is trying to estimate a variance. The second objection is that EB theory gives no guidance as to the optimal choice of the estimator. All that is required is that they be consistent. The final objection is that for complex models, there may be no hope of finding useful sums of squares.

A final problem with EB methodology is that it gives no insight into the sampling error of the estimator. The best it can do is evaluate the error when the variances are known. The additional error introduced by estimating the variances cannot be accounted for. Even if a good estimator of the nuisance parameters can be found (in which case, the method works quite well), the investigator will have no idea of the quality of the estimate. The previous statement that EB methods work well was in reference to alternative methods and does not mean that the results could be considered accurate. That can only be determined by some measure of sampling error. The next method is an attempt to rectify this problem without leaving the EB framework.

1.2.3 Parametric Empirical Bayes

To see the difficulties in determining the variance of the estimator, we need to take a closer look at what we are trying to do. The general Bayes problem is to find $E(\theta|x)$, the posterior mean given only the data. The EB approach uses the result $E(\theta|x) = E[E(\theta|x,g,h)]$. The interior expectation $E(\theta|x,g,h)$ is just the pure Bayes solution with g and h known. The EB approach avoids taking the outer expectation and instead replaces g and h with their estimates. EB theory indicates that this is a reasonable thing to do. A measure of the quality of the result would be the posterior variance, $\mathrm{Var}(\theta|x)$. We have $\mathrm{Var}(\theta|x) = \mathrm{E}[\mathrm{Var}(\theta|x,g,h)] + \mathrm{Var}[\mathrm{E}(\theta|x,g,h)]$. It is apparent that merely inserting estimates of g and h in $\mathrm{Var}(\theta|x,g,h)$ will underestimate the desired variance. The second term reflects the additional variance due to the estimation of g and h. EB theory does not provide any ideas for estimating the second term.

An attempt to resolve this problem is the parametric empirical Bayes theory of Morris [27]. The key ingredient is to have some idea of the variability of the estimators of g and h. His theory requires not only the discovery of good estimators of g and h but also the ability to determine their sampling distributions. In simple cases (normal distribution, equal exposures), it is possible to show that the usual estimators have chi-square distributions. In slightly more complicated cases, the distribution is approximately chi-square. A detailed discussion of the distribution of some commonly used variance estimators is given in Klugman [21]. As should be apparent, there are considerable difficulties associated with putting this method into practice. One that is not apparent, and is often not mentioned in Morris's articles, is that to complete the calculation it is necessary to formulate a prior distribution for g and h. Morris uses $p(g,h) = 1$, but does not provide a justification for that choice. Other choices are supported by an argument that the resulting estimator of the credibility factor is unbiased, a surprising justification for a Bayesian.

The fourth model also requires prior distributions for g and h but proceeds in a more direct Bayesian manner. Before moving on, I should add one final criticism of the parametric empirical Bayes approach. Whatever errors in estimation are introduced cannot be reduced by improving the computational aspects of the method. The errors are due to lack of knowledge of the exact distribution of the estimators of the variances and no amount of computation can resolve that issue.

1.2.4 Hierarchical Bayes

We have seen that the two EB schools are somewhat artificial attempts to resolve the problems of the pure Bayes method. This is mostly due to a lack of recognition of the real problem with the pure Bayes approach. The problem is that the second level distribution is not a prior distribution at all, but is part of the model. In the ratemaking setting, this distribution contains our knowledge of the relationships among the various rating classes. It is not our prior opinion about a particular class. The solution is to reformulate the model into three levels.

Level 1: $p(x|\theta,g)$—Describes variations within each group.
Level 2: $p(\theta|\mu,h)$—Describes variations among the groups.
Level 3: $p(\mu,g,h)$—A true prior distribution on the unknown parameters.

This is once again a pure Bayesian problem. As with any Bayesian analysis, a prior distribution must be established before any data is collected. By displacing the prior to a level further removed from the observations, the choice of the prior will have less influence on the final outcome. To repeat, level 2 describes an underlying (though not directly observable) physical process. Subjective beliefs enter only at the third level. The remaining problems are to select the prior distribution and to select the form of the p.d.f.'s for levels 1 and 2.

Assuming the two problems just mentioned can be resolved, this would appear to be an ideal solution to the credibility problem. With the three densities in hand, it is just a matter of employing the probability calculus to obtain the posterior distribution of θ given x. Any difficulties that will be encountered will be of a computational nature. Once the posterior distribution has been obtained, additional computation will yield the mean and variance. Another useful quantity is the predictive distribution, the p.d.f. (or the mean and variance) of the next observation from the group in question. This is once again obtained by an application of the probability calculus. The other methods do not provide this item.

An additional advantage of this approach is that the tools of Bayesian modeling and inference are all available. For example, one might want to compare various models for the level 2 distribution (e.g., cross-classification vs. one-way classification). Many of these tools are presently in the development stage, but more and more techniques are likely to become available in the future.

1.3 The Normal Model

As mentioned in the previous subsection, it is necessary to specify the probability distributions for the three levels. For levels 1 and 2, multivariate normal models are an appropriate choice. At the end of the section, a suggestion for improving the process is proposed.

It is obvious that individual losses do not follow the normal distribution. Since losses are non-negative quantities, a distribution with support on the entire real line cannot be expected to be a good model. Furthermore, there is considerable evidence that the tails of loss distributions are much heavier than those of the normal distribution. See Hogg and Klugman [15] for a number of examples. One way to minimize the disparity is to work with loss ratios. The distribution at the second level will now reflect group to group variations in the departure from the expected losses. This will be more stable than the group to group variation in the absolute level of losses. In addition, loss ratios are likely to have identical unconditional distributions. That is, if you were given a list of risk classes and a list of loss ratios, you would be unable to do better than chance in attempting to match them up. The loss ratios are likely to be dependent. Knowing that the loss ratio for one class is high increases the chances that the others are also. The multivariate normal model is one of the few multivariate models that allows for dependence in a manner that is easy to construct and interpret.

Despite the fact that the observations are not normally distributed, there are a number of good reasons for employing the normal model. The first, though least appealing, justification is computational convenience. Although the algebra is tedious, as demonstrated in Section 2, a number of results can be obtained analytically. The remaining numerical work will be simple, at least relative to that required for non-normal models. It is likely that as our numerical capabilities increase, this argument for normality will lose its validity. For the present, the following quote due to Novick and Jackson [31] is appropriate.

> "Surely it is better to get some results using a model which is only approximately relevant than to sit twiddling one's thumbs in front of a model which is felt to be more accurate but which one is unable to manipulate."

The second justification for normality is related to the link between the normal model and linear credibility. It was mentioned above that in a particular simple model, the linear least squares solution depended only upon the first two moments. It turns out that the Bayes solution for the same model with normal distributions is identical. Therefore, at least in this case, normality and linear least squares are equivalent. It has been shown that, in general, any model that is a member of the linear exponential family of distributions will produce the same result as the linear least squares solution (Ericson [9] and Jewell [17]). There has been speculation (Goel [11]) that the linear exponential family contains all the distributions with this property. So, to a certain extent, those who are willing to accept linear solutions should be equally comfortable with models from the linear exponential family. As far as choosing the normal distribution as the member to use, a second argument is needed. Most current practitioners estimate the variances using sums of squares. These estimates are unbiased for erable evidence that the tails of loss distributions are much heavier than those of the normal distribution. See Hogg and Klugman [15] for a number of examples. One way to minimize the disparity is to work with loss ratios. The distribution at the second level will now reflect group to group variations in the departure from the expected losses. This will be more stable than the group to group variation in the absolute level of losses. In addition, loss ratios are likely to have identical unconditional distributions. That is, if you were given a list of risk classes and a list of loss ratios, you would be unable to do better than chance in attempting to match them up. The loss ratios are likely to be dependent. Knowing that the loss ratio for one class is high increases the chances that the others are also. The multivariate normal model is one of the few multivariate models that allows for dependence in a manner that is easy to construct and interpret.

Despite the fact that the observations are not normally distributed, there are a number of good reasons for employing the normal model. The first, though least appealing, justification is computational convenience. Although the algebra is tedious, as demonstrated in Section 2, a number of results can be obtained analytically. The remaining numerical work will be simple, at least relative to that required for non-normal models. It is likely that as our numerical capabilities increase, this argument for normality will lose its validity. For the present, the following quote due to Novick and Jackson [31] is appropriate.

> "Surely it is better to get some results using a model which is only approximately relevant than to sit twiddling one's thumbs in front of a model which is felt to be more accurate but which one is unable to manipulate."

The second justification for normality is related to the link between the normal model and linear credibility. It was mentioned above that in a particular simple model, the linear least squares solution depended only upon the first two moments. It turns out that the Bayes solution for the same model with normal distributions is identical. Therefore, at least in this case, normality and linear least squares are equivalent. It has been shown that, in general, any model that is a member of the linear exponential family of distributions will produce the same result as the linear least squares solution (Ericson [9] and Jewell [17]). There has been speculation (Goel [11]) that the linear exponential family contains all the distributions with this property. So, to a certain extent, those who are willing to accept linear solutions should be equally comfortable with models from the linear exponential family. As far as choosing the normal distribution as the member to use, a second argument is needed. Most current practitioners estimate the variances using sums of squares. These estimates are unbiased for use. In Section 4, an indication of how one might "verify" the choice of model and prior is presented.

Three approaches can be taken to specifying $p(\mu,g,h)$. The first is to always use $p(\mu,g,h) = 1$. Morris [27] uses this prior in obtaining his parametric empirical Bayes results. The first thing to note is that since the support of (μ,g,h) is usually unbounded in at least one direction, this prior is not a proper probability distribution. Box and Tiao [3] argue that this is acceptable. Suppose μ is the average loss ratio over all rating classes. We can be virtually certain that this value is between 0 and 10. A uniform distribution over this interval combined with one that tails off slowly outside this interval would reflect the fact that very little is known about the true average loss ratio. Inferences that we would make using this prior would differ very little from those made using $p(\mu) = 1$ for $-\infty < \mu < \infty$. Two features of this approach should be noted. First, there is no guarantee that the posterior distribution of $\theta|x$ will exist. This would make it impossible to determine the posterior mean or variance. Second, the posterior mode is identical to the maximum likelihood estimator (after integrating out all nuisance parameters). In general, when this prior admits a solution it is quite reasonable.

The second school of thought is to find a general way of obtaining prior distributions that reflect minimal prior knowledge. Words such as "noninformative" or "reference" are often attached to such priors. The goal of research in this area is to find a way to automatically generate *the* noninformative prior for a given distribution. The fact that there is still disagreement on the appropriate reference prior for the probability of success in a sequence of Bernoulli trials (Geisser [10]) indicates how much work remains in this area. In the simple univariate case, Box and Tiao [3] support the prior $p(g) = 1/g$ when g is the variance. An extension is given by Tiao and Zellner [33] who argue that if g is a covariance matrix, the appropriate prior density is the inverse of its determinant.

The third belief is that only proper densities (those that integrate to 1) should be allowed for the prior distribution. Proponents of this approach insist that everyone has a prior distribution and it is just a matter of care and effort to bring it out. This makes excellent theoretical sense but is very difficult to implement. It is even more difficult to convince someone else that your opinion, as expressed by your prior distribution, is valid.

I have elected to take the middle ground. For credibility problems, the reciprocal prior for variances seems to be an appropriate choice for the prior density. This prior appears to be more "balanced" than the uniform one. Since the support is the interval from zero to infinity, we should expect that our prior opinion is equally apportioned between points near zero and those near infinity. The prior $1/g$ does this as it bounds an infinite area over all regions of the form $(0,a)$ and (a,∞). The uniform prior puts infinite probability only on the latter region. That is, it seems biased towards larger values of the variance. In Section 3, some brief attention will be given to a proper prior, so those who have one can still employ the methods to be discussed.

All of the ideas presented in this section other than the use of normality are summarized in the following quote (Berger [2]):

> "We would indeed argue that noninformative prior Bayesian analysis is the *single most powerful method of statistical analysis,* in the sense of being the *ad hoc* method most likely to yield a sensible answer for a given investment of effort." (author's italics)

2. THE HIERARCHICAL NORMAL LINEAR MODEL

In this section, the algebraic manipulations required to evaluate the three level hierarchical model are performed. Attention will be restricted to linear versions of the model. This is done mostly for computational convenience.

Before beginning the manipulations, a few notational items will be presented. Scalars will be represented by lower case letters. Vectors will be represented by bold face characters. Matrices will be represented by upper case letters. In classical statistics it is common to use upper case symbols to represent random variables. In a Bayesian analysis the various quantities are sometimes random and are sometimes fixed, so no attempt is made to use notation to identify random quantities. For example, in the model, the data are random and the parameters are fixed, but in the posterior, the parameters are random and the data are fixed. At times, the distribution of some parameters conditioned on others is needed. When examining a density function, the way to tell the fixed quantities from the random ones is to look at the left hand side. For example, $p(\boldsymbol{\theta}|\mathbf{y},G,H)$ indicates that in the function which follows, $\boldsymbol{\theta}$ is the random quantity and is the variable in the density, while $\mathbf{y}$, G, and H are fixed quantities. The density is for the indicated random variable, conditioned on the specific values given. When the two sides are separated by a proportionality symbol ($\propto$), the constant of proportionality may depend upon the conditional items. The constant can always be found by integrating the function with respect to the random elements.

2.1 The Model

The linear version of the three level Bayesian model is usually attributed to Lindley and Smith [24]. The three levels are:

$$y|\boldsymbol{\theta},G \sim N(A\boldsymbol{\theta},G) \qquad (N \times 1),$$

$$\boldsymbol{\theta}|\boldsymbol{\mu},H \sim N(B\boldsymbol{\mu},H) \qquad (k \times 1),$$

and

$$\boldsymbol{\mu} \sim N(\boldsymbol{\rho},C) \qquad (z \times 1),$$

where $\boldsymbol{\rho}$ and C are known and A and B (also known) are of full rank. A special case, and the only one considered here, is obtained by letting $C^{-1} \to 0$. This is equivalent to setting $p(\boldsymbol{\mu}) \propto \mathbf{1}$, the widely accepted noninformative prior for the mean. It is not necessary in this case to make any statement about $\boldsymbol{\rho}$. In most applications the covariance matrices G and H will not be known. It is then necessary to specify a prior distribution for them. Let $p(G,H)$ be the density for this prior distribution.

The standard credibility problem is to make inferences about $\boldsymbol{\theta}$, the expected losses (or loss ratios) for the various groups under consideration. The matrix A reflects the nature of the data collected. For example, there may be data from various years for each group. The second level indicates any relationships between the groups. One particular version of this model is analyzed in Sections 3 and 4; examples of other models are presented in Section 5. In any event, the objective of all the manipulations in this section is to obtain the posterior distribution and moments of $\boldsymbol{\theta}$ given the data y. Of less interest are the posterior distributions of G and H.

2.2 Three Helpful Mathematical Items

The first useful relationship is a matrix equation that is true for any symmetric non-singular matrix G; it will be used for completing the square.

$$x'Gx - 2x'B = (x - G^{-1}B)'G(x - G^{-1}B) - B'G^{-1}B.$$

The second item relates to the multivariate normal density. In general, the multivariate normal p.d.f. for a random variable with mean $\boldsymbol{\mu}$ and covariance C (a positive definite matrix) is

$$f(x) = (2\pi|C|)^{-1/2}\exp[-(x - \boldsymbol{\mu})'C^{-1}(x - \boldsymbol{\mu})/2],$$

where $|C|$ denotes the determinant of the matrix C. This implies that in general

$$\int\exp[-(x - \boldsymbol{\mu})'C^{-1}(x - \boldsymbol{\mu})/2]dx = (2\pi|C|)^{1/2}.$$

The final item is concerned with finding conditional densities. The general problem is the following: Let $f(a_1, \ldots a_m|b_1, \ldots, b_n)$ be proportional to the conditional density of $A_1, \ldots, A_m$ given $B_1 = b_1, \ldots, B_n = b_n$ (n may be 0). To find the conditional density of $A_1, \ldots, A_g$ given $B_1 = b_1, \ldots, B_h = b_h$ (where $g \leq m$ and $h \leq n$ and at least one of the inequalities is strict) evaluate $\int \ldots \int f(a_1, \ldots a_m|b_1, \ldots, b_n)da_{g+1} \ldots da_m$ and then drop all terms that involve only $b_1, \ldots, b_n$. (This latter step can be done first; additional terms can be eliminated after the integration.) The resulting function will be proportional to the desired density.

A related fact is that a conditional density is proportional to the conditional density in which some of the quantities on the left hand side of the "|" are moved to the right hand side. For example, $f(a_1, a_2|b_1, b_2)$ is proportional to the conditional density of A_1 given $A_2 = a_2$, $B_1 = b_1$, $B_2 = b_2$. Any factors that depend only on a_2 can be deleted.

2.3 Two Useful, Non-Bayesian Quantities

In the first level of the model, with G assumed known and $\boldsymbol{\theta}$ taken to represent a fixed, but unknown, parameter, the classical least-squares estimator of $\boldsymbol{\theta}$ is found by minimizing

$$\begin{aligned}&(y - A\boldsymbol{\theta})'\, G^{-1}\, (y - A\boldsymbol{\theta})\\ &= y'\, G^{-1}y - 2\boldsymbol{\theta}'\, A'\, G^{-1}y + \boldsymbol{\theta}'\, A'\, G^{-1}A\boldsymbol{\theta}\\ &= y'\, G^{-1}y +\\ &\quad [\boldsymbol{\theta} - (A'\, G^{-1}A)^{-1}\, A'\, G^{-1}y]'\, (A'\, G^{-1}A)\, [\boldsymbol{\theta} - (A'\, G^{-1}A)^{-1}\, A'\, G^{-1}y].\end{aligned}$$

Let $\Lambda = (A'\, G^{-1}A)^{-1}$, a positive definite matrix. Then the minimum must occur at $\hat{\boldsymbol{\theta}} = \Lambda A'\, G^{-1}y$.

Combining the first two levels gives

$$y|\boldsymbol{\mu}, G, H \sim N(AB\boldsymbol{\mu}, G + AHA').$$

The same manipulations yield

$$\hat{\boldsymbol{\mu}} = [B'A'(G + AHA')^{-1}AB]^{-1}B'A'(G + AHA')^{-1}y$$

and some matrix algebra produces the alternative form

$$\hat{\boldsymbol{\mu}} = [B'(H + \Lambda)^{-1}B]^{-1}B'(H + \Lambda)^{-1}\hat{\boldsymbol{\theta}}.$$

While the theory behind the above development is not germane to a Bayesian analysis, it is comforting to note that a Bayesian analysis often produces results that match those from classical theory. The quantities $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\mu}}$ will appear often in the analysis that follows, but will arise from a different algebraic procedure.

2.4 The Joint Density of $(y,\boldsymbol{\theta},\boldsymbol{\mu},G,H)$

The joint density involving all of the quantities from the three levels is the ideal place to begin. The last concept presented in Section 2.2 indicates that it is also the conditional density of any subset of the five variables given the remaining ones. The density is given by

$$p(y|\boldsymbol{\theta},G)p(\boldsymbol{\theta}|\boldsymbol{\mu},H)p(\boldsymbol{\mu})p(G,H)$$

which is proportional to (recalling that $p(\boldsymbol{\mu}) \propto \mathbf{1}$)

$$p(G,H)(2\pi)^{-(N+k)/2}(|G||H|)^{-1/2}$$
$$\times \exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - (\boldsymbol{\theta} - B\boldsymbol{\mu})'H^{-1}(\boldsymbol{\theta} - B\boldsymbol{\mu})/2].$$

It would be pleasant to proceed directly to the density of $\boldsymbol{\theta}$ given y. However, it is not possible to obtain this density analytically. Instead, begin by obtaining those conditional densities that are reasonably easy to derive. This is done in the next section. In the following section these densities are used to obtain the desired result.

2.5 Several Conditional Densities

The following subsections contain the derivations of a number of important conditional densities. The results are summarized in Section 2.5.6. Readers who are uninterested in the derivations can skip to that point.

2.5.1 Density of $\boldsymbol{\theta}|y,G,H$

This derivation is presented in great detail in order to indicate how these calculations are done. Since the joint density from Section 2.4 is also the conditional density of $(\boldsymbol{\theta},\boldsymbol{\mu}|y,G,H)$ it is only necessary to integrate $\boldsymbol{\mu}$ out of the joint density. Begin by removing terms involving only y,G, H, and constants. The remaining part of the joint density is

$$\exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - (\boldsymbol{\theta} - B\boldsymbol{\mu})'H^{-1}(\boldsymbol{\theta} - B\boldsymbol{\mu})/2]$$
$$= \exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'H^{-1}\boldsymbol{\theta}/2 + 2\boldsymbol{\mu}'\mathrm{B}'\mathrm{H}^{-1}\boldsymbol{\theta}/2$$
$$- \boldsymbol{\mu}'B'H^{-1}B\boldsymbol{\mu}/2].$$

Let $\Xi = (B'H^{-1}B)^{-1}$. Completing the square with respect to $\boldsymbol{\mu}$ in the above expression yields

$$\exp[(-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'H^{-1}\boldsymbol{\theta}/2$$
$$- (\boldsymbol{\mu} - \Xi B'H^{-1}\boldsymbol{\theta})'\Xi^{-1}(\boldsymbol{\mu} - \Xi B'H^{-1}\boldsymbol{\theta})/2 + \boldsymbol{\theta}'H^{-1}B\Xi B'H^{-1}\boldsymbol{\theta}/2].$$

Now integrate with respect to $\boldsymbol{\mu}$. $\boldsymbol{\mu}$ only appears as the quadratic form in the third term and upon integration produces only constants and the determinant of Ξ. Since Ξ is a function of H only, it can be dropped. Therefore the density is proportional to

$$\exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'H^{-1}\boldsymbol{\theta}/2 + \boldsymbol{\theta}'H^{-1}B\Xi B'H^{-1}\boldsymbol{\theta}/2].$$

Expand the first term and remove the part not involving $\boldsymbol{\theta}$. The result is now

$$\exp[-\boldsymbol{\theta}'(A'G^{-1}A - H^{-1}B\Xi B'H^{-1} + H^{-1})\boldsymbol{\theta}/2 + \boldsymbol{\theta}'A'G^{-1}y].$$

Let $V^{-1} = A'G^{-1}A - H^{-1}B\Xi B'H^{-1} + H^{-1}$.
Complete the square on $\boldsymbol{\theta}$ to obtain

$$\exp[-(\boldsymbol{\theta} - VA'G^{-1}y)'V^{-1}(\boldsymbol{\theta} - VA'G^{-1}y)/2 + y'G^{-1}AVA'G^{-1}y/2].$$

Since the second term does not depend on $\boldsymbol{\theta}$ it can be dropped. The final result is

$$p(\boldsymbol{\theta}|y,G,H) \propto \exp[-(\boldsymbol{\theta} - VA'G^{-1}y)'V^{-1}(\boldsymbol{\theta} - VA'G^{-1}y)/2].$$

By inspection it is immediately apparent that

$$\boldsymbol{\theta}|y,G,H \sim N(VA'G^{-1}y,V).$$

Let $\tilde{\boldsymbol{\theta}} = VA'G^{-1}y$ be the conditional mean. It can be rewritten as

$$\tilde{\boldsymbol{\theta}} = (H^{-1} + \Lambda^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\hat{\boldsymbol{\mu}}).$$

This is the customary weighted average common in a Bayesian analysis. It is also a (linear) credibility formula. In fact, this is the result that arises from an EB analysis. As discussed in Section 1, the point of departure is the treatment of the unknown G and H.

2.5.2 *Density of* $\boldsymbol{\mu}|y,G,H$

This calculation is included mostly for completeness. It is not used in any subsequent work. The procedure is exactly the same as that used above, only now integrate out $\boldsymbol{\theta}$ instead of $\boldsymbol{\mu}$. The result is

$$\boldsymbol{\mu}|y,G,H \sim N(\hat{\boldsymbol{\mu}},[B'(\Lambda + H)^{-1}B]^{-1}).$$

2.5.3 *Density of* $\boldsymbol{\theta},G,H|y$

To find this density, integrate $\boldsymbol{\mu}$ out of the joint density. The difference between this calculation and the one in Section 2.5.1 is that terms involving G

and H must be retained. The result is

$$p(\boldsymbol{\theta},G,H|y) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Xi|^{1/2} \times \exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'H^{-1}\boldsymbol{\theta}/2 + \boldsymbol{\theta}'H^{-1}B\Xi B'H^{-1}\boldsymbol{\theta}/2].$$

Let $Q = H^{-1} - H^{-1}B\Xi B'H^{-1}$. Then

$$p(\boldsymbol{\theta},G,H|y) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Xi|^{1/2} \times \exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'Q\boldsymbol{\theta}/2].$$

2.5.4 Density of $\boldsymbol{\mu},G,H|y$

Begin by integrating $\boldsymbol{\theta}$ out of the joint density, once again retaining terms involving G and H.

$$\begin{aligned} p(\boldsymbol{\mu},G,H|y) &\propto p(G,H)|G|^{-1/2}|H|^{-1/2}\int \exp[-yG^{-1}y/2 + 2\boldsymbol{\theta}A'G^{-1}y/2 \\ &\quad - \boldsymbol{\theta}'A'G^{-1}A\boldsymbol{\theta}/2 - \boldsymbol{\theta}'H^{-1}\boldsymbol{\theta}/2 + 2\boldsymbol{\theta}'H^{-1}B\boldsymbol{\mu}/2 \\ &\quad - \boldsymbol{\mu}'B'H^{-1}B\boldsymbol{\mu}/2]d\boldsymbol{\theta} \\ &\propto p(G,H)|G|^{-1/2}|H|^{-1/2}\int \exp[-yG^{-1}y/2 - \boldsymbol{\theta}'(\Lambda^{-1} + H^{-1})\boldsymbol{\theta}/2 \\ &\quad + 2\boldsymbol{\theta}'(A'G^{-1}y + H^{-1}B\boldsymbol{\mu})/2 - \boldsymbol{\mu}'\Xi^{-1}\boldsymbol{\mu}/2]d\boldsymbol{\theta}. \end{aligned}$$

The third term in the exponent can be written $\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\boldsymbol{\mu}$. Complete the square to obtain

$$\begin{aligned} p(\boldsymbol{\mu},G,H|y) &\propto p(G,H)|G|^{-1/2}|H|^{-1/2}\int \exp\{-y'G^{-1}y/2 \\ &\quad - [\boldsymbol{\theta} - (\Lambda^{-1} + H^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\boldsymbol{\mu})]'(\Lambda^{-1} + H^{-1}) \\ &\quad \times [\boldsymbol{\theta} - (\Lambda^{-1} + H^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\boldsymbol{\mu})] \\ &\quad + (\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\boldsymbol{\mu})'(\Lambda^{-1} + H^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\boldsymbol{\mu}) \\ &\quad - \boldsymbol{\mu}'\Xi^{-1}\boldsymbol{\mu}/2\}d\boldsymbol{\theta} \\ &\propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Lambda^{-1} + H^{-1}|^{-1/2} \times \exp[-y'G^{-1}y/2 \\ &\quad - \boldsymbol{\mu}'\Xi^{-1}\boldsymbol{\mu}/2 + \boldsymbol{\mu}'B'H^{-1}(\Lambda^{-1} + H^{-1})^{-1}H^{-1}B\boldsymbol{\mu}/2 \\ &\quad + 2\boldsymbol{\mu}B'H^{-1}(\Lambda^{-1} + H^{-1})^{-1}\Lambda^{-1}\hat{\boldsymbol{\theta}}/2 \\ &\quad + \hat{\boldsymbol{\theta}}'\Lambda^{-1}(\Lambda^{-1} + H^{-1})^{-1}\Lambda^{-1}\hat{\boldsymbol{\theta}}/2]. \end{aligned}$$

Now use the three identities

$$\begin{aligned} &H^{-1}(\Lambda^{-1} + H^{-1})^{-1}H^{-1} = H^{-1} - (\Lambda + H)^{-1}, \\ &H^{-1}(\Lambda^{-1} + H^{-1})^{-1}\Lambda^{-1} = (\Lambda + H)^{-1}, \text{ and} \\ &\Lambda^{-1}(\Lambda^{-1} + H^{-1})^{-1}\Lambda^{-1} = \Lambda^{-1} - (\Lambda + H)^{-1}. \end{aligned}$$

Let $\Psi = \Lambda + H$. Then,

$$p(\boldsymbol{\mu},G,H|\mathbf{y}) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Lambda^{-1} + H^{-1}|^{-1/2} \times \exp[-y'G^{-1}y/2 - \boldsymbol{\mu}'\Xi^{-1}\boldsymbol{\mu}/2 + \boldsymbol{\mu}'B'H^{-1}B\boldsymbol{\mu}/2 - \boldsymbol{\mu}'B'\Psi^{-1}B\boldsymbol{\mu}/2 + 2\boldsymbol{\mu}'B'\Psi^{-1}\hat{\boldsymbol{\theta}}/2 + \hat{\boldsymbol{\theta}}'\Lambda^{-1}\hat{\boldsymbol{\theta}}/2 - \hat{\boldsymbol{\theta}}'\Psi^{-1}\hat{\boldsymbol{\theta}}/2].$$

The second and third terms in the exponent cancel. From Section 2.3, $(B'\Psi^{-1}B)\hat{\boldsymbol{\mu}} = B'\Psi^{-1}\hat{\boldsymbol{\theta}}$. Complete the square to obtain

$$p(\boldsymbol{\mu},G,H|\mathbf{y}) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Lambda^{-1} + H^{-1}|^{-1/2} \times \exp[-y'G^{-1}y/2 - (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'(B'\Psi^{-1}B)(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})/2 + \hat{\boldsymbol{\mu}}'(B'\Psi^{-1}B)\hat{\boldsymbol{\mu}}/2 + \hat{\boldsymbol{\theta}}'\Lambda^{-1}\hat{\boldsymbol{\theta}}/2 - \hat{\boldsymbol{\theta}}'\Psi^{-1}\hat{\boldsymbol{\theta}}/2].$$

2.5.5 Density of G,H|y

Integrate $\boldsymbol{\mu}$ out of the density in Section 2.5.4 to obtain

$$p(G,H|\mathbf{y}) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Lambda^{-1} + H^{-1}|^{-1/2}|B'\Psi^{-1}B|^{-1/2} \times \exp[-y'G^{-1}y/2 + \hat{\boldsymbol{\theta}}'\Lambda^{-1}\hat{\boldsymbol{\theta}}/2 - \hat{\boldsymbol{\theta}}'\Psi^{-1}\hat{\boldsymbol{\theta}}/2 + \hat{\boldsymbol{\mu}}'(B'\Psi^{-1}B)\hat{\boldsymbol{\mu}}/2].$$

In the second term of the exponent write $\hat{\boldsymbol{\theta}}$ in terms of $\mathbf{y}$. Then complete the squares to obtain

$$p(G,H|\mathbf{y}) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Lambda^{-1} + H^{-1}|^{-1/2}|B'\Psi^{-1}B|^{-1/2} \times \exp[-(y - A\hat{\boldsymbol{\theta}})'G^{-1}(y - A\hat{\boldsymbol{\theta}})/2 - (\hat{\boldsymbol{\theta}} - B\hat{\boldsymbol{\mu}})'\Psi^{-1}(\hat{\boldsymbol{\theta}} - B\hat{\boldsymbol{\mu}})/2].$$

The two terms in the exponent are the within and between sums of squares, respectively. Both depend on the unknown variances, G and H. While it is once again comforting to note that frequentist quantities have appeared in the Bayesian development, we should keep in mind that these quantities have no special meaning. The determinants can be rewritten as

$$(|G^{-1}|/|A'G^{-1}A|)^{1/2}(|\Psi^{-1}|/|B'\Psi^{-1}B|)^{1/2}.$$

The two numerator terms can form the basis for a prior distribution on (G,H). This is somewhat consistent with the ideas presented in Box and Tiao [3] and in Tiao and Zellner [33].

2.5.6 Summary

The important matrices and distributions from this section are repeated for convenience:

$$\Lambda = (A'G^{-1}A)^{-1}$$

$$\Xi = (B'H^{-1}B)^{-1}$$

$$\Psi = \Lambda + H$$
$$Q = H^{-1} - H^{-1}B\Xi B'H^{-1}$$
$$V = [A'G^{-1}A - H^{-1}B\Xi B'H^{-1} + H^{-1}]^{-1} = (\Lambda^{-1} + Q)^{-1}$$
$$\hat{\boldsymbol{\theta}} = \Lambda A'G^{-1}y$$
$$\hat{\boldsymbol{\mu}} = [B'(H + \Lambda)^{-1}B]^{-1}B'(H + \Lambda)^{-1}\hat{\boldsymbol{\theta}}$$
$$\tilde{\boldsymbol{\theta}} = (H^{-1} + \Lambda^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\hat{\boldsymbol{\mu}})$$
$$\boldsymbol{\theta}|y,G,H \sim N(\tilde{\boldsymbol{\theta}},V)$$
$$p(\boldsymbol{\theta},G,H|y) \propto p(G,H)|G|^{-1/2}|H|^{-1/2}|\Xi|^{1/2} \times \exp[-(y - A\boldsymbol{\theta})'G^{-1}(y - A\boldsymbol{\theta})/2 - \boldsymbol{\theta}'Q\boldsymbol{\theta}/2]$$
$$p(G,H|y) \propto p(G,H)|G|^{-1/2}|\Lambda|^{1/2}|\Psi|^{-1/2}|B'\Psi^{-1}B|^{-1/2} \times \exp[-(y - A\hat{\boldsymbol{\theta}})'G^{-1}(y - A\hat{\boldsymbol{\theta}})/2 - (\hat{\boldsymbol{\theta}} - B\hat{\boldsymbol{\mu}})'\Psi^{-1}(\hat{\boldsymbol{\theta}} - B\hat{\boldsymbol{\mu}})/2].$$

2.6 *Two Empirical Bayes Approaches to Estimating* $\boldsymbol{\theta}$

As introduced in Section 1.2.2, the EB approach begins by finding the posterior mean of $\boldsymbol{\theta}$ given the covariance terms G and H. In the HNLM this is $\tilde{\boldsymbol{\theta}}$. External estimates are then found for G and H. In this section, two general approaches to finding such estimates are introduced.

The first method uses the posterior density $p(G,H|y)$. Either the mean or the mode could be used as the estimate. The mean is superior in that it is guaranteed to be in the interior of the parameter space. The mode is often easier to compute, but may be on a boundary. Either estimate usually requires a numerical evaluation. Specific formulas for a simple model are presented in Section 3.8.

The second method is an iterative technique. Begin with a preliminary estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$. Then in $p(\boldsymbol{\theta},G,H|y)$ hold $\boldsymbol{\theta}$ fixed at its current value and find the values of G and H that maximize this density. Obtain a revised estimate of $\boldsymbol{\theta}$ by evaluating $\tilde{\boldsymbol{\theta}}$ at the values of G and H just obtained. Repeat this procedure until $\tilde{\boldsymbol{\theta}}$,$G$, and H stabilize. It is not entirely clear what the results mean, but the procedure is similar to that recommended by Morris [27]. Computationally, this tends to be the simplest approach, as the maximization can often be done analytically. This is demonstrated for a simple model in Section 3.5. If an analytical approach is not possible, an all-purpose maximization method like that of Nelder and Mead [29] is likely to provide the answer.

Recall that one of the drawbacks of the usual EB method is its inability to produce variance estimates. In Section 3.8, it is shown how this can be done when G and H are estimated by the posterior mean.

2.7 Finding Posterior Quantities by Integration

All of the integrations done up to this point were easy to accomplish by completing the square. The one needed to obtain the posterior density of $\boldsymbol{\theta}|y$ is found from

$$p(\boldsymbol{\theta}|y) = \int p(\boldsymbol{\theta}|G,H,y)p(G,H|y)dGdH.$$

The first density in the integrand is a multivariate normal density and was obtained in Section 2.5.1. The second density was obtained in Section 2.5.5 up to a constant of proportionality. We must obtain that constant in order to insert the exact density in the integral above. That can be found by integrating the expression found in Section 2.5.5 with respect to both G and H. These two integrals are of equal difficulty and usually must be done numerically. The degree of difficulty will depend on the form of the covariance matrices G and H and the prior density $p(G,H)$. It will be seen in Section 3 that in a specific case the problem can be analytically reduced to a one-dimensional numerical integration. Some excellent procedures for performing multidimensional numerical integration are given in Smith, Skene, Shaw, Naylor, and Dransfield [32].

In most applications, the vector $\boldsymbol{\theta}$ will be of a reasonably high dimension, certainly greater than two. It is unlikely that much insight will be gained by examining the posterior density. The remainder of this section is devoted to obtaining various summary quantities. This will conclude the development of the general hierarchical normal linear model.

2.7.1 Posterior Mean of $\theta_i|y$

One way to obtain this quantity would be to evaluate the following integral:

$$\int \theta_i p(\boldsymbol{\theta}|y)d\boldsymbol{\theta}.$$

Given the fact that a numerical step is necessary to yield each evaluation of the integrand, the cost of performing this integration is likely to be quite high. Instead, employ the following result (this notion was introduced in Section 1.2.3):

$$\mathrm{E}(\theta_i|y) = \mathrm{E}[\mathrm{E}(\theta_i|G,H,y)] = \mathrm{E}(\tilde{\theta}_i|y) = \int \tilde{\theta}_i p(G,H|y)dGdH.$$

Note that $\tilde{\theta}_i$ is a function of G and H.

2.7.2 Posterior Variance of $\theta_i|y$

A similar argument yields the following.

$$\begin{aligned}\mathrm{Var}(\theta_i|y) &= \mathrm{Var}[\mathrm{E}(\theta_i|G,H,y)] + \mathrm{E}[\mathrm{Var}(\theta_i|G,H,y)] \\ &= \mathrm{Var}(\tilde{\theta}_i|y) + \mathrm{E}(v_{ii}|y) \\ &= \int(\tilde{\theta}_i)^2 p(G,H|y)dGdH - [\mathrm{E}(\theta_i|y)]^2 + \int v_{ii}p(G,H|y)dGdH\end{aligned}$$

where v_{ii} is the i^{th} diagonal element of the matrix V introduced in Section 2.5.1.

2.7.3 Posterior Density of $\theta_i|y$

This univariate density could be plotted to provide insight about a particular group mean. An approximate integration needs to be performed to get each point from the posterior density. The formula is

$$p(\theta_i|y) = \int p(\theta_i|G,H,y)p(G,H|y)dGdH.$$

The first density is a univariate normal density with mean $\tilde{\theta}_i$ and variance v_{ii}.

If this calculation appears to be too time-consuming, the posterior distribution may be approximated by a normal distribution with moments as given in Sections 2.7.1 and 2.7.2. This result is given in Berger [2] and is a Bayesian version of the central limit theorem. The same result applies in the following section.

2.7.4 Predictive Density of a Future Observation

In the insurance setting it may be more useful to get information about the losses in a future period than to estimate the class mean. Such a calculation would incorporate both the uncertainty with respect to the group mean and the uncertainty about the experience of next year's insureds.

In general, consider a new observation, $\boldsymbol{x} \sim N(A_x\boldsymbol{\theta},C_x)$ where C_x will depend in some way on the elements of G. A typical example would have A_x be a $1 \times k$ vector of zeros with a one in the i^{th} column. This would make x (a scalar) an observation from the i^{th} group. The matrix C_x would be a scalar of the form σ^2/P where σ^2 is the variance from the original model and P is a measure of exposure for the year to come.

The density of interest is

$$p(\boldsymbol{x}|y) = \int p(\boldsymbol{x}|\boldsymbol{\theta},G,H,y)p(\boldsymbol{\theta},G,H|y)d\boldsymbol{\theta}dGdH.$$

This is likely to be difficult to obtain. It is much easier to get the moments.

$$\begin{aligned} \mathrm{E}[\boldsymbol{x}|\boldsymbol{y}] &= \mathrm{E}[\mathrm{E}(\boldsymbol{x}|\boldsymbol{\theta},G,H,\boldsymbol{y})] \\ &= \mathrm{E}[A_x\boldsymbol{\theta}|\boldsymbol{y}] = A_x\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}]. \end{aligned}$$

The expectation can be found using the formula in Section 2.7.1 since each element of the vector of expectations can be found individually.

For the variance

$$\begin{aligned} \mathrm{Cov}[\boldsymbol{x}|\boldsymbol{y}] &= \mathrm{Cov}[\mathrm{E}(\boldsymbol{x}|\boldsymbol{\theta},G,H,\boldsymbol{y})] + \mathrm{E}[\mathrm{Cov}(\boldsymbol{x}|\boldsymbol{\theta},G,H,\boldsymbol{y})] \\ &= \mathrm{Cov}[A_x\boldsymbol{\theta}|\boldsymbol{y}] + \mathrm{E}[C_x|\boldsymbol{y}] \\ &= A_x\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}]A_x' + \mathrm{E}[C_x|\boldsymbol{y}]. \end{aligned}$$

The covariance requires evaluation of $\mathrm{Cov}(\theta_i,\theta_j|\boldsymbol{y})$. This can be done from $\int v_{ij}p(G,H|\boldsymbol{y})dGdH$. The $(ij)^{\text{th}}$ term of the expected value is evaluated as $\int (C_x)_{ij}p(G,H|\boldsymbol{y})dGdH$.

3. THE ONE-WAY MODEL

In this section a specific hierarchical model is investigated. It is similar to the model treated by Bühlmann and Straub [7] in their EB analysis and is appropriate when there are k identically distributed groups and the goal is the simultaneous estimation of their means. The three levels of the one-way model are

Level 1—$y_{ij}|\theta_i,\sigma^2 \sim N(\theta_i,\sigma^2/P_{ij})$ $i = 1, \ldots, k$ $j = 1, \ldots, n_i$

Level 2—$\theta_i|\mu,\tau^2 \sim N(\mu,\tau^2)$

Level 3—$\mu \sim N(0,\infty)$.

The random variables at each level are conditionally independent and P_{ij} is some measure of exposure. The usual situation is that y_{ij} is the average loss (or loss ratio) in year j for class i. This differs from the Bühlmann-Straub model in just one respect. In their model, the level one variances were allowed to differ from class to class. Their result, however, uses only the average of these variances. That is, at no point is this variability taken into account. An indication of how one could truly account for unequal variances is given in Section 5.

To use the formulas of the previous section it is necessary to identify the various matrices and vectors. Begin by letting $\boldsymbol{y}$ be the $N \times 1$ vector of the observations where $N = \Sigma n_i$. Arrange the observations so $y_{11}, \ldots, y_{1n_1}$ appears first, followed by $y_{21}, \ldots, y_{2n_2}$, and so forth. The matrix A is $N \times k$ and contains only zeros and ones. In the first column, the ones are in the first n_1 rows. In the second column, the ones are in rows $n_1 + 1$ through $n_1 + n_2$, and so forth. The vector $\boldsymbol{\theta}$ is $k \times 1$ and contains the unknown group means, $\theta_1, \ldots, \theta_k$. The covariance matrix G is diagonal with diagonal elements running from σ^2/P_{11} in the upper left corner to σ^2/P_{kn_k} in the lower right corner. At the second level, B is a $k \times 1$ vector consisting entirely of ones. Let $\mathbf{1}$ indicate such a vector. The vector $\boldsymbol{\mu}$ is a scalar and so will be written μ. The covariance matrix H is diagonal with all elements equal to τ^2, that is, $H = \tau^2 I_k$.

The exposition will proceed in four steps. The first is a development of a pair of useful matrix relationships. They will aid in the evaluation of the determinants and inverses. Next, prior distributions for σ^2 and τ^2 are introduced. The third step is to obtain the conditional densities. The final step is to perform the integrations.

3.1 Two Useful Matrix Facts

If A is a nonsingular matrix and $\boldsymbol{c}$ and $\boldsymbol{d}$ are vectors, then

$$|A + \boldsymbol{c}\boldsymbol{d}'| = |A|(1 + \boldsymbol{d}'A^{-1}\boldsymbol{c}).$$

If the determinant is non-zero then

$$(A + \boldsymbol{c}\boldsymbol{d}')^{-1} = A^{-1} - (A^{-1}\boldsymbol{c}\boldsymbol{d}'A^{-1})/(1 + \boldsymbol{d}'A^{-1}\boldsymbol{c}).$$

For the special case where A is diagonal $(a_1, \ldots, a_k)$, $\boldsymbol{c} = c\mathbf{1}$, and $\boldsymbol{d} = \mathbf{1}$, the results are (where $J_k = \mathbf{1}\mathbf{1}'$, a $k \times k$ matrix)

$$|A + cJ_k| = (\Pi a_i)(1 + c\Sigma a_i^{-1})$$

and

$$(A + cJ_k)^{-1} \text{ has } (ii)^{\text{th}} \text{ term } a_i^{-1} - c/(a_i^2 b) \text{ and } (ij)^{\text{th}} \text{ term } -c/(a_i a_j b)$$

where $b = 1 + c\Sigma a_i^{-1}$.

Derivations of these results can be found in Graybill [12] (Theorem 8.9.3).

3.2 Prior Densities for (G,H)

In this section, three noninformative priors and one proper prior will be introduced. Two of the noninformative priors are based on a general theory that can be used in any setting of the HNLM. The third one is particular to the one-way model.

The easiest one to describe is the naive version of a noninformative prior. It is $p(G,H) \propto 1$. In the one-way model the only random elements are σ^2 and τ^2, so the actual prior in this case is $p(\sigma^2,\tau^2) \propto 1$. This prior is used by Morris [28]. While it is convenient for computational purposes, there are good theoretical reasons (Box and Tiao [3]) for not using it. The essence of the argument is that this prior puts too much weight on large values of the parameters. On the other hand, it is often the case that this improper prior will yield a proper posterior (something that must always be checked when using a noninformative prior). Also, the posterior mode is the maximum likelihood estimate. This should give comfort to those who are troubled by Bayesian methods. Call this prior 1.

The second prior is based on the arguments of Box and Tiao [3] for the balanced model. The balanced model is the special case where $n_1 = \ldots = n_k$ and $P_{11} = \ldots = P_{kn_k}$. The first requirement is common in insurance studies, as n_i often is the number of years of observation for the i^{th} class. However, it is extremely unlikely that the exposures will be equal for all years and all classes. In any event, in the one-way balanced model the reasonable noninformative prior is $p(\sigma^2,\tau^2) \propto (\sigma^2)^{-1}(\sigma^2 + Pn\tau^2)^{-1}$ where P is the common value of the P_{ij} and n is the common value of the n_i. A generalization for the unbalanced model is to use $p(\sigma^2,\tau^2) \propto (\sigma^2)^{-1}(\sigma^2 + m\tau^2)^{-1}$. Since m is to play the role of Pn one choice is $m = \Sigma P_{ij}/k$. An alternative is taken from the constant used when creating the unbiased frequentist estimator of τ^2. It is $m = [(\Sigma P_{ij})^2 - \Sigma(P_i)^2]/(k - 1)\Sigma P_{ij}$ where $P_i = \Sigma P_{ij}$. Call this (with arbitrary m) prior 2.

The final noninformative prior is, like the first one, available in all situations. In general, it is

$$p(G,H) \propto |G|^{-p\dim(G)/\dim(G)}|\Lambda + H|^{-p\dim(H)/\dim(H)}.$$

This prior is taken after Box and Tiao [3] and is related to the Fisher information about G and H. In the above expression, $p\dim(G)$ refers to the number of distinct parameters in the matrix G while $\dim(G)$ is the number of rows in G. For the one-way model, $p(\sigma^2,\tau^2) \propto (\sigma^2)^{-1}[\Pi(\sigma^2 + P_i\tau^2)]^{-1/k}$. In the balanced case this prior is identical to prior 2. Call it prior 3.

The fourth and final prior is an attempt to offer a proper distribution. As such, it requires that the investigator have a genuine opinion about the variances. When seeking a proper prior, mathematical convenience is always a high priority. At the very least, the family of proper priors should include a sufficiently large variety of possibilities so as to give the investigator a chance of finding a representative prior. The natural choice for variances is the inverse gamma distribution. The general form of the density is

$$p(x) \propto x^{-\nu}\exp(-\lambda/x).$$

For it to be a proper distribution we must have $\nu > 1$ and $\lambda > 0$. The limiting case of $\nu = 1$ and $\lambda = 0$ is similar to prior 2. Since prior 1 is equivalent to $\nu = 0$ and $\lambda = 0$ we see how far from being a proper distribution prior 1 is. The inverse gamma prior will be referred to as prior 4. A specific version appropriate for the one-way model will be given later.

3.3 Conditional Densities in the One-Way Model

This section contains all the details of the evaluation of the formulas in Section 2 in the special case of the one-way model.

3.3.1 Preliminary Quantities

$\Lambda = \text{diagonal } (\sigma^2/P_1, \ldots, \sigma^2/P_k)$ where $P_i = \Sigma_j P_{ij}$

$\Xi = \tau^2/k$ (a scalar)

$\hat{\theta}_i = \Sigma_j P_{ij} y_{ij}/P_i$

$\hat{\mu} = \Sigma w_i \hat{\theta}_i / w.$ where $w_i = P_i\tau^2/(\sigma^2 + P_i\tau^2)$ and $w. = \Sigma_i w_i$.

3.3.2 $\boldsymbol{\theta}|\boldsymbol{y},G,H$

This is a multivariate normal random variable. The mean vector $\tilde{\boldsymbol{\theta}}$ has i^{th} element

$$\tilde{\theta}_i = w_i\hat{\theta}_i + (1 - w_i)\hat{\mu}.$$

The matrix V^{-1} is

$$\begin{aligned} V^{-1} &= \text{diagonal } (\tau^{-2} + \sigma^{-2}P_i) - (k\tau^2)^{-1}J_k \\ &= \text{diagonal } [\tau^{-2}(1 - w_i)^{-1}] - (k\tau^2)^{-1}J_k, \end{aligned}$$

where as before J_k is a $k \times k$ matrix of 1's. The covariance matrix is V. Using the inversion formula from Section 3.1,

$$v_{ii} = \text{Var}(\theta_i|\boldsymbol{y},\sigma^2,\tau^2) = \tau^2(1 - w_i)[1 + (1 - w_i)/w.] \text{ and}$$

$$v_{ij} = \text{Cov}(\theta_i,\theta_j|\boldsymbol{y},\sigma^2,\tau^2) = \tau^2(1 - w_i)(1 - w_j)/w. \,.$$

3.3.3 $\boldsymbol{\theta},G,H|\boldsymbol{y}$

The two required sums of squares are

$$\begin{aligned}(\boldsymbol{y} - A\boldsymbol{\theta})'G^{-1}(\boldsymbol{y} - A\boldsymbol{\theta}) &= \Sigma_{ij}P_{ij}(y_{ij} - \theta_i)^2/\sigma^2 \\ &= [\Sigma_{ij}P_{ij}(y_{ij})^2 - \Sigma_i P_i(\hat{\theta}_i)^2 + \Sigma_i P_i(\theta_i - \hat{\theta}_i)^2]/\sigma^2\end{aligned}$$

and

$$\boldsymbol{\theta}'Q\boldsymbol{\theta} = [\Sigma_i(\theta_i)^2 - k\bar{\theta}^2]/\tau^2 \text{ where } \bar{\theta} = \Sigma_i\theta_i/k.$$

The second version of the first quantity is useful for computational purposes as the first two sums depend only on the data while the last sum has only k terms. The desired density is

$$\begin{aligned}p(\boldsymbol{\theta},\sigma^2,\tau^2|\boldsymbol{y}) \propto\; & p(\sigma^2,\tau^2)(\sigma^2)^{-N/2}(\tau^2)^{-(k-1)/2} \\ & \times \exp\{-\Sigma_{ij}P_{ij}(y_{ij} - \theta_i)^2/2\sigma^2 - (\Sigma_i(\theta_i)^2 - k\bar{\theta}^2)/2\tau^2\}.\end{aligned}$$

3.3.4 $G,H|\boldsymbol{y}$

Two important matrices are

$$\Psi = \text{diag}(\sigma^2/P_i + \tau^2) = \tau^2\text{diag}(1/w_i)$$

and

$$B'\Psi^{-1}B = w./\tau^2.$$

The two sums of squares are

$$(\boldsymbol{y} - A\hat{\boldsymbol{\theta}})'G^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\theta}}) = \Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2/\sigma^2$$

and

$$(\hat{\boldsymbol{\theta}} - B\hat{\mu})'\Psi^{-1}(\hat{\boldsymbol{\theta}} - B\hat{\mu}) = \Sigma_i w_i(\hat{\theta}_i - \hat{\mu})^2/\tau^2.$$

The desired density is

$$\begin{aligned}p(\sigma^2,\tau^2|\boldsymbol{y}) \propto\; & p(\sigma^2,\tau^2)(\sigma^2)^{-(N-k)/2}(\tau^2)^{-(k-1)/2}[\Pi_i w_i/w.]^{1/2} \\ & \times \exp[-\Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2/2\sigma^2 - \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i/2\tau^2].\end{aligned}$$

For computation the exponent can be written

$$-[\Sigma_{ij}P_{ij}(y_{ij})^2 - \Sigma_i P_i(\hat{\theta}_i)^2]/2\sigma^2 - [\Sigma_i w_i(\hat{\theta}_i)^2 - (\Sigma_i w_i\hat{\theta}_i)^2/w.]/2\tau^2.$$

3.3.5 Prior Distributions for (σ^2,τ^2)

To make this section complete, the four priors developed in Section 3.2 are displayed.

Prior 1—$p(\sigma^2,\tau^2) \propto 1$.

Prior 2—$p(\sigma^2,\tau^2) \propto (\sigma^2)^{-1}(\sigma^2 + m\tau^2)^{-1}$.

Prior 3—$p(\sigma^2,\tau^2) \propto (\sigma^2)^{-1}[\Pi_i(\sigma^2 + P_i\tau^2)]^{-1/k}$.

Prior 4—$p(\sigma^2,\tau^2) \propto (\sigma^2)^{-\nu_1}(\tau^2)^{-\nu_2}\exp(-\lambda_1/\sigma^2 - \lambda_2/\tau^2)$.

Prior 4 uses two independent inverse chi-square random variables. Prior 1 is a special case. In the next section the four priors will be written with one general formula.

3.4 A Transformation

It turns out that calculations are much easier to carry out with a transformation of σ^2 and τ^2. The one to use is

$$\delta = \tau^2/\sigma^2 \text{ and } \alpha = \sigma^2.$$

The Jacobian for this transformation is α. The four prior densities become

Prior 1—$p(\alpha,\delta) \propto \alpha$.

Prior 2—$p(\alpha,\delta) \propto \alpha^{-1}(1 + m\delta)^{-1}$.

Prior 3—$p(\alpha,\delta) \propto \alpha^{-1}[\Pi_i(1 + P_i\delta)]^{-1/k}$.

Prior 4—$p(\alpha,\delta) \propto \alpha^{-(\nu_1+\nu_2-1)}\delta^{-\nu_2}\exp(-\lambda_1/\alpha - \lambda_2/\alpha\delta)$.

A general form that includes all four is

$$p(\alpha,\delta) \propto \alpha^{-q/2}h(\delta)\exp(-\lambda_1/\alpha - \lambda_2/\alpha\delta).$$

The two important conditional densities become (note that since the Jacobian was included in the prior, no other adjustments are needed)

$$\begin{aligned} p(\boldsymbol{\theta},\alpha,\delta|y) \propto\ & (\alpha)^{-(N+k+q-1)/2}(\delta)^{-(k-1)/2}h(\delta) \\ & \times \exp\{-[\lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \theta_i)^2]/2\alpha \\ & - [\lambda_2 + (\Sigma_i(\theta_i)^2 - k\bar{\theta}^2)]/2\alpha\delta\} \end{aligned}$$

and

$$\begin{aligned} p(\alpha,\delta|y) \propto\ & (\alpha)^{-(N+q-1)/2}(\delta)^{-(k-1)/2}h(\delta)\ [\Pi_i w_i/w.]^{1/2} \\ & \times \exp\{-[\lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2]/2\alpha \\ & - [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/2\alpha\delta\}. \end{aligned}$$

The other important quantity is the distribution of $\boldsymbol{\theta}|\boldsymbol{y},\alpha,\delta$. From Section 3.3.2, it is multivariate normal. The i^{th} element of the mean vector is

$$\tilde{\theta}_i = w_i\hat{\theta}_i + (1 - w_i)\hat{\mu} \text{ where } w_i = P_i\delta/(1 + P_i\delta) \text{ and } \hat{\mu} = \Sigma_i w_i\hat{\theta}_i/w. \,.$$

The covariance matrix has diagonal elements

$$v_{ii} = \alpha\delta(1 - w_i)[1 + (1 - w_i)/w.]$$

and off diagonal elements

$$v_{ij} = \alpha\delta(1 - w_i)(1 - w_j)/w..$$

3.5 Iterative EB Estimates

Recall the two-step iterative procedure from Section 2.6. For the one-way model, the first step is to find the values of α and δ that maximize $p(\boldsymbol{\theta},\alpha,\delta|\boldsymbol{y})$ with $\tilde{\boldsymbol{\theta}}$ replacing $\boldsymbol{\theta}$. This density was obtained in the previous section. Let

$$C = \lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \tilde{\theta}_i)^2$$

and

$$D = \lambda_2 + \Sigma_i(\tilde{\theta}_i - \tilde{\mu})^2 \text{ where } \tilde{\mu} = \Sigma_i\tilde{\theta}_i/k.$$

Differentiating the density with respect to α and δ produces the two equations

$$C + D/\delta - (N + k + q - 1)\alpha = 0$$

and

$$Dh(\delta)/\alpha - (k - 1)h(\delta)\delta + 2h'(\delta)\delta^2 = 0.$$

Solve the first equation for $\alpha = (C\delta + D)/(N + k + q - 1)\delta$. Insert this in the second equation to obtain

$$2Ch'(\delta)\delta^2 + [2Dh'(\delta) - (k - 1)Ch(\delta)]\delta + (N + q)Dh(\delta).$$

The solutions for the four priors are

Prior 1—$\delta = (N - 2)D/(k - 1)C$.

Prior 2—δ is the non-negative root of
$(k + 1)mC\delta^2 + [(k - 1)C - NmD]\delta - (N + 2)D$. There is exactly one non-negative root and it is slightly larger than $ND/(k + 1)C$.

Prior 3—This one must be solved numerically.

Prior 4—$\delta = (N + 2\nu_1 - 2)D/(k - 1)C$.

The second step is to obtain the revised estimate of $\boldsymbol{\theta}$. It is

$$\tilde{\boldsymbol{\theta}} = w_i\hat{\boldsymbol{\theta}} + (1 - w_i)\hat{\boldsymbol{\mu}} \text{ where } w_i = P_i\delta/(1 + P_i\delta).$$

A surprising observation is that for prior 4 the estimate does not depend on ν_2. The solution with prior 1 is the ratio of the appropriate sums of squares but, unlike the usual EB estimate, it can never be negative. It is, however, possible to get a value of zero.

3.6 *Density of* $\delta|y$

As indicated in Section 2.6, the iterative algorithm does not provide the other quantities of interest. Before developing the integration formulas, analytically integrate α out of the posterior density of $\alpha,\delta|\boldsymbol{y}$. For future use, let c be the constant of proportionality in the posterior density of $\alpha,\delta|\boldsymbol{y}$.

$$\begin{aligned} p(\delta|y) &= \int c(\alpha)^{-(N+q-1)/2}(\delta)^{-(k-1)/2}h(\delta)[\Pi_i w_i/w.]^{1/2} \\ &\quad \times \exp\{-[\lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2]/2\alpha \\ &\quad - [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/2\alpha\delta\}d\alpha \\ &= c(\delta)^{-(k-1)/2}h(\delta)[\Pi_i w_i/w.]^{1/2} \times \{\lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2 \\ &\quad + [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/\delta\}^{-(N+q-3)/2} \\ &\quad \times \Gamma[(N + q - 3)/2]2^{(N+q-3)/2}. \end{aligned}$$

3.7 *Evaluation by Integration*

Let $f(\delta)$ be the essential part of $p(\delta|y)$. That is,

$$\begin{aligned} f(\delta) &= (\delta)^{-(k-1)/2}h(\delta)[\Pi_i w_i/w.]^{1/2} \\ &\quad \times \{\lambda_1 + \Sigma_{ij}P_{ij}(y_{ij} - \hat{\theta}_i)^2 + [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/\delta\}^{-(N+q-3)/2}. \end{aligned}$$

Let the constant $g = \int f(\delta)d\delta$. So $p(\delta|y) = gf(\delta)$ and the relationship between the constants g and c is

$$c = \{g\Gamma[(N + q - 3)/2]2^{(N+q-3)/2}\}^{-1}.$$

The integral for g must be done numerically. An approach that is not necessarily the most efficient but is sure to work is to use separate numerical integrations on the intervals [0,1], [1,2], [2,4], [4,8], . . . until the contribution from the latest interval is sufficiently small. An iterative Gaussian integration converges fairly quickly. Any numerical analysis text (e.g., Burden, Faires, and Reynolds [8]) is likely to prove useful.

When doing numerical integration, it is important to know in advance if the integral will be finite. For the integral above it is sufficient to look at the balanced case. After removing some constants that depend only upon n and P, the integrand becomes

$$h(\delta)(1 + nP\delta)^{-(k-1)/2}[c_1 + \lambda_2/\delta + c_2/(1 + nP\delta)]^{-(N+q-3)/2}$$

with c_1 and c_2 being positive constants that depend only on the data. The four priors can be generalized to $h(\delta) = \delta^{-\nu_2}(1 + nP\delta)^{-h}$ where $\nu_2 = \nu_2$ for prior 4 and is zero otherwise and h is 1 for priors 2 and 3 and zero otherwise. It is necessary to verify the conditions under which the integral will exist as both $\delta \to 0$ and $\delta \to \infty$. For the first case, the essential part of the integrand is

$$\delta^{-\nu_2}[c_3 + \lambda_2/\delta]^{-(N+q-3)/2}.$$

For existence, either

$$\lambda_2 > 0 \text{ and } N + q - 2\nu_2 > 1 \text{ or}$$

$$\lambda_2 = 0 \text{ and } (\nu_2 > 1 \text{ or } \nu_2 = 0)$$

must hold. This condition is always satisfied for priors 1 through 3. For the second case, the tail behavior is governed by

$$\delta^{-(k-1+2h+2\nu_2)/2}$$

and so the integral will exist if $k - 1 + 2h + 2\nu_2 > 2$. This reduces to $k > 3$ for prior 1, $k > 1$ for priors 2 and 3, and $k > 3 - 2\nu_2$ for prior 4. Keep in mind that *both* conditions need to be satisfied. Rather than repeat these arguments for the integrals that follow, the existence results are summarized in a table at the end of this section.

Returning to the estimation problem, the first quantity to compute is

$$E(\delta|y) = \int \delta f(\delta) d\delta / g.$$

The next, and most useful quantity, is

$$E(\theta_i|y) = \int [w_i\hat{\theta}_i + (1 - w_i)\hat{\mu}] f(\delta) d\delta / g.$$

The next quantity of interest is $\text{Var}(\theta_i|y)$. From Section 2.7.2, two integrals are needed. The first one is similar to the one above. It is

$$E(\theta_i^2|y) = \int [w_i\hat{\theta}_i + (1 - w_i)\hat{\mu}]^2 f(\delta) d\delta / g.$$

The second one is $\int v_{ii} p(\alpha,\delta|y) d\alpha d\delta$. With regard to α (see Section 3.4), v_{ii} contributes a multiplicative constant of α and so the integral with respect to α is similar to the one done in Section 3.6.

$$\begin{aligned}\int v_{ii} p(\alpha,\delta|y) d\alpha d\delta &= \int (1 - w_i)[1 + (1 - w_i)/w.]\delta\alpha p(\alpha,\delta|y) d\alpha d\delta \\ &= \int (1 - w_i)[1 + (1 - w_i)/w.]\delta \\ &\quad \times c(\delta)^{-(k-1)/2} h(\delta)[\Pi_i w_i/w.]^{1/2} \times \{\lambda_1 + \Sigma_{ij} P_{ij}(y_{ij} - \hat{\theta}_i)^2 \\ &\quad + [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/\delta\}^{-(N+q-5)/2} \\ &\quad \times \Gamma[(N + q - 5)/2]2^{(N+q-5)/2} d\delta.\end{aligned}$$

Let
$$\begin{aligned}f^*(\delta) &= (\delta)^{-(k-1)/2} h(\delta)[\Pi_i w_i/w.]^{1/2} \\ &\quad \times \{\lambda_1 + \Sigma_{ij} P_{ij}(y_{ij} - \hat{\theta}_i)^2 \\ &\quad + [\lambda_2 + \Sigma_i(\hat{\theta}_i - \hat{\mu})^2 w_i]/\delta\}^{-(N+q-5)/2} \\ &\quad \div (N + q - 5)\end{aligned}$$

and so

$$\int v_{ii} p(\alpha,\delta|y) d\alpha d\delta = \int (1 - w_i)[1 + (1 - w_i)/w.]\delta f^*(\delta) d\delta / g.$$

In case there is interest in the two variance components, their posterior means are given by

$$E(\sigma^2|y) = \int f^*(\delta) d\delta / g$$

and

$$E(\tau^2|y) = \int \delta f^*(\delta) d\delta / g.$$

The predictive distribution for the next observation in the i^{th} class was discussed in Section 2.7.4. If the next X is $N(\theta_i, \sigma^2/R_i)$ then

$$E(X|y) = E(\theta_i|y)$$

$$\mathrm{Var}(X|y) = E(\sigma^2|y)/R_i + \mathrm{Var}(\theta_i|y).$$

These quantities have already been obtained.

The final item to obtain by integration is the posterior density of θ_i. From Section 2.7.3,

$$\begin{aligned}p(\theta_i|y) &= \int (s_i)^{-1}\delta^{-k/2} h(\delta)[\Pi_i w_i/w.]^{1/2} \\ &\quad \times \{[(\theta_i - m_i)/\delta s_i]^2 + \Sigma P_{ij}(y_{ij})^2 - \Sigma_{ij}(\hat{\theta}_i)^2 P_i w_i \\ &\quad - (\Sigma_i w_i \hat{\theta}_i)^2/w.\delta\}^{-(N+q-2)/2} d\delta \\ &\quad \times \Gamma[(N + q - 2)/2]/\{\Gamma[(N + q - 3)/2] g\sqrt{\pi}\}\end{aligned}$$

where

$$m_i = w_i\hat{\theta}_i + (1 - w_i)\hat{\mu}$$

and

$$(s_i)^2 = (1 - w_i)[1 + (1 - w_i)/w.].$$

TABLE 1

CRITERIA FOR THE EXISTENCE OF THE INTEGRALS

	$\int f(\delta)d\delta$ $\int \bar{\theta}_i f(\delta)d\delta$ $\int \bar{\theta}_i^2 f(\delta)d\delta$	$\int \delta f(\delta)d\delta$	$\int f^*(\delta)d\delta$ $\int v_{ii} f^*(\delta)d\delta$	$\int \delta f^*(\delta)d\delta$
Prior 1	$k > 3$	$k > 5$	$k > 3$	$k > 5$
Priors 2,3	$k > 1$	$k > 3$	$k > 1$	$k > 3$
Prior 4	$k > 3 - 2\nu_2$	$k > 5 - 2\nu_2$	$k > 3 - 2\nu_2$	$k > 5 - 2\nu_2$
and $\lambda_2 > 0$	$N + 2\nu_1 > 3$	$N + 2\nu_1 > 3$	$N + 2\nu_1 > 5$	$N + 2\nu_1 > 5$
or $\lambda_2 = 0$	$\nu_2 > 1$ or $\nu_2 = 0$	$\nu_2 > 2$ or $\nu_2 = 1$	$\nu_2 > 1$ or $\nu_2 = 0$	$\nu_2 > 2$ or $\nu_2 = 1$

3.8 An EB Procedure Based on Integration

To do all the calculations listed in the previous section requires a large number of approximate integrations. It would be helpful if those that are needed for each of the k groups individually could be avoided. A compromise that is reminiscent of EB methodology is presented here.

To proceed it is necessary to obtain a general result for the mean and variance of a function of a random variable. To do this begin with a general random variable X and a function $g(x)$. Use the Taylor series expansion about $\xi = \mathrm{E}(X)$ to write

$$\mathrm{E}[g(X)] \doteq \mathrm{E}[g(\xi) + (X - \xi)g'(\xi)] = g(\xi)$$

and

$$\mathrm{Var}[g(X)] \doteq \mathrm{Var}[g(\xi) + (X - \xi)g'(\xi)] = [g'(\xi)]^2\mathrm{Var}(X).$$

Generally, for these approximations to be reasonable, the random variable X should in some sense be the average of a fairly large number of observations and the function $g(x)$ should be thrice differentiable around ξ. Almost any advanced text on mathematical statistics will contain theorems that make the above results precise. The random variable under consideration here is $\delta|y$ and its density will converge in the same manner in which a sample mean converges as the sample size, N, increases.

To evaluate $\mathrm{E}(\theta_i|y) = \mathrm{E}(w_i\hat{\theta}_i + (1 - w_i)\hat{\mu}|y)$, let $\delta|y$ play the role of X and so $\xi = \mathrm{E}(\delta|y) = \tilde{\delta}$, a quantity obtained in Section 3.7. Then let $\tilde{w}_i = P_i\tilde{\delta}/(1 + P_i\tilde{\delta})$. Finally,

$$\mathrm{E}(\theta_i|y) \doteq \tilde{w}_i\hat{\theta}_i + (1 - \tilde{w}_i)\tilde{\mu}$$

the Japanese do. The real question is not big versus small business. The question is how do we have a prosperous big and small business section.

The final choice we've been given, again a false choice, is high-tech versus basic industries. The fact is, our basic industries, such as auto and steel, are not going to survive unless they get those advanced technologies that can automate their facilities quickly. The reality is that high-tech is going to be the salvation of big industry in this country. What we need is an economic environment where capital investment can proceed much faster than it is now. Once we put these myths aside, once we put these false choices aside, then the question logically is—how do we proceed? How do we go about creating the type of economy that we need? My argument is that what we really need is an economy that is flexible, an economy that's got vigor, one that's dynamic. In fact, you could almost say that there are two schools of economic thought on how to do it. One was expressed by Damon Runyon in his musical "Guys and Dolls." He had Bat Masterson, the protagonist, say—"The race may not go to the swift or the battle to the strong, but that's how to bet your money." That's opposed to the Mae West school. She said, "If something is worth doing well, it's worth doing very slowly."

I personally am of the Damon Runyon school. If you are going to have a swift and strong economy, how do you go about it? What's keeping us from doing what we should be doing? There's a series of obstacles, or choke points, or bottlenecks, that really keep this economy from moving along. The first and the most important today is the fact that we just haven't recognized the fact that we are in an interdependent global economy. Our policies are still trapped in the 1940s and 1950s when we dominated the global economy. When you go back and look, what you find is that the global trade system we now have was designed by us and the British between 1943 and 1947. We built it on three foundations: the International Monetary Fund, the World Bank, and something called the International Trade Organization. The International Trade Organization was to be a supernational organization whose basic purpose was to control unfair trade practices. It was to knock them down and not permit a re-occurrence of the Great Depression. The United States Senate refused to ratify that treaty. Then what we had to do was go into a very complex, very difficult set of negotiations with other nations and create something called the GATT (General Agreement on Tariffs and Trade). When it was first created it worked fairly well. It was based on the assumption that the rest of the world's economies were, in some degree, like those of the United States and England, that in effect it was an Anglo-American rule-driven economy. The government sets the rules;

demonstrate the evaluation of prediction intervals. In both cases, the results will be compared to those obtained from the usual EB formulas.

4.1 Workers' Compensation Frequency Data

The data were supplied by the National Council on Compensation Insurance (NCCI) and comprise observed frequencies from 7 years on 133 rating groups in 36 states. To make this data set somewhat manageable, the years were combined to yield the following:

y_{ij} = relative frequency in state j from group i

P_{ij} = Payroll in state j from group i.
$i = 1, \ldots, k = 133 \quad j = 1, \ldots, n_i$.

The number of states per group (n_i) is not always 36 as some states had no exposures for some groups. The total number of observations was $N = 4{,}572$, indicating that 216 cells had no exposure.

The objective is to estimate θ_i, the relative frequency of claims from insureds in rating group i. While the payrolls were adjusted for inflation, no attempt was made to adjust for any trend in the number of claims. As a final note, only claims resulting in permanent partial disability were included. It is important to recognize that the purpose of these illustrations is not to recommend a specific ratemaking procedure for workers' compensation insurance, but rather to illustrate the calculations using the formulas of Section 3. In particular, one might check the possibility that a cross-classified (state by group) model better describes the process.

For comparison, the formulas recommended by Bühlmann and Straub [7] were evaluated. These are the conventional EB formulas and produce the following results:

$$\hat{\sigma}^2 = \Sigma P_{ij}(y_{ij} - \hat{\theta}_i)^2/(N - k) = 4.746$$

$$\hat{\tau}^2 = [\Sigma P_i(\hat{\theta}_i - \hat{\mu})^2 - (k - 1)\hat{\sigma}^2]/(P - \Sigma P_i^2/P) = -.1123$$
where $\hat{\mu} = \Sigma P_i\hat{\theta}_i/P$ and $P = \Sigma P_i$.

When a negative value is obtained for τ^2, the convention is to use zero. This produces credibility weights of zero and so the grand mean is used as the estimate for each of the group means. As has been mentioned before, this method does not allow for evaluation of the quality of the estimates.

The key function in all the integrations is $f(\delta)$ as displayed in Section 3.7. The only item that involves the index j is $\Sigma(y_{ij} - \hat{\theta}_i)^2$ and it depends only on the data and so is constant. The number of observations per group will not affect the computation, other than the evaluation of a single sum of squares. All of the sums that involve δ have $k = 133$ terms. A second item is the size and shape of the integrand. The function $f(\delta)$ is well-behaved, being small at $\delta = 0$ and zero as $\delta \to \infty$, and having a single mode. However, because a variety of constants were removed from this density, it turned out that the value at the mode was very large. To see just how large, a variety of values of $\ln(f(\delta))$ were computed. In the actual calculations that followed, I worked with $f(\delta)/\exp(4{,}650)$ as the maximal value was near $\exp(4{,}650)$. Since all expressions are the ratio of two integrals involving this function, the adjustments cancel. Finally, it should be noted that $f(\delta)$ was calculated by first obtaining the logarithm of each of its constituent factors, adding them, subtracting 4,650, and then exponentiating the result. This avoided any overflow or underflow problems in the intermediate calculations.

In Table 2, the results for the first three priors are displayed. Note that while the posterior mean of τ^2 is indeed small (so zero was not an unreasonable estimate), it is large relative to σ^2/P_i, and so zero was not a reasonable choice for the credibility weight. This led to results that were considerably different from those obtained by the Bühlmann-Straub formula. Values of μ and the z_i were found by solving the following system of k equations:

$$E(\theta_i|y) \doteq z_i\hat{\theta}_i + (1 - z_i)\tilde{\mu} \text{ where } \tilde{\mu} = \Sigma z_i\hat{\theta}_i/\Sigma z_i.$$

The z_i then take on the role of credibility factors. These do not automatically arise in a Bayesian framework, and except for the fact that actuaries are accustomed to seeing this quantity, there is no reason to compute it. The three classes displayed were the ones with the smallest, median, and largest values of $\hat{\theta}_i$, respectively.

It is not surprising that the three priors produced virtually identical results. The large amount of data overwhelms all of these priors. In addition, I computed the standard deviation of $(\tau^2|y)$. Under prior 2 it is 0.000416. The mean of 0.003087 is over seven standard deviations above zero, indicating that the Bühlmann-Straub estimate is extremely unlikely to be valid.

The same items were evaluated using the approximations from Section 3.8. This was done only for prior 2, since the results will be similar for the others. The results are displayed in the last column of Table 2. In this case, the approximation performed well.

TABLE 2

ESTIMATES BY INTEGRATION IN THE ONE-WAY MODEL

	Prior1	Prior2	Prior3		EBStyle
$E(\delta \mid y)$	.0006613	.0006498	.0006503	$\tilde{\delta}$	.0006498
$E(\sigma^2 \mid y)$	4.755	4.753	4.753		
$E(\tau^2 \mid y)$	.003143	.003087	.003089		
Inferred μ	.07488	.07487	.07487	$\tilde{\mu}$	.07488
Class 107 (Auditors, Accountants, Draftsman)					
P_i	102,471	102,471	102,471		102,471
$\hat{\theta}_i$	.002762	.002762	.002762		.002762
z_i	.9852	.9849	.9850	$\tilde{w}_i$	.9852
$E(\theta_i \mid y)$	.003830	.003848	.003847		.003829
$SD(\theta_i \mid y)$	.006763	.006761	.006761		.006762
Class 68 (Explosives and Ammunition Mfg.)					
P_i	3,018	3,018	3,018		3,018
$\hat{\theta}_i$	.06262	.06262	.06262		.06262
z_i	.6638	.6599	.6600	$\tilde{w}_i$	.6623
$E(\theta_i \mid y)$	.06674	.06679	.06678		.06676
$SD(\theta_i \mid y)$	.03237	.03227	.03227		.03234
Class 89 (Stevedore)					
P_i	11,275	11,275	11,275		11,275
$\hat{\theta}_i$	.3895	.3895	.3895		.3895
z_i	.8800	.8782	.8783	$\tilde{w}_i$	.8799
$E(\theta_i \mid y)$	.3518	.3512	.3512		.3517
$SD(\theta_i \mid y)$	.01980	.01979	.01979		.01980
CPU (sec.)	14.92	15.23	15.11		12.49
Cost ($)	3.91	3.97	3.94		3.50

The computation was done on an IBM 4381 computer. The time did not include that used for setting up the data set (computing and arranging the values of y_{ij} and P_{ij}), so this can be viewed as the increase in cost of the Bayesian method over the Bühlmann-Straub formula (which is essentially free).

In addition, the iterative algorithm (Section 3.5) was employed with prior 2. Eight iterations were required for convergence. The results were $\tilde{\delta} = 0.0005936$, $\tilde{\sigma}^2 = 4.625$, and $\tilde{\mu} = 0.07479$. The results compare favorably with those obtained by integration.

4.2 Workers' Compensation Loss Ratio Data

This data set was taken from Meyers [25]. He provided loss ratios for three years of experience in 319 rating classes in the state of Michigan. In addition, the premium volume was given for each class/year; they will be used as the P_{ij} as in the Meyers paper. In that paper he used the Bühlmann-Straub formulas to obtain the credibility estimates. In view of the success from the previous section, I only computed estimates based on prior 2 and the EB approximation. The results were (the column labelled EB contains the results from the Meyers paper):

	EB	HNLM
σ^2	92,374	101,650
τ^2	0.019237	0.019762
μ	0.5822	0.5799
$K = \sigma^2/\tau^2$	4,801,900	5,143,710

It is not surprising that the results are similar. This also indicates that the Bühlmann-Straub formulas are indeed based on a hidden assumption of normality.

One of the most useful features of the Meyers data set is that it also provided the premiums and actual losses for the year following the three years of experience. This admits an evaluation of the predictive ability of the various procedures. I will begin the evaluation by duplicating the two tests performed by Meyers. In performing the tests, the expected losses based on the estimated loss ratios were adjusted to make the total expected losses equal to the actual losses. This is legitimate, since both credibility procedures were formulated to indicate relativities, not the absolute level of future losses. To do so would require trend factors to be incorporated into the analysis. An indication of how this might be done within the HNLM is given in the next section.

The first test is to measure the squared error of the predicted versus the observed losses. Bayes procedures (of any kind) should do well since the objective is to minimize squared error. The formula is $\Sigma P_i(A_i/E_i - 1)^2/k$ where A_i is the observed losses, E_i is the expected losses, and P_i is the premium. Also available for this test were the losses expected according to the rates promulgated by the NCCI. In addition, the weighted average relative error of the predictions was computed. The formula is $\Sigma P_i|A_i/E_i - 1|/\Sigma P_i$. The results were:

	Mean squared errors	Mean relative errors
NCCI	298,063	0.26776
EB	289,651	0.26396
HNLM	287,416	0.26368

The second test was invented by Meyers [25]. He called it the "Underwriting Test." The idea is to consider an insurer with established rates and a new entrant into the market. The new entrant uses his own method to determine premiums. He then offers insurance only to applicants in those rating classes for which his calculations produce rates less than those of the established insurer. He then charges a slightly lower premium than the established insurer and gets all of this business. If the new entrant's ratemaking methods are superior, he will expect a profit from his actions. Assuming differences only in relativities, but not in overall level, the established insurer will lose the same amount that the new entrant gains. A formalization of this process has A_i as the actual losses, E_i (for established) as the established insurer's expected losses, and N_i (for new) as the new entrant's expected losses. The profit and loss ratio, respectively, for the new entrant will be

$\Sigma(E_i - A_i)$ and $\Sigma A_i/\Sigma E_i$,

where all sums are taken over those classes for which $N_i < E_i$. The comparisons among the three estimators are presented in Table 3.

TABLE 3

THE UNDERWRITING TEST

Established	New Entrant	Profit for New Ent.	Loss Ratio
NCCI	EB	9,751,941	.950
NCCI	HNLM	9,929,786	.949
EB	NCCI	2,311,720	.990
EB	HNLM	7,493,090	.952
HNLM	NCCI	2,584,151	.989
HNLM	EB	−6,753,204	1.026

According to Meyers [25], a loss ratio of less than 0.957 has less than a five percent probability of occurring by chance. The results are consistent with the mean squared error ordering in that HNLM has a significant loss ratio as a new entrant against both EB and NCCI. Neither is significant as a new entrant against HNLM. By the same reasoning, EB is superior to NCCI. There is also an inconsistency present in that HNLM appears to do better versus NCCI than it does versus EB. An examination of the data reveals that the problem is the lack of sensitivity of this approach. Below are the results from two of the 319 classes (all figures in thousands of dollars):

		Expected		
Class	Loss	NCCI	EB	HNLM
8033	4,704	8,135	8,165	8,149
9079	22,208	15,464	16,087	16,108

In these classes HNLM scores a big "win" over EB, although the predictions are indistinguishable. Also note that in class 8033, HNLM defeats EB but loses to NCCI. It happened that there were no similar cases producing great gains under EB with a premium only slightly better than HNLM. The similarity of the expected losses should not have produced such a large overall difference between EB and HNLM. I attribute this to the method itself.

Recall that one of the stated advantages of HNLM is that it also produces prediction intervals for the future observations. This was done for the 319 classes. The standard deviations of the predictions were computed according to the approximation in Section 3.8. The formula is

$$\begin{aligned}\mathrm{Var}(\theta_i|y) \doteq\ & (\hat{\theta}_i - 0.5799)^2(P_i)^2(2.030 \times 10^{-15})/(1 + 1.957 \times 10^{-7}P_i)^4 \\ & + (0.01993 + 3.842 \times 10^{-9}P_i)/(1 + 1.944 \times 10^{-7}P_i)^2.\end{aligned}$$

If all is well, the standardized actual losses (actual minus predicted divided by the standard deviation) should follow a normal distribution with mean zero and variance one. To see if that is so, two plots were prepared. Figure 1 shows a histogram of the 319 standardized losses. It is apparent that there is more skewness present than one would expect from a normal distribution. The chi-square goodness-of-fit test statistic using 20 intervals is 71.55. With 17 degrees of freedom there is clearly a lack of fit. Figure 2 is a plot of the standardized errors against the expected losses. This can be used to check for serial correlations and constant variances. The former is not a problem and that is confirmed by performing a sign test. There are 153 sign changes out of 318 opportunities, clearly close to the expected number of 159. There does appear to be a problem

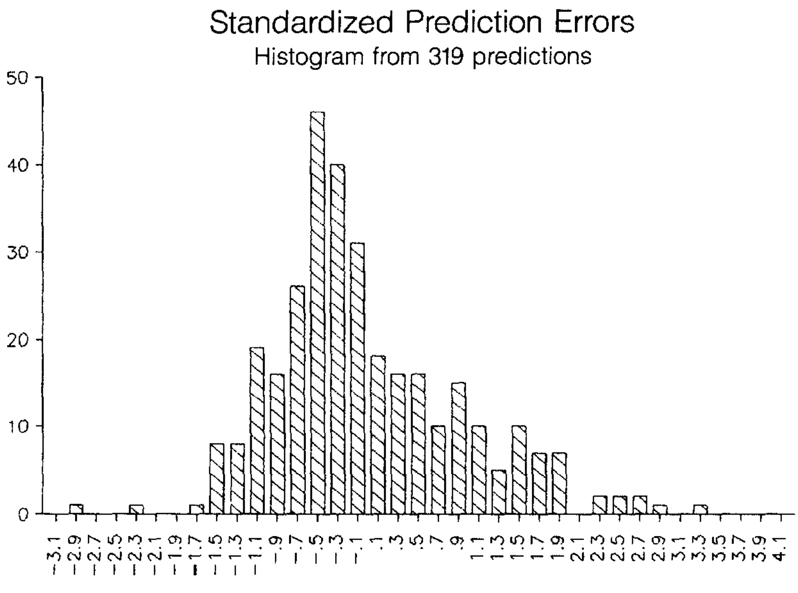


Figure 1

with the variances. For small predicted losses, the points are much too concentrated about the horizontal axis. This would lead us to suspect that we are overstating the variances in this range. One way to allow for this would be to adopt the unequal variance model of Section 5.1.

With these problems in mind, is there any value in attempting to predict these values? I believe there is. First of all, as was stated in Section 1, an inadequate model is almost always better than none at all. Secondly, we have some idea of the shortcomings, and could make some ad hoc corrections in the future.

As an illustration of the benefit of knowing the prediction errors, consider the following analysis which is done in the spirit of the "Underwriting Test." Identify all classes for which the HNLM predicted loss exceeds the NCCI predicted loss by at least k standard deviations. Do not offer insurance to these classes. For $k = 1$ there is only one class, number 4420. In thousands the

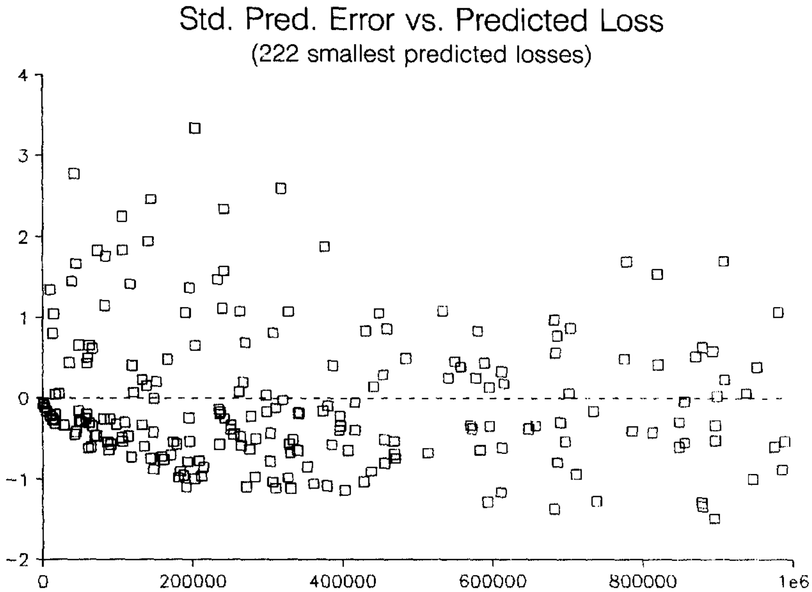


Figure 2

predictions were 4,329 and 3,427. The actual loss turned out to be 3,855. The point here is that this analysis can help identify classes in which the current rate levels are out of line, perhaps inspiring an investigation to see if something unusual has happened, either to the insureds in that class, or to the data in the process of recording it.

A final comment is in order. The above analysis leads us to believe that the normal model is not appropriate for these losses. In most settings we would not have the actual losses available in order to check this out. Can this be done with the original data? Box [4],[5] suggests the following approach to model-checking. In the general Bayesian setting, let X be the marginal distribution of the observations. Its density is computed from $f(x) = \int f(x|\theta)f(\theta)d\theta$ where, as usual, $f(x|\theta)$ is the model density and $f(\theta)$ is the prior density. If the model and prior are reasonable, the observed data x should, in some sense, be a "typical" observation from this density. While Box suggests a specific test, I will just display the standardized observations.

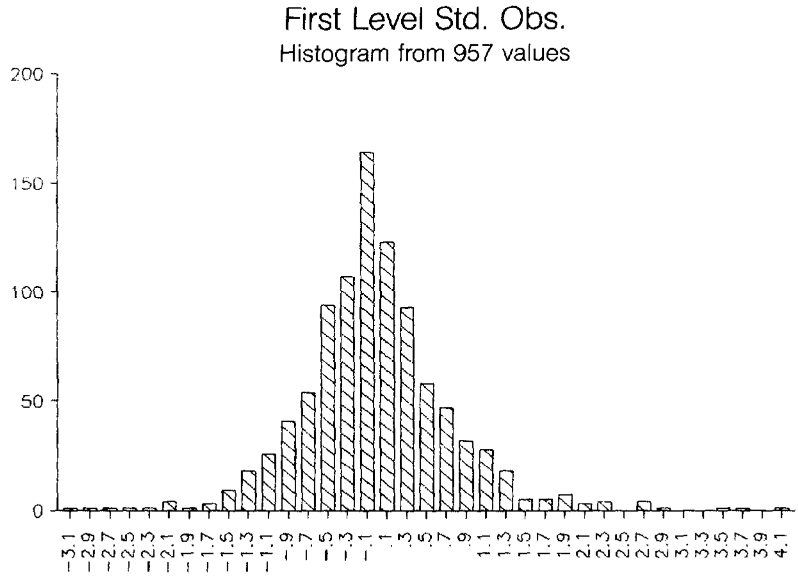


Figure 3

In particular, I will restrict attention only to the assumption of normality and will condition on the other aspects of the model such as constant variance. I will also condition on the estimated values of the variances. In the general HNLM, the distributions of interest are

First level—$y \sim N(A\hat{\theta},\hat{G})$,

Second level—$\hat{\theta} \sim N(B\hat{\mu},\hat{H})$, and

Overall—$y \sim N(\mathrm{AB}\hat{\mu},\hat{G} + A\hat{H}A')$.

In the particular case of the one-way model, these distributions become

First level—$y_{ij} \sim N(\hat{\theta}_i,\hat{\sigma}^2/P_{ij})$,

Second level—$\hat{\theta}_i \sim N(\hat{\mu},\hat{\tau}^2)$, and

Overall—$y_{ij} \sim N(\hat{\mu},\hat{\sigma}^2/P_{ij} + \hat{\tau}^2)$.

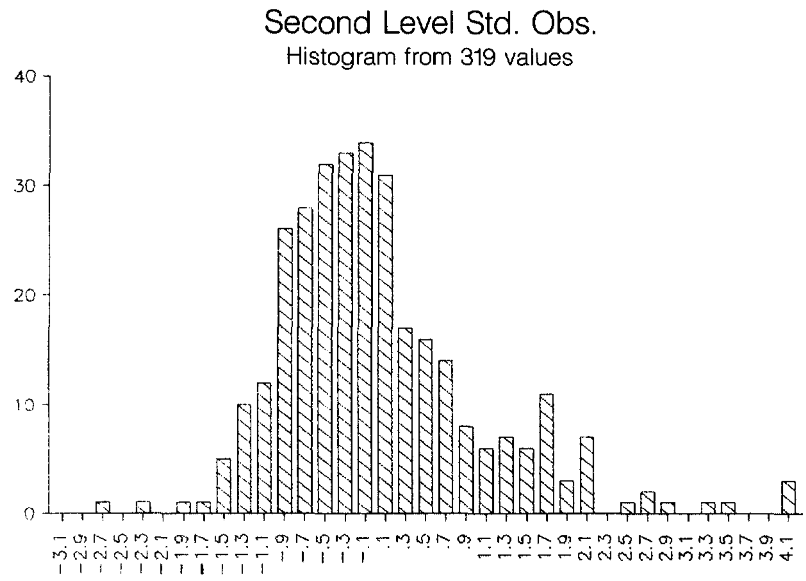


Figure 4

Figures 3, 4, and 5 display the histograms for the three sets of observations. In each case the appropriate values (y_{ij} or $\hat{\theta}_i$) were standardized according to the indicated means and variances. If the normal model was correct, the histograms should correspond to the standard normal distribution. An examination of the figures indicates that normality might indeed hold at the first level, but definitely does not at the second level. As a result, it is clear that the overall model should not be normal and that is indicated by the histogram.

Does the discovery of non-normality invalidate all the work that has been done? I believe the answer is no. We are at least as well off as one who used the EB methodology and we have the additional knowledge that we do not have the optimal solution. It is now a matter of deciding if the extra effort of analyzing a non-normal model is justified. Perhaps the ideas suggested in Section 1.3 are worth investigating.

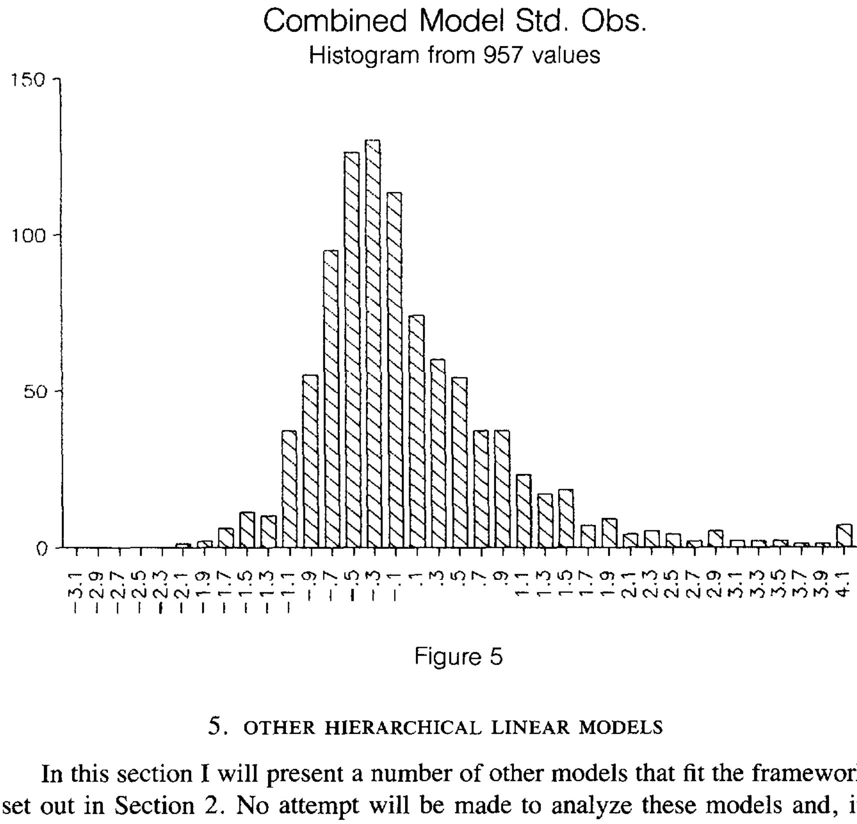


Figure 5

5. OTHER HIERARCHICAL LINEAR MODELS

In this section I will present a number of other models that fit the framework set out in Section 2. No attempt will be made to analyze these models and, in particular, no attempt will be made to assess the computational difficulties of evaluating these models. Unless there is an indication to the contrary it should be assumed that all of the random variables at a given level of the model are conditionally independent.

5.1 Unequal Variances

The process variance within each class may differ from class to class. From year to year within one class it is still assumed that variances are proportional to some exposure measure. The first two levels of the model are

Level 1—$Y_{ij}|\theta_i \sim N(\theta_i, \sigma_i^2/P_{ij})$ and

Level 2—$\theta_i|\mu \sim N(\mu, \tau^2)$.

Noninformative priors would then be placed on μ, τ^2, and $\sigma_1^2, \ldots, \sigma_k^2$.

It is easy to see what the EB approach to this model would be. Each σ_i^2 would be estimated from data in the i^{th} class. This is not in the spirit of credibility analysis where we would expect that information from the other classes can improve the estimation of a particular σ_i^2. A model that would do this would have an additional component at Level 2 such as

$$\sigma_i^2|\nu,\lambda \sim \text{Inverse gamma}(\nu,\lambda).$$

Noninformative priors would then be placed on μ, τ^2, ν, and λ.

5.2 Parameter Uncertainty

Suppose it is possible that the class mean θ_i varies from year to year, but not in any predictable manner. A model for this would be

Level 1—$Y_{ij}|\alpha_{ij} \sim N(\alpha_{ij},\sigma^2/P_{ij})$

Level 2—$\alpha_{ij}|\theta_i \sim N(\theta_i,\gamma^2)$, and

Level 3—$\theta_i|\mu \sim N(\mu,\tau^2)$.

Collapse the first two levels to produce

$$Y_{ij}|\theta_i \sim N(\theta_i,\sigma^2/P_{ij} + \gamma^2)$$

This is similar to a model proposed in Meyers [26]. It is not possible to derive EB estimates of the three variance terms as the within sum of squares is all that is available to estimate both σ^2 and γ^2. There is, however, a least squares approach based on the relationship of the variance in one group to its exposure that can yield estimates of the three parameters. A detailed HNLM analysis of this model is presented in Klugman [23]. The major problem is the evaluation of a two dimensional integral.

5.3 Hierarchical

In this paper, the word hierarchical applies to all the models. In credibility work this term has been reserved for the case where the k classes can be divided into g groups, where the i^{th} group would have m_i classes in it ($m_1 + \ldots + m_g = k$). Begin with a three level model:

Level 1—$Y_{ijt}|\theta_{ij} \sim N(\theta_{ij},\sigma^2/P_{ijt})$

Level 2—$\theta_{ij}|\beta_i \sim N(\beta_i,\gamma^2)$

Level 3—$\beta_i|\mu \sim N(\mu,\tau^2)$

Noninformative priors would be required for μ, τ^2, γ^2, and σ^2. Levels 2 and 3 may be combined to form a single distribution. However, when conditioned only on μ, the θ_{ij} are no longer independent. EB formulas for this model and the one in Section 5.4 are given in Venter [34].

5.4 Cross-Classified

Suppose each rating class is identified by two variables, such as sex and age, or state and occupation. An additive model, with the possibility of error, can be expressed with three levels:

Level 1—$Y_{ijt}|\theta_{ij} \sim N(\theta_{ij},\sigma^2/P_{ijt})$

Level 2—$\theta_{ij}|\mu,\alpha_i,\beta_j \sim N(\mu + \alpha_i + \beta_j,\gamma^2)$

Level 3—$\alpha_i \sim N(0,\tau_1^2)$
$\beta_j \sim N(0,\tau_2^2)$

Noninformative priors would be placed on μ, τ_1^2, τ_2^2, γ^2, and σ^2. The first two levels are easily collapsed to produce the single level

$$Y_{ijt}|\mu,\alpha_i,\beta_j \sim N(\mu + \alpha_i + \beta_j,\gamma^2 + \sigma^2/P_{ijt}).$$

It has been common to set $\gamma^2 = 0$ in analyzing this model. Including it allows for some departure from additivity. Letting τ_1^2 and τ_2^2 become infinite (so uniform priors are placed on all α_i and β_j) produces a simple version of the model. The credibility compromise is between a strict additive model and the use of individual class means.

5.5 Linear Trend

In the one-way model we might observe that there is a year to year trend in the means. A simple linear trend would be modeled as

Level 1—$Y_{ij}|\alpha_i,\beta_i \sim N(\alpha_i + j\beta_i,\sigma^2/P_{ij})$

Level 2—$\alpha_i|\mu_1 \sim N(\mu_1,\tau_1^2)$
$\beta_i|\mu_2 \sim N(\mu_2,\tau_2^2)$

with noninformative priors on μ_1, μ_2, τ_1^2, τ_2^2, and σ^2. This is similar to the well-known model introduced by Hachemeister [13]. It could be generalized to other types of trend by altering the structure of the mean at level 1.

5.6 Time Series

A linear time series model can be formulated in two stages. Here the subscripts indicate observations at a given time t.

Level 1—$\mathbf{Y}_t|\boldsymbol{\theta}_t \sim N(F_t\boldsymbol{\theta}_t, A)$

Level 2—$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1} \sim N(G_t\boldsymbol{\theta}_{t-1},B)$

The matrices F_t and G_t are known while A and B require prior distributions. The first level is the process distribution which explains how the observations relate to the underlying parameters. The second level is the state distribution which explains how the parameters change over time.

As an example, consider the linear trend model from Section 5.4. In this setting it would look like

Level 1—$Y_{ij}|\theta_{ij} \sim N(\theta_{ij},\sigma^2/P_{ij})$

Level 2—$\theta_{ij}|\theta_{i,j-1} \sim N(\beta_i + \theta_{i,j-1},\tau^2)$.

It is not exactly the same, as level 2 implies that there are some disturbances that let the progression of means depart from strict linearity. The parameter α_i in Section 5.5 is unchanging over time. Prior distributions would be needed for $\theta_{i,0}$ (to get the system started) and for β_i, τ^2, and σ^2.

This model is very similar to the Kalman filter. An excellent non-Bayesian application of this model to loss reserving is found in deJong and Zehnwirth [18]. A discussion of its relationship to the usual credibility models is given in deJong and Zehnwirth [19].

6. CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

The intent of this paper was to introduce the hierarchical normal linear model as a tool for classification ratemaking. This model has three advantages over the EB approach. First, methods for estimating the variances do not have to be created on a case-by-case basis. Instead, the estimates fall naturally out of the analysis. Second, estimates of estimation and prediction error are available. Finally, model-checking and model-selection procedures can be employed. The latter was not discussed in this paper, but methods do exist for identifying the most appropriate model when there are several to choose from (for example, a one-way vs. a cross-classified analysis). See Klugman [23] for an application.

Of course, this approach also introduces difficulties of its own. Foremost among them are the intensive computations needed to perform the analysis. In addition, the derivation of formulas for specific models can be very time consuming (although once obtained they can be used over and over). These problems are really another advantage of the HNLM approach; they are all technical in nature and are certain to be solved if there is sufficient interest in doing so.

The major area for future work (other than grinding out the solution to the many models of interest) is the relaxation of the normality assumption. There is overwhelming evidence that insurance data are not normal and so methods to accommodate that fact are most desirable. I envision two ways to attack this problem. One is to create methods that are robust against general departures from normality. To do this, the t-distribution or a mixture of normal distributions could be used in the model. Another way would be to find methods that are superior under specific distribution assumptions that are likely to correspond to insurance experience. In any case, considerable sensitivity testing should be done to any recommended formula.

REFERENCES

[1] R. Bailey, "A Generalized Theory of Credibility," *PCAS* XXXII, 1945, p. 13.

[2] J. Berger, *Statistical Decision Theory and Bayesian Analysis,* Second ed., New York, NY, Springer-Verlag, 1985.

[3] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis,* Reading, MA, Addison-Wesley, 1973.

[4] G. Box, "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society,* Series A, Vol. 143, 1980, p. 383.

[5] G. Box, "An Apology for Ecumenism in Statistics," *Proceedings of the Conference on Scientific Inference, Data Analysis, and Robustness,* G. Box, T. Leonard, and C-F. Wu, eds., New York, NY, Academic Press, 1983.

[6] H. Bühlmann, "Experience Rating and Credibility," *ASTIN Bulletin,* Vol. 4, 1967, p. 199.

[7] H. Bühlmann and E. Straub, "Credibility for Loss Ratios," *ARCH,* 1972.2.

[8] R. Burden, J. Faires, and A. Reynolds, *Numerical Analysis,* Third ed., Boston, MA, PWS, 1981.

[9] W. Ericson, "A Note on the Posterior Mean of a Population Mean," *Journal of the Royal Statistical Society,* Series B, Vol. 31, 1969, p. 332.

[10] S. Geisser, "On Prior Distributions for Binomial Trials (with discussion)," *The American Statistician,* Vol. 38, 1984, p. 244.

[11] P. Goel, "On Implications of Credible Means Being Exact Bayesian," *Scandinavian Actuarial Journal,* 1982, p. 41.

[12] F. Graybill, *Matrices with Applications in Statistics,* Second ed., Belmont, CA, Wadsworth, 1983.

[13] C. Hachemeister, "Credibility for Regression Models with Applications to Trend," *Credibility: Theory and Applications,* P. Kahn, ed., New York, NY, Academic Press, 1975.

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

[15] R. Hogg and S. Klugman, *Loss Distributions,* New York, NY, Wiley, 1984.

[16] Insurance Services Office, *Report of the Credibility Subcommittee: Development and Testing of Empirical Bayes Procedures for Classification Ratemaking,* New York, NY, ISO, 1980.

[17] W. Jewell, "Credible Means are Exact Bayesian for Exponential Families," *ASTIN Bulletin,* Vol. 8, 1974, p. 77.

[18] P. deJong and B. Zehnwirth, "Claims Reserving, State-Space Models and the Kalman Filter," *Journal of the Institute of Actuaries,* Vol. 110, 1983, p. 157.

[19] P. deJong and B. Zehnwirth, "Credibility Theory and the Kalman Filter," *Insurance: Mathematics and Economics,* Vol. 2, 1983, p. 281.

[20] S. Klugman, "Loss Distributions: Estimation, Large Sample Theory, and Applications," *Premium Calculation in Insurance,* F. de Vylder, M. Goovaerts, and J. Haezendonck, eds., Dordrecht, Holland, Reidel, 1984, p. 263.

[21] S. Klugman, "Distributional Aspects and Evaluation of Some Variance Estimators in Credibility Models," *ARCH,* 1985.1, p. 73.

[22] S. Klugman, "Bayesian Credibility with a Noninformative Prior," *Insurance and Risk Theory,* M. Goovaerts, F. de Vylder, and J. Haezendonck, eds., Dordrecht, Holland, Reidel, 1986, p. 195.

[23] S. Klugman, "Credibility Under Parameter Uncertainty: An Illustration of the Hierarchical Normal Linear Model," paper submitted for publication, 1987.

[24] D. Lindley and A. Smith, "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society,* Series B, Vol. 34, 1972, p. 1.

[25] G. Meyers, "Empirical Bayesian Credibility for Workers' Compensation Classification Ratemaking," *PCAS* LXXI, 1984, p. 96.

[26] G. Meyers, "An Analysis of Experience Rating," *ARCH,* 1985.2, p. 373.

[27] C. Morris, "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association,* Vol. 78, 1983, p. 47.

[28] C. Morris, "Parametric Empirical Bayes Confidence Intervals," *Proceedings of the Conference on Scientific Inference, Data Analysis, and Robustness,* G. Box, T. Leonard, and C-F. Wu, eds., New York, NY, Academic Press, 1983.

[29] J. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Computer Journal,* Vol. 6, 1965, p. 308.

[30] R. Norberg, "Empirical Bayes Credibility," *Scandinavian Actuarial Journal,* 1980, p. 177.

[31] M. Novick and P. Jackson, *Statistical Methods for Educational and Psychological Research,* New York, NY, McGraw-Hill, 1974.

[32] A. Smith, A. Skene, J. Shaw, J. Naylor and M. Dransfield, "The Implementation of the Bayesian Paradigm," *Communications in Statistics, Theory and Methods,* Vol. 14, 1985, p. 1079.

[33] G. Tiao and A. Zellner, "Bayesian Estimation of Multivariate Regression," *Journal of the Royal Statistical Society,* Series B, Vol. 26, 1964, p. 277.

[34] G. Venter, "Structured Credibility in Applications—Hierarchical, Multidimensional, and Multivariate Models," *ARCH,* 1985.2, p. 267.

[35] A. Whitney, "The Theory of Experience Rating," *PCAS* IV 1918, p. 274.

RESERVING LONG TERM MEDICAL CLAIMS

RICHARD H. SNADER

Abstract

In this paper, the use of life contingencies to establish reserves for claimants requiring lifetime medical care is explored. In evaluating such claims, consideration should be given to the effects of inflation, discounting for interest, life expectancy, the impact of the claimant's medical condition on life expectancy, and the accurate measurement of medical costs. The evaluation is made in three phases: a claim evaluation, a medical evaluation, and an actuarial evaluation.

The claim evaluation consists of gathering accurate information about the claimant's medical condition and the current cost of providing medical care. The medical evaluation consists of using the medical information obtained from the claim evaluation to estimate the effect on the claimant's life span. Information obtained from the claim and medical evaluations is combined with assumptions regarding interest, inflation, and mortality to produce the actuarial evaluation.

1. INTRODUCTION

Among the most difficult claims to adjust are those requiring medical care for the lifetime of the claimant. These claims usually arise from unlimited medical coverage provided under workers' compensation insurance or in connection with unlimited PIP benefits in certain no-fault auto states. In many cases, the claimants are seriously injured with little likelihood of recovery, and the cost of providing the required care is extremely high.

Although such claims occur infrequently, the potential financial impact of a mere handful of them very often can be catastrophic for an insurer. Consider, for example, the case of a permanently disabled person of age 20 requiring continuing medical care and treatment currently costing $50,000 annually. Despite serious injuries, a normal life span of 50 years might safely be assumed for this individual, suggesting a potential total cost of $2.5 million.

This amount is expressed in current dollars. What of the consequences of inflation? Inflation rates applicable to medical care expenditures are much greater than overall monetary inflation rates. Suppose costs are assumed to increase by 10% annually, a rate that reasonably approximates long term inflation rates applicable to medical care. The potential cost then becomes approximately $150 million. Of course, there are other elements to consider. Suppose the claimant dies early or lives to the age of 100. How can these and other contingencies be dealt with? How can an insurer set reasonable reserves on cases of this nature?

In this paper, an approach is suggested that combines claim adjusting expertise with actuarial processes and with principles employed in underwriting life insurance and annuity contracts. By its nature the suggested approach requires centralization in order to bring together these diverse elements. Claim adjusters in the field can hardly be expected to acquire all of the skills needed to deal with long term medical cases. Many will not see even one such claim in a lifetime of adjusting.

In evaluating long term medical claims, consideration must be given to the effects of inflation, discounting for interest, normal life expectancy, the claimant's medical condition (i.e., the probability of living a normal life span) and the accurate measurement of medical costs. The evaluation can then be made in three steps:

- a claim evaluation;
- a medical evaluation; and,
- an actuarial evaluation.

Taken together, these three steps constitute the financial evaluation of the claim.

The claim evaluation consists of gathering accurate and current data concerning (1) the amount and timing of medical expenditures and (2) the medical condition of the claimant based on the most current medical information available. The claim evaluation should be performed annually. The medical evaluation consists of using the medical information obtained from the claim evaluation to estimate the effect on the claimant's life span. Such an evaluation might be performed by a person with life insurance underwriting expertise or by some other person capable of relating medical information to life expectancies. Ideally, the result of the medical evaluation should be expressed as a multiplier applicable to standard mortality rates taken from an appropriate mortality table.

2. ACTUARIAL EVALUATION

Information obtained from the claim and medical evaluations together with appropriate assumptions regarding inflation and interest rates can be used to make the actuarial evaluation. Before getting too deeply involved in the details of the evaluation, a brief review of the applicable principles of life contingencies will be helpful. It will be assumed that the reader is familiar with standard actuarial notation as presented by Jordan [1] or by Bowers, et al [2].

3. REVIEW

An immediate life annuity of 1 payable to a life aged (x) with the first payment commencing at the end of one year is given by

$$a_x = \sum_{t=1}^{\infty} v^t {}_tp_x.$$

Where $v = (1 + i)^{-1}$ for effective interest rate i and ${}_tp_x$ may be considered the probability of making a payment at the end of year (t), or more simply the "probability of payment."

If the first payment is due at the beginning of the year, the series of payments is called an annuity due and is given by

$$\ddot{a}_x = \sum_{t=0}^{\infty} v^t {}_tp_x = 1 + a_x .$$

A temporary annuity payable for n years is given by

$$a_{x:\overline{n}|} = \sum_{t=1}^{n} v^t {}_tp_x.$$

An annuity deferred for a period of n years is given by

$${}_{n|}a_x = \sum_{t=n+1}^{\infty} v^t {}_tp_x.$$

4. GENERALIZED FORMULAE

Before annuity principles can be applied to reserving lifetime medical claims, the formulas must be generalized in order to gain sufficient flexibility to deal with more complicated situations.

Assume initially that only medical payments are being reserved, as might be the case in dealing with a Michigan no-fault claim.

Let $M_{x,t}$ be the medical payment due to (x) at the end of year (t), expressed in terms of current dollars.

Let i be the interest rate assumption expected to prevail over the lifetime of the claim and j be the inflation rate assumption pertaining to medical care costs.

If the reserve for future medical payments for a life aged (x) is denoted by R_x, an expression for the reserve is

$$R_x = \sum_{t=1}^{\infty} M_{x,t}(1 + i)^{-t}(1 + j)^t {}_tp_x.$$

The expression can be simplified somewhat by letting v_m denote the combined interest rate-inflation rate assumption and defining it by

$$v_m = (1 + i)^{-1} (1 + j).$$

The incurred cost of the claim is determined by adding the payments made to date to the reserve, R_x. If C_x is the incurred cost and P is the amount paid to date,

$$C_x = P + R_x.$$

When dealing with workers' compensation insurance, income replacement benefits as well as medical care costs are usually paid over the claimant's lifetime. In most cases, income replacement benefits are made at a fixed rate, but in some instances they are indexed to the CPI and therefore subject to inflation.

The previous formulation is easily extended to cover this situation by the introduction of a few additional terms.

Let $S_{x,t}$ be the indemnity payment due to (x) at the end of year (t).

Let i be the interest rate assumption, and k be the inflation rate assumption.

The combined interest rate-inflation rate assumption is given by $v_s = (1 + i)^{-1} (1 + k)$, the reserve is

$$R_x = \sum_{t=1}^{\infty} (S_{x,t}v_s^t + M_{x,t}v_m^t){}_tp_x,$$

and the incurred cost is still given by

$$C_x = P + R_x.$$

Since R_x is a discounted reserve, it will have an impact on reserve development observed in Schedule P. Even if mortality assumptions materialize exactly as predicted, adverse development will occur as a result of discount amortization over the lifetime of all claims reserved in this manner. This phenomenon is illustrated in Appendix A.

5. REINSURANCE

Ferguson [3] addressed the problem of calculating retained and ceded reserve components, but did not cover the situation involving lifetime medical care. The generalized formulas can be extended with relative ease to deal with reinsurance.

If Q is the primary insurer's reinsurance retention, the problem is solved by finding an integer, n, such that

$$Q^- = P + \sum_{t=1}^{n} [(1 + k)^t S_{x,t} + (1 + j)^t M_{x,t}] \leq Q, \text{ and}$$

$$Q^+ = P + \sum_{t=1}^{n+1} [(1 + k)^t S_{x,t} + (1 + j)^t M_{x,t}] > Q.$$

n can be thought of as the number of years required to exhaust the primary layer of coverage under the assumption that the claimant is still living.

The primary insurer's reserve is denoted by $R^*_{x:\overline{n}|}$, where

$$R_{x:\overline{n}|} \leq R^*_{x:\overline{n}|} < R_{x:\overline{n+1}|},$$

$$R_{x:\overline{n}|} = \sum_{t=1}^{n} (S_{x,t} \cdot v_s^t + M_{x,t} \cdot v_m^t)\ {}_tp_x, \text{ and}$$

$$R_{x:\overline{n+1}|} = \sum_{t=1}^{n+1} (S_{x,t} \cdot v_s^t + M_{x,t} \cdot v_m^t)\ {}_tp_x.$$

$R^*_{x:\overline{n}|}$ is approximated by linear interpolation:

$$R^*_{x:\overline{n}|} \doteq (R_{x:\overline{n+1}|} - R_{x:\overline{n}|}) \left(\frac{Q - Q^-}{Q^+ - Q^-}\right) + R_{x:\overline{n}|}.$$

And the reinsurer's reserve is given by

$${}_{n|}R_x^* \doteq R_x - R^*_{x:\overline{n}|}.$$

Some reinsurance agreements provide coverage only for so called catastrophe claims, where more than one worker is injured by a single event. Confusion occasionally results from the situation where the lifetime benefits of two or more claimants of different ages must be considered in establishing reinsurance reserves.

A general approach for resolving this problem can be illustrated by the following example.

- Let $A_{x,t}$ be the amount expected to be paid to (x) at the end of year (t).
- Let $B_{y,t}$ be the amount expected to be paid to (y) at the end of year (t).
- Let $C_{z,t}$ be the amount expected to be paid to (z) at the end of year (t).
- Assume $A_{x,t}$, $B_{y,t}$ and $C_{z,t}$ have already been adjusted to the expected inflation level of year (t).
- Assume $A'_{x,t}$, $B'_{y,t}$, and $C'_{z,t}$ have been discounted to the present day for interest.

$$R = \sum_{t=1}^{\infty} (A'_{x,t} \cdot {}_tp_x + B'_{y,t} \cdot {}_tp_y + C'_{z,t} \cdot {}_tp_z).$$

Now find an integer n such that

$$P + \sum_{t=1}^{n} (A_{x,t} + B_{y,t} + C_{z,t}) \leq Q, \text{ and}$$

$$P + \sum_{t=1}^{n+1} (A_{x,t} + B_{y,t} + C_{z,t}) > Q.$$

It follows that the primary insurer's share can be determined by linear interpolation between

$$\sum_{t=1}^{n} (A'_{x,t} \cdot {}_tp_x + B'_{y,t} \cdot {}_tp_y + C'_{z,t} \cdot {}_tp_z), \text{ and}$$

$$\sum_{t=1}^{n+1} (A'_{x,t} \cdot {}_tp_x + B'_{y,t} \cdot {}_tp_y + C'_{z,t} \cdot {}_tp_z).$$

The reinsurer's share is the total reserve less the primary insurer's share.

This method determines n based on the assumption that all three claimants live long enough for their combined payments to exceed the primary retention. It is possible that one or more of the claimants might die earlier than expected, in which case the time required to reach the retention could be much longer. The proposed method is conservative in that it provides the reinsurer with the earliest possible recognition of liability. A technically more exact method conceivably could be constructed based on the multiple life status $(x\ y\ z)$, but such a method would be quite complicated. The procedure outlined above is reasonable for real life situations.

The foregoing procedures can easily be extended to deal with several layers. It is possible, however, that C_x will exceed all layers of reinsurance, in which case the excess over the limit of the uppermost layer will revert to the primary insurer.

6. REVISED MORTALITY ASSUMPTIONS

Usually the very seriously injured claimants have extremely high annual costs associated with their medical care. Paraplegics, quadriplegics and brain stem injuries are examples of cases requiring expensive care. If these individuals could be expected to live normal life spans, the reserve values for their claims could become astronomical. Many times, however, such individuals are not expected to live as long as an unimpaired life. In such cases, a thorough medical evaluation will provide a basis for altering the mortality assumptions inherent in a standard table. Usually the results of such an evaluation are expressed in terms of the relationship to standard mortality. Normally revised values of q'_x are related to the q_x values of a standard mortality table so that

$$q'_x = f \cdot q_x,$$

subject to the restrictions that $f > 1$ and $f \cdot q_x \leq 1$. Then

$$p'_x = 1 - q'_x = 1 - f \cdot q_x, \text{ and}$$

$$_tp'_x = (1 - f \cdot q_x)(1 - f \cdot q_{x+1}) \ . \ . \ . \ (1 - f \cdot q_{x+t-1}).$$

The generalized reserve formulas can now be given as

$$R'_x = \sum_{t=1}^{\infty} (S_{x,t} \cdot v_s^t + M_{x,t} \cdot v_m^t) \ _tp'_x;$$

$$R'_{x:\overline{n}|} = \sum_{t=1}^{n} (S_{x,t} \cdot v_s^t + M_{x,t} \cdot v_m^t) \ _tp'_x; \text{ and,}$$

$$R'_{x:\overline{n+1}|} = \sum_{t=1}^{n+1} (S_{x,t} \cdot v_s^t + M_{x,t} \cdot v_m^t) \ _tp'_x.$$

$R^*_{x:\overline{n}|}$ is found by interpolation and $_{n|}R^*_x$ is found by subtraction.

An alternative might be to relate p'_x to p_x by

$$p'_x = (1 - f)p_x, \ 0 < f < 1.$$

In this case,

$$_tp'_x = (1 - f)^t \, _tp_x.$$

It should be noted that other, possibly more rigorous methods for adjusting mortality can be found in actuarial literature ([1], p. 57). There also are alternatives to adjusting mortality rates from a standard table. For example, one method is to estimate a fixed remaining life, e.g., five years, for the claimant.

Another method would be to construct a separate impaired life table that would reflect the average mortality rates of claimants with certain serious injuries.

7. FURTHER GENERALIZATION

Additional refinements can be made in the reserve formulas. For example, interest rates and inflation rates can be allowed to vary with time. Consider the expression for a medical expense reserve. Suppose i_r is the interest rate expected to be earned in year r, and j_r is the inflation rate expected to be applicable to payments made at the end of year r. Recall that $M_{x,t}$ is expressed in terms of current dollars. Then the present value of $M_{x,t}$ is given by

$$M_{x,t}(1 + j_1)(1 + j_2) \ldots (1 + j_r) \ldots (1 + j_t)(1 + i_1)^{-1}(1 + i_2)^{-1} \ldots \ldots (1 + i_r)^{-1} \ldots (1 + i_t)^{-1},$$

and the present value of expected future payments is

$$\sum_{t=1}^{\infty} \prod_{r=1}^{t} [(1 + i_r)^{-1}(1 + j_r)]M_{x,t} \cdot {}_tp_x.$$

The formulas can be generalized even further by assuming that payments occur at the mid-point of the time intervals instead of at the end-points. This last refinement is achieved by adjusting exponents applicable to inflation and interest and by adjusting mortality factors. For example:

$$R_x = \sum_{t=1}^{\infty} M_{x,t}v_m{}^{t-1/2} \cdot {}_{t-1/2}f_x,$$

where

$${}_{t-1/2}p_x \doteqdot \frac{1}{2}\left({}_{t-1}p_x + {}_tp_x\right).$$

8. THE MODEL IN OPERATION

The general approach should now be obvious. Claims must be reviewed annually to obtain current medical data and cost information. This is the responsibility of adjusters in the field who must provide accurate estimates of the amount and timing of payments expected to be made after the reserve evaluation date.

Timing of payments is the essence of the discount calculation. It should not automatically be assumed that payments will be made in equal amounts each year. For example, it might be that very large payments will be required over the first few years of a claimant's injury when hospital care is needed. Afterwards, however, lower payments may be required if the claimant can be cared for in a less expensive facility or at home.

In addition to providing estimated payments, field adjusters must acquire sufficient medical information to determine if the claimant's life is impaired and, if so, the extent of impairment. Since the claimant's medical condition is subject to change, it must be evaluated frequently. A change in medical condition results in a change in life expectancy and usually is accompanied by a change in expected costs.

Once payment data and medical information have been assembled, it is a fairly easy task to develop a reserving model.

- First, display the expected payment stream by year of expected payment.
- Adjust the individual payments to reflect the level of inflation expected to apply to each year of payment.
- Next, discount the individual payments for interest to the present time.
- Multiply each discounted payment by the probability of payment (${}_tp_x$) to obtain the discounted values of expected payments.
- Sum the discounted expected payments to obtain the reserve.

If the claimant is so badly impaired that a shortened life span is anticipated, it will be necessary to estimate the increase in mortality rates resulting from the impairment. Such an estimate can be made only by a person trained in evaluating medical information and translating such information into revised mortality rates.

Life insurance underwriters are often called upon to make such assessments. Their judgments are usually expressed as multipliers applicable to the q_x values in some standard mortality table. Multipliers might range from 1.2 to 10 or even higher. Once the multiplier has been determined, the payment probabilities can be adjusted and the operational steps described above can be taken.

The following examples are given as illustrations of the model in operation.

Example 1

Assume unimpaired individuals aged (x) are expected to experience the following mortality.

t	Number Living Aged $(x + t)$	Expected Deaths
0	1,000	307
1	693	218
2	475	153
3	322	106
4	216	72
5	144	49
6	95	33
7	62	22
8	40	15
9	25	10
10	15	15
		1,000

Assume a claim is being evaluated for which both medical and indemnity payments are made and that the total amount paid to date is \$230,000, consisting of \$30,000 for indemnity and \$200,000 for medical expense. Assume further that indemnity payments have been awarded at an annual rate of \$15,000 and medical expenses are expected to be \$100,000 each year for the claimant's lifetime. Indemnity payments are discounted at 3.5%. Medical payments are not discounted and are not expected to be increased by inflation. Claim payments in excess of \$1 million are reinsured.

The schedule of payments is shown in the following table.

Year (t)	Indemnity Benefits	Medical Payments	Total Annual Payments	Cumulative Payments
Paid to Date	$30,000	$200,000		$ 230,000
1	15,000	100,000	$115,000	345,000
2	15,000	100,000	115,000	460,000
3	15,000	100,000	115,000	575,000
4	15,000	100,000	115,000	690,000
5	15,000	100,000	115,000	805,000
6	15,000	100,000	115,000	920,000
7	10,435	69,565	80,000	1,000,000
7	4,565	30,435	35,000	1,035,000
8	15,000	100,000	115,000	1,150,000
9	15,000	100,000	115,000	1,265,000
10	15,000	100,000	115,000	1,380,000

Inflation Rate, Indemnity = 0.0%
Inflation Rate, Medical = 0.0%
Interest Rate, Indemnity = 3.5%
Interest Rate, Medical = 0.0%

Details of the reserve calculations are shown in the following table.

Year (t)	Indemnity Discount Factor	Medical Discount Factor	Probability of Payment
1	.9662	1.000	.693
2	.9335	1.000	.475
3	.9019	1.000	.322
4	.8714	1.000	.216
5	.8420	1.000	.144
6	.8135	1.000	.095
7	.7860	1.000	.062
7	.7860	1.000	.062
8	.7594	1.000	.040
9	.7337	1.000	.025
10	.7089	1.000	.015

Year (t)	Discounted Indemnity	Discounted Medical	Discounted Total	Discounted Cumulative
Paid to Date	$30,000	$200,000	$230,000	$230,000
1	10,044	69,300	79,344	309,344
2	6,651	47,500	54,151	363,495
3	4,356	32,200	36,556	400,051
4	2,823	21,600	24,423	424,474
5	1,819	14,400	16,219	440,693
6	1,159	9,500	10,659	451,352
7	509	4,313	4,822	456,174
7	222	1,887	2,109	458,283
8	456	4,000	4,456	462,739
9	275	2,500	2,775	465,514
10	160	1,500	1,660	467,174
	$58,474	$408,700	$467,174	

From these tables it can be seen that:

- the incurred cost of the claim, with future payments discounted for both interest and mortality, is $467,174,

- the present value of future payments, which is the reserve for the claim, is \$237,174,
- claim payments will exceed the retention between the sixth and seventh year,
- the primary insurer's share of the reserve is approximated by \$226,174, and
- the reinsurer's share of the reserve is approximated by \$11,000.

Example 2

In this example, it is assumed that medical payments are subject to an annual inflation rate of 10%, and these payments can safely be discounted at 8%. For convenience, indemnity payments are discounted at a statutory rate of 3.5%.

Inflation Rate, Indemnity	=	0.0%
Inflation Rate, Medical	=	10.0%
Interest Rate, Indemnity	=	3.5%
Interest Rate, Medical	=	8.0%

Year (t)	Indemnity Benefits	Medical Payments	Total Annual Payments	Cumulative Payments
Paid to Date	\$30,000	\$200,000		\$230,000
1	15,000	110,000	\$125,000	355,000
2	15,000	121,000	136,000	491,000
3	15,000	133,100	148,100	639,100
4	15,000	146,410	161,410	800,510
5	15,000	161,051	176,051	976,561
6	7,199	21,609	23,439	1,000,000
6	7,801	155,547	168,717	1,168,717
7	15,000	194,872	209,872	1,378,589
8	15,000	214,359	229,359	1,607,948
9	15,000	235,795	250,795	1,858,742
10	15,000	259,374	274,374	2,133,117

Year (t)	Indemnity Discount Factor	Medical Discount Factor	Probability of Payment
1	.9662	.9259	.693
2	.9335	.8573	.475
3	.9019	.7938	.322
4	.8714	.7350	.216
5	.8420	.6806	.144
6	.8135	.6302	.095
6	.8135	.6302	.095
7	.7860	.5835	.062
8	.7594	.5403	.040
9	.7337	.5002	.025
10	.7089	.4632	.015

Year (t)	Discounted Indemnity	Discounted Medical	Discounted Total	Discounted Cumulative
Paid to Date	$30,000	$200,000	$230,000	$230,000
1	10,044	70,583	80,627	310,627
2	6,651	49,276	55,927	366,554
3	4,356	34,022	38,379	404,933
4	2,823	23,245	26,068	431,001
5	1,819	15,784	17,602	448,603
6	556	1,294	1,435	450,038
6	603	9,312	10,330	460,368
7	731	7,050	7,781	468,149
8	456	4,632	5,088	473,237
9	275	2,949	3,224	476,461
10	160	1,802	1,962	478,423
	$58,474	$419,949	$478,423	

Incurred Cost of the Claim	$478,423
Present Value of Future Payments	248,423
Primary Insurer's Share	220,038
Reinsurer's Share	28,385

Making no assumptions regarding inflation or interest and basing all estimates of future expenditures on current costs, as was done in the first example with respect to medical payments, is often thought to be equivalent to assuming interest earnings will be exactly offset by inflation. It might, therefore, be said that payments have been discounted at an implied interest rate equal to the rate of future inflation.

This oversimplification may be sufficient for estimating the gross reserve but does not work very well when reinsurance is involved. If future costs are projected in current dollars, the time required to reach the reinsurance retention will be overestimated and the reserve will not be accurately divided into reinsurance layers. The proper sequence of calculations is to adjust future payments so that they reflect the levels of inflation expected to apply at the time the payments are made, and then to estimate the retention period, n.

Example 3

This example is the same as the first, except that it is assumed the claimant is subject to a mortality rate 50% greater than normal. Using the relationship

$$q'_x = 1.5\, q_x,$$

the following table can be developed.

t	${}_tp_x$	q_{x+t-1}	q'_{x+t-1}	${}_tp'_x$
1	.693	.3070	.4605	.5395
2	.475	.3146	.4719	.2849
3	.322	.3221	.4832	.1472
4	.216	.3292	.4938	.0745
5	.144	.3333	.5000	.0373
6	.095	.3403	.5105	.0182
7	.062	.3474	.5211	.0087
8	.040	.3548	.5322	.0041
9	.025	.3750	.5625	.0018
10	.015	.4000	.6000	.0007

Reserve values are now calculated as follows:

Year (t)	Indemnity Discount Factor	Medical Discount Factor	Probability of Payment
1	.9662	.9259	.5395
2	.9335	.8573	.2849
3	.9019	.7938	.1472
4	.8714	.7350	.0745
5	.8420	.6806	.0373
6	.8135	.6302	.0182
6	.8135	.6302	.0182
7	.7860	.5835	.0087
8	.7594	.5403	.0041
9	.7337	.5002	.0018
10	.7089	.4632	.0007

Year (t)	Discounted Indemnity	Discounted Medical	Discounted Total	Discounted Cumulative
Paid to Date	$30,000	$200,000	$230,000	$230,000
1	7,819	49,952	57,771	287,771
2	3,989	26,867	30,856	318,627
3	1,991	14,139	16,130	334,757
4	974	7,288	8,262	343,019
5	471	3,717	4,188	347,207
6	107	887	994	348,201
6	115	961	1,076	349,277
7	103	899	1,002	350,279
8	47	432	479	350,758
9	20	193	213	350,971
10	7	76	83	351,054
	$45,643	$305,411	$351,054	

Incurred Cost of the Claim	$351,054
Present Value of Future Payments	121,054
Primary Insurer's Share	118,201
Reinsurer's Share	2,853

Example 4

In this example, (x) and (y) are injured in a common occurrence. Expected medical costs are \$50,000 per year to (x) and \$100,000 per year to (y). Reinsurance is excess over \$1 million. Payments are not discounted and are not expected to be increased by inflation. Expected mortality is shown in the following table.

t	Number Living Aged $(x + t)$	Expected Deaths	Number Living Aged $(y + t)$	Expected Deaths
0	1,000	307	1,000	257
1	693	218	743	201
2	475	153	542	147
3	322	106	395	108
4	216	72	287	75
5	144	49	208	58
6	95	33	150	42
7	62	22	108	31
8	40	15	77	23
9	25	10	54	17
10	15	15	37	12
11			25	9
12			16	6
13			10	4
14			6	3
15			3	3

Details of the reserve calculations are shown in the following table.

Year (t)	Payments to (x)	Payments to (y)	Total Payments	Cumulative Payments
1	$50,000	$100,000	$ 150,000	$ 150,000
2	50,000	100,000	150,000	300,000
3	50,000	100,000	150,000	450,000
4	50,000	100,000	150,000	600,000
5	50,000	100,000	150,000	750,000
6	50,000	100,000	150,000	900,000
7	33,333	66,667	100,000	1,000,000
7	16,667	33,333	50,000	1,050,000
8	50,000	100,000	150,000	1,200,000
9	50,000	100,000	150,000	1,350,000
10	50,000	100,000	150,000	1,500,000
11		100,000	100,000	1,600,000
12		100,000	100,000	1,700,000
13		100,000	100,000	1,800,000
14		100,000	100,000	1,900,000
15		100,000	100,000	2,000,000
			$2,000,000	

Year (t)	${}_tP_x$	${}_tP_y$	Expected Payments to (x)	Expected Payments to (y)	Expected Total	Cumulative Total
1	.693	.743	$ 34,650	$ 74,300	$108,950	$108,950
2	.475	.542	23,750	54,200	77,950	186,900
3	.322	.395	16,100	39,500	55,600	242,500
4	.216	.287	10,800	28,700	39,500	282,000
5	.144	.208	7,200	20,800	28,000	310,000
6	.095	.150	4,750	15,000	19,750	329,750
7	.062	.108	2,067	7,200	9,267	339,017
7	.062	.108	1,033	3,600	4,633	343,650
8	.040	.077	2,000	7,700	9,700	353,350
9	.025	.054	1,250	5,400	6,650	360,000
10	.015	.037	750	3,700	4,450	364,450
11		.025		2,500	2,500	366,950
12		.016		1,600	1,600	368,550
13		.010		1,000	1,000	369,550
14		.060		600	600	370,150
15		.030		300	300	370,450
			$104,350	$266,100	$370,450	

Present Value of Future Payments = $370,450
Primary Insurer's Share = 339,017
Reinsurer's Share = 31,433

9. SENSITIVITY

An additional example based on more realistic mortality assumptions is given in Appendix B. The example also employs plausible inflation and interest rate assumptions. It is constructed from the following hypothetical set of circumstances.

- The claimant is 40 years old.
- Annual medical payments, currently estimated at $50,000, are expected to be required for the lifetime of the claimant. Payments are estimated in current dollars.
- Payments are uniformly distributed throughout each year.
- Mortality follows the 1969–71 U.S. Life Table for the Total Population.
- The primary insurer's retention is $1,000,000. The limit for the first layer of reinsurance is $5,000,000 over the $1,000,000 retention. The limit for the second layer is $5,000,000 excess over $6,000,000.

The problem is to find the reserve required for the primary insurer and the reinsurers. Several inflation rate, interest rate, and mortality scenarios have been constructed. Three of these scenarios are displayed in Appendix B. The first exhibit of the appendix portrays the situation where future payments are not inflated and not discounted. The second exhibit shows the situation where inflation and interest are both assumed to be 6%. The third exhibit portrays "realistic" interest and inflation assumptions.

The "realistic" interest rate assumption is similar to an assumption that a life insurer might use in calculating GAAP reserves for annuities. Specifically, the following assumption is employed.

First 10 years	8%
Next 10 years	7%
Next 10 years	6%
Remaining years	5%

In constructing a "realistic" inflation rate assumption, we assume the medical care component of the CPI is a good indicator. Over a relatively long period the medical inflation rate, as measured by the CPI, ranged between 9% and 10%, but more recently has declined to approximately 7.5%. It is reasonable to anticipate a gradual return to the long term level in the near future. In the very long term, it seems reasonable to assume that medical inflation and interest rates will follow each other fairly closely, with medical inflation (as opposed to general economic inflation) exceeding interest rates by a slight margin. From these considerations, the following "realistic" inflation scenario is constructed.

Year 1	7.5%
Year 2	8.0%
Year 3	8.5%
Years 4 to 10	9.0%
Next 10 years	8.0%
Next 10 years	7.0%
Remaining years	6.0%

In the following tables, we examine the effect that variations in the assumptions have on the reserve calculations. Table 1 illustrates the effect of several different interest/inflation combinations. Mortality is assumed to follow the 1969–71 population table and is referred to as "standard mortality." Reserves are in thousands.

TABLE 1—STANDARD MORTALITY

Insurance Layer	Interest/Inflation Assumption				
	0%, 0%	6%, 6%	8%, 8%	10%, 10%	Realistic
1	$ 941	$ 641	$ 588	$ 545	$ 591
2	760	842	738	651	868
3	0	218	375	505	540
Total	$1,701	$1,701	$1,701	$1,701	$1,999

In the first four columns, interest and inflation are separately but equally quantified. The first column illustrates the notion that using no specific inflation or interest assumption is equivalent to discounting at an implied interest rate equal to inflation.

The total reserves shown for the first four columns are equal, as expected. The fifth column shows the results of realistic inflation and interest assumptions. The total reserve is higher in this case because inflation exceeds interest over most of the payout period.

Except for the 0%, 0% assumption, the primary insurer's reserve is relatively insensitive to changes in interest and inflation. This result occurs when the retention is low compared with possible total payments. The inflation and interest rate assumptions selected are more important to reinsurers.

Table 2 shows the effect of variations in interest rates when mortality and inflation rates are held constant.

TABLE 2—STANDARD MORTALITY, REALISTIC INFLATION

Insurance Layer	Interest Assumption			
	6%	8%	10%	Realistic
1	$ 661	$ 590	$ 530	$ 591
2	1,130	770	534	868
3	662	328	166	540
Total	$2,453	$1,688	$1,230	$1,999

As expected, the total reserve decreases as the interest rate assumption increases. As previously observed, the primary insurer's reserve is low relative to possible total payments.

In Table 3, the effect of changes in inflation rates for constant mortality and interest rates is illustrated.

TABLE 3—STANDARD MORTALITY, REALISTIC INTEREST

Insurance Layer	Inflation Assumption			
	6%	8%	10%	Realistic
1	$ 571	$ 589	$ 605	$ 591
2	646	837	1,010	868
3	185	634	1,593	540
	$1,402	$2,060	$3,208	$1,999

As expected, the total reserve increases with inflation while the primary reserve is not materially affected. The choice of the inflation assumption is obviously a prime concern for the insurer of the third layer.

The final table illustrates the effect of variations in mortality assumptions. The results require no comment.

TABLE 4—REALISTIC INFLATION, REALISTIC INTEREST

Insurance Layer	Mortality Assumption: Standard	2.5 × Standard	5 × Standard	10 × Standard
1	$ 591	$ 566	$528	$460
2	868	649	414	189
3	540	149	27	1
	$1,999	$1,364	$969	$650

REFERENCES

[1] C. W. Jordan, *Life Contingencies*, Chicago, Society of Actuaries, 1967.

[2] N. L. Bowers, Jr., H. U. Gerber, J. C. Hickman, D. A. Jones, C. J. Nesbitt, *Actuarial Mathematics*, Chicago, Society of Actuaries, 1986.

[3] R. E. Ferguson, "Actuarial Note on Workmen's Compensation Loss Reserves," *PCAS* LVIII, 1971.

APPENDIX A

The amortization of discount and its impact on reserve development can be illustrated by assuming a block of claims exists for a large number of annuitants, each one aged (x). Suppose annual payments of one dollar are made to all surviving members of the group with the first payment being made at the end of one year. Payments to survivors are unaffected by inflation.

Denote the number of claimants by l_x, the number of survivors at the end of one year by l_{x+1}, at the end of two years by l_{x+2}, and so on. Denote the undiscounted or "full value" reserve by

$$FV_x = l_{x+1} + l_{x+2} + l_{x+3} + \ldots$$
$$= \sum_{t=1}^{\infty} l_{x+t} \ .$$

The reserve for any one individual is

$$\frac{1}{l_x} \sum_{t=1}^{\infty} l_{x+t} = \sum_{t=1}^{\infty} {}_tp_x = e_x \ ,$$

where e_x is known as the *expectation of life*.

The expected value of payments made at the end of the first year is l_{x+1}, and the reserve for the surviving claimants is given by

$$FV_{x+1} = \sum_{t=1}^{\infty} l_{x+t+1} \ .$$

Development on the initial reserve is

$$l_{x+1} + FV_{x+1} - FV_x$$
$$= l_{x+1} + \sum_{t=1}^{\infty} l_{x+t+1} - \sum_{t=1}^{\infty} l_{x+t}$$
$$= l_{x+1} + (l_{x+2} + l_{x+3} + \ldots) - (l_{x+1} + l_{x+2} + l_{x+3} + \ldots)$$
$$= 0.$$

Denote the discounted or "present value" reserve by

$$PV_x = vl_{x+1} + v^2 l_{x+2} + v^3 l_{x+3} + \ldots$$
$$= \sum_{t=1}^{\infty} v^t l_{x+t} \ .$$

For any one individual, the reserve is

$$\frac{1}{l_x} \sum_{t=1}^{\infty} v^t l_{x+t} = \sum_{t=1}^{\infty} v^t {}_tp_x = a_x \; .$$

The expected value of payments made at the end of the first year is still l_{x+1}. The reserve for the survivors is

$$PV_{x+1} = \sum_{t=1}^{\infty} v^t l_{x+t+1} \; .$$

In this case, development on the initial reserve is

$$\begin{aligned} & l_{x+1} + PV_{x+1} - PV_x \\ & \quad = l_{x+1} + \sum_{t=1}^{\infty} v^t l_{x+t+1} - \sum_{t+1}^{\infty} v^t l_{x+t} \\ & \quad = l_{x+1} + (vl_{x+2} + v^2 l_{x+3} + \ldots) - (vl_{x+1} + v^2 l_{x+2} + v^3 l_{x+3} + \ldots) \\ & \quad = (1 - v)(l_{x+1} + vl_{x+2} + v^2 l_{x+3} + \ldots) \\ & \quad = \left(\frac{1 - v}{v}\right) (vl_{x+1} + v^2 l_{x+2} + v^3 l_{x+3} + \ldots) \\ & \quad = i \sum_{t=1}^{\infty} v^t l_{x+t} = i \, PV_x, \text{ since } \frac{1 - v}{v} = i \; . \end{aligned}$$

It has been shown that, for this set of circumstances, observed development equals the interest earned on the original reserve.

This observation can be extended intuitively to any situation where reserves are discounted. If reserves have been estimated with total accuracy and payments precisely follow the assumed payment pattern, observed development for undiscounted reserves will be zero and for discounted reserves will equal the interest earned on the average value of assets required to secure the reserves during the payment period.

APPENDIX B

SCENARIO 1

Age of Claimant = 40
Interest on Medical Payments = 0 %
Inflation Rate for Medical Payments = 0 %
Mortality Rate = 1 Times Standard Mortality
Retention Limits = 1,000,000

Year	Medical Payments Current Dollars	Medical Payments Inflated Dollars	Disc. Factor	Mort. Factor	Reserve Amount	Cumulative Reserve Amount
1	50,000	50,000	1.0000	.9969	49,843	49,843
2	50,000	50,000	1.0000	.9935	49,673	99,516
3	50,000	50,000	1.0000	.9898	49,489	149,005
4	50,000	50,000	1.0000	.9858	49,289	198,295
5	50,000	50,000	1.0000	.9814	49,071	247,366
6	50,000	50,000	1.0000	.9767	48,833	296,199
7	50,000	50,000	1.0000	.9715	48,576	344,775
8	50,000	50,000	1.0000	.9659	48,297	393,071
9	50,000	50,000	1.0000	.9599	47,995	441,067
10	50,000	50,000	1.0000	.9534	47,670	488,737
11	50,000	50,000	1.0000	.9464	47,318	536,055
12	50,000	50,000	1.0000	.9388	46,938	582,993
13	50,000	50,000	1.0000	.9305	46,527	629,519
14	50,000	50,000	1.0000	.9216	46,081	675,601
15	50,000	50,000	1.0000	.9120	45,601	721,201
16	50,000	50,000	1.0000	.9017	45,083	766,284
17	50,000	50,000	1.0000	.8905	44,525	810,810
18	50,000	50,000	1.0000	.8786	43,928	854,738
19	50,000	50,000	1.0000	.8658	43,291	898,028
20	50,000	50,000	1.0000	.8522	42,611	940,639
21	0	0	1.0000	.8378	0	940,639
21	50,000	50,000	1.0000	.8378	41,889	982,528
22	50,000	50,000	1.0000	.8224	41,122	1,023,650
23	50,000	50,000	1.0000	.8062	40,311	1,063,961
24	50,000	50,000	1.0000	.7890	39,451	1,103,412
25	50,000	50,000	1.0000	.7708	38,541	1,141,953
26	50,000	50,000	1.0000	.7516	37,580	1,179,533
27	50,000	50,000	1.0000	.7313	36,565	1,216,098
28	50,000	50,000	1.0000	.7100	35,498	1,251,596
29	50,000	50,000	1.0000	.6876	34,379	1,285,976
30	50,000	50,000	1.0000	.6642	33,210	1,319,186
31	50,000	50,000	1.0000	.6399	31,995	1,351,181

32	50,000	50,000	1.0000	.6147	30,733	1,381,914
33	50,000	50,000	1.0000	.5884	29,422	1,411,336
34	50,000	50,000	1.0000	.5611	28,056	1,439,392
35	50,000	50,000	1.0000	.5326	26,632	1,466,024
36	50,000	50,000	1.0000	.5031	25,153	1,491,177
37	50,000	50,000	1.0000	.4726	23,629	1,514,806
38	50,000	50,000	1.0000	.4414	22,070	1,536,876
39	50,000	50,000	1.0000	.4098	20,492	1,557,368
40	50,000	50,000	1.0000	.3781	18,905	1,576,273
41	50,000	50,000	1.0000	.3464	17,319	1,593,592
42	50,000	50,000	1.0000	.3148	15,739	1,609,330
43	50,000	50,000	1.0000	.2836	14,182	1,623,512
44	50,000	50,000	1.0000	.2533	12,665	1,636,177
45	50,000	50,000	1.0000	.2241	11,203	1,647,380
46	50,000	50,000	1.0000	.1959	9,795	1,657,175
47	50,000	50,000	1.0000	.1690	8,449	1,665,624
48	50,000	50,000	1.0000	.1437	7,183	1,672,807
49	50,000	50,000	1.0000	.1205	6,023	1,678,830
50	50,000	50,000	1.0000	.0996	4,981	1,683,811
51	50,000	50,000	1.0000	.0812	4,060	1,687,870
52	50,000	50,000	1.0000	.0650	3,252	1,691,123
53	50,000	50,000	1.0000	.0511	2,557	1,693,680
54	50,000	50,000	1.0000	.0395	1,973	1,695,652
55	50,000	50,000	1.0000	.0298	1,492	1,697,145
56	50,000	50,000	1.0000	.0222	1,108	1,698,253
57	50,000	50,000	1.0000	.0162	809	1,699,063
58	50,000	50,000	1.0000	.0117	583	1,699,645
59	50,000	50,000	1.0000	.0083	414	1,700,059
60	50,000	50,000	1.0000	.0058	290	1,700,349
61	50,000	50,000	1.0000	.0040	201	1,700,550
62	50,000	50,000	1.0000	.0028	138	1,700,688
63	50,000	50,000	1.0000	.0019	94	1,700,782
64	50,000	50,000	1.0000	.0013	63	1,700,845
65	50,000	50,000	1.0000	.0008	42	1,700,886
66	50,000	50,000	1.0000	.0006	28	1,700,914
67	50,000	50,000	1.0000	.0004	18	1,700,932
68	50,000	50,000	1.0000	.0002	12	1,700,944
69	50,000	50,000	1.0000	.0002	8	1,700,952
70	50,000	50,000	1.0000	.0001	5	1,700,956

Retention Layer	Reserve Share
1	940,639
2	760,317
Total	1,700,956

SCENARIO 2

AGE OF CLAIMANT = 40
INTEREST ON MEDICAL PAYMENTS = 6 %
INFLATION RATE FOR MEDICAL PAYMENTS = 6 %
MORTALITY RATE = 1 TIMES STANDARD MORTALITY
RETENTION LIMITS = 1,000,000 5,000,000

Year	Medical Payments Current Dollars	Medical Payments Inflated Dollars	Disc. Factor	Mort. Factor	Reserve Amount	Cumulative Reserve Amount
1	50,000	51,478	.9713	.9969	49,843	49,843
2	50,000	54,567	.9163	.9935	49,673	99,516
3	50,000	57,841	.8644	.9898	49,489	149,005
4	50,000	61,311	.8155	.9858	49,289	198,295
5	50,000	64,990	.7693	.9814	49,071	247,366
6	50,000	68,889	.7258	.9767	48,833	296,199
7	50,000	73,023	.6847	.9715	48,576	344,775
8	50,000	77,404	.6460	.9659	48,297	393,071
9	50,000	82,048	.6094	.9599	47,995	441,067
10	50,000	86,971	.5749	.9534	47,670	488,737
11	50,000	92,190	.5424	.9464	47,318	536,055
12	50,000	97,721	.5117	.9388	46,938	582,993
13	50,000	103,584	.4827	.9305	46,527	629,519
14	12,743	27,982	.4554	.9216	11,744	641,263
14	37,257	81,817	.4554	.9216	34,337	675,601
15	50,000	116,387	.4296	.9120	45,601	721,201
16	50,000	123,370	.4053	.9017	45,083	766,284
17	50,000	130,773	.3823	.8905	44,525	810,810
18	50,000	138,619	.3607	.8786	43,928	854,738
19	50,000	146,936	.3403	.8658	43,291	898,028
20	50,000	155,752	.3210	.8522	42,611	940,639
21	50,000	165,097	.3029	.8378	41,889	982,528
22	50,000	175,003	.2857	.8224	41,122	1,023,650
23	50,000	185,503	.2695	.8062	40,311	1,063,961
24	50,000	196,634	.2543	.7890	39,451	1,103,412
25	50,000	208,432	.2399	.7708	38,541	1,141,953
26	50,000	220,938	.2263	.7516	37,580	1,179,533
27	50,000	234,194	.2135	.7313	36,565	1,216,098
28	50,000	248,245	.2014	.7100	35,498	1,251,596
29	50,000	263,140	.1900	.6876	34,379	1,285,976
30	50,000	278,929	.1793	.6642	33,210	1,319,186
31	50,000	295,664	.1691	.6399	31,995	1,351,181

32	50,000	313,404	.1595	.6147	30,733	1,381,914
33	50,000	332,208	.1505	.5884	29,422	1,411,336
34	50,000	352,141	.1420	.5611	28,056	1,439,392
35	50,000	373,269	.1340	.5326	26,632	1,466,024
36	33,304	263,544	.1264	.5031	16,754	1,482,777
36	16,696	132,122	.1264	.5031	8,399	1,491,177
37	50,000	419,405	.1192	.4726	23,629	1,514,806
38	50,000	444,570	.1125	.4414	22,070	1,536,876
39	50,000	471,244	.1061	.4098	20,492	1,557,368
40	50,000	499,519	.1001	.3781	18,905	1,576,273
41	50,000	529,490	.0944	.3464	17,319	1,593,592
42	50,000	561,259	.0891	.3148	15,739	1,609,330
43	50,000	594,935	.0840	.2836	14,182	1,623,512
44	50,000	630,631	.0793	.2533	12,665	1,636,177
45	50,000	668,469	.0748	.2241	11,203	1,647,380
46	50,000	708,577	.0706	.1959	9,795	1,657,175
47	50,000	751,091	.0666	.1690	8,449	1,665,624
48	50,000	796,157	.0628	.1437	7,183	1,672,807
49	50,000	843,926	.0592	.1205	6,023	1,678,830
50	50,000	894,562	.0559	.0996	4,981	1,683,811
51	50,000	948,235	.0527	.0812	4,060	1,687,870
52	50,000	1,005,130	.0497	.0650	3,252	1,691,123
53	50,000	1,065,437	.0469	.0511	2,557	1,693,680
54	50,000	1,129,364	.0443	.0395	1,973	1,695,652
55	50,000	1,197,125	.0418	.0298	1,492	1,697,145
56	50,000	1,268,953	.0394	.0222	1,108	1,698,253
57	50,000	1,345,090	.0372	.0162	809	1,699,063
58	50,000	1,425,796	.0351	.0117	583	1,699,645
59	50,000	1,511,343	.0331	.0083	414	1,700,059
60	50,000	1,602,024	.0312	.0058	290	1,700,349
61	50,000	1,698,145	.0294	.0040	201	1,700,550
62	50,000	1,800,034	.0278	.0028	138	1,700,688
63	50,000	1,908,036	.0262	.0019	94	1,700,782
64	50,000	2,022,518	.0247	.0013	63	1,700,845
65	50,000	2,143,869	.0233	.0008	42	1,700,886
66	50,000	2,272,502	.0220	.0006	28	1,700,914
67	50,000	2,408,852	.0208	.0004	18	1,700,932
68	50,000	2,553,383	.0196	.0002	12	1,700,944
69	50,000	2,706,586	.0185	.0002	8	1,700,952
70	50,000	2,868,981	.0174	.0001	5	1,700,956

Retention Layer	Reserve Share
1	641,263
2	841,514
3	218,179
Total	1,700,956

SCENARIO 3

AGE OF CLAIMANT = 40
INTEREST ON MEDICAL PAYMENTS = REALISTIC
INFLATION RATE FOR MEDICAL PAYMENTS = REALISTIC
MORTALITY RATE = 1 TIMES STANDARD MORTALITY
RETENTION LIMITS = 1,000,000 5,000,000

Year	Medical Payments Current Dollars	Medical Payments Inflated Dollars	Disc Factor	Mort. Factor	Reserve Amount	Cumulative Reserve Amount
1	50,000	51,841	.9623	.9969	49,727	49,727
2	50,000	55,729	.8910	.9935	49,328	99,056
3	50,000	60,188	.8250	.9898	49,146	148,202
4	50,000	65,303	.7639	.9858	49,174	197,376
5	50,000	71,181	.7073	.9814	49,409	246,785
6	50,000	77,587	.6549	.9767	49,626	296,411
7	50,000	84,570	.6064	.9715	49,821	346,232
8	50,000	92,181	.5615	.9659	49,993	396,225
9	50,000	100,477	.5199	.9599	50,141	446,367
10	50,000	109,520	.4814	.9534	50,263	496,629
11	50,000	119,377	.4478	.9464	50,588	547,217
12	43,453	112,045	.4185	.9388	44,018	591,235
12	6,547	16,883	.4185	.9388	6,633	597,868
13	50,000	139,242	.3911	.9305	50,676	648,544
14	50,000	150,381	.3655	.9216	50,660	699,204
15	50,000	162,411	.3416	.9120	50,600	749,805
16	50,000	175,404	.3193	.9017	50,493	800,298
17	50,000	189,437	.2984	.8905	50,335	850,633
18	50,000	204,592	.2789	.8786	50,124	900,757
19	50,000	220,959	.2606	.8658	49,858	950,615
20	50,000	238,636	.2436	.8522	49,534	1,000,149
21	50,000	257,727	.2287	.8378	49,381	1,049,530
22	50,000	275,767	.2158	.8224	48,935	1,098,464
23	50,000	295,071	.2035	.8062	48,421	1,146,886
24	50,000	315,726	.1920	.7890	47,836	1,194,721
25	50,000	337,827	.1812	.7708	47,173	1,241,895
26	50,000	361,475	.1709	.7516	46,430	1,288,325
27	50,000	386,778	.1612	.7313	45,603	1,333,929
28	50,000	413,853	.1521	.7100	44,690	1,378,619
29	50,000	442,822	.1435	.6876	43,690	1,422,309
30	43,794	415,009	.1354	.6642	37,315	1,459,624
30	6,206	58,811	.1354	.6642	5,288	1,464,912
31	50,000	506,987	.1283	.6399	41,627	1,506,539

32	50,000	537,406	.1222	.6147	40,367	1,546,905
33	50,000	569,651	.1164	.5884	39,013	1,585,918
34	50,000	603,830	.1108	.5611	37,555	1,623,473
35	50,000	640,060	.1056	.5326	35,989	1,659,462
36	50,000	678,463	.1005	.5031	34,314	1,693,776
37	50,000	719,171	.0957	.4726	32,542	1,726,318
38	50,000	762,321	.0912	.4414	30,685	1,757,003
39	50,000	808,061	.0868	.4098	28,761	1,785,764
40	50,000	856,544	.0827	.3781	26,788	1,812,552
41	50,000	907,937	.0788	.3464	24,773	1,837,325
42	50,000	962,413	.0750	.3148	22,727	1,860,052
43	50,000	1,020,158	.0714	.2836	20,674	1,880,726
44	50,000	1,081,367	.0680	.2533	18,639	1,899,365
45	50,000	1,146,249	.0648	.2241	16,644	1,916,009
46	50,000	1,215,024	.0617	.1959	14,692	1,930,700
47	50,000	1,287,926	.0588	.1690	12,792	1,943,493
48	50,000	1,365,201	.0560	.1437	10,980	1,954,473
49	50,000	1,447,113	.0533	.1205	9,293	1,963,766
50	50,000	1,533,940	.0508	.0996	7,760	1,971,526
51	50,000	1,625,977	.0484	.0812	6,384	1,977,910
52	50,000	1,723,535	.0461	.0650	5,163	1,983,073
53	50,000	1,826,947	.0439	.0511	4,099	1,987,172
54	50,000	1,936,564	.0418	.0395	3,192	1,990,363
55	50,000	2,052,758	.0398	.0298	2,438	1,992,801
56	50,000	2,175,924	.0379	.0222	1,827	1,994,629
57	50,000	2,306,479	.0361	.0162	1,348	1,995,976
58	50,000	2,444,868	.0344	.0117	979	1,996,955
59	50,000	2,591,560	.0327	.0083	702	1,997,657
60	50,000	2,747,053	.0312	.0058	497	1,998,154
61	50,000	2,911,877	.0297	.0040	348	1,998,502
62	50,000	3,086,589	.0283	.0028	241	1,998,743
63	50,000	3,271,785	.0269	.0019	165	1,998,907
64	50,000	3,468,092	.0256	.0013	112	1,999,019
65	50,000	3,676,177	.0244	.0008	75	1,999,094
66	50,000	3,896,748	.0233	.0006	50	1,999,144
67	50,000	4,130,553	.0222	.0004	33	1,999,177
68	50,000	4,378,386	.0211	.0002	22	1,999,199
69	50,000	4,641,089	.0201	.0002	14	1,999,213
70	50,000	4,919,554	.0191	.0001	9	1,999,223

Retention Layer	Reserve Share
1	591,235
2	868,389
3	539,599
Total	1,999,223

REGRESSION MODELS IN CLAIMS ANALYSIS I: THEORY

GREG C. TAYLOR

Abstract

This paper considers the application of regression techniques to the analysis of claims data. Examples are given to indicate why, in certain circumstances, this might be preferable to traditional actuarial methods.

The various errors of prediction which occur when loss reserves are estimated by regression are classified and discussed.

Formal procedures are discussed for determining which of the available predictors will be entered into a regression, and the drawbacks of these procedures.

Various approaches to the estimation of uncertainty associated with loss reserves estimated by regression are considered.

The effect on regression techniques of outlying data points, and hence the subject of robust/resistant regression, is considered briefly.

1. INTRODUCTION

Regression models have not been prevalent in claims analysis leading to loss reserving. This is evident from a survey of claims reserving methods (Taylor, [23]).

The scarcity arises from the suspicion with which many actuaries regard such models. Their use does not have the "hands on" nature characteristic of methods based on age-to-age factors, for example, with which actuaries tend to feel at ease. There is a feeling of abstractness and loss of control in the estimation of parameters from the data.

This skepticism is justified by the countless misapplications of regression methods which occur in practice. Despite this, it appears that regression techniques have a very definite place in the actuarial repertoire. But they will serve their users effectively only if it is realized that blind and mechanical application of simple least squares regression will, in certain circumstances, be statistically inefficient.

In these circumstances, regression becomes a delicate tool rather than the crude bludgeon as which it is often regarded, and in which role it is even more often used. A proposition which is all too often neglected in practice is that a user can expect effective performance of any body of methodology only if the user is aware of its general properties, its strengths and weaknesses, the circumstances in which it should and should not be applied, the response of its output to input anomalies, the whole array of quirks and pitfalls awaiting the unwary, how to "tune" the model building procedure for maximum results, and so on.

The intention of this paper is to canvass briefly the various aspects of regression modelling. Within this larger purpose, there are two intentions. First, some of the grosser abuses of such modelling will be suitably exposed. Second, from a more positive viewpoint, it is hoped that the exposure of the causes of anomalous regression output will set the procedures in a perspective from which their beneficial aspects can be more clearly seen.

The following sections deal very briefly with such questions as:

(i) Why use regression models as opposed to the "traditional" actuarial ones such as those using age-to-age factors?
(ii) Precisely what criteria are to be satisfied, and how should the extent to which they are satisfied be assessed?
(iii) How many of the available predictors should be included in a regression model, and how should the choice be made?
(iv) What procedures, other than ordinary least squares regression, are available for fitting the selected model to data?
(v) How might the impact on the fitting of isolated rogue data points be assessed, and how might the fitting procedures be modified to reduce this impact?

2. MOTIVATING EXAMPLES

Consider first a relatively complex example. A simpler one will be presented shortly.

In what follows, let

i = year of occurrence of claim;
j = development year, i.e., number of years after year of occurrence;
N_i = number of claims incurred in year of occurrence i;
N_{ij} = number of claims settled in (i,j);
C_{ij} = amount of claim payments (adjusted for claims escalation) in (i,j);
$S_{ij} = C_{ij}/N_{ij}$ = average claim payment per settlement in (i,j);
$F_{ij} = N_{ij}/N_i$ = rate of settlement in (i,j);
$t_i(j) = \Sigma_{k=0}^{j} N_{ik}/N_i$ = proportion of claims from year of occurrence i settled by the end of development year j;
$\bar{t}_{ij}(k) = \min(\frac{1}{2}[t_i(j) + t_i(j+1)], u_k)$ for some partition $\{u_0, \ldots, u_{n+1}\}$ of $[0,1]$.

Suppose that the following model has been suggested:

$$S_{ij} = a + \sum_{k=0}^{n} b_k \bar{t}_{ij}(k) + c/F_{ij} + e_{ij}, \tag{2.1}$$

where a, b_0, . . . , b_n, and c are unknown parameters and e_{ij} is a random error term. This is the invariant see-saw model (Taylor, [22]).

Formula (2.1) expresses S_{ij} as a linear function of the observations $\bar{t}_{ij}(0)$, . . . , $\bar{t}_{ij}(n)$, $1/F_{ij}$ and a random error. Evidently, the unknown parameters may be determined by some form of linear regression of the S_{ij} on these observations.

Indeed, how else might the parameter estimation be carried out? Note that the parameter values a, b_0, . . . , b_n, c are common to all cells (i,j). In contrast with the example below involving age-to-age factors, there is no simple transformation of the dependent variable S_{ij} which will isolate any one of the parameters.

In this example, the very "shape" of the model, the intertwining of dependent and independent variables, virtually demands regression for parameter estimation.

In the next example, a much simpler model is considered but the situation is somewhat subtler. Using the same notation as before, let

$$C_{ij} = N_i u_i r_j + e_{ij}, \tag{2.2}$$

where

u_i = average claim size (adjusted for claims escalation) experienced in year of occurrence i;

r_j = the average proportion of claim payments (again adjusted for claims escalation) deriving from year of origin i which are payable in development year j.

The model (2.2) may be rewritten in the form:

$$\log C_{ij} = \log (N_i u_i) + \log r_j + f_{ij}, \tag{2.3}$$

where f_{ij} is a new random error term. The transformed model (2.3) is linear in the parameters $\log (N_i u_i)$ and $\log r_j$ which may therefore be estimated by regression methods. This indeed is the basis of Kremer's [13] ANOVA approach.

Note also, however, that (2.2) is the prototype for development of age-to-age factors (e.g., Skurnick, [20]; Berquist and Sherman, [3]). This is because it implies

$$C_{i,j+1}/C_{ij} = r_{j+1}/r_j + \text{error term}, \tag{2.4}$$

or more commonly,

$$A_{i,j+1}/A_{ij} = \sum_{k=0}^{j+1} r_k / \sum_{k=0}^{j} r_k + \text{error term}, \tag{2.5}$$

where

$A_{ij} = \sum_{k=0}^{j} C_{ik}$ = total claim payments (adjusted for claims escalation) made in respect of year of occurrence i up to the end of development year j,

and the r_{j+1}/r_j in (2.4), or the $\sum_{k=0}^{j+1} r_k / \sum_{k=0}^{j} r_k$ in (2.5), are the age-to-age factors.

This example is subtler than the previous one in the sense that one has a *choice* as to the method of estimation of its parameters. This choice should be made against reasonable criteria, and therefore one needs to specify these.

Consider, for example, the following three possibilities:

(i) proceed with regression estimation of the log $(N_i u_i)$ and log r_j via (2.3) after appropriate specification of f_{ij};

(ii) ignore the error term in (2.5) and estimate $\sum_{k=0}^{j+1} r_k / \sum_{k=0}^{j} r_k$ by $A_{i,j+1}/A_{ij}$;

(iii) assume the vector (log r_0, log r_1, . . .) to lie within some finite-dimensional vector space spanned by $g_1, g_2, \ldots, g_s$ where $g_m = (g_{m0}, g_{m1}, \ldots)$, and so use regression methods to fit the following adaptation of (2.3):

$$\log C_{ij} = \log (N_i u_i) + \sum_{m=1}^{s} b_m g_{mj} + f_{ij}. \tag{2.6}$$

It is instructive to consider the number of parameters to be estimated in each case.

In case (i), there are $I + J - 1$ parameters if I values of i are considered and r_j is assumed zero for $j = J, J + 1$, etc. The -1 arises from the constraint

$$\sum_{k=0}^{\infty} r_j = 1, \tag{2.7}$$

by definition.

In case (ii), there are again $I + J - 1$ parameters, the only difference between the two cases being that the former approaches parameter estimation in a formal manner whereas the latter takes an *ad hoc* approach.

In case (iii), where the g_m are fully specified in advance, the parameters $b_1, \ldots, b_s$ and the log $(N_i u_i)$ number just $I + s$.

Note that in the last case the number of parameters is independent of J. This contrasts with the first two cases in which increasing J without limit increases the number of model parameters also without limit. For example, consider the case $I = J = 10$, $s = 3$. The numbers of model parameters are:

Case (i): 19
Case (ii): 19
Case (iii): 13.

Case (iii) involves only two-thirds as many parameters as cases (i) and (ii).

Note that the number of parameters in case (iii) may be reduced further by treating the log $(N_i u_i)$ in the same way as the log r_j and representing them in some vector space of reduced dimension.

Every actuary is aware, intuitively at least, of the dangers of over fitting, i.e., fitting a model involving more parameters than are justified by the volume of available data. Generally, the fitting of models which are parsimonious in their use of parameters smoothes out the roughness inherent in the raw observations. Increasing the number of model parameters diminishes this smoothing effect until ultimately, when there are enough parameters that they are in one-to-one correspondence with the observations, the instability of the parameter estimates is equal to that of the observations themselves.

Many actuaries have a distaste for models like (2.6) on the ground that the parameters b_m under estimation are too abstract, that they do not correspond sufficiently with real world objects. This is what I meant in referring at the start of Section 1 to the "hands on" nature of the more traditional actuarial models.

The formal objection to models such as (2.6) is likely to take the form: "What if basis vectors $b_1, \ldots, b_s$ cannot be found (for s sufficiently small to be useful) which capture the more subtle features of the r_j?"

The answer is that any such losses of accuracy cannot be considered in isolation from possible gains in stability accruing from a reduction in the number of model parameters requiring estimation. In formal terms, the approximation of (2.3) by (2.6) may introduce some *bias* into the model, but this bias must be weighed against any reduction in *variability* of the model's predictions.

The distaste for abstraction that individuals may experience is perhaps understandable, but ultimately the relative merits of competing models must be assessed by the models' objective performance, rather than the users' preferences or prejudices.

The above remarks concerning questions of bias versus stability do no more than state the intuitively obvious. However, it is possible, and useful, to formalize the concepts involved so that model selection (such as the choice between cases (i), (ii) and (iii) dealt with above) can proceed on a more rigorous basis.

These matters are pursued in Section 4. A helpful preliminary to this is an examination and classification of the types of error that arise in the prediction of future observations on the basis of a model fitted to past data. This forms the subject of Section 3.

3. ERRORS OF PREDICTIONS

3.1. Illustrative example

Again it will be useful to consider an example which is highly simplified but nevertheless illustrative of the wrong turns that can be taken in a slipshod approach to errors of prediction. Though oversimplified, the essence of the model corresponds to some of the approaches which I have seen in practice.

Suppose it is assumed for the model (2.2) that:

$$V[C_{ij}|\{u_i, r_j\}] = N_i u_i^2 \sigma^2, \tag{3.1.1}$$

where σ^2 is independent of both i and j.

Suppose also, that estimates $\hat{u}_i$, $\hat{r}_j$ of the u_i, r_j have been obtained in the manner described in case (ii) of Section 2. Hence, estimates $\hat{C}_{ij}$ corresponding to the observations C_{ij} have been found. More particularly, though, predictions $\hat{P}_i = \Sigma_{j=T+1}^{\infty} \hat{C}_{ij}$ have been obtained of the future claim payments $P_i = \Sigma_{j=T+1}^{\infty} C_{ij}$ arising from year of occurrence i, where C_{iT} is the latest observation on that year of occurrence.

Suppose that one seeks:

$$V[\hat{P}_i] = \sum_{j=T+1}^{\infty} V[\hat{C}_{ij}] + \text{covariances}. \tag{3.1.2}$$

In practice, the estimation of the covariances may prove awkward. However, let us concentrate for the moment on some of the pitfalls involved in the estimation of the $V[\hat{C}_{ij}]$.

An argument that seems to appeal to some practitioners begins by considering the scaled residuals $(C_{ij} - \hat{C}_{ij})/N_i^{1/2}\hat{u}_i$. If $\hat{C}_{ij}$ is regarded as replaceable by $E[C_{ij}]$, assuming $\hat{C}_{ij}$ to be unbiased, the squares of these residuals become estimators of σ^2, i.e.,

$$\hat{\sigma}^2 = n^{-1} \sum_{i,j} [(C_{ij} - \hat{C}_{ij})^2/N_i\hat{u}_i^2], \tag{3.1.3}$$

the summation running over the n pairs i,j for which observations exist, and perhaps with some reduction of n to reflect loss of degrees of freedom. The required estimate of $V[\hat{C}_{ij}]$ can then be obtained by means of (3.1.1) as:

$$n^{-1}N_i\hat{u}_i^2 \sum_{k,j} [(C_{kj} - \hat{C}_{kj})^2/N_k\hat{u}_k^2]. \tag{3.1.4}$$

While this procedure may appear a reasonable practical solution to the problem, uncluttered by the quibbles of purists, it is suggested that it is in fact far from the truth. It is suggested further to contain a major error of reasoning likely to carry substantial numerical consequences. $V[\hat{P}_i]$ is not even the second moment of interest. Even if it were, material contributions to it have been omitted.

Essentially, the difficulties arise from the cavalier approach to the problem. A more careful and organized approach is required.

3.2. Component errors of prediction

To achieve the requirement of the previous subsection, let us drop the particular problem we have been considering and consider a generalized problem instead. Let Y denote an observable n-vector whose ith component is, apart from random noise, some function of observable quantities $X_{i1}, \ldots, X_{ip}$:

$$Y = f(X) + e, \tag{3.2.1}$$

where X is the $n \times p$ matrix with X_{ij} as (i,j)-element, f: $R^{np} \to R^n$ has the particular (possibly non-linear) form described above, and e is a random error term with zero mean.

Suppose that the functional form f is unknown in this context and consider linear approximations Xb to $f(X)$ where b is a p-vector of parameters. Then (3.2.1) becomes:

$$Y = Xb + [f(X) - Xb] + e. \tag{3.2.2}$$

Suppose further that the exact set of independent variables on which Y depends (the columns of X) is unknown, and that as a consequence Y is modelled as a linear function of a subset of Y, i.e., Y_i is modelled by:

$$\sum_{j \in A} X_{ij} b_j \tag{3.2.3}$$

for some $A \subset \{1, 2, \ldots, p\}$ instead of by $\sum_{j=1}^{p} X_{ij} b_j$.

Let (3.2.3) be denoted by $X_A b_A$, whereupon (3.2.2) decomposes as:

$$Y = X_A b_A + X_B b_B + [f(X) - Xb] + e, \tag{3.2.4}$$

where B denotes the set $\{1, 2, \ldots, p\} - A$.

Let $\hat{b}_j$ denote the regression estimate of b_j, where the term "regression estimate" is deliberately left vague for the moment. Let X^* denote an $m \times p$ matrix, each column of which represents m further values of the relevant predictor. The task is to predict the m-vector

$$Y^* = f^*(X^*) + e^*, \tag{3.2.5}$$

where now $f^*: R^{mp} \to R^m$.

Corresponding to (3.2.4):

$$Y^* = X_A^* b_A + X_B^* b_B + [f^*(X^*) - X] + e^* \tag{3.2.6}$$

Let $\hat{Y}^*$ be the regression prediction of Y^*:

$$\hat{Y}^* = X_A^* \hat{b}_A, \tag{3.2.7}$$

so that the prediction error is:

$$\begin{aligned} Y^* - \hat{Y}^* &= X_A^*(b_A - \hat{b}_A) + X_B^* b_B + [f^*(X^*) - X] + e^* \\ &= X_A^*(E\hat{b}_A - \hat{b}_A) + [X_A^*(b_A - E\hat{b}_A) + X_B^* b_B] \\ &\quad + [f^*(X^*) - X] + e^*. \end{aligned} \tag{3.2.8}$$

In many applications X represents observation of the predictors in the past, and X^* represents values to be assumed by the same predictors in the future.

At this point it is convenient to stop and consider the components of prediction error appearing on the right side of (3.2.8). They are:

(i) the *specification error* $[f^*(X^*) - X]$ essentially due to unmodeled nonlinearity;
(ii) the *selection error* $[X_B^* b_B + X_A^*(b_A - E\hat{b}_A)]$ due to incorrect selection of predictors;
(iii) the *estimation error* $X_A^*(E\hat{b}_A - \hat{b}_A)$ arising from the fact that even the most efficient estimators of the regression coefficients are still only random variables; and,
(iv) the *statistical error* e^* reflecting the inherent random noise in the process.

The terminology in (i), (iii), and (iv) is taken from Bartholomew [2]. The terminology in (ii) is taken from Miller [15].

By the first version of (3.2.8), it might appear simpler to regard $X_A^*(b_A - \hat{b}_A)$ as estimation error and $X_B b_B^*$ as selection error. Note, however, that the selection of the set A of (linear) predictors instead of $A \cup B$ introduces a bias in $\hat{b}_A$ as estimator of b_A. For example, in the case of ordinary least squares regression with no specification error,

$$\hat{b}_A = (X_A^T X_A)^{-1} X_A^T Y,$$

whence (3.2.4) yields

$$E\hat{b}_A = b_A + (X_A^T X_A)^{-1} X_A^T X_B b_B. \tag{3.2.9}$$

In the case of claims analysis, it is possible to characterize the four contributions to prediction error as follows.

In the fitting of a model to past claims data, the wrong algebraic model structure may be chosen. This will lead to specification error.

Suppose that the true underlying model is in fact linear, and all of the relevant predictors are identified, so that there is no specification error. Still it will usually be necessary to use past data (incorporating its random noise) to decide which of the available predictors are included in the model. The noise in the process may lead to wrong decisions; relevant predictors may be omitted, and irrelevant ones included. This will result in selection bias.

Suppose that the true underlying model is linear and is correctly selected, so that there is neither specification nor selection error. Still it will be necessary to estimate the parameters of the linear model by reference to past data. As these data contain random noise, so will the parameter estimates. The deviation of these estimates from their true values constitutes estimation error.

Suppose that, by some unspecified means, it were possible to select the correct (linear) model form and estimate its parameters precisely, so that there were no specification, selection, or estimation error. Even then future claims experience could not be predicted with precision because the inherent randomness of the claims process would generate deviations of experience from expected values. These deviations constitute statistical error.

3.3. Prediction bias and mean square error of prediction

Let us now consider the *prediction bias* $E\hat{Y}^* - EY^*$ and the *mean square error of prediction* (MSEP)

$$E(Y^* - \hat{Y}^*)^2 = E(Y^* - \hat{Y}^*)^T (Y^* - \hat{Y}^*).$$

By (3.2.6) and (3.2.7) the prediction bias is:

$$E\hat{Y}^* - EY^* = X_A(E\hat{b}_A - b_A) - X_B - [f^*(X^*) - X]. \tag{3.3.1}$$

In finding an expression for the MSEP, it will be advantageous to decompose the prediction error as:

$$\begin{aligned} Y^* - \hat{Y}^* &= (Y^* - EY^*) - (\hat{Y}^* - E\hat{Y}^*) - (E\hat{Y}^* - EY^*) \\ &= (Y^* - EY^*) - (\hat{Y}^* - E\hat{Y}^*) - \text{prediction bias}. \end{aligned} \tag{3.3.2}$$

In any form of linear regression of Y on X_A, Y^* and $\hat{Y}^*$ will be uncorrelated (easily checked from first principles since the former depends on future observations and the latter on past), so that (3.3.2) yields:

$$\begin{aligned} \text{MSEP} &= E(Y^* - EY^*)^2 + E(\hat{Y}^* - E\hat{Y}^*)^2 + (\text{prediction bias})^2 \\ &= E(e^*)^2 + E[X^*(\hat{b}_A - E\hat{b}_A)]^2 + (\text{prediction bias})^2, \end{aligned} \tag{3.3.3}$$

by (3.2.5) and (3.2.7).

The MSEP is thus seen to comprise three identifiable contributions deriving from:

(i) statistical error;
(ii) estimation error; and,
(iii) prediction bias (incorporating specification error and selection error).

It is convenient at this point to revert to the example of Section 3.1, recalling particularly the critical remarks made at the end of that section.

With the benefit of the more formal analysis of Section 3.2 and the present subsection, it is possible to recognize that the expression (3.1.4) for $V[\hat{P}_i]$ is essentially only estimation error. Both statistical error and prediction bias are omitted.

3.4 Components of selection error

Section 3.2 defined selection error as the term $[X_B^* b_B + X_A^*(b_A - E\hat{b}_A)]$ in (3.2.8). As seen in (3.3.1), this is the part of prediction bias *not* arising from nonlinearity. It was shown in Section 3.2 that the first member represents the bias introduced directly by the omission of the set B of predictors; the second member is the bias in $\hat{b}_A$ arising from this omission.

It must now be recognized that $E\hat{b}_A$ has been implicitly regarded as an unconditional expectation in the above. This would be appropriate if the set A were chosen without reference to the data Y. In practice, however, and particularly in claims analysis, this will not be the case. Usually, A will be chosen because it produces a better fit of model to data than certain other sets.

In this case,

$$E\hat{b}_A = E[\hat{b}_A|P] + \{E\hat{b}_A - E[\hat{b}_A|P]\}, \tag{3.4.1}$$

where $E\hat{b}_A$ is now explicitly the unconditional expectation of $\hat{b}_A$ and P denotes the procedure for subset selection. Substitution of (3.4.1) in the expression for selection error given at the start of this subsection yields:

$$\text{selection bias} = X_B^* b_B + X_A^*\{b_A - E[\hat{b}_A|P]\} + X_A^* \{E[\hat{b}_A|P] - E\hat{b}_A\}. \tag{3.4.2}$$

There are now three contributions to selection bias:

(i) *omission bias,* consisting of the first two members on the right of (3.4.2), and representing the bias due to the omission of the set B of predictors;
(ii) *stopping rule bias,* consisting of that part of the final member of (3.4.2) which arises from the limitation imposed by P on the number of predictors included in A; and
(iii) *competition bias,* consisting of that part of the final member of (3.4.2) which, for a given size of set A, arises from the manner in which P selects A from subsets of $A \cup B$ of that size.

These components of selection error are discussed in some detail by Miller [15] (pp. 400–405), who gives various other references.

Miller also gives a simple example of competition bias in a case in which:

(i) $A \cup B$ consists of just 2 predictors;
(ii) A consists of just a single predictor;
(iii) P consists of selection of the single predictor according to ordinary least squares;
(iv) $E\hat{b}_1 = 1$ and $V[\hat{b}_1] = V[\hat{b}_2]$; and,
(v) the size of the sample of observations is large (presumably, results will be worse otherwise).

Values of $E[\hat{b}_1|\text{variable 1 selected}]$ are calculated for varying values of $E\hat{b}_2$, $V[\hat{b}_i]$, and $C[\hat{b}_1,\hat{b}_2]$, and range from 1.02 to 1.53, compared with $E\hat{b}_1 = 1$. The value of $E[\hat{b}_1|$ variable 1 selected$]$ increases with increase in each of the variables $E\hat{b}_2$, $V[b_i]$, and $C[\hat{b}_1,\hat{b}_2]$.

Thus, it is apparent that selection bias may be substantial. Miller suggests a number of possible remedies for competition bias, though not for stopping rule bias. They are:

(i) using half of the data to select predictors and the other half to fit the model;
(ii) using jackknife or bootstrap methods (see Section 5 of this paper);
(iii) using shrunken estimators of the ridge or Stein type;
(iv) using simulation to estimate bias; or,
(v) using maximum likelihood estimation of regression coefficients taking the subset selection procedure P into account in the likelihood.

4. SUBSET SELECTION

4.1. General

Consider the method by which the subset A of predictors (in the terminology of Section 3) might be chosen. What criterion might be adopted?

Starting at the most naive end of reasoning, consider (3.2.8) and the identification of the different types of prediction error in the passage immediately following. If all available predictors are included in A, then B is empty and selection error falls to zero.

However, such a suggestion is likely to introduce the very practical problem (and, we shall see shortly, the theoretically objectionable fact) that the number of predictors runs literally into hundreds. Moreover, the evidence may be that the majority are statistically insignificant.

Alternatively, then, one might consider including in A only those predictors which can be demonstrated as statistically significant, and specifically as significant not only in isolation but also in conjunction with the other members of A. This, typically, is the type of procedure followed by stepwise regressions (Efroymson, [6]).

Certainly, this alternative procedure might reduce selection error to quite tolerable levels. It is necessary to recognize, however, that reduction of this type of error does not of itself result in efficient prediction. High efficiency in fact requires a low MSEP.

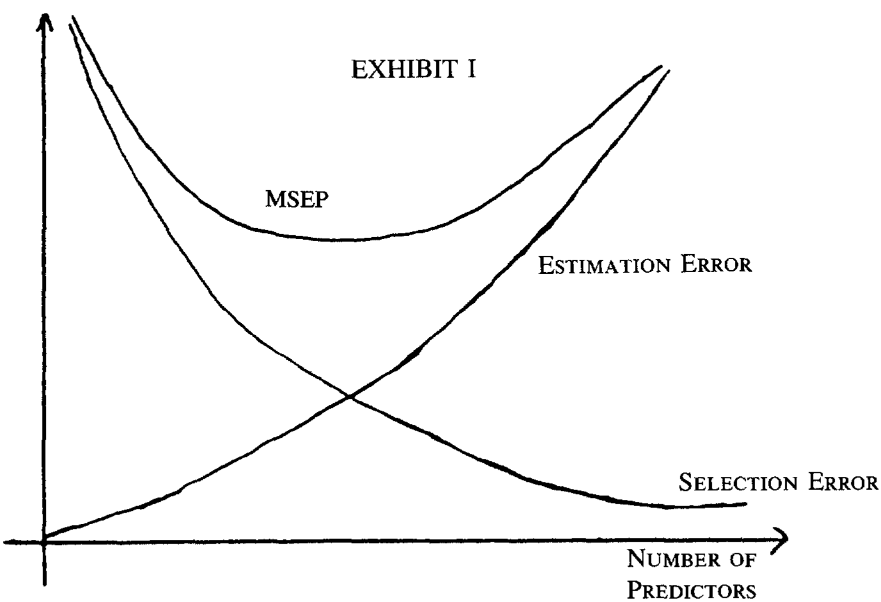
Recall from (3.3.3) and the text just following it that the MSEP consists of three components, only one of which is selection error. Of the remaining two, statistical error is independent of the model selected. Hence, an examination of the prediction efficiency of various models amounts to an examination of the respective effects of increasing the number of predictors on:

(i) selection error (as noted above, this decreases); and,
(ii) estimation error.

It turns out that, broadly, estimation error increases as the set of predictors increases. This is intuitive. The more predictors that need to be fitted to a fixed number of data points, the more difficult the fitting becomes. As the number of predictors becomes too large, the phenomenon of over fitting mentioned in Section 2 becomes more in evidence.

In the extreme case in which the numbers of data points and predictors are roughly equal, the whole fitting procedure is concentrated on achieving adherence of the model to past observation. The model is then being fitted to the random noise of past observation as well as the underlying signal, with consequent loss of predictive power. That is, estimation error is increased.

The opposite effects on selection error and estimation error of increasing the number of predictors are illustrated by Exhibit I.



This indicates the existence of an optimal subset of available predictors in the sense of minimizing MSEP. The next couple of subsections deal with simple statistics aimed at facilitating the selection of the subset which is optimal or, more realistically, which is not too far sub-optimal.

4.2 Mallows' C_p statistic

Consider once again the situation introduced in Section 3.2, but assume now the underlying algebraic structure $f(.)$ in (3.2.1) is linear. In this case (3.2.2) becomes:

$$Y = Xb + e, \tag{4.2.1}$$

where, as before, e has zero mean, and is further assumed to have stochastically independent components all with equal variance σ^2.

Recall the decomposition of MSEP:

$$E(Y^* - \hat{Y}^*)^2 = E(Y^* - EY^*)^2 + E(\hat{Y}^* - E\hat{Y}^*)^2 + (\text{prediction bias})^2. \tag{3.3.3}$$

A somewhat simplified version of this is:

$$\begin{aligned}\Delta = E(EY^* - \hat{Y}^*)^2 &= E(\hat{Y}^* - E\hat{Y}^*)^2 + (\text{prediction bias})^2 \\ &= \text{estimation error} + \text{prediction error}.\end{aligned} \tag{4.2.2}$$

The left side of (4.2.2) is a measure of deviation of the expected values of future observations from predictions, whereas MSEP is a measure of deviation of the actual values of future observations from predictions.

The difference between the two measures is the statistical error $E(e^*)^2$. Since this is independent of the model chosen, subset selection according to minimum MSEP is the same as minimizing Δ. This is the basis of Mallows' C_p statistic introduced by Mallows [14] and discussed by Seber ([19], pp. 364–369).

In the following, let a subscript q indicate that the quantity under consideration relates to a model based on q of the available predictors (one of them representing a constant term, i.e., a constant column of X). Seber shows that:

$$\Delta_q = q\sigma^2 + (PB)_q^2, \tag{4.2.3}$$

with PB denoting prediction bias.

Now the usual definition of *residual sum of squares* (RSS) is:

$$\text{RSS} = (Y - \hat{Y})^2,$$

and as is well-known,

$$E(\text{RSS}_q) = (n - q)\sigma^2 + (PB)_q^2. \tag{4.2.4}$$

By (4.2.3) and (4.2.4),

$$E(\text{RSS}_q) + (2q - n)\sigma^2 = \Delta_q.$$

Therefore, if

$$C_q = \text{RSS}_q/\hat{\sigma}^2 + 2q - n, \tag{4.2.5}$$

with $\hat{\sigma}^2$ a suitable estimator of σ^2, C_q will be an approximately unbiased estimator of Δ_q/σ^2. Then minimization of MSEP, equivalently of Δ_q, will be approximately achieved by selection of the subset of predictors which minimizes C_q defined by (4.2.5).

In the case in which the number of predictors included in the model is denoted by p (recall that this symbol has been reserved for the total number of available predictors), (4.2.5) becomes C_p. This is the name by which it is usually known—Mallows' C_p statistic.

4.3. Breiman and Freedman S_p statistic

Breiman and Freedman [4] consider a situation similar to that of Section 4.2. In their case, however, the elements of the design matrix X in (4.2.1) are random variables.

It is assumed, in addition to the assumptions of Section 4.2, that e and the columns of X are jointly normal with zero mean and that e is stochastically independent of the columns of X. As before σ^2 denotes $V(e^*)$, and in addition we adopt the notation:

$$\sigma_q^2 = V[X_B b_B | X_A], \tag{4.3.1}$$

where X_A, X_B, have the same meaning as in Section 3, the set A now containing q predictors.

Just as in Section 4.2, the quality of the regression is assessed by reference to the MSEP, though in the presence of random variation of X this requires further definition. Breiman and Freedman define

$$\text{MSEP} = E[E[(Y^* - \hat{Y}^*)^2|X,Y]], \tag{4.3.2}$$

where the outer expectation operator is unconditional, i.e., averages over the data X, Y. The algebra is developed in terms of the case $m = 1$ (i.e., the vector Y^* has a single component) though this does not result in any loss of generality in the S_p statistic presented below.

The algebraic development is rather similar to that of Section 4.2. The extended MSEP (4.3.2) may be written in the form, parallel to (3.3.3):

$$\begin{aligned}\text{MSEP} = \text{statistical error} &+ E[\text{estimation error}|X,Y]\\ &+ E[(\text{prediction bias})^2|X,Y].\end{aligned} \tag{4.3.3}$$

Now, apart from the averaging over data, the final two terms of (4.3.3) are those appearing as Δ on the right side of (4.2.2). Hence, (4.3.3) becomes:

$$\begin{aligned}\text{MSEP} &= \sigma^2 + E[(PB)^2|X,Y] + E[\text{estimation error}|X,Y]\\ &= \sigma^2 + \sigma_q^2 + E[\hat{\mathrm{b}}_A - E\hat{\mathrm{b}}_A)^T X_A^T X_A(\hat{b}_A - E\hat{b}_A)|X,Y],\end{aligned} \tag{4.3.4}$$

where use has been made of (4.3.1).

With a little further development, Breiman and Freedman show that:

$$\text{MSEP} = (\sigma^2 + \sigma_q^2)\ [1 + q/(n - 1 - q)]. \tag{4.3.5}$$

The first bracketed term on the right is estimated by $(n - q)^{-1}$(RSS), whence MSEP is estimated by

$$S_q = (n - q)^{-1}(\text{RSS})[1 + q/(n - 1 - q)]. \tag{4.3.6}$$

The paper by Breiman and Freedman goes on to demonstrate certain optimality properties of S_q.

In the case in which the number of predictors included in the model is denoted by p (recall that in the present paper this symbol has been reserved for the total number of available predictors), (4.3.6) becomes S_p. This is the name by which it is usually known—Breiman and Freedman S_p statistic.

In application of S_p, the subset of regression predictors is selected from those available in such a way as to minimize S_p.

The choice between C_p and S_p in regressions arising from claims analysis is not always easy. In the first example of Section 2, the values entering the design matrix, $\bar{t}_{ij}(k)$ and $1/F_{ij}$ will indeed be random variables, as allowed by S_p (but not C_p). On the other hand, however, their mean values will be necessarily non-zero, contrary to the assumption underlying S_p (but not C_p).

This will be particularly true of any constant term in the regression equation, such as a in (2.1). It is perhaps desirable to examine the behavior of both C_p and S_p as the subset of predictors entered into the regression is varied.

Miller [15] (pp. 406–407) suggests in strong terms that the efficacy of stopping rules such as those based on C_p and S_p is very much limited by the existence of competition bias (Section 3.4):

"the vast literature on stopping rules . . . is an irrelevant academic exercise until the problems of estimation have been overcome."

He points out that competition bias can easily be of the order of two standard errors when the same data set is used for subset selection and parameter estimation. He provides a simulated example in which the true MSEP is compared with that estimated, ignoring competition bias, by the formula:

$$\text{MSEP (false)} = [1 + (q + 1)/n]\ \text{RSS}/(n - 1 - q),$$

for a model containing q predictors and a constant term. The results were as shown in Exhibit II.

4.4. *Spjøtvoll's goodness-of-fit*

Spjøtvoll [21] provides a test of the goodness-of-fit of one subset of predictors relative to another. This is dealt with in reasonable detail by Miller [15] (pp. 397–399).

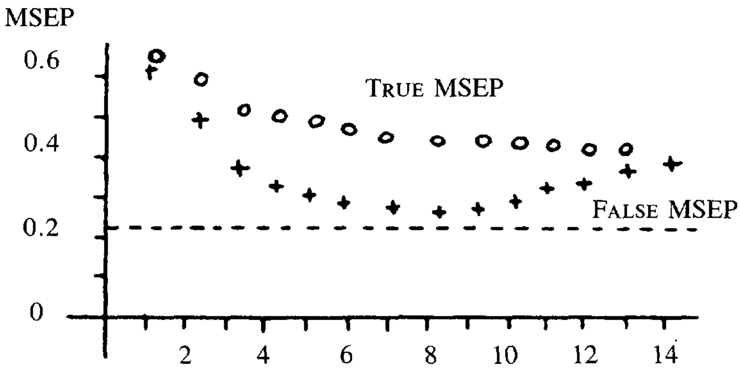
Spjøtvoll's measure of goodness-of-fit is:

$$(Xb - X_A E\hat{b}_A)^T (Xb - X_A E\hat{b}_A) = (Xb)^T(Xb) - (Xb)^T X_A(X_A^T X_A)^{-1} X_A^T(Xb). \tag{4.4.1}$$

Since the first member of this last expression is independent of the subset of predictors selected, Spjøtvoll chose to use just:

$$(Xb)^T X_A(X_A^T X_A)^{-1} X_A^T(Xb). \tag{4.4.2}$$

EXHIBIT II



Miller points out that, if goodness-of-fit is to be assessed for prediction purposes, (4.4.1) might reasonably be modified by the inclusion of a statistical error term. (See Section 3.2 for explanation.) Then (4.4.1) is replaced by:

$$(Xb - X_A\hat{b}_A)^T(Xb - X_A\hat{b}_A) = (Xb)^T(Xb) - (Xb)^T X_A(X_A^TX_A)^{-1}X_A^T(Xb) + \sigma^2 \text{ trace } [X_A(X_A^TX_A)^{-1}X_A^T],$$

where $\sigma^2 I = Ve$. This extra term is equal to $q\sigma^2$ (just as in (4.2.3)) when there are q linear predictors including a constant term, so that (4.4.1) is replaced by:

$$(Xb - X_A\hat{b}_A)^T(Xb - X_A\hat{b}_A) = (Xb)^T(Xb) - (Xb)^T X_A(X_A^TX_A)^{-1}X_A^T(Xb) + q\sigma^2, \tag{4.4.3}$$

and (4.4.2) by:

$$(Xb)^TX_A(X_A^TX_A)^{-1}X_A^T(Xb) - q\sigma^2. \tag{4.4.4}$$

Note that (4.4.3) is identical to Δ_q defined in (4.2.2) in the development of Mallows' C_p with the exception that in the latter case it is based on the future design matrix X^* whereas (4.4.3) is based on the past X.

By (4.4.4), different subsets of predictors, say M and N, are compared by means of the statistic:

$$G_{MN} = (Xb)^T[X_M(X_M^TX_M)^{-1}X_M^T - X_N(X_N^TX_N)^{-1}X_N^T](Xb) - (q_M - q_N)\sigma^2,$$
$$= b^TC_{MN}b - (q_M - q_N)\sigma^2, \qquad (4.4.5)$$

where C_{MN} is the appropriate $p \times p$ matrix. We note that the final member of this expression was not used by Spjøtvoll.

Spjøtvoll goes on (summarized by Miller) to develop maximum and minimum values for G_{MN} conditional upon b lying within a $(1 - \alpha)$ confidence set of the form:

$$\Pr[(b - \hat{b})^TX^TX^T(b - \hat{b}) \leq k] = 1 - \alpha,$$

where $\hat{b}$ is the regression estimate of b in the full model.

These limits on G_{MN} may be used to test whether M provides a significantly better or worse fit than N to the data.

5. METHODS OF ESTIMATION OF SECOND MOMENTS OF LOSS RESERVES

5.1. *General*

This section will consider methods by which MSEP of loss reserves can be estimated.

First note that this will not consist merely of estimating (3.3.3). Typically, Y^* will be some vector of future claim payments, subdivided for example according to year of occurrence and development year. In such a case, the estimated loss reserve would be:

$$\hat{R} = 1^T\hat{Y}^*, \qquad (5.1.1)$$

where 1 is an m-vector with every component equal to unity.

Then (3.3.3) is replaced by:

$$\text{MSEP}(R) = 1^T\text{e}(e^*)^2\ 1 + 1^TE[X^*(\hat{b}_A - E\hat{b}_A)]^2\ 1 + (\text{prediction bias})^2. \qquad (5.1.2)$$

This last equation shows that the MSEP of loss reserve R consists of separate terms representing statistical error, estimation error and prediction bias respectively.

There is little that can be said as to the formal inclusion of the last of these components in any estimate of MSEP. To the extent that it is perceptible, it should be removed from the estimated loss reserve (i.e., first moment thereof) rather than allowed for in MSEP estimation. Some components of prediction bias, e.g., specification error (Section 3.2), are by their very nature, likely to defy any reliable formal evaluation.

The usual situation is therefore that the first two members on the right of (5.1.2) can be evaluated in systematic manner, but only informal allowance can be made for the third, bearing in mind Miller's remarks quoted in Section 4.3.

There are several approaches to this evaluation. They are discussed in detail by Ashe [1]. Brief details are given in the next few subsections.

5.2. *Parametric estimation*

The linear model (4.2.1) will be referred to here as the parametric model—parametric in the sense that the error term e is assumed to have certain (usually parametric) properties.

If e is well-defined, then its parameters (e.g., σ^2) may be estimated from the data, and hence the first two components of MSEP(R) in (5.1.2) estimated. Logically, this is straightforward even if the algebraic manipulation involved may be cumbersome occasionally. The algebraic details are provided by Taylor and Ashe [24].

The calculations involved in this procedure are quite manageable with just about any reputable regression package. Naturally, the results are reliable only to the extent that the parametric assumptions underlying the procedure may be relied upon. Care is therefore necessary in dealing appropriately with the co-variance structure of e. See, for example, the weighting procedure used by Taylor and Ashe [24] in their regressions.

5.3. *Jackknife*

The jackknife algorithm was introduced by Quenouille [17] and is now found in many standard texts, e.g., Mosteller and Tukey [16]. The purpose of the algorithm was to reduce bias in parameter estimates based on limited data.

An outline of the method is as follows. Suppose that some parameter θ is estimated by a statistic S. This statistic may be a complicated function of the data. The precise properties of S are either unknown or difficult to compute. It is known, however, that the bias contained in S is of order n^{-1} for sample size n.

Let S be denoted by $S(n)$ for sample size n. Now, for each $i = 1, 2, \ldots,$ n, define $S_i(n)$ as the value of S based on the $(n - 1)$-sample obtained by deletion of the ith observation. Then define a pseudo-value:

$$P_i(n) = nS(n) - (n - 1)\, S_i(n),\ i = 1, 2, \ldots, n. \tag{5.3.1}$$

By assumption,

$$_ES(n) = \theta + a/n + o(n^{-1}).$$

Hence

$$_EP_i(n) = \theta + o(n^{-1}),$$

and so

$$\overline{P}(n) = \sum_{i=1}^{n} P_i(n)/n = \theta + o(n^{-1}) + \text{error term} \tag{5.3.2}$$

contains a bias of order less than n^{-1} as an estimator of θ.

The variance of $\overline{P}(n)$ is estimated by (Mosteller and Tukey, 1977, p. 135):

$$\{\overline{P^2}(n) - [\overline{P}(n)]^2\}/(n - 1), \tag{5.3.3}$$

where

$$\overline{P^2}(n) = \sum_{i=1}^{n} P^2(n)/n. \tag{5.3.4}$$

This algorithm may be applied to the present context by setting $S(n)$ equal to the estimated loss reserve obtained from a regression claims model based on n data points (a single data point being, for example, the observed claim payments in a given development year of a given year of occurrence). This can be generalized by taking $S(n)$ to be the vector of loss reserves for the different years of occurrence; or the vector of claim payments projected for each of the years of run-off; or, indeed, any one of the many cross-sections which might be taken from the regression forecast of future cash flows according to year of occurrence and development year.

In practical application, it might seem reasonable to adapt the jackknife estimates (5.3.1) to (5.3.4) to weighted regression. Possible replacement formulas are:

$$P_i(n) = [WS(n) - (W - w_i)S_i(n)]/w_i \tag{5.3.1a}$$

where w_i is the weight applied to observation i in the weighted regression and

$$W = \sum_{i=1}^{n} w_i;$$

$$\overline{P}(n) = \sum_{i=1}^{n} w_i P_i(n)/W; \tag{5.3.2a}$$

$$\{\overline{P}^2(n) - [\overline{P}(n)]^2\} \times \sum_{i=1}^{n} w_i^2/W^2 \tag{5.3.3a}$$

$$\overline{P}^2(n) = \sum_{i=1}^{n} w_i P_i^2(n)/W. \tag{5.3.4a}$$

Despite the seeming reasonableness of (5.3.1a) to (5.3.4a), Ashe [1] (p. S108) points out that the response of the weighted jackknife to his particular numerical examples is wild. It is possible that the bias assumption underlying the jackknife is incorrect and that the adoption of unequal weights w_i magnifies this in $\overline{P}(n)$. Indeed, Miller [15] (p. 404) provides a semi-rigorous argument that competition bias is of order $n^{-1/2}$, not n^{-1} as required for the jackknife to be valid.

Ashe [1] (p. S110) points out the usefulness of the pseudo-values in their own right as providing an indication of the influence of individual data points. A deviant value of $P_i(n)$ indicates that the whole regression is strongly influenced by data point i. Further discussion of the influence function and the appropriate response to it will appear in Section 6.

There are two shortcomings of the jackknife.

First, the entire procedure is dependent on the assumption that bias in the statistic S is of order n^{-1}. In practical applications, this may not be known with any certainty.

Secondly, variance estimates (5.3.3) and (5.3.3a) are in fact estimates of estimation error only. Presumably, regression estimates $\hat{\sigma}^2(n)$ of statistical error could also be jackknifed. The results would however be dubious since the assumption of a bias of order n^{-1} would be even more uncertain in the case $\hat{\sigma}^2(n)$ than $S(n)$.

5.4. Bootstrap

The bootstrap (Efron, [5]) is a procedure which makes use of data resampling. Application of the technique to regression problems is discussed by Freedman and Peters [7].

Consider the model:

$$Y = Xb + e, \tag{5.4.1}$$

where X is a given design matrix and e is a random vector with mean zero and covariance matrix V.

As in previous sections, let $\hat{b}$ denote the regression estimate of b. Then let:

$$\hat{e} = Y - X\hat{b}, \tag{5.4.2}$$

and

$$\xi = V^{-1/2}\hat{e}, \tag{5.4.3}$$

where the meaning of $V^{-1/2}$ is the conventional one for a positive definite matrix V.

Note that the components of ξ are independent, identically distributed (i.i.d). Let $F(.)$ denote the empirical distribution function obtained by assigning equal masses to them. It is now possible to generate pseudo-data sets:

$$Y^{(i)} = X\hat{b} + e^{(i)}, \ i = 1, 2, \ldots \tag{5.4.4}$$

where

$$e^{(i)} = V^{1/2}\xi^{(i)},$$

and $\{\xi^{(i)}\}$ is a random sample drawn from $F(.)$. Each set of pseudo-data leads to a new estimate $\hat{b}^{(i)}$ of b.

Let X^*, Y^* have the same meaning as in earlier sections. Then each estimate $\hat{b}^{(i)}$ leads to an estimate $\hat{Y}^{*(i)}$ of Y^* where

$$\hat{Y}^{*(i)} = X^*\hat{b}^{(i)}, \ i = 1, 2, \ldots \tag{5.4.5}$$

The collection $\{\hat{Y}^{*(i)}\}$ provides an empirical distribution of the random variable

$$\hat{Y}^* = X^*\hat{b}. \tag{5.4.6}$$

This distribution may be used to study the mean, variance, non-normality, confidence limits, etc. of (5.4.6). Note that:

$$\hat{Y}^{*(i)} - X^*b = X^*[\hat{b}^{(i)} - b], \tag{5.4.7}$$

which contains only estimation error. More pertinent to forecasting is a collection of forecasts of $\hat{Y}^*$ which contains statistical error also.

This is obtained by replacing (5.4.5) and (5.4.6) by:

$$\hat{Y}^{*(i)} = X^*\hat{b}^{(i)} + e^{*(i)}, \tag{5.4.5a}$$

$$\hat{Y}^* = X^*\hat{b} + e^*, \tag{5.4.6a}$$

where

$$e^* = W^{1/2}\xi^*, \tag{5.4.8}$$

$$e^{*(i)} = W^{1/2}\xi^{*(i)}, \tag{5.4.9}$$

for a known matrix W, the components of ξ, ξ^* are i.i.d., and $\{\xi^{(i)}, \xi^{(i)*}\}$ is a random sample drawn from $F(.)$, i.e., a particular $\xi^{(i)}$ and $\xi^{(i)*}$ are stochastically independent.

In this case (5.4.7) is replaced by:

$$\hat{Y}^{*(i)} - X^*b = X^*[\hat{b}^{(i)} - b] + e^{*(i)}, \tag{5.4.7a}$$

which includes both estimation and statistical error.

Freedman and Peters [7] (p. 99) deal with the case in which V is unknown and provide an iterative scheme for its estimation simultaneously with the generation of pseudo-data.

It is to be emphasized that the whole procedure assumes the validity of the basic model (5.4.1). If the model is invalid, estimates of second moments will probably be enlarged but not necessarily in the correct way.

For example, if prediction bias is present in model (5.4.1), it will be absorbed into $\hat{e}$ of (5.4.2) and hence ξ of (5.4.3). The components of ξ will then have non-zero mean and will not in general be identically distributed as assumed in the generation of pseudo-data (5.4.4).

5.5. Comparison of the estimation procedures

The advantages and disadvantages of the three estimation procedures considered in Sections 5.2 to 5.4 are summarized by Ashe [1] (p. S112) as follows:

Parametric estimation
- small number of calculations
- estimation error and statistical error available
- accurate if the parametric assumptions are correct

Jackknife:
- influence of individual data points on the estimate is available
- only estimation error is available
- estimate of loss reserve possible has reduced bias

Bootstrap:
- non-parametric
- estimation error and statistical error available
- distribution of loss reserve given

6. ROBUSTNESS

6.1. Influence function

The concept of an influence curve was introduced by Hampel [9]. It is discussed by Mosteller and Tukey [16] (pp. 351–356). A generalization to an influence function, a multi-dimensional version of the influence curve, is discussed by Rey [18] (pp. 15, 16).

The influence function of data points $y_1, \ldots, y_n$ on statistic $S(y_1, \ldots, y_n)$ is defined as the vector,

$$I(y_1, \ldots, y_n) = \frac{\partial S}{\partial \mathbf{y}} (y_1, \ldots, y_n), \tag{6.1.1}$$

with $\mathbf{y}$ denoting the vector $(y_1, \ldots, y_n)$. It indicates the influence on S of small variations in the data points.

A single component $\partial S/\partial y_i$ of (6.1.1), plotted as a function of y_i, with $y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n$ fixed at their observed values, provides the influence curve of y_i.

In the context of loss reserving by regression methods $S(y_1, \ldots, y_n)$ may be taken as the forecast (5.1.1):

$$\hat{R} = 1^T\hat{Y}^* = 1^T X^* \hat{b}_A, \tag{6.1.2}$$

where

$$\hat{b}_A = \hat{b}_A(Y_1, \ldots, Y_n)$$

is the regression estimate of b_A as in (3.2.7) and is a function of the data vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$.

Since $I_i(.)$ measures the effect of small variations of y_i on S, and the jackknife pseudo-estimate $P_i(n)$ measures the effect of removing y_i from the data, the two are related, as foreshadowed in Section 5.3. As suggested there, the pseudo-values perhaps serve as some kind of proxy for the influence function.

6.2. Robust regression

Regression need not be carried out by means of least squares, weighted or unweighted. Indeed, the importance of least squares regression derives, through the Gauss-Markov theorem (Graybill, [8]), from the oft-made assumption that random error terms in the data are normally distributed. When this assumption does not hold, least squares regression may not be appropriate.

There is no doubt that most classes of insurance involve long tailed claim size distributions. The basic data of any claims analysis, such as claim payments subdivided by year of occurrence and year of development, are therefore likely to incorporate error terms with long tailed distributions. Under weighted least squares regression, one or two rogue data points might well drag the entire regression away from the estimates which it would otherwise provide.

Robust regression encompasses procedures for fitting linear models whose properties are relatively insensitive to the distribution of these error terms. *Resistant* regression includes procedures leading to estimates which are not greatly distorted by extreme cases.

The latter of these two concepts is evidently related to the influence function. The smaller the influence function of a particular data point, the more resistant the regression to outlying values at that point.

Various methods have been used to reduce the influence function from that associated with least squares regression. For a summary, see Huber [11], [12]. An actuarial reference is Hogg [10]. Most of these methods can be viewed as fairly simple modifications of weighted least squares regression.

Consider the model,

$$Y = Xb + e, \tag{6.2.1}$$

where the notation is as in previous sections and, in particular, e is not necessarily normal although it is assumed to have zero mean. Under weighted least squares regression, b is estimated by that $\hat{b}$ which minimizes the weighted sum of squares (WSS):

$$\text{WSS} = (Y - X\hat{b})^T W (Y - X\hat{b}), \tag{6.2.2}$$

for some $n \times n$ matrix W which is independent of Y. Under resistant regression (6.2.2) is replaced by:

$$\text{WSS} = (Y - X\hat{b})^T W(\hat{Z})(Y - X\hat{b}), \tag{6.2.3}$$

where the weight matrix W depends on an estimate $\ddot{Z}$ of the vector of standardized residuals,

$$\hat{Z} = \text{diag}\,(\hat{\sigma}_1^{-1}, \ldots, \hat{\sigma}_n^{-1})(Y - X\hat{b}), \tag{6.2.4}$$

with $\hat{\sigma}_i^2$ an estimate of $V[Y_i]$.

Most commonly, the form of $W(\hat{Z})$ is:

$$\begin{aligned} w_{ij} &= h_i(\hat{Z}_i),\ j = i; \\ &= 0, \qquad j \neq i; \end{aligned} \tag{6.2.5}$$

for some function h_i which decreases as $\hat{Z}_i$ departs from zero. Thus, outlying observations, generating large values of $\hat{Z}_i$, are assigned little weight in WSS.

Typical choices of the attenuating function $h_i(.)$ are:

$$\begin{aligned} h_i(z) &= w_i, \qquad |z|\ ; \leqslant 2; \\ &= 4w_i/|z|^2,\ |z| \geqslant 2, \end{aligned} \tag{6.2.6}$$

where diag $(w_1, \ldots, w_n)$ is the weight matrix which would have been used for weighted least squares regression; or alternatively,

$$\begin{aligned} h_i(z) &= w_i z^{-1} \sin\,(2z/3),\ |z| \leqslant 3\pi/2; \\ &= 0, \qquad\qquad |z| \geqslant 3\pi/2; \end{aligned} \tag{6.2.7}$$

or again,

$$\begin{aligned} h_i(z) &= w_i\,[1 - (z/5)^2]^2,\ |z| \leqslant 5; \\ &= 0, \qquad\qquad |z| \geqslant 5. \end{aligned} \tag{6.2.8}$$

It is apparent that any system (6.2.3) in which the weight matrix $W(\hat{Z})$ depends on $\hat{b}$ renders WSS non-quadratic in $\hat{b}$. Then the solution b is nonlinear in the data Y. It will usually be necessary, therefore, for (6.2.3) to be minimized iteratively. At each iteration, the $\hat{\sigma}_i^2$ need to be recalculated on the basis of the residuals at the preceding iteration. Then $W(\hat{Z})$ can also be calculated on the basis of the same residuals, and (6.2.3) minimized with the new $W(\hat{Z})$ treated as independent of $\hat{b}$.

REFERENCES

[1] F. R. Ashe, "An Essay at Measuring the Variance of Estimates of Outstanding Claim Payments," *Astin Bulletin,* Vol. 16, 1986, p. S99.

[2] D. J. Bartholomew, "Errors of Prediction in Markov Chains," *Journal of the Royal Statistical Society,* Series B, Vol. 37, 1975, p. 444.

[3] J. R. Berquist and R. E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive Systematic Approach," *PCAS* LXIV, 1977, p. 123.

[4] L. Breiman and D. Freedman, "How Many Variables Should be Entered in a Regression Equation?" *Journal of the American Statistical Association,* Vol. 78, 1983, p. 131.

[5] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics,* Vol 7, 1979, p. 1.

[6] M. A. Efroymson, "Multiple Regression Analysis," *Mathematical Methods for Digital Computers,* Vol. 1 (A. Ralston and H. S. Wilf, eds), 1960, p. 191. John Wiley & Sons, New York.

[7] D. A. Freedman and S. C. Peters, "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association,* Vol. 79, 1984, p. 97.

[8] F. A. Graybill, *An Introduction to Linear Statistical Models,* Vol. 1, 1961, McGraw-Hill.

[9] F. E. Hampel, "The Influence Curve and its Role in Robust Regression," *Journal of the American Statistical Association,* Vol. 69, 1974, p. 383.

[10] R. V. Hogg, "Statistical Robustness: One View of its Use in Applications Today," *Actuarial Research Clearing House,* 1979.3 Issue, p. 9.

[11] P. J. Huber, "Robust Statistics: A Review," *Annals of Mathematical Statistics,* Vol. 43, 1972, p. 1041.

[12] P. J. Huber, "Robust Statistical Procedures," *Society of Industrial and Applied Mathematics,* 1977.

[13] E. Kremer, "IBNR-Claims and the Two-Way Model of ANOVA," *Scandinavian Actuarial Journal,* 1982, p. 47.

[14] C. L. Mallows, "Some Comments on C_p," *Technometrics,* Vol. 15, 1973, p. 661.

[15] A. J. Miller, "Selection of Subsets of Regression Variables," *Journal of the Royal Statistical Society, Series A,* Vol. 147, 1984, p. 389.

[16] F. Mosteller and J. W. Tukey, *Data Analysis and Regression,* Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

[17] M. H. Quenouille, "Notes on Bias in Estimation," *Biometrika,* Vol. 43, 1956, p. 353.

[18] W. J. J. Rey, "Robust Statistical Methods," *Lecture Notes in Mathematics,* Vol. 690, 1979, Springer-Verlag, Berlin.

[19] G. A. F. Seber, *Linear Regression Analysis,* John Wiley and Sons, Inc., New York, 1977.

[20] D. Skurnick, "A Survey of Loss Reserving Methods," *PCAS* LX, 1973, p. 16.

[21] E. Spjøtvoll, "Multiple Comparison of Regression Functions," *Annals of Mathematical Statistics,* Vol. 43, 1972, p. 1067.

[22] G. C. Taylor, "An Invariance Principle for the Analysis of Non-Life Insurance Claims," *Journal of the Institute of Actuaries,* Vol. 110, 1983, p. 205.

[23] G. C. Taylor, *Claims Reserving in Non-Life Insurance,* North-Holland, Amsterdam, 1986.

[24] G. C. Taylor and F. R. Ashe, "Second Moments of Estimates of Outstanding Claims," *Journal of Econometrics,* Vol. 23, 1983, p. 37.

A NOTE ON THE GAP BETWEEN TARGET AND EXPECTED UNDERWRITING PROFIT MARGINS

EMILIO C. VENEZIAN

Abstract

Profit margins experienced by insurance companies are, on average, considerably lower than the "target" margins used to compute the premiums. The difference has been attributed to a variety of factors, ranging from errors in actuarial projections, to regulatory delays, to regulatory and competitive pressures. This note examines the potential impact of the procedure used to "mark up" the projected cost per policy on the gap between two quantities, the intended or "target" margin and the expected value of the realized profit margin.

The analysis shows that the practice of dividing the expected loss cost by a "permissible loss ratio" computed by deducting the anticipated expenses and a profit provision from unity will produce an expected underwriting profit margin that is, on average, lower than that built into the rates.

1. INTRODUCTION

A stylized view of actuarial ratemaking involves a provision for profit. Mathematically, the provision is made by dividing the expected cost of servicing an insurance contract by a number which represents unity minus the "target" profit margin.[1] The results of this computation may be used directly in the market, as in situations in which rates are promulgated by a department or bureau, or may be merely estimates of the marginal cost of providing insurance which guide management in its pricing policies. In either event, the result is a key input into the pricing decision.

Over extended periods of time in most jurisdictions, the average underwriting profit margins achieved by the industry as a whole, or by individual firms, differ substantially from the targets ostensibly built into the rates. This "gap" has been

[1] In practice the numerator may include the costs of losses and loss adjustment services and the denominator reflects other anticipated expenses (such as commissions, administrative expenses, and premium taxes) as well as the target profit margin.

emphasized by various authors [2,3,4]. One view is that the gap exists because the target profit built into the rates is excessive [2]. In this view, the gap represents the difference between improperly regulated prices and prices that would hold in a competitive economic system. Others attribute the gap to difficulties in ratemaking and to "cutbacks and delays in implementing rate increases" [3]. Still others remark on the existence and importance of these gaps but don't provide a rationale for their existence [4].

This note analyzes the gap in terms of the stylized procedure described above. The emphasis is on the difference between the underwriting margin that is incorporated into the ratemaking formula, here called the "target" margin, and the margin that would be expected on a statistical basis from the direct use of the formula. The analysis shows that the procedure used in developing premiums from actuarial projections is responsible for at least part of the difference between the target and expected profit margins, even if the projections used in making rates are unbiased.

2. A STYLIZED MODEL OF RATEMAKING AND PROFIT DETERMINATION

For our purposes, a very simple stylized model of actuarial ratemaking is adequate. The simple model presented here would be applicable directly to state-mandated rates or to bureau rates with no deviations. Trivial extensions would be needed for situations in which uniform deviations from promulgated rates are permissible. The analysis would also be applicable to rates developed through management discretion as long as the actuarial projections of needed rates are a major determinant of the rates ultimately adopted by management. We view ratemaking as consisting of the following steps:

1. The forecast cost per policy, F, is developed from past data.
2. The target margin, T, is determined.
3. The price, P, is calculated from

$$P = \frac{F}{1 - T} . \tag{2.1}$$

Policies are sold at this price and will, eventually, prove to involve a cost per policy of C. The underwriting profit per policy during the period in question will be the difference between the revenue, P, and the cost, C. The underwriting profit margin during the period will, accordingly, be

$$m = \frac{P - C}{P} . \tag{2.2}$$

Combining these equations, the underwriting profit margin may be expressed as

$$m = 1 - \frac{C}{F}(1 - T). \tag{2.3}$$

Denoting the true expected value of the cost per policy as C_T, the observed cost per policy will be a random variable whose expected value is C_T. Accordingly, we write

$$C = C_T(1 + y), \tag{2.4}$$

where y is a random variable. By definition the expected value of y is zero.

From these equations it follows that the achieved underwriting margin can be expressed as

$$A = \mathrm{E}(m) = 1 - (1 - T)\,\mathrm{E}\left(\frac{C}{F}\right) \tag{2.5}$$

$$= 1 - (1 - T)\,\mathrm{E}\left(\frac{C_T(1 + y)}{F}\right). \tag{2.6}$$

If we were to view the value of F as being identical to C_T, then we could use equation 2.6. In view of the fact that the expected value of y is zero, we would then conclude that the achieved margin is the same as the target margin. This appears to be the origin of the conventional wisdom that the two are equal in the absence of effects such as competition or regulatory lags. The value of F is, however, a forecast rather than the true value of the cost per policy. It is, accordingly, a random variable whose value depends on the unobservable value of the true cost. Assuming that F is fixed is tantamount to assuming that the actual cost per policy will tend to cluster around the forecast rather than around its true expected value.

In order to recognize the effect of forecast errors, we denote the forecast cost per policy as:

$$F = C_T(1 + x), \tag{2.7}$$

where x is a random variable measuring the prediction error. Since the premium is always greater than zero we can guarantee that $x > -1$. The expected value of x will be zero if the estimators used in ratemaking are unbiased, but this is not assured. We assume that the values of x and y are independent unless the errors in the forecast affect actual experience.[2]

Recognizing the random elements in the forecast we must write

$$A = \mathrm{E}(m) = 1 - (1 - T)\,\mathrm{E}\left(\frac{1 + y}{1 + x}\right). \tag{2.8}$$

In view of the independence of x and y, this can be simplified to

$$= 1 - (1 - T)\,\mathrm{E}(1 + y)\,\mathrm{E}\left(\frac{1}{1 + x}\right), \tag{2.9}$$

and since the expected value of y is zero

$$= 1 - (1 - T)\,\mathrm{E}\left(\frac{1}{1 + x}\right). \tag{2.10}$$

This can be written in a more suggestive form as

$$A = 1 - \frac{(1 - T)}{\mathrm{E}(1 + x)} - (1 - T)\left[\mathrm{E}\left(\frac{1}{1 + x}\right) - \frac{1}{\mathrm{E}(1 + x)}\right]. \tag{2.11}$$

If the actuarial estimates are unbiased, as they strive to be, the expected value of $1 + x$ will be one. The first two terms on the right hand side will, accordingly, be equal to T. The quantity in brackets in the third term will always be positive due to the fact that the harmonic mean[3] of a positive variable (in this case, $1 + x$) is always less than the arithmetic mean [1]. Since $1 - T$ is also positive and the third term has a negative sign, it follows that in general the expected value of the underwriting margin will be less than the target margin.

3. AN EXAMPLE

A concrete example may serve to illustrate the relationship. Let us consider the situation when the logarithm of variable $1 + x$ is normally distributed with mean m and standard deviation s. This is realistic in that it corresponds to a

[2] Dependence can arise in a number of ways. Two deserve mention: the self-selection of the purchasers of insurance in response to changing effective prices and the easing or tightening claims settlement practices by management as profit margins change.

[3] The harmonic mean is defined as the reciprocal of $\mathrm{E}(1 + x)^{-1}$

situation in which rates are always positive and have lognormally distributed errors. In this case M_n, the expected value of $(1 + x)^n$, is given by

$$M_n = \exp(nm + \tfrac{1}{2}n^2s^2) \tag{3.1}$$

for any value of n.

Since the rate estimator is presumed to be unbiased, the expected value of $1 + x$ must be one. This requires that M_1 be one and, accordingly, that $m = -\tfrac{1}{2}s^2$.

Imposing this condition and obtaining M_{-1} from Equation 3.1, we find that Equation 2.11 may be written as

$$A = T - (1 - T)\,[\exp(s^2) - 1]\,. \tag{3.2}$$

It is worth noting that the factor multiplied by $1 - T$ is the variance of the relative error in the forecast. Thus, if the standard deviation of the relative forecast error is 10 percent, the bias in the underwriting profit margin will be very close to one percentage point. If the standard deviation of the relative forecast error were as high as 30 percent, the bias would be nine percentage points.

4. CONCLUSION

When premiums are set by marking up unbiased predictions of cost per policy by dividing them by one minus a target margin, it can be guaranteed that there will be a gap between the "target" and the "expected" underwriting profit margin. Mathematically, the gap is generated by the difference between the expected value of the reciprocal of a random variable and the reciprocal of the expected value of the variable. If projected loss ratios estimate the "true" expected loss ratios at current rates but are subject to random error, the same results apply when the premiums are derived by dividing the projected loss ratio at current rates by a "permissible loss ratio" that incorporates a target provision for underwriting profit.

This paper does not present estimates of the magnitude of the effect. Direct estimates of the difference could be calculated if there were records of the actual forecasts that could be compared with realized values. That data is not generally available. Even with that data, additional assumptions would be required in order to develop exact estimates. The example provided illustrates that the gap may be large. Extensive simulation based on distributions other than the log-normal and approximations based on publicly available data indicate that for workers' compensation insurance, the difference between intended and achieved margins attributable solely to the effect described in this paper is larger than one percentage point in most states, and may well reach five percentage points.

While errors of this magnitude are not uncommon, it must be remembered that this is a systematic, not a random effect. It is also important to keep in mind that the regulatory process and the rigors of competition may well result in estimators that are biased downward. In fact, if the estimates given above are correct, then for workers' compensation insurance, the effect of biased estimators may be three to four times larger than the statistical gap described in this note. Attempts to collect better data and refine the estimation procedure are in progress.

REFERENCES

[1] E. Beckenbach and R. Bellman, *An Introduction to Inequalities*, Random House, New York, NY, 1961, p. 52.

[2] W. Fairley, "Investment Income and Profit Margins in Property-Liability Insurance: Theory and Empirical Results," *Bell Journal of Economics*, Vol. 10, 1979, p. 192.

[3] Joint Industry Task Force, *Automobile Insurance Rates and Investment Income*, Insurance Information Institute, New York, NY, 1981, p. 15.

[4] I. H. Plotkin, "Separate Comments on the Investment Income Advisory Committee Report," *Report of the Advisory Committee to the NAIC Task Force on Profitability and Investment Income*, Vol. II, Allstate Insurance Group, Northbrook, IL, 1983, p. 1.

[5] E. C. Venezian, "Comments of the Exchange Between Hedges and Witt," *Journal of Risk and Insurance*, Vol. XLIX, 1982, p. 618.

SOME CONSIDERATIONS ON AUTOMOBILE RATING SYSTEMS UTILIZING INDIVIDUAL DRIVING RECORDS

LESTER B. DROPKIN

REPRINTED FROM VOLUME XLVI

1. INTRODUCTION

With the recent introduction of automobile rating systems which modify an otherwise applicable rate by utilizing some form of individual driving record, a number of questions presented themselves. On the one hand, it was felt that a mathematical description of a phenomenon—in this case risk distributions by number of accidents—is intrinsically of value and constitutes an advance. The first part of this paper is concerned with the presentation of such a description. A frequency distribution known as the negative binomial distribution is utilized in these first sections.

Of considerable and immediate importance is the question: What is the probability that an individual rated according to a given "driving record sub-classification" has been correctly classified? The answer to the question as phrased is actually an objective and, as such, is not specifically answered here. Rather, we have utilized a simple type of segregating system, based on the number of traffic violations only without regard to the type of violation involved.[1] In the concluding parts of this paper an analysis of this simple model is made and conclusions are drawn. As is there pointed out, this paper has as one of its prime intents, the introduction and utilization of certain approaches to the problem. While an extrapolation of some of these conclusions to the actual rating systems currently being introduced by the rating bureaus and others is made, this paper is by its nature preliminary. It is hoped that the near future will produce more extensive investigations.

2. THE RATIONALE OF USING THE NEGATIVE BINOMIAL DISTRIBUTION

Of those individuals who have no accidents during an experience period some will be persons with a high loss-causing propensity but have been "lucky",

[1] For a description of the California study which constitutes the basic data for this paper, see Harwayne [1].

some will be persons with a very low propensity and have seen their "expectations" realized, and conversely. All this we know (or assume). The attempt is made here to unravel some of these threads and to gain a means of approach whereby some of the probabilities involved may be set forth.

In discussions of the distributions of risks by number of accidents it has been traditional to base such discussions on the Poisson frequency function, $P(x)$. That is, if we let n be a random variable (equal to the number of accidents) we have assumed that the probability that $n = x$, where $x = 0, 1, \ldots$ is given by

$$P(x) = \text{probability that } n \text{ equals } x = (m^x e^{-m})/x!. \tag{2.1}$$

In dealing with a given body of experience, the parameter m is set equal to the observed mean because in the Poisson distribution $E(x) = m$.

A test of goodness of fit by use of the chi-square distribution will, however, often indicate a significant deviation. A much improved fit will often result by considering that n is distributed in accordance with the two parameter frequency function

$$N(x) = \text{probability that } n \text{ equals } x = \left(\frac{a}{1+a}\right)^r \binom{-r}{x} \left(\frac{-1}{1+a}\right)^x, \tag{2.2}$$

where $x = 0, 1, \ldots$

This frequency function is known as the *negative binomial distribution.*[2] For this function $E(x) = r/a$ and $\sigma^2 = (r/a)\,[(a+1)/a]$ as will be shown subsequently. In fitting observed data to equation (2.2), the observed mean and variance are set equal to r/a and $(r/a)\,[(a+1)/a]$ respectively, whence the parameters r and a can be determined by solving the two equations simultaneously. Upon solving we get that $r = m^2/(\sigma^2 - m)$ and $a = m/(\sigma^2 - m)$. In actually using $N(x)$ with a given body of data, it is usual to use the following expanded form in which the values are obtained when

$$\left(\frac{a}{1+a}\right)^r$$

[2] See Appendix B for a comparison of the fit achieved by the use of the negative binomial and by the Poisson. The Chi-square test on the Poisson and the very good fit of the negative binomial was called to my attention by F. Harwayne.

is multiplied by the terms of the sequence

$$\left\{1, \frac{r}{1+a}, \frac{r(r+1)}{2!\,(1+a)^2}, \frac{r(r+1)(r+2)}{3!\,(1+a)^3}, \ldots\right\}.$$

That is, the probability that $n = 0$ is

$$\left(\frac{a}{1+a}\right)^r;$$

and that $n = 2$ is

$$\left(\frac{a}{1+a}\right)^r \frac{r(r+1)}{2!\,(1+a)^2}.$$

The rationale of the applicability of $N(x)$ to distributions by number of accidents results from the following considerations. If we assume that the parameter m in equation (2.1) is itself a (continuous) random variable with the frequency function $T(m)$ then the probability that n takes on any given value x is

$$\int_0^\infty P(x)T(m)\,dm. \tag{2.3}$$

Without for the moment specifying the form of $T(m)$, the introduction of a variable m can be interpreted as a way of accounting for the variation of risk among the members of a given population. That is, it is assumed that

(a) the individual chances vary from one person to another but (for the given individual) remain constant throughout the experience period, and
(b) these initial propensities are distributed in the population in a simple curve, $T(m)$.

The negative binomial, $N(x)$ results from assigning to $T(m)$ the specific form

$$T(m) = \frac{a^r}{\Gamma(r)} m^{r-1} e^{-am} \qquad (a, r \text{ positive}) \tag{2.4}$$

which is a Pearson Type III. The Type III curve is selected because of its skew form and because it leads to conveniently simple equations for fitting. It is also possible if a frequency is expressible by a Type III curve to express the chance of a variation within a given limit by utilizing Pearson's *Tables of the Incomplete Gamma Function*. This enters into later considerations. The mathematics of these considerations is given in Appendix A.

3. THE EFFECT OF SEGREGATING BY DRIVING RECORD

As indicated in the Introduction, we have dealt here only with a simple segregation by traffic violation; i.e., we have used only the data appearing in the California Study.

While the average accident involvement generally increases with increasing number of violations (see F. Harwayne, op. cit.) it does appear that for the groups with 5, 6, 7, 8 and 9 or more violations, the mean accident frequencies have bĕcome relatively stable. (The respective means are .557, .508, .502, 545, and .656).

The fact that the negative binomial fits the data for the total group indicates that there is a real spread, that is, a distribution, of the probability of having an accident. From the construction of the negative binomial we have seen that this distribution is describable by a Type III curve.

Now it is clearly the function of a segregating system to split up the total heterogeneous group into homogeneous groups. The question is therefore raised as to whether or not, or to what degree, a segregating system based on traffic violations does split up the total group. If the system we are dealing with here accomplished this purpose totally, then the distributions by number of accidents of the individual groups should be describable by Poisson curves. Now if the variance of the separate groups were less than the Binomial variance,[3] then Poisson curves would indeed be indicated. However, Appendix C shows that this is not the case. In every instance, the variance is greater than the Binomial variance. This would seem to indicate that the desired segregation was not achieved.

We can, however, go further. Since a Poisson distribution is not indicated for the distributions by number of accidents, a negative binomial is indicated. But a negative binomial for the distribution by number of accidents is describable by a Type III curve. Now if we can picture these individual Type III curves, we can see in which groups, if any, the probability of having an accident is highly concentrated about the mean probability for that group. In other words, if we can determine what portion of the distribution is within stated deviations from the mean, then we can see how closely a given mean probability (of having an accident) approximates a constant probability and thus how closely the segregating system under consideration achieves its aim.

[3] The Binomial variance is equal to the product of the mean and the complement of the mean.

The required areas (or rather portion of total area) under the various Type III curves can be determined through a utilization of Pearson's *Tables of the Incomplete Gamma Function.* (See Appendix D for details.) Appendix E sets forth, by individual group, the portion of the distribution within stated deviations from a given mean probability of having an accident. The deviations utilized are plus and minus 20%, 30%, 40%, and 50%.

4. REVIEW, SUMMARY AND CONCLUSIONS

We have, in a certain sense, conceptually separated this paper into two parts. This was done in order to emphasize what to us seems to be the importance of the negative binomial distribution as a valuable instrument in its own right. It is our belief that this distribution can be an equally useful tool in attacking numerous other actuarial problems. It is also believed that many worthwhile results can flow from a utilization of the general approach illustrated by equation (2.3). This equation is typical of the general theory of processes random in time (stochastic processes) and we believe that this theory will come to be of particular value to the actuary.

It is also important to emphasize here that there are two distributions which enter into our considerations. On the one hand, there is the distribution of the probability of having an accident. On the other hand, there is the distribution of risks by number of accidents. If the first distribution is a constant, then the second is a Poisson. If the first is a Type III, then the second is a negative binomial. Since the two parameters of the negative binomial are also the two parameters of the component Type III we can use the sample mean and variance to determine them. From a knowledge of the values of these parameters we can determine the spread about the mean probability of having an accident. If there is little spread then the segregating system has performed its function. A review of the figures shown in Appendix E indicates that in no group was there a real concentration about the mean. Thus for the group with 1 violation only about 25% of the group can be expected to lie within plus or minus 20% of the mean, 62% can be expected to fall outside of an interval of plus or minus 30% of the mean, etc. Notice, too, that for the group having no violations, which represents 58.7% of the total number of individuals in the study, only a little over 25% of the group can be expected to lie in an interval of plus and minus 40% of the mean.

It is also very instructive to look at the question of overlapping. We see that about 25% of each of the groups having 1, 2, 3 or 4 violations can be expected to have a probability of having an accident greater than or equal to the mean probability for the succeeding group. As examples: The mean probability in group 3 is .354; the portion of group 2 having a probability of .356 (= 1.3 times the mean of group 2) or more is .25 (= 1 − .75). The mean probability in group 5 is .553; the portion of group 4 having a probability of .554 (= 1.3 times the mean of group 4) or more is .26 (= 1 − .74).

There is, in addition, considerable overlapping in the other direction. Thus, for example, the mean probability in Group 1 is .194; the portion of Group 2 having a probability of .192 (= .7 times the mean of group 2), or less is .36. For Group 2, therefore, about 60% of the group may be expected to have a probability of having an accident which is either less than the mean of the preceding group or greater than that of the following group. Similar figures obtain for other groups.

If, in asking these questions we were to think of an interval about the means of the preceding and following groups, the amount of overlapping would of course be greater.

Having now performed these calculations, what are our conclusions? We are, it would seem, to conclude that the segregating system here considered does not function to effectively separate the total into groups sufficiently homogeneous to merit modifications of the rate.

We may well expect, a priori, that a segregating system which is based on only certain violations rather than all violations, that introduces a weighting process for these violations and that includes accident record as well as violation record, will produce a separation into groups more homogeneous than we have seen here. We must, however, also note that the use of 2 years' experience instead of the 3 years which form the data for this study, will act to decrease whatever sharpness of separation the foregoing will presumably introduce.

While it is dangerous to extrapolate, it would appear from the results presented in this paper that two conclusions of general application may be drawn. These are that:

(a) after a certain point an increase in the number of violations does not contribute proportionately to an increase in the average number of accidents; and,
(b) the effect of segregating according to driving record is less effective than might be heretofore thought.

It is clear that the general area with which this paper is concerned is of current importance and is obviously a fertile field for many future papers. Presentations dealing with models more closely approximating the actual rating systems in use and with utilizations of the negative binomial distribution in other areas are earnestly to be desired.

REFERENCES

[1] Frank Harwayne, "Merit Rating in Private Passenger Automobile Liability Insurance and the California Driver Record Study," *PCAS* XLVI, 1959, p. 189.

APPENDIX A

Mathematics of the Negative Binomial

We display here the mathematics of the considerations set forth in the first part of the paper. By substituting in equation (2.3) the specific forms $P(x)$ and $T(m)$ given by equations (2.1) and (2.4), we derive therefrom the equation for $N(x)$ given by equation (2.2). Following this, we show that:

(a) $\sum_{x=0} N(x) = 1,$

(b) $E(x) = r/a,$

(c) $E(x^2) = \frac{r}{a}\left(\frac{a + r + 1}{a}\right)$ and that therefore,

(d) $\sigma^2 = E(x^2) - [E(x)]^2 = \frac{r}{a}\left(\frac{a + 1}{a}\right).$

Derivation of N(x)

From equation (2.1), $P(x) = (m^x e^{-m})/x!$ and from equation (2.4), $T(m) = (a^r m^{r-1} e^{-am})/\Gamma(r)$ we are to derive $N(x)$. We proceed as follows:

$$N(x) = \int_0^\infty \frac{m^x e^{-m}}{x!} \frac{a^r m^{r-1} e^{-am}}{\Gamma(r)} dm$$

$$= \int_0^\infty \frac{a^r}{x!\,\Gamma(r)} m^{(x+r-1)} e^{-m(1+a)} dm$$

$$= \frac{a^r}{x!\,\Gamma(r)} \frac{(x + r - 1)!}{(1 + a)^{x+r}} \qquad \text{[see Pierce \#493]}$$

$$= \left(\frac{a}{1 + a}\right)^r \frac{1}{(1 + a)^x} \frac{(x + r - 1)!}{x!\,\Gamma(r)}$$

Now, since the last factor in this equation can be transformed as follows:

$$\frac{(x + r - 1)!}{x!\,\Gamma(r)} = \frac{(r + x - 1)!}{x!(r - 1)!} = \frac{r(r + 1) \ldots (r + x - 1)}{x!}$$

$$= \frac{(-r)[-(r + 1)][-(r + 2)] \ldots [-(r + x - 1)](-1)^x}{x!}$$

$$= (-1)^x \binom{-r}{x},$$

we have that

$$N(x) = \left(\frac{a}{1 + a}\right)^r \frac{1}{(1 + a)^x} (-1)^x \binom{-r}{x},$$

which is equation (2.2):

$$N(x) = \left(\frac{a}{1 + a}\right)^r \binom{-r}{x} \left(\frac{-1}{1 + a}\right)^x.$$

From this it immediately follows that

$$\sum_{x=0} N(x) = \left(\frac{a}{1 + a}\right)^r \left(1 - \frac{1}{1 + a}\right)^{-r} = \left(\frac{a}{1 + a}\right)^r \left(\frac{a}{1 + a}\right)^{-r} = 1.$$

Derivation of $E(x)$, $E(x^2)$ and σ^2

By definition, $E(x) = \sum_{x=0} x\, N(x)$, whence

$$E(x) = \sum_{x=0} x\, N(x) = 0 + \sum_{x=1} x\, N(x)$$

$$= \sum_{x=1} \left(\frac{a}{1 + a}\right)^r \left(\frac{r}{1 + a}\right) \binom{-(r + 1)}{x - 1} \left(\frac{-1}{1 + a}\right)^{x-1}$$

$$= \left(\frac{a}{1 + a}\right)^r \left(\frac{r}{1 + a}\right) \left(1 - \frac{1}{1 + a}\right)^{-(r+1)}$$

$$= \left(\frac{a}{1 + a}\right)^r \left(\frac{r}{1 + a}\right) \left(\frac{a}{1 + a}\right)^{-(r+1)} = \frac{r}{1 + a} \left(\frac{a}{1 + a}\right)^{-1} = \frac{r}{a}.$$

Similarly, from $E(x^2) = \Sigma x^2 N(x)$ and $\sigma^2 = E(x^2) - [E(x)]^2$, we have

$$E(x^2) = \sum_{x=0} x^2 N(x) = 0 + \left(\frac{a}{1+a}\right)^r \left(\frac{r}{1+a}\right) + \sum_{x=2} x^2 N(x).$$

By writing $[x(x-1)+x]$ for x^2, we get

$$E(x^2) = \left(\frac{a}{1+a}\right)^r \left(\frac{r}{1+a}\right) + \sum_{x=2} x(x-1)N(x) + \sum_{x=2} x\, N(x).$$

But, since $\sum_{x=1} x\, N(x) = r/a$, it follows that

$$\sum_{x=2} x\, N(x) = r/a - \left(\frac{a}{1+a}\right)^r \left(\frac{r}{1+a}\right).$$

Accordingly,

$$E(x^2) = \frac{r}{a} + \sum_{x=2} x(x-1) \left(\frac{a}{1+a}\right)^r \binom{-r}{x} \left(\frac{-1}{1+a}\right)^x$$

$$= \frac{r}{a} + \sum_{x=2} \left(\frac{a}{1+a}\right)^r \frac{r(r+1)}{(1+a)^2} \binom{-(r+2)}{x-2} \left(\frac{-1}{1+a}\right)^{x-2}$$

$$= \frac{r}{a} + \left(\frac{a}{1+a}\right)^r \frac{r(r+1)}{(1+a)^2} \left(1 - \frac{1}{1+a}\right)^{-(r+2)}$$

$$= \frac{r}{a} + \frac{r(r+1)}{a^2} = \frac{r}{a}\left(\frac{a+r+1}{a}\right).$$

From this it immediately follows that

$$\sigma^2 = \frac{r}{a}\left(\frac{a+r+1)}{a}\right) - \left(\frac{r}{a}\right)^2$$

$$\sigma^2 = \frac{r}{a}\left(\frac{a+1}{a}\right).$$

APPENDIX B

COMPARISON OF FIT BY POISSON AND NEGATIVE BINOMIAL FOR TOTAL GROUP

	Observed Freq.		Theoretical Frequency			
			Negative Binomial		Poisson	
Number of Accidents	No.	%	No.	%	No.	%
0	81714	86.07	81726	86.086	80655	84.959
1	11306	11.91	11273	11.874	13147	13.848
2	1618	1.71	1647	1.735	1072	1.129
3	250	.26	245	.258	58	.061
4	40	.04	37	.039	3	.003
5 or more	7	.01	7	.008	—	—
	94935		94935		94935	

Mean = .163 $\sigma^2 = .193$ Binomial Variance = .136

For fitting the neg. binomial: $r = .8927$; $a = 5.472$; $\frac{a}{1 + a} = .8455$

For fitting the Poisson: $e^{-.163} = .84959$.

APPENDIX C

Group (Violations)	Mean	Variance	Binomial Variance[4]
0	.087	.096	.079
1	.194	.207	.156
2	.274	.299	.199
3	.354	.395	.229
4	.426	.501	.245
5 or more	.553	.610	.247
Total	.163	.193	.136

[4] Equals the product of the mean and its complement.

APPENDIX D

The determination of the ratios of $\int_0^t T(m)dm$ to $\int_0^\infty T(m)dm$ with $T(m)$ as defined in equation (2.4), is accomplished by utilizing the *Tables of the Incomplete Gamma Function* prepared under the direction of Karl Pearson in 1922.

The complete gamma function $\Gamma(p + 1)$ is defined as $\int_0^\infty e^{-x}x^p dx$ while the incomplete gamma function $\Gamma_x(p + 1)$ is defined as $\int_0^x e^{-x}x^p dx$. If $I(x,p)$ denotes the ratio of the incomplete to the complete gamma function, then $I(x,p)$ gives the portion of the curve to the left of x. However, $I(x,p)$ has not been published. Instead, a variable $u = x/\sqrt{p + 1}$ is used and it is these equivalent tables of $I(u,p)$ which were prepared by Pearson. That is

$$I(u,p) = \frac{\displaystyle\int_0^{u\sqrt{p+1}} v^p e^{-v} dv}{\displaystyle\int_0^\infty v^p e^{-v} dv}.$$

In order to use the tabulated values of $I(u,p)$ it is necessary to proceed as follows:

We first recall that $\int_0^\infty T(m)dm = 1$ so that we are looking for values of $\int_0^t T(m)dm$ and recall that

$$\int_0^t T(m)dm = \int_0^t \frac{a^r m^{r-1} e^{-am}}{\Gamma(r)} dm.$$

Now let $v = am$ so that $m = a^{-1}v$ and $dm = a^{-1}dv$. The integral thus becomes

$$\int_0^{at} \frac{v^{r-1} e^{-v}}{\Gamma(r)} dv.$$

Now let $p = r\text{-}1$; we then have

$$\int_0^{at} \frac{v^p e^{-v}}{\Gamma(p + 1)} dv.$$

But this is precisely $I(u,p)$ with $at = u\sqrt{p + 1}$; from this we get that

$$u = at/\sqrt{p + 1} = at/\sqrt{r}.$$

Since we know a and r from the data for a given t we have the values of u and p with which to enter the tables. One could for example determine values with $t =$ mean, mean $\pm$ 5%, mean $\pm$ 10%, mean $\pm$ 20%, etc.

APPENDIX E

When the procedures indicated in Appendix D are carried out, for values of t = 50%, 70%, 80%, 120%, 130%, 140% and 150% of the mean, separately for each individual group, the following results are obtained:

PORTION OF CURVE WITHIN INTERVAL SHOWN FOR GROUP SHOWN

	Group (Violations)					
Interval	0	1	2	3	4	5 or more
0 to $.5\bar{x}$	.45	.18	.20	.19	.23	.10
0 to $.6\bar{x}$	.50	.25	.28	.27	.31	.17
0 to $.7\bar{x}$	.54	.32	.36	.35	.39	.26
0 to $.8\bar{x}$	.59	.40	.43	.43	.46	.36
0 to $1.2\bar{x}$	.71	.65	.70	.70	.70	.72
0 to $1.3\bar{x}$	.73	.70	.75	.75	.74	.78
0 to $1.4\bar{x}$	.76	.74	.79	.79	.78	.83
0 to $1.5\bar{x}$	.78	.78	.83	.83	.81	.88
$.5\bar{x}$ to $1.5\bar{x}$	.33	.60	.63	.64	.58	.78
$.6\bar{x}$ to $1.4\bar{x}$	.26	.49	.51	.52	.47	.66
$.7\bar{x}$ to $1.3\bar{x}$	.19	.38	.39	.40	.35	.52
$.8\bar{x}$ to $1.2\bar{x}$	.12	.25	.27	.27	.24	.36

DISCUSSION BY ROBERT A. BAILEY
REPRINTED FROM VOLUME XLVII

As Mr. R. E. Beard, secretary and editor of ASTIN, said,[1]

> "The literature in the English language relating to analytical expressions of the risks involved in general insurance is scanty and largely limited to papers presented to International Congresses of Actuaries and the *Proceedings of the Casualty Actuarial Society*. There are, however, a number of contributions to the subject in various other languages, scattered over various journals, mainly, insurance publications of European countries, e.g. *Skandinavisk Aktuarietidskrift* and a few books."

The C.A.S. can rightfully be proud of its contributions in this field which have been ably enhanced by Mr. Dropkin's treatment of the negative binomial distribution.

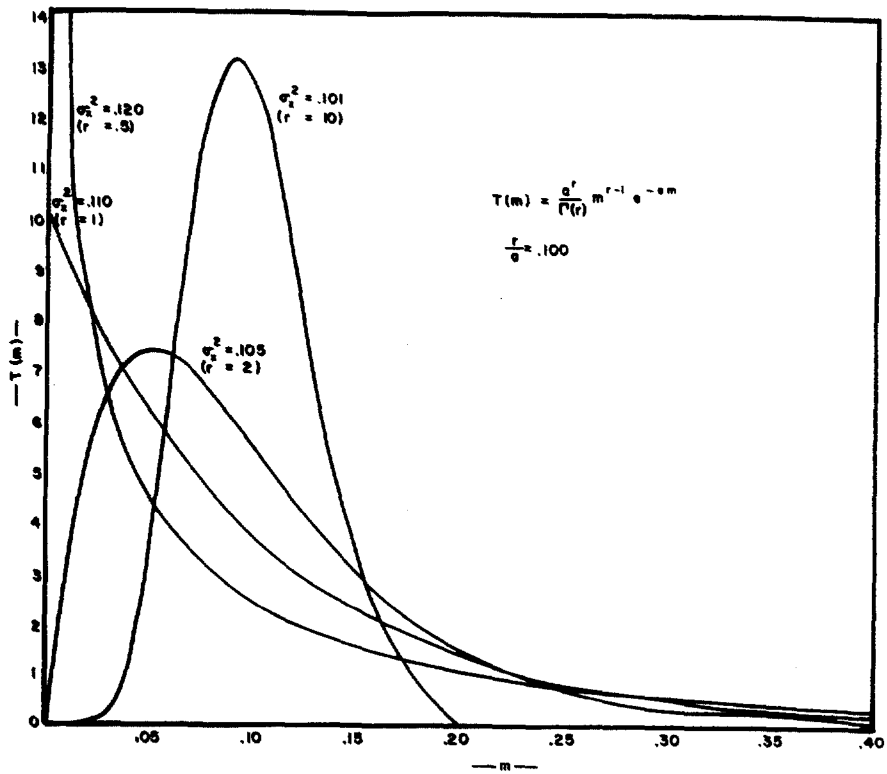
The analytical expression of risk distributions provides a valuable insight into many practical problems. One of the important results of Mr. Dropkin's paper is a realization of the large amount of variation among individual risks. Automobile risks even within a single class or merit rating group are far from being all alike. In order to help visualize this variation, there are shown in Figure 1 the graphs of the distribution of risks which Mr. Dropkin shows to be inherent in the negative binomial distribution. Four graphs are shown, all for an average accident frequency $\frac{r}{a} = .100$, and with variances of the accident frequency (not the variances of m, the inherent hazard) of $.120 (r = \frac{1}{2})$, $.110 (r = 1)$, $.105 (r = 2)$ and $.101 (r = 10)$.

One of the many practical applications to which Mr. Dropkin's development can be applied is the calculation of the discount for n accident-free years. This application was suggested to the writer by Mr. Dropkin's paper because it provided a means of deriving mathematically what had been derived empirically in the paper presented at the same time as Mr. Dropkin's, "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car," since the discount from the overall average rate for n accident-free years is equal to the "credibility" as defined in the paper just cited.

The chance that any individual risk with inherent hazard (m) will be accident-free for 1 year is e^{-m} where e^{-m} is the value of the Poisson distribution

[1] *Transactions of the XVth International Congress of Actuaries*, Volume II, 1957, p. 230.

FIGURE 1



$P(x) = \dfrac{m^x e^{-m}}{x!}$ when $x = 0$. Mr. Dropkin shows that the total distribution of individual risks can be described by the distribution

$$T(m) = \frac{a^r}{\Gamma(r)} m^{r-1} e^{-am} .$$

Therefore, the distribution of risks with 1 or more accident-free years is

$$T_1(m) = \frac{T(m)e^{-m}}{\int_0^\infty T(m)e^{-m}dm} = \left(\frac{a+1}{a}\right)^r T(m)e^{-m} .$$

Likewise the distribution of risks with 2 or more accident-free years is

$$T_2(m) = \left(\frac{a+2}{a}\right)^r T(m)e^{-2m} .$$

This provides us a means of immediately calculating the expected claim frequency of claim-free risks. Mr. Dropkin shows that the claim frequency for all risks $= E(x)$

$$= \sum_{x=0}^{\infty} x \int_0^{\infty} \frac{m^x e^{-m}}{x!} \frac{a^r m^{r-1} e^{-ma} dm}{\Gamma(r)}$$

$$= \frac{r}{a} .$$

Therefore, the claim frequency for risks with 1 or more accident-free years

$$= \sum_{x=0}^{\infty} x \int_0^{\infty} \frac{m^x e^{-m}}{x!} \frac{(a+1)^r m^{r-1} e^{-m(a+1)} dm}{\Gamma(r)}$$

$$= \frac{r}{a+1} .$$

Similarly, the expected claim frequency for risks with 2 or more accident-free years is $\frac{r}{a+2}$, and for 3 or more accident-free years is $\frac{r}{a+3}$, and so on.

Therefore, the expected claim frequency for risks accident-free for n or more years relative to the expected claim frequency for all risks, assuming that the inherent hazard (m) for each individual risk remains unchanged from one year to the next, is $\frac{a}{a+n}$ and the corresponding discount from the average rate is $\frac{n}{a+n}$. This is the same as saying that these risks are $\frac{n}{a+n}$ better than average.

The expression $\frac{n}{a+n}$ is equal to the "credibility" of risks accident-free for n or more years, as defined in the paper cited above, and it is the same result obtained independently by Dr. F. Bichsel, in a paper entitled "Une méthode pour calculer une ristorne adéquate pour années sans sinistres" (A method of calculating an adequate no-claim bonus for years without accidents) presented at the ASTIN Colloquy in La Baule, France, in June, 1959. Furthermore, if this expression for the credibility of the experience of an individual risk for n years,

$$Z = \frac{n}{a + n},$$

is multiplied in the numerator and denominator by the premium for one car year, it becomes

$$Z = \frac{P}{P + K},$$

where P is the premium during the experience period and where K is a constant which equals the parameter a multiplied by the premium for one car year. This is the credibility formula derived by Mr. A. W. Whitney in "The Theory of Experience Rating," *PCAS*, Vol. IV, and used ever since in almost all experience rating plans.

Another application which Mr. Dropkin's development suggested is a comparison of the variation of hazard among licensed drivers and among licensed automobiles. In Appendix B, Mr. Dropkin fits the negative binomial to the total distribution of California drivers and obtains $r = .8927$. From the graphs shown in Figure 1 and also from an analysis of the formula for $T(m)$ it can be seen that when $0 < r \leq 1$, $T(m)$ is a "J" shaped curve with a maximum height at $m = 0$. ($T(m)$, it should be remembered, is the distribution of the inherent hazard of the individual drivers and is to be distinguished from $N(x)$, the distribution of the resulting accidents.) It is reasonable that the California data should be described by a "J" shaped curve since some drivers licensed in California do not drive in California for a number of reasons, such as they do not have a car or they live outside the state. Since such licensed drivers will have an inherent hazard $m = 0$, a "J" shaped curve is a reasonable distribution of hazard for licensed drivers. On the other hand, however, the distribution of hazard for licensed automobiles should not be a "J" shaped curve, since practically no automobiles have a hazard $m = 0$ and therefore for the distribution of hazard for licensed automobiles, r should be greater than 1.

This proposition can be tested by using the Canadian merit rating experience for insured automobiles. By setting the one-year credibility for Class 1 cars of

$.055^2$ equal to the expression derived above for the one-year credibility, $\frac{1}{a+1}$, we obtain $a = 17.2$. Since the average frequency for Class 1 $= .087 = \frac{r}{a}$, we obtain $r = 1.50$ which is greater than 1 as we would expect. From this we can draw the conclusion that there is more variation of hazard among drivers than among cars.

There are undoubtedly many other applications which can be made of Mr. Dropkin's work and we are fortunate to have a development of the negative binomial distribution in the *Proceedings,* especially at this time when merit rating is of such great concern. We are entering a time of great competitive effort in the search for more accurate classification systems, not only in private passenger automobile insurance but in other lines as well, as Mr. Pruitt pointed out so forcefully last November in his presidential address, "St. Vitus's Dance." The negative binomial distribution, which has also been called the "accident proneness" distribution, provides a valuable tool for that search.

[2] R. A. Bailey and L. J. Simon, "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car," *PCAS* XLVI, Table 4, p. 163.

DISCUSSION BY SHOLOM FELDBLUM

Actuaries have generally used the Poisson distribution to model accident frequencies for "repeatable" risks—that is, where more than one accident is possible in an exposure period. The reasons for this are both theoretical and practical. Theoretically, if the following conditions are true, then the Poisson distribution is indicated:

1. The probability of exactly one accident in an infinitesimal unit of time dt is approximately equal to $k\ dt$, where k is a constant, during any interval of time in the exposure period.
2. The probability of more than one accident in an infinitesimal unit of time is negligible compared to the probability of exactly one accident.
3. The distribution is "memoryless"; that is, the numbers of accidents in distinct intervals of time are independent.

Practically, the Poisson distribution is mathematically convenient in numerous ways:

1. The mean and variance of the Poisson distribution are equal, so the variance may be estimated along with the mean from a simple averaging of raw results.
2. Since the mean and variance are equal, their ratio is unity, a known constant. This makes "classical pure premium credibility" easier to calculate, as discussed by Mayerson, Jones, and Bowers [1].
3. The Poisson distribution is conjugate to the gamma, a distribution both convenient and realistic for modeling the mean accident frequency among individuals in a population. This makes Bayesian estimation of future mean accident frequency distributions particularly convenient.
4. The Poisson claim frequency distribution can be combined with a claim size distribution to form a "compound Poisson" aggregate claim distribution. The compound Poisson distribution has advantages over other compound distributions. For instance, if the claim size distribution is discrete (or can be realistically modeled by a discrete distribution), the aggregate claim distribution can be determined by a recursive procedure, which facilitates the mathematics of determining this distribution [2].

Lester Dropkin's paper complicates this simplified Poisson world [3]. He points out that if the accident frequency is Poisson distributed for each individual in a population, but the mean accident frequencies vary by individual, then the accident frequency for the population as a whole is no longer Poisson distributed.

In particular, if the mean accident frequencies among individuals are gamma distributed, and the accident frequency for each individual is Poisson distributed, then the accident frequency for the population as a whole has a negative binomial distribution.

For the Poisson frequency, the mean and variance are equal. For the negative binomial, the variance is always greater than the mean. Using Dropkin's notation, the mean of the negative binomial is r/a, while the variance is $r\,(a + 1)/a^2$. (In this notation, r is the scale parameter and a is the shape parameter of the underlying gamma distribution.)

This analysis shows that the wider the dispersion of mean accident frequencies among individuals, the greater the variance of the total population accident frequency. For instance, suppose that the mean accident frequency for the population as a whole is 1; that is, $r = a$. For the underlying gamma distribution, the mean is a/r and the variance is a/r^2; thus, the ratio of the variance to the square of the mean is $1/a$. That is, as a decreases, the values are more widely dispersed relative to the mean, and as a increases, the values are more closely situated to each other. By examining the variance of the negative binomial distribution, we note that as a decreases, the variance of the population accident frequency increases, and vice versa as a increases.

Surprisingly, this result is not generally true. In life insurance, the accident frequency is generally modeled as a Bernoulli random variable, since at most one claim is possible per individual per exposure period. The mean death rate of the population, that is, the parameter of the Bernoulli random variable, may be determined from actual data by dividing the number who die at a given age by the number exposed in the population at that age. For example, if there are 1,000 individuals at age 50 in the population, and 20 individuals die at age 50, then the mean death rate at age 50 is 2%.[1]

Using the actual data, we hypothesize that the mean death rate is 2%. Assuming also that the death rate is 2% for each individual, the variance of the death rate for each individual is (0.02) (0.98) = 0.0196. The variance of the estimate of the mean death rate is (0.02) (0.98)/1000 = 0.0000196.

After reading Dropkin's analysis, one may question this: since the individual death rates vary about 2%, and only average to 2% for the population as a whole, should not the population variance differ from 0.0196? Should it not be

[1] To be exact, we should assume that there is no migration; i.e., there are no new entrants or withdrawals at age 50. Thus, there were 1,000 individuals who attained age 50; 20 of these died during the course of the year; and 980 individuals attained age 51.

similar to the Poisson case, where if the population mean accident rate is 2%, but the individual mean accident rates vary about 2%, the population variance is greater than 2%?

The answer is no, as can be seen by a simple example, as well as by a more formal mathematical proof. Suppose there are two individuals, with a mean population death rate of 50%. Assume two cases:

1. Each individual has a death rate of 50%.
2. One individual has a death rate of 75%, and the other individual has a death rate of 25%.

For each case we determine the first two moments for each individual, the moments of the "mixture" distribution, and the variance of the "mixture" distribution.

1. Each individual has a death rate of 50%. For each individual, both the first and second moments are 0.50, and so the first and second moments of the mixture distribution are also 0.50. Therefore the variance of the mixture distribution is 0.50 − (0.50) (0.50) = 0.25.
2. One individual has a death rate of 75%, and the other individual has a death rate of 25%. For the first individual, the first and second moments are 0.75, while for the second individual, they are 0.25. Therefore, the first and second moments of the mixture distribution are 0.50, and the variance is 0.25.

The general proof follows the same reasoning. Suppose each individual has a mean death rate of p_i, and over the population as a whole these average to m. Then the second moments for each individual are also p_i, and over the population as a whole these also average to m. Therefore, the variance of the population mean death rate is $m\ (1 - m)$.

The Bernoulli distribution allows only one occurrence, while the Poisson distribution has no limit on the number of occurrences. What if the number of possible occurrences is finite but is greater than one, such as with the binomial distribution?[2]

[2] The following result for the binomial distribution was shown to me by Dr. Rodney Kreps, an actuary at Fireman's Fund Insurance Companies.

Theorem: Suppose the accident frequency is modeled by a binomial distribution with parameters p_i and n, i.e.,

$$f(x) = \binom{n}{x} p_i^x (1 - p_i)^{n-x}$$

Further, suppose n is fixed for all individuals in the population, but p_i varies according to a p.d.f. $g(p)$, which has mean m and variance s^2. For each individual, the mean is $n\,p_i$, and the variance is $n\,p_i\,(1 - p_i)$. For the population as a whole, the mean is $n\,m$, and the variance is $n\,m\,(1 - m) + s^2\,n\,(n - 1)$.

Proof: For each individual, the mean is $n\,p_i$, and the second moment is $n\,p_i - n\,p_i^2 + n^2\,p_i^2$.

Therefore, the mean for the population is

$$\int_0^1 (n\,p)\,g(p)\,dp = n\,m.$$

The second moment for the population is

$$\int_0^1 (n\,p - n\,p^2 + n^2\,p^2)\,g(p)\,dp$$

$$= n\,m - n\,SM + n^2\,SM \text{ (where } SM \text{ is the second moment of } g(p))$$
$$= n\,m + n\,(n - 1)\,(SM - m^2) - n\,(n - 1)\,m^2$$

Subtracting the square of the mean, we get

$$= n\,m\,(1 - m) + n\,(n - 1)\,s^2, \text{ which is the desired result.}$$

Thus, the more that the number of possible occurrences for each individual (n) increases, the more the variance for the population as a whole depends upon the variance of $g(p)$.

A useful application of this result is in Bayesian estimation. Generally, in performing a Bayesian estimation, the accident frequency is chosen as Poisson or binomial, and the prior distribution as gamma or beta. However, the problem of selecting the parameters of the prior distribution can be serious, and the choice of these parameters will influence the resultant posterior distribution [4].

The above result provides a method of selecting parameters. Suppose the binomial distribution is chosen for accident frequency, with a given n. Then if the population mean is u, the mean of the prior distribution of the p_i is $u/n = m$. Similarly, if the variance of observed results is VAR, the variance of the prior distribution of the p_i is

$$s^2 = (VAR - n\, m\, (1 - m)) \,/\, (n\, (n - 1)).$$

Given these values of m and s^2, the two parameters of the prior beta distribution are easily determined.

This discussion may help clarify two apparently misleading items in Dropkin's paper. First, Dropkin's criterion for choosing whether to model accident frequency by a Poisson or negative binomial distribution is the observed relation of the population variance to the population mean. If the population variance is approximately equal to the binomial variance, i.e., $p\ (1 - p)$, where p is the population mean, then use a Poisson distribution; if it is significantly larger, then use a negative binomial distribution.

Presumably, this criterion should be, "If the population variance is approximately equal to the Poisson variance, i.e., p, where p is also the population mean, then use a Poisson distribution . . ." Dropkin's statement at first seems logical if he is referring to the probability of having one or more accidents, rather than to the number of accidents per exposure unit. But then the accident frequency is a Bernoulli distribution, and the population variance will be independent of the underlying distribution of mean accident frequencies.

Second, Dropkin implies that the only choices for modeling the accident frequency are the Poisson and the negative binomial. He shows that his data has a variance and mean incompatible with the Poisson distribution, and he concludes:

> We can, however, go further. Since a Poisson distribution is not indicated for the distributions by number of accidents, a negative binomial is indicated [3].

This is hardly so. His actual data only indicates that the Poisson distribution does not provide a perfect fit. It in no way indicates that a negative binomial distribution is better than other two-parameter distributions. The negative binomial distribution is only "indicated" if one assumes that each individual has a Poisson accident frequency and the mean accident frequencies among individuals are gamma distributed. One may test this by calculating the third moment of the observations and comparing it to the hypothetical third moment of the negative binomial distribution; unfortunately, Dropkin does not do this.[3]

[3] George Phillips, an actuary with the Transamerica Corporation, has recommended to me that use of other statistical methods, such as percentile matching, may give better results than examination of the third moment.

Other users of Dropkin's results have adopted this reasoning, such as Mayerson, et al., in "On the Credibility of the Pure Premium" [1] (though since the authors' purposes there are heuristic and not intended for practical applications, one can hardly fault them). They take the first two moments from Dropkin's accident frequency distribution, assume that it can be modeled by a negative binomial distribution, and calculate the third moment. But until one compares the derived third moment with the observed third moment, there is no evidence that the negative binomial provides an appropriate model. (I must reiterate, though, that the purpose of this paper is only to show how to apply a theory, not to provide firm credibility tables, and for such heuristic purposes, the assumptions are entirely plausible.)

To sum up: the Poisson distribution is a theoretically appealing model for accident frequencies for each individual. The accident frequency distribution for the population as a whole will depend upon the distribution of mean accident frequencies among the individuals in the population. The negative binomial distribution for the population accident frequency is indicated only if the individual mean accident frequencies are gamma distributed. The form of the mean accident frequency distribution may depend upon the line of insurance, class of risk, and so forth; in any case, there is no easy way to test it. Rather, one may test the first three (or more) moments of the observed results. In Dropkin's case, the first two moments provide the parameters of the negative binomial distribution as well as of the underlying gamma distribution. The observed third moment would then test whether the negative binomial provides an appropriate model. If not, a different two parameter population accident frequency distribution may be assumed. If the individual accident frequencies are Poisson distributed, this implies an underlying distribution of mean accident frequencies among individuals that is not gamma. Once more, the observed third moment can test whether this population accident frequency model is appropriate.

Of course, the more complex the distribution chosen, the better it may agree with observed results, but the less mathematically tractable it may be—and a mathematically intractable model is hardly useful. A great advantage of the Poisson distribution is its simplicity; the negative binomial distribution is also quite versatile. Nevertheless, there is a need to test the hypothetical models, to strike a balance between simplicity and accuracy.

REFERENCES

[1] A. L. Mayerson, D. A. Jones, and N. L. Bowers, Jr., "On the Credibility of the Pure Premium," *PCAS* LV, 1968, p. 175.

[2] Society of Actuaries, "Risk Theory," Part V Examination Study Note.

[3] Lester B. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* XLVI, 1959, p. 165.

[4] Dick London, *Graduation: The Revision of Estimates*, Winsted and Abington, Connecticut, ACTEX Publications, 1985, ch. 5.

DISCUSSION BY STEPHEN W. PHILBRICK

Mr. Dropkin's paper consists of two parts. The first is a discussion of the "importance of the negative binomial distribution as a valuable instrument in its own right." Second, this tool is used to comment on the use of the number of traffic violations to "split up the total heterogeneous group into homogeneous groups."

The author succeeds admirably in his first endeavor. A concise explanation of the rationale for the use of the negative binomial distribution is given. The arguments are intuitively appealing, since the choice of a Poisson distribution for an individual risk is desirable, and the notion that the parameters for the individuals vary from person to person is certainly more reasonable than the assumption that all drivers have identical accident propensities. The author is also to be commended for the algorithm for the calculation of the probabilities of $N(x)$, which is much more convenient than evaluating the traditional formula.

The second section contains some problems. The author concludes, "the fact that the negative binomial fits the data for the total group indicates that there is a real spread, that is, a distribution, of the probability of having an accident." Unfortunately, this conclusion cannot be supported by this argument. To demonstrate this, I randomly sampled from a Poisson distribution and attempted to fit both a Poisson and a negative binomial to the sample data. A Poisson distribution would have absolutely no "spread" of the parameter since the parameter is a fixed constant. The distribution of the parameter should not be confused with the distribution of the number of accidents. As Dropkin correctly points out, "It is important to emphasize here that there are two distributions which enter into our considerations. On the one hand, there is the distribution of the probability of having an accident. On the other hand, there is the distribution of risks by number of accidents. If the first distribution is a constant, then the second is a Poisson."

The example in Table 1 shows the result of 10,000 trials from a Poisson distribution with the parameter of .274. The parameter value was selected to equal the mean of the group with two violations. The sample mean and variance are shown, as well as the values of a and r as calculated by Dropkin's formulas. Comparison of the actual results with the expected results of a Poisson indicates a reasonably good fit (as expected), which is further corroborated by calculating the chi-square statistic and noting that it is significant at the 5% level. Note, however, that the negative binomial provides an even better fit. This should not be completely surprising, since the negative binomial is a two-parameter distribution and, more importantly, the Poisson can be thought of as the "limit" of a negative binomial. (Let a and r go to infinity such that r/a remains constant, and the result is a Poisson with parameter r/a.) This concept is given added intuitive appeal by examining the formula for a; the denominator is $(\sigma^2 - M)$ whose expected value is zero. Hence, calculations of the parameter a for samples from a Poisson would be expected to produce large values, which is borne out by observation.

TABLE 1

FIXED VALUE OF M

Expected Mean	.274		
Number of Trials	10,000		
Sample Mean	.2749	Poisson Chi-square	9.75
Sample Variance	.2801		
Sample a	52.5622	Negative Binomial Chi-square	5.58
Sample r	14.4494		

Number of Claims	Actual	Expected Poisson	Expected Negative Binomial
0	7613	7603.32	7616.12
1	2061	2083.31	2054.58
2	296	285.41	296.31
3	24	26.07	30.33
4 or more	6	1.89	2.65

Therefore, we see that a good fit of a negative binomial does *not* imply a *real* spread of the parameter, since a good fit is expected when there is no spread.

It should not be inferred that I disagree that there is a real spread of the parameter. I merely disagree with his proof. Indeed, the sample variance of .193 is too much larger than the mean of .163 to be accounted for by process variance.[1]

[1] This could be shown mathematically, but my statement is made upon empirical observations. A sample of 95,000 trials from a Poisson distribution with a mean of .163 produced variances generally no more than .002 higher than the mean.

He next suggests that "the function of a segregating system is to split the total heterogeneous group into homogeneous groups." I generally agree with this except I would prefer to replace "homogeneous" with "more homogeneous."

He states, "If the system we are dealing with here accomplished this purpose totally, then the distributions by number of accidents of the individual groups should be describable by Poisson curves." This statement is too strong. He has hypothesized that the accident propensities are describable by a Type III curve, which is continuous. He proposed to partition this curve into six discrete groups and measure the results against a standard (the Poisson) which requires that each group have a single-valued accident propensity. This is clearly impossible with a discrete partitioning. I would prefer that he would test to see if the result were *closer* to a Poisson curve.

His test is to compare the sample variance to the binomial variance. I am at a loss as to the reasoning behind this. If the results were Poisson, I would expect the variance to be close to the mean, not to the binomial variance which is always less than the mean.

He then concludes, "since a Poisson distribution is not indicated for the distributions by number of accidents, a negative binomial is indicated." This statement does not follow at all. This statement is equivalent to the following reasoning: "I have shown that the total group is negative binomial. This means that the distribution of parameters, $T(m)$, is describable by a Type III curve. The segregating system can be thought of as assigning individuals, hence their particular parameter, to various groups. Define $T_i(m)$ as the resulting distribution of m for the i^{th} group. If the distribution of *accidents* for each group is Poisson, then the associated $T_i(m)$ is a constant. *If the distribution is not Poisson, then the associated $T_i(m)$ is Type III.*" It should be clear that this is not true. Even if one accepts that the distribution of parameters of the total group is Type III, it is unreasonable to assume that the only possible partitions of $T(m)$ into $T_i(m)$ are either constants or Type III curves. This error is serious, since he uses it to draw conclusions about the overlap of parameter between groups.

He has made two errors of implication:

1. If the underlying $T_i(m)$ are not constants, they must be Type III. (This is equivalent to the statement that if the accident distributions are not Poisson, they must be negative binomial.)

   If we could analyze the actual distribution of accidents within each group and find that, indeed, it is closely fit by a negative binomial, then the above problem would be moot. But he still could not draw his conclusions. To see this, we have to examine the second (and most critical) error of implication.

2. If a distribution is closely fit by a negative binomial, then the distribution of parameters, $T(m)$, is closely fit by a Type III curve. Furthermore, the parameters of the Type III curve can be estimated from the mean and variance of the accident distribution.

This is basically a sensitivity question. How sensitive is the resulting distribution to the form of $T(m)$? How "close" to a Type III must $T(m)$ be to cause the accident distribution to be "close" to a negative binomial? The fact is that many reasonable forms of $T(m)$ other than a Type III curve will produce a distribution which is fit very well by a negative binomial. Table 2 shows the result of 10,000 trials from a Poisson distribution whose parameter is uniformly distributed between .194 and .354 (hence, has mean .274). Notice that the result is fit quite well by a negative binomial.

TABLE 2

M IS UNIFORMLY DISTRIBUTED OVER (.194, .354)

Expected Mean	.274		
Number of Trials	10,000		
Sample Mean	.2667	Poisson Chi-square	10.71
Sample Variance	.2776		
Sample a	24.5329	Negative Binomial Chi-square	.23
Sample r	6.5429		

Number of Claims	Actual	Expected Poisson	Expected Negative Binomial
0	7698	7603.32	7699.67
1	1978	2083.31	1973.08
2	287	285.41	291.44
3	33	26.07	32.50
4 or more	4	1.89	3.31

Table 3 is another example where the parameter could take on the value .184 or .364 with equal probability. Again, the negative binomial fits well.

TABLE 3

M HAS EQUAL PROBABILITY OF BEING .184 OR .364

Expected Mean	.274		
Number of Trials	10,000		
Sample Mean	.2690	Poisson Chi-square	4.57
Sample Variance	.2806		
Sample *a*	23.1120	Negative Binomial Chi-square	2.60
Sample *r*	6.2171		

Number of Claims	Actual	Expected Poisson	Expected Negative Binomial
0	767	760.33	768.48
1	203	208.33	198.15
2	24	28.54	29.65
3	6	2.80	3.72

These examples were not chosen arbitrarily. Note that in the example used in Table 2, there is no overlap as defined by Dropkin, i.e., no value of the parameter falls outside the mean of the neighboring groups. On the other hand, also in the example in Table 2, the possible values of the parameter are always outside the means of the neighboring groups. Hence, the conclusions he reaches concerning overlap are not well-founded.

Let me reemphasize: Although the form of the distribution of $T(m)$ needs to be Type III for the negative binomial to follow, a distribution of $T(m)$ which is significantly different from Type III will produce an accident distribution which can be fit very closely by a negative binomial. Hence, it is improper to conclude that a good fit of a negative binomial necessarily implies that the underlying $T(m)$ is Type III.

This result is certainly unfortunate, particularly with the recent furor over classifications. To my knowledge, the questions of overlap are currently unresolved, since the true accident propensities are unknown and only the resulting accident distributions are known. For a particular individual, the expected frequency is so low that process variance overpowers the information contained in the results. Dropkin's paper provides a novel approach to the problem. His approach, in brief, is to observe the distribution of accidents and, together with an assumed knowledge of the accident producing process, make inferences about the underlying distribution of accident propensities. *The concept is theoretically sound; unfortunately, the low sensitivity of the resulting distribution to the form of $T(m)$ makes it impossible to draw meaningful conclusions about* $T(m)$. The approach, however, should not be quickly discarded. Is there another way of looking at our data? Can we find some function of our data that is dependent on the form of the distribution of the accident propensities *and is highly sensitive to the form?* If so, then we could draw valid conclusions about accident propensities.

In conclusion, this paper has given an excellent discussion of the propensities of the negative binomial, and an interesting approach to the solution of a knotty problem, although this specific application of the approach was less than conclusive.

DISCUSSION BY CHRIS SVENDSGAARD

Abstract

When the chi-square test is used in the manner suggested by Mr. Dropkin, the hypothesis tested is that all insureds have the same Poisson frequency distribution (with identical means). Rejection of the hypothesis does not necessarily imply a non-Poisson frequency distribution, if individual insureds have different mean frequencies.

The "Binomial Variance" test suggested by Mr. Dropkin is invalid.

The author gratefully acknowledges the suggestions for improvements to this discussion made by Michael Fusco, Paul Braithwaite, and Dr. John Cozzolino.

1. INTRODUCTION

Mr. Dropkin's 1959 paper is important in two major respects. It shows conclusively that classifying risks solely on their driving records is not correct. (Class and territory cannot be ignored.) And it introduced the negative binomial distribution to the casualty actuarial community.

Since the paper is still on the syllabus of examinations, it is clearly relevant today. However, it has been discovered that one of the hypothesis-testing techniques used by Mr. Dropkin is invalid. Additionally, even when using the correct technique (as Mr. Dropkin does for his main results), pitfalls in interpreting the result may trap the unwary.

A common practical application of Mr. Dropkin's paper has been to apply either the chi-square or "Binomial Variance" test to a frequency distribution, reject the Poisson assumption, and conclude that frequency is negative binomial. In this review, I will show that the "Binomial Variance" test is invalid. In addition, I will show that the negative binomial is not necessarily a correct conclusion, even if the particular Poisson hypothesis tested is rejected using a valid test. Finally, I will make a quibble.

2. THE "BINOMIAL VARIANCE" TEST IS INVALID

Mr. Dropkin states, "The Binomial Variance is equal to the product of the mean and the complement of the mean[1]." His test is to compare the "Binomial Variance" to the "variance." If the latter were greater than the former, Mr. Dropkin would reject the Poisson hypothesis.

The semantics are a little fuzzy here—are we talking about parameters or estimates of parameters? Is the "mean" μ or $\Sigma X_i/n$? Parameters are unknowable. Hence Mr. Dropkin's "mean" must be the sample mean. Similarly, the "variance" must be one of the two popular sample estimates of the variance.

In the Appendix, I demonstrate that the "Binomial Variance" is always less than or equal to either sample estimate of the variance. The inequality is strict if any insured has more than one claim. This means that (if any insured has as many as 2 claims) the Poisson hypothesis would *always* be rejected using Mr. Dropkin's "Binomial Variance" test.

An alternative might be to compare the sample mean to the sample variance. Since the mean and the variance of the Poisson distribution are equal, the sample mean and the sample variance ought to be close to one another. This comparison would be a valid (although perhaps not the most powerful) test if care were taken to calculate the significance levels associated with various differences between the sample mean and sample variance. (Simply rejecting the Poisson hypothesis if the sample variance were greater than the sample mean would probably lead to a 50% rejection rate—even if the data were generated by a Poisson process, no matter how extensive the data were.) However, the chi-square test is already available—why go to the trouble of calculating significance levels for another test?

3. NEGATIVE BINOMIAL NOT NECESSARILY CORRECT

What hypothesis is being tested? Even assuming the chi-square test is performed, the answer is not "Poissoneness." The null hypothesis is really: "Each driver's number of accidents is Poisson AND each driver has the same mean frequency (Poisson parameter)." When we reject the null hypothesis using the chi-square test, we know that either the distribution is not Poisson, OR the drivers do not have the same mean frequency (Poisson parameter), OR both.

The data must all come from a single territory, a single class, and a single time period. Otherwise, drivers will have different mean frequencies, and the chi-square test will be useless, whether or not it rejects the null hypothesis.

For example, suppose that our portfolio consists of 200 drivers. Suppose 100 Class A drivers have mean claim frequency 1 per year and 100 Class B drivers have mean claim frequency 100 per year. Finally, suppose each driver's claim frequency distribution is Poisson, and each driver's claim frequency is independent of every other driver's claim frequency.

In this case, it is extremely unlikely that the chi-square test as applied by Mr. Dropkin would accept the Poisson hypothesis.[1] Yet each driver's distribution is Poisson, and the total number of claims, being the sum of independent Poisson random variables, is Poisson. The negative binomial distribution is clearly inappropriate, and the prior distribution is clearly not gamma (or any other continuous distribution).

4. A QUIBBLE

Mr. Dropkin makes the assumption (repeated by other authors) that the claim propensity must be constant over time in order to have a Poisson distribution. This assumption is false. Consider a risk that has constant Poisson frequency with mean 1 for the first half of the year, and constant Poisson frequency with mean 2 in the second half of the year. The sum of independent Poisson random variables is Poisson, so the frequency for the whole year is Poisson with mean 3. (Note that the claim propensity is not even differentiable with respect to time, let alone constant.) Nor is it necessary that claim propensity be piecewise constant. Bühlmann [2], for instance, shows this conclusion.

[1] On the average, the estimated mean would be 50.5. If every insured had this mean, the expected number of insureds having between 40 and 60 claims would be over 170. In fact, in this situation the expected number of insureds in the range is (much) less than one.

REFERENCES

[1] L. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* XLVI, 1959, p. 168.

[2] H. Bühlmann, *Mathematical Methods in Risk Theory*, Springer-Verlag, New York, 1970, pp. 49–51.

APPENDIX

Proof that "Binomial Variance" is *always* less than or equal to variance (equality) only if no insured has more than 1 claim).

Situation:

m = Number of insureds

n_i = Number of claims for ith insured

$$\text{Variance estimate} = \frac{\Sigma n_i^2}{m} - \left(\frac{\Sigma n_i}{m}\right)^2$$

$$\text{“Binomial variance”} = \left(\frac{\Sigma n_i}{m}\right)\left(1 - \frac{\Sigma n_i}{m}\right) = \frac{\Sigma n_i}{m} - \left(\frac{\Sigma n_i}{m}\right)^2$$

$$\text{Variance Estimate} - \text{“Binomial Variance”} = \frac{\Sigma n_i^2}{m} - \frac{\Sigma n_i}{m}$$

Since the number claims n_i is always an integer, $n_i^2 \geq n_i$ always.

The variance estimate used in the above has m as a denominator. This variance estimate is less than the variance estimate which has $m - 1$ in the denominator. Hence the above proof shows that the Variance Estimate minus the "Binomial Variance" is always non-negative. (Note also that when the mean number of claims is greater than 1, the "Binomial Variance" is negative.)

ADDRESS TO NEW MEMBERS—NOVEMBER 16, 1987

TWENTY-TWENTY HINDSIGHT

LEROY J. SIMON

To each of the new Fellows and new Associates of the Casualty Actuarial Society, let me extend a personal "Welcome!" And to the spouses and accompanying persons that are a part of this great accomplishment, let me say, "Thank you for the perseverance and for the sacrifice that each of you has had to make to allow this to be possible."

In the next few minutes, I'm going to make five quick points, so listen carefully, or one may zip right by you.

Point number one is borrowed from Malcolm Young's presidential address to the CPCU in 1955, which I will paraphrase to say, "It is easier to become an actuary than to be one." You now have inherited a mantle from your profession. The developments of the past 73 years by some of the actuarial giants have built the Casualty Actuarial Society to its position today where you inherit it. Your task is simple—work hard for your profession; pay it back; carry it forward to even greater heights.

The second point that I would like to make is that actuarial work is an international activity. Close to one thousand United States actuaries belong to the International Actuarial Association, and about five hundred of these belong to ASTIN. ASTIN is the acronym for Actuarial Studies in Non-Life Insurance and is the international casualty actuarial society. However, only an average of about eight United States casualty actuaries have attended ASTIN meetings over the recent past. Whether you're a Fellow or an Associate, at your earliest opportunity, consider joining IAA and simultaneously signing up for ASTIN. The literature produced and disseminated by these organizations is the principal motivation for most of the Americans that belong, and I'm sure you will find it worthwhile also. But ASTIN offers much more than that, and I urge you to try very hard to get involved in the meetings. The misinformation on ASTIN is that it is too theoretically inclined, but the facts from those who attend ASTIN meetings are that, just as in the United States, there is a mixture of theoretical and applied matters. As a special introductory one-time offer, the 1989 meeting of ASTIN will be held in the United States, linked to the November CAS meeting, so hurry now and join and be ready to register when you get your first chance.

The third point is that the world is changing—which may not be news to most people, but you, who have been poring over the textbooks so long and hard, may not have noticed. A magazine article on the ten greatest inventions this century got me to thinking about the five great events that have happened in the United States in just the span of my business career thus far. The first of these is the introduction of *Playboy*. I start with this because it tends to get your attention. I saw the publication of the first issue of *Playboy* and, for better or for worse, it represents a turning point in the cultural values guiding personal conduct and attitudes. The second great event I have seen was Howdy Doody. To me, Howdy Doody symbolizes the introduction, on a mass basis, of television into homes throughout the country. It has revolutionized communication and the use of leisure time. The third great event was the rise of Elvis Presley. I saw Elvis appear on the Ed Sullivan show when his music and his gyrations were near scandalous. But Elvis Presley, and others around that time, revolutionized the way music is written and performed. The fourth great event was Stopette. Stopette was the sponsor of a television program, "What's My Line," that had enormous popularity because of its stars: Bennett Cerf, Steve Allen, Dorothy Kilgallen, Martin Gabel, and Arlene Francis. Their product was an underarm deodorant that was dispensed through a small squeezable plastic bottle. It revolutionized the personal hygiene habits of the country, and most particularly those of men, who now even use cologne and hairspray, which would have been unheard of in past eras. Finally, the fifth great event that happened during my business career was Pacman. I credit Pacman with having taught us very subtly that computers can be fun because people suddenly realized that this very popular arcade game was, in fact, a computer. Suddenly, instead of computers being derided and joked about, the world saw that they could be fun to play with. This led to the personal computer and the discovery that they could be fun to work with as well as extraordinarily productive.

There will be at least five great events during the course of your business careers. I became an FCAS 33 years ago, and as you look back on your business careers 33 years from today—in the year 2020—with your twenty-twenty hindsight, you will picture this year as strangely as I now picture a time without *Playboy*, Howdy Doody, Elvis, Stopette, and Pacman. The lesson here is to prepare for change. Work with it; roll with it; but most of all, manage it.

Manage change in a way that will build on the foundations of the past, but with a focus on the future.

The fourth point I would like to make today is to get with a class outfit. You have obviously worked hard and diligently and sacrificed a good deal to obtain this position today. You owe it to yourself to hire yourself a good boss and be associated with an organization that you can be proud of. When Walter Wriston, the retired Chairman and CEO who led Citibank to such great heights, was asked recently to list the characteristics that he looked for in selecting people for promotion throughout the organization, his first answer was integrity. Combine integrity with the pride of accomplishment that you feel today, then blend in confidence in yourself with actuarial work, an employment you can be proud of because of the principles and the ethic for which it stands, and you have an unbeatable formula for career success. Use your best instincts and sound judgments to select wisely those few critical branches in your career path that will keep you on that first-class track.

Thus, I come to my final point: *You are the future*. It is out of this group of 1987 Fellows and Associates, and others like you, that the future leadership of the Casualty Actuarial Society must come. Individually, you will go many different ways in your careers, but wherever they may lead, you have this actuarial specialty as a safe harbor, as a solid rock, as a secure base. Develop your careers with careful planning and foresight and be sure to include a large measure of payback to the Casualty Actuarial Society—I have tried; I'm still trying. I'll be watching you because in some small part, you're *my* Class of 1987.

PRESIDENTIAL ADDRESS—NOVEMBER 17, 1987

CAS AT A CROSSROADS

MICHAEL A. WALTERS

Over the past fifteen years, the CAS has experienced remarkable growth. With over 1,300 members at present, we can almost supply one actuary for each insurance company. At this same rate of growth, we'll admit our 2,000th fellow in the year 2000, and have over 3,000 members.

At this prosperous point in our history, we are nonetheless at a crossroads with regard to our growth, direction, and identity. To navigate our way through some difficult decisions, we need to understand ourselves, our profession, and the clients we serve.

Fortunately, our long range planning committee and board of directors are well on their way to charting our path. A new mission statement reflecting input from the members has been drafted, and we have a working definition of a casualty actuary. To construct a profile of the insurance industry of the future, we have begun a series of interviews with chief executives that focus on the problems to be solved, the people needed to solve those problems, and the strengths and weaknesses of actuaries and the qualities they can bring to bear on the solutions.

What follows is a summary of the status of these projects, with some personal observations as well.

INDUSTRY OF THE FUTURE

A consolidation of insurers is inevitable in the future, with strong carriers becoming even stronger. But some new ventures are likely to succeed, thus helping the industry remain very competitive. Regulation, technology, the environment, the tort system, the economy, and changes in consumer and business attitudes will all contribute to making an already complex business among the most complex of American industries. Increased retentions and self-insurance, including the growth of risk retention groups and risk purchasing groups, will make commercial insurance even more complex than at present and increasingly based on leveraged risk.

How will insurers thrive in this environment? The successful ones will learn how to unbundle and sell services when the prices for assuming risk are too low. This implies that they will also learn how to detect the approach of those threshold points.

SKILLS OF THE FUTURE

What kind of skills are needed for success in this world of increased risk? Strong quantitative and financial skills are the key to solving underwriting and pricing problems in the future. In addition, of course, organizations must look for individuals with leadership qualities, a code of ethics, and interpersonal, communication, and other management skills.

Toward this end, casualty insurers have, no doubt, experimented with hiring MBA's over the past decade, possibly with mixed success. MBA's certainly understand the financial side of the business and what drives Wall Street. However, learning the complex world of casualty insurance may take more time than some MBA's are allowing. Recently, we asked one of the CEO's we interviewed what particular skills are needed in the future. He gave us an interesting variation on the "brains and guts" combination that traditionally is the hallmark for business success. He called it "brains and no guts." What is needed, in other words, is thoughtful pessimism, grounded in strong analytical skills.

ACTUARY OF THE FUTURE

Where does the actuary of the future fit into the industry of the future? Leaving aside for the moment the opportunities with government, buyers of insurance, or consulting firms, what does an FCAS bring to the table in managing the supply of insurance?

First and foremost, actuaries are eclectics rather than scientific specialists. The application of calculus and statistics in no way puts us on the cutting edge of those subjects. Similarly, exposure to economics, accounting, and law makes us conversant with, but not experts in, those subjects. Being tested on policy forms or regulation doesn't license us to practice in those areas without additional training. Even in the areas of pricing and reserving, where extensive testing is done, without an apprenticeship in a particular area, a new FCAS would not qualify as expert enough to give opinion testimony in a rate case or to sign a reserve statement.

So what does an FCAS designation imply about a person's overall qualifications? Given the diversity of subjects studied, the combination of which few others have likely encountered, actuaries are trained in the fundamentals of all the areas that define the practical operations of insurance, including some exposure to life insurance. This set of knowledge, obtained in, say, a seven year post-graduate span, might otherwise require twenty years of on-the-job training.

Next, an FCAS brings a high level of intelligence to apply to new problems, plus such qualities as ambition, dedication, competitiveness, and persistence. For more than a few, it also may bring a bit of arrogance for having overcome so many obstacles, and at the same time a certain trust or even gullibility for having persevered so long with the confidence that the end would be worth the struggle. Hence the syllabus committee has a great responsibility to make sure the exam process is relevant.

The designation also implies discipline, for this career has most of the characteristics of the "road less traveled"—from delaying gratification, to acceptance of responsibility, and dedication to truth.

Thus the making of an FCAS is as much a selection process as an educational one. And although attaining the FCAS doesn't guarantee integrity, it at least suggests it, for who would risk losing it all for a small and transitory advantage?

CAREERS AND TRAINING

What role does this education and selection process qualify an actuary to fulfill? Given a mastery of the fundamentals, the discipline, and the dedication, an actuary's potential should be limitless. In addition to the obvious areas of pricing and reserving, several CEO's suggested that the CAS offers an excellent training program for senior underwriting positions. In fact, the shortage of casualty actuaries has no doubt stifled a natural pipeline of talent into the underwriting function as well as senior management positions. When we raised the hypothetical question of tripling the number of casualty actuaries, we received favorable responses from the senior executives interviewed thus far.

Expanding the actuary's career path beyond traditional actuarial work, however, is not an easy and automatic transition. The commitment to the fundamentals of the insurance business and emphasis on technical information may put some actuaries at a disadvantage with regard to leadership and management advancement. The very analytical orientation of actuarial training is in partial conflict with a basic tenet of management training. Management is "the art of getting things done through others" and requires a certain letting go of technical details and ability to trust in the process to achieve the right results. Actuaries, on the other hand, are trained to cut through to the heart of a problem and try to solve it themselves.

Does this suggest management training should be on the Syllabus? Almost ten years ago, the Society of Actuaries (SOA) met with the CAS on joint education topics and suggested the idea of management topics because of a perception that new actuaries made bad managers. When the problem was studied, however, it appeared that some life companies were justifying the high salaries of their new fellows by immediately putting them in managers' positions. With little or no supervisory training or experience, no wonder these individuals struggled.

Now the SOA is seriously talking again about management training in the context of flexible education. I still think it would be better to let employers tackle that subject. What the CAS can do, however, is highlight the importance of management training, and cover the subject in continuing education workshops, rather than expanding an already comprehensive syllabus.

SYLLABUS OF THE FUTURE

The syllabus is perhaps already too detailed for the actuary of the future. If you accept that an FCAS can't remain current in every practice area, then what should be tested are the fundamentals and just enough of the details to assess overall problem solving skills. Rather than testing in some cases the ability to memorize facts, why not have a few open book exams to test the application of facts and understanding of concepts?

Some details being memorized today may have to give way to a greater breadth of topics such as finance and, possibly, the fundamentals of life insurance and pension plans. Although the content and method of education has served us well throughout our growth, we must now ask whether a ten-exam sequence of mostly self-study is the right mode for future actuaries. The traditional syllabus may not be the way to develop other skills that actuaries will need, particularly in the areas of communications—both oral and written—and public relations.

These skills depend heavily on empathy or audience awareness, traits that unfortunately are not well developed in many of today's actuaries who are trained in logic and quantitatively oriented. Actuaries may have to learn to pay as much attention to the audience receiving the message as to the message itself. Examples, analogies, or even parables do not, of course, constitute valid proofs, but they can go a long way towards helping the listener get the message. As for writing ability, essay questions on a revised Part 3 might help send the right message to students about future professional skills.

A focus on public relations is equally important. For some actuaries, public relations has almost a pejorative ring to it. And yet, unless we reach out to our two principal publics—the employers of actuaries and the students we recruit—our best laid plans may not succeed.

We have been trying to set up a joint public relations effort with the Council of Presidents. We interviewed a public relations firm that suggested we start with a communications audit of college placement directors. They also suggested contacting our employers, a process we had already begun with our CEO interviews. The SOA, which had not really considered employers as one of their publics, has recently expressed a lot of interest in these interviews.

DISCIPLINE AND STANDARDS

Another issue we must address at this crossroads is discipline. We must overcome our reluctance to "blow the whistle" when obvious transgressions occur. Poor quality work by anyone, even a competitor, hurts us all. And high quality work, even by a competitor, helps us all.

You may even be doing a favor for an actuary who is breaching the standards by reporting the violation early. A private warning or admonition by the discipline committee might help prevent a more serious violation and public sanction later in an actuary's career.

An effective discipline system is also perhaps a good way to deal with the issue of qualifications. Trying to define specific preconditions to practice in certain areas can create an unnecessary bureaucracy or result in some inequities. The issue of continuing education recognition is a good example of the latter.

Putting an asterisk next to an actuary's name in the *Yearbook* to designate a higher level of qualification implies that the outside study is crucial to one's area of practice. It also implies that those without the recognition code have not stayed current in their own specialty area and may not be qualified in any area of actuarial practice.

As an alternative to a "prior approval" system of qualification, why not use the honor code method of discipline? This means essentially a self-policing by actuaries to practice only within their areas of competence. The fear of disciplinary action by a vigilant and active group of monitoring actuaries would serve as a deterrent to actuaries practicing in areas in which they are not competent.

The formation of a permanent actuarial standards board or council in 1988 will signal the start of this new phase of professionalism. Anticipation of this has already accelerated the formulation of actuarial principles in several of our key practice areas. These principles will constitute the foundation by which the standards body will start issuing casualty standards of practice.

Without standards by which our practice areas can be judged, our body of knowledge could be viewed as neither unique nor relevant enough to serve as the foundation of a profession. Nor is discipline really possible without written standards. Such standards not only define quality service in the key areas of practice, but also offer a "safe harbor" for those who follow them in the event of an adverse outcome (when disappointed clients may vent their frustrations on the actuary).

Writing down these standards can also help enrich the knowledge of those actively involved in the process, as was the case when the actuarial principles were being articulated.

THE ORGANIZATION OF THE PROFESSION

Another fork in the road for the CAS will clearly be the question of merger with other actuarial organizations. Whether it's called unification, consolidation, or greater coordination, the idea implies some need for give and take for a greater overall good. Ideally, it would allow us to retain our autonomy and responsiveness to our casualty constituents, while improving efficiency, gaining greater leverage, and achieving synergy to address such common issues as public relations, recruiting, and discipline (particularly the triple jeopardy that is currently possible with tribunals in three different actuarial organizations that could hand down three different verdicts for the same alleged crime).

Improved efficiency does not necessarily mean that the profession's leaders can avoid a few extra meetings a year. Those who are motivated to consolidate in order to streamline meetings should realize that having others look out for their interests is false economy. Liaison representatives or their equivalent will probably still be needed, no matter what the structure of the profession.

Emotions run high when merger is discussed, as our name, heritage, and pride in our success are not easily discarded. Consequently, those CAS representatives charged with studying the possibilities will no doubt try to find alternative ways to accomplish the above goals short of actual merger. However, whatever the outcome, we must not lose the cooperation and sense of community we have built up over the years with our sister actuarial organizations.

We have much in common with our actuarial brethren in North America and much we can learn. There is also much that can be learned from what the CAS has done in its soon-to-be-celebrated 75 years in existence. We can take pride in creating a unique body of knowledge and forging an identity as the only sole-purpose casualty organization of its kind in the world. This pride gives us an extra competitive push to see where the future will take us. As Phil Ben-Zvi mentioned last year at this time, the tenets of casualty actuarial science know no national boundaries. We are hosting the ASTIN meeting in 1989 right after our Diamond Jubilee; we will then have the opportunity to celebrate the world-wide nature of the casualty business.

The idea of six specific casualty exams in addition to a common core of actuarial knowledge is appropriate for a world with more than three hundred billion dollars in casualty and property insurance premiums—more than for life and health insurance combined. So there is ample reason to believe that casualty actuaries might ultimately outnumber life actuaries. Perhaps early in the next century a CAS president will be standing here addressing a profession that has dramatically increased in size to match its stature.

Thank you for the privilege of serving this honorable profession as president.

KEYNOTE ADDRESS—NOVEMBER 17, 1987
THE TROUBLE WITH THE FUTURE

ROBERT M. EVANS

Let me begin with a brief word of gratitude, to extend to you my appreciation for inviting me to share this time with you today. I want to do something that may seem a little bit obscure or odd. I want to talk about tomorrow. I want to talk about the world of the future. Now, to many people that seems to touch on the bizarre, because the process of education you and I have grown up in would imply that we can read about the past, we can live in the present, but how can anybody possibly know about the future?

I recall when I first got a job in a newsroom in a broadcast station. There was an editor there who said it was all right if we wanted to write about the problems of tomorrow just as long as we didn't use "the future" in whatever we wrote. When I was in the seventh grade, I had to memorize Patrick Henry's "Give Me Liberty or Give Me Death" speech for a presentation before the Virginia House of Delegates. There was a phrase in it that still sticks in my mind. He said, "I have but one lamp by which my feet are guided and that is the lamp of experience. I know of no way of judging the future but by judging the past." And there was a French poet by the name of M. Paul Valery who had a most intriguing thing to say about the world of tomorrow. He said, "The trouble with the future is that it is no longer what it used to be."

Now, that is an entertaining aphorism, but all of that relates to the concept of what it is that people who deal with the future do, and how they choose to perform whatever services they offer to deal with the future. It has nothing to do with previews or predictions. There is nothing by way of forecast that comes close to that. There is nothing that is magical or mystical about it. There is no kind of a glass ball that you hold up high and gaze deep into imperturbable depths and look at some dim, vague outline of a horizon of tomorrow. There is nothing by way of witchery or sorcery in this. Not at all. It is a process of projection and there is a very simple, even scientific premise involved in it. That is to say, you take the facts of today, project them down a time frame, and, the projection will come out to the desired point in time with its obvious result. If that doesn't please you, you do the obvious; you intercede. You make some kind of interruption to make sure that that event will not take place in that way.

I find it intriguing, for example, that both John Naisbitt, author of *Megatrends*, and Tom Peters, author of *In Search of Excellence,* had virtually the same thing to say about the process that they use. They make a point of the fact that they have nothing new or unusual. They've had no incredible flash of something unrevealed. They speak of it as being an accumulation of obvious things, that is to say, a blinding flash of the obvious. Which is to say that we deal with current circumstances, project them down a time frame, and yield what that future will bring, unless you wish to change that future result.

Now, if you look beyond these years of the 1980's into the decade of the 1990's, we are talking about a common denominator of the decade of the 1990's. They will probably focus around the single concept of change. There are other concepts, of course, other things occurring, other directions, other trends. But, if there is one commonality under which we could group the trends of the decade of the 1990's, it will be that elemental concept of change. We are talking about changes in the way we work, changes in the way we live, changes in our family structure and in our professional responsibilities. We are talking about changes in competition, changes in the market place, changes in our customer base, changes in services, changes in products—quite literally—changes in virtually everything we do.

I have had several academicians make the point to me that the kind of change we are talking about for the 1990's could probably be characterized by four words, each beginning with the letter "p." First, we deal with changes that are literally without precedent. We have had no known parallel of it in the experience of our work lives in decades gone by. Second, we are talking about changes that are profound. Third, we are talking about changes that are pervasive. They surround us on every side. Fourth and last, we are talking about changes that are permanent. That is, the style of life that you and I grew up in, the way of working and accommodating the business problems that we have become accustomed to by our own traditions and habits, is now a world of the past. That world will no longer return as we used to know it.

There are a lot of things that go into any response to the question of what is happening, and a lot of them involve technology. Again, there is nothing that is new or magical about this. A great deal of this I'm sure you are familiar with. Very briefly, we now talk about an increasing era of high technology, "High Tech," that magic phrase we see in headlines and on book jackets and hear on the air all the time. To any person dealing with it alone, it probably would be difficult to perceive the entire vista of what is happening.

If you go back to the beginning of the industrial revolution, two or three centuries ago, there may have been some who perceived part of what was happening. But the difficulty, the quandary, of trying to get an accumulation of everything that is occurring is an enormously difficult exercise to engage in. Any one of us might have a perception that we have our own thin sliver of awareness and, therefore, we know what is behind it all. So any one of us, particularly people who are involved in the insurance industry as you are, could immediately say, "Well, we know what is causing the changes of today. It is the computer, isn't it?" That is valid. But it is not just the computer. Someone else might say, with a different sliver of perception, "Well, I know what is causing it. It is this new laser beam technology." And what an extraordinary impact that will have on all of us. That's also valid. But it's not just the new laser beam technology. And someone else will say, "I know exactly what it is. It is those new optical fibers that are coming, this optical fiber network that will quite literally girdle this globe we live on." That's also valid. But it is not just optical fibers alone. Someone else will say, "Well, it must be the new genetics, the new biotechnological revolution to come." And that, too, is valid. But it is not just biotech alone.

Going beyond that, someone else will say, "It's got to be the new energy forms: energy, funding and fueling everything we do, quite literally." That, too, is true. But it is not just the new energy forms alone. Someone else might make the judgment, "Well, it must be these new satellite communications." And that is equally valid. After all, satellite communications will literally change the way we contact each other. But it is not just satellite communications alone. And, as many of you, I think, begin to perceive, it is not only one or some of these. It's all of these together. It is all of these and dozens of others. Other things like credit cards, video games, localism, globalism, stereo, microchips, cable television, electronic banking, word processing, information workers, flex time, video recording, robotics, artificial intelligence. And what's happening? All of these things together are convening and converging and quite literally assaulting the basic framework that you and I have grown up in, the industrial world that the United States has built for itself. The analogy is not to a hurricane, which with an enormous force comes in and flattens structures but leaves the ground unchanged. Rather, the analogy is much more akin to an earthquake, where the ground beneath our feet trembles and there is an enormous upheaval, and the terrain on which we walk is never again the same.

I'm sure all of you have heard descriptions of the new information era to come, so let's dispense with those. One thing you may not be as aware of, however, is an intriguing aspect of the new information era to come. I refer to the time frame, the *collapse* of the time frame.

If you go back over the course of the last two or three centuries, there was indeed an effort wherein we moved away from agriculture and moved toward industry. If you would go back over this entire 20th century of the United States' economy, and if you allow the over-simplification involved, you could describe the growth and development in only three words: farmer, laborer, and clerk. Go back to the first decade of the 20th century. The great majority of American people worked on the land. We tilled the soil; we sowed our seeds; and we raised the produce we needed to survive. Jump ahead, if you would, to the mid-point of the century, the decade of the 1950's. The great preponderance of American people labored in factories. We were on assembly lines; we used machine tools to fabricate things.

Come now, if you will, to this decade of the 1980's. The great majority of American people are clerks. Without being denigrating by the use of that concept, it means that basically we work in information pursuits with ideas or concepts. Now note, if you will, the progression: from farmer to laborer to clerk, that is to say, from agriculture to industry to information. What is extraordinary is the remarkable collapse of the time frame. Because, whereas that first progression from agriculture to industry might have taken two or three centuries, the progression today from industry into information is taking place in only a decade, or, in some specific technologies, in some particular industries, in just a few years. Change is coming so remarkably fast that we scarcely have time to cope with it. We scarcely have time to reflect on what our reactions and interactions should be down that corridor, the decade of the 1990's.

If you would accept the analogy that there is a great tome that sits on a shelf and the title is *The Story of Mankind,* I think you would agree that we are coming to the end of a chapter. The chapter that has been entitled "The Age of Industry" is drawing to a close. And we are now in the opening paragraphs of the age of information.

We are not really certain what it is going to be like. We are not absolutely certain how it is going to reshape our lives. We can't give you any kind of detail on what its impact is going to be. I dare say that a great part of the American people perhaps are not even aware that it is taking place, but it may be the single most significant change in the nature of our lives and the way our economy functions since we first began that factory process that led to the rise of the age of industry. An academician described it to me in these terms: The changes in the next fifty years will have a greater impact and make a far more monumental change in our lifestyles than the changes that took place from the time primitive man lived in caves through the medieval era in central Europe. That is an enormous change, and it begins to emphasize some of the elements of what the decade of the 1990's is going to be like.

Now focus on things that are more related to your immediate work responsibilities and to your own industry's interests. Let's talk for just a little bit about insurance and some of the changes that are portended for the decade of the 1990's. Bear in mind, coming back to both Tom Peters and John Naisbitt, we talk not about startling new revelations that are anything undreamed of before: we talk about the blinding flash of the obvious. And yet, when strung together like beads upon a string to compose a necklace, they begin to give you a perspective on what is taking place.

There was a time, I'm sure you are well aware of it, when the insurance dollar used to belong to the insurance company alone, just as the banking dollar belonged to the bank, the real estate dollar belonged to the real estate broker and, of course, the investment dollar belonged to the investment broker. No longer. Because today all of those things begin to compete for the same dollar.

If you look at what has been happening to financial services, whoever dared to dream that at the mid-point of the 1980's, the largest seller of financial services in all the United States would not be the large life insurance companies like Prudential, not the largest banks in the U.S. like Citicorp, not the big accounting firms or Wall Street brokerage houses? The largest seller of financial services is a company that began as a mail order warehouse firm and then expanded to become a retail department store chain, Sears Roebuck. You've seen the sign. You go into a store at a shopping center and go past the ladies' lingerie and the men's ready to wear, beyond sportswear, beyond the kitchenware department with the pots and pans and dishes and saucers, and there you see the sign—financial services. They do the entire range of interests for people who come to inquire. You can acquire investment vehicles; you can buy insurance coverage; you can acquire real estate. And what that specifies is that the whole basis of competition is changing. It is not simply insurance; it's not simply banking; it's not simply real estate; it's not simply brokerage services. It all becomes a part of that vast anomolous vagary called financial services.

Now, that results in immense changes in the marketplace, does it not? First and most dramatically, it brings new, hitherto unrealized players into the game. It changes the roles and the rules under which the old players conduct their business as well. There has been an upheaval in the way your industry functions and operates. The boundaries between product lines in the industry begin to blur and then slowly disappear. New products pop up with increasing, almost alarming, frequency, and old products begin to fade out with increasing rapidity as well. And, of course, the baseline of the consumer to whom this industry has to sell is changing too. The market place is vastly different today.

People are living longer; people are living differently; lifestyles are changing; the work force is new and different today. Females are coming into the marketplace to hold jobs. Other minorities are rising to a position of prominence. It is interesting that 1984 was the first year that blacks were no longer the largest minority in the United States. Hispanics now outnumber blacks as the largest minority in the country.

People today are better educated. They understand more. In a financial services marketplace they are far more aware, far more informed about the array of choices that they have in front of them. They are becoming much more technologically aware. They are becoming familiar bit by bit (no pun intended) with computers. Some of them are even, perish the thought, becoming computer literate.

I had a young son that I sent off to Duke University to be a freshman this past year. When Jason was in the eighth grade he was introduced to computers. For him it was a source of fun because of the games that were played on it. When Jason got to Duke, he, of course, blew economics, and physics was a mystery to him, but computers he was able to zap away. They had a final exam with seven questions, four required, three optional. Jason answered all seven. He got 155 on the final exam and they then gave him an invitation to become an instructor in the freshman computer lab. The relevance of that? Jason represents a generation only in their teens, a generation in their twenties, a generation in their thirties and some beginning to spill into their forties, a generation that understands what so many of us would regard as technical engineering vocabulary. We are not familiar with vocabulary like peripherals and megabytes and random access memory.

Those things are as commonplace in the everyday jargon of that generation as are the designations on a baseball score card that indicate shortstop, second base, strike out, and home run. Which is to say that the vocabulary, which so many of us regard as industrial jargon that we are not familiar with, will be as commonplace in the world of tomorrow as baseball and sports terminology is today.

Will that have an impact upon you and your company and how it operates? You can bet your bottom megabyte that it will, because the present structure in corporate America will have to be realigned and changed. The definition of executive responsibilities will have to be redefined and restated. Some of you may well have jobs, portions of which may change, and portions of which may totally disappear. Just as there will be the rise of new jobs and new job responsibilities to take their place, there is going to be a demand to develop new products and new services. There will be the necessity to differentiate between present products and present services, the necessity to develop new marketing positions, the necessity to develop entirely new systems of distribution.

I must confess that I find all of that a little bit intimidating. I frequently find myself encountering corporate officials and executives from a variety of companies, across a whole cluster of different industries, who are concerned, who are anxious, and some of whom even confess to a sense of fright. That is very commonplace; it is part of human nature. *I* feel that way. I think all of us share in that sense of intimidation. After all, by definition, change has to be new; it has to be strange; it has to be untried. By definition it has to have uncertainty in it. But a time of change can be seen not just as a problem to cope with or a dilemma to be avoided. If it can be seen as an opportunity to be challenged by, it means there are extraordinary new opportunities that you open for yourselves and your personal careers, indeed, that you open for your companies in terms of realizing so much more of your corporate goals for the future.

Adapting to change can enable you to achieve those kinds of new goals for self and for company, but perhaps it is more important to reflect upon the converse of that, that is to say, the failure to change. The absence of adaptation to this new challenge can put you and all of the people who work with and for you at an enormous disadvantage, possibly even imply the risk of severe financial penalty. Which is to say, the company that can adapt to these changes can create for itself enormous advantages in reaching new markets, in reacting to new market demands and changes, in developing alternative channels for distribution and alternative products and services and, perhaps most important of all, it can begin to lock in a substantial customer base because of the sorts of services and new things it can provide for them.

If you can learn to respond to these challenges of change, if you can be aware of the potential as well as the problems and the limitations, if you can learn to adapt and adjust to these changes that will come tomorrow, those that will be most relevant to you and your needs, you will be far better armed to compete effectively and efficiently in today's marketplace, of course, but even better prepared to compete in tomorrow's marketplace that is ever growing, ever widening, ever deepening, ever expanding, ever challenging, and, of course, ever changing.

If you look back over the past half century of American business, there was so much about American business activity that was comfortable, prosperous, and predictable. Some industries became staid, proper, and tradition-bound. If there was any room for change, it was very limited. If there was any allowance of competition, it was quite controlled and constrained as well. Obviously, nobody in the 1980's could begin to describe an American marketplace with those sorts of characteristics, because what is happening in our economy today? It's exploding with entrepreneurship. It's being prodded by competition. It's being unleashed by deregulation. It is being sparked by the new marketplace. It is being spurred on by the new consumer. It is being powered by computers and fueled by a whole host of new technologies. Our world of business is changing. Indeed, a new society is struggling to be sculpted, molded, or shaped into a form that we can't quite perceive yet. I love Bob Hawkins's phrase, "the next economy." The next economy is now struggling to be born.

Let's take that into some aspects of demographics. Again, some of this I'm sure you may have familiarity with. There are certain things in basic demographics that constitute part of the marketplace in which you must compete that I think are most intriguing. I made reference to the first already. What is happening with women? 1983 was the last year, at least in terms of count, that a minority of women held jobs. In 1984 more than 53% of American females were actively involved in the marketplace, held employment. The projection is that, by the year 1990, 60% of American women will hold jobs, and by the year 2000, a little more than a dozen years away, 75% of American women will be gainfully employed. Virtually every woman under the age of thirty-five today holds a job. Or, to put it another way, under age thirty-five, virtually as many women as men are in the marketplace and are at work.

Move to a second aspect of change of demographics, the extraordinary entrepreneurial explosion that the American economy is going through. If you go back to the height of the industrial development of the U.S., the decade of the 1950's, the American economy was able to create 72,000 new business entities in one calendar year. At the time that happened we thought that was gang busters. In 1984, the American economy created over 500,000 new business entities. In 1985, it was 560,000. In 1986, it has risen to 640,000 and it is still rising. 640,000 new businesses compared with the "height" of industrial development in the U.S. of only 72,000. What is remarkable is the female role in this. More than 300,000 of those new entities were women-created, women-owned, women-run. That is to say, compared with the height of what our industrial economy was able to make, women in the 1980's have created four and a half times that volume of new business activity.

The other aspect of demographics that I suspect is of relevance to people who are involved with insurance is the aging of America, the seniors boom. The American society is growing older as the years go by. It is extraordinary. In this very year, people over age sixty-five as a group are expanding at two times the rate of the general American population. Taking it a step further, by the year 1990, the projection is that one third of the American people will be over age sixty-five. Ladies and gentlemen, that is a little staggering: one third of the American people over the age of sixty-five. The estimate is that by the time we get into the next century, when the baby boomers from 1946 and the next two decades begin to reach age sixty-five, by the year 2025, the ratio of people over 65 to people in their teens will be two to one. That is also staggering.

I suppose the best contemporary evidence of it is that 1984 marked the last time that teenagers were a dominant factor in the American population in terms of sheer numbers. 1984 was the first time in American history that people over age sixty-five were greater in number than teenagers. It is intriguing because they are easily located. If we wanted to take a survey of where the oldest senior citizens were in the U.S., seven states are where the senior citizens tend to congregate. Some are obvious; they are the sunbelt states of Florida, Texas, and California. Added to that are the four population centers of the mid-west and the northeast: New York, Pennsylvania, Ohio, and Illinois. It is intriguing, though, that there are states to which older senior citizens move, five in number that are attracting the greatest preponderance of them. Four of them, I am sure we know. I'm not sure many of us could say what the fifth state is. The four states with which we are familiar that the older senior citizens move to are Florida, Texas, California, and Arizona. But what is the fifth state to which senior citizens move? Anyone care to take a guess? Hawaii? North Carolina? Any other suggestions? It is interesting, because the fifth state to which senior citizens move, perhaps most improbably, is New Jersey. I must seek some counsel and advice from people who live there as to why that brings smiles and laughter to your faces.

What is there about New Jersey that attracts people who are retiring senior citizens? If you think about where the American population is concentrated, from Boston in the north to Washington in the south (a concentration that at some time will grow into one immense, lengthy, oblong community that will carry the name of Boswash), the central point is the New Jersey coastline. Out of all of those states on the Atlantic seacoast, that is the state that has the longest coastline. But that is only part of the reason. The major reason, quite apart from the central location between the population centers of New York to the north and Philadelphia to the south, from what we can gather, is the rise of a new aspect of real estate development.

I'm talking about a concept of congregate housing. We are familiar with the fact that elderly citizens, when they have a place, a condominium, that they buy for the future, want to have shopping close at hand. They would like to have some recreation facility close by. What is intriguing is that not only do they need financial services, they also need health facilities. The new concept of congregate housing for older citizens implies the existence of a health care center, right in whatever other facilities are constructed for the people who live there. Not necessarily an extraordinary or amazing new development, but it does account for New Jersey's value to the community of Boswash of the future.

There are a lot of marketing implications that arise from those kinds of facts, as I'm sure you would agree. This kind of congregate living will increase over time, and, as time goes on, older Americans will tend to be female. You are well aware of the fact of nature that women have a longer life expectancy than do men. It is intriguing because today, of people over the age of seventy-five in the U.S., more than seventy percent are widows, are females. I dare say that the yet unperceived political issue for the early part of the 1990's is going to be possibly the "feminization of poverty" as the elderly senior citizen community becomes increasingly female and increasingly poor. There are a lot of other marketing opportunities that are realizable in terms of a projection of an aging older population to cover a whole host of different concerns of marketing specialization for people in your business.

There is a final aspect of financial services that I want to touch on. I suspect again that it may be within some of your ranges of awareness, but I dare say many people have not really focused on it yet. I want to talk about the coming accumulation of enormous financial resources. Spectacular investments will be made in areas of the economy that have never existed before. It comes back to a concept of banking and what is happening with the change of banks. I'm sure you have heard the kind of elementary question that commentators sometimes raise—what is a bank? Everybody today is providing banking services: not only the Sears Roebuck I mentioned, but also other retailers like J. C. Penney, from American Express on one end to 7-11 convenience stores on the other.

What is a bank? What would you say to take account of the fact that a supermarket chain, the Safeway supermarkets, in the New England states alone, in any given period of time, cashes more checks than do Citicorp, Bank of America, and Chase Manhattan combined? A supermarket rendering the fundamental banking service of cashing checks for people has led to the rise of a whole host of non-banks doing banking, as we are all familiar with. A lot of that has been due to the rise of a new technology, the new automated teller machine doing banking business electronically, and there are some extraordinarily intriguing kinds of projections of what an electronic marketplace the future will yield.

Since the courts have already settled the issue of whether or not an automated teller machine is a branch bank, saying it is not, does it raise the prospect that a lot of neighborhood branch banks will become obsolete and disappear? Reflect for one moment that the great majority of banking transactions at a branch bank do not require negotiation with a bank officer. The great majority of transactions involve the elementary functions of either putting money in or taking money out. That can be done with an ATM, and it raises the prospect that branch banks will begin to disappear. Instead of having two or three banks on the north side of a city, the bank may well choose to have two or three dozen or more ATMs spread across the suburban landscape. After all, not only are they more widespread, they're available to their customers twenty-four hours a day, weekends and holidays included.

I am sure all of you are familiar, if you go into airports or shopping centers with teenagers, with what are called video game arcades. What would you say to the prospect of ATM arcades? They could be located in places like hotels, airports, shopping centers, office towers, factories, or even convention centers and hotels. An arcade might have two to four dozen ATMs from a whole host of different banks. You walk out on a street, or down a hallway, and find what look like telephone booths but are ATM booths, and your local bank has perhaps a dozen booths for banking transactions.

The outcome of that relates to another aspect of what electronic banking is bringing us. I am referring now to the aspect we call plastic money, not a credit card, but the converse of that, a debit card, that is to say, a point of sale transaction. You go to the cashier in a department store; you give them not your credit card, but your debit card. You have deposited a small amount of money with that store. The debit card is a direct transaction. Your account with that store is debited and their account is credited. There is no paperwork trail; there is no banking institution that has to participate in it. At a retail point of sale, a debit card can be a direct transaction, and all that is required is for the customer to make an advance deposit of funds and for the retail outlet then to pay interest on whatever the unused balance is at the end of the month. There is no reason that this is applicable only to a retail department store. It could be used by a restaurant, a department store, a shopping center, a hotel, a fast food outlet. You could go on a trip and, with your debit card, pay for your room at a Holiday, Ramada, or Hilton Inn. You could drive up to a Texaco, Gulf, or Exxon station and, having made a deposit with that oil company, use your debit card to pay for it. You could do the same with a shopping center, or a restaurant. On a Saturday, you could drive into McDonalds or Wendy's and treat the kids to a Big Mac, fries, and a milkshake by use of the debit card. After all, collectively, these kinds of outlets extend a far greater dollar volume of consumer credit than do banks. As I say, all they need is permission to hold a balance and to pay interest on it.

We are talking about what will become an incredibly intense, competitive marketplace. What organization, what retail outlet will be allowed to hold the money of the people who shop with them? It is going to run into hundreds of billions of dollars. We are talking essentially about banking transactions that have nothing to do with a bank. We are talking about the accumulation of vast new capital, capital resources for investment and for expenditure in a lot of different directions, many choices that have nothing to do with banks, but the accumulation of immense financial reserves that have so much to do with insurance interests and where insurance investment dollars go as well. As my French poet M. Paul Valery said about the trouble with the future, the trouble with the future is that it is no longer what it used to be.

I want to focus just a few moments on some changes that are taking place in the global arena. Raise your gaze from the horizon of the U.S. and focus on some things that are transpiring in the world around us. We have lived the better part of a half century in a world structured on certain basic assumptions that, quite frankly, flowed out of the end of WWII. Now, the basic structures of global life, what happens economically, financially, socially, militarily, and in security matters, are almost a half century old. They have never been opened to challenge or question by anyone. Yet changes are under way today that will quite fundamentally alter the basic shape of the globe as we have structured it.

The very elemental question comes to one's lips: Are we creating the shape of a world to come that is beyond our ability to imagine today?

Let me talk about a couple of concepts again, things you have heard, nothing magically new or revolutionarily different, but things that begin to change the perception of what is happening in the world around us, that alter the basic assumptions of a half century of human life on this minor planet we call Earth. Talk about a concept we have labelled Western Europe; Western Europe is no longer tied to and dominated economically by the U.S. Talk about the military strategic area; the Western alliance has become an empty shell. Talk about the disintegration of the Iron Curtain; talk about a deterioration in basic Soviet control and domination of Eastern Europe. Talk about closer links between East and West; talk about a possible reunification of East Germany with West Germany in the central part of the European continent. We are talking about changes that alter the basic assumptions for what a half century of our life has been like. As M. Paul Valery said, the trouble with the future is that it is no longer what it used to be.

Very briefly, in terms of NATO and the Western alliance, the Undersecretary of State for Political Affairs in the first Reagan administration, Lawrence Eagleberger, stated on several occasions that he thought the NATO alliance had only a 60/40 chance of survival: better than even, but not a great deal better. His concept was not that there would be any kind of legislative repeal of the Western alliance, no kind of a pronunciamento from a podium that any man would make in front of his own parliament, but that in terms of basic structure it would merely come to be a skeleton, an empty shell, in which there was no substance or content. And the obvious question is why. What is happening to imply that kind of trend for change? First, we, as a people in the U.S., are beginning to shift our attention and our focus away from the Atlantic and onto the Pacific, a most remarkable and monumental change in our basic attitude towards what we do in life. Second, the demographic focus of the U.S. is moving westward too. Population and prosperity go from the east to the west. John Naisbitt said in *Megatrends* that this very state we are in, Texas, is not southern, but much more western. Of the ten most rapidly growing cities in the U.S., eight lie in western states. We are, therefore, becoming much more attuned to what happens in the Pacific countries. Third, Western Europe as a competitive factor is opting out of a lot of high tech competition that is rising in the marketplaces of today.

It comes down to these as conclusions: a Western Europe no longer dominated by the U.S.; a Western alliance that begins to fade and disappear; a disintegration of the Iron Curtain; a deterioration of Soviet control; the reunification of East and West. We talk about changes that are generational, and M. Paul Valery's comment about the trouble with the future becomes relevant again.

Move your gaze a little further beyond Europe to what is happening in the Soviet Union. Enormous, monumental changes are taking place there. Mikhail Gorbachev is his name. He came to power in March of 1985. He was the youngest member of the Politburo, the youngest Soviet leader since Joseph Stalin assumed power before WWII, and the first Soviet leader since Nikolai Lenin to have an academic degree from college, holding a baccalaureate in law from Moscow University. Mr. Gorbachev has made us familiar with the new words glasnost (openness) and perestroika (restructuring). It is extraordinary. If you go back to 1917, the beginning of that Russian revolution, the ideology of communism had a ghostlike, haunting impact on so many parts of the world outside of Soviet borders. For three-quarters of a century, one of the major political preoccupations of dozens of governments has been that haunting concern with the ideology of communism. If you would, indulge a gross oversimplification involving two or three basic premises. The Soviets talk about the inevitability of conflict with imperialism, which means us. They talk about the inevitability of their triumph in that conflict. And they talk about the inherent superiority of their system. Conflict, triumph, and their superiority: that has been basic to every Soviet leader since 1917. It has been the fundamental tool of that society.

With Mr. Gorbachev, that basic belief is beginning to be challenged, and it's beginning to be changed. When Gorbachev talks about these things to his people, they often have a compelling, alarming quality. If you were a Soviet citizen, consider how you would react to a leader who takes a public podium and on television says to you, "Everybody in this society must change; everybody, from the worker to the minister to the Secretary of the Central Committee must change; and those who do not intend to change will be swept out of the road, violently if need be." How much change? How wide, as long as everyone has to participate in it, is the change to become? Mr. Gorbachev says the current perestroika (restructuring) embraces not only the economy but all other facets of public life: social relations, the political system, the educational system, the health care system, the ideological sphere, the political work of the party. Said

Gorbachev, "I would equate restructuring with the same import as the word revolution."

Now to you and me, who have spent so many years of our lives hearing how much the word revolution is in the holy pantheon of communism, even a sacrosanct deity kind of belief, the idea of restructuring is important to revolution, as it implies the extent, the enormousness, of the things Gorbachev is attempting to create there. If you go back over the course of these decades gone by, the Soviet state was created in the belief that they could create jobs, accumulate power, aspire to world leadership, dominate this planet we call Earth, and that, in time, their ideology would sweep the globe. All of mankind would become Marxist: their projection of the future.

Now in this decade of the 1980's, the Soviet economy cannot begin to keep pace with the U.S. The Soviets don't even purport to be competitive with the Japanese! The Soviet economy cannot even keep up with South Korea today. That does indeed begin to challenge the basis of their ideological belief, the very fundamental foundation of their creed, their role in life. A lot of us in the Western world could say there is a way to do it, and perhaps Gorbachev has the key to it: restructuring, change, accommodation, adaptation to this new high technology world.

The Soviets don't give a great deal of promise of being able to do it. It is a society in which merely having a mimeograph machine can be an act of treason. They can't reproduce printed material without having the approval of the central political organization that controls all materials published there. When I say all materials published, I mean even the print work that goes on the face of a matchbook cover, even a street sign, even a store sign that tells what kind of shop it is. Everything that is reproduced for reading has to be approved by the central authority. Is that the kind of society that can entertain the concept of a computer, a modem, data banks, and printers? It is beyond their capacity to really reflect upon. They are being asked to change, and yet the current gerontocracy, that pantheon of leaders who are in their seventies and eighties, cannot change the system. They have spent their lifetimes constructing it and building it to what it is today. They are being asked to replace it with an entirely new concept of operations that is totally unfamiliar to them. They have had no experience with it. Perhaps most important of all, for any of the leaders to change means that he has to become irrelevant and unimportant. He has to remove himself from power.

I think it would be beyond human nature for us to imagine any person being willing to collaborate in the elimination of his own access to power. It does go to the heart of the system. They were able to transform themselves from a 19th century agricultural state into a 20th century industrial giant. Can they move into the 21st century of high technological sophistication? They have given themselves a bureaucratic system that is swollen, immobilized. Their system encourages mediocrity; it rewards it. It discourages innovation. It puts down any concept of initiative. It chokes off experimentation. It rewards deceit. It is rife with corruption. You could postulate, then, that change and reform may conceivably be impossible for the Soviet Union as they march down this corridor to the decade of the 1990's.

I want to close with a focus on a concept that, again, many of you have heard of: the Pacific basin, the Pacific rim countries. It has an enormous impact and import for American business, for American corporate efforts. It is startling. We talk about a basic shift of interests in the United States away from the Atlantic alliance, away from being an Atlantic civilization to being a Pacific civilization. 1980 was the first year that the dollar volume of American trade with Pacific nations was higher than the dollar volume of trade with the Atlantic countries. That is extraordinary. Not only has that persisted through the 1980's, but in each and every year the gap between the Pacific and Atlantic trade has widened, as America more and more comes to focus on a Pacific civilization, on an economy dominated by Pacific economic relationships rather than by the Atlantic countries.

It is interesting because, in the White House, Ronald Reagan is well aware that that is taking place. He has said on several occasions words to the equivalent of, "The future of our world lies much more in the Pacific than it does in the Atlantic." The concept is not new. It goes all the way back to President Roosevelt. President Roosevelt was one who said that the dawn of the Pacific Ocean era will give rise to a new focus in man's life. Ladies and gentlemen, the Roosevelt who said that was not Franklin Delano in the 1930's or 40's. It was Theodore Roosevelt in 1903.

These projections by academicians, analysts, planners, commentators, policy makers, corporate officials, leaders, politicians, and government leaders are all coming true. It is equally relevant to Canada as it is to the U.S. The Canadians experienced in 1983, for the first time, an increase in dollar volume of trade with Pacific nations. For us in the U.S., this is an enormous change of focus and emphasis because you and I, as a people, come from antecedents who are basically European. The founding fathers crossed that body of water they called the Atlantic. The original thirteen colonies all had waters of the Atlantic lapping on their beaches and their coastlines. The waves of immigrants that increased the American population came from European countries as well. We speak a European language—English. We have a common law structure that governs our marketplace, our corporate lives, our daily work responsibilities, and our individual relationships that is derived from the European common law system. Our basic religious patterns are from European Christianity. Our family conventions, customs, and traditions of daily life are all European in their origin.

Now, after two centuries of life as a republic, we are shifting away from being an Atlantic to being a Pacific civilization. That is an extraordinary change. The Pacific rim countries, all of those countries around that great arc of the Pacific, contain forty-four percent of the world's population. It has forty-five percent of the world's gross industrial product. As a marketplace, it is in the neighborhood of three trillion dollars; not millions, not billions, but trillions and growing rapidly today.

I had it put to me in an intriguing way in Washington. Someone had commented that, for two centuries, every time American leaders looked up from their desks and gazed across a body of water, it would be the Atlantic. When Ronald Reagan and the Californians whom he has chosen as advisors and counselors raise their eyes from the desks in their studies and gaze across a body of water, it is the Pacific Ocean that they look upon. Something to reflect on: the dawn of the Pacific Ocean era.

All of this implies what I hope may be a little obvious: we are on the threshold of a startlingly new and dramatically different era. It is an era that is different from anything in our past human experience. There was once a Chinese classical philosopher, Confucius. He was able to assemble a large group of men and women, and he was very direct and to the point when he said, "All of you are cursed to live in the midst of very interesting times indeed." John Naisbitt, the closing phrase of that best selling publishing event of the 1980's, *Megatrends,* said "My God, what a fantastic time in which to be alive." Whereas that French poet, M. Paul Valery, said that the trouble with the future is that it is no longer what it used to be.

MINUTES OF THE 1987 ANNUAL MEETING

November 15–18, 1987

THE HYATT REGENCY HOTEL, SAN ANTONIO, TEXAS

Sunday, November 15, 1987

The Board of Directors held their regular quarterly meeting from 1:00 P.M. to 4:00 P.M.

Registration was held from 4:00 P.M. to 6:30 P.M.

A presentation to the new Associates and their guests on the workings of the Casualty Actuarial Society was held from 5:30 P.M. to 6:30 P.M. The Vice Presidents made short presentations concerning their areas of responsibility and the workings of the committees which report to each of them.

A general reception for all members and guests was held from 6:30 P.M. to 7:30 P.M.

Monday, November 16, 1987

Registration continued from 7:30 A.M. to 8:30 A.M.

The meeting opened with general remarks from the C.A.S. President, Michael A. Walters. Mr. Walters then announced the results of the elections of Officers and Directors:

President-Elect
    Kevin M. Ryan
Directors
    Albert J. Beer
    Alan C. Curry
    Charles L. McClenahan
    Jerome A. Scheibl

Mr. Walters recognized the twenty-five new Associates and presented diplomas to the thirty-nine new Fellows, who were introduced by Mr. David Hartman, President-Elect. The names of these individuals follow.

FELLOWS

Richard V. Atkinson	William R. Gillam	Mark W. Scully
Roger A. Atkinson, III	Catherine B. Harwood	John Slusarski
Robert S. Bennett	Jeffrey R. Jordan	Richard A. Smith
Gary S. Bujaucius	Andrew E. Kudera	Bruce R. Spidell
Andrew R. Cartmell	David L. Miller	Phillip A. Steinen
Susan J. Comstock	Susan M. Miller	Gerald R. Visintine
Martin W. Deede	Francis X. Murphy, Jr.	Steven M. Visner
Thomas J. DeFalco	Anthony Peraine	Robert H. Wainscott
Janet B. Dezube	Roberta J. Pflum	Michael C. Walsh
Jacques Dufresne	Jeffrey H. Post	Kelly A. Wargo
Dennis D. Fasking	Pamela S. Reale	Guy H. Whitehead
Sholom Feldblum	Jeffrey R. Scheuing	Robert L. Willsey
Robert W. Gardner	Peter J. Schultheiss	James W. Yow

ASSOCIATES

Charles T. Bell	Mary Jean King	Mark E. Schultze
Kathleen N. Casale	Christopher P. Maher	Susanne Sclafane
Angela F. Elliott	Blaine C. Marles	Lisa A. Slotznick
John S. Ewert	Rade T. Musulin	Linda D. Snook
Randall A. Farwell	Walter R. Naylor	Judith E. Stoffel
Nathan J. Gendelman	Rudy A. Palenik	Mary Jane Styczynski
Anne G. Greenwalt	Bruce Paterson	Thomas A. Wallace
Marshall J. Grossack	Steven C. Peck	Edward M. Wrobel, Jr.
Norman P. Hebert		

Mr. Walters then introduced LeRoy J. Simon, who delivered a brief address to the new members.

Michael Fusco, Vice President of Programs, gave a brief summary of the program content.

Mr. Walters next introduced Stephen Philbrick, Chairman of the Committee on the Review of Papers, who gave a brief summary of the new *Proceedings* papers. Mr. Philbrick announced the joint winners of the Dorweiler Prize, Howard C. Mahler for a discussion on "An Analysis of Experience Rating," and Ronald F. Wiser for "The Cost of Mixing Reinsurance." Mr. Walters then called for reviews of prior papers from those in the audience. There were none.

Mr. Walters concluded the business session at 9:45 A.M.

At 10:30 A.M., Mr. Joseph A. Herbers of Tillinghast/Towers Perrin moderated a panel entitled "Natural Catastrophes." His panel consisted of:

Donald Seagraves
Executive Director, AIRAC

Leo Jordan
Associate General Counsel, State Farm

Michael Wacek
Vice President & Actuary, E. W. Blanch

The panelists presented their views on the impact of natural catastrophes on the insurance industry.

Lunch was served from 12:00 P.M. to 1:30 P.M. A luncheon for new Fellows was hosted by members of the Executive Council.

Beginning at 1:45 P.M., there were a series of concurrent sessions, including four *Proceedings* paper presentations, a discussion of a previous *Proceedings* paper, and five workshops.

The new *Proceedings* papers presented were:

1. "Reserving Long Term Medical Claims"
   Author: Richard H. Snader
   United States Fidelity and Guaranty Company

2. "An Analysis of Excess Loss Development"
   Authors: Emanuel Pinto
   Metropolitan Reinsurance Company
   Daniel F. Gogol
   Metropolitan Reinsurance Company

3. "On the Gap Between Target and Expected Underwriting Profit Margins"
   Author: Emilio C. Venezian
   Venezian Associates

4. "Credibility for Classification Ratemaking Via the Hierarchical Normal Linear Model"
   Author: Stuart Klugman
   The University of Iowa

Also presented were discussions of a previous paper:

"Some Consideration on Automobile Rating Systems Utilizing Individual Driving Records" by Lester B. Dropkin
Discussions by: Sholom Feldblum
Stephen W. Philbrick
Christian Svendsgaard

The workshops covered the following topics:

1. "Questions and Answers with the CAS Board of Directors"
   Moderator: Michael Fusco
   Vice President, CAS Programs
   Panelists: Alan C. Curry
   Board of Directors, CAS
   David G. Hartman
   President-Elect, CAS
   Allan M. Kaufman
   Board of Directors, CAS
   Mavis A. Walters
   Board of Directors, CAS

2. "Commercial Lines Simplification"
   Moderator: Michael Averill
   Vice President, Home Insurance Company
   Panelists: Barry C. Lipton
   Assistant Actuary, Fireman's Fund Insurance Co.
   William E. Sleeper
   Principal, Sleeper, Sewell & Co.

3. "Insuring the Long Haul Trucking Industry"
   Moderator: Jay Deragon
   President, National Risk Management
   Panelists: Lana Batts
   Vice President, American Trucking Association
   Gene S. Yerant
   President, Transport Insurance Company

4. "Flexible Education: Which Way the CAS?
   Panelists: Michael L. Toothman
   Vice President, CAS Membership
   Steven G. Lehmann
   Chairman, CAS Syllabus Committee

5. Limited Attendance Workshop: "Tax Planning"
   Moderator: Christopher P. Garand
   Vice President, General Reinsurance Corporation

The Officers held a reception for new Fellows and their guests from 5:30 P.M. to 6:30 P.M.

The President's Reception was held from 6:30 P.M. to 7:30 P.M.

Tuesday, November 17, 1987

Tuesday morning from 8:30 A.M. until 9:30 A.M. was devoted to the business session of the American Academy of Actuaries.

Ms. Mavis Walters, Vice President, American Academy of Actuaries, moderated a panel entitled, "The CAS and the AAA: Working Together." Her panel consisted of:

Albert Beer
Chairperson, Committee on Property and Liability Insurance Issues, American Academy of Actuaries

Stephen Lowe
Chairperson, Committee on Property and Liability Insurance Financial Reporting, American Academy of Actuaries

M. Stanley Hughey
Member of IASB and Past President, American Academy of Actuaries

At 9:30 A.M., Mr. Walters introduced the guest speaker, Robert Evans, Senior Associate of the Naisbitt Group, who presented his views on the future.

From 11:00 A.M. to 12 P.M., Jeffrey H. Post, Senior Actuarial Officer, St. Paul Fire and Marine, moderated a panel entitled "Update on Medical Malpractice Insurance." The panel consisted of:

James O. Wood
Consulting Actuary, Tillinghast/Towers Perrin

Raymond Scalettar, M.D.
Board of Directors, American Medical Association

Lunch was served from 12:45 P.M. to 2:15 P.M. during which Mr. Walters delivered his Presidential Address.

Open CAS committee meetings were held from 2:30 P.M. to 5:00 P.M.

From 6:30 P.M. to 7:30 P.M., a general reception was held.

Wednesday, November 18, 1987

Concurrent sessions were held from 8:15 A.M. to 9:30 A.M.

At 10:00 A.M., there was a panel discussion on "Wall Street's Perspective on the Insurance Industry." The moderator was Martin Bondy, Senior Vice President, Home Insurance Company. The panelists were:

Thomas V. Cholnoky
Securities Analyst, Goldman, Sachs & Company

Leandro S. Galban, Jr.
Senior Vice President, Donaldson, Lufkin & Jenrette

David Seifer
Vice President—Research, First Boston Corporation

The closing remarks were made by Mr. Walters after which the Annual Meeting was adjourned at 11:45 A.M.

In attendance as indicated by registration records were 223 Fellows; 79 Associates; and 38 guests, subscribers, and students. The list of their names follows.

FELLOWS

Alff, G. N.
Asch, N. E.
Atkinson, R. A. III
Atkinson, R. V.
Bailey, V. M.
Balcarek, R. J.
Barclay, D. L.
Bartlett, W. N.
Bass, I. K.
Baum, E. J.
Bear, R. A.
Beer, A. J.
Bellusci, D. M.
Bennett, R. S.
Bensimon, A. S.
Ben-Zvi, P. N.
Berquist, J. R.
Bethel, N. A.
Beverage, R. M.
Bill, R. A.
Blanchard, R. S. III
Blivess, M. P.
Boone, J. P.
Bornhuetter, R. L.
Bradshaw, J. G., Jr.
Braithwaite, P.
Brannigan, J. F.
Briere, R. E.
Brooks, D. L.
Bryan, C. A.
Bujaucius, G. S.
Bursley, K. H.
Captain, J. E.
Carbaugh, A. B.
Carponter, J. D.
Carter, E. J.
Cartmell, A. R.
Comstock, S. J.
Conger, R. F.
Connell, E. C.
Crowe, P. J.
Curry, A. C.
Daino, R. A.
Dean, C. G.
Deede, M. W.
DeFalco, T. J.
Degerness, J. A.
Dembiec, L. A.
Dezube, J. B.
Dolan, M. C.
Donaldson, J. P.
Dufresne, J.
Dyck, N. P.
Dye, M. L.
Easton, R. D.
Eyers, R. G.
Faber, J. A.
Fasking, D. D.
Feldblum, S.
Ferguson, R. E.
Fiebrink, M. E.
Finger, R. J.
Fisher, R. S.
Fitzgibbon, W. J., Jr.
Flaherty, D. J.
Foote, J. M.
Ford, E. W.
Forde, C. S.
Foster, R. B.
Furst, P. A.
Fusco, M.
Gallagher, C. A.
Gallagher, T. L.
Gannon, A. H.
Garand, C. P.
Gardner, R. W.
Giambo, R. A.
Gibson, J. A., III
Gillam, W. R.
Gillespie, J. E.
Gluck, S. M.
Goldberg, S. F.
Golz, J. F.
Grady, D. J.
Graham, T. L.
Grannan, P. J.
Greco, R. E.
Hafling, D. N.
Hale, J. B.
Hall, A. A.
Hall, J. A., III
Hallstrom, R. C.
Hartman, D. G.
Harwayne, F.
Harwood, C. B.
Hebert, B. J.
Heer, E. L.
Hennessy, M. E.
Hewitt, C. C., Jr.
Higgins, B. J.
Hoppe, K. J.
Hough, P. E.
Howald, R. A.
Hoylman, D. J.
Hughey, M. S.
Hutter, H. E.
Jean, R. W.
Jordan, J. R.
Kaliski, A. E.
Kallop, R. H.
Kane, A. B.
Kaplan, R. S.
Kaufman, A. M.
Kelly, A. E.

FELLOWS

Kleinman, J. M.
Kneuer, P. J.
Knilans, K.
Kollar, J. J.
Koupf, G. I.
Krause, G. A.
Kudera, A. E.
Lamb, R. M.
LaRose, J. G.
Lehmann, S. G.
Levin, J. W.
Linden, O. M.
Lindquist, P. L.
Lipton, B. C.
Lommele, J. A.
Lonergan, K. F.
Lowe, S. P.
Lyle, A. C.
Lyons, D. K.
Makgill, S. S.
Martin, P. C.
Masterson, N. E.
McClenahan, C. L.
McDonald, G. P.
McMurray, M. A.
Mendelssohn, G. A.
Meyer, R. E.
Miccolis, R. S.
Miller, D. L.
Miller, R. A., III
Miller, S. M.
Mohl, F. J.
Munt, D. S.
Murphy, F. X., Jr.
Murrin, T. E.
Muza, J. J.
Nester, K. L.
Nichols, R. S.
Niswander, R. E., Jr.
Oakden, D. J.
Parker, C. M.
Patrik, G. S.
Peraine, A.
Pflum, R. J.
Philbrick, S. W.
Phillips, H. J.
Pinto, E.
Post, J. H.
Potts, C. M.
Pruiksma, G. J.
Purple, J. M.
Reale-Sealand, P.
Rodgers, B. T.
Rosenberg, D. M.
Roth, R. J., Jr.
Ryan, K. M.
Scheibl, J. A.
Scheuing, J. R.
Schultheiss, P. J.
Schwartz, A. I.
Scully, M. W.
Sherman, O. L., Jr.
Siewert, J. J.
Simon, L. J.
Skurnick, D.
Slusarski, J.
Smith, L. M.
Smith, R. A.
Snader, R. H.
Spidell, B. R.
Steeneck, L. R.
Steer, G. D.
Steinen, P. A.
Strug, E. J.
Suchoff, S. B.
Swift, J. A.
Terrill, K. W.
Tiller, M. W.
Toothman, M. L.
Truttmann, E. J.
Van Ark, W. R.
Venter, G. G.
Visintine, G. R.
Visner, S. M.
Wacek, M. G.
Wainscott, R. H.
Walker, G. M.
Walsh, M. C.
Walters, M. A.
Walters, M. A.
Wargo, K. A.
Warthen, T. V., III
Webb, B. L.
White, C. S.
Whitehead, G. H.
Williams, P. A.
Willsey, L. W.
Willsey, R. L.
Wilson, J. C.
Winkleman, J. J.
Withers, D. A.
Woll, R. G.
Wood, J. O.
Woods, P. B.
Wulterkens, P. E.
Yow, J. W.
Zatorski, R. T.
Zicarelli, J. D.
Zubulake, T. J.

ASSOCIATES

Anderson, B. C.
Andler, J. A.
Bell, A. A.
Bell, C. T.
Brathwaite, M. E.
Cadorine, A. R.
Canetta, J. A.
Carlton, K. E., III
Cascio, M. J.
Cathcart, S. B.
Chorpita, F. M.
Christhilf, D. A.
Cohen, A. I.
Connor, V. P.
Costner, J. E.
Covitz, B.
Crifo, D. A.
Dashoff, T. H.
DeGarmo, L. W.
Elliott, P. L.
Evans, D. M.
Ewert, J.
Farwell, R. A.
Fiebrink, D. C.
Flanagan, T. A.
Gendelman, N. J.
Gogol, D. F.
Goldberg, T. L.
Greenwalt, A. G.
Groh, L. M.
Grossack, M. J.
Halpert, A.
Harbage, R. A.
Head, T. F.
Hebert, N. P.
Herbers, J. A.
Hobart, G. P.
Hurley, P. M.
Jensen, J. P.
King, M. J.
Kolojay, T. M.
Kulik, J. M.
Kuo, C. K.
Levine, G. M.
Limpert, J. J.
Lis, R. S., Jr.
Maher, C. P.
Marles, B. C.
McGovern, E.
Morgan, S. T.
Musulin, R. T.
Naylor, W. R.
Newell, R. T.
Ollodart, B. E.
Paterson, B.
Peck, S. C.
Quintano, R. A.
Salton, J. C.
Sansevero, M., Jr.
Schultze, M. E.
Schwab, D.
Sclafane, S.
Silverman, J. K.
Slotznick, L. A.
Smith, B. W.
Snook, L. D.
Snow, D. C.
Somers, E. C.
Stoffel, J. E.
Styczynski, M. J.
Svendsgaard, C.
Taylor, A. E.
Tucker, W. B.
Turner, G. W., Jr.
Varca, J. J.
Wallace, T. A.
Webster, P. J.
Whatley, M. W.
Wrobel, E. M., Jr.

GUESTS–SUBSCRIBERS–STUDENTS

Altschuler, M.
Berry, B.
Brant, J. F.
Brassier, D.
Cunningham, J.
Demarle, G. P.
Dumontet, F.
Eversmann, T.
Feldmeier, J.
Fenrich, K.
Galban, L. S., Jr.
Gale, E.
Graves, G.
Guarini, L.
Gutman, E.
Hopkovitz, M.
Kellison, S. G.
Kido, C. T.
Knox, F.
Laberge, C.
Lemaire, J.
Little, D.
Metzner, C.
Mitchell, K.

GUESTS–SUBSCRIBERS–STUDENTS

Morrison, G.
Narayan, P.
Salton, M. A.
Scruggs, M.
Simms, G.
Smith, D.
Smith, S.
Spangler, J.
Van Leer, P.
Werland, D.
Wheaton, K.
Wills, M. E.
Wilson, G.
Winn, J. G.

REPORT OF THE VICE PRESIDENT—ADMINISTRATION

This report is intended to provide the membership with a summary of the significant activities of the CAS during the past year.

During 1987, the CAS continued to grow with 86 new members admitted and 61 Associates becoming new Fellows. Total membership now stands at 1,365. A new regional affiliate, Casualty Actuaries of the Southeast (CASE), was formed.

The Board of Directors, with primary responsibility for setting overall CAS policy, met four times during 1987. Several policy decisions were made. The significant actions taken by the Board were published in *The Actuarial Review.*

The Executive Council, with primary responsibility for day to day activities, also met four times during the year. The April meeting of the Executive Council was held in conjunction with a committee chairpersons meeting. In addition, the Executive Council met with the leadership of the regional affiliates at the Annual Meeting in November.

1987 was an unusually active year for the CAS. The activities of the Board, the Executive Council, and the CAS Committees included the following items.

Enhancement of the Body of Actuarial Knowledge

- Actuarial Principles

  Final exposure documents for statements of principles on ratemaking and reserving were released to the membership in October. A statement of principles on actuarial valuation has been drafted and will be released to the membership in 1988.

- Actuarial Textbook

  Progress is being made on a textbook on casualty actuarial science. All chapters except one have been written and are being reviewed. Exposure to the membership is expected during 1988, and publication is expected in 1989.

- Bibliographies

  Bibliographies on ratemaking, reserves, management information, risk classification, and risk theory were published. The bibliographies will be updated and expanded during 1988.

- *Actuarial Forum*

  The publication of a non-refereed journal was authorized. Entitled *Actuarial Forum,* the publication is scheduled for release in November.

- Loss Reserve Discounting

  Two white papers were completed on the topic of loss reserve discounting. One is a discussion paper prepared by the Committee on Theory of Risk. The other is a paper on techniques and considerations prepared by the Committee on Reserves. Both will appear in the *Actuarial Forum.*

Examinations, Education, and Continuing Education

- Canadian Part 8

  The first CAS examination with separate Canadian content was given.

- Flexible Education

  A proposal to extend the Flexible Education System to CAS Part 4 was considered but deferred until 1989. The CAS and the Society of Actuaries are jointly considering proposals by the Canadian Institute of Actuaries for several changes in Parts 3 and 4 of both organizations.

- Seminars

  A special interest seminar on ratemaking was held in March, and a Canadian loss reserve seminar was held in April.

- Continuing Education Catalog

  A catalog of continuing education opportunities was distributed to the membership.

Programs

- Actuarial Centennial

The actuarial centennial will be held in June, 1989 to celebrate the 100th anniversary of the actuarial profession in North America. In connection with the centennial, a history of the profession is being written. A task force of casualty actuaries has been appointed to help with the writing and editing.

- Diamond Jubilee

The CAS Diamond Jubilee will take place in November, 1989. A steering committee is actively working on plans. The 1989 ASTIN meeting will be hosted by the CAS in conjunction with the Diamond Jubilee. A history of the first 75 years of the CAS is being written.

Organization and Staffing

- Organizational Review

An Organizational Review Task Force was formed to review the organizational structure of the CAS that was established in 1983. The Task Force's objective is to determine if the structure is working as effectively as originally intended and what changes, if any, might be required to increase its effectiveness. The Task Force expects to conclude its review in 1988.

- Meeting Planner

A part-time meeting planner will be added to staff.

Planning

- Organization of the Profession

A task force under the direction of the Council of Presidents has been formed to explore ways to strengthen the actuarial profession as a whole and to consider whether restructuring the organization of the profession would help achieve this goal. The task force has three casualty representatives. A separate ad hoc CAS committee has also been formed to review the activities of the task force and formulate recommendations for the CAS Board of Directors.

- CEO Interviews

  Interviews are being conducted with CEO's of major companies for the purpose of surveying industry needs in the 1990's, management's assessment of what skills are needed to deal with those needs, and management's perspective on how well actuaries provide those skills.

Communication

- Publicity

  An arrangement was made to obtain publicity services from the American Academy of Actuaries' public information staff. Efforts have been directed toward coverage in the trade press for CAS meetings and special interest seminars.

- Membership Survey

  A comprehensive membership survey was completed and the results have been compiled for the Board of Directors, the Executive Council, and the Long Range Planning Committee. Copies were sent to committee chairmen and regional affiliates.

For 1988, the Board of Directors re-elected the following Vice Presidents:

Vice President—Administration	Richard H. Snader
Vice President—Development	Charles A. Bryan
Vice President—Membership	Michael L. Toothman
Vice President—Programs	Michael Fusco

The membership elected Kevin Ryan to President-Elect and four new members to the Board: Albert Beer, Alan Curry, Charles McClenahan, and Jerome Scheibl.

The CAS financial condition remained strong in 1987. The surplus increase in 1987 was greater than anticipated, primarily due to unexpected revenues from the ratemaking seminar and CLRS. Despite the favorable results achieved in 1987, both a dues increase and an exam fee increase will be needed in 1988 to cover major new items of expense such as the *Actuarial Forum*. Dues will be increased to $150 for both Fellows and Associates. Exam fees will be $100.

The Audit Committee examined the CAS books for fiscal year 1987 and found the accounts to be properly stated. The year ended with an increase in surplus of $40,287.71. Members' equity now stands at $373,208.41, subdivided as follows:

Michelbacher Fund	$ 67,175.45
Dorweiler Fund	9,168.31
CAS Trust	2,327.53
Scholarship Fund	7,225.48
CLRS Fund	5,000.00
CAS Surplus	282,311.64
Total Members' Equity	373,208.41

Respectfully submitted,

RICHARD H. SNADER
Vice President-Administration

FINANCIAL REPORT
FISCAL YEAR ENDED 9/30/87

OPERATING RESULTS BY FUNCTION

FUNCTION	INCOME	DISBURSEMENTS	NET RESULTS
Exams	$186,722.03	$137,563.29 (a)	$ 49,158.74
Member Services (b)	182,057.93	253,997.67	(71,939.74)
Programs	321,212.76	298,612.96	22,599.80
Other (c)	40,468.91	0.00	40,468.91
Total	$730,461.63	$690,173.92	$ 40,287.71 (d)

Notes: (a) Does not include exam related expenses incurred by the development function.
(b) Areas under supervision of VP-Administration & VP-Development.
(c) Investment income less Foreign Exchange and Miscellaneous Bank Debits.
(d) Change in CAS Surplus.

BALANCE SHEET

ASSETS	9/30/86	9/30/87	CHANGE
Checking Account	$ 63,313.38	$105,648.35	$ 42,334.97
Money Market Fund	175,786.78	107,559.25	(68,227.53)
Bank Certificates of Deposit	100,000.00	0.00	(100,000.00)
U.S. Treasury Notes & Bills	243,247.83	443,631.19	200,383.36
Accrued Interest	9,443.40	15,934.63	6,491.23
CLRS Fund	5,000.00	5,000.00	0.00
Total Assets	$596,791.39	$667,773.42	$ 80,982.03
LIABILITIES			
Office Expenses	$ 34,965.00	$ 46,000.00	$ 11,035.00
Printing Expenses	146,133.06	137,389.59	(8,743.47)
Prepaid Exam Fees	55,712.20	39,211.00	(16,501.20)
Prepaid Reading Fees	0.00	117.00	117.00
Meeting Expenses & Prepaid Fees	10,435.38	(7,529.31)	(17,964.69)
Diamond Jubilee Expense Reserve	18,729.37	84,210.03	65,480.66
Other	253.00	5,166.70	4,913.70
Total Liabilities	$266,228.01	$304,565.01	$ 38,337.00
MEMBERS EQUITY			
Michelbacher Fund	$ 64,351.65	$ 67,175.45	$ 2,823.80
Dorweiler Fund	9,703.83	9,168.31	(535.52)
CAS Trust	2,195.78	2,327.53	131.75
Scholarship Fund	7,288.19	7,225.48	(62.71)
CLRS Fund	5,000.00	5,000.00	0.00
CAS Surplus	242,023.93	282,311.64	40,287.71
Total Equity	$330,563.38	$373,208.41	$ 42,645.03

Richard H. Snader,
Vice President-Administration

This is to certify that the assets and accounts shown in the above financial statement have been audited and found to be correct.

Audit Committee

David M. Klein, Chairman
Albert J. Quirin
William J. Rowland
Charles Walter Stewart

1987 EXAMINATIONS—SUCCESSFUL CANDIDATES

Examinations for Parts 4, 6, 8, and 10 of the Casualty Actuarial Society were held on May 5, 6, 7, and 8, 1987. Examinations for Parts 5, 7, and 9 were held on November 4, 5, and 6.

Examinations for Parts 1, 2, and 3 (SOA courses 100, 110, 120, 130, and 135) are jointly sponsored by the Casualty Actuarial Society and the Society of Actuaries. These examinations were given in May and November of 1987. Candidates who passed these examinations were listed in the joint releases of the two societies.

The Casualty Actuarial Society and the Society of Actuaries jointly awarded prizes to the undergraduates ranking the highest on the Calculus and Linear Algebra examination. For the May, 1987 examination, the $200 prize was awarded to Giuseppe Russo. The additional $100 prize winners were Martin Leroux, Steven P. Lindblad, David W. Littleton, and Robert S. Manning.

For the November, 1987 examination, the $200 prize was awarded to Daniel C. Testa. The additional $100 prize winners were Daniel B. Finn, Michael J. Johnson, Timothy A. Kelley, and Andrew A. Samwick.

The following candidates will be admitted as Fellows and Associates at the May, 1988 meeting as a result of their successful completion of the Society requirements in the November, 1987 examinations.

FELLOWS

Richard V. Atkinson	Sholom Feldblum	Jeffrey H. Post
Roger A. Atkinson, III	Robert W. Gardner	Pamela S. Reale
Robert S. Bennett	William R. Gillam	Jeffrey R. Scheuing
Gary S. Bujaucius	Catherine B. Harwood	Peter J. Schultheiss
Andrew R. Cartmell	Jeffrey R. Jordan	Mark W. Scully
Susan J. Comstock	Andrew E. Kudera	John Slusarski
Martin W. Deede	David L. Miller	Richard A. Smith
Thomas J. DeFalco	Susan M. Miller	Bruce R. Spidell
Janet B. Dezube	Francis X. Murphy, Jr.	Phillip A. Steinen
Jacques Dufresne	Anthony Peraine	Gerald R. Visintine
Dennis D. Fasking	Roberta J. Pflum	Steven M. Visner

By (4.4.4), different subsets of predictors, say M and N, are compared by means of the statistic:

$$G_{MN} = (Xb)^T[X_M(X_M^TX_M)^{-1}X_M^T - X_N(X_N^TX_N)^{-1}X_N^T](Xb) - (q_M - q_N)\sigma^2,$$
$$= b^TC_{MN}b - (q_M - q_N)\sigma^2, \quad (4.4.5)$$

where C_{MN} is the appropriate $p \times p$ matrix. We note that the final member of this expression was not used by Spjøtvoll.

Spjøtvoll goes on (summarized by Miller) to develop maximum and minimum values for G_{MN} conditional upon b lying within a $(1 - \alpha)$ confidence set of the form:

$$\Pr[(b - \hat{b})^TX^TX^T(b - \hat{b}) \leq k] = 1 - \alpha,$$

where $\hat{b}$ is the regression estimate of b in the full model.

These limits on G_{MN} may be used to test whether M provides a significantly better or worse fit than N to the data.

5. METHODS OF ESTIMATION OF SECOND MOMENTS OF LOSS RESERVES

5.1. *General*

This section will consider methods by which MSEP of loss reserves can be estimated.

First note that this will not consist merely of estimating (3.3.3). Typically, Y^* will be some vector of future claim payments, subdivided for example according to year of occurrence and development year. In such a case, the estimated loss reserve would be:

$$\hat{R} = 1^T\hat{Y}^*, \quad (5.1.1)$$

where 1 is an m-vector with every component equal to unity.

Then (3.3.3) is replaced by:

$$\text{MSEP}(R) = 1^T\text{e}(e^*)^2\ 1 + 1^TE[X^*(\hat{b}_A - E\hat{b}_A)]^2\ 1 + (\text{prediction bias})^2. \quad (5.1.2)$$

This last equation shows that the MSEP of loss reserve R consists of separate terms representing statistical error, estimation error and prediction bias respectively.

Greenwood, Deborah A.
Gross, Marian R.
Gusler, Terry D.
Harr, Steven T.
Heise, Mark A.
Hemerick, Mary B.
Highet, Thomas H.
Ikeda, Joanne K.
Ill, Jeffrey R.
Jones, Brian A.
Jones, Terrell A.
Kangas, Patricia L.
Kelso, Kevin E.
Kerin, Allan A.
Kerner, Michael G.
Lavallee, Normand
Lavrey, Paul W.
Lefebvre, Christine
Letourneau, Roland D.
Lew, Allen
Loisel, Andre
Lomartire, Katherine A.
MacKenzie, Kathleen A.
McCarty, Jeffrey F.
McFarlane, Liam M.
McIntosh, Heather L.
McMillan, Liming S.
McShea, Christopher J.
Mercier, Mark F.
Mitchell, H. Elizabeth
Murphy, Daniel M.
Musulin, Rade T.
Naigles, Mark
Naylor, Walter R.
Norton, Jonathan
Nystrom, Keith R.
Ottone, Joanne M.
Patschak, Susan J.
Perez, Andre
Phifer, Robert C.
Poe, Michael D.
Pouliot, Lisa M.
Prescott, Richard W.
Redding, Scott E.
Regnier, Steven J.
Retterath, Robin M.
Roberge, Linda
Rodrigue, Michel
Samson, Pierre
Scanlon, Edmund S.
Schmidt, Jeffrey W.
Schultze, Mark E.
Shampo, Jonathan N.
Share, Robert D.
Shook, Gary E.
Speedling, Michael P.
Stahley, Barbara A.
Stephenson, Karin L.
Suchar, Christopher M.
Tang, Lee M.
Taylor, Rae M.
Teetsel, Marianne
Teng, Ting-Shih
Tremblay, Paul
Vanier, Anne-Marie
Van Laar, Kenneth R., Jr.
Wagner, Rebecca A.
Walker, Christopher P.
Wallace, Thomas A.
Wanner, Gregory S.
Watkins, Nancy P.
Weinstein, Scott P.
Wenitsky, Russell B.
Winslow, Martha A.
Yocius, Richard P.

Part 6

Allen, Danny M.
Ayres, Karen F.
Ayres, William P.
Bauer, Bruno P.
Bell, Charles T.
Book, Steven W.
Bowman, David R.
Briggs, Steven A.
Bryant, Deborah H.
Burns, Patrick J.
Byington, Jennifer S.
Cain, Mark J.
Casale, Kathleen N.
Caudill, Teresa J.
Chaffee, Janet L.
Cloutier, Denis
Cofield, Joseph F.
Conley, Kevin J.
Crowe, Alan M.
Darby, Robert N.
Elliott, Angela F.
Ely, James
Emmons, William E.
Evans, Karen F.
Ewert, John S.
Flannery, Nancy G.
Fontaine, Andre F.
Gagnon, Luc
Gendelman, Nathan J.
Gibson, John F.
Gozzo, Susan M.
Grossack, Marshall J.
Gruenhagen, Todd A.

Hebert, Norman P.
Higdon, Barbara A.
Higgins, James S.
Hill, Robert C.
Hrozinecik, George A.
Jasper, Jane E.
King, Mary Jean
Lamy, Mathieu
Leveille, Jean-Marc
Maher, Christopher P.
Mahoney, Michael W.
Marchena, Eduardo P.
Marles, Blaine C.
McCreesh, James B.
Meyer, Robert J.
Mitchell, William H.
Mohler, Elena D.
Moylan, Thomas G.
Nemlick, Kenneth J.
Nyce, G. Christopher
Palenik, Rudy A.
Patel, Bhikhabhai C.
Paterson, Bruce
Peck, Steven C.
Penick, Robert L.
Popejoy, Kathy
Premont, Andre
Schmid, Valerie L.
Schug, Richard D.
Schwartz, Arthur J.
Sclafane, Susanne
Skov, Steven A.
Slotznick, Lisa A.
Snook, Linda D.
Spieler, David
Steinberg, Karen F.
Stoffel, Judith E.
Strommen, Douglas N.
Styczynski, Mary Jane
Swanstrom, Ronald J.
Van de Water, John V.
Vetter, Barbara A.
White, Lawrence
Wildman, Peter W.
Wrobel, Edward M.
Yow, Heather E.

Part 8

Abell, Ralph L.
Allaire, Christiane
Artes, Lawrence J.
Bellafiore, Leonard A.
Blakinger, Jean M.
Bourdon, Theresa A.
Bradley, J. Scott
Brathwaite, Malcolm E.
Brehm, Paul J.
Carlson, Christopher S.
Caron, Louis P.
Chabarek, Paul
Conway, Ann M.
Crawshaw, Mark
Donnelly, Vincent T.
Feldblum, Sholom
Fletcher, James E.
Francis, Louise A.
Frank, Jacque B.
Gillam, William R.
Girard, Gregory S.
Graves, Nancy A.
Griffith, Ann V.
Groh, Linda M.
Haefner, Larry A.
Hays, David H.
Herbers, Joseph A.
Johnson, Eric J.
Johnson, Wendy A.
Joyce, John J.
Keatinge, Clive L.
Kudera, Andrew E.
Lalonde, David A.
Lamb, Dean K.
Lamb, John A.
Leccese, Nicholas M., Jr.
Lessard, Alain
MacKinnon, Brett A.
McDermott, Sean P.
Miller, David L.
Miller, Mary F.
Mueller, Nancy D.
Muller, Robert G.
Mulvaney, Mark W.
Murphy, Francis X., Jr.
Ng, Wai Hung
Overgaard, Wade T.
Phillips, George N.
Privman, Boris
Radau, Christine E.
Schwab, Debbie
Sornberger, George C.
Tucker, Warren B.
Wacker, Gregory M.
Whitehead, Guy H.
Whitlock, Robert G.
Wilk, Roger A.
Wilson, Ernest I.

Part 10

Atkinson, Richard V.
Atkinson, Roger A., III
Anderson, Mary V.
Bennett, Robert S.
Boudreau, Joseph J.
Brown, Brian Y.
Bujaucius, Gary S.
Cartmell, Andrew R.
Comstock, Susan J.
Cross, Susan L.
Deede, Martin W.
DeFalco, Thomas J.
Dekle, James N.
Dezube, Janet B.
Dufresne, Jacques
Earwaker, Bruce G.
Fasking, Dennis D.
Feldblum, Sholom
Gardner, Robert W.
Gorvett, Richard W.
Greaney, Kevin M.
Handte, Malcolm R.
Harwood, Catherine B.
Jordan, Jeffrey R.
Lacroix, Marthe A.
Miller, Susan M.
Ollodart, Bruce E.
Peraine, Anthony
Pflum, Roberta J.
Post, Jeffrey H.
Reale, Pamela S.
Sandman, Donald D.
Scheuing, Jeffrey R.
Scholl, David C.
Schultheiss, Peter J.
Schultz, Roger A.
Scully, Mark W.
Shepherd, Linda A.
Slusarski, John
Smith, Richard A.
Spidell, Bruce R.
Steinen, Phillip A.
Sutter, Russel L.
Turner, George W., Jr.
Visintine, Gerald R.
Visner, Steven M.
VonSeggern, William J.
Wainscott, Robert H.
Walsh, Michael C.
Wargo, Kelly A.
Weber, Dominic A.
Whitehead, Guy H.
Willsey, Robert L.
Yow, James W.

The following candidates will be admitted as Fellows and Associates at the May, 1988 meeting as a result of their successful completion of the Society requirements in the November, 1987 examinations.

FELLOWS

Mary V. Anderson
Brian Z. Brown
William M. Carpenter
Sanders B. Cathcart
Bruce G. Earwaker
Kenneth R. Kasner
Eric R. Keen
Marthe A. Lacroix
Patrick Mailloux
William J. Miller
George N. Phillips
Donald D. Sandman
Roger A. Schultz
Linda A. Shepherd
Russel L. Sutter
Jean Vaillancourt
William J. VonSeggern
James C. Votta
Patricia J. Webster

ASSOCIATES

Jeffrey Adams
Lawrence J. Artes
Robert K. Bender
Kay E. Bennighof
Steven W. Book
Michael Caulfield
Denis Cloutier
Joseph F. Cofield
Steven L. Colin
Kevin J. Conley
Alan M. Crowe
Michael K. Curry
Robert N. Darby
Donna R. Dickinson
Mark DiGaetano
James Ely
Karen F. Evans
William G. Fanning
Beth E. Fitzgerald
Richard J. Gergasko
Richard N. Gibson
Susan M. Gozzo
Nancy A. Graves
Bruce H. Green
James W. Haidu
James S. Higgins
Alan M. Hines
Jane E. Jasper
Steven J. Johnston
John J. Joyce
Chester T. Kido
Constantine G. Koufacos
David A. Lalonde
Susan E. LaPointe
Richard Lebrun
Cecilia M. LePere
Roland D. Letourneau
David J. Macesic
Michael W. Mahoney
James B. McCreesh
William H. Mitchell
Elena D. Mohler
Thomas G. Moylan
Chris E. Nelson
Kenneth J. Nemlick
Kwok C. Ng
Christopher G. Nyce
George N. Phillips
Denis Poirier
Sasikala Raman
Srinivasa Ramanujam
Thomas E. Schadler
Valerie L. Schmid
Richard D. Schug
Steven A. Skov
John A. Stenmark
Douglas N. Strommen
Ronald J. Swanstrom
Guy Vezina
Debra L. Werland
Peter W. Wildman
Heather E. Yow

The following is the list of successful candidates in examinations held in November, 1987.

Part 5

Abellera, Daniel N.
Adams, Jeffrey
Adams, Lawrence E.
Allen, Nancy S.
Ayres, Karen F.
Ayres, William P.
Barnes, Walter B.
Becker, Allan R.
Belleau, Richard
Blank, Cara M.
Bodiford, William T., III
Bourassa, Pierre
Brassier, Dominique E.
Brauner, Yaakov B.
Burt, Richard F., Jr.
Cain, Mark J.
Carroll, Lynn R.
Carter, Victoria J.
Cauchon, Martin
Caudill, Teresa J.
Chang, Jessalyn
Chu, Cindy C.
Cloutier, Jean
Cole, Jeffrey R.
Colford, Cynthia S.
Connor, Kathleen F.
Cossette, Charles
Costello, Dianne
Czabaj, Daniel J.
Daigneault, Wayde A.
Denoncourt, Germain
Doyon, Yves
Duffy, Timothy B.
Dumulon, Denis
Ely, James
Fallon, Steven R.
Felisky, Kendra M.
Finnerty, Deborah C.
Fitzpatrick, Kerry L.
Francoeur, Yves
Gleason, Bradley J.
Goldstein, Laurence B.
Golec, Matthew E.
Grim, Cynthia M.
Groshong, Susan J.
Gruenhagen, Todd A.
Harr, Steven T.
Hartzen, Gayle L.
Hausserman, Diane K.
Hemerick, Mary B.
Hess, Thomas G.
Highet, Thomas H.
Huberman, Gloria A.
Hwang, Li Hwan
Ill, Jeffrey R.
Jasper, Jane E.
Kelso, Kevin E.
Kenyon, Deborah E.
Kerner, Michael G.
Kim, Changseob
Kincaid, Bryan J.
Kirste, Richard O.
Koufacos, Constantine G
Kozlowski, Ronald T.
Kretsch, David J.
Laurin, Michel
Lavrey, Paul W.
Lee, Ramona C.
Lemieux, Eric F.
LePere, Cecilia M.
Lew, Allen
Luker, Christopher J.
Mackenzie, Kathleen A.
Mailhot, Susan C.
Main, William G.
Marcinko, Carole F.
McCarty, Jeffrey F.
McIntosh, Heather L.
McNeese, Dennis T.
McPadden, Matthew S.
Merlino, Paul M.
Meyer, Stephen J.
Michel, Aaron E.
Michelson, Jon W.
Miller, Brett E.
Mitchell, H. Elizabeth
Mitzel, Charles B.
Moynihan, Kevin J.
Murphy, Daniel M.
Murphy, Marianne M.
Nerone, Anthony J.
Nimick, Anne H.
Norton, Jonathan
Olszewski, Laura A.
Ondrich, Naomi S.
Palmer, Donald D.
Patschak, Susan J.
Perez, Andre
Petersen, Loren V.
Premont, Andre
Prescott, Richard W.
Price, Debbie
Quinn, Timothy P.
Raguse, Jeffrey C.
Rahardjo, Kay K.

Ramsey, Deborah L.
Reddig, Scott E.
Revilla, Victor U.
Rominske, Steven C.
Rundle, Timothy J.
Sadwin, Stuart G.
Scott, Robert F.
Shadman-Valavi, Ahmad
Sheng, Michelle G.
Simon, Christy L.
Skov, Steven A.
Speedling, Michael P.
Spiegler, David
Steenken, Lisa N.
Steinert, Lawrence J.
Stobo, Deborah A.
Strauss, Frederick M.
Struzzieri, Paul J.
Tang, Lee M.
Taylor, Rae M.
Teetsel, Marianne
Teng, Ting-Shih
Turvill, Melanie A.
Vanier, Anne-Marie
Vasek, William
Wagner, Rebecca A.
Walker, Christopher P.
Wenitsky, Russell B.
Wickenden, Leigh F.
Wischmeyer, Chad C.

Part 7

Allen, Danny M.
Artes, Lawrence J.
Barnes, Katharine E.
Bender, Robert K.
Bennighof, Kay E.
Book, Steven W.
Bradley, J. Scott
Burns, Patrick J.
Burns, William E.
Caulfield, Michael
Clark, David R.
Cloutier, Denis
Cofield, Joseph F.
Colin, Steven L.
Conley, Kevin J.
Crowe, Alan M.
Curry, Michael K.
Darby, Robert N.
Davenport, Edgar W.
Dickinson, Donna R.
DiGaetano, Mark
Evans, Karen F.
Evensen, Philip A.
Fanning, William G.
Fitzgerald, Beth E.
Gergasko, Richard J.
Gibson, Richard N.
Gill, Bonnie S.
Gozzo, Susan M.
Graves, Nancy A.
Gray, Margaret O.
Green, Bruce H.
Greenhill, Eric L.
Haidu, James W.
Heise, Mark A.
Higgins, James S.
Hines, Alan M.
Hurley, John M.
James, Peter H.
Jeffery, Philip W.
Johnston, Steven J.
Jones, Brian A.
Joyce, John J.
Kaufman, David L.
Kido, Chester T.
Konopa, Milan E.
Laberge, Christian
Lalonde, David A.
LaPointe, Susan E.
Lebrun, Richard
Letourneau, Roland D.
Macesic, David J.
Mahoney, Michael W.
McCreesh, James B.
McShea, Christopher J.
Meyer, Robert J.
Mitchell, William H.
Mohler, Elena D.
Moylan, Thomas G.
Nelson, Chris E.
Nemlick, Kenneth J.
Nesmith, Robin
Ng, Kwok C.
Nyce, G. Christopher
Ottone, Joanne M.
Paddock, Timothy A.
Phillips, George N.
Pino, Susan L.
Poirier, Denis
Provencher, Yves
Radau, Christine E.
Raman, Sasikala
Ramanujam, Srinivasa
Rech, James E.
Roberts, Jonathan S.
Rosenbach, Allen D.
Schadler, Thomas E.
Schill, Barbara J.

Schmid, Valerie L.
Schug, Richard D.
Seiter, Margaret E.
Stenmark, John A.
Stone, Edward C.
Strommen, Douglas N.
Sublett, Sharon
Suchar, Christopher M.
Swanstrom, Ronald J.
Tinkler, William P.
Vezina, Guy
Weihrich, Leslie D.
Wellington, Elizabeth A.
Werland, Debra L.
White, Lawrence
Wildman, Peter W.
Winslow, Martha A.
Yow, Heather E.

Part 9

Allaire, Christiane
Anderson, Mary V.
Bakel, Leo R.
Boisvert, Paul, Jr.
Boucek, Charles H.
Boudreau, Joseph J.
Bourdon, Theresa A.
Brahmer, John O.
Brehm, Paul J.
Brown, Brian Z.
Cardoso, Ruy A.
Carlson, Christopher S.
Caron, Louis P.
Carpenter, William M.
Cathcart, Sanders B.
Conley, Kevin J.
Conway, Ann M.
Crawshaw, Mark
Cross, Susan L.
DiDonato, Anthony M.
Earwaker, Bruce G.
Ericson, Janet M.
Girard, Gregory S.
Goldberg, Leonard R.
Gorvett, Richard W.
Graves, Gregory T.
Greene, Alex R.
Griffith, Ann V.
Groh, Linda M.
Grossack, Marshall J.
Gunn, Christy H.
Haefner, Larry A.
Hays, David H.
Hebert, Norman P.
Hill, Anthony D.
Hofmann, Richard A.
Johnson, Wendy A.
Kadison, Jeffrey P.
Kasner, Kenneth R.
Keatinge, Clive L.
Keen, Eric R.
Kryczka, John R.
Lacroix, Marthe A.
Lamb, Dean K.
Lebens, Joseph R.
Lewandowski, John J.
Maher, Christopher P.
Mailloux, Patrick
Math, Steven E.
McCoy, Mary E.
Miller, Mary F.
Miller, William J.
Mucci, Robert V.
Overgaard, Wade T.
Paterson, Bruce
Peck, Steven C.
Peterson, Steven J.
Procopio, Donald W.
Proska, Mark R.
Rice, Denise E.
Sandman, Donald D.
Schultz, Roger A.
Shapland, Mark R.
Shepherd, Linda A.
Stoffel, Judith E.
Sutter, Russel L.
Taylor, Angela E.
Theisen, Joseph P.
Thompson, Robert W.
Tistan, Ernest S.
Vaillancourt, Jean
Veilleux, Andre
Volponi, Joseph L.
VonSeggern, William J.
Votta, James C.
Wachter, Christopher J.
Wacker, Gregory M.
Webster, Patricia J.
Whitlock, Robert G., Jr.
Wilson, Ernest I.
Woerner, Susan K.
Woodruff, Arlene F.
Yit, Bill S.
Yunque, Mark A.

NEW FELLOWS ADMITTED MAY, 1987 (Left to Right): First Row: Michael A. Walters (President), Rajagopalan K. Raman, Denis G. Guenthner, David A. Withers, D. Lee Barclay, Charles I. Petit, Layne M. Onufer, Neil C. Aldin, Grover M. Edie; Second Row: Howard M. Eagelfeld, Myron L. Dye, Timothy L. Schilling, Ruth A. Howald, Allan Chuck, Wayne S. Keller, Warren D. Montgomery; Third Row: Steven A. Gapp, Paul J. Kneuer, Frederick F. Cripe, Charles Gruber, Robert H. Lee, Kenneth Easlon, Mark J. Homan.

NEW ASSOCIATES ADMITTED MAY, 1987 (Left to Right): First Row: Dean K. Lamb, Donald W. Procopio, Alex R. Greene, Larry A. Haefner, David H. Hays, Malcolm E. Brathwaite, Karen Pichler Valenti, Susan L. Cross, Bill S. Yit, William Der, Sara E. Schlenker, Michael A. Walters (President); Second Row: Charles H. Boucek, David R. Heyman, Mark R. Proska, Mark R. Shapland, Ralph L. Abell, John W. Buchanan, James C. Votta, Gregory T. Graves, Lincoln B. Williams, R. Glenn Taylor, Robert A. Weber; Third Row: Nicholas M. Leccese, Richard S. Brutto, Paul J. Brehm, Kim Scott, Ray Cardoso, Brett A. MacKinnon, Ann V. Griffith, Norman E. Donelson; Fourth Row: Jean-Luc Allard, Debbie Schwab, Gregory S. Girard, Eric R. Keen, Mark Crawshaw, Sean P. McDermott, Elise C. Liebers, Peter G. Wick, Joseph J. Boudreau; Fifth Row: Anthony M. DiDonato, Chyen Chen, Craig P. Taylor, Leonard R. Goldberg, Jerome F. Klenow, Robin Williams, Christiane Allaire, Ann M. Conway; Sixth Row: Clive L. Keatinge, Paul Boisvert, Jr., Carol Desbiens, Walter P. Cieslak, Pierre G. Laurin, Pierre Fromentin, Mary F. Miller, Theresa Wilson Bourdon; Seventh Row: Andre Veilleux, Janet M. Ericson, Sam F. Licitra, Peter J. Siczewicz, Paul E. Lacko, Kenneth R. Krissinger, John J. Lewandowski, Ernest I. Wilson; Not pictured: Joseph Lebens, Richard Plano, Frederic Schnapp

NEW FELLOWS ADMITTED NOVEMBER, 1987 (Left to Right): First Row: Michael A. Walters (President), Catherine Harwood, Pam Reale, Roberta Pflum, Susan Miller; Second Row: Kelly Wargo, Gerald Visintine, Michael Walsh, Phillip Steinen, Jeff Jordan; Third Row: James Yow, Sholom Feldblum, Janet Dezube, Anthony Peraine, Andrew Kudera, Jeffrey Scheuing; Fourth Row: Gary Bujaucius, Robert Willsey, Susan Comstock, Bruce Spidell, Robert Wainscott; Fifth Row: Martin Deede, Guy Whitehead, Bob Bennett, Roger Atkinson, Dennis Fasking; Sixth Row: David Miller, Jeff Post, Peter Schultheiss, William Gillam; Seventh Row: John Slusarski, Andrew Cartmell, Thom DeFalco, Steve Visner; Eighth Row: Mark Scully, Jacques Dufresne, Rich Atkinson, Rich Smith, Rob Gardner; Not pictured: Fancis X. Murphy, Jr.

NEW ASSOCIATES ADMITTED NOVEMBER, 1987 (Left to Right): First Row: Michael A. Walters (President), Lisa Slotznick, Anne Greenwalt, Judith Stoffel, Linda Snook; Second Row: Nathan Gendelman, Mary King, Mark Schultze, Suzanne Sclafane, Angela Taylor; Third Row: Christopher Maher, Walter Naylor, Mary Stycznyski, Bruce Paterson; Fourth Row: Norman Hebert, Thomas Wallace, Rade Musulin; Fifth row: Marshall Grossack, Blaine Marles, John Ewert, Steven Peck, Randall Farwell; Not pictured: Charles Bell, Kathleen Casale, Rudy Palenik, Edward Wrobel

OBITUARIES

Edward C. Andrews
Russell O. Hooker
Ralph M. Marshall
Nels M. Valerius
Max S. Weinstein
John C. Wooddy

EDWARD C. ANDREWS
1908–1987

Edward C. Andrews, an Associate of the Casualty Actuarial Society since 1955, died March 22, 1987 in West Hartford, Connecticut at the age of 83.

"Andy" Andrews was born in Norwich, Connecticut; graduated from Norwich Free Academy; and received his degree from Amherst College in 1926. Soon thereafter, he joined The Travelers Insurance Companies where he worked for more than 40 years and became an Associate Actuary.

At Travelers, Andy worked most of his career in the Casualty Actuarial Department. He established himself as an expert in federal and state taxation of fire and casualty insurance companies and served on many industry committees in that capacity. He was also much in demand as a panelist at tax seminars.

As a young man, Andy was an accomplished musician and sat in with many of the name bands of that time. He has also served in many capacities with the Boy Scouts of America.

Andy is survived by his son Hugh of Orleans, Massachusetts and one grandson, Edward C. Andrews, II of Marquette, Michigan.

RUSSELL O. HOOKER
1899–1987

Russell O. Hooker, a Fellow of the Society of Actuaries since 1933, a Fellow of the Casualty Actuarial Society since 1924, and a Charter Member of the American Academy of Actuaries, died on March 31, 1987 at his home in West Hartford, Connecticut at the age of 88. He was a 1920 graduate of Cornell University.

Mr. Hooker, a nationally known insurance executive, was Actuary and Director of Examinations for the Connecticut Insurance Department for 28 years and was recognized throughout the insurance industry by his representation on many committees of the National Association of Insurance Commissioners. In 1956, he established an actuarial consulting firm known as Russell O. Hooker, Consulting Actuary. This expanded to Russell O. Hooker and Associates, which in 1969 became Hooker & Holcombe, Inc.

Mr. Hooker was a ninth generation descendant of the Reverend Thomas Hooker, founder of Hartford, Connecticut.

Mr. Hooker is survived by his wife, Gertrude, a son, John Hooker, and two daughters, Barbara Thorpe and Elise Sirman.

RALPH M. MARSHALL
1896–1987

Ralph M. Marshall, a Fellow of the Casualty Actuarial Society since 1928, died August 28, 1987 in Easton, Maryland. He was 91.

Born in Pennsylvania, Mr. Marshall was a United States Army veteran of World War I. He was a 1918 graduate of Worcester Polytechnic Institute in Massachusetts. He spent his entire actuarial career at the National Council on Compensation Insurance in New York City. Mr. Marshall is remembered by most CAS members as the author of the basic text on workers' compensation ratemaking. Although some of the procedures have been revised since 1954, the paper is still a fundamental reference work on WC ratemaking.

In 1929, Mr. Marshall became a member of the CAS Education Committee, and from 1932 to 1936 served on the Examination Committee. He served as General Chairman of that committee in 1936. Mr. Marshall was a member of the Council (now the Board) from 1936 to 1939. He served on the Special Committee on Reserves for Fidelity and Surety lines in 1937 and 1938, and was a member of the Special Committee on Mortality for Disabled Lives from 1937 to 1945 and from 1955 to 1957.

In addition to his paper on WC ratemaking, Mr. Marshall co-authored a paper in 1933 on remarriage tables, and authored discussions of papers in 1932, 1934, 1939, and 1941 on WC and loss reserving topics. He retired to Maryland in 1961.

Mr. Marshall's wife, Maude (Wells), died in 1985. He is survived by a brother, Paul, of Avon, Connecticut, and several nieces and nephews.

NELS M. VALERIUS
1904–1986

Nels M. Valerius became a Fellow of the Casualty Actuarial Society in 1928. He worked for Aetna Life and Casualty for 44 years before retiring in 1969. Nels died on August 4, 1986 in Cheshire, Connecticut at the age of 82.

Born in Muskegan, Michigan, Nels lived in Connecticut most of his life. He graduated valedictorian of the Class of 1925, Trinity College, and was elected to Phi Beta Kappa.

The "big Aetna" was the scene of Mr. Valerius's actuarial life. In 1928, when he became a Fellow, he was in the Accident and Liability Department of the Aetna Life Insurance Company. He became Assistant Actuary of the Aetna Casualty and Surety Company in 1947. In 1959, he was named Associate Actuary. By the time he retired, the company was known as Aetna Life and Casualty.

Nels's contributions to the CAS were notable, both in service and in publications. He was a member of the Council (now the Board) from 1939 to 1942, from 1945 to 1947, and from 1953 to 1956. He served on the Examination Committee from 1934 to 1938 as Chairman of the Associate Part, then Chairman of the Fellowship Part, and in 1938 as General Chairman. From 1939 to 1948, he served on the Education Committee; from 1947 to 1952 on the Committee on Review of Papers; and from 1955 to 1957 on the Committee on Mortality for Disabled Lives. He was Chairman of the Committee on Review of Papers from 1949 to 1952.

Nels also provided the Society a wealth of knowledge through his papers and discussions. He authored a 1933 paper on workers' compensation reserves; co-authored a 1934 paper on WC ratemaking; authored another WC ratemaking paper in 1939, a retrospective rating paper in 1942, and notes on the Whittaker-Henderson Formula A in 1967. His discussions were published in 1934, two in 1935, another in 1941, and again in 1947.

Nels was also very active in his community and served as a member of the Newington, Connecticut Board of Finance; as a lay leader of the Bethel Baptist Church; and as a director of the Hartford Hearing League and the Elm Park Baptist Home.

Nels Valerius is survived by his wife, Gunhild Gunnarson Valerius of Cheshire, Connecticut; and three daughters, Sunnie Bachelder of Granville, New York, Cynthia Burns of Newington, Connecticut, and Sylvia Matthews of New Britain, Connecticut. He also leaves two brothers, Eric of Newington, Connecticut, and Erling, a missionary in Brazil; a sister, Esther Smith of Cheshire, Connecticut; 12 grandchildren; and 13 great-grandchildren.

MAX S. WEINSTEIN
1903–1988

Max S. Weinstein, an Associate of the Casualty Actuarial Society since 1932, and a Fellow of the Society of Actuaries, died on March 12, 1988. He was 84 years old.

Born in Brooklyn, New York, Mr. Weinstein was the first Chief Actuary of the New York State Employees Retirement System, a position he held from 1945 until his retirement in 1965. After retirement, he was a consultant to Jack Bigel Associates of New York City on union pension plans. He was instrumental in organizing the Actuarial Bureau of the New York State Teachers' Retirement System.

Mr. Weinstein held a Bachelors Degree in Electrical Engineering from the Cooper Union in New York City. Besides being a Fellow of the Society of Actuaries and an Associate of the Casualty Actuarial Society, he was a member of the Conference of Actuaries in Public Practice, The Adirondack Actuaries Club, the American Statistical Association, and the Mathematical Association of America.

Surviving family is located in the Albany, New York area.

JOHN C. WOODDY
1915–1987

John Culver Wooddy, an Associate of the Casualty Actuarial Society since 1950, and a Fellow of the Society of Actuaries since 1954, died on November 9, 1987 after a long illness. He was 72 years old.

Born September 9, 1915 in Houston, Texas, Mr. Wooddy graduated from the University of Chicago in 1936. Before enlisting in the Army in 1941, he was employed by Lumberman's Mutual Casualty Company in Chicago and passed the first two examinations of the Casualty Actuarial Society. His Army service, with rank of lieutenant, continued until 1946.

After his return to civilian life, Mr. Wooddy continued to associateship in the Casualty Actuarial Society and to fellowship in the Society of Actuaries. His first post-war employment was as staff actuary of the American Telephone and Telegraph Company in New York. In 1954, he became associated with the North American Reassurance Company, first as Assistant Actuary, and ultimately as Senior Vice-President. He retired in 1979 and entered actuarial consulting.

Mr. Wooddy's service to the profession was extraordinary in both quality and scope. He participated on many of the Society of Actuaries committees including the Examination Committee, the Committee on Research, and the Risk Theory Committee. He was elected twice to the Society of Actuaries Board of Governors.

Mr. Wooddy was also active in the Casualty Actuarial Society and published discussions of papers written by J. Lange and L. Simon. In 1971 he was a member of the Committee on Forms of Amalgamation.

Mr. Wooddy represented the United States in several International Actuarial Association posts, culminating as Vice-President for the United States between 1980 and 1984. John Wooddy will be remembered for his exemplary standards of integrity and responsibility, for his graciousness and patience in dealing with others, for the loyalty and affection he inspired, and for the courage with which he continued his life's work during years of debilitating illness. These exceptional qualities were fully shared by his wife Lucy who, until her death early in 1987, mastered the handicaps of her own illness so she could care for her husband.

In John's honor, the John Culver Wooddy annual prize for actuarial research in reinsurance and transfer of risk is to be administered by the Actuarial Education and Research Fund and is being funded by his admirers.

Mr. Wooddy is survived by his sister, Jane Wooddy Wright, of Dana Point, California.

INDEX TO VOLUME LXXIV

INDEX—CONTINUED

Page

INDEX—CONTINUED

Page