TITLE: RATING CLAIMS-MADE INSURANCE POLICIES

AUTHOR: Mr. Joseph Marker

Mr. Marker is Actuary with the Westfield Companies. He received his FCAS in 1977 and holds a B. A. from the University of Michigan as well as an M. A. from the University of Minnesota.

Mr. F. James Mohl

Mr. Mohl is Senior Actuarial Officer for the St. Paul Fire and Marine. He received his FCAS in 1975 and serves on the CAS Education Committee.

REVIEWER: Mr. Michael F. McManus

Mr. McManus is Associate Actuary with Chubb Insurance Company. He received his B. S. from St. Peter's College in New Jersey and earned his FCAS in 1977. Mike is a member of the American Academy of Actuaries and is a member of the CAS Examination Committee.

I. Introduction

In this paper, we propose to discuss the claims-made approach to pricing Medical/Professional Liability insurance. We will begin with a brief summary of the historic context which lead the largest medical malpractice writer in the country (St. Paul Fire and Marine) to switch its book of business to claims-made. Then we will describe in depth the claims-made concept itself: how it works, how it differs from traditional occurrence coverage, what its inherent advantages are, and what special problems it presents and how these might be resolved. In particular, we will compare the accuracy of claims-made and occurrence ratemaking under varying assumptions about a changing claim environment. We will outline special features of The St. Paul filings which distinguish them from previous claims-made filings by other carriers. Finally, we will highlight special analytic tools which were developed to price claims-made coverages, and will show how these same tools can aid the actuary in pricing and reserving occurrence coverage as well. Let us look back at the time before claims-made to see how the decision to offer this coverage evolved.

II. Claims-Made: A Historic Perspective

The 1950's were an era of steady growth and moderate inflation. Insurance companies generally did well. And "malpractice insurance", as it was then called, was particularly favorable, although it was such a small part of most companies' total book of business that they did not bother to distinguish it from other general (non-automobile) liability lines. Rates were very low and stable--or even falling--throughout the period. Suing a doctor was almost unheard of; suing and winning was even more unusual. Medical Liability insurance was regarded "peace of mind" coverage, if it was thought of at all.

In the 1960's the situation began to change, as inflation gradually accelerated throughout the decade. Moreover, "social inflation"--a term coined to describe the inflation in value of a tort in the minds of plaintiffs, attorneys, judges and juries--consistently ran at a higher rate than economic inflation, adversely affecting claim severity in all liability lines. Compounding the increase in severity was an increase in frequency, brought about in part by a "psychology of entitlement"--a feeling that an injured party should be compensated even if negligence had not been proven. This took the form of the erosion of traditional tort defenses, especially in the malpractice area. Still, at the end of the decade, Medical Liability did not appear to have deteriorated as much as some other liability lines. Appearances were deceiving, however, since Medical Liability insurers simply failed to recognize the impact of claims that had been Incurred But Not Reported.

In the early 1970's the insurance industry was hit with a triple whammy: severe recession resulting in the steepest plunge in the stock market since the 1930's, soaring economic inflation, and price controls which held back rate increases while doing nothing about social inflation. The combination of inflation and price controls lead to inadequate rates on current business. Worse yet, reserves on prior years--particularly reserves for Incurred But Not Reported claims--had to be increased at the same time. Only much later did it become apparent how unprofitable results from the late 1960's and early 1970's really were. This placed further pressure on the companies' surplus, already dwindling due to the stock market collapse. There were some who charged that insurers were trying to make up for stock market losses by raising rates for their policyholders. All lines were affected by these conditions to some degree but the "long tail" Medical/Professional Liability lines were particularly susceptible due to their high ratio of reserves to premium.

Some Medical Liability insurers responded to the "malpractice crisis" by seeking astronomical rate increases. However, company actuaries had difficulty in estimating IBNR and justifying it to regulators. Where rate increases were granted, cries of "unaffordability" could be heard. Where they were not granted, carriers pulled out of the market or, in at least one case, went bankrupt. Availability at any price became a real concern.

For The St. Paul the situation was critical, even assuming we could obtain any rate increase sought--a highly questionable assumption. Costs were spiralling at such a rate that what had once been a minor line was now large enough to place the entire company in jeopardy if we guessed wrong about the IBNR. Something had to be done to cut the exposure presented by the tail: either get out of the business or find a way to expose ourselves to that risk a year at a time and price it a year at a time. Out of that idea grew the decision to try a claims-made approach.

III. Claims-Made Coverage Concepts

The basic idea of claims-made coverage is simple: a claims-made policy covers claims reported ("made") during the policy period, regardless of when the underlying accident occurred. This contrasts with an occurrence policy which covers claims occurring during the policy period. "Claims-made" is not a new concept. Insurers have traditionally written some professional liability lines and many bonds on a claims-made basis.

The St. Paul has modified the claims-made policy concept to meet the specific needs of professional liability insureds. First, an insured who is in his first year of professional practice does not need coverage for acts which occurred prior to his beginning practice. The same is true of an insured who begins claims-made coverage after letting an "occurrence basis" policy expire. These insureds need coverage restricted to accidents occurring on or after the date that they first began insuring on a claims-made policy basis. This need is met by placing a "retroactive date" on a claims-made policy and restricting policy coverage to accidents occurring on or after that date. Second, the insured needs coverage for claims reported after he retires from his occupation -- "tail coverage" policies provide the necessary coverage. This coverage is also needed in case of death, disability, or simply changing insurance carriers.

At this point, we will define some of the coverage terms which will appear throughout the remainder of the paper. A convenient way to explain the coverages is to define the occurrences covered in terms of accident period covered and reported period covered. This turns out to be convenient in pricing because insurance loss data contains both accident date and reported date. If we can define the occurrences covered in terms of these dates, then we can price the policies using insurance loss data, even though that data may be collected under "occurrence basis" rather than claims-made policies.

The matrix of losses by accident year and reported year that we use is indexed differently than the traditional system of using reported year for one dimension and accident year for the other. For convenience we replaced accident year by "accident year lag" which is computed:

Accident Year Lag = Reported Year - Accident Year

We visualize losses in the Figure 1 matrix:

LAG (I)	REPORT YEAR (J) 1	2	3	4	5	6	7
0	$L_{0,1}$	$L_{0,2}$	$L_{0,3}$	$L_{0,4}$	$L_{0,5}$	$L_{0,6}$	$L_{0,7}$
1	$L_{1,1}$	$L_{1,2}$	$L_{1,3}$	$L_{1,4}$	$L_{1,5}$	$L_{1,6}$	$L_{1,7}$
2	$L_{2,1}$	$L_{2,2}$	$L_{2,3}$	$L_{2,4}$	$L_{2,5}$	$L_{2,6}$	$L_{2,7}$
3	$L_{3,1}$	$L_{3,2}$	$L_{3,3}$	$L_{3,4}$	$L_{3,5}$	$L_{3,6}$	$L_{3,7}$
4	$L_{4,1}$	$L_{4,2}$	$L_{4,3}$	$L_{4,4}$	$L_{4,5}$	$L_{4,6}$	$L_{4,7}$

FIGURE 1

NOTE: For convenience, we do not show lags greater than four years in the matrix. In practice, we group all losses with longer lags in the LAG 4 row.

With this configuration $L_{i,j}$ is the loss reported in year j with accident year lag i, (the accident year is j-i).

Notice that a complete accident year consists of a Northwest-Southeast diagonal in the matrix. For example:

Accident year 1

= (Acc. year 1, report year 1) + (Acc. year 1, report year 2) + ...

= (Report year 1, lag 0) + (Report year 2, lag 1)

+ (Report year 3, lag 2) + ...

= $L_{0,1} + L_{1,2} + L_{2,3} + L_{3,4} + \cdots$

We are now in a position to describe some of the coverage concepts in terms of the matrix above (in the examples which follow, all policies are assumed written at the beginning of year j *for a one-year term*):

"Mature" claims-made policy. A policy which covers claims reported during the policy period, regardless of accident date. Such a policy written at the beginning of year j will cover the jth column of matrix L in Figure 1.

First-year claims-made policy. A policy which covers only the "lag 0" row of the jth column of L. An insured in his first-year of a claims-made insurance program would purchase this coverage.

Second-year claims-made policy. A policy which covers the "lag 0" and "lag 1" row of the jth column of L. An insured in his second year of the claims-made insurance program would purchase this coverage.

Occurrence policy. A policy which covers claims arising from accidents occurring during the policy period. Such a policy would cover a Northwest-to-Southeast diagonal of matrix L. This is the traditional form of coverage in most liability lines.

Tail policy. A policy written for an insured who leaves the claims-made program. It covers losses whose accident date lies in the period during which the claims-made coverage was in force, and whose reported date is after the insured's last claims-made policy expired.

Retroactive date. The earliest accident date for which coverage is provided under a claims-made policy. Normally this would be the date on which an insured's first claims-made policy commences. Only claims with accident date subsequent to the retroactive date are covered by any subsequent claims-made or tail policy.

We will illustrate these coverages with the example of a hypothetical insured who begins practice at the start of year 1 and retires four years later. He buys an occurrence policy to cover his first year of practice, then switches to the claims-made program. He purchases first-year, second-year, and third-year claims-made policies for years 2, 3, 4 respectively. At the end of year 4, he retires and purchases a tail policy. The policies cover the Figure 1 loss matrix in the manner shown in Figure 2.

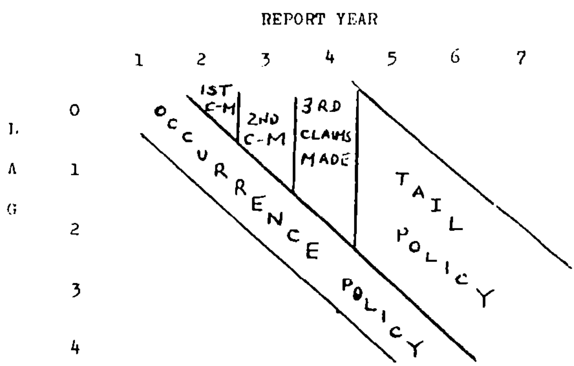

FIGURE 2

An important point to note is that the coverage above is equivalent to the coverage provided under 4 occurrence policies, as shown in Figure 3 below:

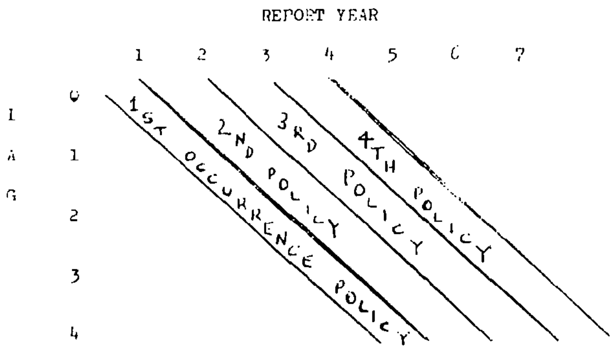

FIGURE 3

Although the coverage to the insured is the same under the claims-made system as under the traditional occurrence system, there is an important difference in the timing of the premium determination. To illustrate, the losses for report year 4 lag 2 are covered by the occurrence policy written at (and priced no later than) the beginning of year 2. For our claims-made insured, these losses would be covered by the third-year claims-made policy written at the beginning of year 4. The claims-made system allowed the insurer an extra two years to price this "lag 2" loss element.

IV. Claims-Made Ratemaking Principles

As noted previously, the major differences between the claims-made and the occurrence policy lies not in the coverage provided, but in the timing of pricing decisions affecting that coverage. Under claims-made we are always pricing next year's claims. Under occurrence pricing we must take into account claims to be reported many years in the future. The accuracy of any forecast is a direct function of how far beyond the data the projection is to be carried. A series of simple examples will illustrate this principle as it applies to claims-made and occurrence policies.

Let reported year J=0 represent the last year of history, J=1 represent the claims-made year we are pricing, and let (0,1), (1,2), (2,3), (3,4), (4,5) represent the components of the occurrence year we are simultaneously pricing.* In terms of the diagram we have:

	Future (Projections)				
History	1	2	3	4	5
$L_{0,0}$	$L_{0,1}$				
$L_{1,0}$	$L_{1,1}$	$L_{1,2}$			
$L_{2,0}$	$L_{2,1}$		$L_{2,3}$		
$L_{3,0}$	$L_{3,1}$			$L_{3,4}$	
$L_{4,0}$	$L_{4,1}$				$L_{4,5}$
	(C-M)				(Occ)

FIGURE 4

* Actually it might be more accurate to state that reported year J=-1 is the last year of history, since 1) there is typically a six month lag between the end of the experience period and the effective date of a filing, and 2) an additional six months between the effective date and the average date when the new price is in effect. Since this would have the same effect on pricing both occurrence and claims-made policies, it will be ignored for the sake of simplicity here.

Now let us assume $L_{0,0} = L_{1,0} = L_{2,0} = L_{3,0} = L_{4,0} = \$200$; that is, losses reported in the last year were produced in equal proportions from occurrences in the last five years. Also let's say we forecast that losses will increase at a rate of $20 per year for each lag. Then our diagram becomes:

	Future (Projections)				
History	1	2	3	4	5
$200	$220				
$200	$220	$240			
$200	$220		$260		
$200	$220			$280	
$200	$220				$300
$1,000	$1,100 (C-M)				$1,300 (Occ)

FIGURE 5

It is immediately apparent that next year's occurrence policy is more expensive than next year's claims-made policy, in this case by $1,300 - $1,100 = $200. The First Principle of Claims-Made Ratemaking states: <u>A claims-made policy should always cost less than an occurrence policy, as long as claim costs are increasing</u>. Furthermore, the greater the trend, the greater the difference will be. For example, suppose we underestimated inflation by $10 per year per lag. Then our diagram would become:

	Future (Assuming Change in Trend)				
History	1	2	3	4	5
$200	$230				
$200	$230	$260			
$200	$230		$290		
$200	$230			$320	
$200	$230				$350
$1,000	$1,150 (C-M)				$1,450 (Occ)

FIGURE 6

Now the difference is \$1,450 - \$1,150 = \$300. But consider what happened to the relative rate levels. The claims-made rate level proved to be inadequate by \$1,150 - \$1,100 = \$50 or 4.5%. The occurrence rate level turned out to be inadequate by \$1,450 - \$1,300 = \$150 or 11.5%.

The result is obvious when you think about it. But it is fundamental to understanding the difference between claims-made and occurrence ratemaking. In fact, it deserves restating as the Second Principle of Claims-Made Ratemaking: <u>Whenever there is a sudden, unpredictable change in the underlying trend, claims-made policies priced on the basis of the prior trend will be closer to the correct price than occurrence policies priced in the same way</u>. Stated another way, the confidence interval about the projected losses for a claims-made policy is narrower than for an occurrence policy priced at the same time.

In addition to a sudden unexpected change in the underlying trend there is another type of change that plagues actuaries pricing long-tailed lines: a sudden unexpected shift in the reporting pattern. Let us see how this would affect pricing accuracy under the two types of policies. First, recall our projections by referring back to Figure 5.

Now let's see what happens if we have a \$20/per year/per lag shift toward later reportings; that is, \$20 of what would normally be reported in lag 0 is not reported until lag 1, \$20 from lag 1 moves to lag 2, etc. (Note that only the first and last lags are affected since the others have the same dollars shifting in and out, and the same total dollars are reported.) Then our example looks like this:

History	Future (Assuming Change in Reporting Pattern) 1	2	3	4	5
$200	$200	$200	$200	$200	$200
$200	$220	$240			
$200	$220		$260		
$200	$220			$280	
$200	$240	$280	$320	$360	$400
$1,000	$1,100				$1,380

FIGURE 7

Under these circumstances, the mature claims-made policy is still priced correctly (as we would expect since the total dollars reported is unchanged), although a first year claims-made policy would have been slightly over-priced. But the occurrence policy is under-priced by $1,380 - $1,300 = $80, or 6.2%. The Third Principle of Claims-Made Ratemaking states: Whenever there is a sudden unexpected shift in the reporting pattern, the cost of mature claims-made coverage will be affected very little if at all relative to occurrence coverage.

If we put the two types of errors together, the result is even more dramatic.

History	Future (Assuming Change in Trend & Shift in Reporting Pattern) 1	2	3	4	5
$200	$210	$220	$230	$240	$250
$200	$230	$260			
$200	$230		$290		
$200	$230			$320	
$200	$250				$400
$1,000	$1,150				$1,530

FIGURE 8

The claims-made policy is under-priced by $50 or 4.5% as before. But the occurrence policy is under-priced by $1,530 - $1,300 = $230 or 17.7%. By now, it should be obvious that claims-made rates are both more accurate (because of a shorter forecast period) and more responsive to changing conditions (because external changes affect losses as they are reported). Two other points deserve emphasis. First, claims-made policies incur no liability for IBNR claims so the risk of reserve inadequacy is greatly reduced. (Principle Number Four). For example, a company writing occurrence policies for five years at the end of the period marked "history" in Figure 5 would carry an IBNR reserve of 4 x $220 + 3 x $240 + 2 x $260 + 1 x $280 = $2,400. A company writing claims-made for the same period would have an IBNR reserve of $0. The occurrence IBNR reserve needed under varying assumptions would be $2,600 (Figures 6 and 7) or $2,800 (Figure 8), so either of the two unfavorable developments would result in an IBNR reserve inadequacy of 8.33% for the occurrence policy. The IBNR needed for the claims-made policy is always 0.

The final point follows directly from the above. Because there is no need for IBNR, the time lapse between the collection of premiums and the payment of claims is greatly reduced. Consequently, the investment income earned from claims-made policies is substantially less than under occurrence policies. (Principle Number 5). The longer the reporting lag, or the shorter the settlement lag, the greater the difference will be.* The point is, as we reduce risk of inadequate

* Algebraically, the reduction may be expressed as R/(R+S+1/2), where R is the mean reporting lag in years, S is the mean settlement lag and 1/2 represents the 1/2 year lag between payment of premium and the occurrence of a claim on average. Of course integrals rather than averages should really be used, but this approach produces a reasonably accurate answer given the uncertainties about R and S.

rates and insufficient reserves by switching to claims-made coverage, we pay for it with reduced investment income. On the other hand, the reduced risk should allow us to write more policies for a given amount of capacity, thus making up for the reduction in expected profitability per policy.

Summarizing the five Principles of Claims-Made Ratemaking discussed in this section:

1) A claims-made policy should always cost less than an occurrence policy, as long as claim costs are increasing.
2) Whenever there is a sudden, unpredictable change in the underlying trend, claims-made policies priced on the basis of the prior trend will be closer to the correct price than occurrence policies priced in the same way.
3) Whenever there is a sudden unexpected shift in the reporting pattern, the cost of mature claims-made coverage will be affected very little if at all relative to occurrence coverage.
4) Claims-made policies incur no liability for IBNR claims so the risk of reserve inadequacy is greatly reduced.
5) The investment income earned from claims-made policies is substantially less than under occurrence policies.

Now that the advantages of the claims-made approach are apparent, we will discuss how pure premium data for claims-made pricing is compiled, even where claims-made coverage has never been written.

V. Historical Pure Premium Collection

As explained above, our approach to ratemaking requires that we compute historical pure premiums by reported period and lag. To do this we collect the loss data and the exposure data and form the quotient.

Collection of Losses. It is easy enough to categorize losses by reported period and lag using the coded reported date and accident date. Since we use pure premiums on an "ultimate value" basis, development factors are applied to the most recent loss valuations. The development factors used in our approach have these features:

1. They are a function of report period only.
2. The development factors are applied only to the case reserve portion of the loss, not to the paid component.
3. The factors are determined through a "backward recursive" formula, described in Appendix A.

Because the factors develop reported period losses to ultimate value, they provide for anticipated shortages or redundancies in case reserves, but they do not provide for IBNR (Incurred But Not Reported) losses. There is no need for IBNR losses in the claims-made ratemaking process, since the primary focus is on losses by reported period.

Collection of Exposure. Determining the number of exposures for each reported period and lag is more difficult than tabulating the losses. This is especially the case when the data base consists of a mixture of occurrence and claims-made policies. The best way to see the difficulty is to look at hypothetical premium transactions and see how much each transaction contributes "earned exposure" to each report period by lag combination. Keep in mind that the goal in processing exposure data is to provide a suitable denominator for the pure premium calculations.

In the examples we will use the report period by lag matrix indexing system developed in Section III. We will call $E_{i,j}$ the exposure to loss reported in period j, with accident year lag i. (See Figure 9 below)

REPORT YEAR (J)

		1	2	3	4	5	6	7
	0	$E_{0,1}$	$E_{0,2}$	$E_{0,3}$	$E_{0,4}$	$E_{0,5}$	$E_{0,6}$	$E_{0,7}$
L	1	$E_{1,1}$	$E_{1,2}$	$E_{1,3}$	$E_{1,4}$	$E_{1,5}$	$E_{1,6}$	$E_{1,7}$
A G	2	$E_{2,1}$	$E_{2,2}$	$E_{2,3}$	$E_{2,4}$	$E_{2,5}$	$E_{2,6}$	$E_{2,7}$
(I)	3	$E_{3,1}$	$E_{3,2}$	$E_{3,3}$	$E_{3,4}$	$E_{3,5}$	$E_{3,6}$	$E_{3,7}$
	4	$E_{4,1}$	$E_{4,2}$	$E_{4,3}$	$E_{4,4}$	$E_{4,5}$	$E_{4,6}$	$E_{4,7}$

FIGURE 9

Example 1: A "mature" claims-made policy on one insured written at the beginning of year j. This contributes one exposure to all matrix elements with report year = j (i.e., the jth column of the matrix). This is because the policy covers all losses reported in year j, regardless of the lag.

Example 2: An occurrence policy written at the beginning of year j. This policy contributes one exposure to the following matrix elements:

$E_{i,j+i}$ for i = 0, 1, 2, ...

Example 3: A mature claims-made policy written 1/3 of the way through year j. This policy contributes:

2/3 exposure to $E_{i,j}$ for i = 0, 1, 2, ...

1/3 exposure to $E_{i,j+1}$ for i = 0, 1, 2, ...

This is the familiar "uniform earning" which also characterizes occurrence policies and most other policies in property and casualty insurance.

Example 4: A second-year claims-made policy written at the beginning of year j. This policy generates one exposure for only lag 0 and 1 portions of reported year j, (i.e., $E_{0,j}$ and $E_{1,j}$).

Example 5: A first-year claims-made policy written 1/3 of the way (i.e., May 1) through year 1. Before jumping to any conclusions about the amount of exposure look at Figure 10 on the next page. In Figure 10, we introduce the term "difference" as the difference between reported date and accident date. (This is in contrast to "lag", which is the difference between reported year and accident year.)

In Figure 10, the solid lines delineate regions represented by report year-lag combinations. These are parallelograms except for the "lag 0" region, which is a right triangle. The shaded area is the triangular region covered by the policy of Example 5. We can see that the policy covers the following

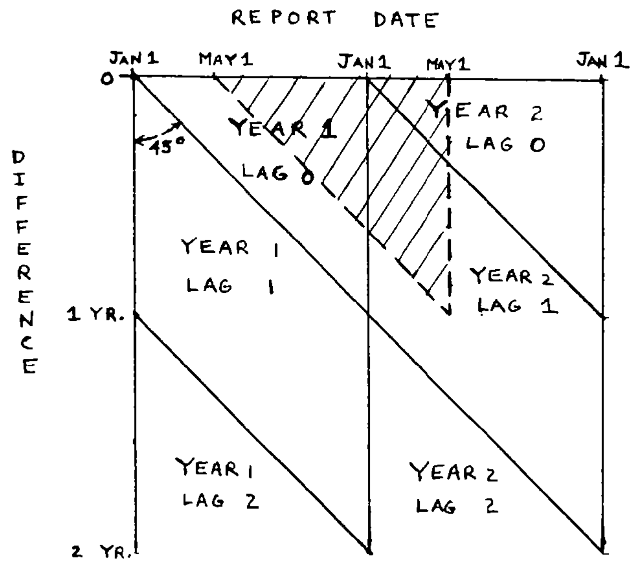

FIGURE 10

THIS FIGURE ILLUSTRATES THE COVERAGE OF A FIRST-YEAR CLAIMS-MADE POLICY WRITTEN ON MAY 1 OF YEAR 1.

SHADED AREA REPRESENTS THE COVERAGE OF THIS POLICY.

SOLID LINES REPRESENT BOUNDARIES OF "REPORTED YEAR - LAG" CELLS.

"DIFFERENCE" (VERTICAL AXIS) REPRESENTS THE TIME DIFFERENCE BETWEEN DATE OF ACCIDENT AND DATE OF REPORTING.

"LAG" IS REPORTED YEAR MINUS ACCIDENT YEAR.

proportion of these regions:

$2/3 \times 2/3 = 4/9$ of report year 1, lag 0;

$1/3 \times 1/3 = 1/9$ of report year 2, lag 0;

$1/3 \times 2/3 = 2/9$ of report year 2, lag 1.

These proportions are the earned exposure contributions to $E_{0,1}$, $E_{0,2}$, and $E_{1,2}$ respectively.

We can see, from Example 5, that the determination of exposure by report year and lag can be a fairly complex problem. This is especially so for "non-mature" claims-made policies and tail policies. However, the graphical technique used in Figure 10 is general and can be applied to any type of policy.

Before going on to a discussion of pure premium projection we will make some observations about the earned exposure calculations. We have concentrated on the general theory of how to make the exposure calculation, given the "maturity" of a claims-made policy transaction and the actual commencement date. In reality one may have to make these calculations using only summarized written premiums and earned premiums by type of policy and by time period (rather than using detailed transaction data). If this is this case, accuracy is greatly improved if the time periods are as fine as possible. Another problem which arises is the actual determination of the "maturity" of a claims-made policy. This requires the coding of the date on which an insured first purchases claims-made coverage (the "retroactive date"). This date is crucial and must be accurately recorded.

A simplification was made in the exposure calculation argument based on Figure 10. Using the proportional areas of the figure was equivalent to assuming a "uniform claim potential" within each year-lag parallelogram or triangle in the figure.

Summary. In this section we have attempted to describe the process by which historical pure premiums (quotient of loss and exposure) by reported period and lag can be computed. The tabulation of loss is straightforward, since insurance loss data contains reported date and accident date. The tabulation of exposure is much more complex, since different claims-made and *occurrence policies contribute to different report period-lag* exposure "cells".

Once these historical pure premiums are computed, the actuary can begin the projection of future pure premiums.

VI. Future Pure Premium Projection

Once historic pure premiums have been calculated, future pure premium projection proceeds in two steps. First, the future "mature" claims-made pure premium is determined. Second, the total pure premium is distributed back to lags and hence to policies at different levels of maturity. We will discuss each of these steps in turn.

In its simplest form, mature pure premium projection consists of nothing more than polynomial or exponential regression, using time as the independent variable.* This is suitable for countrywide data and perhaps for a few high volume states. It is not suitable for most states, however, as random fluctuation and the distortion of changes in legal or social climate can produce very poor fits and unreliable estimates. The actuary must be careful to check for this every time he does his analysis, even for the largest states, since a sudden surge or drop in claims being reported can occur in a single state at any time, destroying a previously stable trend. Normally, these events average out when countrywide data is used, although the evidence of recent years indicates that the experience in all states is becoming more highly correlated with one another.

If the actuary decides a particular state's trend is <u>not</u> sufficiently stable or reliable to use for projecting its future mature pure premiums, he may project them through a two-stage process. First,

---

* Other curves, such as log or power functions, have been proposed as alternatives. Unfortunately, the results derived from fitting these functions are highly dependent on the time index chosen, since the regression is done against the log of the index rather than the index itself.

the actuary generates countrywide fitted pure premium using polynomial regression as described above. Second, he applies linear regression or "regression through the origin"*, with state pure premium as the dependent variable and countrywide fitted pure premium as the independent variable. This approach assumes that, in the long run, a consistent relationship exists between state and countrywide pure premium. In other words, it assumes that the state will have the same percentage change from one year to the next that all states do, while using the individual state's own experience to determine its "relativity" to the countrywide rate. Linear regression is similar, but adds a constant term which allows partial recognition of the state's own apparent trend. Of course, if linear regression is used in both stages one and two, the result is the same as using linear regression against time directly.

Two points merit emphasis about the above procedure. First, as always, it is the actuary's task to strike the delicate balance between stability and responsiveness. This is done directly through his choice of a projection method, rather than indirectly through the choice of a credibility formula. The question to be kept constantly in mind is : How reliable is this data as an indicator of the claim process in this state? Fortunately, a wealth of information about the quality of the regression is available to help answer that question. Second, all projections are done on the experience itself. No "outside" frequency or severity trend information is superimposed on the data, thus avoiding the problem of explaining two or three sets of data and reconciling them with one another. There is no reason why this procedure should be limited to claims-made; its advantages apply equally well to any type of coverage.

---

* See Appendix B for a technical description.

Once the future mature pure premium has been determined, the problem of distributing it to lags may be approached in several ways. The original approach taken was to regress individual lag pure premiums ("a row" in the pure premium matrix) against time in the same way we regressed the total of all lags. As might be imagined from the above discussion, this method is highly sensitive, so much so that some lags will shoot upward at high rates while others are trending downward, in some cases even projecting negative values. Even if such trends were accurate reflections of what was going on in the real world, they would be undesirable for projecting pure premiums and rates since a smooth transition between rates for policies at succeeding maturities is very important in helping insureds understand the steps in claims-made coverage. A less-sensitive method was clearly needed. One simple approach we tried was to calculate the historic proportion for each lag, as follows:

$$(1) \quad b_i = \sum_j X_{i,j} \Big/ \sum_{i,j} X_{i,j},$$

where $X_{i,j}$ is the pure premium for report period j, accident period lag i.

The problem with this approach is that it does not recognize trends in relative pure premiums between lags. It was decided that a weighted proportion - with greater weight going to the larger, and presumably more recent, observations - would be a better representation. Surprisingly, it turned out that regression through the origin was the answer again. In this case, the historic proportion for lag i turns out to be:

$$(2) \quad b_i = \sum_j X_{i,j} \hat{X}_j \Big/ \sum_j (\hat{X}_j)^2,$$

where $\hat{X}_j$ is the fitted report period total pure premium.

Let's see how this compares to the historic proportion calculated above. Note that:

$$X_j \equiv \sum_i X_{i,j} \quad \text{and} \quad \sum_j \hat{X}_j = \sum_j X_j = \sum_{i,j} X_{i,j}.^{*}$$

Therefore, (1) can be re-written as follows:

$$b_i = \sum_j X_{i,j} \Big/ \sum_{i,j} X_{i,j}$$

$$b_i = \sum_j X_{i,j} \Big/ \sum_j X_j$$

$$b_i = \sum_j X_{i,j} \Big/ \sum_j \hat{X}_i$$

Thus, we see the difference between (1) and (2) is simply that the $\hat{X}_j$'s are used as weighting factors to place greater weight on larger pure premiums. It is important to note that $\sum_i b_i = 1$ since the b_i's are the fractions of the total pure premium associated with each lag.

Summing up, the projection of pure premiums may be viewed as a two-step process. First, project the total ("mature") pure premium ignoring lags. Second, distribute the total pure premium back to lags. Several methods for carrying out each step have been suggested in this section.

There is no one "right" method for all circumstances. In fact, once the data is collected into a historic pure premium matrix, the possibilities for projection methods are limited only by the actuary's imagination and the flexibility of his statistical software package. For example, both econometrics and time series analysis merit exploration since the pure premium data by reported year seems to indicate distinct cycles about the long term trend line roughly corresponding

---

* This is true, if and only if, the $\hat{X}_j$'s were arrived at through linear regression. For regression through the origin, the residuals do not sum to zero.

to the economic cycle. This is logical since the incidence of malpractice should vary only with the utilization of medical services, while the reporting of a claim has a lot to do with how the claimant feels about his own economic situation. In any case, we suggest a "simulation" approach be used as a means of sensitivity analysis.

At the St. Paul, we divide states into four categories:

A - States, with highly stable patterns, where we use regression on their total pure premiums to determine their own trend;

B - States, where we use regression through the origin for both the total and lag pure premium;

C - States, where we use regression through the origin for the total pure premium but use the countrywide lag pattern and;

D - States, with very thin data, where we use a judgmental relativity to the countrywide pure premium.

As noted earlier, the categorization of each state must be reviewed each year to make sure changes in claim environment have not materially altered the data's reliability. Sensitivity analysis provides valuable insights in this process as well.

The ability to project pure premiums allows the actuary to determine more than prices for claims-made policies. Specifically, it can also be used to price occurrence policies, and to predict IBNR emergence and reserve adequacy. We will have more to say on this in Section VIII. But first, we will briefly discuss some special features of The St. Paul filings which distinguish them from those of other claims-made writers.

VII. Special Features of St. Paul Claims-Made Filings

Not all claims-made policies are alike in coverage. The St. Paul claims-made form contained several unique (at the time) coverage features which presented the actuaries with special pricing problems. Also, we chose several pricing techniques which were not traditional to facilitate the process of claims-made ratemaking from occurrence data. We will briefly discuss each of these special features in turn.

Several unique features of The St. Paul filings have already been discussed in Section III. One of these was the retroactive date; i.e., the earliest accident date for which coverage is provided under the claims-made policy. Previous claims-made or "discovery" policies treated all insureds the same, even if they had no prior exposure (e.g., just coming out of medical school) or if they were previously covered by an occurrence policy.

Another concept mentioned earlier was "tail coverage". When we introduced claims-made we felt it was the wave of the future. Someday all insureds would be using it and insureds would move from one carrier to another, carrying their retroactive dates with them. However, we recognized that this might not occur for many years and decided that we would have to offer our claims-made insureds guaranteed coverage for the "tail"; i.e., for claims which occurred while the insured was covered by claims-made but were not reported until after the last claims-made policy had expired. This was considered a rather dangerous step, since it in effect gave the insured the right to convert his coverage from claims-made to occurrence at any time. Were we leaving ourselves open to the same pricing problems we had had under occurrence? We argued that the risk could be greatly reduced by selling the tail coverage in three annual installments, or reporting endorsements, reserving the right to price them one at a time. The first reporting endorsement would be just like the insured's next claims-made policy, providing coverage for claims reported during that year only, except that accidents occurring in that year would be excluded. The second reporting endorsement would be similar, except that accidents occurring in the two year period after expiration of the last claims-made policy would not be covered. Only the third reporting endorsement would provide the kind of perpetual coverage that the occurrence policy did, with similar pricing hazards. It was argued that that hazard was acceptable since 1) we would be pricing it at least three years later than we would have priced the comparable coverage under an occurrence policy; 2) each insured would buy this "occurrence" coverage only once instead of every year, while the great majority of insureds were still buying claims-made; and 3) by the time we reach the third reporting endorsements the proportion of claims remaining to be reported is fairly small, so even a large percentage error in the rate would not result in a large dollar loss.

As we discussed the claims-made concept with our insureds, it became apparent that the three-pay reporting endorsement concept was acceptable to the majority but could work a real hardship for a few. So we added an additional option: We would sell a single-payment reporting endorsement to any insured terminating coverage due to death, disability or retirement.

The pricing of reporting endorsements - both three-pay and single-pay - poses no special problems. It merely requires trending the projected pure premiums further into the future. In fact, since the policy is essentially selling IBNR coverage, the pricing of reporting endorsements is equivalent to the determination of IBNR reserves, which will be discussed in Section VIII. Before proceeding to that discussion, however, we will briefly mention three special features of The St. Paul rate filings not directly linked to the coverage provided.

The St. Paul's claims-made policy is on an annual basis. But semi-annual reporting periods and lags were used in calculating and projecting pure premiums. The advantages of using this approach are twofold:

1. Less distortion calculating the earned exposures by "cell" (See Section IV), and
2. More data points for use in the regression.

Underlying the whole idea of pricing claims-made coverage from occurrence data is the implied assumption that the same body of claims would be reported at the same time under either policy. The more we thought about it, the less reasonable this assumption seemed. We decided that two changes were likely to occur at the transition point, both due to insureds understanding that coverage for a particular claim would not commence until the claim had been reported.

First, we assumed that, on average, claims would be reported sooner. Specifically, we assumed a two-month "shift" forward in claim reporting; algebraically,

$$L'_{0,1} = L_{0,1} + 1/6\ L_{1,2}$$

$$L'_{1,2} = L_{1,2} - 1/6\ L_{1,2} + 1/6\ L_{2,3} \text{ etc.}$$

Second, we assumed that there would be some additional reporting of incidents that would never have come in under the occurrence policy. Few, if any, of these incidents would result in loss payment, but they would require investigation and hence loss expense payment. Specifically, we assumed 5% additional claim dollars would be reported, and that all of this additional activity would come in at Lag 0. Algebraically,

$$L''_{0,1} = L'_{0,1} + .05 \times \quad L_{i,1}$$

We can never know what would have been reported had we continued with the occurrence policy, so it is impossible to test whether or not the "shift" and the additional incident reporting actually occurred. Now that several years of claims-made experience is contained in the data base, the need for this special adjustment no longer exists and it has been dropped from the filing.

The final special feature of the St. Paul filings involves the treatment of company expense. It was obvious that the pure premium and hence the rate for a first-year claims-made policy would be much less than for a mature policy. It did not follow that all expenses would be proportionately lower. In fact, most company expenses are probably fixed: i.e., they do not vary with the size of the premium. This was recognized by splitting the expense dollar into two parts: fixed and variable. This affected not only the relativities between different policy maturities, but between different classes of risk as well: the higher the rate, the lower the expense ratio. Algebraically, the rate calculation changed from:

$$R = PP/(1-E-P)$$

to $$R = (PP + FE)/(1-VE-P)$$

where R is the rate, PP is the pure premium, P is the profit allowance and E = FE + VE is the expense, broken down into its fixed and variable components. The following example will illustrate how this early instance of "expense flattening" works.

	Pure Premium (Relativity)	Fixed Expenses (% of Rate)	Variable Expenses and Profit (25% of Rate)	(Relativity)
Class 1 Physician, First-Year Claims-Made Policy	$100 (1.00)	$35 (19.4%)	$45	$180 (1.00)
Class 1 Physician, Mature Claims-Made Policy	$500 (5.00)	$35 (4.9%)	$178	$713 (3.96)
Class 7 Surgeon, First-Year Claims-Made Policy	$800 (8.00)	$35 (3.1%)	$278	$1,113 (6.19)
Class 7 Surgeon, Mature Claims-Made Policy	$4,000 (40.00)	$35 (0.56%)	$1,345	$5,380 (29.89)

FIGURE 11

VIII. Other Uses of Analytical Tools Developed

The techniques discussed in this article were developed specifically to price the claims-made coverages. However, once we develop a method to project pure premiums by future report year and lag, we have developed a tool which we can use to solve a variety of insurance problems. We can price occurrence coverages by adding up the appropriate elements from Figure 1 of Section III. For example, the projected pure premium for an occurrence policy commencing at the beginning of year 3 is straightforward:

$$X_{0,3} + X_{1,4} + X_{2,5} + X_{3,6} + \ldots,$$

where $X_{i,j}$ is the projected pure premium for reported year j, lag i.

Another area where the methods have application is in loss reserve determination. The "pure IBNR" (Incurred But Not Reported loss) for a company writing occurrence policies falls out of the projected loss calculation. For example, the IBNR reserve at the end of year 2 is the following area from the Figure 12 loss matrix below:

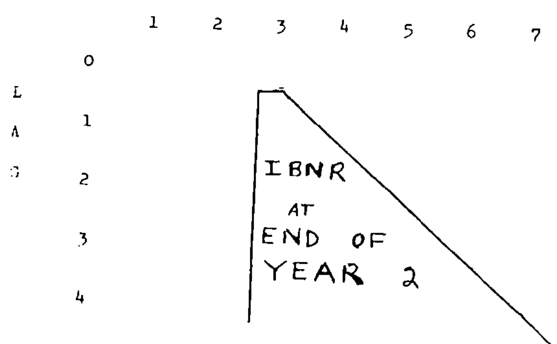

FIGURE 12

That is, the IBNR as of the end of year 2 on occurrence policies is the sum of all losses with reported year greater than 2 and accident year less than 3.

Determining "pure IBNR" is only half the problem of determining a total loss reserve. One must also project the additional development to be incurred on case reserves (reserves on losses already reported). This latter problem must be solved before we even begin to project the pure premiums in Section VI. It turns out that the case reserve development is easier to project once the loss data is collected into the report year by lag format of Figure 1. Appendix A discusses the precise method by which we project this case development.

Thus, the method of analyzing data which was developed to price claims-made policy gives us a convenient way of separating loss development into its two major components and projecting each separately:

Anticipated loss development = IBNR + Case Development.

Moreover, the method also projects an emergence pattern for the IBNR Loss.

IX. Summary

We began this paper by discussing the historical situation which led to the decision to write medical malpractice on a claims-made basis (Section II). Next we translated the problem of pricing the claims-made coverages into the problem of determining pure premiums by report period and accident period "lag" (Section III). Section IV presented a discussion of claims-made ratemaking principles.

Next followed a technical discussion of how to calculate historical pure premiums by report period and lag given insurance loss and exposure data (Section V). Once this is accomplished there are a variety of techniques available to project future pure premiums, and hence rates (Section VI). The St. Paul claims-made program and pricing techniques have several unique features (Section VII). Finally, the analytical tools used in claims-made ratemaking can also be applied to the general problem of IBNR determination for occurrence policies (Section VIII).

Appendix A

The Backward Recursive Reserve Development Method

In claims-made ratemaking the losses for each reported period must be developed to their ultimate value. We used a "Backward Recursive" reserve development method to accomplish this.

This method requires that loss data be available by reported period and "age". (Age 0 means the valuation as of the end of the reported period, age 1 is the valuation one period later, etc.) It also requires that the losses be separated into paid and case reserve components.

The "Backward Recursive" method calculates development factors which are applied to the reserve component of loss only. The determination of these factors proceeds in two steps:

1. "One-step" factors are calculated to develop losses as of each age to the next age. Two factors are calculated for each age k. P_k is the proportion of reserves of age k which will be paid by age k+1. R_k is the ratio of reserves at age k+1 to reserve at age k.
2. Ultimate factors are generated from the "one-step" factors. These factors apply to the reserves at age k to bring them to ultimate valuation.

The calculation of the "one-step" factors is a straightforward tabulation of the data. The factors are simply the following:

P_k = (Paid as of age k+1 - Paid as of age k)/(Age k reserve)

R_k = (Reserve as of age k+1)/(Reserve as of age k).

In order to generate the development factors to take reserves to their ultimate valuation, we need to assume an "ultimate" age, that is, an age N after which no further development occurs. The calculation proceeds in a "backward" fashion from the ultimate age in the following order:

$D_{N-1,N} = P_{N-1} + R_{N-1}$

$D_{N-2,N} = P_{N-2} + R_{N-2} \times D_{N-1,N}$

$D_{N-3,N} = P_{N-3} + R_{N-3} \times D_{N-2,N}$

*
*
*

$D_{0,N} = P_0 + R_0 \times D_{1,N}$

Here $D_{n,N}$ is the development factor which brings reserves at age n to their valuation at age N. Each equation above merely says that the ultimate development on reserves of age n is the sum of payments during the time period n+1 and the ultimate value of reserves of age n+1. The first equation says that if N is assumed to be the ultimate age then reserves of age N-1 are either paid within one time period or remain outstanding at age N.

An example will illustrate the principles. Suppose that the following "single period" development factors have been determined and that age 3 is "ultimate".

Age (k)	0	1	2
P_k	.300	.500	.400
R_k	.800	.500	.500

Recall the meaning of these factors. For example, of all the reserves at age 0, 30% will be paid by age 1 and 80% will remain as reserve. The compound factors to apply to reserves are calculated in "backward" fashion:

$$D_{2,3} = .400 + .500 = .900$$

$$D_{1,3} = .500 + .500 \times (.900) = .950$$

$$D_{0,3} = .300 + .800 \times (.950) = 1.060$$

Appendix B

Regression Through the Origin

"Regression through the Origin" is a least-squares statistical technique similar to linear regression, except that the line of best fit is constrained to pass through the "Origin". Like linear regression this technique uses a line of "best fit" to fit a set of observations of some dependent variable to a set of observations of an independent variable. Unlike linear regression the line of best fit is constrained so that, when the value of the independent variable is zero, the fitted value of the dependent variable is also zero. The criterion for "best fit" is the same for both techniques: the line of best fit is chosen to minimize the sum of the squares of the differences between the observed and fitted values of the dependent variable.

There are two situations where Regression through the Origin might be substituted for linear regression. The first is the case where, a priori, the value of the dependent variable must be zero when the independent variable is zero. The second situation is one where linear regression has been run, but the "intercept" is not significantly different from zero, so that it can be dropped without hurting the accuracy of the model.

An example of a problem where Regression through the Origin might be used is the problem of projecting one company's output as a function of an industry's total output, given historical annual figures. If the second variable takes on a value of zero, the first must also.

The structure and mechanics of Regression through the Origin are similar to linear regression. The modeler has at hand observed values of a dependent variable ($Y_1, Y_2, \ldots Y_N$) and observations of an independent variable ($X_{1,2}, \ldots X_N$). The task is to calculate a parameter b such that

$$\hat{Y} = b X$$

gives the expected value of the variable Y given any observed value of the independent variable X. (Recall that in linear regression we look for parameters a,b to use in an expression $\hat{Y} = a + bX$.)

The parameter b is chosen so that the sum of squares $\sum_{i=1}^{N} (Y_i - bX_i)^2$ is minimized. The formula for b is given by

$$b = \frac{\sum X_i Y_i}{\sum X_i^2}$$

The statistic b has the property that

$$\sum (Y_i - kX_i)^2 = \sum (Y_i - bX_i)^2 + (b-k)^2 \sum X_i^2$$

for any constant k. We can see that the last expression is minimized when k = b, so that b is optimal in the "least squares" sense. For a fuller discussion of the statistical properties of the model, consult John Neter and William Wasserman's Applied Linear Statistical Models (Richard D. Irwin, Inc., 1974).

Although it appears that the Regression through the Origin model is a special case of linear regression, the reverse is actually true! This is because any linear regression model $\hat{Y} = a + bX$ can be rewritten

$$(\hat{Y} - \bar{Y}) = b\,(X - \bar{X}),$$

where $\bar{X}$ and $\bar{Y}$ are the sample means of X and Y respectively.

With this formulation we can see that any linear regression model is a special case of Regression through the Origin in which each variable has zero mean.