CASUALTY ACTUARIAL SOCIETY FORUM

Fall, 1987 Edition

CASUALTY ACTUARIAL SOCIETY
ORGANIZED 1914

CASUALTY ACTUARIAL SOCIETY

Charles A. Bryan
Vice President-Development

USAA Building
San Antonio, TX 78288
(512) 498-2597

July 27, 1987

TO: CAS Members

RE: Introduction of the CAS Forum

Dear CAS Members:

For several years you have expressed the need for some communications vehicle to complement the Proceedings, the Year Book and the Actuarial Review. This periodical fills the need. It is a non-refereed journal and will be used to provide a convenient means of communication for our members. We encourage authors of papers, notes or discussions to submit material to the VP-Development.

We expect future editions will contain these items:

- papers and discussions from seminars,
- work products of CAS committees,
- correspondence on actuarial topics,
- preliminary copies of papers that will eventually appear in the Proceedings, and
- reprints of important articles which are currently out of print.

In this issue we have four papers that were presented at the May CAS meeting. Please send in any comments or discussions on these papers to the Chairman of the Committee on Review of Papers.

We hope you enjoy this Actuarial Journal. We would be most happy to hear your comments and suggestions for improvements and enhancements.

Yours truly,

Charles A. Bryan

CHARLES A. BRYAN

/ss-11.68

TABLE OF CONTENTS

A DISCUSSION OF INDEXED RETENTIONS

There is a 1974 PCAS paper called, "Nonproportional Reinsurance And The Index Clause." This paper has recently resulted in some interesting correspondence. We are distributing this correspondence in the CAS Forum as a convenient way of expanding the insights available from this paper. We expect a continuing feature of the CAS Forum will be such correspondence on recent papers. This is a good method for our members to be able to present their views without the rigor necessary for the Proceedings.

We are publishing this correspondence in an unedited version. The views presented are strictly those of the writers of the letters. Since this is a non-refereed document, there has not been any attempt made to review the validity of the various statements or to decide which arguments are most compelling.

We believe this is a good opportunity for CAS members to make their views known to the entire membership and at the same time increase the value of papers. We encourage this sort of correspondence on other papers. At the same time, we hope that this sort of correspondence will grow into formal discussions submitted to the Proceedings.

Indexed Retentions And Multiple Claimant Losses In Nonproportional Reinsurance

By Sholom Feldblum

INDEXED RETENTIONS AND MULTIPLE CLAIMANT LOSSES IN NONPROPORTIONAL REINSURANCE

BY SHOLOM FELDBLUM

Ronald Ferguson, in his "Nonproportional Reinsurance and the Index Clause," discusses the motivation for and procedures of using an index clause in nonproportional reinsurance.[1] When there is a single claim subject to the retention, determining the indexed retention is straightforward: one adjusts the retention by the change in the inflation index between the policy inception date and the claim settlement date. When more than one claim or payment is subject to a single retention, the procedure becomes more difficult, since there is no single date to which the retention should be indexed. Ferguson recommends deflating all loss values to time 0, summing the deflated values, determining the percentage of the total deflated value that the original retention forms, and then applying that percentage to the settlement values to determine the primary insurer's share of the loss. Ferguson's example, which succintly illustrates the entire procedure, is reproduced below in Figure 1; the example assumes an original retention of $50,000 in a policy effective in 1974, and inflation of 10% a year.[2]

	Year Settled	Loss Amount	Index	Deflated Value
Claimant A	1975	$ 10,000	1.10	$ 10,000/1.10 = $ 9.091
Claimant B	1976	15,000	1.21	15,000/1.21 = 12,397
Claimant C	1980	150,000	1.77	150,000/1.77 = 84,746
Total		175,000		106,234

Original retention as a percent of deflated losses: 50,000/106,234 = 0.4707
Excess recovery (deflated basis): 56,234/106,234 = 0.5293

Thus, the $175,000 should be allocated as follows:

Retention: 50,000 * 175,000 / 106,234 = 175,000 * 0.4707 = $82,372
Recovery: 56,234 * 175,000 / 106,234 = 175,000 * 0.5293 = $92,628

Figure 1. Index clause procedures for a multiple claimant loss

There are two problems with this procedure. First, as Ferguson himself notes, this procedure assumes that the retention amount is not determined until after all claimants have been paid. One can solve this problem by having the retention determined as soon as enough loss payments have been made to exceed the indexed retention. For instance, if the $150,000 payment in 1980 in the example above were replaced by two payments of $75,000 each in 1979 and 1981, the retention would be determined after the 1979 payment, since

(1) Ronald E. Ferguson, "Nonproportional Reinsurance and the Index Clause," PCAS LXI, 1974, p. 141.

(2) Ferguson, op. cit., Table VI on p. 151 and Table VII on p. 152.

$$10,000 / 1.10 + 15,000 / 1.21 + 75,000 / 1.61 = 68,071 > 50,000$$

The second problem is more serious. In the example above, the primary insurer pays the first two claims in 1975 and 1976, and receives a recovery from the reinsurer only in 1980. Since inflation is positive, the value of the dollar declines in real terms over the years. The retention and excess percentages of 0.4707 and 0.5293 refer to amounts in real dollars, but they are applied to amounts in nominal dollars (the settlement values). Yet the nominal dollars paid by the primary insurer are worth more in real terms that the nominal dollars paid by the reinsurer, since they are paid earlier.

To determine the amount retained by the primary insurer, one should index the original retention to the date of the first claim settlement, subtract the amount of the settlement, index the remaining retention to the time of the second settlement, subtract the amount of the second settlement, and so forth. This procedure is illustrated for Ferguson's example in Figure 2.

Year	Retention	Loss Amount	Remaining Retention	Paid by Insurer	Paid by Reinsurer
1975	50,000 * 1.100 = 55,000	10,000	45,000	10,000	0
1976	45,000 * 1.100 = 49,500	15,000	34,500	15,000	0
1980	34,500 * 1.464 = 50,508	150,000	0	50,508	99,492
Total		175,000		75,508	99,492

Figure 2. Revised calculation of retention for multiple claimant loss

The difference between the two procedures is quite large. Using the first method, the primary insurer pays 9% more than it actually should.

Both procedures may require detailed arithmetic calculations in a large multiple claimant loss. However, if we wish to persuade American primary insurers to accept index clauses in their nonproportional reinsurance contracts, we must be careful not to allocate to them more of an indexed retention than is justified. Ferguson's reasons for using an index clause are entirely convincing; this modification of the procedure for determining the retention in a multiple claimant loss should remove one possible inequity in the application of this clause.

Letter To Mr. Feldblum From Mr. Fisher

Dated 4/21/87

April 21, 1987

Mr. Sholom Feldblum

Dear Mr. Feldblum:

Mr. Ferguson asked me to take a look at your note of 4/10/87.

Your method is interesting and different, but it doesn't accomplish a true sharing of inflation between primary company and reinsurer.

The average effect of inflation on the loss used in your example is 64.8%.

Year	Nominal Dollars	Inflation Factor	Real Dollars
1975	$ 10	1.10	9.1
1976	15	1.21	12.4
1980	150	1.77	84.7
TOTAL	175	1.648	106.2

The comparison below shows that your approach results in an inequitable sharing of inflation.

Feldblum

	Nominal Dollars	Real Dollars	Inflation Effect
Primary	$ 75.5	$ 50	+51.0%
Reinsurer	99.5	56.2	+77.1
Total	$175.0	$106.2	+64.8

Ferguson

	Nominal Dollars	Real Dollars	Inflation Effect
Primary	$ 82.4	$ 50	+64.8%
Reinsurer	92.6	56.2	+64.8
Total	$175.0	$106.2	+64.8

Interestingly enough, your point, I think, (page 2, first paragraph, last sentence) is that it is not equitable to share inflation equally. But are you confusing the leveraged effect of inflation with the leveraged effect of investment income?

Yes, the reinsurer holds his portion of the loss longer than the primary company - but the paper makes no presumption of total return pricing.

Sincerely,

Russ

Russell S. Fisher

cc: R.E. Ferguson

Letter To Mr. Fisher
From Mr. Feldblum

Dated 4/27/87

Dear Mr. Fisher,

I received your letter of April 21 contrasting the two procedures of dividing a multiple claimant loss between the ceding and assuming insurers on a non-proportional reinsurance contract with an indexed retention. You note that the ratio of settlement value to discounted value of the loss payments is the same for the ceding and assuming insurers using Ferguson's procedure, but differs in the revised procedure that I suggested.

Inflation, however, is a time-dependent concept; it considers both the nominal value of a given object at two points of time, as well as the length of time between those points. For instance, suppose a given object sells for $100 at time A and for $120 at time B. If A and B are separated by 2 years, then inflation is 9.5% per year; if there are five years between A and B, then inflation is 3.7% per year.

Thus, to examine the effect on inflation on the ceding and assuming insurers, one must look at the annual rate of inflation experienced by each. The exhibits in your letter are reproduced below, with the annual rate of inflation added.

Original Procedure:	Real Dollars	Nominal Dollars	Ratio	Average Time to Settlement	Annual Rate of Inflation
Primary insurer	50,000	82,372	1.647	4.45 years	11.9%
Reinsurer	56,234	92,628	1.647	6.00 years	8.7
Revised Procedure:					
Primary insurer	50,000	75,508	1.510	4.32 years	10.0%
Reinsurer	56,234	99,492	1.769	6.00 years	10.0

Figure 1. Annual rates of inflation experience by ceding and assuming insurers

To calculate average time to settlement for the primary insurer, I have used both discounted and nominal values of loss payments. The annual rate of inflation is equal to

ratio (of nominal to discounted values) ** (1 / average time to settlement)

Since "ratio" is a mixture of nominal and discounted values, I have determined the average time to settlement using both nominal and discounted values, and then used the average of these. Thus, for Ferguson's procedure, the primary insurer pays $10,000, $15,000, and $57,372 at 1, 2, and 6 years, respectively. These yield discounted values for the loss payments of $9,091, $12,397, and

$32,385, using a 10% discount rate. The average time to settlement is 4.66 years using the nominal values as weights, and 4.24 years using the discounted values as weights, for an average value of 4.45 years. Using the revised procedure, the nominal value of the third payment is $50,508, yielding a discounted value of $28,510. The average time to settlement is 4.54 years using the nominal values as weights, and 4.10 years using the discounted values as weights, for an average value of 4.32 years. There are other methods of determining an "average time to settlement," but the conclusion would not change: using Ferguson's method, the primary insurer experiences a significantly higher rate of inflation than the reinsurer does.

I have based the above description upon the ratio of nomical to real values used in your exhibits. In truth, this ratio is not really meaningful, since it does not take into consideration the time of each payment. Instead, financial theory currently recommends calculating the internal rate of return of each cash flow stream; this is shown in Figure 2.

Original procedure - paid by primary insurer:

Year	Nominal Payment	Time to Settlement	Internal Rate of Return	Real Value	Internal Rate of Return	Real Value
1975	10,000	1 year	11.9%	$ 8.937	12.0%	$ 8,929
1976	15,000	2 years	11.9	11,979	12.0	11,958
1980	57,372	6 years	11.9	29,223	12.0	29,067
Total:				$50,139		$49,953
		- paid by reinsurer:				
1980	92,628	6 years	8.6%	$56,463	8.7%	$56,152

Revised procedure - paid by primary insurer:

Year	Nominal Payment	Time to Settlement	Internal Rate of Return	Real Value	Internal Rate of Return	Real Value
1975	10,000	1 year	9.9%	$ 9,099	10.0%	$ 9,091
1976	15,000	2 years	9.9	12,419	10.0	12,397
1980	50,508	6 years	9.9	28,667	10.0	28,511
Total:				$50,185		$49,998
		- paid by reinsurer:				
1980	99,492	6 years	9.9%	$56,468	10.0%	$56,161

Figure 2. Internal Rates of Return for ceding and assuming insurers

---

Using the original procedure, the internal rate of return is approximately 12.0% for the primary insurer but only 8.7% for the reinsurer. Using the revised procedure, the internal rate of return is 10.0% for both insurers. Clearly, the cash flow that produces the same internal rate of return for the two insurers is the one that equitably shares the effects of inflation.

Sholom Feldblum

cc: R. Ferguson
S. Philbrick
S. Lehmann

Letter To Mr. Fisher
From Mr. Kreps

Dated 5/1/87

May 1, 1987

Mr. Russell Fisher

Dear Mr. Fisher,

I have recently seen a modification by Mr. Feldblum to the seminal paper by Ferguson on indexing retentions. I have been impressed by this sufficiently to wish to share with you an example that I find fairly persuasive.

Consider a risk with two losses, each indexed to $50,000 at time t=0. Let the direct insurer's retention also be $50,000, and the index grow at, say, 10%. Consider scenarios when the losses occur at various points in time:

1) Both losses are at t = 0. Then the direct and the reinsurer each get one loss, an equal sharing.

2) Both losses are at t = 1 year. Then the losses are $55,000, and without the indexing the reinsurer would pay $60,000 and the direct $50,000. However, with the index they each again take one loss.

3) Both losses are at t = 10 years. Then the losses are $129,687 each, and because of the indexing they are still shared.

4) One loss is at t = 1 year and one loss is at t = 10 years. This time, because of the way the index allocation is done, the direct pays not just for the first loss, but also for 29% of the second loss.

5) Even more dramatically, suppose the first loss occurs at t = 0. Then the reinsurer is in the embarrassing position of having to say, "Well, yes, your retention is $50,000 and you have in fact paid it in uninflated dollars. But, instead of covering amounts over your retention we are going to ask you for more. In fact, the longer we wait, the bigger a piece of the subsequent loss we are going to ask you to pay."

This is hardly equitable or a sharing of inflationary effect, and it is perhaps not surprising that there has been some resistance to buying into such a relationship. On the other hand, if we try Mr. Feldblum's notion, which is essentially to regard the retention as a cash flow, then in all of the above scenarios each insurer covers one loss, independent of when they happen. Since the claims follow the index and therefore have equal economic value, this is exactly what is meant by "equitably sharing the effect of inflation" or "retaining their relative monetary value".

The easiest way of stating the procedure is as follows: As claims come in, they are deflated to t = 0, and then subtracted from the retention. When the deflated claim values exceed the retention, then the reinsurer takes over.

This procedure also removes two nagging problems, especially referenced by Levin in his review, from the excellent Ferguson paper. First, it is no longer necessary to wait until all claims come in to know who will pay what. Second, multiple payments on a claim are simply indexed as they come in, and present no difficulties. In fact, this procedure makes indexed retentions as easy to work with and understand as regular retentions.

I feel that this idea has much merit, and is worthy of your consideration.

Sincerely yours,

Rodney Kreps

Rodney Kreps, A.C.A.S, Ph.D.

cc: Feldblum, Ferguson, Lehman, Philbrick

Controlling The Cycle

By Robert A. Bailey

CONTROLLING THE CYCLE

(In Cattle Prices, with Analogies to Insurance Prices)

By Robert A. Bailey, FCAS

THE PROBLEM

There is a well known cycle in the prices of cattle. This cycle causes serious risks for producers, leading occasionally to bankruptcy of producers, packing houses, and their lenders. It also leads occasionally to shortages and high prices for consumers. The uncertainty of the cycle raises costs for everyone. Lenders must charge higher rates to cattle producers to reflect the possibility of bankruptcy due to unforeseen and untimely reductions in prices. Producers must either hedge against price declines to the extent possible by use of the futures markets, or must bear the full risk themselves. In all cases, producers must pass along the costs of borrowing, hedging and risk bearing to the consumers if they are to survive and remain profitable over the long run.

BENEFITS OF CONTROLLING THE CYCLE

If the cycle were controlled and prices were stabilized, all segments of the cattle market - producers, lenders, packers, and consumers - would benefit from lower costs, lower and more stable prices, and more stable supplies.

IMPORTANCE OF IDENTIFYING THE CAUSE OF THE CYCLE

How we try to control the cycle will be greatly influenced by what we perceive to be the causes of the cycle. Much controversy and speculation has arisen over various theories of what causes the cycle. Indeed, the fact that the cycle has continued untamed for centuries - as long as there are records of cattle prices and markets - suggests either that we have not yet discovered the true cause of the cycle or that we have not yet applied adequate controls.

Realizing that many theories have already been advanced, nevertheless, I propose to advance another hypothesis (which may not be entirely new) on what causes the cycle and then to suggest controls appropriate for that hypothesis. I will call this theory the "Uncertainty Due to Time Lag" theory.

THE CAUSE - UNCERTAINTY DUE TO TIME LAG

For cattle, the time lag between breeding and slaughter is about two years. Actually, the time lag can be longer than that because one has to have breeding stock to breed before breeding can commence. But under normal or average conditions, demand can be met with a two-year time lag.

For the producer, the time lag means that he must commit his resources - breeding stock, feed, maintenance, labor, and capital for two years before he knows what he will receive for it. If a producer could know for sure what price he would receive two years later, he would know whether to raise cattle and how many to raise. But the price depends on how many cattle other producers breed concurrently - something he doesn't know in advance. So his decision, and other producers' decisions, on how many cattle to breed reflects their estimate of future supply and prices. If producers are optimistic in the aggregate, supply will be larger and prices lower than expected. If they are pessimistic, supply will be smaller and prices higher than expected.

It is the two-year time lag in creating supply, combined with the uncertainty on what the supply will be, that causes cycles in cattle supplies and prices. When prices are high, producers tend to be optimistic about future prices and tend to breed too many cattle, which depresses prices two years later. And, conversely, pessimism leads to shortages and higher prices.

This is similar to the problem that insurance producers face. They know the price in advance but there is about a two-year time lag before they find out what the costs will be. When producers are pessimistic about costs, supply shrinks immediately and prices rise. When producers are optimistic, supply rises and prices fall. It takes two years, on average, to find out whether the pessimism is justified or not. In the meantime, producers are reluctant to risk their capital unless the price is high enough to overcome their pessimism. It is that time lag in determining costs, combined with the uncertainty on what costs will be, that causes cycles in supply and prices.

REMEDY ONE - SHORTEN THE TIME LAG

If the time lag could be shortened, cycles would be shorter and less severe. So shortening the time lag would help tame the cycle. How can we shorten the time lag? There are several options. It has been observed that the time lag for livestock is related to size. So switching from cattle to hogs or rabbits would significantly shorten the time lag. But even though hogs and rabbits have long been available in the market, are less costly to produce and sell at a lower price, there are still many consumers who prefer beef.

Perhaps an advertising compaign could persuade more consumers that rabbit is better than beef. That should be at least as effective as the efforts of the insurance companies to persuade cattle producers that "claims made" insurance is better than "occurrence" insurance.

REMEDY TWO - RECOGNIZE INVESTMENT INCOME

Inasmuch as part of the price a producer receives represents interest on the capital he has invested in the cattle over the two-year lag in the form of breeding stock, feed, labor, buildings, and equipment, much of which he must borrow, perhaps his risks would be reduced if the amount of interest on his investment could be recognized and defined in advance.

A federal commission could be established to determine how much investment is required, the length of time required for each component of the investment, and a reasonable rate of investment income for cattle production. From this study, a federal rule could be promulgated that would specify what percentage of cattle prices represents a reasonable allowance for interest - more for cattle, less for hogs and rabbits.

Although it would be possible to specify interest allowances at the state level, the federal level is clearly more appropriate because 50 state commissions are unlikely all to reach the same conclusion. That would confuse both the cattle producers and the consumers as to why beef produced in one state but sold in another should have different interest allowances depending on which state it is sold in.

Unfortunately, there are critics who suggest that allocating a fixed proportion of cattle prices for interest will not affect the length of the time lag nor the uncertainty about what the supply will be. And, consequently, the cycle will continue unabated with no effect on prices and no reduction in the cost of risk. But it might help. After all, it has never been tried before. And, everything else that has been tried has been unsuccessful. So why not try it? It may have a beneficial psychological effect by distracting everyone into thinking about interest rates instead of prices. And, it would increase employment - in government, which is more stable than cattle production.

REMEDY THREE - REGULATE PRICES

Inasmuch as the cycle is caused by uncertainty over what the prices will be when the cattle are ready to be sold, we could tame the cycle by fixing the prices by government regulation. This is such an effective remedy that it has been adopted many times in many industries. One of the important benefits of price fixing by government (it is illegal and unfair for anyone else to do it), is that questions of "fairness" (defined by majority vote) are allowed to override cold, hard economics. The result is either that prices are too high and the consumers won't buy all that is supplied, or that prices are too low and producers withdraw from production. The result in both cases is that government gains the opportunity to take up the slack, either by buying the unwanted production or by supplying the unmet demand. Both result in increased government expenditures or obligations, a small price to pay for cheaper beef, more stable prices, and increased employment (in government).

REMEDY FOUR - GOVERNMENT CONTROL OF PRODUCTION

Government control of production would be even more effective than regulation of prices. This would take the uncertainty out of supply and thereby tame the cycle and eliminate the cost of risk. Although this could be achieved by assigning each cattle producer a quota and penalizing the producer if he exceeds or does not meet the quota, control of production is normally most efficient when government owns and manages the production facilities. That eliminates waste, discrimination, and inefficiency, like the Post Office and Social Security. That would enable government to give every state its fair share of cattle production facilities and end unfair discrimination among the states on the basis of climate, land costs, labor costs, and proximity to feed production.

A world without risk would be a tremendous achievement - even if it might also be a world without incentives.

REMEDY FIVE - LEAVE CATTLE PRICES ALONE

For those who find fault with all other remedies, the only remaining alternative would be to allow cattle markets to remain uncontrolled and cyclical. But that would be unamerican to leave a known problem in the hands of individuals - to allow consumers to decide what products they prefer, to allow producers to base supply solely on what they think consumers will be willing and able to pay, to require each consumer to pay for all that he buys even though many are unable to pay for all that they want, to expect politicians running for reelection to turn a deaf ear to pleas for "fairer" prices.

It would also be unfair to allow cycles in insurance prices to be controlled so successfully for so long, and not to extend the same benefits of control to cattle prices.

The Casualty Actuary's Role In Risk Management

By Oakley E. Van Slyke

THE CASUALTY ACTUARY'S ROLE IN RISK MANAGEMENT

OAKLEY E. VAN SLYKE

This paper presents an overview, a book review, and a challenge to you, the reader.

1. An <u>overview</u> of the role of the casualty actuary in risk management today.

2. A <u>book review</u> of a remarkable new work by a philosopher of science which provides a framework we actuaries can use to develop better actuarial methods for risk managers and others.

3. A <u>challenge</u>--actually, two of them--to the reader to 1) criticize the suggestions made in the second section about how to develop better methods, and 2) follow the guidelines as revised to develop better actuarial methods for risk managers.

I. CASUALTY ACTUARIES IN RISK MANAGEMENT

Casualty actuaries are increasingly involved in risk management. Although a few actuaries are risk managers for major corporations, most actuarial services are provided by consulting actuaries. Also, in the U.S. "risk management" usually refers to the handling of property and casualty risks, so most consulting actuaries active in risk management are casualty actuaries.

Actuaries' skills are most often used for the following types of problems:

- Projecting loss costs.
- Evaluating liabilities for outstanding casualty losses.
- Evaluating alternative financing arrangements for property and casualty risks.
- Recommending particular funding levels and risk financing plans.

The typical client is usually a public agency or private corporation. The actuary's contact is usually the risk manager. Although job descriptions vary, the risk manager is typically a middle manager with some background in insurance or, less often, finance. His or her responsibilities usually include the organization of the risk financing program, the placement of insurance, and the recordkeeping associated with insurance and self-insurance programs. The risk manager typically reports to the chief financial officer, but there are many exceptions.

Many actuarial concepts have obvious applications to risk management. A partial list is as follows:

- Credibility
- Distributions of loss by size
- Collective risk (distribution of aggregate losses)
- Interest theory

There are about 3,000 risk managers in the U.S. There are about 100 casualty actuaries actively involved in risk management. This 30-to-1 ratio has existed for some time. As a result of the small number of actuaries, most risk managers project losses, estimate loss ratios on excess insurance policies, and allocate costs among cost centers without the benefit of advice from actuaries. Indeed, all of the members of the Casualty Actuarial Society together do not have the time or resources to perform all of the actuarial work in risk management.

Most risk managers have never used the services of an actuary, but they have used the services of a public accounting firm, often many times. Because of the overlap between the actuary's services and the accountant's, accountants sometimes provide estimates of outstanding losses, allocations of loss costs, or other figures of an actuarial nature.

The Actuary's Most Constructive Role

The actuary's most constructive role in risk management is to provide risk managers and accountants with the basics of actuarial science. As much as we might wish to play an integral role in the day-to-day actuarial element of risk management, the work to be done is too extensive for the small number of actuaries to undertake, and it is also too routine to command the credentials of membership in the Casualty Actuarial Society.

Many risk managers are interested in learning the basics. Examples of this interest are:

- Sessions on using computers in risk management have been well attended at recent meetings of the Risk and Insurance Management Society (RIMS).

- Sessions on risk analysis, excess insurance, and loss development at recent RIMS meetings have been well received.

- Professor John Cozzolino of the Wharton School has held a number of seminars teaching risk managers several aspects of risk analysis.

- This author's assignments have often included, at the risk manager's suggestion, providing methods to calculate outstanding losses at future dates.

The major public accounting firms in the U.S. have split along two schools of thought. One group of firms does not perform actuarial services. Firms in the other group actively solicit actuarial work and employ members of the C.A.S., typically in their management consulting divisions. Smaller accounting firms, perhaps less concerned about the niceties of the big firms' philosophies, are providing loss projections and cost allocations with increasing frequency.

Although actuaries cannot and should not seek to be consulted about all actuarial matters, actuaries are uniquely positioned and qualified to improve risk managers' and accountants' actuarial skills. First, we have the respect of the leadership of the risk management profession. Second, we have enough in-depth understanding of the actuarial issues to develop the methods the risk managers ought to use. Third, we have the motivation; we will command the highest possible price in the marketplace if we provide the highest and best use of our skills.

Our role is like that of the scientist in medicine. The scientist in medicine develops the principles on which medical practice advances. Although most physicians and surgeons are not scientists, all physicians and surgeons owe the success of their profession to scientists. These scientists may be practicing physicians, teachers, or researchers, just as actuaries may be risk managers, teachers, or researchers. The difference between the scientist and the practitioner is not one of skill or experience; it is a difference in goal: the scientist is concerned with developing and testing problem-solving strategies, while the practitioner is concerned only with applying them.

II. OUR BEST, MOST TEACHABLE SCIENCE

If, as we have argued, our long-term challenge is to provide risk managers and accountants with the basics of actuarial science, then our first task is to identify the basics of actuarial science that apply to risk management. The basics we choose to teach--the guidelines for lay practice, if you will--must also be reasonably teachable.

The basic science we provide must be timeless and it must be practical. Risk managers should be able to take hold of our methods and apply them without major change for a period of many years. Moreover, change, where it does come, should come because the risk managers learn to do still more, not because the methods were poor, just as the theory of relativity is used instead of Newtonian mechanics when its refinements are important, and not because Newtonian mechanics was wrong all along. If risk managers develop actuarial skills that help them throughout their careers, they will have more respect for actuarial science and for actuaries as well.

It is not important for the risk manager to learn the theoretical underpinnings of actuarial science, any more than it is important for the construction engineer to learn quantum mechanics. It is only important for him or her to know practical methods. Actuaries have the professional responsibility to make sure the methods taught to risk managers have the appropriate theoretical foundation. Having done so, we should not limit the actuarial methods we advance to those that have a simple theoretical basis. We should advance the best methods that are still easy enough to apply to win acceptance.

Good Science

At first glance, it seems difficult to tell what parts of actuarial science are the best science. Fortunately, a thorough study of what makes a scientific theory valuable suggests we can identify our best science. In an important new book, Kitcher (1982) sets forth for the layman the findings of philosophers of science about what makes good science.

Kitcher explains that good science has three attributes:

1. "A science should be unified....Good theories consist of just one problem-solving strategy, or a small family of problem-solving strategies, that can be applied to a wide range of problems." (p.47)

2. Application of the basic problem-solving strategies requires additional hypotheses, at least about the process by which results are observed, but "an auxiliary hypothesis ought to be testable independently of the particular problem it is introduced to solve, independently of the theory it is designed to save." (p.46)

3. "A great scientific theory opens up new areas of research....Fecundity grows out of incompleteness when a theory opens up new and profitable lines of investigation." (pp.47-48)

These guidelines sound like no more than common sense, but they are not common knowledge or common practice. That is, they aren't widely known or widely used. With a few exceptions, actuaries' methods in risk management appear to be a patchwork of special methods to solve particular problems. We shouldn't be surprised to find that risk managers find our methods odd. We will have more credibility when we give risk managers a small number of problem-solving strategies that solve a large number of problems and that lead risk managers to a greater understanding of their problems.

Teachable Science

My limited experience suggests that for a scientific method to be readily learned, it must have two qualities:

1. The method should be directly applicable to some set of problems. Approaches that require the risk manager to develop his or her own auxiliary assumptions, or to perform mathematical analyses (in addition to computations), will not be used and will soon be forgotten.

2. The method should not require an unreasonable amount of calculations. Methods to solve small problems should involve small amounts of computation. Methods to solve big problems may require lengthy calculations, but if so, the methods should show the meaning of various figures derived in the course of the calculations. Nothing discourages an effort to calculate a value more than hours of work with no apparent result.

Our methods are often inappropriately tedious. We must develop methods that are susceptible to reasonable calculation, even though we must sacrifice precision, reliability, or unbiasedness to do so.

Current Actuarial Science

Kitcher's arguments are so persuasive that his three points will probably hold up under further scrutiny. The two points about teachability are tentative and may be revised substantially as time goes on. Still, it is interesting to apply these five tests to several actuarial methods and see how our methods rate.

Exhibit 1 shows this author's answers for the five points for ten different actuarial methods. The ratings for the first method, abbreviated "the actuarial equation," illustrates the application of the five points:

1. Unified?

   Yes. The actuarial equation, which states the present value of the expected value of a set of costs, provides a simple set of problem-solving strategies that can be applied to a wide variety of risk-management problems.

2. Assumptions testable?

   Yes. Assumptions about probabilities, interest rates, and amounts of loss costs can be checked independently of the actuarial equation.

3. Fecund?

   Yes. Shortcomings in the method suggest important areas of interest. For example, the method makes no accommodation for the costs associated with risk itself, but methods 2 and 3 (and others) are suggested by it.

4. Direct application?

   Sometimes. A more direct approach would be to deal with either the payment pattern or the distribution of loss amounts first, and then consider the other, but often enough the number of events (i,t) is small enough that this method seems adequately direct.

5. Reasonable calculation?

   Usually. The risk manager seldom needs to identify so many events (i,t) that computation will require more than a single page of ledger paper or a small computer spreadsheet.

III. TWO CHALLENGES

We have suggested several standards for determining the core of the science of our profession. If this approach--identifying objective criteria by which to judge actuarial methods--is valid, the analysis completed so far leaves us with two challenges.

1. Clean up the table in Exhibit 1. There are at least three areas of further work:

   a. Correct the entries in the body of the table. For example, is it true that credibility rules based on Bayes' rules seldom lead to simple calculations?

   b. Correct the column headings, especially those that assess what is teachable. Is there more? Or do these two qualities miss the mark? We need to review actuarial work the way Ehrenberg (1981) has reviewed the preparation of tabular data.

   c. Add to the topics. We need to be sure that all actuarial concepts are given a fair hearing.

2. In those cases where a method is good science but hard to teach, we need to develop simpler and more direct methods that retain the basic advantages of the methods.

For example, credibility rules based on Bayes' rule haven't replaced intuition in most risk management applications. They won't until simpler and more direct calculations replace the current formulas. This can be achieved by introducing either 1) additional assumptions for special cases (i.e. the variance of workers' compensation claims is ten times the square of the mean), or 2) alternative calculations (i.e. determining the credibility of each cost center's claims from just exposure and a list of the five largest claims in each cost center).

As these two challenges are met, we can respond to the greater challenge to rise to our most constructive role as scientists who develop better methods for ourselves and others to use.

REFERENCES

A.S.C. Ehrenberg (1981), "The Problem of Numeracy," The American Statistician (American Statistical Association), May, 1981.

Philip Kitcher (1982), Abusing Science (MIT Press).

Exhibit 1

HOW GOOD IS CURRENT ACTUARIAL SCIENCE (TO THE RISK MANAGER)?

Actuarial Method	Unified?	Assumptions Testable?	Fecund?	Direct Application?	Reasonable Calculation?
1. The "actuarial equation" $\sum_{i,t} p_{i,t} v_t a_{i,t}$	Yes	Yes	Yes	Usually	Usually
2. Exponential utility $c \ln \sum_{i,t} p_{i,t} e^{v_t \frac{a_{i,t}}{c}}$	Yes, more than #1	Yes	Yes	Less so than #1	Usually, but less than #1
3. Utility in general $\sum_{i,t} p_{i,t} U(a_{i,t})$	Yes, more than #2	Yes	Yes	No	Usually, same as #2
4. Ad hoc methods of estimating loss development	No	No	Somewhat	Yes	Yes
5. Least-squares methods of estimating loss development	Yes	Yes	Yes	Yes	Yes, if a computer is available
6. Ad hoc credibility rules	No	Yes, but assumptions usually incorrect	No	Yes	Yes
7. Credibility formulas based on Bayes' rule	Yes	Yes	Yes	Sometimes	Seldom
8. Pure Bayesian approach to credibility	Yes	Yes	Yes	No (except in cases when #7 is equivalent.)	No
9. Monte Carlo simulation	Yes	Yes	Yes	Usually	Yes, if a computer is available
10. Computation of convolutions	Somewhat	Yes	Somewhat	Often	Often

THEORY OF RISK DISCUSSION PAPER

The CAS Forum will be used to provide to our membership various committee work products. The first such committee work product is "Risk Theoretic Issues In The Discounting Of Loss Reserves." This is an important discussion document which allows us to focus in on several issues that have not been adequately addressed in the discussion of the discounting of loss reserves. Please forward any comments or discussion to Gary Patrik, Chairman of the Committee on Theory of Risk.

In future issues of the CAS Forum, we will be publishing additional committee work products. Please feel free to correspond with committee members about these work products.

Risk Theoretic Issues In The Discounting Of Loss Reserves

By The CAS Committee
On Theory Of Risk

RISK THEORETIC ISSUES IN THE DISCOUNTING OF LOSS RESERVES

A DISCUSSION PAPER BY THE CAS COMMITTEE ON THEORY OF RISK

BACKGROUND

The discounting of property/casualty loss reserves to reflect the time value of money has been a controversial issue for some time and recent activity in this area has been significant. In 1986 Congress passed landmark legislation to require discounting for income tax purposes. The National Association of Insurance Commissioners has formed a study group to further explore the advisability of discounting for statutory reporting purposes. Some state Insurance Departments have already begun to permit discounting in the statutory Annual Statement for some lines of business in which discounting had traditionally been prohibited. The AICPA is also studying the implementation of reserve discounting as it relates to GAAP financial reporting. Many insurance companies have been engaging in de facto discounting to some degree by means of overly optimistic reserving assumptions and/or by the purchase of financial reinsurance.

In the public debate over discounting it has been pointed out, though not always appreciated, that a fundamental feature of property/casualty loss liabilities is their uncertainty. Opponents of discounting have argued that carrying loss reserves on an undiscounted basis is in implicit recognition of this uncertainty or risk. According to this argument the amount by which undiscounted reserves exceed their discounted value provides a buffer against this uncertainty, a "risk margin" of sorts.

For several years now, the CAS Committee on Theory of Risk has been studying and discussing the issue of uncertainty in loss liabilities, particularly as it relates to the discounting of loss reserves. The Committee takes no official position on the discounting issue itself other than to agree with those observers who state that the issue can only be considered in the context of the purpose for which the financial statement is prepared; the issue can conceivably have a different resolution for statutory purposes, for example, than for tax purposes. Moreover, the Committee takes no official position on the proper accounting treatment to reflect uncertainty in reserves, regardless of the accounting context. Rather, our focus has been in the areas of: i) identifying the sources of uncertainty, ii) mathematically modeling and measuring the uncertainty, and iii) expressing the uncertainty in dollar value terms. We hope that this status report on our activities to date will be of value to those professional committees currently debating the discounting issue and its accounting treatment and also to the regulatory bodies ultimately responsible for the resolution of the debate. We also hope to receive feedback from these audiences to assist us in directing and focusing our further research.

FUNDAMENTAL ISSUES

The largest liability item on the balance sheet of virtually all insurance companies is also, arguably, the most uncertain. Often, the dollar amount of the liability for losses and loss adjustment expenses is not known until several years after the liability has been incurred and accounted for. This liability is subject to future uncertain events beyond the control of the insurance company, such as the socio-legal climate, jury sentiments, attitudes toward claim settlement, etc. that will prevail when the claims that give rise to the liability reach their ultimate disposition. A loss reserve is simply an estimate of this liability as of a given point in time, based on currently available information. These estimates are often in error. Since the amount of the loss reserve is typically several times the company's net worth, uncertainty in the reserve estimate can translate into considerably more uncertainty in the financial well-being of the company.

It is generally true that the reserves for the longer-tailed lines of business (i.e., those with greater-than-average time lags between claim incident and disposition) are the more uncertain. It is also a fact that these same lines of business present the greater opportunity for investment income on the assets supporting the reserves and thus for greater amounts of reserve discounting. There is some correlation then between reserve uncertainty and discount potential, and this gives some support to the idea that undiscounted reserves give implicit recognition to risk. The Committee believes that while this correlation exists it does not represent a sufficiently fundamental relationship to be used as a basis for measuring risk. It is, though, the Committee's position that discounting loss reserves does remove a substantial risk margin, however implicit and imprecise, and makes more pronounced the need to develop an explicit measure of risk.

Once a method for measuring and representing risk is developed, it remains to determine the proper method to report it in financial statements. As mentioned above, the resolution of this issue is outside the scope of the Committee's charter, however there are some considerations we would like to highlight for the benefit of those professional committees charged with this responsibility. A fundamental concern is whether a "risk margin" should be derived separately from the loss reserve and whether such a margin should be reported "above or below the line", i.e., as a liability item or as part of surplus. There are two different and somewhat conflicting accounting philosophies that influence the decision on how to report risk margins. According to one, the emphasis is on insurer solvency and on the balance sheet. Including a risk margin as a liability item (separately or not from the loss reserve) would be consistent with the conservatism inherent in this philosophy as it would serve to delay the flow of profits into surplus until the existence of such profits was sufficiently certain. The second philosophy has a going-concern emphasis and the focus is on the income statement. Including a risk margin as earmarked surplus is more consistent with this philosophy as it leaves losses "pure" and allows more direct matching of income and outgo. As is the case with the issue of discounting loss reserves, the Committee believes that the issue of accounting for risk margins depends on the purpose of the accounting document under consideration. A goal of our research is to provide methods of measuring and representing risk that will have sufficient flexibility to accommodate either of the above accounting philosophies.

SOURCES OF UNCERTAINTY

The sources of reserve uncertainty are many and arise principally from the following elements:

1. the ultimate value of claims reported but unpaid as of the evaluation date

2. the ultimate number and value of claims incurred but unreported as of the evaluation date

3. the ultimate value of claims closed as of the evaluation date but reopened subsequently

4. the payment timing of all unpaid claims for which a liability exists as of the evaluation date

5. investment yields

6. asset values

(Note that this list is not exhaustive.)

Contributing to the uncertainty surrounding these elements are:

- inflation

- judicial and legal climate

- changes in company practice, e.g., with respect to:
  - asset management
  - claims administration

- currency fluctuations

- the interaction of the various items, e.g.:
  - interest rates vs. inflation
  - claim severity vs. payment lag

SYNOPSIS OF COMMITTEE ACTIVITIES

The Committee has examined a number of approaches for modeling and measuring risk in loss reserving, some promising, some not so promising. We believe that a discussion of all approaches considered should be included here since the reasons for deciding against some of them may provide some insight to readers.

We have discussed whether risk could be measured by means of an empirical study of loss development history. Some methods along these lines have already been developed by practicing actuaries. These include measuring variations in historical age-to-age loss development factors and modeling the factors by means of distribution functions. These methods are relatively straightforward and the necessary data is easy to obtain. However, methods based only on historical development data are likely to underestimate potential future variation since, in simple terms, not everything that could have happened has happened. On the other hand, the potential for adverse development could be overstated in the historical data since recent adverse development may be more reflective of earlier implicit discounting than of failure to reserve correctly. The Committee believes that historical development patterns alone are not sufficient to measure reserving risk but that this history is invaluable in testing and validating the models we will discuss below.

We discussed whether risk could be measured in terms of mean and variance concepts. We also discussed whether estimating a given percentile of the distribution of losses could be sufficient to quantify risk. For several reasons, the Committee believes these measures are insufficient. Many important aspects of a probability distribution are not captured by the first two moments or by a given percentile. (For example, very different excess loss premium factors can be generated from two different loss distributions that happen to have the same first two moments.) This discussion did convince us of the importance of estimating the complete distribution of ultimate aggregate losses before attempting to quantify risk.

A discussion of the construction of such an aggregate loss distribution including treatment of the risks associated with investment yields and the timing of loss payments is presented in the Appendix.

We have discussed approaches by which the distribution of loss liabilities (discounted or undiscounted), assuming this distribution could be determined, would be incorporated into the quantification of risk. One approach popular in European countries is ruin theory. In the reserving applications of this theory, the loss distribution is incorporated into a stochastic financial model of the entire insurance company and the company's surplus is considered to be stochastic process over time. The appropriate loss reserve incorporating reflection of risk is the smallest amount such that the probability of the company's technical insolvency is reduced to a specified level. One distinct advantage of this approach is that the implied necessary risk load is not independent of the company's current financial condition. There are some practical problems with ruin theory, however. The selection of an acceptable probability of ruin is problematical. U.S. company managements are understandably uncomfortable with the concept of an "acceptable probability of ruin". Also, the risk load determined via ruin theory is extremely sensitive to the choice of the probability of ruin.

One approach which offers the prospect of incorporating what can be learned from ruin theory (for example, the use of the entire loss distribution, and the financial modeling of the entire company) for determining risk-adjusted reserves is utility theory. An acceptable ruin probability need not be specified, since utility theory assigns a utility function to the entire continuum of financial outcomes. Once the distribution of aggregate losses has been estimated, utility theory can be used to compute its "certainty equivalent". This is the loss amount which, if known with certainty, would be regarded as equivalent to the uncertain distribution of outcomes. Specifying the utility function is non-trivial as is the question of whose utility function to model (shareholders, management, regulators, etc. would generally have different utility functions). Moreover, deriving a single utility function to represent a consensus among people with similar viewpoints (e.g., shareholders) is a problem still not fully solved.

[Digression: The capital asset pricing model (CAPM) was discussed by the Committee and discarded as an explicit means of reflecting risk in reserves, however the discussion did identify a concept that might be useful to those committees addressing the issue of accounting for risk margins. In CAPM theory, a central maxim is that "diversifiable risk" should not be "rewarded". In the context of loss reserving, the corresponding rule might be that margins arising from "diversifiable risk" (e.g., due to the use of poor reserving techniques) should not be reported "above the line" but should be reflected in a segregated surplus account.]

SUMMARY OF CURRENT COMMITTEE OPINIONS

As a result of our research and discussions to date, the Committee has formed the following opinions:

- Regardless of the method by which reserves are discounted and uncertainty is measured, and regardless of the accounting treatment, full disclosure in public documents of the methods, measurements and treatments is advisable.

- Measurement of the uncertainty in loss liabilities is an essential part of the estimation of those liabilities, regardless of the context in which the liability estimates and risk measurements are presented. The discounting of loss reserves, by eliminating the implicit risk margin, makes the need for explicit measurement of risk more pronounced.

- While the ultimate application of the theories the Committee is developing may take the form of simple rules of thumb, it is necessary to more fully develop the theory (including a reasonable methodology for estimating the complete distribution of loss liabilities and a start on building a comprehensive financial model) before such rules can be promulgated.

- The development of the necessary theory is a long-term effort, but events, accelerated now by the discounting issue, will not await the perfect theory. The Committee recognizes that, as a practical matter, methods may need to be introduced prior to the full development of the underlying theory. The Committee hopes that the ideas presented herein will assist other bodies (actuarial, accounting, regulatory, etc.) in the development of those methods and further pledges its intention to be actively involved in the effort.

FUTURE DIRECTIONS

The Committee intends to pursue the development of methods for the quantification of risk. To this end, work is under way to:

- estimate probability distributions for the items listed above under "sources of uncertainty"
- develop an overall company stochastic model to incorporate these distributions
- determine a method for calculation of a risk margin from this model

These are clearly long-term projects. In this effort, and in the development of practical alternatives in the intermediate term, we expect to work closely with (at least) the CAS Committee on Reserves, the CAS Committee on Financial Analysis and the AAA Committee on Property and Liability Financial Reporting Principles.

CAS COMMITTEE ON THEORY OF RISK

Gary Patrik, Chairman	Stephen Philbrick
Roger Hayne	Lewis Roberts
Glenn Meyers	Gary Venter
Jerry Miccolis	Richard Woll

CAS COMMITTEE ON THEORY OF RISK
DISCUSSION DRAFT ON DISCOUNTING OF LOSS RESERVES

APPENDIX

COMMENTS UPON MODELING DISCOUNTED AGGREGATE LOSS LIABILITIES

The loss process can be thought of being made up of many probability distributions, arising from all the sources of uncertainty mentioned in the text. From the point of view of setting an appropriate discounted loss reserve for the current loss liability, we may be interested in a representation of the liability L such as the following:

$$L = V(1)*L(1) + V(2)*L(2) + \ldots$$

where L(i) is the aggregate loss to be paid in the ith year, and V(i) is an appropriate discount factor to present value. (Obviously, time periods other than one year can be used.)

The V(i)'s may have at least three different meanings:

1. The V(i)'s could be what the IRS tells you they are.

2. The V(i)'s could depend upon the asset portfolio supporting the loss reserve and upon future investment returns.

3. The V(i)'s could be the current utility value to you of future payments to be made by you.

The L(i)'s can be modeled by first writing each as the sum of individual L(i)'s for fairly homogeneous exposure groups, accident years, etc. Let us assume that L(i) now represents such a piece of the total. Traditional risk theory models the aggregate loss process by modeling claim counts and amounts and taking the obvious sum:

$$L(i) = X(i;1) + X(i;2) + \ldots + X(i;N(i))$$

where N(i) = number of claims (or occurrences)
and X(i;j) = amount of the jth claim

Given appropriate models for N(i) and X(i;j) and suitable independence assumptions, we can write the moments of L(i) in terms of the moments of N(i) and X(i;j), and we can approximate the distribution of L(i). There are many good papers in the actuarial literature about this.

An advantage of using a claims count/claims severity model is that we can contemplate intuitively satisfying models for various lag distributions, such as the time from loss event occurrence until first report, the time from first report until payment, etc.. And an appropriate model for the claim count could be constructed as follows:

> Suppose that the commonly used Poisson distribution, with parameter n say, is a good model for the total claim count N. Then the number of claims settled in the ith year N(i) will also be Poisson with parameter n*p(i), where p(i) is the lag probability for the ith year, that is, p(i) is the probability that a claim will settle i years after occurrence.

Thus the aggregate losses paid in the ith year of run-off can be modeled via the standard risk theoretic model under suitable assumptions for the claim sizes. This kind of model also allows us to better understand claim size reserves under changing conditions, such as changing policy limits or changes in retentions net of reinsurance. This model is a powerful tool for describing loss liability.

COMMENTS UPON USING PRICING ASSUMPTIONS FOR RESERVING

The loss payment run-off and thus the loss reserves for a given coverage year should relate to the original pricing model distributions as conditional distributions. Suppose that the original pricing model for the loss process said that the total loss payments would have a certain distribution F and that the loss payment run-off would be according to some time series < F(L(t)) >. As of a any time t thereafter, the information on reported and paid and settled claims should conditionalize the original distributional assumptions in order to update future loss payment predictions.

CTR72C.DOC
3/10/87

PAPER PRESENTED AT

1987 RATEMAKING SEMINAR

A very successful ratemaking seminar was held in March of 1987. A number of interesting papers were presented at this seminar. We have chosen one of these papers to be in this issue of the CAS Forum.

Generally, the CAS Forum will publish documents from various seminars that are felt to be of interest to the entire Casualty Actuarial Society. These documents will generally be accepted in the form in which the author initially presented them. This should improve our ability to publish these documents quickly in order to disseminate the available information.

The issue of ratemaking for underinsured motorists coverage has increased in importance as a result of the increasing availability of underinsured motorists coverage and continued pressure on the tort system. It is also a somewhat complicated coverage because assumptions must be made regarding the distribution of limits for vehicles in the general population.

This presentation provides an interesting viewpoint on some of the techniques useful in pricing underinsured motorists coverage.

Underinsured Motorist Coverage Pricing Models

By Gary Grant

UNDERINSURED MOTORIST COVERAGE

PRICING MODELS

BY GARY GRANT

Underinsured Motorist Coverage enables a person to have coverage for bodily injury in the event they are injured by an individual with liability limits inadequate to cover the damages. As recently as 15 years ago, Underinsured Motorist Coverage was relatively unknown. Now, 41 jurisdictions require some form of the coverage.

There are two basic types of coverage -- a difference in limits form of coverage that simply provides a limit of liability that is the difference between the underinsured coverage limit of liability and the tort-feasor's Bodily Injury liability limit. The second form is an excess form of coverage which provides a layer of coverage on top of the Bodily Injury liability limits of the tort-feasor. The two components that define the type of coverage are (1) when the coverage shall be provided or the "trigger" and (2) the limit of liability.

Underinsured Motorist Coverage is a long tail, low frequency, high average cost coverage for which little data is available for pricing. Depending on the form of coverage, the claim frequency can be as low as 1% of the BI liability frequency whereas the average cost may equal 5 to 6 times the BI liability average cost. The purchaser of Underinsured Motorist Coverage essentially buys a portion of the BI liability coverage for the tort-feasor. With this in mind, one approach to pricing the coverage is to use the BI liability coverage data. Individual company data may be used to determine the rates. However, since the tort-feasor could be anyone from the insured population, industry data may be more appropriate.

In the models that follow, the information needed to price the Underinsured Motorist Coverage is (1) the indicated BI liability rates, (2) the percent of the uninsured population, (3) the BI limits factors and (4) the BI limits distribution. For this type of information, industry data is at best difficult and in some cases impossible to obtain. It is generally necessary to use company data or a mixture of company and industry data. For example, the All Industry Research Advisory Committee study included bodily injury limits distributions. One option could be to use this distribution along with the available company data.

Following are four pricing models ranging from a simple difference in limits form of coverage to a rather complicated excess form of coverage. For each model, that portion of the law that defines the coverage is also shown. (It should be noted that these are not the only forms of coverage available. A careful reading of the law to determine the "trigger" and the limit of liability is necessary to determine the form of coverage.) I've chosen to show the Underinsured Motorist Coverage price in terms of dollars and cents rather than as a factor applied to some base premium. Either approach could be used.

UNDERINSURED MOTORIST COVERAGE LAWS
Standard Difference in Limits Model

Situation That Triggers The Coverage:

". . .Where the limits of coverage available for payment to the insured under all bodily injury liability bonds and insurance policies covering persons liable to the insured are less than the limits for the insured's uninsured (underinsured) motorists coverage at the time of the accident."

Limits of Liability

"The limits of liability for an insurer providing underinsured motorist coverage shall be the limits of such coverage, less those amounts actually recovered under all applicable bodily injury liability bonds and insurance policies covering persons liable to the insured."

UNDERINSURED MOTORIST COVERAGE PRICING
STANDARD DIFFERENCE IN LIMITS MODEL

Assume: 10% uninsured
Indicated 15/30 BI Rate = $50

(1) UIM Limit	(2) Industry BI Limits Dist.	(3) BI Limits Factors	(4) Average Limit Factor For Tortfeasor w/ Lower Limit	(5) Relative UIM Exposure	(6) Relative Average Cost	(7) Expo x Cost	(8) Indicated Additive
15/30	30%	1.00	0.000	.00	1.000	.000	$ 0.00
20/40	5	1.10	1.000	.30	.100	.030	1.35
25/50	20	1.20	1.014	.35	.186	.065	2.93
50/100	20	1.40	1.082	.55	.318	.175	7.88
100/300	25	1.60	1.167	.75	.433	.325	14.63

(4) = Average of Col (2) x Col (3) for all Lower Limits
(5) = Sum of Col (2) for all Lower Limits
(6) = Col (3)−Col (4)
(7) = Col (5) x Col (6)
(8) = Col (7) x Indicated BI Rate x (1−% Uninsured)

UNDERINSURED MOTORIST COVERAGE LAWS
Standard Excess Coverage Model

Situation That Triggers The Coverage:

". . .(When) an injured person . . . agrees to settle a claim with a liability insurer and its insured for the limits of liability, and such settlement would not fully satisfy the claim for personal injuries or wrongful death. . ."

Limits of Liability:

". . .And (uninsured motorist coverage) shall cover the difference, if any, between the sum of (all benefits available) and the damages sustained, up to the maximum amount of such coverage provided under this section. The amount of coverage available under this section shall not be reduced by a setoff against any coverage, including liability insurance."

UNDERINSURED MOTORIST COVERAGE PRICING
STANDARD EXCESS COVERAGE MODEL

Assume: 10% uninsured
Indicated 15/30 BI Rate = $50

(1)	(2) TORTFEASOR'S BI LIMIT AND DISTRIBUTION						(3)	(4)
	BI Limit:	15/30	20/40	25/50	50/100	100/300		
UIM	BI Lim Factor:	1.00	1.10	1.20	1.40	1.60	Weighted	Indicated
Limit	Ind. BI Dist:	30%	5%	20%	20%	25%	Average	Additive
15/30	Total Limit	30/60	35/70	40/80	65/130	115/330		
	BI Factor	1.25	1.30	1.35	1.45	1.65		
	Excess Cost	.25	.20	.15	.05	.05	.138	$ 6.21
20/40	Total Limit	35/70	40/80	45/90	70/140	120/340		
	BI Factor	1.30	1.35	1.35	1.50	1.65		
	Excess Cost	.30	.25	.15	.10	.05	.165	7.42
25/50	Total Limit	40/80	45/90	50/100	75/150	125/350		
	BI Factor	1.35	1.35	1.40	1.50	1.65		
	Excess Cost	.35	.25	.20	.10	.05	.190	8.55
50/100	Total Limit	65/130	70/140	75/150	100/200	150/400		
	BI Factor	1.45	1.50	1.50	1.60	1.70		
	Excess Cost	.45	.40	.30	.20	.10	.280	12.60
100/300	Total Limit	115/330	120/340	125/350	150/400	200/600		
	BI Factor	1.65	1.65	1.65	1.70	1.80		
	Excess Cost	.65	.55	.45	.30	.20	.423	19.04

(2) Ind. BI Dist = Industry BI Distribution
Total Limit = UIM Limit + Tortfeasor's BI Limit
Excess Cost = BI Factor–Tortfeasor's BI Limit Factor
(3) = Industry BI Distribution x Excess Cost
(4) = Col (3) x Indicated BI Rate x (1–% Uninsured)

UNDERINSURED MOTORIST COVERAGE LAWS
Special Excess Limits Model #1

Situation That Triggers The Coverage:

"An underinsured motor vehicle is one for which there may be bodily injury liability in effect, but the limits of bodily injury liability coverage under all bonds and insurance policies applicable at the time of the accident total less than the limits provided by the uninsured (underinsured) motorist coverage."

Limits of Liability:

"Acceptance of (higher limits uninsured and underinsured motorist) coverage shall operate to amend the policy's uninsured coverage to pay for bodily injury damage that the insured or his legal representative are legally entitled to recover from the driver of an underinsured motor vehicle."

UNDERINSURED MOTORIST COVERAGE PRICING
SPECIAL EXCESS COVERAGE MODEL #1

Assume: 10% uninsured
Indicated 15/30 BI Rate = $50

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
UIM Limit	Industry BI Limits Dist.	BI Limits Factors	Ave. Limit Factor For Tortfeasor w/ Lower Lim.	Relative UIM Exposure	Ave. Tot. Limits Factor	Relative Average Cost	Expo. x Cost	Indicated Additive
15/30	30%	1.00	.000	.000	0.000	0.000	.000	$ 0.00
20/40	5	1.10	1.000	.300	1.300	.300	.090	4.05
25/50	20	1.20	1.014	.350	1.350	.336	.118	5.31
50/100	20	1.40	1.082	.550	1.473	.391	.215	9.68
100/300	25	1.60	1.167	.750	1.663	.496	.372	16.74

(4) = Average of Col (2) x Col (3) for all Lower Limits
(5) = Sum of Col (2) for all Lower Limits
(6) = Average Limit Factor for Sum of UIM Limit and all Lower Limits.
This is developed on the next exhibit.
(7) = Col (6)–Col (4)
(8) = Col (7) x Col (5)
(9) = Col (8) x Indicidated BI Rate x (1–% Uninsured)

UNDERINSURED MOTORIST COVERAGE PRICING
SPECIAL EXCESS MODEL #1
AVERAGE TOTAL LIMITS FACTOR

(1)	(6A) TORTFEASOR'S LIMIT AND DISTRIBUTION						(6)
UIM Limit	BI Limit Industry BI Dist:	15/30 30%	20/40 5%	25/50 20%	50/100 20%	100/300 25%	Ave. Total Limits Factor
15/30	Total Limit	N/A	N/A	N/A	N/A	N/A	
	BI Factor						0.000
20/40	Total Limit	35/70	N/A	N/A	N/A	N/A	
	BI Factor	1.30					1.300
25/50	Total Limit	40/80	45/90	N/A	N/A	N/A	
	BI Factor	1.35	1.35				1.350
50/100	Total Limit	65/130	70/140	75/150	N/A	N/A	
	BI Factor	1.45	1.50	1.50			1.473
100/300	Total Limit	115/330	120/340	125/350	150/400	N/A	
	BI Factor	1.65	1.65	1.65	1.70		1.663

(6A) Total Limit = UIM Limit + Tortfeasor's BI Limit when UIM Limit is less than Tortfeasor's BI Limit.

(6) = Weighted Average of BI Factors and Industry BI Distribution.

UNDERINSURED MOTORIST COVERAGE LAWS
Special Excess Coverage Model #2

Situation That Triggers The Coverage:

"(Underinsured) motorists insurance shall provide coverage. . . if the limits of liability under all bodily injury liability bonds and insurance policies of another motor vehicle liable for damages are in a lesser amount than the bodily injury liability insurance limits of coverage provided by such policy."

Limits of Liability:

"Any such policy shall, at the option of the insured, also provide (underinsured) motorists insurance for bodily injury, in an amount up to the bodily injury liability insurance limits of coverage provided under such policy, subject to a maximum of (100/300)."

UNDERINSURED MOTORIST COVERAGE PRICING
SPECIAL EXCESS COVERAGE MODEL #2

Assume: 10% uninsured
Indicated 15/30 BI Rate = $50

(1)	(2) TORTFEASOR'S LIMIT						(4)	(5)
	BI Limit:	15/30	20/40	25/50	50/100	100/300	Average	
UIM	BI Lim Factor:	1.00	1.10	1.20	1.40	1.60	Cost X	Indicated
Limit	Ind. BI Dist:	30%	05%	20%	20%	25%	Exposure	Additive
15/30	Total Limit	30/60	35/70	40/80	65/130	115/330	.108	$ 4.86
	BI Factor	1.25	1.30	1.35	1.45	1.65		
	Excess Cost	.25	.20	.15	.05	.05		
20/40	Total Limit	35/70	40/80	45/90	70/140	120/340	.137	6.17
	BI Factor	1.30	1.35	1.35	1.50	1.65		
	Excess Cost	.30	.25	.15	.10	.05		
25/50	Total Limit	40/80	45/90	50/100	75/150	125/350	.162	7.29
	BI Factor	1.35	1.35	1.40	1.50	1.65		
	Excess Cost	.35	.25	.20	.10	.05		
50/100	Total Limit	65/130	70/140	75/150	100/200	150/400	.242	10.89
	BI Factor	1.45	1.50	1.50	1.60	1.70		
	Excess Cost	.45	.40	.30	.20	.10		
100/300	Total Limit	115/330	120/340	125/350	150/400	200/600	.373	16.79
	BI Factor	1.65	1.65	1.65	1.70	1.80		
	Excess Cost	.65	.55	.45	.30	.20		

(2) Excess Cost = BI Factor–Tortfeasor's BI Limit Factor
(4) See Col (3) on next exhibit for calculation
(5) Col (4) x Indicated BI Rate x (1–% Uninsured)

UNDERINSURED MOTORIST COVERAGE PRICING
SPECIAL EXCESS COVERAGE MODEL #2
CALCULATION OF AVERAGE COST X EXPOSURE

(1)	(3)						(4)
	SELECTED BI LIMIT AND DISTRIBUTION						
UIM Limit	BI Limit Co. BI Dist.	15/30 5%	20/40 5%	25/50 15%	50/100 25%	100/300 50%	Weighted Average
15/30		.000	.075	.085	.115	.125	.108
20/40		N/A	.090	.103	.133	.153	.137
25/50		N/A	N/A	.118	.158	.178	.162
50/100		N/A	N/A	N/A	.215	.255	.242
100/300		N/A	N/A	N/A	N/A	.373	.373

(3) BI Limit = BI Limit of Injured Party
Co. BI Dist. = BI Limits Distribution of Injured Party's Company
= Sum of: (Col (2) Excess Cost x Col (2) Industry BI Dist.) where Tortfeasor's Limit is less than Selected BI Limits

(4) = Weighted Average of each row with ~~selected~~ Co. BI Distribution

DRAFT COPY OF A CHAPTER

FROM THE CASUALTY ACTUARIAL SOCIETY TEXTBOOK

The Textbook Steering Committee and individual authors have been in the process of drafting chapters for the Casualty Actuarial Society Textbook. We currently have in process a number of chapters. We expect that we will complete work on this during fiscal 1988.

In this issue of the CAS Forum, we are publishing the draft of the chapter on Credibility. This provides an opportunity to our membership to give the author comments on the chapter. It also provides an opportunity to get an idea of the level of material that will be in the Casualty Actuarial Society Textbook. We encourage you to provide substantive comments on the chapter to Gary Venter and format and other comments regarding the textbook to Irene Bass, Chairman of the Textbook Steering Committee.

Credibility

By Gary Venter

CREDIBILITY

BY GARY VENTER

Section 1 - Introduction and History

Until recent years, classical statistics had focussed on estimating a quantity based only on directly relevant observations; peripherally relevant or seemingly unrelated series which may provide further information had been excluded. Since at least the early 1900's, however, casualty actuarial practice has incorporated related information, sometimes in a fairly ad hoc manner, under the name of "credibility."

Classical statistical procedures estimate a value, such as the average age of a group, by taking a sample from the group and using the mean value of that sample as the estimate. Credibility estimation makes use of the sample value, but may incorporate other information as well, such as the average age of similar groups. In ratemaking, for example, the experience of the latest period might be regarded as a sample from all possible time periods. Rather than using this by itself, even properly adjusted for premium and loss levels, to determine the new rate, other information might be incorporated, such as the old rate, or rates for related exposures.

If the new rate is taken to be a weighted average between the indication from the data and the old rate or some other estimate, the weight applied to the data is called the credibility weight, or sometimes, more loosely, the credibility of the data. The latter terminology may be misleading, however, in that it seems to imply that the credibility weight is an inherent property of the data. This will turn out to not be the case. In addition to any features of the data itself, the context in which it will be used, including what it is to be weighted against, will explicitly or implicitly influence the credibility to be assigned.

Credibility theory incorporates the entire study of this weighting process, including development of the formulas for assigning the credibility weights, as well as estimation of the parameters or values that appear in these formulas.

Although pragmatically motivated, credibility weighting now has both theoretical and practical justification. Credibility formulas can be derived from statistical assumptions, and they have proven useful in application. This chapter outlines the background and use of credibility theory. Being an overview, the results are in many cases given without proof, or the proofs are just outlined. Underlying assumptions are included, however. As with many disciplines, the real world is often more complex than the initial assumptions, and more intricate models are often needed in order to be truly practical. The more practical models are presented in the later sections, but their exposition will benefit from the simpler paradigms covered first.

To illustrate the type of related information that may be useful, imagine that an estimate is desired for the quantity of ice cream a particular person will consume next year. The average consumption for that individual for the last few years might be selected as the estimate. However this estimate could probably be improved by giving some degree of weight to the average consumption of the population at large.

Another example, examined in greater detail below, is to estimate a baseball player's season batting average from the early season performance. In this case it has been shown that giving weight to the early season averages of other players can considerably improve the estimate.

More typical insurance examples include estimation of claim frequency, severity, or total loss cost for an insured, a class, or a rating territory. Experience for other insureds, classes, states, insurers, etc. may be the auxiliary data incorporated.

In all of the above cases the auxiliary information comes from a wider, more stable population. There is, however, another type of credibility application; rather than incorporating a wider population, earlier observations of a single series may be used. For something like claim frequency countrywide, for example, the latest observation by itself could be regarded as sufficient; however if this is subject to significant random fluctuations, some weight may be given to prior years, perhaps with the weights decreasing to zero after some time. Credibility formulas used in this case are somewhat different from those incorporating a wider population.

Limited Fluctuation vs. Greatest Accuracy

By the mid 1920's, two fairly different approaches to credibility had been established. The terminology noted was introduced by Arthur Bailey in his far reaching 1945 paper. Basically speaking, the limited fluctuation approach aims to limit the random component of an estimate; the greatest accuracy approach attempts to make the estimation error as small as possible. The example below shows how each of these might be applied in a single series case.

The series in question, N_i, could be anything of interest, e.g., state loss ratios, countrywide frequency, etc. To have a concrete example, let N_i denote the number of doctor visits made by members of the U.S. Congress in year i. C_i will denote the credibility estimate of N_i made based on the data through year i-1. In a single series situation, both the limited fluctuation and greatest accuracy approaches to credibility make use of a credibility weight Z_i between 0 and 1 so that:

$$C_{i+1} = (1-Z_i)C_i + Z_iN_i \quad (1.1)$$

This weight, however, has a different purpose and derivation in the two approaches.

The limited fluctuation approach seeks to limit the fluctuations in the series of estimates C_j, at least insofar as those fluctuations are due to the randomness inherent in the series of observations N_j. The greatest accuracy approach, on the other hand, seeks to minimize the estimation errors. To be more precise, this approach specifies then seeks to minimize an error function. Usually the expected squared difference between the estimated and actual value is the function to be minimized. In the current example this would be denoted as $E(C_i-N_i)^2$. With this error function, the greatest accuracy approach is referred to as "least squares" credibility.

To illustrate the formulas for Z_i that arise from these two approaches, a few additional assumptions will be introduced. N_i is hypothesized to be approximately normally distributed with mean M_i and constant variance v. (Constant in that it is the same for each year.) The mean M_i is hypothesized to change each year by the random amount D_i, that is, $M_{i+1}=M_i+D_i$. D_i is a random variable with mean zero and variance d. The D's are assumed to be independent of each other and of M_1. Because of the mean zero, each M_j has the same unconditional expected value, denoted by m, i.e., $E(M_i)=m$. (M_i is treated as a random variable because its value is not known, which in part is due to the random term D_i.) Because the M's change each year by the D's, the variance of M increases each year by d. E.g., $Var(M_{i+1})=Var(M_i)+d$. Thus $Var(M_j)=w+(j-1)d$, with $w=Var(M_1)$.

The estimation process has to start somewhere, and so C_1 is the estimate of N_1 made before any of the observations N_j are available. This estimate could have been based on previous knowledge of similar processes, for example. C_1 can also be considered as an estimate of M_1. The variance w can be interpreted as an expression of the uncertainty about the value of M_1 before N_1 is observed; as such it may influence how willing we will be, when estimating N_2, to give up on C_1 in favor of N_1 once it becomes available.

Given these model assumptions, the calculation of the credibility factor Z_i under the two approaches can be addressed. The limited fluctuation approach calculates Z_i based on the conditional distribution of N_i given M_i. It seeks to limit the impact on the credibility estimator (1.1) of random deviations of the observation N_i from its conditional expected value M_i. In other terms, it seeks to guarantee, at least to an acceptably high probability, that the quantity:

$$Z_i(N_i - M_i) \quad (1.2)$$

stays within certain bounds.

The criterion for limiting the deviation is established by first specifying a probability level p, e.g., p=.95, and then requiring, with a probability of at least p, that (1.2) be no greater than some prespecified maximum. In this case that maximum will be taken to be km, where k is a selected small number, e.g., k=.05. Recall that m is the unconditional mean $E(M_i)$. In other words, Z_i is sought so that $Pr\{Z_i(N_i-M_i)\leq km\}=p$.

To see the impact of this criterion, the credibility estimate (1.1) can be rewritten as $C_{i+1}=(1-Z_i)C_i+Z_iM_i+Z_i(N_i-M_i)$. These three terms can be regarded as representing stability, truth, and random noise. Since truth and noise cannot be observed separately, the same factor Z_i applies to both. The highest possible factor is sought, so that truth will be emphasized, as long as noise can be kept within acceptable bounds. Thus the value of Z_i is sought that will keep $Z_i(N_i-M_i)$ below 100k% of the expected value m with probability p.

Since N_i has a symmetric distribution about its mean M_i, that Z_i will also ensure that the absolute value of the random component is less than km with probability 2p-1. Limited fluctuation credibility as so formulated emphasizes the conditional distribution of N_i given M_i, but the conditioning is not always noted explicitly.

The value of Z_i that meets the criterion is $Z_i= km/y\sqrt{v}$, where y is the 100pth percentile of the standard normal distribution. To show this, by hypothesis, given M_i, $(N_i-M_i)/\sqrt{v}$ has the standard normal distribution, and so from the definition of y, we have

$Pr[(N_i - M_i)/\sqrt{v} < y] = p$. Multiplying both sides of the inequality by $Z_i\sqrt{v}$ then gives $(Pr[Z_i(N_i - M_i) < km] = p$, as desired.

In most applications of this approach, Z_i is regarded as a function of m, and the values of m that lead to different credibility levels are sought. The variance v is often taken to be proportional to m, e.g., v=cm. This yields $Z_i = (k/y)\sqrt{m/c}$.

Z_i is capped at 1, even though the formula value may be higher. For selected p and k, the value of m that yields $Z_i=1$ is referred to as the full credibility value, and is given by $m_F = c(y/k)^2$. Then for $m < m_F$, Z_i can be conveniently computed by the **square root rule**:

$$Z_i = \sqrt{m/m_F} \qquad (1.3)$$

which can be verified by substituting $c(y/k)^2$ for m_F in (1.3).

The least squares approach for determining Z_i does not start with formula (1.1), but derives it as the result of a more general estimation problem: N_{i+1} is to be estimated as a linear combination of the previous observations $N_1, \ldots N_i$, with the expected squared error to be minimized. That is, coefficients b_j are sought to minimize:

$$E[N_{i+1} - (b_0 + \Sigma_{j=1}^{i} b_j N_j)]^2 \qquad (1.4)$$

It turns out, after much algebra, that the solution to this estimation problem can be expressed in the form (1.1), that is, as a credibility formula.

The Z_i that do this are computed recursively by:

$$Z_1 = 1/[1 + K] \tag{1.5a}$$

$$Z_{i+1} = 1/[1 + 1/(J+Z_i)] \tag{1.5b}$$

where K=v/w and J=d/v. The details of the derivation, including more general conditions under which (1.4) leads to (1.1), are found in Gerber and Jones (1974). For the interested reader, a sketch of the proof is below.

The minimization of (1.4) is accomplished by first setting its partial derivatives with respect to the b_j to zero. This produces i+1 equations, one for each b_j. For example the partial of (1.4) with respect to b_1 produces the equation:

$$E[N_1(N_{i+1}-b_0-\Sigma_{j=1}^{i} b_j N_j)]=0.$$

All these equations involve terms like $E(N_j)$ and $E(N_j N_h)$, which are then evaluated in order to solve for the b_j's.

To illustrate this procedure, evaluating the $E(N_j N_h)$ type term is outlined. Note that, given M_j, N_j and N_{j+h} are independent, and the conditional expected value of each is M_j, i.e., $E(N_j|M_j)=M_j$ and $E(N_{j+h}|M_j)=M_j$. This is because $M_{j+h}=M_j+D_j+...+D_{j+h-1}$, and the D_i's have mean zero. Then it follows that $E(N_j N_{j+h}|M_j)= M_j^2$, and eventually that $E(N_j N_{j+h})= Var(M_j)+m^2= w+(j-1)d+m^2$. After evaluating all such terms and combining them algebraically, (1.5) is produced. (End of sketch of proof).

A heuristic interpretation of (1.5) can be made. Note that Z_1 is an increasing function of w and a decreasing function of v. The uncertainty about M_1 is measured by w; thus the greater this uncertainty, the greater is the weight given to the observation N_1. But the uncertainty about M_1 is not the only thing considered, the stability of N_1 is greater with lower v and this also leads to greater weight on N_1.

Z_{i+1} is an increasing function of Z_i and J. Greater stability (low v) continues to give greater credibility through a higher J; a higher d also increases the credibility which makes sense as follows: a high d indicates that the M's are greatly subject to change, so the older estimates should be given less weight, with more to the current observation, i.e., higher Z_j's.

It might also be noted that if d happens to equal $w^2/(v+w)$, all the Z's are the same. This can be verified by finding Z_2 from (1.5); it seems an unlikely coincidence, however.

The formulas above are fairly representative of what is produced by the least squares and limited fluctuation approaches to credibility. The limited fluctuation approach will always involve a full credibility value, representing the degree of random fluctuation deemed acceptable. The square root rule for partial credibility is also fairly typical of this approach. The only variance explicitly treated is v, which represents the random fluctuation of a single observation around its own generally unknown mean.

Formula (1.5a) is fairly typical of least squares credibility; often it is somewhat generalized to $Z=P/[P+K]$, where P is a measure of the volume of data observed. Besides recognizing the random fluctuation measured by v, this formula also incorporates the relevancy of the previous estimate, which w quantifies. Also quantifying changes in the process over time, which d achieves, is a further step not always incorporated into least squares analysis. Thus (1.5b) is a less typical but more general example of a least squares credibility formula.

It should be noted that while the limited fluctuation approach does not explicitly recognize the relevance of the previous estimate or the degree of likely process change over time, judgments about these issues may be incorporated into the selection of the degree of random fluctuation deemed acceptable, as specified by the fluctuation k allowable with probability p.

Historical Perspective

Credibility as known today is generally traced to Mowbray (1914), writing in volume I of the *Proceedings of the Casualty Actuarial Society*. Decidedly in the limited fluctuation camp, Mowbray's article approximates an assumed binomial claim count process by the normal distribution to derive the full credibility standard relative to p and k. The goal of the limited fluctuation approach as practiced today is suggested by the title of Mowbray's article: "How extensive a payroll exposure is necessary to give a dependable pure premium?"

The greatest accuracy approach was introduced by Whitney (1918), writing in volume IV of the CAS Proceedings. Whitney assumed that the number of claims for an employer with P employees is binomially distributed with parameters (P,M), and that M itself is normally distributed. The resulting credibility for that employer's experience can be expressed as $Z=P/[P+K]$, with K a function of the binomial and normal variances. The complement 1-Z is applied to the experience of the entire class, as indicated by the class rate, rather than to the previous experience of the employer.

Both Mowbray and Whitney were addressing Workers Compensation experience rating. An application of the limited fluctuation paradigm to automobile classification ratemaking can be found in Stellwegen (1925). Group life insurance experience rating using greatest accuracy credibility was explored by Keffer (1929), who assumed a Poisson claim count distribution with a gamma distribution on the Poisson parameter. Perryman (1932) addressed a number of then current issues, including an interpretation of the limited fluctuation square root rule similar to that discussed above, i.e., a way to give the credibility estimate no larger a random component than a risk with full credibility would have.

The least squares approach to greatest accuracy credibility was established in Bailey (1945), although the notation was cumbersome. Buhlmann and Straub (1970) formalized the derivation of $Z=P/[P+K]$, $K=v/w$, from a least squares error criterion, showed that this was valid for all finite variance distributions, and discussed a method of estimating the variances v and w.

Least squares credibility was recognized by Bailey (1950) to replicate the Bayesian posterior mean for the normal-normal and beta-binomial models. Keffer's result essentially shows this for the gamma-Poisson case, and it is also known for the gamma-gamma pair. Essentially, the posterior mean is the best least squares estimator; credibility provides the best **linear** least squares estimator. Thus when the posterior mean is a linear function of the observations, the two estimators are the same. Ericson (1970) characterized a family of distributions for which this is the case.

More recent research has emphasized generalizations and applications of the original models. Topics include improved estimation of parameters, credibility for trend and regression models, credibility incorporating more than one type of prior estimate, credibility weighting of the prior with a "hyperprior", methods of incorporating more complex relationships between firms of different sizes, methods of improving estimates for distributions with nonlinear posteriors, and treating parameters that may change over time. Many of these generalizations arise because the real world is more complicated than the original models assume; thus in addition to requiring more theory, they are more practical as well.

Section 2 - Review of Statistical Concepts

An understanding of some basic statistics will be presumed in this chapter. A few of the topics most germane to credibility theory will be briefly reviewed in this section, but reference to statistical texts may be required if some material has been unused recently.

Two concepts that will be called upon frequently are covariance and conditional distributions. To review, for two random variables X and Y, the covariance of X and Y is defined as:

$$Cov(X,Y) = E[(X-EX)(Y-EY)] \qquad (2.1)$$

and often can be calculated more conveniently by:

$$Cov(X,Y) = E(XY)-(EX)(EY) \qquad (2.2)$$

Thus Cov(X,X)=Var(X). The covariance of X and Y divided by the product of their standard deviations yields the correlation coefficient. The covariance is zero when X and Y are independent, but not necessarily vice versa.

Recall that if $f(x,y)$ is the joint density for X and Y, then the marginal density for X is defined as:

$$f_X(x) = \int f(x,y)dy \tag{2.3}$$

The integral is taken over the entire support of Y, and the resulting marginal density is basically the probability density function for X. The same thing can be done for Y. The conditional density of Y given X is defined by:

$$f(y|x) = f(x,y)/f_X(x) \tag{2.4}$$

This is interpreted as the density function for Y given that X takes on the value x.

Substituting $f_Y(y)f(x|y)$ for $f(x,y)$ in (2.3), and substituting the result of that for $f_X(x)$ in (2.4) yields Bayes' rule:

$$f(y|x) = f_Y(y)f(x|y)/\int f_Y(y)f(x|y)dy \tag{2.5}$$

which is used to get from one conditional distribution to another. Once x is fixed, the denominator of (2.5) is the constant needed to make the entire right hand side a probability density, i.e., make it integrate to unity. In many applications this constant can be computed later, or not at all, and so Bayes' rule can be written:

$$f(y|x) \propto f(x|y)f(y) \tag{2.6}$$

where "$\propto$" is read "is proportional to". In (2.6) and hereafter, the subscript on the marginal density is dropped unless it is needed to avoid confusion.

Conditional moments can be defined by using the conditional densities in the usual moment definitions. For instance:

$$E(Y|X=x) = \int y f(y|x)dy \tag{2.7}$$

$$Var(Y|X=x) = \int (y-E(Y|X=x))^2 f(y|x)dy \tag{2.8}$$

Since $EY = \int\int y f(x,y)dydx = \int E(Y|X=x)f_X(x)dx$,

$$EY = E[E(Y|X)] \tag{2.9}$$

Similarly it can be shown that

$$VarY = EVar(Y|X) + VarE(Y|X) \tag{2.10a}$$

$$Cov(Y,Z) = ECov(Y,Z|X) + Cov(E(Y|X),E(Z|X)) \tag{2.10b}$$

In applications of Bayes' rule, some distributions are described as prior, conditional, posterior, or predictive. To introduce and illustrate this terminology, an example using some well known distribution functions is given.

Example 2.1

A population of drivers is insured by XYZ insurance company, and each driver has a Poisson distribution for the number of physical damage claims to be submitted each year. For a given driver, let N_i denote the claim count random variable for year i, and y the driver's Poisson parameter, which is assumed not to vary over time. Then the conditional density is $f(n|y)=e^{-y}y^n/n!$, which has mean and variance both equal to y, and skewness of $y^{-1/2}$. In this example it is supposed that y is not known, but it is a random variable having the gamma distribution with parameters b and c, which has the density $f(y) = y^{c-1}e^{-y/b}/b^c(c-1)!$. Here and throughout the chapter a! will be used to denote $\Gamma(a+1)$, as they agree on integers and this can be used to define a! at other points.

This distribution is considered the prior distribution for Y, which is now capitalized to signify that it is a random variable. The gamma distribution in b,c has mean bc and variance b^2c, and in general, $EY^j=b^jc(c+1)\cdots(c+j-1)$ when j is a positive integer, and $EY^j=b^j(c+j-1)!/(c-1)!$ for any real $j>-c$. The shape of the distribution is determined by c; b is referred to as the scale parameter.

The unconditional or mixed distribution for N is its marginal distribution with density $f_N(n) = \int f(n,y)dy=\int f(n|y)f(y)dy$. This is the distribution the insurer faces for the driver's claim counts, as it combines the process distribution for N given Y with the parameter distribution for Y. It is sometimes referred to as the mixture of the process distribution by the parameter distribution. Doing the integration finds this to be a negative binomial distribution, with parameters c and $p=1/(1+b)$. The negative binomial density with parameters c and p is $f(n)= (c+n-1)!p^c(1-p)^n/n!(c-1)!$. This has mean $c(1-p)/p$, variance $c(1-p)/p^2$, and skewness $(2-p)(c-cp)^{-1/2}$.

From the mixed distribution it can be found that $EN_1=cb$, since for $p=1/(1+b)$, $(1-p)/p=b$. This could have been calculated using $EN_1=EE(N_1|Y)$, because $E(N_1|Y)=Y$, from the Poisson distribution, and $EY=bc$ from the gamma distribution. Similarly, $VarN_1=cb(b+1)$. From (2.10) this should equal $EVar(N_1|Y)+VarE(N_1|Y)=EY+VarY=cb+cb^2$. These two components of the total variance are sometimes referred to as "expected value of process variance" and "variance of hypothetical means", respectively. The latter terminology considers $E(N_1|Y)$ as hypothetical, since Y is not a known quantity.

The posterior distribution is the density for Y given N_1, as calculated by Bayes' rule, and can be used to update the prior distribution once an observation is available. By Bayes' rule, $f(y|n) \propto f(n|y)f(y)$. The proportionality means that any factors not involving y can be computed later, as the integral of $f(y|n)dy$ must equal 1. Thus $f(y|n) \propto e^{-y}y^{n}y^{c-1}e^{-y/b} = y^{n+c-1}e^{-y(1+1/b)}$. But from the gamma density above, the gamma distribution in parameters $b/(b+1)$ and $(n+c)$ is proportional to this same quantity, so that must be the posterior distribution of Y.

A measure of the dispersion of a random variable relative to its mean is the coefficient of variation, or CV, which is the ratio of the standard deviation to the mean. For the gamma in b,c this is given by $1/\sqrt{c}$, and so reduces to $1/\sqrt{n+c}$ for the posterior gamma.

Finally, the predictive distribution is the marginal distribution of N_2 resulting from the mixture of the Poisson model by the posterior gamma distribution for Y given N_1. Since a Poisson mixed by a gamma in b,c gives a negative binomial in c,1/(b+1), the Poisson mixed by the posterior gamma in b/(b+1),n+c can be seen to give negative binomial parameters n+c,(b+1)/(2b+1). This is the distribution for N_2 the insurer faces for this driver after observing N_1=n. It has mean (n+c)b/(b+1), which can be written as Zn+(1-Z)bc, with Z=b/(b+1). This can be interpreted as a credibility weighting between the observation n and the previous mean bc.

The usefulness of the predictive distribution goes beyond estimating the subsequent expected value. It gives the probabilities for N_2=j for all values of j, and thus quantifies the possible divergence of actual from expected results.

Exercise

a. Calculate $EE(N_2|Y)$, where the outer expected value uses the posterior gamma above.

b. Calculate Var N_2:

   1. As $EVar(N_2|Y) + VarE(N_2|Y)$.

   2. Directly from the predictive distribution.

When, as in this example, the posterior distribution is of the same type as the prior, just with different parameters, the prior and conditional distributions are said to be conjugate. Since the posterior of N_1 becomes the prior of N_2, etc., conjugate distributions allow for continued updating of the parameters of a single distribution type as subsequent data becomes available.

Thus the gamma-Poisson combination is a conjugate pair. Another is the inverse gamma-gamma pair, as the next example illustrates.

<u>Example 2.2</u>

In this example, the total workers compensation losses X_i for a certain factory in year i are assumed to be gamma distributed with parameters y,c. Here, however the scale parameter y is not known, but is specified by the prior distribution

$$f(y)=y^{-r-1}e^{-b/y}b^{r}/(r-1)! \tag{2.12}$$

This is referred to as the inverse gamma distribution in b,r because Y^{-1} is gamma distributed in b^{-1},r. The moments are given by $E(Y^j) = b^j/(r-1)(r-2)\cdots(r-j)$ for positive integers $j<r$ and $E(Y^j) = b^j(r-j-1)!/(r-1)!$ for any real number $j<r$. If $j\geq r$, the jth moment does not exist. In particular $E(Y) = b/(r-1)$ for $r>1$ and $VarY = b^2/(r-1)^2(r-2)$ for $r>2$. Note that this prior can be specified simply as $f(y)\propto y^{-r-1}e^{-b/y}$, and the conditional by $f(x|y)\propto e^{-x/y}x^{c-1}y^{-c}$.

The posterior can then be calculated as $f(y|x) \propto f(x|y)f(y) \propto e^{-x/y}y^{-c}y^{-r-1}e^{-b/y} \propto e^{-(x+b)/y}y^{-c-r-1}$. But this is the inverse gamma in (x+b), (c+r). This shows the conjugate nature of the pair.

The mixed distribution is $f(x) = \int f(x|y)f(y)dy$, and turns out to be $f(x) = b^r x^{a-1}(c+r-1)!/(b+x)^{c+r}(c-1)!(r-1)!$ This is a generalization of both the F-distribution and the shifted Pareto, and has been called different names. Here it will be referred to as the Beta2 in b,c,r, following McDonald ( ). The moments are given by:

$$E(X^j) = b^j c(c+1)\cdots(c+j-1)/(r-1)(r-2)\cdots(r-j) \qquad (2.11a)$$

for positive integers $j<r$ and

$$E(X^j) = b^j(c+j-1)!(r-j-1)!/(c-1)!(r-1)! \qquad (2.11b)$$

for any real j, $-c<j<r$.

In particular $E(X_1) = cb/(r-1)$ and $Var(X_1) = b^2c(c+r-1)/(r-1)$.

Exercise

Calculate EX and VarX via (2.9) and (2.10).

The predictive distribution for X_2 given $X_1=x$ is the conditional gamma mixed by the posterior inverse gamma and is thus the Beta2 in (x+b),c,(c+r). For $r>1$, this has mean $E(X_2|X_1=x) = (x+b)c/(c+r-1)$. Letting $Z = c/(c+r-1)$, $1-Z = (r-1)/(c+r-1)$, and then the predictive mean can be expressed as $E(X_2|X_1=x) = Zx + (1-Z)E(X_1)$. It is also possible to write $Z = 1/(1+K)$, by letting $K = (r-1)/c$. Thus again a credibility formula arises for the predictive mean. As will be seen below, this does not always happen, but it does for an important class of distributions.

Diffuse Priors

In the above example the prior distribution could have come from information about the distribution of risks within the class. Lacking such information a prior can be developed by actuarial judgement. When information and judgement lack precision, it is often felt best to make the prior as nonspecific as possible. One method that has been developed to do this is to use so called diffuse priors. One class of diffuse priors for a positive parameter y is specified by $f(y) \propto y^p$. There is no value of p for which the integral of y^p over the positive reals is finite; thus no constant can be calculated to make f(y) a proper density function. Nonetheless, $f(x|y)y^p$ may have a finite integral, and if so, a posterior distribution can be calculated. A more detailed discussion of diffuse priors may be found in Berger (   ).

For instance, $p=0$ specifies a uniform prior on the positive reals. For this p, the integral from 0 to M is finite, while that from M to infinity is not, for any number M, no matter how large. This may seem to give too much weight to large possible values of y. For example, the likelihood of y being between 1 and 2 is the same as for it being between 1,000,000,000,001 and 1,000,000,000,002.

The infinite part of the integral of y^p is from M to infinity for $p>-1$, and from 0 to ε for $p<-1$. In the latter case the weight is on values of y near zero. For $p=-1$ neither the interval 0 to ε

nor the interval M to infinity has a finite integral. Thus for $p=-1$, even though the probability is concentrated in unlikely places (near zero and infinity), there is no clearcut pull by the prior to higher or lower values of y.

Example 2.3

In Example 2.2, suppose the prior had been specified as $f(y) \propto y^p$. Then $f(y|x) \propto e^{-x/y} y^{p-c}$. As long as $p<c-1$ this is an inverse gamma posterior, with parameters x and $c-p-1$. The predictive distribution will thus be the Beta2, with parameters x, c, and $c-p-1$. Thus the predictive mean is $cx/(c-p-2)$. For $p=-2$ this is equal to the observation x, which is an appealing result in that it takes the observation at face value. As mentioned above, $p=-1$ seems to make more sense as a prior; for the predictive mean this increases the observation by a factor of $c/(c-1)$, as long as $c>1$. This also has a logical interpretation, in that $c/(c-1)$ is the ratio of the conditional mean to conditional mode, which is the most likely observation. For $p=-1$, the posterior inverse gamma has parameters x_1,c and the predictive Beta2 is in x_1,c,2c. Repeated application after n observations yields a predictive Beta2 in $\sum_{i=1}^{n} x_i$,c,nc. If $c<1$, this will eventually have a finite predictive mean when $nc>1$.

Example 2.4

In Example 2.1, taking $f(y) \propto y^p$ yields the posterior $f(y|n) \propto e^{-y}y^{n+p}$. This is a gamma distribution in 1, n+p+1 as long as $p>-n-1$. The predictive mean is n+p+1, which for p=-1 yields the observation n.

Note that in both of these examples the posterior and conditional distributions are conjugate, and so can then be used to begin the Bayesian updating process as more observations become available.

Aggregate Claims Distributions

The application of credibility to insurance problems often involves a decomposition of the total losses into frequency and severity components. This part of the statistical preliminaries will be the calculation of the moments and percentiles of aggregate claims from those for frequency and severity.

The definition of the aggregate claims T for a given period is:

$$T=X_1+\cdots+X_N \tag{2.12}$$

where N is the number of claims in the period and X_i is the amount of the ith claim. It is usually assumed that the X_i are independent of each other and of N, and that all the claims follow a common severity distribution. Thus, the subscripts can be dropped when referring to the severity random variable X.

The moments of T are given by:

$$E(T)=E(N)E(X) \quad (2.13a)$$

$$Var(T)=E(N)Var(X)+E(X)^2Var(N) \quad (2.13b)$$

$$Skw(T)=(Skw(X)CV^3+3n_2CV^2+n_3)/\sqrt{E(N)(CV^2+n_2)^3} \quad (2.13c)$$

Here CV denotes the severity coefficient of variation, and $n_i=E(N-EN)^i/EN$. (2.13a) is proven from (2.9) by noting $E(T)=EE(T)|N)=E(NE(X))=E(N)E(X)$. (2.13b) follows similarly from (2.10) since:

$$\begin{aligned} Var(T)&=EVar(T|N)+VarE(T|N)\\ &=E(NVar(X))+Var(NE(X))\\ &=E(N)Var(X)+E(X)^2Var(N) \end{aligned}$$

This could alternatively be computed by evaluating $E(T^j|N)$ via (2.9), which is what is used to derive (2.13c).

One method of estimating the percentiles of T is to assume a particular distributional form, e.g., T is normal or gamma distributed. If the moments of X and N are given, the distribution for T can then be estimated from its moments, which are computed via (2.13). The normal distributional assumption incorporates a skewness of zero. The gamma has a skewness of twice its coefficient of variation. This is probably more realistic for property-casualty lines, but neither distributional form is likely to be correct.

Several approaches to improved estimation of percentiles of aggregate claims have been developed. One is to incorporate a third parameter so that the first three moments can be matched. For instance for the normal distribution the so called normal power approximation (NP) incorporates a skewness correction as follows. Let σ_T denote the standard deviation of T, and t_p the pth percentile. Then the normal approximation estimates t_p by $ET+\sigma_T y_p$, where y_p is the pth percentile of the standard normal. The *NP* approximation for t_p is:

$$t_p=ET+\sigma_T y_p+\sigma_T SkwT(y_p{}^2-1)/6 \qquad (2.14)$$

This *NP* formula is derived using a power series expansion for t_p. Pentikainen (   ) recommends its accuracy only for SkwT<1, after which the NP tends to exaggerate the difference between the percentiles t_p and their normal approximation estimates.

Another approximation for aggregate claims is offered by Seal (====), who adds a third parameter to the gamma that shifts the origin to the left or right. The percentiles are calculated using a fairly simple modification to the gamma distribution function. Pentikainen ( ) finds the accuracy of this approximation comparable to that of the NP.

Another way of adding a third parameter to the gamma is to use a power transform, i.e., to assume T^a is gamma distributed for some real number a. If a=-1 this gives the inverse gamma distribution used above. Applications of this method can be found in Venter (1984).

It is also possible to compute the aggregate distribution function without making a distributional assumption for the aggregate claims. However this usually requires knowledge of the density functions for frequency and severity, not just their moments. One such method is simulation. A possible number of claims n is generated according to the frequency distribution, then n possible claim sizes are drawn from the severity distribution. This gives one possible realization of T. This process can be repeated many thousands of times to estimate the distribution function of T. While conceptually simple, this process is often expensive and time consuming.

Another method is to build up the aggregate distribution function recursively, i.e., the probability that $T<t$ is computed from the probabilities that T is less than t-1, t-2, etc. Panjer ( ) shows a fairly efficient way to do this for a discrete severity distribution and a Poisson, negative binomial, or binomial frequency. For the Poisson frequency, dePril ( ) finds an even more efficient algorithm for a piecewise linear distribution function.

Finally, a method of calculating aggregate claim probabilities based characteristic functions is becoming widely used. The characteristic function is a complex analog of the moment generating function, and can be computed for aggregate claims from the moment generating function of frequency and the characteristic function of severity. The distribution function for aggregate claims can be recovered from its characteristic function via numerical integration. The calculation is thus somewhat intricate, but once programmed it is fairly efficient. One difficulty is calculating the severity characteristic function, as it is not usually of closed form. This method was pioneered by Mong ( ), who used a gamma severity. Heckman and Meyers ( ) extended it to a step function probability density, and Venter ( ) generalized this to a piecewise linear density. The latter two severity functions can be used to approximate other distributions, thus making this method of quite general application.

Section 3 - Limited Fluctuation Credibility

The limited fluctuation credibility estimator can be expressed as:

$$C=(1-Z)M+ZT \qquad (3.1)$$

where T is the observation and M is a previous estimate. M is generally supposed to be the estimate one would use if the observation T were not available, and it could come from previous experience and/or related data. Typically T will be the loss ratio, pure premium, frequency, or severity for a class, state, or risk for a certain time period, and C estimates its value for another, usually future, period. Here, to be specific, T will be the aggregate losses for a one year period, thus $T=X_1+\cdots+X_N$ as above, with the usual independence assumptions.

For limited fluctuation theory, (3.1) can be rewritten as:

$$C=(1-Z)M+ZET+Z(T-ET) \tag{3.2}$$

only the last term is considered random, and the goal of the theory is to keep its contribution within specific bounds. In particular, k and p are selected and then Z is sought so that $Pr(Z(T-ET)<kET)=p$. For example, p=.95 and k=.05 are typical choices and result in requiring that the random component of (3.2) be less than 5% of the expected value ET with 95% probability.

Actually this requirement is only an upper bound on Z(T-ET), but in applications T is always assumed to be symmetric or slightly skewed to the right, so that this upper bound requirement guarantees that $Pr(|Z(T-ET)|<kET) > 1-2(1-p)$. Thus for p=.95, k=.05, the credibility requirement provides that the random component has 90% probability of being less than 5% of ET in absolute value.

The criterion can be restated as $Pr(T<ET+kET/Z)=p$, or $t_p=ET + kET/Z$, where t_p again is the pth percentile of T. To find Z, different methods of computing t_p can be invoked. Under the

normal approximation, $t_p=ET+y_p\sqrt{VarT}$, and so $Z=kET/y_p\sqrt{VarT}$. In terms of frequency and severity,

$$Z^2=(k/y_p)^2(ENEX)^2/\ (ENVarX+(EX)^2VarN) \tag{3.3a}$$

$$=(k/y_p)^2(EN)^2/(ENCV^2+VarN) \tag{3.3b}$$

$$=(k/y_p)^2EN/(CV^2+n_2) \tag{3.3c}$$

Where again CV is the severity coefficient of variation and n_2 is the frequency ratio of variance to mean. Z=1 when EN = $(CV^2+n_2)(y_p/k)^2$. This value of EN is called the full credibility value, denoted as n_F. The value of EN that produces credibility Z, n_Z, can be seen to follow $n_Z=Z^2n_F$, or $Z=\sqrt{n_Z/n_F}$.

This "square root rule" holds only for the normal approximation. For the NP,

$$t_p=ET+\sqrt{VarT}(y_p+SkwT(y_p^2-1)/6) \tag{3.4}$$

and so $kET/Z=\ \sqrt{VarT}(y_p+SkwT(y_p^2-1)/6$. This can be solved for Z in term of frequency and severity moments using (2.13) to yield:

$$Z=k/[y_p\sqrt{m_2/EN}+(m_3/m_2)(y_p^2-1)/6EN] \tag{3.5}$$

where m_2 and m_3 are aggregate claim shape descriptors defined by:

$$m_2=n_2+CV^2$$

$$m_3=CV^3SkwX+3n_2CV^2+n_3$$

The normal approximation formula (3.3c) can then be seen to be the special case $m_3=0$, i.e., SkwT=0, which the normal approximation assumes, but which is unlikely in practice. The square root rule does not apply for the NP credibilities; rather they must be calculated from (3.5) directly. It is possible to solve (3.5) for EN by considering it a quadratic in $\sqrt{EN}$. This produces a formula for n_Z, the value of EN needed for credibility Z:

$$n_Z=(Z^2/4k^2)(y_p\sqrt{m_2}+\sqrt{y_p^2m_2+2(k/Z)(y_p^2-1)m_3/3m_2})^2 \qquad (3.6)$$

Both (3.3) and (3.5) have an important invariance property: the calculation of Z from EN is affected neither by simple monetary inflation nor the addition of independent identical distributed exposure units. In fact, without the latter invariance, Z could not really be regarded as a function of EN. The former allows credibility standards to remain constant until the shape of the severity distribution changes.

The invariance under simple monetary inflation results because the severity coefficient of variation and skewness do not depend on scale. The latter invariance follows because the frequency mean, variance and third central moment are all additive functions; that is, the additional units will increase these moments all by the same factor. Thus n_2 and n_3 will not be affected. (Anyone who thinks this is because these three moments are all cumulants is probably correct.)

An example may help clarify these concepts.

Example 3.1

Commercial fire losses for a state are assumed to have a Poisson frequency distribution and a lognormal severity, with CV=7. For the Poisson, n_2 and n_3 both equal 1, and so with this CV, $m_2=50$. The skewness of the lognormal is given by $SkwX=CV^3+3CV$, and so for this example $SkwX=343+21=364$. Thus $m_3=364 \cdot 343+3 \cdot 49+1=125,000$. The credibility requirements are specified by $p=.95$ and $k=.05$, which gives $y_p=1.645$ from a normal table.

The normal approximation n_F is given by $n_F=m_2(y_p/k)^2$, and thus in this case is $50(1.645/.05)^2=54,120$. For the NP, n_F can be calculated via (3.6) to be 80,030. Thus considering skewness has a substantial impact in this case, basically because the severity distribution is highly skewed. The assumption of a CV of 7 for commercial fire is consistent with the findings of Simon (1969). The skewness of aggregate claims may be calculated as $SkwT = m_3/m_2^{1.5}\sqrt{EN}$, which in this case is 1.25. This is somewhat above Pentikainen's recommendation for the boundary of the accuracy of the NP, and thus the NP n_F estimate may be somewhat too high.

Instead of the lognormal severity, it is interesting to consider a constant severity. This could arise, for example, in a group of life insurance policies all with the same benefit. In this case, CV=0, and $m_2=m_3=1$. For the normal approximation, n_F then becomes 1082, which has been a widely used credibility standard.

For the NP approximation, n_F is 1094, via (3.6). Thus for the Poisson alone, the skewness correction is not substantial.

The negative binomial frequency could have been used instead of the Poisson. With parameters c and p, $n_2=1/p$ and $n_3=(2-p)/p^2$. In a study of automobile claims, Dropkin (1959) found $n_2=1.184$. This implies $p=.8446$, and so $n_3=1.620$. For the constant severity case, $m_2=n_2$ and $m_3=n_3$; the normal approximation then yields $n_F=1282$ and the NP gives $n_F=1297$. For the lognormal above, n_F increases to 54,320 under the normal approximation, and to 80,150 with the NP. Thus the negative binomial assumption with these parameters seems to have some impact in the frequency only case, but little when a highly skewed severity has already been included.

Exercise

Verify the calculations in the paragraph above.

Meyers and Schenker (1983) discuss the possibility that the negative binomial n_2 may be substantially larger than 1 for individual large commercial risks. In their model, exposure units are not independent, so some of the above reasoning does not apply. However it is instructional to explore the implications of a large n_2. Thus suppose a negative binomial distribution is given with $n_2=51$. Then $p=1/51$, and $n_3=5151$. For the above severity, m_2=100, and m_3=137,500. Then the normal and NP n_F's are 108,200 and 123,400 respectively.

Exercise

Verify these n_F's. What would they be for frequency only? How many claims would be neede for 50% credibility under the normal and NP approximations?

The limited fluctuation Z depends only on the distribution of T, and treats the previous estimate M as a constant. Thus Z does not depend on how good this estimate may be or where it comes from, although such matters could influence the selection of p and k, *on which Z depends.* If T is the aggregate losses for a state, M could be the previous year's estimate. *If T represents* only a single class or territory, M could be the statewide estimate for the same year. In general, M is supposed to be the best estimate available without the particular observation T, and *in fact may be formed as a combination of other estimators.*

The nondependence of Z on the properties of the previous estimator is both a strength and a weakness. It provides flexibility and a simple algorithm for routine application, and does not require the estimation of additional parameters. However it may ignore or only judgementally consider elements that can be quantified with some additional research. The least squares methodology, to be reviewed next, takes such an approach.

Section 4 - Least Squares Credibility

In the least squares theory, the previous estimator applied to the complement of credibility is specified much more explicitly. Consequently, more details of the estimation problem need to be modelled. This requires some notation. To have a particular problem to work with, it will be supposed that the losses for N risks are observed for a period of n years. The pure premium for the ith risk in year u is denoted as X_{iu}. Pure premium is loss divided by exposure; for now all risks are assumed to have the same number of exposure units, which is constant over time. In Section 6, application of credibility theory to risks of different sizes will be made.

The pure premium for a future time period, time 0, is to be estimated for the gth risk. This will end up being estimated as a credibility weighting of the average observed pure premium for risk g over the n years, denoted as $X_{g.}$, with the grand average of all the risks for those years, denoted as $X_{..}$. In formulas, $X_{g.}=\Sigma_u X_{gu}/n$, and $X_{..}=\Sigma_g X_{g.}/N$.

The credibility given to the risk experience will depend in part on the stability of that experience, as in limited fluctuation theory, but it will also depend on the relevance of the grand mean to the individual risk, which is quantified by the variance across risks of the individual risk means. The greater this variance, the more diverse are the risks, and thus the grand mean provides less relevant information about an individual risk. This will in turn lead to greater credibility assigned to the risk's own experience, and less to the grand mean. The explicit consideration of the relevancy of the estimator applied against the complement of credibility is one of the distinctive features of least squares credibility.

The least squares credibility estimator could be dervied by finding the weight Z that minimizes $E[X_{g0}-(ZX_{g.}+(1-Z)X_{..})]^2$, and this approach will in fact be followed in Section 7. However, the same estimator also arises as a result of a more general estimation problem as follows. X_{g0} is estimated as <u>any</u> linear combination of all the observations X_{iu}, not just a weighted average of $X_{g.}$ with $X_{..}$, with the expected squared error to be minimized. The general linear combination of the observations can be expressed as $a_0+\Sigma_{i,u}a_{iu}X_{iu}$, so the credibility criterion will be to find the weights (a's) that minimize:

$$E[X_{g0}-(a_0+\Sigma_{i,u}a_{iu}X_{iu})]^2 \qquad (4.1)$$

It will turn out that the resulting weights can be combined into a simple credibility formula, which gives further justification for such a formula.

There are Nn+1 weights a_{iu} to find, and this is approached by setting the partials of (4.1) with respect to these variables to zero. Doing this, with some algebraic manipulation, produces the following system of Nn+1 equations:

$$EX_{g0} = a_0 + \Sigma_{i,u} a_{iu} EX_{iu} \qquad (4.2a)$$

$$Cov(X_{g0}, X_{jv}) = \Sigma_{i,u} a_{iu} Cov(X_{iu}, X_{jv}) \qquad (4.2b)$$

There are Nn equations expressed by (4.2b), one for each j,v combination.

Exercise

Derive (4.2). Hint: The partial with respect to a_0 will give (4.2a). Set the partial with respect to a_{jv} to zero and subtract (4.2a) multiplied by EX_{jv} from this equation.

In order to solve this system for the a's, more model assumptions are needed, so that the covariances can be evaluated. As an example, a fairly simple model will be investigated first. It will be assumed that the risk i loss ratio for time u can be decomposed as follows:

$$X_{iu} = m + R_i + Q_{iu} \qquad (4.3)$$

Here m is the overall average, R_i is a risk effect that does not vary over time, and Q_{iu} is a random fluctuation. The R's and Q's are treated as random variables, as their values are not known. The average over all risks of the R_i's is assumed zero, i.e., $ER_i=0$. Also it is assumed that $EQ_{iu}=0$, and so $EX_{iu}=m$. This is an overall expected value; $E(X_{iu}|R_i)=m+R_i$ is the conditional expected value for the ith risk. Finally, it is assumed that different Q's and R's are independent random variables with $VarR_i=t^2$ and $VarQ_{iu}=s^2$.

To compute $Cov(X_{iu},X_{jv})$ under these assumptions, it will be convenient to introduce the following notation: $\delta_{ij}=1$ if $i=j$; otherwise $\delta_{ij}=0$.

With this in hand, note that $E(R_iR_j)=\delta_{ij}t^2$: since different R's are *independent*, *if* $i\neq j$, $E(R_iR_j)=ER_iER_j=0$; also, $ER_i{}^2=VarR_i+(ER_i)^2=t^2$. Similarly, $E(Q_{iu}Q_{jv})=\delta_{ij}\delta_{uv}s^2$.

Now, by definition of covariance,

$$\begin{aligned}Cov(X_{iu},X_{jv})&=E[(X_{iu}-EX_{iu})(X_{jv}-EX_{jv})]\\&=E[(X_{iu}-m)(X_{jv}-m)]\\&=E[(R_i+Q_{iu})(R_j+Q_{jv})]\\&=E[R_iR_j]+E[Q_{iu}Q_{jv}]\text{ (by independence of R's and Q's)}\end{aligned}$$

And thus,

$$Cov(X_{iu},X_{jv}) = \delta_{ij}t^2 + \delta_{ij}\delta_{uv}s^2 \qquad (4.4)$$

The notation says this covariance is *zero unless* $i=j$, *in which case it is* t^2, *unless also* $u=v$, *in which case it is* t^2+s^2. This means that $VarX_{iu}=s^2+t^2$, which can also be expressed as $VarX_{iu}=EVar(X_{iu}|R_i)+VarE(X_{iu}|R_i)$ - the expected process variance plus the variance of the hypothetical means.

Exercise

Show that $EVar(X_{iu}|R_i)=s^2$ and $VarE(X_{iu}|R_i)=t^2$.

Because so many of the covariances are zero, plugging (4.4) back into (4.2b) will make many terms drop out, and in fact produces the equation:

$$\delta_{gj}t^2 = \Sigma_u a_{ju}t^2 + a_{jv}s^2 \quad (4.5)$$

There is still one such equation for every j,v combination; for fixed j summing all the v-equations (n of them) produces:

$$n\delta_{gj}t^2 = n\Sigma_u a_{ju}t^2 + \Sigma_u a_{ju}s^2 \quad (4.6)$$

and so,

$$\Sigma_u a_{ju} = n\delta_{gj}t^2/(s^2+nt^2) \quad (4.7)$$

Plugging this into (4.5) will yield, after some algebra:

$$a_{jv} = \delta_{gj}t^2/(s^2+nt^2) \quad (4.8)$$

This says the weight is zero unless j=g, and then it is:

$$a_{gv} = t^2/(s^2+nt^2) \quad (4.9)$$

To find a_0, substitute

$$EX_{iu} = m \quad (4.10)$$

into (4.2a) to yeild $m=a_0+\Sigma_{i,u}a_{iu}m$, and so $a_0=m(1-\Sigma_{i,u}a_{iu})=m(1-\Sigma_u a_{gu})=ms^2/(s^2+nt^2)$. Finally, since the estimator of X_{g0} is $a_0+\Sigma_{i,u}a_{iu}X_{iu}$, which simplifies to $a_0+\Sigma_u a_{gu}X_{gu}$, the credibility estimator can be written as:

$$\tilde{X}_{g0} = ms^2/(s^2+nt^2) + \Sigma_u X_{gu}t^2/(s^2+nt^2) \quad (4.11)$$

Now $\Sigma_u X_{gu}$ may be written as $nX_{g.}$; defining $Z=nt^2/(s^2+nt^2)$ produces $\tilde{X}_{g0} = (1-Z)m + ZX_{g.}$; here a natural estimate for m would be $X_{..}$, and in fact this is the minimum variance unbiased estimate of m (see ISO (1983)). Substituting this estimate gives:

$$\tilde{X}_{g0} = (1-Z)X_{..} + ZX_{g.} \quad (4.12)$$

Thus the best linear estimate of X_{g0} turns out to be a credibility formula. This formula can alternatively be derived as the least squares linear estimate having $a_0=0$ but constrained to be unbiased (see ISO (1983)).

From the definition of Z it can be seen that if t^2 is higher, so is the credibility given to the risk experience. Since t^2 measures the dispersion of individual risk conditional means around the grand mean m, it can be seen that greater dispersion leads to greater credibility; the more different a risk is likely to be from the average, the greater credence will be placed on a risk's own experience. On the other hand, Z reduces as s^2 increases; higher s^2 means that the risks are less stable over time, and thus less reliance can be placed on their individual results. This was also seen in limited fluctuation credibility.

By defining $K=s^2/t^2$, Z can be written as $Z=n/(n+K)$, which is basically Whitney's 1918 formula.

The credibility formulas illustrated by this simple model will be found to hold in more general situations. In fact, if (4.4) and (4.10) are satisfied, the rest of the development will be the same, ending up with (4.12) with the same definition of Z.

This example is typical in one respect, which is in the division of the uncertainty about X_{iu} into two components: a time invariant risk specific component (here R_i), and a random fluctuation in each time period (Q_{iu}). Some observers may feel that this distinction is somewhat artificial, because neither component is ever observed in isolation; however it is an intuitively reasonable distinction, and leads to a model that seems to have practical value.

A More General Model

In the simple model above, each risk had one parameter R_i which described the risk, and then a random fluctuation. More generally it is now assumed that each risk has a vector of parameters, denoted by R_i, that describe the risk, which still nonetheless is subject to random fluctuation. For example, a risk with a negative binomial frequency distribution and an inverse gamma severity would have four parameters describing these distributions, and random fluctuation from year to year as provided by those distributions. Letting R denote the vector $<R_1, R_2, \ldots, R_N>$, it is assumed that X_{iu} and X_{jv} are conditionally independent given R. Each risk has its own conditional mean and variance, which may be denoted by $E(X_{iu}|R)=E(X_{iu}|R_i)=m_i$ and $Var(X_{iu}|R)=Var(X_{iu}|R_i)=s_i^2$

It is assumed that for different i the R_i are independent identically distributed random vectors with $E(m_i)=m$, $Var m_i=t^2$, and $Es_i^2=s^2$. This implies that $EX_{iu}=m$ and $VarX_{iu}=s^2+t^2$ (why?).

Here again, s^2 is the expected process variance and t^2 is the variance of the hypothetical means.

In order to apply (4.2), it is necessary to compute $Cov(X_{iu},X_{jv})$ from these assumptions. By (2.10),

$$Cov(X_{iu},X_{jv})=ECov(X_{iu},X_{jv}|R)+Cov[E(X_{iu}|R),E(X_{jv}|R)] \quad (4.13)$$

Now by the conditional independence of the X's, the first term is zero unless i=j and u=v, in which case it is $EVar(X_{iu}|R)=Es_i{}^2=s^2$. The second term is also $Cov[E(X_{iu}|R_i),E(X_{jv}|R_j)]$, which by the independence of R_i and R_j is zero unless i=j, in which case it is $Varm_i=t^2$. Thus:

$$Cov(X_{iu},X_{jv}) = \delta_{ij}\delta_{uv}EVar(X_{iu}|R) + \delta_{ij}VarE(X_{iu}|R) \quad (4.14)$$

$=\delta_{ij}\delta_{uv}s^2+\delta_{ij}t^2$, which is (4.4). Plugging this back into (4.2) will then yield (4.12) by the same reasoning used for the original simplified model above.

<u>Example 4.1</u>

Suppose severity is constant at one unit, frequency is Poisson in R_i, *and exposure is one. Then the pure premium* X_{iu} *is the number of claims for risk i in time u; by the Poisson hypothesis,* $m_i=R_i$, and $s_i{}^2=R_i$ as well. If R_i is gamma distributed in b,c, then $t^2=Varm_i=VarR_i=b^2c$, and $s^2= Es_i{}^2=ER_i=bc$. Thus $K=s^2/t^2=1/b$, and $Z=n/(n+K)=nb/(nb+1)$. For n=1 this gives the predictive mean computed in Example 2.1.

<u>Example 4.2</u>

X_{iu} is assumed to be gamma distributed in $R_i=<Y_i,c>$. Thus $m_i=Y_ic$, and $s_i{}^2=Y_i{}^2c$. Y_i is assumed inverse gamma in b,r; so $t^2=Varm_i=c^2VarY_i=c^2b^2/(r-1)^2(r-2)$, and $s^2=Es_i{}^2=cEY_i{}^2=$

$cb^2/(r-1)(r-2)$; then $K=s^2/t^2=(r-1)/c$, and $Z=n/(n+K)= nc/(nc+r-1)$. Thus $\tilde{X}_{g0}$ agrees with the predictive mean from Example 2.2.

Section 5 - Estimation of K

Up until now, s^2 and t^2 were treated as known constants, but in practice they usually have to be estimated. One approach is to estimate s^2 based on observed deviations of risk annual results from risk means, and t^2 from observed deviations of risk means from the grand mean. Sometimes it is more convenient to estimate the total variance $VarX_{iu}=s^2+t^2$ from the deviations of individual risk observations from the grand mean, and then get s^2 or t^2 by subtraction. This is simplified when the conditional distribution is Poisson, because then the conditional mean and variance are equal, so $s^2=EVar(X_{iu}|R_i)=ER_i=m$, the grand mean.

Example 5.1

A group of 300 car owners in a high crime area submit the following number of theft claims in a one year period:

Number of Claims:	0	1	2	3	4	5
Number of Owners:	123	97	49	21	8	2

Each owner is assumed to have a Poisson distribution for X_{i1}, the number of thefts, but the mean number may vary from one owner to another. A credibility estimate is desired for X_{i0}, the number of claims for each driver for the next year.

The average number of claims per driver can be calculated to be 1.0. By the Poisson assumption, this is also s^2. The average value of $X_{i1}{}^2$ can be found to be 2.2, so $s^2+t^2=VarX_{i1}$ can be estimated to be $2.2-1.0^2=1.2$. This implies $t^2=0.2$, and so $K=5$, and $Z=1/6$. The credibility estimate for X_{i0} is thus $5/6+X_{i1}/6$. In general, estimating s^2 and t^2 separately can be approached by calculating the statistics $S_i=\sum_u(X_{iu}-X_{i.})^2/(n-1)$, $S=\sum_i S_i/N$, and $T=\sum_i(X_{i.}-X_{..})^2/(N-1)$. The expected value of these statistics can be calculated (laboriously) from (4.4) and (4.10). As a hint of how that might proceed, multiplying out the squares in S and T result in a whole lot of terms of the form $X_{iu}X_{jv}$, whose expected values then need to be evaluated. This is done using (4.4) and (4.10), which together imply that $E(X_{iu}X_{jv})=m^2+\delta_{ij}(t^2+\delta_{uv}s^2)$. The answers are: $E(S_i|R_i)=s_i^2$; $ES_i=EE(S_i|R_i)=Es_i^2=s^2$; $ES=s^2$; $ET=t^2+s^2/n$.

The formula for S_i looks like a fairly usual statistical result. T looks like it should be something like t^2, but probably a little bit higher, because some extra fluctuation is added from the use of the estimated means rather than m_i and m. Thus the formula for T looks about right also. From these formulas, S is an unbiased estimator of s^2, nT is an unbiased estimator of s^2+nt^2, and $T-S/n$ is an unbiased estimator of t^2. Since $1-Z=s^2/(s^2+nt^2)$, it could be estimated by S/nT, as both numerator and denominator are unbiased. Such an approach may be satisfactory in many cases, and is supported by the independence of S and T, as shown by Klugman (1985).

Example 5.2

Table 5.1 displays the pure premium experience for 9 risks, all with the same constant number of exposure units, for a 6 year period. $X_{i.}$ and S_i are calculated from this experience, as shown, and $X_{..}$=.563, S=.357, and T=.066 can then be computed from the formulas above. These yield S/nT=.899, which can be used as the estimate of 1-Z, and so Z is estimated to be .101.

Table 5.1

Risk	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Xi.	Si
1	.430	.375	2.341	.175	1.016	.466	.801	.649
2	.247	1.587	1.939	.712	.054	.261	.800	.615
3	.661	.237	.063	.250	.602	.700	.419	.072
4	.182	.351	.011	.022	.019	.252	.139	.021
5	.311	.664	1.002	.038	.370	2.502	.815	.792
6	.301	.253	.044	.109	2.105	.891	.617	.622
7	.219	1.186	.431	1.405	.241	.804	.714	.251
8	.002	.058	.235	.018	.713	.208	.206	.071
9	.796	.260	.932	.857	.129	.349	.554	.121
							.563	.357

An important issue in credibility theory is the accuracy of this estimate of Z. For this example, Table 5.1 was generated by taking random draws from assumed gamma distributions for each risk. The parameters of these gamma distributions are shown below, along with the risk conditional means and variances.

Risk	b	c	Mean	Variance
1	.6159	1.0476	.6452	.3974
2	.8001	0.9063	.7251	.5802
3	.6098	0.9654	.5887	.3590
4	.2391	0.9219	.2204	.0527
5	.5206	1.0184	.5302	.2760
6	.6768	1.0937	.7402	.5010
7	.9575	1.1395	1.0911	1.0447
8	.1999	1.0153	.2030	.0406
9	.5083	0.9320	.4737	.2408
			.5797	.3880

Thus m=.5797, which is not too different from $X_{..}$, and s^2=.388, which again is fairly close to S. The variance of the above conditional means can be found to be t^2=.0664, and thus t^2+s^2/n= .1311, which is fairly different from T. Thus the "population" value of 1-Z of $s^2/(s^2+nt^2)$=.493 and Z=.507 is quite a bit different from that estimated by the data.

This experiment was repeated twice more, that is, six years of data were simulated two more times, with the following results:

Experiment	S	T	1-Z	Z
1	.357	.066	.899	.101
2	.274	.103	.443	.557
3	.219	.172	.211	.789

Note:Calculations based on unrounded values

Thus this method does not seem to be able to produce a close estimate of Z with this quantity of data when the process is this unstable. However the average of the three estimates, .482, is just slightly below the underlying value of Z, .507, which gives hope that with just somewhat more data good estimates are possible. Estimating the variance of the estimated Z would help provide an understanding of the accuracy of the calculation, and this is discussed further in Section 7.

Empirical Bayesian Approaches - the (N-1)/(N-3) Correction

Estimating 1-Z by S/nT has a drawback in that while the numerator and denominator are both unbiased, 1-Z is not. This is a typical problem for quotients of unbiased estimators. In this case it arises because E(1/T)>1/ET (see exercise below). This implies that $E(1/nT)>1/s^2+nt^2$, and thus $E(S/nT) > s^2/(s^2+nt^2)$, i.e., S/nT overstates 1-Z, and thus understates Z, on the average.

Exercise

Show that $E(1/T) \geq 1/ET$. Hint: Schwartz' inequality says that $[\int g(t)h(t)dt]^2 \leq \int g(t)^2 \int h(t)^2$. Take g^2 and h^2 to be $tf(t)$ and $f(t)/t$. Equality occurs only in degenerate cases.

The excess of $E(1/T)$ over $1/ET$ varies from one distribution to another, so it is not possible to find a general correction for the bias in Z. This excess is greater for heavy tailed distributions, however, so an approximate lower bound could be found by computing its value in the normal distribution case.

The calculation of 1-Z when both the conditional and prior distributions are normal has been the focus of a field known as parametric Empirical Bayes statistics. A classic article in this field is Efron and Morris (====). Following Morris (====), Klugman (1985) shows that S and T are independent random variables, with S gamma distributed in $2s^2/N(n-1)$, $N(n-1)/2$ and T gamma in $2(t^2-s^2/n)/(N-1)$, $(N-1)/2$. By the gamma moment formula, if Y is gamma in b, c then $E(1/Y)=1/b(c-1)$ as long as $c>1$, and is non-existent otherwise. This is greater than $1/EY$ by a factor of $c/(c-1)$. For T the value of c is $(N-1)/2$, so $ET^{-1}=ET(N-1)/(N-3)$, as long as $N>3$. Thus $E(S(N-1)/(N-3)nT)=1-Z$, and so $S(N-1)/(N-3)nT$ is an unbiased estimator of 1-Z. This is the above credibility estimator of 1-Z adjusted by the factor $(N-1)/(N-3)$. For $N<4$, credibility weighting would not be recommended by this school, as then $E(S/nT)$ would not be finite, and so no correction factor could make S/nT unbiased.

Evaluating the (N-1)/(N-3) Correction

For other conditional and prior distributions, the excess of $E(1/T)$ over $1/ET$ is likely to be greater than for the normal, so the $(N-1)/(N-3)$ correction factor is probably a lower bound. However, there is a potential problem with this correction factor which could cause it to actually overcorrect for bias, namely that it does not take into account the usual pratice of capping $1-Z$ at 1. The true value of $1-Z$, $s^2/(s^2+nt^2)$, must be in the range $[0,1]$. In practice, however, the calculated value of nT may be less than S, which would make $S/nT>1$. Typically $1-Z$ would be capped at 1 in this case, giving $Z=0$. However by this practice the estimator of $1-Z$ has effectively become $\min[1,S/nT]$, which has a lower expected value than S/nT. That is, the capped estimator has lower bias than S/nT, and may even be unbiased or be biased in the other direction.

Even knowing the distributions for S and T, as in the normal-normal case, $E\min[1,S/nT]$ does not have a closed form expression. It can be calculated numerically by:

$$E\min[1,S/nT] = \int \{F_T(s/n) + (s/n)\int_{s/n} t^{-1} f_T(t)\,dt\} f_S(s)\,ds$$

A practical problem with this expression is that the distribution functions of S and T depend on s^2 and t^2, and these cannot be brought out explicitly; however in one special case - when s^2 is a known constant - Emin[1,S/nT], the expected value of the uncorrected estimate of 1-Z, can be calculated as a function of 1-Z. Some results are shown in the table below:

		Emin[1,S/nT]	
N	1-Z: .800	.500	.200
3	.799	.673	.426
18	.809	.557	.227

Thus when 1-Z is large, the capped estimator does not seem to be upwardly biased, although this is not true for smaller factors. Even for N=3, the bias is finite, and thus credibility weighting is a useful possibility even in that case.

Exercise

A population of risks with X_{iu} gamma in b_i, c_i is determined by independent draws of the b's from a uniform distribution on [0,1], and the c's from a uniform distribution on [.85,1.15]. What is K? (Hint: the uniform with width a has variance a/12; use $E(b^2)=Var(b)+(Eb)^2$ and independence of b and c to find $E(b^2c)$; compute Var(bc) via $E(b^2c^2)-(EbEc)^2$.)

Example 5.3

The answer to the previous exercise is K=3.88. As a test of the correction factor in a non-normal distribution case, 100 risks were drawn from such a population, and the average values of b_ic_i and $b_i^2c_i$ were found to be .500 and .330. These compare to the expected values from the uniform distribution of .500 and .333.

The variance of the b_ic_i's (hypothetical means) was .085, compared to a theoretical value of .086. For each risk, 6 years of data were simulated, as in Example 5.2, and this process was repeated for five different experiments. For n=6, 1-Z=.393 from the uniform prior, and 1-Z=.392 from the b's and c's actually drawn, so this is the target value to which the estimate values of 1-Z can be compared. For the five experiments, 1-Z was estimated using the (N-3)/(N-1) factor, as the impact of the capping by 1 did not seem large. The following results occurred:

Experiment	X..	S	T	1-Z
1	.504	.434	.164	.432
2	.482	.302	.109	.452
3	.455	.303	.122	.406
4	.510	.362	.141	.419
5	.471	.277	.102	.443

For comparison, the anticipated value of T is .086+.333/6=.142 from the uniform prior, and .085+.330/6=.140 from the 100 risks actually selected. The small but consistent overstatement of 1-Z may be due to ET^{-1} being greater for these distributions than the normal-normal (N-3)/(N-1) correction contemplates. The Bayesian methods discussed below give slightly higher estimates of 1-Z, but they also provide estimates of the potential error, as will be addressed later.

Bayesian Estimates of Z

In the normal-normal case, some work has been done on Bayesian estimates of s^2 and t^2. This could be done to reflect some prior belief, however faint, in where s^2 and t^2 are likely to be: it also produces an entire posterior distribution for these parameters, rather than just point estimates.

Since S and T are both conditionally gamma distributed given s^2 and t^2, inverse gamma priors, as discussed in Example 2.2, could be postulated. As an example, s^2 will be taken to be inverse gamma distributed in p, 2, and the quantity t^2+s^2/n is given an independent inverse gamma in q, 2. While this approach ends up providing reasonable estimates, it does have a theoretical problem in that some possibility that t^2 is negative is allowed.

With the shape parameter 2, the inverse gamma is an infinite variance distribution with mean equal to the scale parameter. This approach then does not tie down the possible values of s^2 and t^2+s^2/n too precisely, but it does specify an expected value for each.

Following Example 2.2, the posterior distributions are:

$$s^2|S \sim \text{Inverse Gamma in } p+N(n-1)S/2,\ 2+N(n-1)/2 \quad (5.1a)$$

$$t^2+s^2/n|T \sim \text{Inverse Gamma in } q+(N-1)T/2,\ (N+3)/2 \quad (5.1b)$$

Thus $E(s^2|S)=[2p+N(n-1)S]/[2+N(n-1)]$ and $E[(t^2+s^2/n)^{-1}|T]=(N+3)/(2q+(N-1)T)$, from the inverse gamma moment formulas. Also, the prior expected value of $(t^2+s^2/n)^{-1}$ is 2/q, and so the prior expected value of 1-Z is 2p/nq. $E(s^2|S)$ can be seen to be between the prior expectation p and the observation S, and much closer to the latter. Similarly, $E[(t^2+s^2/n)^{-1}|T]$ can be seen to fall between the prior expectation 2/q and the observation 1/T, and is probably somewhat closer to the latter.

Setting the prior expected value of 1-Z to .5 gives q=4p/n, which is a way of picking q once p has been selected. Alternatively, p=nq/4 could be used to set p after q has been selected. Since S gets greater weight than T, the selected q probably has more bearing on the resulting Z than does p. As an example, suppose q=.2 is selected for Example 5.2. This is in the general area of $t^2+s^2/n=.1311$, but not particularly close. Since n is 6, p can be taken as .3. This gives posterior expected values of $E(s^2|S)=(.6+45S)/47$ and $E[(t^2+s^2/n)^{-1}]= 12/(.4+8T)$. For the three experiments, the following values are then generated, and the process is repeated for p=.6, q=.4:

Experiment	$E(s^2\|S)$	$E[(t^2+s^2/n)^{-1}\|T]$	1-Z
p=.3, q=.2			
1	.355	12.93 (=1/.077)	.765
2	.275	9.80 (=1/.102)	.449
3	.222	6.76 (=1/.148)	.250
p=.6, q=.4			
1	.367	9.04 (=1/.111)	.553
2	.288	7.39 (=1/.135)	.355
3	.235	5.51 (=1/.181)	.216

Either selection of priors seems to improve the estimation, but this test is somewhat unrepresentative, as the population Z is close to .5.

Diffuse priors could also be taken for s^2 and t^2+s^2/n, as in Example 2.3. With parameter p for this prior:

$$s^2|S \sim \text{Inverse Gamma: } N(n-1)S/2,\ -p-1+N(n-1)/2 \qquad (5.2a)$$

$$t^2+s^2/n|T \sim \text{Inverse Gamma: } (N-1)T/2,\ -p-1+(N-1)/2 \qquad (5.2b)$$

As discussed in Example 2.2, p=-1 makes the most sense for a diffuse prior. For comparison, p=-2 is also given below:

	$E(s^2\|S)$	$E[(t^2+s^2/n)^{-1}\|T]$	$1-Z$
p=-1	$N(n-1)S/(N(n-1)-2)$	$1/T$	$S/nT[1-2/N(n-1)]$
p=-2	S	$(N+1)/(N-1)T$	$S(N+1)/n(N-1)T$

Both are somewhat greater than S/nT. Note that if p=-2 for S and p=0 for T, 1-Z is the unbiased estimator (N-3)S/n(N-1)T. Neither of these estimates take into account the possible capping of 1-Z.

Regression Interpretation

Least squares credibility can be thought of as a least squares regression estimate in which the dependent variable has not yet been observed. The credibility estimate (4.12) can be rewritten as $\tilde{X}_{g0}-X_{..}=Z(X_{g.}-X_{..})$. Since the expected squared error is minimized by Z, this is similar to a no constant regression for $\tilde{X}_{g0}-X_{..}$, with $X_{g.}-X_{..}$ as the independent variable, where there is an observation for each risk g. The regression estimate of Z is computed by minimizing the sum of the actual square errors once X_{g0} is observed. A test of different methods of developing the credibility estimate then would be to compare Z to the regression estimate once the data is in.

Example 5.4

Efron and Morris (1975) computed the arcsin transforms of the batting averages for 18 players for their first 45 at bats in the 1970 season, as shown below, and used credibility methods to estimate the similar figure for the rest of the season. The reason for the arcsin transform is that it results in an approximately normal distribution with $s^2=1$. Thus only t^2 need be estimated to get Z.

Player	First 45	Rest of Season
Alvarado	-3.26	-4.15
Alvis	-5.10	-4.32
Berry	-2.60	-3.17
Campaneris	-4.32	-2.98
Clemente	-1.35	-2.10
Howard	-1.97	-3.11
Johnstone	-2.28	-3.96
Kessinger	-2.92	-3.32
Munson	-4.70	-2.53
Petrocelli	-3.95	-3.30
Robinson	-1.66	-2.79
Rodriquez	-3.95	-3.89
Santo	-3.60	-3.23
Scott	-3.95	-2.71
Spencer	-2.60	-3.20
Swodoba	-3.60	-3.83
Unser	-3.95	-3.30
Williams	-3.95	-3.43

From the data, $X_{..}=-3.317$, and T=1.115. Since n=1, S/nT=.897, and as the (N-3)/(N-1) factor is 15/17, an unbiased estimate of 1-Z is .791, or Z=.209. The regression estimate of Z is .186, which appears reasonably close. Relying on capping alone to correct S/nT would give Z=.103, which is not as close in this case. The inverse gamma prior for t^2+s^2/n with the prior Z of .5 gives Z=.221, which again is not as close as the factor approach.

Looking at just 3 batters at a time gives a different picture. Without considering capping, the unbiased estimate would be Z=1. For just 3 players, capping S/nT at 1 may in itself produce an unbiased estimate, however. Six different groups of 3 were selected from the above table, namely first 3, second 3, etc. For each of these 6 cases, the capped regression estimate of Z is compared to the capped crediblity estimate and the Z from the inverse gamma prior.

Case:	1	2	3	4	5	6
T:	1.679	2.455	3.072	1.748	0.304	0.041
1-Cap S/nT:	.404	.593	.675	.428	0	0
Inv Gamma:	.359	.535	.576	.478	.303	.257
Regression:	.378	.199	0	.351	0	0

It should be noted that the diffuse prior with p=-1 gives the capped estimate in this example. There is not an unambiguous winner between these two estimators of Z; it is not even clear whether the goal should be the regression Z from the 3 points, or the estimate of .186 from the wider population. It is apparent, however, that the unbiased estimate which ignores capping, i.e., Z=1, is not as close as the others.

Section 6 - Incorporating Risk Size

Up to this point, the exposure was assumed to be the same by risk and over time. In many applications (e.g., territory or class ratemaking, commercial lines experience rating), this is not a viable assumption, and it is removed in this section. For instance, in experience rating the formulation Z=E/(E+K) is often used to assign credibility to risks of different sizes, where E is expected losses. Larger risks will receive greater credibility, as their pure premiums, loss ratios, etc. will have lower variances than for smaller risks. The E/(E+K) formula is based on a particular relationship between the variances of risks of different sizes, namely $Var(X_{iu}|R_i)=s^2/E_{iu}$. That is, the variance is *inversely proportional* to risk size. With this assumption, it can be shown that $K=s^2/t^2$, where again t^2 is the variance of the hypothetical means.

It will be shown below that the inverse relationship of variance to exposure is a reasonable assumption, but that in fact it does not appear to hold in practice. A few other relationships will be explored to see which best accord with observation. Each of these will lead to different credibility formulas. In order to arrive at these formulas, a general formula will be developed that will hold for any relationship of variance to risk size, then the particular relationship desired can be just plugged in.

For the sake of concreteness, let X_{iu} be the pure premium for risk i time u, with L_{iu} the losses, and P_{iu} the exposure. By changing the definitions of P and/or L, X could just as easily be frequency, severity, loss ratio, etc. E.g., taking P_{iu} as the expected losses E_{iu} gives the experience rating credibility formula above.The general credibility formula is:

$$\tilde{X}_{g0} = (1-Z_g)m_g + Z_g X_{g.} \tag{6.1a}$$

$$Z_g = P_{g.}/(P_{g.}+K_g) \tag{6.1b}$$

$$K_g = P_{g.}/t_g^2 \Sigma_u s_{gu}^{-2} \tag{6.1c}$$

$$Z_{gu} = (1-Z_g)t_g^2/s_{gu}^2 \tag{6.1d}$$

Here $m_g=EX_{gu}$, $P_{g.}=\Sigma_u P_{gu}$, $X_{g.}=\Sigma_u Z_{gu}X_{gu}$, $t_g^2=VarE(X_{gu}|R)$, and $s_{gu}^2=EVar(X_{gu}|R)$.

To use this general formula, expressions are needed for m_g, t_g^2, and s_{gu}^2. These expressions will come from model assumptions, mainly assumptions about the relationship between variance and risk size.

Relationship of Variance to Risk Size

Since $X_{iu}=L_{iu}/P_{iu}$, the dependence of $Var(X_{iu}|R)$ on risk size will be approached by formulating the conditional variance of L_{iu} under different assumptions. This conditional variance can then be divided by P_{iu}^2 to yield $Var(X_{iu}|R)$.

L_{iu} is assumed to be the sum of the losses from P_{iu} exposure units. Let L_{iau} denote the losses from exposure unit a. If the exposure units are independent, then $Var(L_{iu}|R)=\Sigma_a Var(L_{iau}|R)$. If these units are conditionally identically distributed given R_i, $Var(L_{iau}|R)$ does not depend on a or u, and so can be denoted as $s(R_i)^2$. Then $Var(L_{iu}|R)=P_{iu}s_i(R)^2$. Thus $Var(X_{iu}|R)=s(R_i)^2/P_{iu}$. Letting $s^2=Es(R_i)^2$ gives $EVar(X_{iu}|R)=s^2/P_{iu}$. Hence, assuming that the risk is a collection of independent identically distributed exposure units yields that the expected conditional variance for a risk decreases in proportion to the exposure.

Hewitt (1967) showed that for a body of risks, the variance did not decrease this fast. The first two columns below derive from that paper.

Average Premium	Estimated Variance	13,150/Prem	.172+ 9900/Prem	$\frac{1837}{Prem^{.773}}$	$\frac{12,230+.133Prem}{254+Prem}$

296	26.3	44.4	33.6	22.7	22.3
628	12.3	20.9	15.9	12.7	14.0
869	10.4	15.1	11.6	9.80	11.0
1,223	7.58	10.7	8.27	7.58	8.39
1,924	5.35	6.83	5.32	5.32	5.73
3,481	3.07	3.78	3.02	3.36	3.40
6,050	2.18	2.17	1.81	2.19	2.07
8,652	1.59	1.52	1.32	1.66	1.50
12,265	1.15	1.07	.980	1.27	1.11
18,944	.749	.694	.695	.906	.769
33,455	.610	.393	.468	.585	.495
68,758	.345	.191	.316	.335	.310
220,786	.163	.060	.217	.136	.188

The variance in this case was not of the pure premium, but of the entry ratio, which is the loss ratio normalized to average to 1. The dollars are at 1958 levels. The other columns are fits of the variance by various functions of premium. The first of these functions specifies that the variance decreases by the inverse of premium. It can be seen that the actual variances are lower than this model would predict for small risks, and higher for large risks.

The difficulty for other functions of premium, however, is *finding models that explain them. Such models would have to* incorporate exposure units that are not conditionally independent given the risk parameter R_i.

One such model is provided by including the possibility that there are varying conditions that affect the risk, so that the loss probabilities are not the same in every year. For instance the risk parameters R_i could specify a distribution from which another parameter H_{iu} is determined each year. If the exposure units are conditionally independent given H_{iu}, then given only R_i they are not independent; they have some correlation due to the common parameter H_{iu}. By the above reasoning, $Var(X_{iu}|H_{iu}) = s(H_{iu})^2/P_{iu}$. Then $Var(X_{iu}|R_i) = EVar(X_{iu}|H_{iu}) + VarE(X_{iu}|H_{iu})$, which can be written as $Var(X_{iu}|R_i) = s^2(R_i)/P_{iu} + y^2(R_i)$.

Thus with the inclusion of varying conditions, the conditional variance becomes a linear function of 1/P. The constant term essentially measures how much variance there is over time.

The second fit of the variance shown above uses this linear function. A much better fit to the risk variances is produced, although the smallest and largest risks still do not fit very well. It could be that the large risks are qualitatively different, and that linear functions could be used with different parameters for large and small risks. In a similar application of the linear model, Meyers and Schenker (====) do just that.
The final two columns represent (1) Hewitt's fit to this data based on $Var = s^2/P^C$, and (2) the function $Var = [y^2+s^2/P]/[1+C/P]$.

Neither of these is based on a model decomposing L_{iu} into exposure units, but improved fits are provided. The latter formula approaches a linear function of 1/P for large risks, but is below that line for the small risks. For all the curves, the parameters were selected to minimize squared errors in the log of the variance, so that percentage errors in the variance would be as small as possible.

To review, then, four formulas relating conditional variance to risk size have been considered. The first two are based on models of the risk process, and the second two are just curves providing better fits. Since the conditional, or "process", variance of X_{gu} is a fuction of the exposure P_{gu}, then the expected value of this variance will be also. That is, $s_{gu}^2 = EVar(X_{gu}|R)$, the expected process variance for the gth risk at time u, is a function of P_{gu}. For the four curves these functions are as follows:

1. $s_{gu}^2 = s^2/P_{gu}$
2. $s_{gu}^2 = y^2 + s^2/P_{gu}$
3. $s_{gu}^2 = s^2/P_{gu}^{.773}$
4. $s_{gu}^2 = [y^2 + s^2/P_{gu}]/[1+C/P_{gu}]$

Each of these can be put into (6.1) to produce a credibility formula. This is done below, after two examples of negative binomial claim frequency distributions corresponding to the first two models.

Example 6.1

In this example, L_{iu} will be the number of claims, so that X_{iu} is claim frequency. The parameter R_i is the ordered pair $\langle V_i, Q_i \rangle$, and L_{iu} is assumed to be negative binomially distributed with parameters $P_{iu}V_i$ and Q_i. (The sum of the claims for P_{iu} independent exposure units, each negative binomial in V_i, Q_i is itself negative binomial in $P_{iu}V_i$, Q_i.) These assumptions yield:

$$E(L_{iu}|R_i) = P_{iu}V_i(1-Q_i)/Q_i$$

$$E(X_{iu}|R_i) = V_i(1-Q_i)/Q_i$$

$$Var(L_{iu}|R_i) = P_{iu}V_i(1-Q_i)/Q_i^2$$

$$Var(X_{iu}|R_i) = V_i(1-Q_i)/P_{iu}Q_i^2$$

Thus the conditional variance of X_{iu} is inversely proportional to the exposure P_{iu}.

Example 6.2

The claims for each exposure unit are assumed to be Poisson with parameter H_{iu}, so that L_{iu} is Poisson in $Y_{iu}=P_{iu}H_{iu}$. H_{iu} is in turn gamma distributed in $R_i=\langle B_i,C_i\rangle$, and so Y_{iu} is gamma in $P_{iu}B_i$, C_i. Thus from Example 2.1, L_{iu} is negative binomial in C_i, $1/(1+P_{iu}B_i)$. Thus:

$$E(L_{iu}|R_i) = P_{iu}B_iC_i$$

$$E(X_{iu}|R_i) = B_iC_i$$

$$Var(L_{iu}|R_i) = P_{iu}B_iC_i(1+P_{iu}B_i)$$

$$Var(X_{iu}|R_i) = B_iC_i/P_{iu} + B_i^2C_i$$

This is then an example of the second variance formula, a linear function of 1/P.

Credibility Formulas Varying By Risk Size

Once an expression relating the variance for different risk sizes has been selected, (6.1) can be used to produce a credibility formula. If $s_{gu}^2 = s^2/P_{gu}$, as in the first model above, then $\Sigma_u s_{gu}^{-2}=P_{g.}/s^2$, and so $K_g=s^2/t^2$. Thus K_g is a constant, as in the constant exposure case, and $Z_g=P_{g.}/(P_{g.}+K)$.

For the other models, K_g is more complex. However, a fairly simple expression is possible in the case of just one observed time period. In the second model, $s_{gu}^{-2}= P_{gu}/(P_{gu}y^2+s^2)$ and $P_{gu}=P_{g.}$, so $K_g= (P_{g.}y^2+s^2)/t^2$, which can be written $K_g=P_{g.}A+B$, i.e., a linearly increasing function of the exposure. In this case $Z_g=P_{g.}/((1+A)P_{g.}+B)$.

If $s_{gu}^2=s^2/P_{gu}^{.773}$, in the case of one exposure period, s^2 is given by $s_{gu}^{-2}= P_{gu}^{.773}/s^2$, so $K_g= P_{g.}^{.227}(s^2/t^2)$, or $K_g= BP_{g.}^{.227}$, *again an increasing function of* $P_{g.}$. *The formula for* Z becomes $Z=P_{g.}^{.773}/(P_{g.}^{.773}+B)$.

Finally, if $s_{gu}^2=[y^2+s^2/P_{gu}]/[1+C/P_{gu}]$, and there is only one exposure period, $t_g^2 s_{gu}^{-2}= [1+C/P_{g.}]/[(y_g^2/t_g^2)+s_g^2/t_g^2P_{gu}]$, so $K_g= [AP_{g.}+B]/[1+C/P_{g.}]$. With this, $Z=P_{g.}/(P_{g.}+K_g)$ yields, after some algebra, $Z_g=[P_{g.}+C]/[P_{g.}(1+A)+B+C]$. By redefining the constants, this can also be written as $Z_g=[P_{g.}+C]/[AP_{g.}+B]$. An interpretation of this formula based on heterogeneity of exposure units within a risk is given by Mahler (1987).

An important difference between (4.12) and (6.1a) is that the complement of credibility goes to $X_{..}$ in the former and in the latter to $m=EX_{g0}=EE(X_{g0}|R)$. In (4.12) $X_{..}$ is also the minimum variance unbiased linear estimate of m.

In the unequal exposure case a weighted average of the X_{iu}'s can be used to estimate m. However the usual exposure weighted average is not optimal. At least for the simplest model $s_{gu}{}^2=s^2/P_{gu}$, it turns out that the minimum variance unbiased linear estimator of m, which will again be denoted $X_{..}$, is $X_{..}=\Sigma_i Z_i X_{i.}/Z_.$, where $Z_.=\Sigma_i Z_i$ (see ISO (1983)). This is sometimes referred to as the credibility weighted average of the $X_{i.}$'s. Standard statistical practice advocates weighting observations in inverse proportion to their variances. In this case $Var(X_{i.}) = t^2 + s^2/P_i = t^2/Z_i$., so the credibility is inversely proportional to the variance.

<u>Estimation of Z</u>

To estimate s^2, y^2, t^2, etc., extensions of the methods used in the equal exposure case can be used. First, the model with $s_{iu}{}^2=s^2/P_{iu}$ will be addressed. Let $S_i=\Sigma_{u=1}^n P_{iu}(X_{iu}-X_{i.})^2/(n-1)$, where here $X_{i.}= \Sigma_u P_{iu}X_{iu}/P_{i.}$, and let $S=\Sigma_{i=1}^N S_i/N$.

By repeated use of the formula $cov(X_{iu},X_{iv}|R_i)=\delta_{uv}s^2(R_i)/P_{iu}$, enough algebra (Appendix 2) will show that $E(S_i|R_i)= s^2(R_i)$. Thus $E(S_i)=s^2$, and $ES=s^2$ as well. S is a lower variance unbiased estimator of s^2 than is S_i.

Buhlmann and Straub(====) propose the following to estimate t^2. Let $W=\sum_{i,u}P_{iu}(X_{iu}-X)^2/(Nn-1)$, where X is the usual exposure weighted average of the X_{iu}'s. It can be shown that $EW=s^2+qt^2$, where $q=\sum_g P_{g.}(1-P_{g.}/P_{..})/(Nn-1)$. Thus $(W-S)/q$ is an unbiased estimator of t^2. As they point out, this can sometimes be negative, in which case they assign $t^2=0$, and so $Z=0$.

Klugman (1985) gives an alternative approach, which appears to be more accurate. Let $T=\sum_{i=1}^{N} Z_i(X_{i.}-X_{..})^2/(N-1)$. In Appendix 2 it is shown that, given the Z_i, $ET=t^2$. T cannot be considered an estimator of t^2, because t^2 is needed to compute Z_i in the formulas for $X_{..}$ and T. However if Z_i is initially set to 1, an iterative procedure can be used to compute $X_{..}$ and T, estimate t^2, compute new Z_i's, etc., until the estimate for t^2 stabilizes (usually quickly). DeVylder (1981) uses the term pseudo-estimator for such a T, and suggests another one.

Klugman (1986) details several Bayesian approaches, and shows that these can give dramatically improved credibilities. One of these generalizes the diffuse priors used in (5.2), by specifying that the joint prior density of s^2 and t^2 is proportional to $s^{-2}[\Pi_i(s^2+P_{i.}t^2)]^{-1/N}$. This particular prior is taken after Box and Tiao (1973, p. 426). Introducing the variable $r=t^2/s^2$, and defining $w_i=rP_{i.}/(1+rP_{i.})$ and $w=\sum w_i$, the posterior distribution for r given the observations X_{iu} turns out to be proportional to:

$$f_1(r) = [r\Sigma_{i,u}P_{iu}(X_{iu}-X_{i.})^2 + \Sigma_i w_i(X_{i.}-X_{..})^2]^{-(N-1)/2} \times$$
$$\Pi_i[(1+P_{i.}r)^{-1/N}(w_i/w)^{.5}].$$

This must be integrated numerically from zero to infinity to find the constant of proportionality. Dividing $f_1(r)$ by this constant gives the conditional density $f(r)$. Then $E(r|\text{the } X_{iu}\text{'s}) = \int rf(r)dr$, which again is done numerically. This gives an estimate for r, and K can then be estimated by 1/r. Alternatively, E(1/r) could be calculated directly by numerical integration.

To estimate K for the model $s_{gu}{}^2=y^2+s^2/P_{gu}$, some algebra will show $E(S_i)= s^2+y^2(P_{i.}{}^2-\Sigma_{u=1}^n P_{iu}{}^2)/(n-1)P_{i.}$. Thus if a linear regression is done for S_i against $(P_{i.}{}^2-\Sigma_u P_{iu}{}^2)/(n-1)P_{i.}$, the slope and intercept can be used as estimators of s^2 and y^2.

Estimation of t^2 in this case could perhaps be done as follows. Let $w_i=P_{i.}t^2/(P_{i.}t^2+s^2+y^2\Sigma_u P_{iu}{}^2/P_{i.})$ and $w=\Sigma_i w_i$. Define $X_{..}= \Sigma_i w_i X_{i.}/w$ and $T=\Sigma_{i=1}^N w_i(X_{i.}-X_{..})^2/(N-1)$, where again $X_{i.}$ is the exposure weighted average of the X_{iu}'s. In this case it can be shown that $ET=t^2$, and so T is an unbiased pseudo-estimator of t^2. However, for this and the more complex models, the Bailey-Simon method is often used instead, as discussed below.

Bailey and Simon (1959) presented the idea of estimating Z by seeing which values of Z would have worked best in the past. In their example, each risk had one unit of exposure, namely a single private passenger car. For the risks with no claims, the credibility estimate of X_{g0} is just $(1-Z)X_{..}$. This can be compared to the average experience for these risks in the next year to see what Z should have been. Since only a fixed number of years (usually 1 to 5) are used in automobile experience rating, this value of Z can then be used in the future.

Meyers (1985) uses a similar retrospective approach to estimate A and B in $Z_g=P_{g.}/(AP_{g.}+B)$ in commercial insurance experience rating. Rather than focusing on the zero loss risks, Meyers creates a test statistic for the overall performance of the plan, and optimizes the test statistic. NCCI adopted a similar procedure with a different test statistic to estimate A, B, and C in $Z_g=(P_{g.}+C)/(AP_{g.}+B)$ for workers compensation experience rating.

<u>Section 7 - *How Good Is Least Squares Credibility*</u>

As discussed earlier, the function of the X_{iu}'s that optimizes the expected squared error in X_{g0} is the conditional expectation $E(X_{g0}|\text{the } X_{iu}\text{'s})$. The best <u>linear</u> function in this sense is the least squares credibility estimate.

Discussion Of "A Bayesian Credibility Formula For IBNR Counts"

By Gary Venter

Discussion of
A Bayesian Credibility Formula for IBNR Counts

by Gary Venter

If an indicator of a significant paper is that it opens the door for further research, Dr. Robbin's paper should stand the historical test. This review will emphasize generalizing the Poisson assumptions of the paper; attention to optimal parameter estimation and other model assumptions may also prove fruitful, as may the quantification of uncertainty in the IBNR estimates.

The three way credibility weighting for IBNR is an interesting result of the paper. Credibility weights are specified for three estimators of IBNR:

(i) the original (e.g., pricing) expected claims less the observed claims to date

(ii) the observed claims to date times a development factor

(iii) the original expected claims less the expected claims to date.

To see the origin of these credibility weights, a slightly more general framework will be used here. A vector of parameters, u, is postulated to determine the distribution of N, the ultimate number of claims, M, the observed claims to date, and R, the IBNR claims.

It is assumed that M and R are conditionally independent given u. Further, n and q are functions of u, and s^2 is a positive constant with:

$$E(N|u) = n$$
$$E(M|u) = n(1-q)$$
$$E(R|u) = nq$$
$$EV(M|u) = s^2$$

This last assumption generalizes the Poisson assumption of the paper, where the expected conditional variance of M was EnE(1-q).

It is also assumed that u is a vector of random variables such that n and q are independent.

The fundamental credibility formula from Robbin, section III.1, is then invoked to estimate R:

$$R^* = ER + (M-EM)C(M,R)/VM.$$

From the assumptions, ER = EnEq and EM = EnE(1-q) = En(1-Eq) = En-EnEq. Also $VM = EV(M|u) + VE(M|u) = s^2 + V(n(1-q)) = s^2 + E(n^2(1-q)^2) - E(n(1-q))^2$. Then by the reasoning of B.3.(ii) of the paper, $VM = s^2 + E(n^2)V(1-q) + E(1-q)^2Vn$.

These three components of the variance of the observed claims, when divided by that variance, will turn out to be the three credibility weights to be applied to the three IBNR estimators (i) - (iii) above. To see this, a general formula on covariances is used to compute C(M,R):

$$C(M,R) = EC(M,R|u) + C(E(M|u),E(R|u)).$$

Because of the conditional independence of M and R, the first term is zero, and so $C(M,R) = C(n(1-q),nq)$

$$= E(n(1-q)nq) - E(n(1-q))E(nq)$$

Then, by the reasoning of B.3.(i) of the paper, $C(M,R) = VnEqE(1-q)-E(n^2)V(1-q)$. Plugging all of this back into the original credibility formula gives:

$$R^* = EnEq + (M + EnEq - En)\ [VnEqE(1-q) - E(n^2)V(1-q)]/VM.$$

This is regrouped into Robbin's three way credibility formula as follows: first combine the EnEq terms; apply M-En to the second term in brackets to yield $\underline{(En-M)E(n^2)V(1-q)/VM}$. When applied to the first term in brackets the M and En are separated, giving

a) En combined with Eq and adding to the EnEq component; and

b) $\underline{M[Eq/E(1-q)]VnE(1-q)^2/VM}$. The underlined terms are the IBNR estimators (i) and (ii) times credibility weights, where the weights are the second and third components of the variance VM above, divided by VM.

This interprets Eq/E(1-q) as a development factor, and in fact by the hypotheses above, ER/EM = Eq/E(1-q) and EN/EM = 1/E(1-q). This corresponds to the method of estimating LDF's from several accident years' data by $\sum N_i\ /\ \sum M_i$, as recommended by Stanard (PCAS 1985). With this definition of the LDF, the mathematically imprecise estimate of the development factor used by Dr. Robbin becomes unnecessary.

Finally the remaining terms of R^* can be algebraically combined to yield the credibility weight of s^2/VM applied to EnEq. Writing EnEq as En - EnE(1-q) shows this term to be the original expected claims less the expected claims to date.

The assumption that M and R are conditionally independent may be somewhat limiting. The possibility that some claims come in earlier than usual, so fewer come in later, or vice versa, suggest that R and M are not unconditionally independent. Assuming they are conditionally independent then attributes their correlation to non-independent parameters. But this suggests that the parameters are different from year to year. If the claims reported before and after a given point are each modelled as conditionally independent draws from a fixed, possibly unknown, report lag distribution, a negative correlation between reported and unreported claims would not be anticipated.

Dr. Robbin is to be congratulated for this thought provoking and potentially useful paper. He has proven his main point: a Bayesian credibility formula for IBNR does count.

Adjusting Loss Development Patterns For Growth

By Charles McClenahan

Adjusting Loss Development Patterns for Growth

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.

Charles McClenahan

Adjusting Loss Development Patterns for Growth

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 deals with the specific impact of such growth patterns upon exposure-based IBNR factors. (LeRoy J. Simon, "Distortion in IBNR Factors" *PCAS LVII*, 1970 p.64)

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.

HYPOTHETICAL CASE STUDY

Appendix I 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	% Error
1984	$1,323	$1,329	-0.5%
1985	177,979	175,723	1.3
1986	944,466	879,471	7.4
Total	$1,123,768	$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.

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 II 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 $\mathcal{L}_i^g$ to be the observed proportion of ultimate losses at accident year age i:

$$\mathcal{L}_i^g = \int_{i-1}^{i} (1+g)^{i-x} (1-a^x)\, dx \qquad i \geq 1$$

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

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

$${}_{i-1}\mathcal{F}_i^g = \frac{\mathcal{L}_i^g}{\mathcal{L}_{i-1}^g} \qquad i \geq 2;\ g \neq 0$$

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

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

$${}_{i-1}\mathcal{F}_i^g = \frac{c - ba^i}{c - ba^{i-1}} \qquad [2]$$

In the special case where g=0:

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

$$_{i-1}\mathcal{F}_i^0 = \frac{\ln(a) + a^{i-1}(1-a)}{\ln(a) + a^{i-2}(1-a)} \qquad i \geq 2 \qquad [3]$$

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.

EXAMPLE OF PROCESS

Going back to the hypothetical case outlined in Appendix I, 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 [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 [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	Theoretical Development Factors		
Year	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 I.

Application of [3] provides the following "growth-free" factors:

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

Implying the following factors to adjust to a "growth-free" basis:

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

And the following factors to 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	% 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%.

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 III 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

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 repid growth or withdrawal occur, however, this approach provides a relatively simple and efficacious basis for improving estimates of ultimate losses.

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 on Sheet 2, 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	705,367	1,318,846	1,469,650	
1985	832,041	1,542,366		
1986	875,722			

Appendix I
Sheet 2

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

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 $1 - a^6 = .8386 \Rightarrow a = .7379$

Factors to Adjust to "Growth-Free" Basis

g	a=.250 1-2	a=.250 2-3	a=.250 3-4	a=.600 1-2	a=.600 2-3	a=.600 3-4	a=.800 1-2	a=.800 2-3	a=.800 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
0	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

Revisions In Loss Reserving Techniques Necessary To Discount Property-Liability Loss Reserves

By Stephen P. D'Arcy

Revisions in Loss Reserving Techniques Necessary to Discount Property-Liability Loss Reserves

by

Stephen P. D'Arcy, F.C.A.S.
Assistant Professor
Department of Finance
University of Illinois at Urbana-Champaign

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 the 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 and 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.

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. However, months prior to August, 1986, 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.

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 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 1960's 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 number 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 all known claims. 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 claims 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. However, any inaccuracy in the loss ratio estimate 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 influence 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 below:

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.

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

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

where $\frac{C \cdot u}{C \cdot j}$ = 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 a 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:

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

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 represents the following:

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

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 the following:

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

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.

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

$P_{\cdot j}$ = 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 $\sum_{j=1}^{n} P_{\cdot j} = 1$. 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:

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

where $E_{i,j+1}$ = 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:

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

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:

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

where f is the first accident year with losses still outstanding

ℓ is the latest accident year

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

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 ratios 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]. However, the risk adjustment factors 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 primarly 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 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 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, are 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 a 5 percent and 10 percent interest rate 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 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 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 <u>reported</u> 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.

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 reduces from 161.5 percent to an almost profitable 100.8 percent.

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.

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 was 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 changes in interest rates do not affect the accuracy of statutory loss reserve levels for almost all cases. 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.

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 reduce 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] Anderson, D. R., "Effects of Under and Overevaluations in Loss Reserves," Journal of Risk and Insurance, Vol. 38, No. 4 (December, 1971), pp. 585-600.

[3] Balcarek, R. J., "Loss Reserving in the Sixties," Proceedings of the Casualty Actuarial Society, Vol. LIX (1972), pp. 18-25.

[4] Berquist, J. R. and Sherman, R. E., "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," Proceedings of the Casualty Actuarial Society, Vol. LXIV (1977), pp. 123-184.

[5] Cummins, J. David and Chang, Lena, "An Analysis of the New Jersey Formula for Including Investment Income in Property-Liability Insurance Ratemaking," Journal of Insurance Regulation, Vol. 1, No. 4 (June, 1983), pp. 555-573.

[6] Cummins, J. David and Harrington, Scott, "Property-Liability Insurance Rate Regulation: Estimation of Underwriting Betas Using Quarterly Profit Data," Journal of Risk and Insurance, Vol. 52, No. 1 (March, 1985), pp. 16-43.

[7] D'Arcy, Stephen P., "Discussion of 'Duration' by Ronald E. Ferguson," Proceedings of the Casualty Actuarial Society, Vol. 71 (1984), pp. 8-25.

[8] Ferguson, Ronald E., "Duration," Proceedings of the Casualty Actuarial Society, Vol. 70 (1983), pp. 265-288.

[9] Fisher, M. J., "Industry Trade Groups Hit GAO Report on Taxation," National Underwriter Property & Casualty Insurance Edition, Vol. 88, No. 4 (March 2, 1984), pp. 1, 32.

[10] Fisher, M. J., "P&C Tax Study Questions Loss Reserves, Expenses," National Underwriter Property & Casualty Insurance Edition, Vol. 88, No. 5 (February 3, 1984), pp. 1, 43.

[11] Fogler, H. Russell, "Bond Portfolio Immunization, Inflation, and the Fisher Equation," Journal of Risk and Insurance, Vol. 51, No. 2 (June 1984), pp. 244-264.

[12] Forbes, S. W., "Loss Reserving Performance Within the Regulatory Framework," Journal of Risk and Insurance, Vol. 37, No. 4 (December, 1970), pp. 527-38.

[13] "From Treasury's Proposal on Discounting of Loss Reserves," National Underwriter Property & Casualty Insurance Edition, Vol. 88, No. 50 (December 14, 1984), pp. 3,39.

[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] Greenwald, J., "Insurers' Sad Song May Play One More Time," Business Insurance, Vol. 18, No. 13 (March 26, 1984), pp. 1, 32-33.

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

[17] Kahane, Yehuda and Porat, M. Moshe, "Financial Analysis of Underwriting Results--A Corrected Approach to Loss Ratio Analysis, with a Special Reference to Inflation," Journal of Risk and Insurance, Vol. 51, No. 4 (December, 1984), pp. 710-719.

[18] McClenahan, Charles L., "A Mathematical Model for Loss Reserve Analysis," Proceedings of the Casualty Actuarial Society, Vol. 62 (1975), pp. 134-153.

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

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

[21] Pointek, S., "Despite Record Underwriting Loss, Industry Netted $2 Billion in 1985," National Underwriter Property & Casualty Insurance Edition, Vol. 90, No. 13 (March 28, 1986), pp. 1, 24.

[22] Pointek, S., "Industry Loss Reserves Are 10 Percent Short: ISO," National Underwriter Property & Casualty Insurance Edition, Vol. 88, No. 3 (January 20, 1984), pp. 1, 37.

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

[25] Skurnick, David, "A Survey of Loss Reserving Methods," Proceedings of the Casualty Actuarial Society, Vol. 60 (1973), pp. 16-58.

[26] Smith, B. D., "An Analysis of Auto Liability Loss Reserves and Underwriting Results," Journal of Risk and Insurance, Vol. 47, No. 2 (June, 1980), pp. 305-20.

[27] Smith, Barry D., "Analyzing the Tax Deductibility of Premiums Paid to Captive Insurers," Journal of Risk and Insurance, Vol. 53, No. 1 (March 1986), pp. 85-103.

[28] Smith, Michael L. and Witt, Robert C., "An Economic Analysis of Retroactive Liability Insurance," Journal of Risk and Insurance, Vol. 52, No. 3 (September 1985), pp. 379-401.

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

[30] Weiss, Mary, "A Multivariate Analysis of Loss Reserving Estimates in Property-Liability Insurers," Journal of Risk and Insurance, Vol. 52, No. 2 (June, 1985), pp. 199-221.

D/292

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

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

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

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.[illegible]
Expense Ratio	32.2	32.2	32.2	17.2	17.2	17.2	4.7	4.7	4.[illegible]
Combined Ratio	127.9	118.5	111.3	103.0	100.9	99.2	161.5	125.7	100.[illegible]
Net Written Premium to Surplus Ratio	1.60	1.24	1.06	0.96	0.93	0.90	3.71	0.68	0.[illegible]

Position Paper On The Methodologies And Considerations Regarding Loss Reserve Discounting

By The CAS Committee
On Reserves

A DISCUSSION PAPER

BY THE COMMITTEE ON RESERVES

POSITION PAPER ON THE METHODOLOGIES AND CONSIDERATIONS REGARDING LOSS RESERVE DISCOUNTING

The appropriateness of discounting loss reserves for property and casualty insurance company financial statements has been discussed and debated in many forums over the years. The insurance industry, insurance regulators, legislators, insurance accountants and actuaries have all contributed to the evaluation of this controversial issue. However, there are technical aspects of recognizing the time value of money that may not be well understood by all those involved in the implementation of loss reserve discounting.

The purpose of this paper is to describe and discuss the methodologies and considerations pertaining to loss reserve discounting. The broader issue of the appropriateness of discounting is not addressed in this paper. Instead, this statement of the Casualty Actuarial Society's Committee on Reserves is a discussion of the technical and theoretical considerations underlying the process given that discounting is deemed appropriate.

In most circumstances, the reserve discounting process is largely determined by the underlying loss reserve evaluation process and the governing accounting principles. The discounting issues surrounding the different reserve evaluation techniques and accounting contexts will thus be discussed. It is assumed that the reader of this paper is familiar with the Committee on Reserves' "Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves" and has a general understanding of statutory and generally accepted accounting principles for property and casualty insurance entities.

This paper consists of two sections:

I. Definitions

II. Methodologies and Considerations

Section II is further divided into four subsections:

A. Ultimate Payment Values

B. Payment Timing

C. Interest Rates for Discounting

D. Other Considerations

SECTION I - DEFINITIONS

The definitions in this section are specific to the discounting process. Definitions pertaining to loss reserve evaluations, which are contained in the previously referenced Statement of Principles on Loss Reserves, are not included here but are assumed to be understood by the reader.

A full value reserve is defined as a provision for the payment of outstanding claims at the anticipated future settlement amount. A full value reserve reflects future inflation as it may affect unpaid claim amounts. It does not reflect the time value of money.

A discounted reserve is defined as a full value reserve reduced for future investment income earnings that can be generated by funds held between the date of the valuation and the date of the final payment on outstanding claims. A discounted reserve is thus a full value reserve adjusted to reflect the time value of money.

In discussing discounted reserves, there are two general types of investment yields or interest rates that arise. One is the market interest rate, which corresponds to the possible yield on new money invested in the current market. Such an interest rate is therefore dependent on the current performance of the selected security(ies). The other is the portfolio interest rate, which corresponds to the average yield on an existing investment portfolio.

Consideration of investment yields for discounting purposes also gives rise to two additional categories of risk. One is investment risk, which corresponds to the uncertainty surrounding the realization of a specified investment income stream. Two elements of the investment risk include uncertain investment yields and uncertain investment liquidity.

The other category of risk is default risk. This corresponds to the possibility of a complete and total loss on a chosen investment security.

SECTION II - METHODOLOGIES AND CONSIDERATIONS

Typically, the loss reserve discounting process follows these steps:

- Estimate the full value reserves for the group of claims under consideration as of the specified valuation date.

- Estimate the future loss payout patterns for the same group of claims.

- Apportion the full value reserves to the future payment periods using the estimated payout patterns.

- Determine an appropriate discounting rate of return.

- Calculate the present value, as of the valuation date, of the projected payments for each future payment period using the selected discounting rate of return.

- Cumulate the present value payments for all future payment periods.

There are, of course, many variations on this discounting process. In fact, the initial calculation of a full value reserve is not always necessary. Even so, the process can appear deceptively simple. A detailed analysis of each of the steps described above indicates that the process can be much more complex, depending on the volatility of the lines of business and the accounting context.

The most critical issues concerning the discounting process are discussed in the following four subsections.

A. Ultimate Payment Values

The calculation of a discounted reserve usually involves, as a starting point, an estimate of the full value reserve. The reader is referred to the previously mentioned Statement of Principles for a discussion of the principles and considerations involved in evaluating full value reserves. The same principles and considerations would apply to the calculation of a full value reserve which will form the basis of a discounted reserve.

The reserve provision finally recommended by an actuary will frequently be selected after a review of the results of several different reserving techniques. When a full value reserve is the sole objective, the aggregate reserve amount is of primary importance. When a full value reserve is to be used as the starting point for the calculation of a discounted reserve, the selection of the reserve provision by its component parts (e.g., accident year) may have a material effect on the amount of reserve discounting. Special attention should be paid to the projections by component in this instance.

The actuary calculating the discounted reserve provision should be fully aware of the assumptions and considerations underlying the selection of the full value reserve. Many of these assumptions and considerations may have a material effect on the determination of the expected payment pattern. The selection of the payment pattern should be consistent with the selection of the full value reserve, to the extent that the key assumptions can be identified and their effect quantified. This point is discussed in further detail in the Payment Timing subsection.

While many of the same principles and considerations of reserve evaluation apply to both full value and discounted reserves, the materiality of those considerations may differ significantly. For example, the selection of a development factor for estimated development at an advanced maturity (i.e., a "tail factor") may be very significant in the determination of a full value reserve but may not be nearly as significant for a discounted reserve. On the other hand, a change in the settlement rate of claims may not materially affect the amount of a full value reserve, but could be important in the calculation of a discounted reserve. To the extent that the materiality of a reserve consideration determines the amount of analysis that item receives, the evaluation of a discounted reserve may require a change in emphasis on the items analyzed.

The accounting treatment of reserves where discounting is permitted frequently requires the disclosure of the amount of discounting; i.e., the amount of difference between the full value and the discounted reserves. Where disclosure is not required, it is recommended that such a disclosure still be included, as it permits the use of financial evaluation ratios or other indicators which traditionally use non-discounted reserve amounts.

Although most discounted reserves are calculated from an initial full value reserve, there are instances where this is not the case. For example, some discounted reserves are calculated using an assumed difference between future claim cost trend and future interest rates. In this case, a meaningful disclosure may require that the reserve method and assumptions be identified and a representative interest rate be used to calculate a full value reserve, for illustrative purposes.

B. Payment Timing

To determine the timing of the loss and loss adjustment expense payments for an insurance enterprise, the entity's own historical payment data should be used to the extent that credible data is available. If necessary, this data should be supplemented by appropriate data from a broader source, such as insurance industry composites. Any such supplementary data should reflect the payment timing characteristics of the category of business under consideration; i.e., the data should be drawn from the same line (or sub-line) of business and policy type (e.g., claims-made vs. occurrence, primary vs. excess, etc.) to the extent possible.

The data used in the estimation of payment timing should comprise several exposure periods and evaluation dates. Ideally, a complete development triangle of payment data should be evaluated. Specific techniques for estimating payment timing from this data include, for example:

- determining payment development factors directly from the triangle and deriving a pattern by inverting the resulting paid-to-ultimate development factors; and

- calculating a triangle of paid-to-ultimate ratios from the payment triangle and selecting a payment pattern by examination of these ratios by exposure period.

Methods such as these that use triangular data are less subject to distortion than cross-sectional methods that use only a single payment valuation date to derive a payment pattern.

In many situations, publicly available data, e.g., statutory annual statement data, may be the source of the payment timing experience. The actuary should be aware of the limitations of such data in selecting a loss payment pattern for discounting purposes.

As indicated before, the loss payment timing estimates should be reconcilable to the estimates of ultimate amounts to be paid even if these latter estimates have not been derived by techniques based on paid losses. Of particular importance is the allocation of the amount of reserve discount to exposure period. Such distribution is dependent on the allocation of the full value reserve. The actuary should determine if the resulting reserve discount allocation is consistent with that implied by the payment timing estimates.

It is possible that future loss payments could be subject to influences not present during the period when historical data is available. The actuary should determine whether this might be the case. Estimates of payment timing should reflect conditions (both internal and external) expected to prevail during the future payment period. If such conditions are different from those prevailing during the historical evaluation periods, attempts should be made to adjust the payment timing indications from the historical data.

Payment timing information should be examined periodically as data becomes available. Should information become available that would cause a material change in the estimated timing of payments for a particular category of business, the change should be reflected immediately.

The actuary should determine whether optimal payment timing estimates would be obtained by treating losses, allocated loss adjustment expenses and unallocated loss adjustment expenses separately or in some combination. This determination typically is influenced by the nature of the available data.

In estimating discounted loss reserves on a basis net of ceded reinsurance, the timing of the expected reinsurance recoveries should be considered. In particular, it should be determined whether the timing of such recoveries will affect the entity's investment income on its net business and thus its discounted net reserves. Adjustments to the amount of reserve discount should be made as necessary.

Special consideration should be given to loss payments which will be made according to a fixed schedule, such as structured settlements and workers' compensation lifetime benefit cases. If the volume of such cases is significant, separate treatment of fixed schedule claims may be appropriate.

In evaluating the timing of loss payments, a range of reasonable payout patterns may become evident. In selecting from within the range, the actuary should consider the volatility of the line of business and the investment risk. Further, consideration should be given to the purpose of the discounted reserve within the specific financial reporting context.

C. Interest Rates for Discounting

Since the amount of reserve discount depends heavily upon the interest rate used in discounting the expected loss payments, the rate must be chosen carefully. Because the insurer's actual asset portfolio will usually involve various types of risk (specified below), the interest rate should be based upon a hypothetical asset which negates these risks. Several elements must be considered:

1. Since interest rates vary through time, the actuary must select an interest rate consistent with the reserve evaluation date. Generally, a market rate is preferable to a portfolio (based on amortized value) rate. This is because the economic value of a loss reserve is its worth in exchange, through reinsurance or some other medium. In such a case, the appropriate interest rate is one for which a cash amount could be invested at the evaluation date to exactly liquidate the future loss payments as they become due. Further, the actuary should determine whether the existing assets could be converted to sufficient cash for this purpose.

2. The actuary should consider the duration of the expected loss payments measured from the evaluation date. Preferably, the term of the hypothetical investment security should match the average payment duration. Otherwise, the underlying

security's value would be insufficient to pay the losses if either interest rates dropped and the term to maturity was less than the loss duration, or if interest rates increased and the term was greater than the loss duration.

3. The actuary should recognize the possibility of default risk in the hypothetical investment security. In other words, if the losses are in fact paid out as expected, the invested assets set aside to fund the payments must accumulate to the desired amount. The only investment securities having no default risk are those issued or backed by the U.S. Government.

4. The effect of federal income taxes should be taken into account. If the discounted reserve is the same as that used for taxation, a pre-tax interest rate would be appropriate. If the discounted reserve for financial reporting differs from that used for taxation, the actuary should determine to what extent variance from a pre-tax interest rate may be appropriate.

5. The actuary may consider introducing an explicit risk adjustment to lower the interest rate. This would be done in order to protect against the possibility of adverse loss development or of earlier than anticipated loss payment. If the undiscounted reserve already includes such a margin, then an interest rate reduction may be

unnecessary. The risk adjustment should vary according to the payment volatility of the type of reserve being discounted, and would be particularly important in valuing reserves for a sale or acquisition.

To summarize, the interest rate chosen may bear no relationship to the investment practices of the insurer whose reserves are discounted. However, it should incorporate the insurer's actual loss payment characteristics. Finally, the interest rate might differ depending on the type of financial statement or valuation used.

D. Other Considerations

This paper has concentrated solely on the discounting of loss and loss adjustment expense reserves. However, other balance sheet items would be directly affected if the net loss and loss adjustment expense liabilities were discounted for future investment income. These include the following:

Contingent commissions - In many cases, the current liability for contingent commissions is dependent on loss experience as measured by the current loss reserve. Discounting of loss reserves would then imply that discounting of the contingent commission reserve is a consideration.

Retrospective premium adjustments - Many insurance and reinsurance contracts have provisions for premium adjustments based on actual loss experience generated under the applicable policies. Discounting of liabilities for such premium adjustments may be an appropriate consideration if the underlying loss reserves are discounted.

Unauthorized reinsurance - Under statutory accounting, reinsurance of loss liabilities with unauthorized reinsurance companies may not, in some situations, be taken as an offset to a company's loss reserves. Discounting of this provision would become a consideration if loss reserve discounting were deemed appropriate for statutory statements.

Adjustments for foreign exchange rates - An adjustment to the insurance company's balance sheet is made for liabilities subject to changes in foreign exchange rates. To the extent that loss reserves are discounted, the issue of future changes in foreign exchange rates becomes a relevant consideration.

As discussed in the Statement of Principles on Loss Reserves, a number of evaluation techniques are frequently used and a range of reasonable full value reserve estimates are often developed. The range concept is equally applicable to the derivation of discounted reserves. Given the two additional components of the discounted reserve evaluation; i.e., payment timing and interest rates, the range of estimates may be significantly expanded. Determination of the most appropriate estimate can then become a more difficult process under discounting.

The Construction Of Automobile Rating Territories In Massachusetts

By Robert F. Conger

THE CONSTRUCTION OF AUTOMOBILE RATING TERRITORIES IN MASSACHUSETTS

Robert F. Conger

THE CONSTRUCTION OF

AUTOMOBILE RATING TERRITORIES IN MASSACHUSETTS

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.

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, 1965 [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, 1978 [16]) and elsewhere (SRI International, 1976, [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), 1976 [23]): Equity, Homogeneity, Discrimination, Reliability, Stability and Compatibility. Briefly, in the context of territory definitions, these criteria were defined as follows:

<u>Equity</u> — 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.

<u>Homogeneity</u> — All the towns in a territory should have approximately the same expected insured losses per car.

<u>Discrimination</u> — The probability of a town being placed in the wrong territory should be minimized.

<u>Reliability</u> — 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 of the involved parties to develop an effective way to estimate the expected insured losses per car in each town. These estimated 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.

1. A regular review, typically biennial, of all territory definitions.

2. The use of detailed insurance data, by town, as the basic information underlying the determination of which towns should be grouped together.

3. 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.

4. 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.

5. 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 evolutionary development of that methodology. The evolutionary steps described in this paper all 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, 1985 [14]) and in the resulting decision of the Commissioner (Massachusetts Division of Insurance, 1985 [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).

[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.

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 Council (AIRAC), 1982 [1]):

Greater New York, Boston and Chicago; and

Remainder of the United States

By 1917, the country was divided into eleven rating territories (DuMouchel, 1983 [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, 1982 [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 report but are described in other sources (California Department of Insurance, 1978 [2]; McDonald and Thornton (Texas), 1983 [24]; New Jersey Department of Insurance, 1981 [25]; Rhode Island Ad Hoc Committee on Territorial Rating, 1980 [26]; AIRAC, 1982 [1]).

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.

- 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.

---

[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.

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 <u>highest</u> liability rates, while Physical Damage Territory was charged the <u>lowest</u> 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 confusion and costs. 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 <u>methodologies</u> used for devising liability and physical damage rating territories also were independent (SRB, 1976 [23]). For physical damage coverages, twenty-four territories had been established on a geographic basis similar to that used in other states currently.

---

[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.

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 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's moving too many territories in any one revision or reversing direction from its movement in the previous revision.

---

[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.

[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.

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," and some recent years the Commissioner has been offered proposals for partial imposition of geographic constraints (Massachusetts Division of Insurance, 1980 [17]; AG, 1981 [10]). However, each of the reviews since 1977 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, 1976 [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;

- 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 include 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, 1976 [23]). The new approach diverged from past methods in several respects, summarized below.

First, claim frequency[6] rather than pure premium data were used. The exclusion of claim severity data was justified on the basis

---

[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.

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, 1976 [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, 1983 [22]); MARB, 1983 [13]; MARB, 1985 [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 "frequency" was

[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.

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 PLD data, as in earlier years, was tainted by the effects of numerous statutory coverage changes, and thus was excluded from the methodology. However, the SRB analysts tested the performance of the constructed frequency 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 have 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. However, the 1977 resolution of the credibility was not entirely satisfactory in that it provided no partial credibility and provided no systematic basis for the treatment of the "non-credible" 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. However, the review for 1977 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.

Finally, the numerous constraints on town movements were removed, and as a result many towns were affected sharply by the territory

---

[8] Except for a few coverages that have rates not varying by 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 Sub-committee 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, 1980 [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

$PP_{t,c}$ 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 <u>more</u> 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 <u>not</u> 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

---

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

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:

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

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)

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

It will be noted that 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 be used 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 <u>model</u> 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:

---

[10] The relationshop of traffic density to geographical variations in insurance experience has been observed in the literature (e.g., HLDI, 1985 [9]; AIRAC, 1982 [1]), as well as in some of the methodologies used in other states.

$$M_t = (a)\ D_t + (b)$$

where D_t is the town "traffic density"

$= E_t$/Road-miles in town

(a),(b) are regression coefficients

M_t is the model synthetic pure premium for the town.

A similar (but not identical) model has been used in all subsequent reviews.

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).

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 was 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.

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

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 <u>conceptual</u> 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)

---

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

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, 1980 [11]). However, the following interpretations may be placed on the credibility formula and formula for K (see, for example, Hewitt, 1975 [7]; and Hickman, 1975 [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 denomination of K is small, K is large, and the credibilities assigned to I_t are relatively small.

(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

---

[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.

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.

The State Rating Bureau recommendations for 1981 (as described in Massachusetts Division of Insurance, 1980 [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 <u>and</u> located within a common geographic regional of the state.

The Commissioner of Insurance, faced with this methodological dispute, chose the simple updating for 1981 and directed the

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

parties to undertake a cooperative review and development of methodological changes (Massachusetts Division of Insurance, 1980 [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, 1981 [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 the average classification relativity ("ACRF") in the town; and a dummy variable that allowed the abberant data in Boston to be included in the parameterization calculations without distorting the density regression coefficient.[16]

The ACRF available 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

---

[15] Claim frequency index = $\dfrac{\text{Town claim frequency}}{\text{Statewide claim frequency}}$

[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.

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

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

where A_0, A_1, A_2, A_3 are the regression parameters

Boston Dummy = 1 in Boston

0 elsewhere

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.

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 (see below).

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

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. However, with H_t, 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 (1981 [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.

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.

---

[17] This credibility methodology also has been adapted for use in calculating Massachusetts private passenger class-territory rate relativities (MARB, 1984 [15]).

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

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

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

This index calculated 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, 1981 [12]; AG, 1981 [10]; SRB, 1981 [21]), and the Commissioner adopted that recommendation (Massachusetts Division of

Insurance, 1981 [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, 1981 [10]).

The addition of the continuity 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 state-wide 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 (MARB, 1983 [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 <u>county</u>[18], rather than the <u>statewide</u> 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.

The Commissioner (Division of Insurance, 1984 [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 also were employed. The selected sources were the countywide[21] average claim severities

---

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

[20] These recommendations were developed by the MARB (MARB, 1985 [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 stand-alone severity treatment used in the revision for 1984.

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.

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

---

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

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 5 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, 1985 [20]) employed the MARB index but, mirroring the SRB's concerns, imposed tight constraints on allowed reassignments of towns.

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.

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 the 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, 1984 [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

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

particular geographical area, small changes in data or methodology may 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 yet 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.

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).

However, judgment 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 unac-

---

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

ceptably 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 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 1985 territories.

Homogeneity Measures

Appendix B details several quantitative measures that have been designed to compare the relative homogeneity of alternative Massachusetts automobile 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. Page 2 of 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, Page 2 whether this difference is due to shifting claims experience or due to improving methodologies, but the inclusion on Exhibit 6, Page 1 of the updated calculations based on the 1984 methodology makes it apparent that much of the difference is due to shifting claims experience.

---

[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 (see above).

Index Methodology. Page 1 of Exhibit 6 compares the homogeneity of territories produced by the 1984 town index methodology and by the 1986 town index methodology. Each is 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 measure probably have little useful value in comparing the homogeneity of the final territories produced by the two methods.

Territory Groupings. Exhibit 6, Page 2 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 1985 territories. However, the additional constraints imposed by the Commissioner nearly recreate, in 1986, the 1984-85 territory definitions and thus bear a non-trivial 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, and 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.

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.

ACKNOWLEDGMENTS

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.

REFERENCES

[1] All-Industry Research Advisory Council (AIRAC), 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] DuMouchel, William H., "The 1982 Massachusetts Automobile Insurance Classification Scheme," The Statistician 32 (1983).

[4] DuMouchel, William H., and Harris, Jeffrey E., "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] Garrison, C.B. and Paulson, A.S., "An Entropy Measure of the Geographic Concentration of Economic Activity," Economic Geography, Vol. 49 (1973), pp.319-324.

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

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

[8] Hickman, J.C., "Introduction and Historical Overview of Credibility in Credibility Theory and Applications, ed. 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 (AG), Preliminary Filing for the 1981 Hearing on Automobile Insurance Territories (proposed territories to be used in 1982), August 7, 1981.

[11] Massachusetts Automobile Rating and Accident Prevention Bureau (MARB), Filing for 1981 Massachusetts Automobile Rating Territories, July 7, 1980.

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

[13] Massachusetts Automobile Rating and Accident Prevention Bureau (MARB), Filing for 1984 Private Passenger Territory and Classification Definitions, July 25, 1983.

[14] Massachusetts Automobile Rating and Accident Prevention Bureau, Filing for 1986 Private Passenger Territory and Classification Definitions, July 26, 1985.

[15] Massachusetts Automobile Rating and Accident Prevention Bureau Recommendations with Respect to Rate Relativites for Private Passenger Vehicles (for 1985), November 2, 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 23, 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 (SRB), 1982 Massachusetts Automobile Rating Territories -- Recommendation (proposed territores to be used in 1982), August 7, 1981.

[22] Massachusetts State Rating Bureau, Recommendations With Respect to Rating Territories and Classification for 1984 Automobile Insurance, July 25, 1983.

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

[24] McDonald, James L., and Thornton, John H., "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 30, 1980.

[27] Siczewicz, Peter, "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] Stern, Philipp K., "Ratemaking Procedures for Automobile Liability Insurance," Proceedings of the Casualty Actuarial Society LII, 1965.

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

Exhibit 2

EXAMPLE OF TRAFFIC DENSITY CALCULATIONS

Town Name	(1) 1983 PDL Exposures	(2) Road Miles	(3) Traffic Density (1)÷(2)
Hampden	2968.9	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	10022.2	121	82.8
Easthampton	8556.3	83	103.1
Northampton	13633.4	178	76.6
South Hadley	8218.0	99	83.0
Ware	4745.5	121	39.2
Belchertown	4679.7	147	31.8
Hadley	3034.7	80	37.9
Hatfield	2023.0	59	34.3
Huntington	1098.6	54	20.3
Williamsburg	1492.7	48	31.1
Chesterfield	555.1	56	9.9
Cummington	439.5	62	7.1
Goshen	403.7	43	9.4
Granby	3104.1	70	44.3
Middlefield	212.1	37	5.7
Pelham	658.7	43	15.3
Plainfield	294.4	49	6.0
Southampton	2527.3	71	35.6
Westhampton	678.1	49	13.8
Worthington	622.1	64	9.7
Cambridge	30201.7	142	212.7
Lowell	37722.5	239	157.8
Everett	15385.0	63	244.2
Malden	23400.1	107	218.7
Medford	26637.5	151	176.4
Newton	44546.0	311	143.2
Somerville	26169.1	105	249.2
Waltham	27703.2	152	182.3
Watertown	17042.1	77	221.3
Arlington	25049.7	121	207.0
Belmont	14010.8	82	170.9
Chelmsford	19038.3	186	102.4
Concord	9960.8	109	91.4
Dracut	12546.6	112	112.0
Framingham	35207.4	233	151.1
Hudson	9345.9	79	118.3
Lexington	18274.3	153	119.4
Marlborough	17001.5	130	130.8
Melrose	15376.9	82	187.5
Maynard	5420.1	41	132.2

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.

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)

	Index						PDL	Observed Pure Premium (latest three years)	
Town Name	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	3035	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	2023	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	81045	$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	9961	73.91	198.98
Westminster	0.8263	0.7813	0.6824	0.8120	0.4102	0.7041	3378	79.75	186.29
TOTAL (33 towns)						0.6900	94683	$77.00	$188.88
TERRITORY 3									
Rockport	0.5922	0.7192	0.8276	0.7717	0.5292	0.7068	3920	$69.03	$187.59
Southampton	0.8541	0.9494	0.6603	0.7553	0.4671	0.7071	2527	69.55	179.61
Williamsburg	0.7943	0.7382	0.7546	0.7765	0.4272	0.7091	1493	83.44	187.00
Amherst	0.6887	0.6539	0.7270	0.8517	0.6172	0.7298	10022	78.67	204.44
Northampton	0.7930	0.9033	0.8373	0.7261	0.4467	0.7299	13633	85.53	180.99
Easthampton	0.8141	0.8874	0.8341	0.7618	0.3950	0.7325	8556	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	8218	85.80	183.24
Chelmsford	0.7609	0.6756	0.8297	0.7881	0.5589	0.7438	19038	86.94	218.69
Brimfield	0.9462	0.9789	0.7243	0.7987	0.3833	0.7467	1336	96.42	195.70
TOTAL (43 towns)						0.7313	192994	$81.48	$200.11
TERRITORY 4									
Dalton	0.8673	0.7354	0.8972	0.7410	0.4462	0.7518	3427	$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	5420	89.74	203.76
Lexington	0.7151	0.6436	0.8919	0.8103	0.5826	0.7612	18274	88.87	222.39
Ware	0.9327	0.9597	0.7733	0.8745	0.3557	0.7694	4746	85.02	194.39
Hampden	0.8769	1.0627	0.7932	0.8385	0.4141	0.7708	2969	100.69	219.45
Granby	0.8929	1.0815	0.7670	0.8488	0.4533	0.7763	3104	93.16	205.79
Belchertown	0.9374	1.0049	0.7505	0.8813	0.4590	0.7879	4680	95.83	212.63
TOTAL (52 towns)						0.7688	244421	$88.71	$219.64

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(For Selected Towns)

Exhibit 5
Sheet 2

Town Name	Index BI	PIP	PDL	Coll.	Comp.	Combined	PDL Exposures	Observed Pure Premium (latest three years) Liability	Package
TERRITORY 5									
Freetown	0.7255	1.0134	0.7048	0.8121	0.8952	0.7923	4429	$77.97	$214.40
Huntington	0.8181	1.0210	0.7228	0.9715	0.5001	0.7923	1099	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	1858	114.34	230.16
TOTAL (58 towns)						0.8133	377184	$93.36	$225.84
TERRITORY 6									
Pittsfield	0.9237	1.0037	1.0512	0.7716	0.4955	0.8405	23215	$103.34	$213.53
Belmont	0.8108	0.7705	0.9853	0.9192	0.7551	0.8772	14011	94.29	242.68
Groveland	1.0068	0.9376	0.9187	0.9202	0.6729	0.8897	2997	121.45	275.98
TOTAL (29 towns)						0.8623	246318	$99.29	$237.42
TERRITORY 7									
Lynnfield	0.7559	0.7815	0.8793	0.9499	0.9536	0.8903	7022	$89.51	$278.18
Hudson	1.0119	0.9559	0.9603	0.9667	0.5376	0.8935	9347	113.63	262.89
Newton	0.8075	0.7563	1.0137	0.9686	0.7953	0.9076	44546	97.82	265.62
Arlington	0.8520	0.8296	1.0094	0.9641	0.8836	0.9330	25050	97.88	257.57
Framingham	0.8947	0.9556	1.0863	0.9804	0.6821	0.9359	35207	110.40	270.40
Taunton	0.9870	1.1305	0.8896	0.8941	0.9970	0.9433	21884	108.03	254.07
TOTAL (29 towns)						0.9155	316422	$103.49	$259.77
TERRITORY 8									
Norwood	1.0023	0.8238	0.9998	0.9462	0.8767	0.9473	16092	$107.88	$267.20
Marlborough	1.0850	1.0861	1.0098	0.9880	0.6390	0.9526	17002	117.22	271.45
Wilmington	1.0063	0.9694	1.0451	0.9645	0.9674	0.9938	10232	124.82	299.34
Tewksbury	1.0542	1.0173	1.0043	1.0518	0.8375	0.9973	13205	118.43	301.59
TOTAL (11 towns)						0.9701	133231	$110.64	$280.22
TERRITORY 9									
Marshfield	0.8233	1.0737	0.9463	1.0817	1.0827	1.0006	11537	$104.63	$298.99
Dracut	1.0308	1.2132	1.0090	0.9948	0.9184	1.0045	12547	128.22	295.68
Melrose	1.0408	0.9126	1.0519	0.9753	1.1626	1.0369	15377	110.18	290.56
Waltham	1.0843	1.0354	1.1057	1.0299	0.9376	1.0412	27703	118.94	288.15
Holyoke	1.3525	1.1436	1.1678	0.9313	0.8089	1.0597	17069	132.77	261.19
TOTAL (21 towns)						1.0345	351465	$116.81	$287.84
TERRITORY 10									
Haverhill	1.2649	1.1987	1.0834	0.9720	0.9866	1.0679	21905	$132.68	$294.11
Watertown	1.0245	0.8988	1.1275	1.1080	1.0918	1.0819	17042	111.42	293.96
Worcester	1.2873	1.0872	1.2669	1.0435	0.8536	1.1113	63452	134.67	292.29
TOTAL (6 towns)						1.0914	149630	$126.23	$297.59

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(For Selected Towns)

Exhibit 5
Sheet 3

Town Name	BI	PIP	PDL	Coll.	Comp.	Combined	PDL Exposures	Liability	Package
	Index							Observed Pure Premium (latest three years)	
TERRITORY 11									
Holbrook	1.1936	1.0925	1.1007	1.1166	1.1444	1.1276	6136	$130.64	$325.24
Lowell	1.2058	1.2939	1.1749	1.1143	1.0629	1.1488	37723	138.40	324.54
Quincy	1.1494	1.0244	1.1279	1.1782	1.4034	1.1907	39832	121.99	330.82
TOTAL (7 towns)						1.1604	138765	$124.57	$330.05
TERRITORY 12									
Springfield	1.6355	1.6528	1.2426	1.0985	0.9862	1.2463	62300	$153.62	$309.41
Medford	1.0835	1.0678	1.1427	1.1778	1.7861	1.2576	26638	125.71	365.73
Brockton	1.3352	1.2859	1.2262	1.2281	1.4207	1.2832	41920	143.62	360.58
TOTAL (3 towns)						1.2604	130857	$144.74	$337.27
TERRITORY 13									
Cambridge	1.1122	1.1053	1.1675	1.3506	1.7395	1.3130	30202	$124.92	$362.73
Malden	1.2725	1.2228	1.2689	1.2197	1.7768	1.3440	23400	140.38	378.93
Lynn	1.2561	1.1420	1.3693	1.1942	1.8444	1.3690	33217	144.02	369.84
TOTAL (4 towns)						1.3415	91504	$136.84	$370.15
TERRITORY 14									
Lawrence	1.4980	1.6148	1.3072	1.1900	1.9389	1.4421	23935	$162.76	$394.42
Everett	1.3023	1.2718	1.3755	1.2769	2.2025	1.4757	15385	147.74	424.12
TOTAL (3 towns)						1.4549	48036	$154.29	$407.97
TERRITORY 15									
Somerville	1.3551	1.2398	1.3910	1.4373	2.3460	1.5588	26169	$151.50	$436.19
TOTAL (1 town)						1.5588	26169	$151.50	$436.19
TERRITORY 16									
Revere	1.5148	1.4559	1.4578	1.5443	3.1677	1.7956	17396	$162.89	$547.65
Chelsea	1.6758	1.6544	1.5957	1.6970	3.3012	1.9318	7307	183.39	573.08
TOTAL (2 towns)						1.8359	24704	$168.95	$555.17
TERRITORY 17 - West Roxbury (Boston)									
	1.2349	1.0351	1.1328	1.1949	1.4951	1.2311	12867	$118.80	$335.09
TERRITORY 18 - Roslindale (Boston)									
	1.4514	1.2973	1.4038	1.5675	2.4886	1.6536	10769	$150.58	$457.66
TERRITORY 19 - Jamaica Plain (Boston)									
	1.5896	1.6845	1.4404	1.7914	3.1083	1.8918	10400	$155.81	$512.46
TERRITORY 20 - Hyde Park (Boston)									
	1.3877	1.6163	1.3734	1.5910	2.4247	1.6534	11481	$151.63	$463.63
TERRITORY 21 - Dorchester (Boston)									
	2.1882	2.3175	1.6118	2.3901	4.2874	2.4664	33479	$213.34	$727.76
TERRITORY 22 - Roxbury (Boston)									
	2.4487	2.8154	1.8756	2.6857	4.8291	2.7791	6538	$245.62	$818.45
TERRITORY 23 - Boston Central (Boston)									
	1.6152	1.6337	1.4941	1.9746	3.6202	2.0513	16904	$162.83	$613.69

MASSACHUSETTS
INDICATED 1986 RATING TERRITORIES
(For Selected Towns)

Exhibit 5
Sheet 4

Town Name	Index BI	Index PIP	Index PDL	Index Coll.	Index Comp.	Index Combined	PDL Exposures	Observed Pure Premium (latest three years) Liability	Observed Pure Premium (latest three years) Package
TERRITORY 24 - Brighton (Boston)									
	1.2748	1.2898	1.3588	1.5606	2.0844	1.5361	15873	$137.77	$422.43
TERRITORY 25 - South Boston (Boston)									
	1.6416	1.6759	1.4753	1.9179	3.2148	1.9269	6491	$168.47	$572.51
TERRITORY 26 - East Boston (Boston)									
	1.7229	1.6995	1.8241	1.9951	3.9162	2.2038	10780	$197.54	$669.83
TOTALS ALL TERRITORIES						1.0000	2783010	$109.48	$279.68

Homogeneity Measures for Territory Groupings - Index Based on 1984 Index Method Updated For New Data vs. 1986 Index Method

Homogeneity Measure*	Territory Division Index Intervals; no capping								1984-85 Territory Grouping
	5%		5.5%		6%		6.5%		
	1984 Method	1986 Method	1984 Method	1986 Method	1984 Method	1986 Method	1984 Method	1986 Method	
1. P.P. Squared Diff. (Absolute)									
a) Liability[2]	78.18	61.71	75.98	61.06	74.57	60.54	68.22	59.56	88.03
b) Package[3]	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.

#3548

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[2]	61.71	61.06	60.54	59.56	59.15	60.05	88.03	127.07
b) Package[3]	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.

Exhibit.
Page 2

THE CONSTRUCTION OF AUTOMOBILE RATING TERRITORIES IN MASSACHUSETTS

APPENDIX A. CALCULATION OF THE TOWN INDEX AND TOWN RANKING[1]

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 A1 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

[1]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.

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, below.

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, below.

The resulting claim frequency and claim cost indications by town are combined to produce a pure premium index by town and coverage. These pure premium indexes 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 A2 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 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 indexes are rebalanced 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.

The traffic density in a town is calculated as the ratio of insured exposures[2] 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

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

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

d. <u>Model and Credibility Parameters</u>

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,

---

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

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 A2.

The regression model parameters estimated in the latest review are:

	A-1	A-2	PDL	Comprehensive	Collision
Intercept ($A_{o,c}$)	-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 (x.002647)	24.7	97.5	0.0
(2)	ACRF (x.8816)	.9682	1.0528	1.0014
(3)	Boston Dummy (x1.3393)	0	0	1
(4)	Intercept (-.2486)	-0.2486	-0.2486	-.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^2_c = A measure of the year to year variation in claim frequencies (see Exhibit A2)

σ^2_c = A measure of the extent to which actual claim frequencies differ from model claim frequencies (see Exhibit A2)

The credibility parameters (τ^2_c, σ^2_c), determined in the latest review in accordance with the formulas outlined in Exhibit A2, 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	Wilm-ington	Brighton
(1)	Model Claim Frequency ($MFI_{c,t}$)	.6791	.9499	1.9994
(2)	Exposures ($E_{c,t}$)	1629.5	21480.6	35513.6
(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

$AFI,_{c,t}$ = 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)x(1) + (1.0-(3))x(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[3], and the statewide claim severity relativities indicated by historical data. The credibility parameters are determined by a Two Layer Hierarchical Empirical Bayes Method.

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

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

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 A3 and A4. The calculated parameters are shown in Exhibit A5. Illustrative examples of the credibility-weighting process are included in Exhibit A6.

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[4]

$$\frac{\text{average claim cost by coverage by year by town}}{\text{average claim cost by coverage by year, statewide}} \div \frac{\text{average age/symbol relativity by cov. by town}}{\text{average age/symbol relativity by cov., statewide}}$$

3. County or county group assignments of the towns.

As shown in Exhibit A6, for the three town Collision example the methodology yields

	Holland	Wilmington	Brighton
Estimated Relative Severity	1.0846	1.0508	.9011

[4]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.

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.

However, any town-to-town variation in pure premiums that is captured by other rating variables 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[5] as measured by the ACRF described above.

The resulting town net pure premium indices are rebalanced to unity within each coverage.

In the three town example for collision:

	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) x (2)	.8064	1.0156	1.5630
(4) Average Class Rating Factor	.9682	1.0528	1.0014
(5) Net pure premium index, rebalanced to unity = ((3) ÷ (4)) ÷ 1.00015	.8328	.9645	1.5606

---

[5] 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.

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:

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

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:

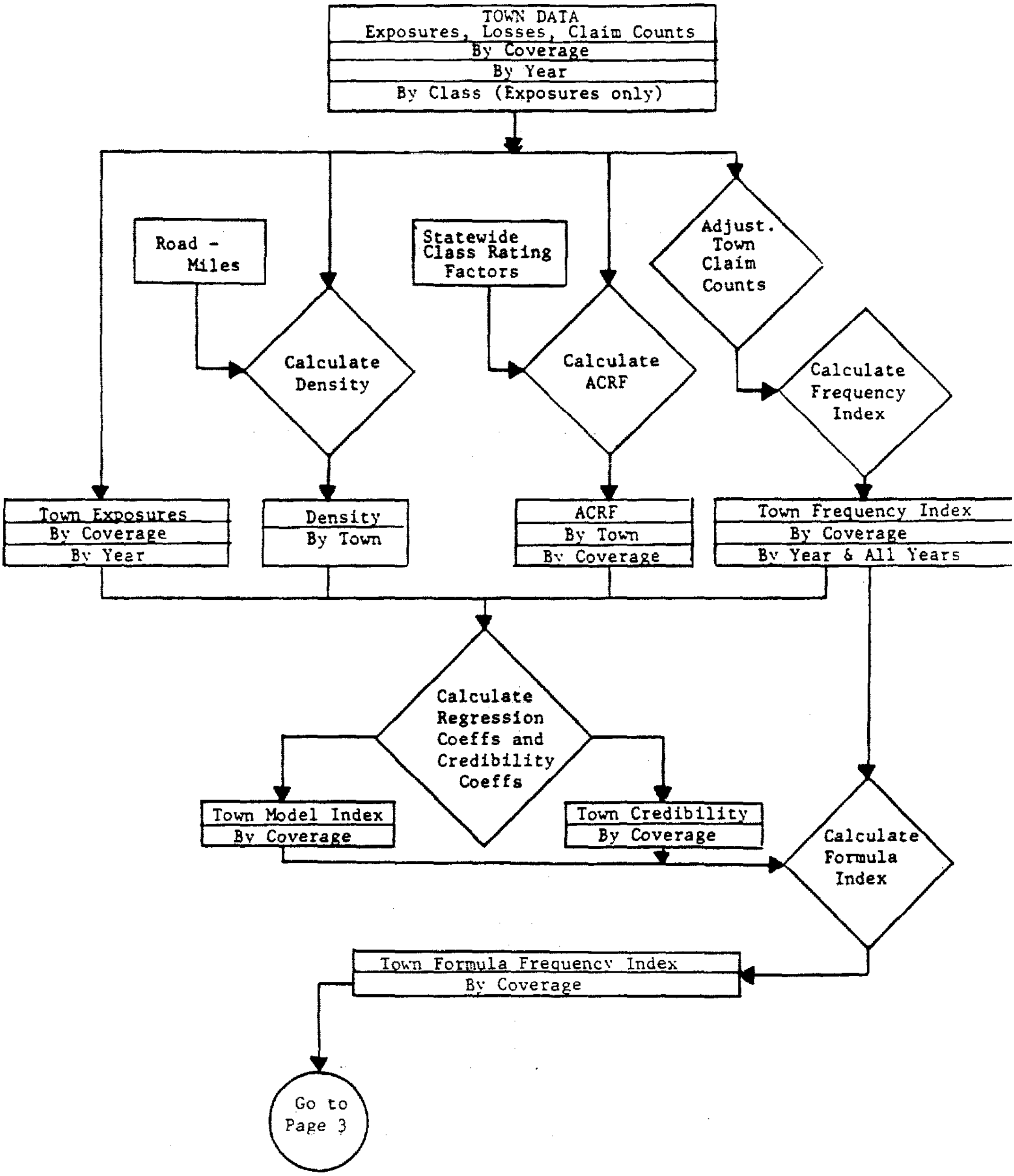
	Holland	Wilmington	Brighton
(1) Exposure (latest year)			
A-1, A-1, PDL	914.0	10232.4	15872.8
Comprehensive	533.7	7176.6	11806.0
Collision	422.2	5794.7	9679.2
(2) Net Pure Premium Index			
A-1	.5397	1.0063	1.2748
A-2	.8712	.9694	1.2898
PDL	.6477	1.0451	1.3588
Comprehensive	.4253	.9674	2.0844
Collision	.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:

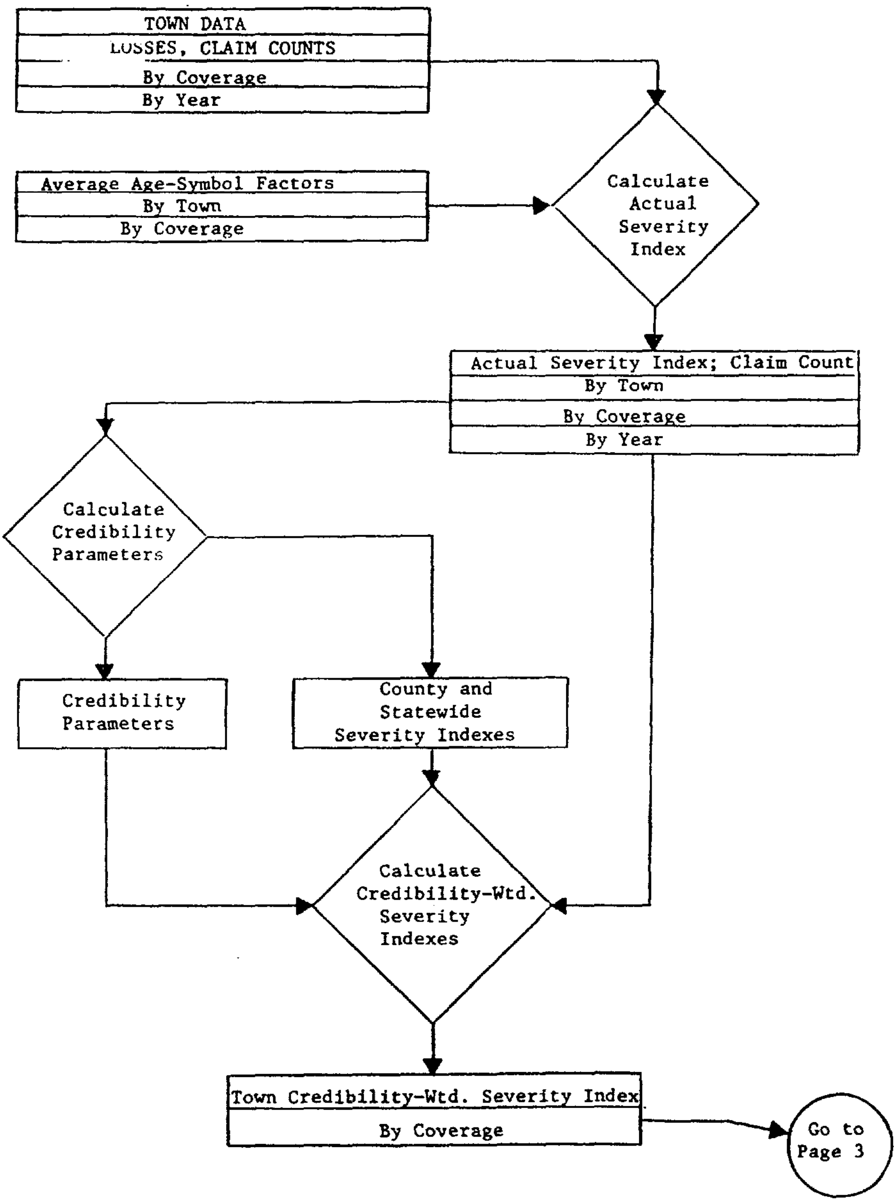
	Rank
Holland	30
Brighton	349
Wilmington	302

Exhibit A1
Page 1

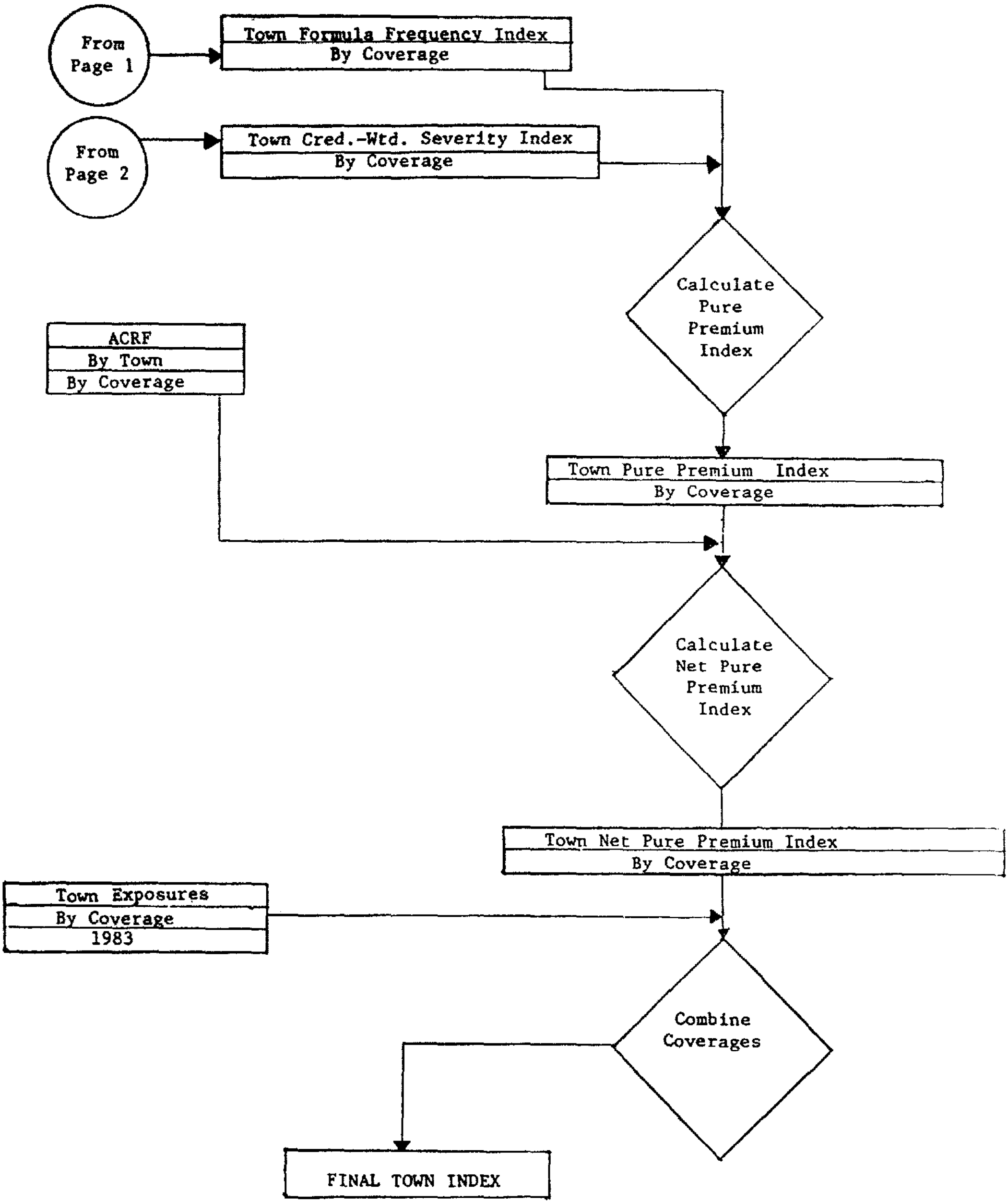
OUTLINE OF TOWN INDEX CALCULATIONS



OUTLINE OF TOWN INDEX CALCULATIONS



OUTLINE OF TOWN INDEX CALCULATIONS



CALCULATION OF CLAIM FREQUENCY INDEX MODEL PARAMETERS AND CREDIBILITY PARAMETERS FORMULAS*

The basic structure of the claim frequency model is:

$$\widehat{\text{Frequency Index}}_{c,t} = A_{o,c} + A_{1,c} \times \text{Density}_t + A_{2,c} \times \text{ACRF}_{c,t} + A_{3,c} \times \text{Boston Dummy}_t$$

where: the subscripts c and t refer to coverage and territory

Density_t is the town density for non-Boston Towns

$\text{ACRF}_{c,t}$ is the Average Class Rating Factor for the coverage and town

Boston Dummy_t = 1 in towns which are part of Boston, 0 elsewhere

$A_{o,c}$; $A_{1,c}$; $A_{2,c}$; $A_{3,c}$; 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 x 1 vector of actual town claim frequency indices

$\hat{y}$ is a 360 x 1 vector of model claim frequency indices

x is a 360 x 4 matrix of the independent variables in the claim frequency model.

Column 1 is unity

Column 2 is Density_t

Column 3 is ACRF_t

Column 4 is Boston Dummy_t

* For a more detailed exposition of these formulas refer to 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. Also see DuMouchel and 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.

x^T is the 4 x 360 transpose of x

A is a 1 x 4 matrix of the regression coefficients

W is a 360 x 360 diagonal matrix of the weights to be applied to each town in the weighted least squares regression.

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

$$\tau^2/H_t$$

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

$H_{t,y}$ = $\text{Exposures}_{t,y}/Y_t$

Y_t = Claim frequency index for town t, all years combined

$Y_{t,y}$ = Claim frequency index for town t, year y

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

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

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:

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

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

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:

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

The credibility assigned to the actual town frequency index is:

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

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

Calculation of the Claim Cost Index Using The Two Layer
Hierarchical Empirical Bayesian Credibility Model
Summary of Formulas[1]

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.[2]) The intentis 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:

$$P_{C.}(.) = \sum_{g,t} P_{CG}(t)$$

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

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

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

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

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

---

[1] These formulas and their derivations 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.

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

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

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

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

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

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

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

where 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.

Let $I(P) = 0$ if $P = 0$
$\qquad\qquad 1$ 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 <u>estimated</u> 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:

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

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

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

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 A3 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.

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

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

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

and the weights[1], w_g, are

$$w_g = \frac{[\sum_t I(P_{Cg}(t))] - I(P_{Cg}(.))}{\sum_{g,t} I(P_{Cg}(t)) - \sum I(P_{Cg}(.))}$$

[1] 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 some years would receive less weight.

and 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.

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

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

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

The resulting values of $\tilde{s}_C^2$ which were used in the review for 1986 are displayed in Exhibit A5, Page 2.

<u>Limitations on k values</u>

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 k_C for each county) are limited by the imposition of a maximum value and a minimum value. When the calculated value was less than the minimum, the value of the parameter was set equal to that minimum.[2] 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 A5, Page 1.

[2] 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 A5
Page 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

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

CREDIBILITY PARAMETER S-SQUARED, 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 A5
Page 2

PRICING EXAMPLE, SEVERITY
BRIGHTON TOWN NUMBER 822
SUFFOLK TOWN'S PDL EXPOSURES 15872.8

	B.I.	PIP	PDL	COMP.	COLL.
(1) CLAIMS FOR TOWN	657.0000	1197.0000	5569.0000	6770.0000	6388.0000
(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	1026.1405
(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)X(4) + (5)X1-(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)	1004.7278	331.8594	1156.6019	696.4471	2500.0000
(9) CRED. FOR TOWN = (1)/((1)+(8))	.3954	.7829	.8280	.9067	.7187
(10) EST. REL. SEV. FOR COUNTY GP. = (6)	1.0849	1.1138	1.0684	1.3657	.9443
(11) EST. REL. SEV., TOWN = (7)X(9) + (10)X1-(9)	1.0946	1.0919	1.0265	1.1962	.9011

PRICING EXAMPLE, SEVERITY
HOLLAND TOWN NUMBER 494
HAMPDEN TOWN'S PDL EXPOSURES 914.0

	B.I.	PIP	PDL	COMP.	COLL.
(1) CLAIMS FOR TOWN	14.0000	56.0000	164.0000	105.0000	12[illegible].[illegible]000
(2) CRED. WEIGHTED MEAN FOR COUNTY GROUP	.9334	.9513	.9325	.7344	[illegible].[illegible]438
(3) OVERALL K (CLAIMS)	548.0728	740.3648	515.3687	142.6349	102[illegible].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)X(4) + (5)X1-(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)	1812.2624	355.5324	2500.0000	184.7182	342.3312
(9) CRED. FOR TOWN = (1)/((1)+(8))	.0077	.1361	.0616	.3624	.2737
(10) EST. REL. SEV. FOR COUNTY GP. = (6)	.9387	.9575	.9340	.7402	1.0483
(11) EST. REL. SEV., TOWN = (7)X(9) + (10)X1-(9)	.9388	.9513	.9394	.7195	1.0846

PRICING EXAMPLE, SEVERITY
WILMINGTON TOWN NUMBER 652
MIDDLESEX TOWN'S PDL EXPOSURES 10232.4

	B.I.	PIP	PDL	COMP.	COLL.
(1) CLAIMS FOR TOWN	419.0000	734.0000	3010.0000	2386.0000	2170.0000
(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	1026.1405
(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)X(4) + (5)X1-(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)	2500.0000	543.7111	2500.0000	124.9303	198.6265
(9) CRED. FOR TOWN = (1)/((1)+(8))	.1435	.5745	.5463	.9502	.9161
(10) EST. REL. SEV. FOR COUNTY GP. = (6)	1.0060	.9866	1.0305	.8841	1.0413
(11) EST. REL. SEV., TOWN = (7)X(9) + (10)X1-(9)	1.0111	.9781	1.0499	.9923	1.0508

THE CONSTRUCTION OF AUTOMOBILE RATING TERRITORIES IN MASSACHUSETTS

APPENDIX B: HOMOGENEITY AND HOMOGENEITY MEASURES[1]

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

---

[1]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.

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.[2] 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.

Homogeneity Measures

This section defines several measures of the homogeneity of a territory group 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

[2] 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.

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 B1. 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 to compare concentrations in economic activity over time.[3]

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 <u>entropy</u> or expected information of the system is defined by:

$$H = \sum_{i=1}^{k} p_i \text{ Log } p_i$$

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[4]

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

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

---

[3]Garrison, C.B. and Paulson, A.S., "An Entropy Measure of the Geographic Concentration of Economic Activity," <u>Economic Geography</u>, Vol. 49 (1973), 319-324.

[4]As usual, if $n_i = 0$ then (n_i/n) Log $(n_i/n) = 0$.

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, defining:

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

when e_t = Town Index_t - Territory $\text{Index}_{(t)}$ 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.

Homogeneity Measure Definitions

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

Homogeneity Measure Definitions

MEASURE	DEFINITIONS
5. Pure Premium Cred. Wgtd. Percentage Max. Diff HM 5	$\sum_{\text{Terr }(i)} 83\ \text{EXP}_{(i)}\ \underset{i}{\text{Max}}\ \text{Max Cred}_i \left\lvert \frac{\text{Town PP}_i - \text{Terr PP}_{(i)}}{\text{Town PP}_i} \right\rvert \div \sum_{\text{Town } i} 83\ \text{EXP}_i$
5a. Pure Premium Percentage Max. Diff HM 5A	$\sum_{\text{Terr }(i)} 83\ \text{EXP}_{(i)}\ \underset{i}{\text{Max}} \left\lvert \frac{\text{Town PP}_i - \text{Terr PP}_{(i)}}{\text{Town PP}_i} \right\rvert \div \sum_{\text{Town } i} 83\ \text{EXP}_i$
6. Index Max. Diff. HM 6	$\sum_{\text{Terr }(i)} 83\ \text{EXP}_{(i)}\ \underset{i}{\text{Max}} \left\lvert \text{Town Ind}_i - \text{Terr Ind}_{(i)} \right\rvert \div \sum_{\text{Town } i} 83\ \text{EXP}_i$
7. Error Entropy HM 7	$-\sum_{e_i} (\text{Exp}_{(e_i)}/\text{EXP})\ \text{LOG}\ (\text{EXP}_{(e_i)}/\text{EXP})$

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.